

Westfälische Wilhelms-Universität Münster

Institut für Theoretische Physik

Constraining nuclear parton distributions with data from light and heavy meson production

Inaugural-Dissertation zur Erlangung des Doktorgrades
der Naturwissenschaften im Fachbereich Physik
der Mathematisch-Naturwissenschaftlichen Fakultät
der Westfälischen Wilhelms-Universität Münster

vorgelegt von:
Pit Duwentäster
aus Ahlen

Münster, Mai 2022

Dekan: Prof. Dr. Michael Rohlfing
Erster Gutachter: Prof. Dr. Michael Klasen
Zweiter Gutachter: PD Dr. Karol Kovařík
Tag der Disputation: _____
Tag der Promotion: _____

Zusammenfassung

Ein präzises Verständnis von Partonverteilungsfunktionen ist entscheidend in der Interpretation von Daten aufgenommen in hadronischen Kollisionsexperimenten wie beispielsweise am Relativistic Heavy Ion Collider (RHIC), dem Large Hadron Collider (LHC) und dem zukünftigen Electron-Ion Collider (EIC). Analysen von nuklearen Partonverteilungsfunktionen (PDFs) sind jedoch noch mit deutlich größeren Unsicherheiten belastet als jene von freien Protonen. Insbesondere die nuklearen PDFs von Gluonen in der Regionen von niedrigem Bjorken- x sind nur wenig eingeschränkt von den Daten aus tief inelastischer Streuung und dem Drell-Yan Prozess, die den Großteil der Daten in nuklearen PDF Analysen wie nCTEQ15 darstellen. Um dieses Problem anzugehen, präsentieren wir zwei neue globale Analysen nuklearer PDFs.

Die erste Analyse ergänzt Daten von inklusiver Pion, Kaon und η Meson Produktion und untersucht die Abhängigkeit von den nicht störungstheoretisch berechenbaren Fragmentationsfunktionen (FFs), welche die Hadronisierung des Endzustandes beschreiben. Diese FFs werden gegen Daten aus Proton-Proton Kollisionen getestet, um geeignete Wahlen für die Skalen und kinematischen Schnitte zu bestimmen. Des weiteren wird der Einfluss der FF Wahl vermindert, indem die Unsicherheiten der FFs zu denen der Daten addiert werden. Das Ergebnis ist ein neues Set an nuklearen PDFs, genannt nCTEQ15WZ+SIH, welches reduzierte Gluon-Unsicherheiten in der Region um $x \approx 0.2$ vorweist. Sowohl die neuen, als auch die alten Daten in der Analyse werden gut von den resultierenden PDFs beschrieben.

Der zweite Teil erweitert die nCTEQ15WZ+SIH Analyse weiter durch das Hinzufügen von Daten zur Produktion von Mesonen mit offenem schwerem Flavor oder von schweren Quarkonia. Um das Problem zu umgehen, dass es keine universell akzeptierte Theorie zum Produktionsmechanismus von Quarkonia gibt, verwenden wir einen neuen datengetriebenen Ansatz, der potentiell für viele verschiedene Prozesse eingesetzt werden kann. Dieser Ansatz erlaubt eine gute Beschreibung der Proton-Proton Daten, die benutzt werden, um die Theorie-Parameter zu bestimmen und ist kompatibel mit störungstheoretischen QCD Berechnungen für D^0 und J/ψ Produktion. Mit diesem Ansatz fügen wir der neuen globalen PDF Analyse eine weitreichende Menge an Daten zur Produktion schwerer Quarks hinzu, während wir die Unsicherheiten ähnlich wie in der vorherigen Analyse beachten. Das Ergebnis ist eine signifikante Reduktion der Gluon-Unsicherheiten über einen weiten kinematischen Bereich, der sich nach unten bis zu $x \approx 10^{-5}$ erstreckt.

Die Ergebnisse und Methoden dieser beiden Analysen sind ein wichtiger Schritt auf dem Weg zu zukünftigen nuklearen PDF Analysen mit beispielloser Genauigkeit.

Abstract

A precise understanding of parton distribution functions (PDFs) is crucial in the interpretation of data taken in hadronic collider experiments like the Relativistic Heavy Ion Collider (RHIC), the Large Hadron Collider (LHC) and the upcoming Electron-Ion Collider (EIC). However, analyses of nuclear parton distribution functions (nPDFs) still have significantly larger uncertainties than those of free protons. Particularly the nuclear gluon PDF and the low Bjorken- x region are only weakly constrained by the deep inelastic scattering and Drell-Yan process data, that make up the bulk of the data in analyses like nCTEQ15. To address this problem, we present two new global nPDF analyses.

The first analysis introduces data on inclusive pion, kaon and η meson production and investigates the dependence on the non-perturbative fragmentation functions (FFs), which describe the hadronization of the final state. These FFs are tested against proton-proton collision data to identify appropriate scale choices and kinematic cuts. The impact of the ambiguous FF choice on the nPDF fit is further mitigated by adding their uncertainties to the data. The result of the analyses is a new set of nuclear PDFs called nCTEQ15WZ+SIH with reduced gluon uncertainties, particularly in the region around $x \approx 0.2$. This new set of nPDFs yields a good description of all included data sets.

The second part extends the nCTEQ15WZ+SIH fit further by including data on open heavy flavored meson and heavy quarkonium production. To circumvent the problem that there is currently no universally accepted model for quarkonium production, we introduce a data-driven approach that can potentially describe a wide variety of processes. This approach delivers an accurate description of the proton-proton collision data, which is used to determine the theory parameters and is compatible with perturbative QCD calculations for D^0 and J/ψ production. Using this approach, a vast body of heavy quark production data is added to a new global nPDF fit, while accounting for the uncertainties in a similar manner as the previous analysis. This results in a significant reduction of the gluon PDF uncertainty over a wide kinematic region down to $x \approx 10^{-5}$.

Both the findings and the methods used by these two analyses represent important steps towards future nuclear PDF fits with unprecedented accuracy.

List of publications

The work presented in this thesis has contributed to a number of publications listed below. Additionally, results from this thesis have been presented in talks at various conferences. The main authors are underlined.

Articles

- **Impact of LHC vector boson production in heavy ion collisions on strange PDFs.**

A. Kusina, T. Ježo, D.B. Clark, P. Duwentäster, E. Godat, T.J. Hobbs, J. Kent, M. Klasen, K. Kovařík, F. Lyonnet, K.F. Muzakka, F.I. Olness, I. Schienbein, J.Y. Yu
Eur.Phys.J.C 80 (2020) 10, 968

- **Extending nuclear PDF analyses into the high- x , low- Q^2 region.**

E.P. Segarra, T. Ježo, A. Accardi, P. Duwentäster, O. Hen, T.J. Hobbs, C. Keppel, M. Klasen, K. Kovařík, A. Kusina, J.G. Morfín, K.F. Muzakka, F.I. Olness, I. Schienbein, J.Y. Yu
Phys.Rev.D 103 (2021) 11, 114015

- **Impact of inclusive hadron production data on nuclear gluon PDFs.**

P. Duwentäster, L.A. Husová, T. Ježo, M. Klasen, K. Kovařík, A. Kusina, K.F. Muzakka, F.I. Olness, I. Schienbein, J.Y. Yu
Phys.Rev.D 104 (2021) 9, 094005

- **Impact of heavy quark and quarkonium data on nuclear gluon PDFs.**

P. Duwentäster, T. Ježo, M. Klasen, K. Kovařík, A. Kusina, K.F. Muzakka, F.I. Olness, R. Ruiz, I. Schienbein, J.Y. Yu
Submitted to Phys.Rev.D

- **Compatibility of neutrino DIS data and its impact on nuclear parton distribution functions.**

K.F. Muzakka, P. Duwentäster, T.J. Hobbs, T. Ježo, M. Klasen, K. Kovařík, A. Kusina, J.G. Morfín, F.I. Olness, R. Ruiz, I. Schienbein, J.Y. Yu
Submitted to Phys.Rev.D

Proceedings

- **Constraining the nuclear gluon PDF with inclusive hadron production data.**

P. Duwentäster, L.A. Husová, T. Ježo, M. Klasen, K. Kovařík, A. Kusina, K.F. Muza-

kka, F.I. Olness, I. Schienbein, J.Y. Yu
Published in SciPost Physics Proceedings DIS2021.

- **Impact of W and Z Production Data and Compatibility of Neutrino DIS Data in Nuclear Parton Distribution Functions.**

K.F. Muzakka, P. Duwentäster, T.J. Hobbs, T. Ježo, M. Klasen, K. Kovařík, A. Kusina, J.G. Morfín, F.I. Olness, R. Ruiz, I. Schienbein, J.Y. Yu
Published in SciPost Physics Proceedings DIS2021.

Talks given

- **Impact of inclusive hadron production data on nuclear gluon PDFs**
XXVIII International Workshop on Deep-Inelastic Scattering and Related Subjects (2021), online
- **Impact of Fragmentation Function choice on nCTEQ nuclear PDFs**
Fragmentation Functions and Nucleon Structure (2021), online
- **Impact of heavy quarkonium data from the LHC on nuclear PDFs**
Physics Opportunities with Heavy Quarkonia at the EIC (2021), online
- **Constraining the nuclear gluon PDF with quarkonium production data**
Quarkonia As Tools (2022) in Aussois, France
- **Update on nCTEQ PDFs**
XXIX International Workshop on Deep-Inelastic Scattering and Related Subjects (2022) in Santiago de Compostela, Spain

Acknowledgements

First of all, I would like to thank Prof. Michael Klasen for giving me the opportunity to work on this thesis in his group, as well as for the supervision and support during these years.

I am also grateful towards my secondary supervisor Dr. Karol Kovařík, who introduced me to the topic of parton distribution functions and always provided valuable feedback on my work.

Furthermore, I want to thank all other members of the institute for a pleasant office atmosphere and many helpful discussions.

Further thanks go towards the entire nCTEQ collaboration for all the valuable advice, explanations of many topics related to my work and generally very enjoyable meetings.

Finally, I want to express my gratitude towards the friends and family, who accompanied me throughout the past 8 years of university. In particular, I want to thank Jens Bickmann for countless hours of (un)productive discussions and for proofreading this work.

Contents

I	Introduction and theoretical framework	1
1	The parton model	4
1.1	Deep inelastic scattering	4
1.2	The structure of nucleons	6
1.3	Parton distribution functions	7
1.3.1	Nucleon structure functions in the parton model	9
1.3.2	Sum rules	11
1.4	DGLAP evolution	13
1.5	Nuclear modification of PDFs	17
1.5.1	Shadowing	18
1.5.2	Antishadowing	19
1.5.3	EMC effect	21
1.5.4	Fermi motion	22
1.5.5	Short range correlations	23
1.5.6	Target mass corrections	23
1.6	Statistical hadronization model	24
1.6.1	Statistical hadronization of the QGP	24
1.6.2	SHM with charmed hadrons	26
2	Data modelling and the nCTEQ framework	30
2.1	PDF parameterization	30
2.2	The χ^2 cost function	32
2.3	Minimization	34
2.4	Estimation of uncertainties	38
2.4.1	The Hessian method	40
2.4.2	Choosing the tolerance T	42
2.4.3	Treatment of nuisance parameters	44
2.4.4	Monte Carlo replicas	45
2.5	Optimizations	45
2.5.1	K-factors	46
2.5.2	Gridding	46

2.6	Correlation between data sets and PDFs	47
2.6.1	Cosine of correlation angle	47
2.6.2	Effective χ^2 difference	47
2.6.3	Parameter scans	48
2.7	Bayesian reweighting	48
II	Nuclear PDF analyses with light and heavy meson production data	51
1	Overview of relevant nuclear PDF fits	53
1.1	Historical nuclear PDF analyses	53
1.2	Nuclear PDFs in the nCTEQ framework	53
1.3	Nuclear PDF analyses by other groups	55
2	Single inclusive hadron production	57
2.1	Theoretical framework	58
2.1.1	Calculating the cross section in practice	59
2.1.2	Overview of available fragmentation functions	60
2.2	Available data	62
2.3	Fragmentation function and scale dependence	64
2.3.1	Comparing fragmentation functions in proton-proton collisions .	64
2.3.2	Comparing different factorization and renormalization scales .	65
2.3.3	Comparing fragmentation functions in dAu and pPb collisions .	66
2.3.4	Fragmentation kinematics	70
2.4	Fits with SIH data	73
2.4.1	Main fits	74
2.4.2	Alternative fits	81
2.5	Conclusions	82
3	Heavy quark production	85
3.1	Theoretical framework	86
3.1.1	Different schemes for the production of open heavy flavored mesons	87
3.1.2	Perturbative models for the production of quarkonia	91
3.1.3	Empirical determination of effective scattering-matrix elements .	93
3.2	Proton-proton baseline	95
3.2.1	Kinematic cuts and excluded data	95
3.2.2	Baseline fit	97
3.2.3	Alternative parameterizations of the effective scattering matrix element	98
3.2.4	Comparison with prompt D^0 production in the GMVFNS	99
3.2.5	Comparison with J/ψ production in NRQCD	99
3.3	Fits with heavy quark data	101
3.3.1	Data selection	102

3.3.2	The nCTEQ15HQ nuclear PDFs	102
3.3.3	Fit quality	105
3.4	Conclusions	111
4	Impact on other observables	113
4.1	Rapidity dependence of heavy quark production	113
4.2	Cold nuclear effects in lead-lead collisions	113
4.3	Impact on single inclusive hadron and vector boson production	114
4.4	Prompt photon production	117
4.5	Comparison with other nuclear PDF fits	118
III Conclusions and Outlook		123
1	Summary	123
2	Prospects for future studies	125
IV Appendix		129
1	Feynman rules	129
1.1	Quantum electrodynamics	129
1.2	Quantum chromodynamics	130
2	Definition of kinematic variables	132
3	Additional tables	133
4	Additional figures	139
4.1	Single inclusive hadron production	139
4.2	Heavy quark production	140
4.2.1	Heavy quark production in proton-proton collisions	141
4.2.2	Heavy quark production in proton-lead collisions	147
Bibliography		153

I. Introduction and theoretical framework

The history of scattering experiments began more than a century ago with the famous gold foil experiment by Geiger, Marsden and Rutherford [1], in which the existence of a nucleus in the center of atoms was shown. Scattering experiments have remained one of the fundamental pillars of particle physics ever since and many of the standard model particles, like the top quark [2, 3] or the Higgs boson [4, 5], were confirmed to exist in such experiments.

While the gold foil experiment could only give a vague estimate of the size of the atomic nucleus without any indication of deeper substructure, further research found that it is composed of protons and neutrons, which are themselves made up of quarks held together by gluons. The structure of nuclei in terms of these quarks and gluons, collectively called *partons*, is the focus of this thesis.

Parton distribution functions (PDFs) describe the structure of hadrons in terms of the momentum fraction x carried by a specific parton. The PDFs are not just interesting because of our inherent curiosity about the nuclear structure, but they are also a required ingredient in any calculation for cross sections of processes with hadronic initial states, *i.e.*, those measured at the Relativistic Heavy Ion Collider (RHIC), the Large Hadron Collider (LHC) or the planned Electron-Ion Collider (EIC) [6, 7]. However, while these cross sections can be calculated in perturbation theory at the partonic level, the PDFs themselves are inaccessible by perturbative means and therefore need to be determined empirically in a procedure known as global QCD analysis [8]. The combination of perturbative calculation of the partonic subprocesses with vast bodies of data through sophisticated statistical tools creates a vast field of research with endless room for improvement.

In this thesis, we are interested in obtaining a more precise understanding of how the nucleon number A modifies the PDFs beyond just summing the proton and neutron PDFs. Such nuclear parton distribution functions (nPDFs) have been studied in many previous analyses by various groups [9–16], but the uncertainties of these nuclear PDFs are still far larger than those of free proton PDFs [15–29]. Nuclear PDF analyses currently lag behind those of protons in many regards, including the concept of generalized PDFs [30], *e.g.*, PDFs accounting for transverse structure of hadrons [31], polarized PDFs [32, 33] and double scattering PDFs [34, 35]. There are also QED enhanced proton PDFs with photon distributions [36, 37] and calculations have long been performed at next-to-next-to-leading order (NNLO) accuracy [38]. Finally, the proton PDF analyses include a far

greater number of physical processes in their fits, which gives them tighter constraints across a larger kinematic range [17–19].

Many of the points mentioned above cannot be addressed until a sufficient amount of the required data is taken on nuclear targets. Therefore, we focus on the last point, since there are still many available data sets not yet included in global analyses of nuclear PDFs, which can provide new insights on poorly understood kinematic regions. Specifically, we will extend the nCTEQ15(WZ) analyses [9, 39] with data of light and heavy meson production taken at RHIC and the LHC to put stronger constraints on the nuclear gluon PDF at low x , where the uncertainties are particularly large.

The gluon PDF at low x is interesting for a variety of reasons: It contributes a large fraction to the overall cross section in a vast number of processes measured at the LHC, including vector boson production [40], prompt photon production [41] and production of heavy quarks [42]. Secondly, it is a significant source of uncertainty in more general models requiring PDFs as an ingredient, like the Statistical Hadronization Model [43], which has been used to describe the freezeout of quark-gluon plasmas (QGP) with great success. Furthermore, the different parton species mix as described by the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) evolution equations [44–47], which means that the gluon uncertainty directly affects the uncertainties of the quark PDFs as well. Finally, knowledge of the nuclear PDFs across the entire x range provides a starting point for comparisons with microscopic models of the nuclear modification of PDFs [48–58].

Before we can perform the new global nuclear PDF analyses, however, we need to understand the theoretical background, statistical tools and other technical aspects. This thesis is therefore structured as follows: We begin Sec. I.1 with a historical derivation of the basic parton model from observations in deep inelastic scattering experiments. Then we introduce the DGLAP evolution equations as a next-to-leading order (NLO) correction and provide a first look at nuclear modification of PDFs.

Section I.2 then introduces the statistical tools, beginning with the definition of the χ^2 cost function as a measure of fit quality. We then continue with the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm [59–62] used to minimize this cost function in our new analyses. Finally, we introduce the Hessian framework [63] in which the uncertainties of the PDFs are determined and some other tools to judge the fit quality with respect to individual data sets. Throughout the section we also introduce various extensions and optimizations of these methods.

The new analyses are presented in Chap. II, which begins with an overview of relevant existing nPDF analyses in Sec. II.1. Section II.2 then presents the first new nuclear PDF analysis, which includes data of pion, kaon and η meson production in a new global fit. Special care is taken in this analysis to mitigate the dependence on fragmentation functions (FFs), which are non-perturbative objects very similar to PDFs, and the dependence on

unphysical renormalization and factorization scales. The result is a new nuclear PDF set called nCTEQ15WZ+SIH [64].

The second analysis introduces data of heavy quark production and is presented in Sec. II.3, which begins with a discussion of different theoretical frameworks before introducing an empirical framework for the calculation of the cross sections. After showing that this framework can accurately reproduce the cross sections measured in proton-proton collisions, it is also verified against perturbative calculations for the same processes. The empirical predictions are then used to include LHC data on the production of mesons containing heavy quarks in another global fit. We compare the resulting nuclear PDF set, nCTEQ15HQ [65], to previous fits in the nCTEQ framework and similar fits performed by other groups using different methods.

We conclude in Chap. III with a summary of the developments and results. Finally, we give a brief outlook towards potential future improvements of nuclear PDF analyses over different time scales.

1. The parton model

Parton distribution functions (PDFs) are immensely important to modern particle physics because they are required to perform any calculation of cross sections for collisions involving hadrons. To explain the need for these PDFs in more detail, we begin with the historical derivation of the parton model [66] from experimental observations in deep inelastic scattering (DIS). This data shows an effect called *Bjorken scaling* [67, 68], which was the first evidence for point-like constituents making up nucleons.

We then give a proper definition of the PDFs and derive some initial expectations for these PDFs simply by assuming that the proton contains three valence quarks (two up and one down quark) held together by gluons. This assumption is sufficient to predict the general shape of the PDFs and even some observables in certain kinematic limits.

After this, we give a brief outline of the impact of NLO QCD corrections on PDFs, which violate the previously introduced Bjorken scaling. Included is a sketch of the derivation of the DGLAP evolution equations [44–47], which predict a logarithmic dependence on the momentum scale Q^2 and a mixing of PDF flavors with this scale.

We conclude the section with an introduction to the nuclear modification of PDFs and related observables, while providing an overview of microscopic models aiming to explain these modifications.

1.1. Deep inelastic scattering

In the 1960s electron-proton scattering was measured at the Stanford Linear Accelerator Center (SLAC). Fig. I.1 shows the observed cross section against the missing mass W . The leftmost peak shows elastic scattering, as evident by the fact that it is positioned at $W = m_{\text{proton}}$. The other nearby peaks are due to the excitation of Δ -resonances. The trail towards higher energies without any noticeable peaks marks the region of deep inelastic scattering, which is described by the Feynman diagram in Fig. I.2. Instead of scattering elastically, the electron transfers enough energy to the proton to break it up into a final state that is not known a priori.

Calculating the cross section of inelastic electron-proton scattering as an electromagnetic process requires the contraction of the most general¹ hadronic tensor

$$W^{\mu\nu} = W_1 \left(-g^{\mu\nu} + \frac{q^\mu q^\nu}{q^2} \right) + \frac{W_2}{M^2} \left(p^\mu - \left(\frac{p \cdot q}{q^2} \right) q^\mu \right) \left(p^\nu - \left(\frac{p \cdot q}{q^2} \right) q^\nu \right), \quad (1)$$

with the leptonic tensor $L_{\mu\nu}^e$ in the spin-averaged scattering matrix element

$$\overline{|\mathcal{M}|^2} = 4\pi M \frac{e^4}{q^4} L_{\mu\nu}^e W^{\mu\nu}. \quad (2)$$

¹A more general, parity violating version exists when taking into account *electroweak* interactions.

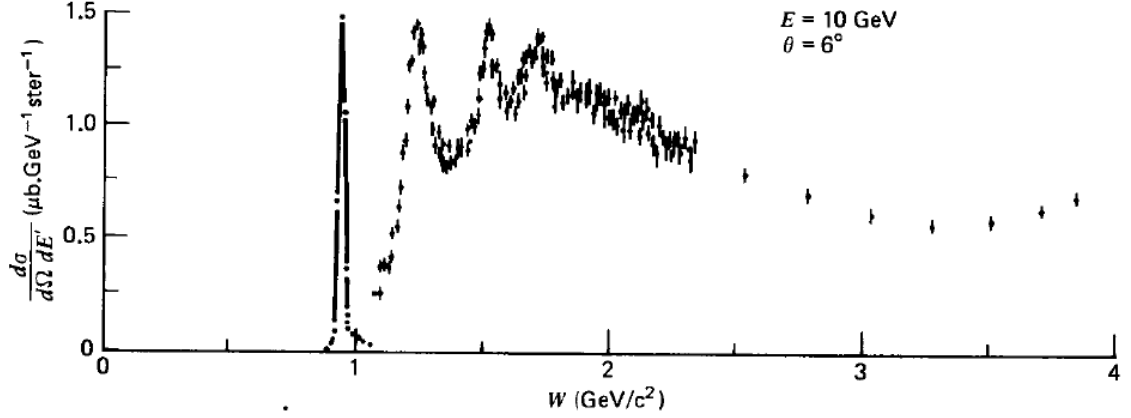


Figure I.1.: Cross section of electron-proton scattering as a function of missing mass W as measured at SLAC. The leftmost peak has been scaled down by a factor 8.5. Figure taken from Ref. [69].

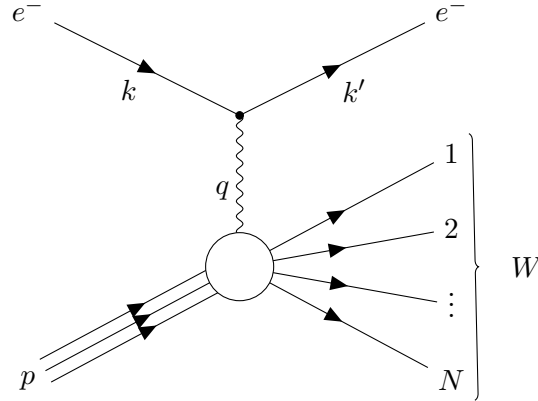


Figure I.2.: Leading order diagram of deep inelastic electron-proton scattering.

The coefficients W_1 and W_2 are scalars depending only on the other scalar variables at the vertex². Performing the calculation of the cross section with this, one obtains

$$\frac{d\sigma}{dE' d\Omega} = \frac{4\alpha^2 E'^2}{q^4} \left\{ W_2(\nu, q^2) \cos^2 \frac{\theta}{2} - 2W_1(\nu, q^2) \sin^2 \frac{\theta}{2} \right\} \quad (3)$$

at leading order. Attempts at experimentally determining the coefficients $W_1(\nu, q^2)$ and $W_2(\nu, q^2)$ in the 1960s lead to the discovery that in the high-energy limit the cross section

²A summary of the kinematic variables is shown in App. IV.2

(3) takes on the form

$$\frac{d\sigma}{dE'd\Omega} = \frac{4\alpha^2 E'^2}{q^4} \left(\cos^2 \frac{\theta}{2} - \frac{q^2}{2M^2} \sin^2 \frac{\theta}{2} \right) \delta \left(\nu - \frac{Q^2}{2m} \right). \quad (4)$$

This is the same form that can be derived for pointlike lepton-lepton scattering at leading order. This observation was what lead R. Feynman to the invention of the parton model [66], which postulates that the proton is built up from pointlike constituents called partons. Subsequently, these partons were identified with the quarks of M. Gell-Mann's quark model [70] that was created to explain isospin symmetry of hadrons.

1.2. The structure of nucleons

The core statement of the parton model is that at sufficiently high energies the electron no longer scatters on the proton as a whole but may instead perform an elastic scattering on a “free” quark within the proton. By comparing the cross sections for lepton-lepton scattering (4) and lepton-proton scattering (3), we see that in the parton model the proton structure functions become

$$2W_1 = \frac{Q^2}{2m^2} \delta \left(\nu - \frac{Q^2}{2m} \right) \quad \text{and} \quad W_2 = \delta \left(\nu - \frac{Q^2}{2m} \right), \quad (5)$$

where m is the mass of the quark. The structure functions W_1 and W_2 can be rewritten further so that they only depend on the dimensionless ratio $\frac{Q^2}{2m\nu}$ but not on Q^2 and ν separately:

$$\begin{aligned} 2mW_1(\nu, Q^2) &= \frac{Q^2}{2m\nu} \delta \left(1 - \frac{Q^2}{2m\nu} \right), \\ \nu W_2(\nu, Q^2) &= \delta \left(1 - \frac{Q^2}{2m\nu} \right). \end{aligned} \quad (6)$$

This behaviour is called *Bjorken scaling*, named after J. Bjorken who was the first to predict this behavior in the late 1960s [67, 68]. By introducing the new dimensionless variable $\omega = \frac{2q \cdot p}{Q^2} = \frac{2M\nu}{Q^2}$, we can summarize the behavior for large Q^2 as:

$$\begin{aligned} MW_1(\nu, Q^2) &\xrightarrow{\text{large } Q^2} F_1(\omega), \\ \nu W_2(\nu, Q^2) &\xrightarrow{\text{large } Q^2} F_2(\omega), \end{aligned} \quad (7)$$

Note that we used the proton mass M instead of the quark mass m to define ω because it will be more convenient later on. The assumption of pointlike, spin- $\frac{1}{2}$ particles inside the proton therefore leads to the prediction that looking at either structure function W_1 or W_2 for a fixed ω , while varying Q^2 , should result in a constant value. Figure I.3 shows

a sample of electron-proton scattering data from the BCDMS experiment [71] showing no signs of Q^2 dependence and thereby providing strong evidence for the parton model. As we will see later on in Sec. I.1.4, specifically Fig. I.7, this Bjorken scaling is actually violated and there is a logarithmic Q^2 dependence arising from higher order QCD corrections.

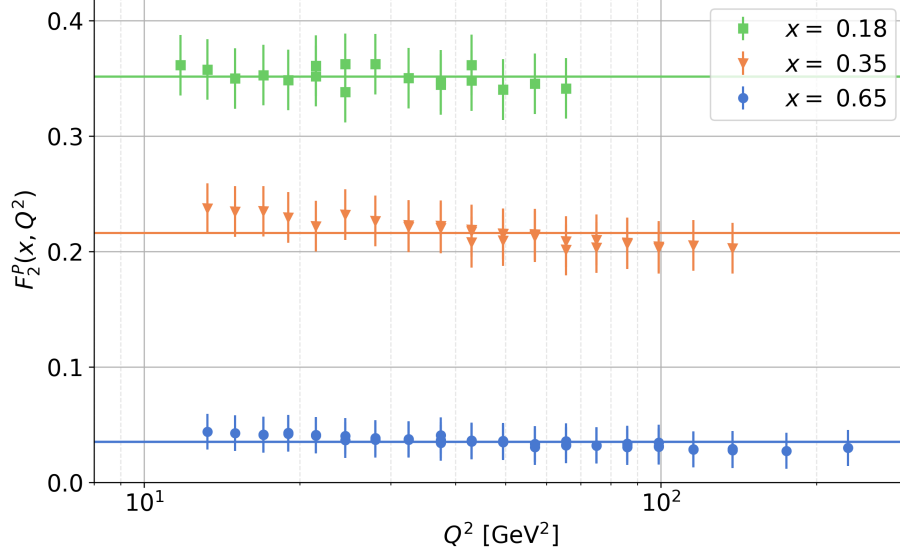


Figure I.3.: Proton structure function F_2 as a function of Q^2 for three different values of the Bjorken variable x . The data is taken from the BCDMS experiment [71] and the lines represent fitted constants. Figure taken from Ref. [72].³

1.3. Parton distribution functions

In the parton model the deep inelastic scattering process can be viewed as

$$E, p \rightarrow \left[\begin{array}{c} \text{Diagram: A proton (circle) with an incoming photon (wavy line) and an outgoing photon (wavy line). The proton is split into partons (arrows) labeled with momentum fractions x and xp. The parton i is shown interacting with the photon. The diagram is enclosed in large square brackets. \end{array} \right] = \sum_i \int dx e_i^2 \quad , \quad (8)$$

where $i \in \{\text{u-quark, d-quark, ..., gluon}\}$ and the respective charges are given by e_i . Since gluons have an electric charge of 0, they do not contribute at leading order. We can then introduce the parton momentum distribution $f_i(x)$ which describes the probability that a parton i that interacts with the photon carries a fraction x of the proton's total momentum

³Unless otherwise noted, all plots were created using Matplotlib [73].

p . Denoting the total chance of scattering on a parton i as P_i , we can define

$$f_i(x) \equiv \frac{dP_i}{dx} = p \quad \left\{ \begin{array}{l} xp \\ (1-x)p \end{array} \right.$$

$$\sum_i \int_0^1 dx \, x f_i(x) = 1. \quad (9)$$

Table I.1.: Parton kinematics.

	Proton	Parton
Energy	E	$x E$
Momentum	p	$x p$
Mass	M	$m = (x^2 E^2 - x^2 p^2)^{\frac{1}{2}} = x M$

$$\begin{aligned} \frac{2F_1(\omega)}{M} &= \frac{Q^2}{2m^2\nu} \delta\left(1 - \frac{Q^2}{2m\nu}\right) \implies F_1(\omega) = \frac{Q^2}{4xm\nu} \left(1 - \frac{Q^2}{2m\nu}\right) = \frac{1}{2x\omega} \delta\left(x - \frac{1}{\omega}\right), \\ F_2(\omega) &= \delta\left(1 - \frac{Q^2}{2m\nu}\right) \implies F_2(\omega) = x\delta\left(x - \frac{1}{\omega}\right), \end{aligned} \quad (10)$$
$$\begin{aligned} F_1(\omega) &= \sum_i \int dx \, e_i^2 \frac{1}{2\omega x} f_i(x) \delta\left(x - \frac{1}{\omega}\right), \\ F_2(\omega) &= \sum_i \int dx \, e_i^2 x f_i(x) \delta\left(x - \frac{1}{\omega}\right). \end{aligned} \quad (11)$$

⁴Actually, the mass of the partons cannot be variable like $m = xM$ suggests. Assuming the parton's momentum is xp , then the energy can only be xE if the masses are $m = M = 0$. We can justify this, however, by working in a frame of reference, where $|p| \gg m, M$. In a frame where the protons momentum is infinite, these kinematics then become exact.

$$\begin{aligned}
MW_1(\nu, Q^2) &\xrightarrow{\text{large } Q^2} F_1(x) = \frac{1}{2x} \sum_i e_i^2 x f_i(x), \\
\nu W_2(\nu, Q^2) &\xrightarrow{\text{large } Q^2} F_2(x) = \sum_i e_i^2 x f_i(x),
\end{aligned} \tag{12}$$

with

$$x = \frac{1}{\omega} = \frac{Q^2}{2M\nu}. \tag{13}$$

Thus, the structure functions of the proton depend only on the dimensionless variable x , often called Bjorken x .

1.3.1. Nucleon structure functions in the parton model

Using the results of the previous section can already grant us some insight into the quark structure of nucleons. Ignoring the heavy charm, bottom and top quarks for now, we can expand $F_2(x)$ as

$$\begin{aligned}
\frac{1}{x} F_2^{ep}(x) &= \left(\frac{2}{3}\right)^2 [u^p(x) + \bar{u}^p(x)] + \left(\frac{1}{3}\right)^2 [d^p(x) + \bar{d}^p(x)] + \left(\frac{1}{3}\right)^2 [s^p(x) + \bar{s}^p(x)], \\
\frac{1}{x} F_2^{en}(x) &= \left(\frac{2}{3}\right)^2 [u^n(x) + \bar{u}^n(x)] + \left(\frac{1}{3}\right)^2 [d^n(x) + \bar{d}^n(x)] + \left(\frac{1}{3}\right)^2 [s^n(x) + \bar{s}^n(x)],
\end{aligned} \tag{14}$$

for protons and neutrons, respectively. Assuming isospin symmetry, we can relate their quark contents by writing

$$\begin{aligned}
u^p(x) &= d^n(x) \equiv u(x), \\
d^p(x) &= u^n(x) \equiv d(x), \\
s^p(x) &= s^n(x) \equiv s(x).
\end{aligned} \tag{15}$$

Further, it is instructive to divide the quarks into valence quarks q_v and sea quarks q_s . Sea quarks are generated inside any hadrons as quark-antiquark pairs from the gluons binding the valence quarks. Since they always appear in $q\bar{q}$ -pairs, they do not affect the total quantum numbers of the nucleons. Assuming zero mass for up, down and strange quark, it is convenient to write

$$\begin{aligned}
S(x) &\equiv u_s(x) = \bar{u}_s(x) = d_s(x) = \bar{d}_s(x) = s_s(x) = \bar{s}_s(x), \\
u(x) &\equiv u_v(x) + u_s(x), \\
d(x) &\equiv d_v(x) + d_s(x),
\end{aligned} \tag{16}$$

1. The parton model

so that we can simplify F_2^{ep} and F_2^{en} into

$$\frac{1}{x}F_2^{ep} = \frac{1}{9}(4u_v + d_v) + \frac{4}{3}S \quad \text{and} \quad \frac{1}{x}F_2^{en} = \frac{1}{9}(u_v + 4d_v) + \frac{4}{3}S. \quad (17)$$

Since the gluons create the quark-antiquark pairs of the sea, it can be shown that at leading order the sea quark distributions grow logarithmically as $x \rightarrow 0$. Therefore, they are expected to overshadow the three valence quarks in this region, such that

$$\frac{F_2^{en}(x)}{F_2^{ep}(x)} \xrightarrow{x \rightarrow 0} 1. \quad (18)$$

On the other end of the x spectrum, it becomes increasingly unlikely for the sea quarks to be created with a very large momentum fraction x . Hence, the valence quarks dominate in interactions:

$$\frac{F_2^{en}(x)}{F_2^{ep}(x)} \xrightarrow{x \rightarrow 1} \frac{u_v + 4d_v}{4u_v + d_v}. \quad (19)$$

Both predictions are consistent with the data shown in Fig. I.4. Additionally this data suggests that $u_v \gg d_v$ at large x as the ratio (19) tends towards $\frac{1}{4}$.

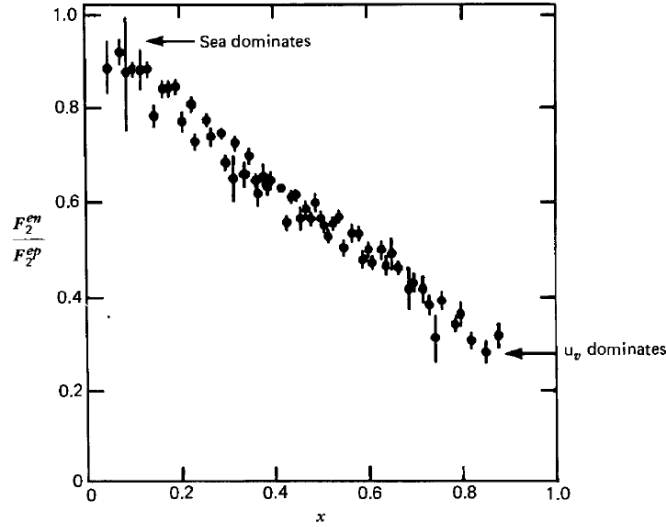


Figure I.4.: The ratio of structure functions of neutron over proton F_2^n/F_2^p as a function of x . Data taken at SLAC. Figure taken from Ref. [69].

A qualitative description of F_2 for various classes of proton structures is shown in Fig. I.5. Free quarks would produce sharp peaks at $x = \frac{1}{N_{\text{quarks}}}$ (first and second panel), which get smeared out when the interaction that binds the quarks into one hadron is considered (third panel). The splitting of gluons into sea quarks then leads to a non-vanishing structure function at low x as discussed above (fourth panel). This can be compared to the fitted

quark distributions shown in Fig. I.6, where the $xu(x)$ and $xd(x)$ distributions in the left panel are indeed qualitatively described by the last panel of Fig. I.5, while the valence quark distribution in the right panel of Fig. I.6 replicates the behaviour predicted in the third panel of Fig. I.6.

1.3.2. Sum rules

To reproduce the proton's quantum numbers and momentum in the parton model, the following four sum rules must be fulfilled:

$$\int_0^1 dx x[u(x) - \bar{u}(x)] = 2, \quad (20)$$

$$\int_0^1 dx x[d(x) - \bar{d}(x)] = 1, \quad (21)$$

$$\int_0^1 dx x[s(x) - \bar{s}(x)] = 0, \quad (22)$$

$$\sum_i \int_0^1 dx x f_i(x) = 1. \quad (23)$$

Therefore, Fig. I.6 also contains information about the gluons, even though they do not impact the structure functions F_2 from which the quark distributions were extracted. Integrating these fitted structure functions over x (while neglecting the small strange-quark contributions), we get

$$\begin{aligned} \int dx F_2^{ep}(x) &= \int_0^1 dx \frac{4}{9}x(u + \bar{u}) + \frac{1}{9}x(d + \bar{d}) = 0.18, \\ \int dx F_2^{en}(x) &= \int_0^1 dx \frac{1}{9}x(u + \bar{u}) + \frac{4}{9}x(d + \bar{d}) = 0.12. \end{aligned} \quad (24)$$

With the total momentum fraction carried by a specific flavour

$$\epsilon_q \equiv \int_0^1 dx x(q + \bar{q}), \quad (25)$$

this results in

$$\epsilon_u = 0.36, \quad \epsilon_d = 0.18, \quad \epsilon_g = 1 - (\epsilon_u + \epsilon_d) = 0.46. \quad (26)$$

Hence, the gluons mediating the strong force between the quarks carry roughly half of the protons total momentum.

1. The parton model

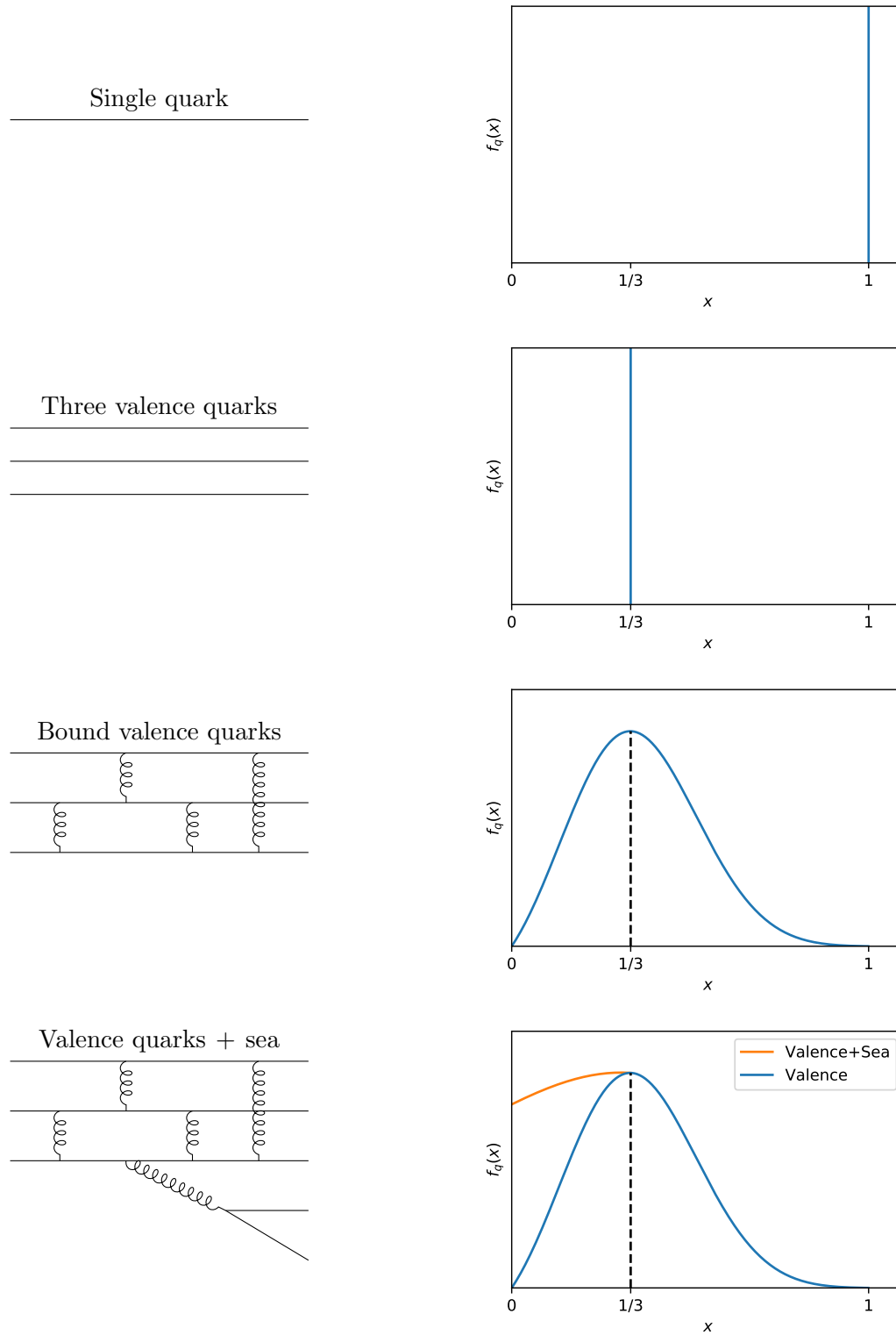


Figure I.5.: Qualitative expectation for the structure function F_2^p based on different assumed proton compositions.

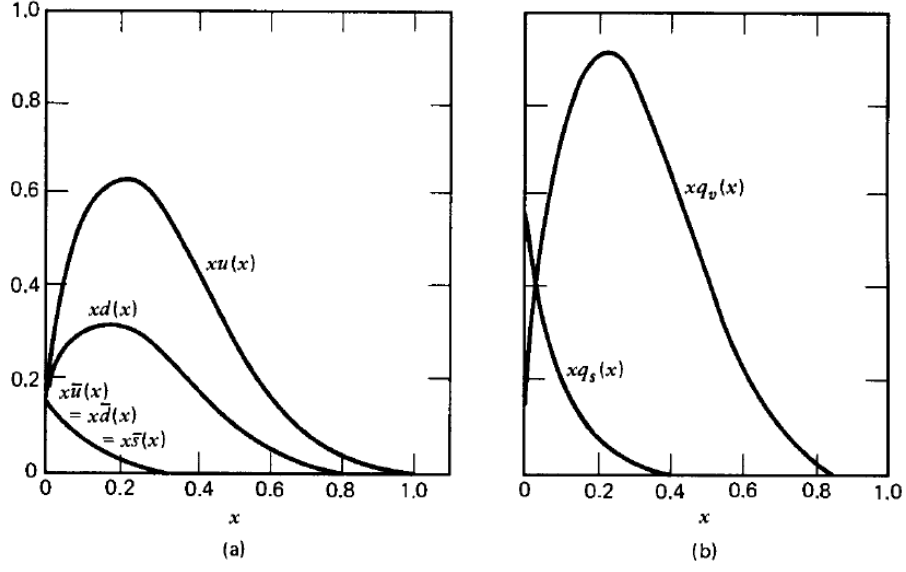


Figure I.6.: Quark momentum distributions fitted to F_2 data taken at SLAC. The left panel shows the distributions for individual flavours, while the right panel shows the total valence and sea quark distributions. Figure taken from Ref. [69]

1.4. DGLAP evolution

When performing experimental measurements of the proton structure function F_2 at higher precision and over a larger kinematic range than was possible in the early days at SLAC, it was observed that Bjorken scaling is violated, as shown in Fig. I.7. The structure function shows a clear logarithmic dependence on the energy scale Q^2 that is caused by higher order QCD effects. To show this, we start by writing hadronic tensor $\hat{W}^{\mu\nu}(z, Q^2)$ in terms of its partonic contributions $W^{\mu\nu}(x, Q^2)$, where $z \equiv \frac{Q^2}{2p_i \cdot q}$. If the probability of interacting with a parton of momentum $p_i^\mu = \xi p^\mu$ for any $0 \leq \xi \leq 1$ is given by a parton distribution function $f_i(\xi)$, we can write

$$\begin{aligned}
 W^{\mu\nu}(x, Q^2) &= \sum_i \int_0^1 dz \int_0^1 d\xi f_i(\xi) \hat{W}^{\mu\nu}(z, Q^2) \delta(x - z\xi) \\
 &= \sum_i \int_x^1 \frac{d\xi}{\xi} f_i(\xi) \hat{W}^{\mu\nu}\left(\frac{x}{\xi}, Q^2\right).
 \end{aligned} \tag{27}$$

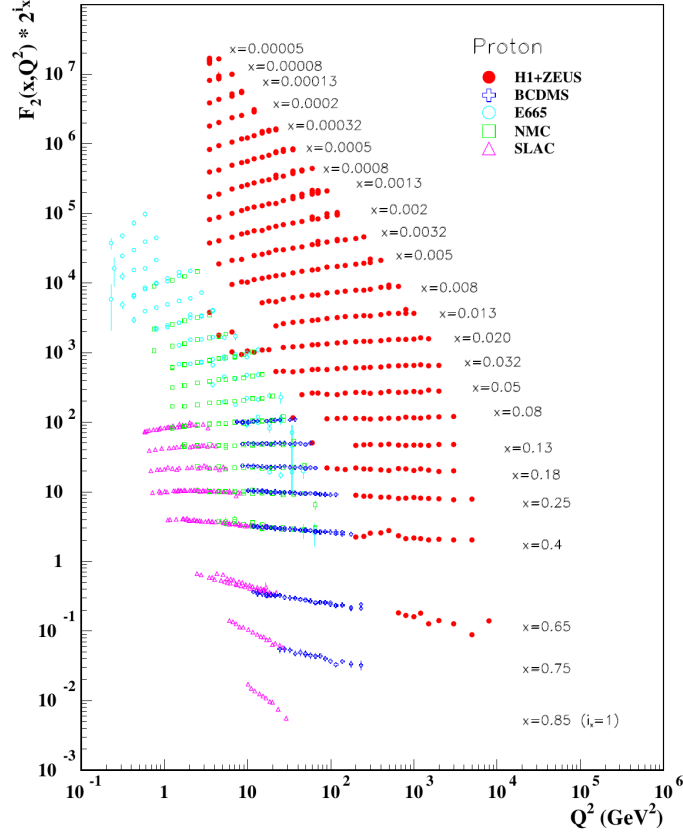


Figure I.7.: F_2^p measured in various experiments. The values of F_2^p for the various x are offset from each other by multiplication with 2^i , where i counts the x bins in descending order. Figure taken from Ref. [74].

At leading-order, the process $\gamma^* q \rightarrow q$ is the only contribution to the partonic tensor $\hat{W}^{\mu\nu}(z, Q^2)$:

$$\begin{aligned} \hat{W}^{\mu\nu}(z, Q^2) &= \frac{e_i^2}{2} \int \frac{d^3 \vec{p}_f}{(2\pi)^3} \frac{1}{2E_f} \text{Tr} \left[\gamma^\mu \not{p}_i \gamma^\nu \not{p}_f \right] (2\pi)^4 \delta^4(p_i + q - p_f) \\ &= 2\pi Q_i^2 \left[\left(-g^{\mu\nu} + \frac{q^\mu q^\nu}{q^2} \right) + \frac{4z^2}{Q^2} \left(p_i^\mu - \frac{p_i \cdot q}{q^2} q^\mu \right) \left(p_i^\nu - \frac{p_i \cdot q}{q^2} q^\nu \right) \right] \delta(1-z). \end{aligned} \quad (28)$$

Comparing this to the hadronic tensor (1), we obtain

$$\hat{W}_1 = 2\pi e_i^2 \delta(1-z) \quad \text{and} \quad \hat{W}_2 = \frac{8\pi e_i^2}{Q^2} z^2 \delta(1-z). \quad (29)$$

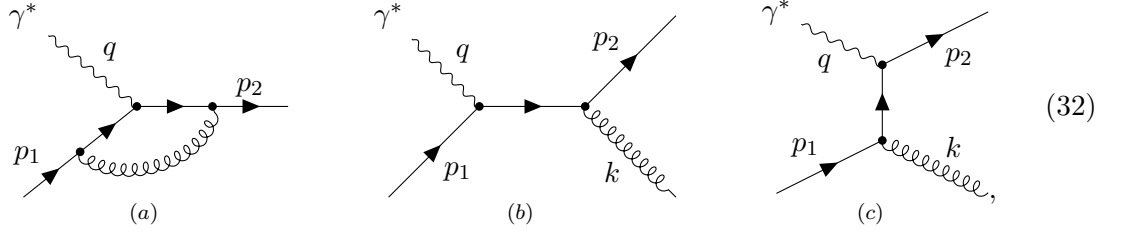
Next, we introduce the quantity $W_0 \equiv -g^{\mu\nu}W_{\mu\nu}$, which is proportional to the unpolarized cross section for the process $\gamma^*p \rightarrow X$. In the high energy limit $Q \gg m_p$, this gives

$$W_0(x, Q^2) = 3W_1(x, Q^2) - \frac{Q^2}{4x^2}W_2(x, Q^2) = 2W_1(x, Q^2) \quad (30)$$

$$\longrightarrow W_0(x, Q^2) = 4\pi \sum_i e_i^2 f_i(x) = \frac{4\pi}{x} F_2(x, Q^2). \quad (31)$$

This equation can be used to define PDFs even beyond leading-order giving us a convenient way to calculate the Q^2 dependence of the PDFs without the need to look at the individual structure functions. At leading order Eqs. (29) and (30) result in $W_0^{\text{LO}} = 4\pi e_i^2 \delta(1-z)$.

At next-to-leading order (NLO) the following three graphs contribute to \hat{W}_0 :



where (a) is the virtual $\gamma^*q \rightarrow q$ and (b) + (c) are real $\gamma^*q \rightarrow qg$ graphs. A detailed calculation of these contributions can be found in Ref. [72], resulting in the following \hat{W}_0 at NLO:

$$\begin{aligned} \hat{W}_0 &= \hat{W}_0^{\text{LO}} + \hat{W}_0^V + \hat{W}_0^R \\ &= 4\pi e_i^2 \left\{ \left[\delta(1-z) - \frac{1}{\epsilon} \frac{\alpha_s}{2\pi} P_{qq}(z) \left(\frac{4\pi\mu^2}{Q^2} \right)^\epsilon \frac{\Gamma(1-\epsilon)}{\Gamma(1-2\epsilon)} \right] + \frac{\alpha_s}{2\pi} C_F \left[-\frac{3}{2} \left[\frac{1}{1-z} \right]_+ \right. \right. \\ &\quad \left. \left. + (1+z^2) \left[\frac{\ln(1-z)}{1-z} \right]_+ - \frac{1+z^2}{1-z} \ln z + 3 + 2z - \left(\frac{9}{2} + \frac{1}{3}\pi^2 \right) \delta(1-z) \right] \right\}, \end{aligned} \quad (33)$$

where P_{qq} is a so-called splitting function

$$P_{qq}(z) = C_F \left[(1+z^2) \left[\frac{1}{1-z} \right]_+ + \frac{3}{2} \delta(1-z) \right]. \quad (34)$$

The fact that a $\frac{1}{\epsilon}$ pole with residue proportional to $P_{qq}(z)$ remains in the parton-level cross section does not pose a problem because it is not a physical observable. Calculating the hadronic cross section (27) from the partonic results (33), we get

$$W_0(x, Q^2) = 4\pi \sum_i e_i^2 \int_x^1 \frac{d\xi}{\xi} f_i(\xi) \left[\delta\left(1 - \frac{x}{\xi}\right) - \frac{\alpha_s}{2\pi} P_{qq}\left(\frac{x}{\xi}\right) \left(\frac{1}{\epsilon} + \ln \frac{\tilde{\mu}^2}{Q^2} \right) + \text{finite} \right], \quad (35)$$

as the physical observable. Since the splitting function satisfies

$$\int_0^1 P_{qq}(z) dz = 0, \quad (36)$$

the pole vanishes and the total DIS cross section at any given Q^2 is finite when integrated over x . However, for any fixed x the cross section is still divergent because the calculation was only performed in massless QCD, where the infrared divergence does not get cut off due to a physical scale such as the quark mass. Nonetheless, by taking differences at different scales Q^2 and Q_0^2 , we can obtain a finite result:

$$W_0(x, Q^2) - W_0(x, Q_0^2) = 4\pi \sum_i e_i^2 \int_x^1 \frac{d\xi}{\xi} f_i(\xi) \left[\frac{\alpha_s}{2\pi} P_{qq} \left(\frac{x}{\xi} \right) \ln \frac{Q^2}{Q_0^2} \right]. \quad (37)$$

The pole in the partonic cross section (33) therefore leads to a logarithmic Q^2 dependence in the physical hadronic cross section that explains the violation of Bjorken scaling.

We can also express observables in terms of renormalized quantities to remove the necessity of taking differences. This can be done, for example, by defining some reference scale Q_0^2 at which Eq. (31) holds exactly:

$$W_0(x, Q^2) \equiv 4\pi \sum_i e_i^2 f_i(x, \mu^2 = Q^2). \quad (38)$$

The renormalized PDFs then have to fulfill

$$f_i(x, \mu_1^2) = f_i(x, \mu^2) + \frac{\alpha_s}{2\pi} \int_x^1 \frac{d\xi}{\xi} f_i(\xi, \mu_1^2) P_{qq} \left(\frac{x}{\xi} \right) \ln \frac{\mu_1^2}{\mu^2} \quad (39)$$

to be consistent with Eq. (37). Taking the derivative w.r.t. the scale μ then gives a DGLAP [44–47] evolution equation

$$\mu^2 \frac{d}{d\mu^2} f_i(x, \mu^2) = \frac{\alpha_s}{2\pi} \int_x^1 \frac{d\xi}{\xi} f_i(\xi, \mu^2) P_{qq} \left(\frac{x}{\xi} \right). \quad (40)$$

Further DGLAP evolution equations can be obtained by considering other processes that exist at NLO such as $\gamma^* g \rightarrow q\bar{q}$ with an initial state gluon. These other DGLAP evolution equations lead to a mixing of PDFs that can be fully expressed for quarks and gluons by

$$\mu^2 \frac{d}{d\mu^2} \begin{pmatrix} f_i(x, \mu^2) \\ f_g(x, \mu^2) \end{pmatrix} = \sum_j \frac{\alpha_s}{2\pi} \int_x^1 \frac{d\xi}{\xi} \begin{pmatrix} P_{q_i q_j} \left(\frac{x}{\xi} \right) & P_{q_i g} \left(\frac{x}{\xi} \right) \\ P_{g q_j} \left(\frac{x}{\xi} \right) & P_{gg} \left(\frac{x}{\xi} \right) \end{pmatrix} \begin{pmatrix} f_j(\xi, \mu^2) \\ f_g(\xi, \mu^2) \end{pmatrix}, \quad (41)$$

where i is one quark flavour and the sum runs over all quark flavours. The other splitting

functions can be derived from the cross sections of other processes in a similar manner to the way we derived $P_{qq}(z)$. At leading order they are given by

$$P_{qq}(z) = C_F \left[\frac{1+z^2}{[1-z]_+} + \frac{3}{2} \delta(1-z) \right], \quad (42)$$

$$P_{qg}(z) = T_F [z^2 + (1-z)^2], \quad (43)$$

$$P_{gq}(z) = C_F \left[\frac{1+(1-z)^2}{z} \right], \quad (44)$$

$$P_{gg}(z) = 2C_A \left[\frac{z}{[1-z]_+} + \frac{1-z}{z} + z(1-z) \right] + \delta(1-z) \left[\frac{11}{6} C_A - \frac{2}{3} T_F n_f \right]. \quad (45)$$

A demonstration of the effect that DGLAP evolution has on actual PDFs is shown in Fig 2.10 of Ref. [72].

1.5. Nuclear modification of PDFs

Naively, one might expect that the structure functions measured in deep inelastic scattering on heavier nuclei might just be the sum of the structure functions of the contained protons and neutrons. However, measurements of DIS on iron targets revealed that the structure functions of these nuclei are different from this expectation. The first such measurement was performed by the european muon collaboration (EMC) at CERN [75] in 1982 and observed a suppression of F_2 around $x \approx 0.6$. This observation was therefore named the *EMC effect*. Figure I.8 shows more recent data of $\frac{F_2^{Fe}}{F_2^D}$ from SLAC and BCDMS with isoscalar corrections applied to the structure function of iron. In the absence of nuclear effects this ratio would therefore be expected to be one.

The fitted curve clearly shows four distinct regions of the nuclear modification. First, the *shadowing* region at low x shows a suppressed F_2 in the heavier nucleus. At x values just above that, we see a slight enhancement of the structure function in the so called *anti-shadowing* region and then another suppression in the region of the aforementioned EMC effect. Finally, at very high x , the nuclear structure function is enhanced again in the *Fermi-motion* region. Microscopic models have been proposed for the shadowing region [48–51], the antishadowing region [51–53], the EMC effect [51, 54–58], and ferm-motion [83]. Additionally, at very small x , the gluon density is expected to become very large until it reaches a saturation, such that the assumptions underlying factorization break down. This saturated state is described by the theory of color-glass condensates (CGC) [84, 85] and there are models interpolating between the CGC at small x and collinear factorization at large x , as explained in Ref. [86] and references therein.

In the following we will give a brief overview of the currently available theoretical explanations of these effects.

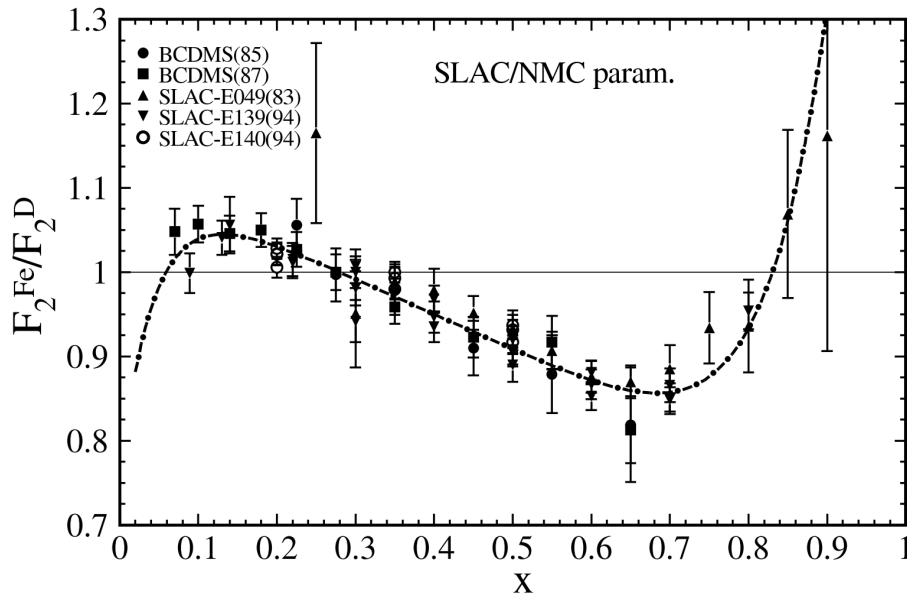


Figure I.8.: Nuclear correction ratio, F_2^{Fe}/F_2^D as a function of x . The parameterized curve has been fitted to the shown SLAC and BCDMS data [76–81]. Figure taken from Ref. [82]

1.5.1. Shadowing

As mentioned before, shadowing is generally observed in the low- x region and increases in magnitude as x gets smaller. Additionally, the effect is stronger at low Q^2 and rises with the mass number A of the nucleus. Recent reviews on the topic are given by Refs. [48, 49].

At moderate energies, nuclear shadowing is a consequence of destructive interference among the scattering amplitudes of multiple interactions between an incoming particle and nuclear target. This effect was originally derived as an explanation for the observed cross sections in scattering processes with deuterium targets by Glauber in 1955 [87], long before parton distributions were a known quantity. However, at higher energies this description is no longer valid, since the interference terms cancel in this limit [88]. Instead, one needs to consider the effects of an arbitrary projectile, *e.g.*, photon, neutrino or hadron interacting with the target hadrons through multiple fluctuations into long-lived quark-gluon states [89]. This series of interactions with the target nucleons is shown in Fig. I.9. These interactions involve diffractive processes, *i.e.* processes where the target nucleus stays intact. Diffractive processes are subject to a different factorization theorem [90], in which diffractive parton distributions $f_{j/N}^D$ are defined. As a secondary consequence of this factorization theorem, the interactions where $N = 2$ nucleons participate can be expressed in terms of the regular proton PDF and this diffractive proton PDFs. In this framework, interactions with $N \geq 3$ nucleons, can be parameterized by the effective soft cross section σ_{soft}^j . This can be used to derive a master formula for the nuclear shadowing

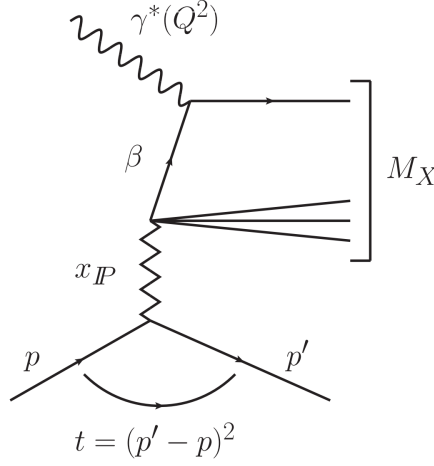


Figure I.9.: Diffractive DIS at leading order. Figure taken from Ref. [91].

$\delta x f_{j/A}(x, Q_0^2) \equiv x f_j(x, Q_0^2) - x f_{j/N}^{\text{IA}}(x, Q_0^2)$, where $x f_{j/N}^{\text{IA}}(x, Q_0^2)$ is the nuclear PDF in the impulse approximation. The full details of this derivation are given in Ref. [49], but the result is:

$$\begin{aligned} \delta x f_{j/A}(x, Q_0^2) = & -8\pi A(A-1) \Re e \frac{(1-i\eta)^2}{1+\eta^2} \int_x^{x_0} dx_P \beta f_j^{D(3)}(\beta, Q_0^2, x_P) \int d^2b \int \int_{-\infty}^{\infty} dz_1 dz_2 \\ & \times \rho_A(\vec{b}, z_1) \rho_A(\vec{b}, z_2) e^{i(z_1-z_2)x_P m_N} e^{-\frac{A}{2}(1-i\eta)\sigma_{\text{soft}}^j(x, Q^2) \int_{z_1}^{z_2} dz' \rho_A(\vec{b}, z')}. \end{aligned} \quad (46)$$

The diffractive parton distributions $f_j^{D(3)}$ depends on the variables $\beta = x/x_P$ and $x_P = (M_X^2 + Q_0^2)/(W^2 + Q_0^2)$, where M_X is the total mass of the diffractively produced final state. The parameter η is the ratio of the real to imaginary parts of the diffractive $\gamma^* N \rightarrow XN$ scattering amplitude. Finally, the quantity ρ is the nuclear density parameterized by the transverse and longitudinal coordinates \vec{b} and z , respectively.

Since this shadowing correction can be defined entirely in terms of leading twist operators, it is itself a leading twist quantity and the Q^2 dependence of the nuclear PDFs derived from it should therefore obey the standard DGLAP evolution.

1.5.2. Antishadowing

Antishadowing describes the observation that the structure function $F_2(x, Q^2)$ is enhanced between $0.05 \lesssim x \lesssim 0.2$ in the nuclear case, compared to the proton case. It was observed in Ref. [91] that the same effects leading to shadowing at low x can also lead to antishadowing

at higher values of x . From momentum conservation they derive

$$2 \sum_{q=u,d,s,c} \int_0^{0.1} dx \delta x \bar{q}_A(x, Q_0^2) + \int_0^{0.1} dx \delta x g_A(x, Q_0^2) + \int_0^{0.2} dx \delta x g_A^{\text{anti}}(x, Q_0^2) = 0. \quad (47)$$

where the first term describes the sea quark sum shadowing, the second one corresponds to gluon shadowing and the final one describes gluon antishadowing. Valence quark contributions are neglected due to their numerically small modification. Antishadowing contributions of sea quarks are considered to be 0 since the nuclear modification of measured Drell-Yan cross sections is consistent with unity for $x > 0.1$ [91]. The upper limit of $x = 0.2$ for the gluon antishadowing comes from the fact that coherence effects vanish quickly as the coherence length $l_c \approx 1/(2m_n x)$ becomes comparable to the average distance $r_N N \approx 1.7$ fm between two nucleons in a nucleus, which corresponds to $x \approx 0.2$. It can be shown [91], that sum rule (47) is fulfilled locally on the interval $\ln(x/x_P) \leq 1$. Therefore, it is useful to introduce an explicitly x_P -dependent way of expressing the gluon antishadowing:

$$\int_0^{0.1} dx_P \delta x g_A^{\text{anti}}(x, x_P, Q_0^2) = \delta x g_A^{\text{anti}}(x, Q_0^2), \quad (48)$$

which we can insert into Eq. (47) to obtain

$$\sum_j \int_0^{0.1} dx \int_x^{0.1} dx_P \delta x f_{j/A}(x, x_P, Q_0^2) + \int_0^{0.2} dx \int_0^{0.1} dx_P \delta x g_A^{\text{anti}}(x, x_P, Q_0^2) = 0, \quad (49)$$

where j sums over the sea quarks and gluons. The shadowing correction $\delta x f_{j/A}(x, x_P, Q_0^2)$ is the one derived in Eq. (46), without the integration over x_P . Changing the order of the integrations and requiring the sum rule to be satisfied for each x_P , the final result is

$$\sum_j \int_0^{x_P} dx \delta x f_{j/A}(x, x_P, Q_0^2) + \int_0^{0.2} dx \delta x g_A^{\text{anti}}(x, x_P, Q_0^2) = 0. \quad (50)$$

While this does not provide an explicit form for the gluon shadowing factor, it does provide a framework, in which it can be parameterized and determined without having to introduce an explicit A dependence. This was done in Ref. [91] for two different parameterizations of the antishadowing obtaining the results shown in Fig. I.10. The nuclear modification obtained from the leading twist model generally agrees with the fitted nPDF, but the two parameterizations do not share overlapping uncertainties, which indicates that they may be underestimated.

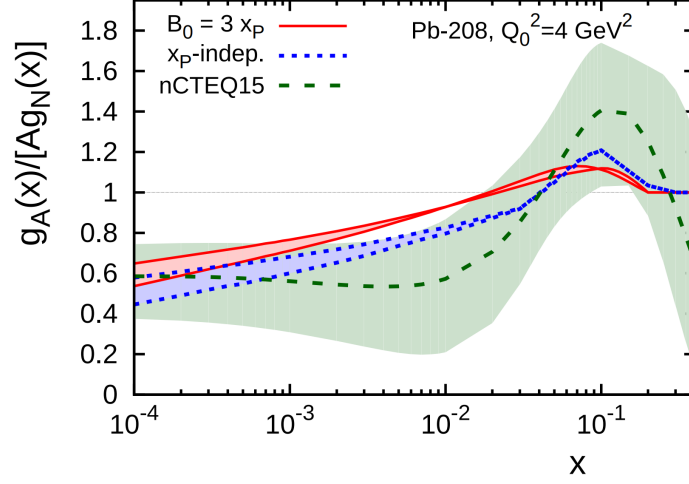


Figure I.10.: Leading twist nuclear gluon shadowing and antishadowing compared to nCTEQ15 nuclear PDFs. Figure taken from Ref. [91].

1.5.3. EMC effect

The suppression of the structure function F_2 in the x region above the antishadowing region is called the EMC effect. Currently there are multiple theoretical models aiming to explain the nuclear modification in this region: nuclear binding [92, 93], pion excess [94, 95], multi-quark clusters [96, 97], dynamical rescaling [98, 99], medium modification [100, 101] and short range correlations (SRC) [102, 103]. Discussing all of these models in detail is far beyond the scope of this section and therefore we limit the discussion to a brief overview of each. We refer to the review Ref. [57] and the references therein for more detailed explanations.

The first model considers the scattering on nuclear targets as a two step process, where the virtual photon first scatters from a quark bound in an arbitrary hadron present in the nucleus. These hadrons themselves follow a distribution inside the nucleus that is defined in analogy to the parton distribution functions. The full nuclear PDF is then a convolution of the free quark PDFs inside the hadrons with the distribution functions of the hadrons inside the nucleus.

The pion excess model is motivated by the fact that pions mediate the long-range part of the nucleon-nucleon interaction. If the amount of pions in the nucleus is sufficient, the contributions of the pion PDFs could explain the EMC effect in this model, but this would also suggest an enhancement of sea quark distributions in nuclei, which is inconsistent with pion Drell-Yan experiments [104].

In the model of multi-quark clusters, the quarks inside the nucleus do not stay bound in individual nucleons, but instead a portion of the quarks forms multiquark states with 6 or more quarks. This shifts the valence quark momenta towards higher x values, such that

momentum sum rules then dictate a suppression in the EMC region.

The dynamical scaling model is based on the experimental observation that the nuclear structure function of iron resembles that of the nucleons at a larger value of Q^2 . The model therefore attempts to explain nuclear modification of PDFs through the following relation

$$F_{2A}(x, Q^2) \approx F_{2N}(x, \xi_A(Q^2)Q^2), \quad (51)$$

with a rescaling factor $\xi_A(Q^2) > 1$. While this model indeed explains the nuclear modification in the EMC region, it also suggests a significant enhancement in the region $x \lesssim 0.1$, which is not observed experimentally.

In the medium modification model, the nuclear modification is caused by the meson field, which binds the nucleons together. However, in contrast to the pion excess model, the meson field directly modifies the quark distributions inside of the nucleons and there is no direct interaction between the photon and the mesons.

Since the model is relevant beyond the EMC effect, SRCs receive a dedicated discussion in Sec. I.1.5.5.

1.5.4. Fermi motion

In the proton case, the Bjorken variable x is limited to the region $0 \leq x \leq 1$. In the case of a nucleus with mass number A , however, the individual nuclei can exchange momentum through their interactions, such that the range is extended to $0 \leq x_N \leq A$. Since the PDFs vanish at the boundaries of this region, this naturally leads to a rapidly rising ratio $R_A \equiv F_2^A/F_2^p > 1$ as $x_A \rightarrow 1$. More generally, it is expected that in this region $F_2^{A_1} > F_2^{A_2}$ for $A_1 > A_2$. Measurements have been performed at Hall C of Jefferson Lab of F_2 for various nuclei in the large x region up to $x \lesssim 1.4$ [105, 106]. They fit the high- x behavior with an exponential $\propto \exp(-sx)$ and find $s \approx 15 \pm 2$ with a mild increase at higher Q^2 values and a mild decrease for heavier nuclei with larger A .

For $s = 14$, the ratio of F_2 in the region $x \in [0.65, 1]$ compared to $x \in [1, \infty]$ is roughly $\lesssim 1\%$. Given that the partons with $x \in [0.65, 1]$ carry roughly 2% of the total nucleus momentum, this can be used to estimate the total momentum carried by partons with $x > 1$ to be $1\% \times 2\% \approx 0.02\%$ of the total momentum.

However, even minor differences in this exponential parameter explain the observed $F_2^{A_1}/F_2^{A_2}$ ratio in the $x \rightarrow 1$ region.

While some of the effects listed in the sections above were originally defined in the context of nuclear structure functions, the same names are also commonly used to describe the corresponding observations in the nuclear parton distribution functions themselves. Examining these effects in terms of PDFs with nuclear modification has been the goal of many analyses [9–16, 107–109] and this thesis aims to improve the understanding of these nuclear modifications by including new data in nuclear PDF analyses in the nCTEQ

framework. Before we can begin with such analyses, however, we need to understand the technical framework in which they are performed, which will be the subject of Sec. I.2.

1.5.5. Short range correlations

Short range correlations offer a unified approach to explain the high- x behavior of the nuclear modification. The existence of short range correlations between nucleons is well motivated from theories as far back as in the 1950s [110] and finally confirmed experimentally in the late 1980s [111, 112]. Short range correlations mostly manifest as pairs of two nuclei with strongly overlapping wavefunctions. This leads to a large momentum relative to each other, while the pair's center of mass has a low momentum relative to the full nucleus. The majority of these pairs are formed as proton-neutron pairs, which are around 6 times more likely than proton-proton or neutron-neutron pairs [113]. Groups of more than two nucleons are also possible, but are significantly less likely and have therefore not been confirmed experimentally thus far [114].

It has been observed experimentally, that there is a linear correlation between the strength of the EMC effect and the strength of SRCs in different nuclei [115], which suggests that SRCs might be the cause of the EMC effect. The dominance of pn pairs among the SRC pairs would then suggest that nuclei with a large asymmetry in number of protons and neutrons would see a larger relative EMC modification in up valence quarks than in down valence quarks, since a larger fraction of the protons participates in SRC pairs.

Preliminary studies [116] show that nuclear PDF parameterizations motivated by these SRCs indeed show good agreement with the available data, particularly at large x . In this study, it is assumed that the full nuclear PDF is the sum of a free proton PDF and a SRC PDF and the A dependence is encoded in the ratio of the two PDFs. Imposing this separately on the proton and neutron PDFs it was found that the total number of protons and neutrons in the SRC pair is indeed equal, which is compatible with the expectations due to the dominance of pn pairs.

1.5.6. Target mass corrections

As the name implies, target mass corrections (TMC) are necessary to account for the fact that the proton mass was assumed to be negligible in the initial derivation of the structure function F_2 . In the case of heavier nuclei, the error due to this assumption is therefore expected to increase. The leading TMC is simply given by

$$F_2^{\text{TMC}}(x, Q) = \frac{x^2}{\xi^2 r^3} F_2^{(0)}(\xi, Q), \quad (52)$$

where $\xi = 2x/(1+r)$ is the so-called Nachtmann variable with $r = \sqrt{1 + 4x^2 M^2/Q^2}$ and $F_2^{(0)}(\xi, Q)$ is the structure function in the massless limit. The obvious question at this point

is whether the TMCs actually scale with the mass of the nucleon M or the mass of the full nucleus M_A , which could cause significant differences, especially for heavy nuclei like lead with $A = 208$. Tracing the steps of the TMC derivation in the operator product expansion (OPE) performed in Ref. [117] it can be seen that the A dependence of the Bjorken x for the nucleus cancels that of the mass, such that TMCs in the nucleus are still suppressed by Mx/Q , where M is the nucleon mass and x the Bjorken variable for a single nucleon. There are, however, potential complications due to the extension of the OPE and DGLAP formalisms into the nuclear case, which are currently being investigated [118].

Subleading terms of the TMCs have been derived in Ref. [117] and in Ref. [119] comparisons between the leading and full TMC have been performed. The results show that the subleading terms cause a change in F_2 up to $\sim 25\%$ at high x and low Q . The subleading terms are also weakly dependent on the nuclear A , but this dependence is negligible compared to current data uncertainties. Since all of the DIS data sets used in nuclear PDF fits are given as ratios of cross sections between two nuclei, these subleading TMC terms do not produce a meaningful effect in current analyses, but would become important if data on absolute cross sections is used in the future.

1.6. Statistical hadronization model

The statistical hadronization model (SHM) [120–122] describes the production of hadrons in the freeze-out of a quark-gluon plasma (QGP). Applying a statistical approach to the hadron gas produced in heavy-ion collisions has proven to accurately reproduce the multiplicities of the produced particles, even though the number of particles involved in the process may be as low as $\mathcal{O}(10)$. One of the key ingredients in predictions of heavy quark production in the SHM is the gluon PDF of the initial state hadrons and the uncertainty of current nuclear gluon PDFs is one of the limiting factors for the predictivity of the current models, which is one of the main motivations of this thesis. In the following we will briefly explain the most important aspects of the basic SHM approach for light quarks and present some key results. Afterwards we will outline how charmed hadrons can be introduced to the model, which is where the nuclear PDFs enter the picture.

1.6.1. Statistical hadronization of the QGP

The equilibrium behavior of thermodynamic observables can be evaluated as the average over statistical ensembles. Therefore, the equilibrium distribution is given by the average over the accessible phase space and the ensemble corresponding to thermodynamic equilibrium is given by the one with the most uniform phase space density. Conversely, this means that the agreement of measured observables with predictions using the statistical operator implies that the observed system is in equilibrium. The statistical hadronization model generally works in the grand-canonical (GC) ensemble, which is fully determined by

the temperature T and the chemical potential μ_B . If any additional parameters, *e.g.*, a strangeness suppression factor γ_s are required, that implies a deviation from the chemical equilibrium.

The basic quantity required to compute any observable in the statistical hadronization model is the partition function $Z(T, V)$ in the GC ensemble,

$$Z^{\text{GC}}(T, V, \mu_Q) = \text{Tr} \left[e^{-\beta(H - \sum_i \mu_{Q_i} Q_i)} \right], \quad (53)$$

where H is the Hamiltonian of the system, Q_i are the conserved charges and μ_{Q_i} the chemical potential guaranteeing charge conservation in the whole system. The temperature is commonly replaced by its inverse $\beta = \frac{1}{T}$. The Hamiltonian was originally chosen to describe a hadron resonance gas with contributions from mesons with masses below 1.5 GeV and baryons with masses below 2 GeV, but more recent analyses [123] include the entire hadron spectrum contained in the 2008 list from the particle data group [124]. The medium is considered to interact only via strong interaction, such that electric charge, baryon number and strangeness are conserved. The partition function can then be written as the sum over the partition functions Z_i of the individual hadrons and resonances

$$\ln Z(T, V, \vec{\mu}) = \sum_i \ln Z_i(T, V, \vec{\mu}), \quad (54)$$

with $\vec{\mu} = (\mu_B, \mu_S, \mu_Q)$ being the vector of chemical potentials for baryon number B , strangeness S and electric charge Q . For particles with energy $\epsilon_i = \sqrt{p^2 + m_i^2}$ and spin-isospin degeneracy factor g_i , the individual partition functions are given by

$$\ln Z_i(T, V, \vec{\mu}) = \frac{V g_i}{2\pi^2} \int_0^\infty \pm p^2 dp \ln \left[1 \pm \lambda_i e^{-\beta \epsilon_i} \right] \quad (55)$$

$$= \frac{V T g_i}{2\pi^2} \sum_{k=1}^\infty \frac{(\pm 1)^{k+1}}{k^2} \lambda_i^k m_i^2 K_2 \left(\frac{k m_i}{T} \right), \quad (56)$$

where the upper signs are for fermions and the lower ones for bosons⁵ and K_2 is the modified Bessel function of the second kind. The fugacity λ_i is defined as

$$\lambda_i(T, \vec{\mu}) = \exp \left(\frac{B_i \mu_B + S_i \mu_S + Q_i \mu_Q}{T} \right). \quad (57)$$

From the partition function Eq. (56) we can obtain the density of particle i :

$$n_i(T, \vec{\mu}) = \frac{\langle N_i \rangle}{V} = \frac{T g_i}{2\pi^2} \sum_{k=1}^\infty \frac{(\pm 1)^{k+1}}{k^2} \lambda_i^k m_i^2 K_2 \left(\frac{k m_i}{T} \right). \quad (58)$$

⁵We will keep the convention of upper signs for fermions and lower signs for bosons throughout the section.

To fully capture the production of particles in heavy ion collisions, however, we need to take resonance decays into lighter particles into account, such that the average number $\langle N_i \rangle$ of particles i in the Volume V at temperature T becomes

$$\langle N_i \rangle(T, \vec{\mu}) = \langle N_i \rangle^{\text{therm}}(T, \vec{\mu}) + \sum_j \Gamma_{j \rightarrow i} \langle N_j \rangle^{\text{therm}, R}(T, \vec{\mu}). \quad (59)$$

The first term is simply the thermal production, while the second one sums over all resonances j that decay into the particle i with a branching ratio $\Gamma_{j \rightarrow i}$. Particularly for lighter hadrons at higher temperatures these resonance contributions represent the dominant part of the total yield. In the region of large T or μ_B , where the density is high, the repulsive interactions between the hadrons can become important and need to be accounted for in the partition function, which is usually done by treating the hadrons as hard spheres, which leads to a shift in the baryon chemical potential. At low temperatures, typically below $T < 100 \text{ MeV}$, the widths of the resonances become important and the approximation by a δ function is no longer justified. The details of the implementation both of these corrections can be found in Ref. [122]. Fig. I.11 shows the yield of various particles as predicted in the SHM compared to ALICE data from lead-lead collisions at $\sqrt{s} = 2.76 \text{ TeV}$. There is a very good agreement across the entire mass scale even though the the y -axis of the plot spans nine orders of magnitude and the particles include not just mesons and baryons, but also (hyper-)nuclei. The large distance between the actual π prediction and the dashed line indicating purely thermal production highlights the importance of the previously mentioned resonance decays.

1.6.2. SHM with charmed hadrons

The SHM explained thus far works well for hadrons including any combination of the light up, down and strange quarks. Charm quarks, however, are significantly heavier and therefore the thermal production of these particles is strongly suppressed in heavy-ion collisions. Even at LHC energies, the thermal production can be neglected compared to the direct hard production [126, 127]. Therefore, the SHM can be extended towards charmed hadrons (SHMC) by a statistical recombination model [127], which assumes that all charm quarks are produced in the initial hard scattering and then equilibrate thermally in the QGP. Further, it is assumed that the QGP completely screens the color charges [128], such that no quarkonia, *i.e.*, $c\bar{c}$ states are produced within the volume V . The quarks instead thermalize in the QGP and hadronization only happens via the same statistical process as that of the lighter flavors. In this approach, the total number of open charm hadrons is then adjusted by an enhancement factor g_c to the number of $c\bar{c}$ pairs directly produced in

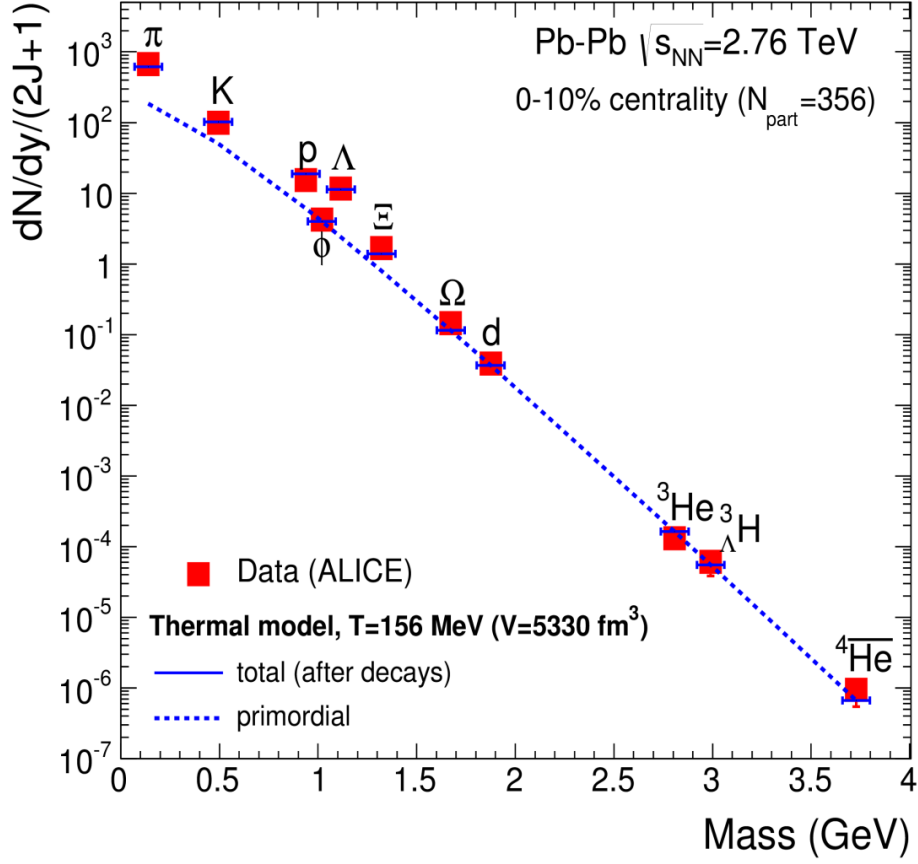


Figure I.11.: Yield of various particles measured in lead-lead collisions at $\sqrt{s} = 2.76$ TeV by ALICE compared to SHM predictions. Figure taken from Ref. [125]

the hard collision. This adjustment factor is defined by

$$N_{c\bar{c}}^{\text{direct}} = \frac{1}{2} g_c N_{oc}^{\text{therm}} \frac{I_1(g_c N_{oc}^{\text{therm}})}{I_0(g_c N_{oc}^{\text{therm}})}, \quad (60)$$

where the modified Bessel functions of the first kind I_n appear if the calculation is performed in the canonical ensemble. The switch to the canonical ensemble may be necessary depending on the beam energy, due to the number of (anti-)charm quarks being smaller than 1. In the grand-canonical limit, which should be reached at LHC energy, the ratio of the Bessel functions tends to unity. Equation (60) neglects contributions from charmonia due to their low relative abundance. Once the factor g_c is determined from this, the yield for a specific particle i is then given by

$$N_i = g_c N_i^{\text{therm}} \frac{I_1(g_c N_{oc}^{\text{therm}})}{I_0(g_c N_{oc}^{\text{therm}})}, \quad (61)$$

for hadrons with one (anti-)charm quark and $N_i = g_c^2 N_i^{\text{therm}}$ for those with two (anti-)charm quarks. In modern analyses [129], the number of directly produced charm quarks is taken from perturbative calculations of $\frac{d\sigma^{AA' \rightarrow c\bar{c}}}{dy}$, which are performed in the FONLL scheme (see Sec. II.3.1) and scaled by the nuclear overlap function. This cross section is where the nuclear PDFs for the nuclei A and A' enter. If both nuclei are heavy, *e.g.*, in the case of lead-lead collisions, this leads to a considerable uncertainty on g_c . Especially in the case of charmonium production, which scales with g_c^2 this causes the theoretical uncertainty to be far larger than the data, as illustrated in Figs. I.12 and I.13. Figure I.12 shows a similar spectrum as Fig. I.11, but for the SHMC and Fig. I.13 shows the ratio between J/ψ production in proton-proton and lead-lead collisions as predicted by the SHMC and measured by ALICE.

Assuming that the even heavier bottom quark also thermalizes in the QGP, the same approach can also be applied to hadrons with bottom quarks and bottomonia. This has been done in Ref. [43], where compatibility with experimental data was found in the case of central lead-lead collisions. However, as was the case with charmonia, the theoretical uncertainty on the enhancement factor g_b due to the nuclear PDFs is far larger than the data uncertainty. Therefore, more precise nuclear PDFs are required to test whether or not the SHM is compatible with data in the case of heavy quark production.

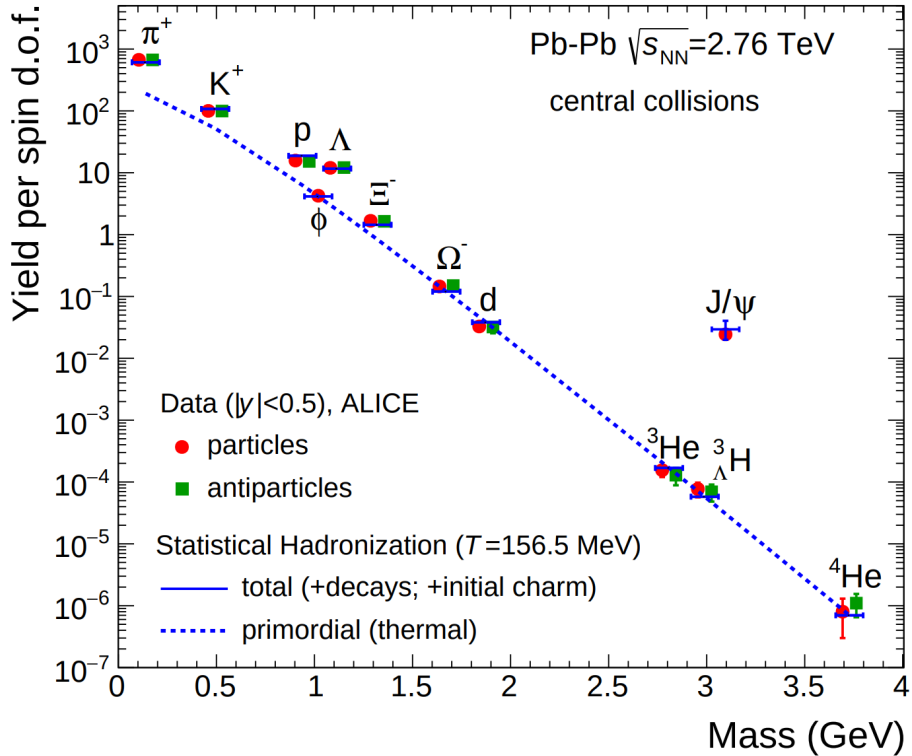


Figure I.12.: Yield of various particles measured in lead-lead collisions at $\sqrt{s} = 2.76$ TeV by ALICE compared to SHMC predictions. Figure taken from Ref. [130]

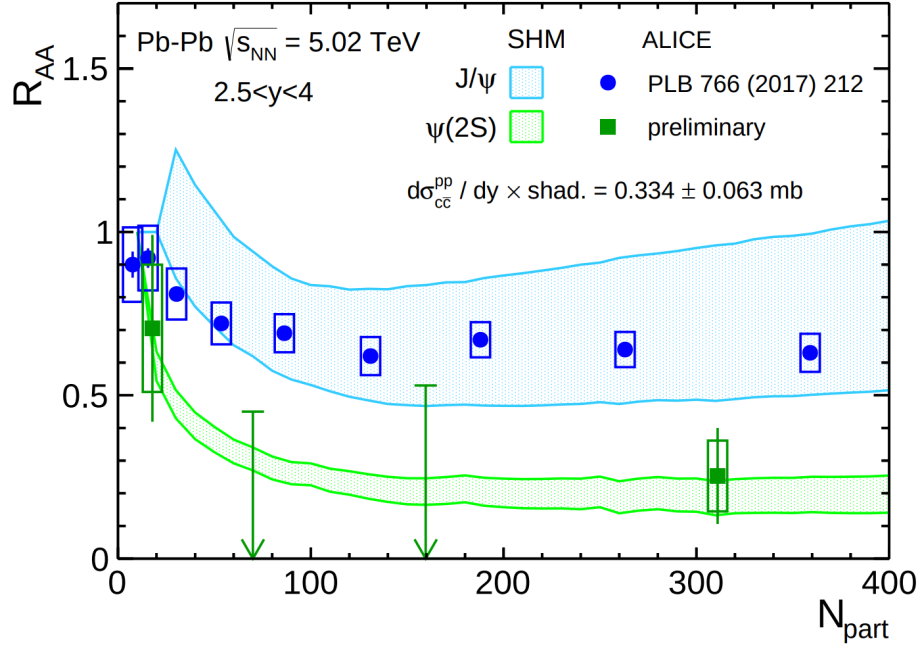


Figure I.13.: Nuclear modification factor R_{PbPb} of charmonia as measured by ALICE and predicted in the SHMC. Figure taken from Ref. [129]

2. Data modelling and the nCTEQ framework

As explained in detail in the previous section, the parton distribution functions are non-perturbative objects. Therefore, the only way to determine them is via lattice QCD calculations or empirically from available data of hadronic collisions. The former method has made steady progress over the recent years [28], but is still far away from reaching the necessary precision to make useful predictions for many observables measured in current experiments. Therefore, we focus on the latter approach. In principle, this process is quite simple and can be broken down into three steps:

1. Define a parameterization for the PDFs and choose a set of starting parameters.
2. Calculate observables from those PDFs and compare to available data.
3. Change the parameters until the optimal description of the data is found.

However, hidden inside these steps are significant complexities. For example, the difficulty of the second step should be obvious by now, given the derivation for one of the simplest observables at NLO in the preceding chapter, *i.e.*, the inclusive DIS cross section. The third step in particular can be broken down further into three components. At first, we need to define what it means for a prediction to give a good description of the data. This is done in the derivation of the so called χ^2 function in Sec. I.2.2, which is a measure of the likelihood that a given set of parameters accurately represents the data. Next, we will require an efficient way to traverse the highly multidimensional parameter space efficiently to find the best set of parameters, which is given by the BFGS algorithm explained in Sec. I.2.3. Once the optimal parameters are found, the question remains how large the uncertainties on these parameters are. This question is answered in the Hessian formalism in Sec. I.2.4, using a quadratic expansion of the χ^2 function around the optimal parameters. Before going on to these topics, however, we will quickly go over the first point regarding the parameterization in Sec. I.2.1.

2.1. PDF parameterization

The nCTEQ framework extends the methods used to perform global proton PDF fits into the nuclear realm. In the past, nuclear data was used to calculate correction factors which were then applied to proton PDF fits. In contrast, the nCTEQ framework is built such that there is direct communication between nuclear and proton data by taking the CTEQ proton PDF parameterization [131–133] and introducing nuclear A dependence into the parameters. The CTEQ approach parameterizes the PDFs at an initial scale $Q_0 = 1.3 \text{ GeV}$

as

$$\begin{aligned}
 x f_i^p(x, Q_0) &= c_0 x^{c_1} (1-x)^{c_2} e^{c_3 x} (1 + e^{c_4 x})^{c_5}, \\
 &\text{for } i = u_v, d_v, g, \bar{u} + \bar{d}, s + \bar{s}, s - \bar{s}, \\
 \frac{\bar{u}^p(x, Q_0)}{\bar{d}^p(x, Q_0)} &= c_0 x^{c_1} (1-x)^{c_2} + (1 + c_3 x)(1-x)^{c_4}.
 \end{aligned} \tag{62}$$

These quantities are then transformed into their equivalents of an effective bound proton⁶ $x f_i^{p/A}(x, Q_0)$ by introducing nuclear A dependence into the parameters

$$c_k \longrightarrow c_k(A) \equiv p_k + a_k(1 - A^{-b_k}). \tag{63}$$

The parameters p_k are then taken from the CTEQ6.1m [133], which is similar to CTEQ6.1, but with minimal influence of nuclear data in the proton fit. Some of the normalization coefficients c_0 can be determined from the sum rules (20)-(23) and can therefore be excluded from the fit. Furthermore, since the parameterization allows more flexibility than the available data can reasonably constrain, the nCTEQ15 fit is limited to the following set of open parameters:

$$\{a_1^{u_v}, a_2^{u_v}, a_4^{u_v}, a_5^{u_v}, a_1^{d_v}, a_2^{d_v}, a_5^{d_v}, a_1^{\bar{u}+\bar{d}}, a_5^{\bar{u}+\bar{d}}, a_1^g, a_4^g, a_5^g, b_0^g, b_1^g, b_4^g, b_5^g\}.$$

Additionally, the strange distribution is parameterized as

$$s^{p/A} = \bar{s}^{p/A} = \frac{\kappa(A)}{2} (\bar{u}^{p/A} + \bar{d}^{p/A}), \tag{64}$$

with

$$\kappa(A) = p_0^{s+\bar{s}} + a_0^{s+\bar{s}}(1 - A^{b_0^{s+\bar{s}}}). \tag{65}$$

The full nucleus is then given by the sum of the Z effective bound protons and $(A - Z)$ effective neutrons:

$$f_i^{(A,Z)}(x, Q) = \frac{Z}{A} f_i^{p/A}(x, Q) + \frac{A - Z}{A} f_i^{n/A}(x, Q), \tag{66}$$

where the neutron PDFs are determined assuming exact isospin symmetry. One more noteworthy detail is the fact that x is technically no longer limited to the interval $[0,1]$, but can instead span the interval $[0,A]$ in the nuclear case. However, since the structure functions fall off rapidly in this region [105, 134], this is generally neglected in nuclear PDF fits. An increasing number of these restrictions may be lifted as more data is included in

⁶It is important to stress that these effective bound protons are merely a tool for internal bookkeeping and sum rules. The quantities that are determined in the fit are the full nuclei.

the fits.

An illustration of the effect that each parameter has on the shape of the resulting PDF is given in Fig. 4.3 of Ref. [72].

2.2. The χ^2 cost function

The goal is to find the best possible PDFs for a given a set of N_D data points with measured values D_i and statistical uncertainty σ_i . First, we define $T_i(\mathbf{x})$ to be the theoretical prediction for the data at point i , given the PDFs defined by the vector of parameters \mathbf{x} ⁷. The optimal PDF parameters are then determined by minimizing a cost function χ^2 as described by the diagram shown in Fig. I.14.

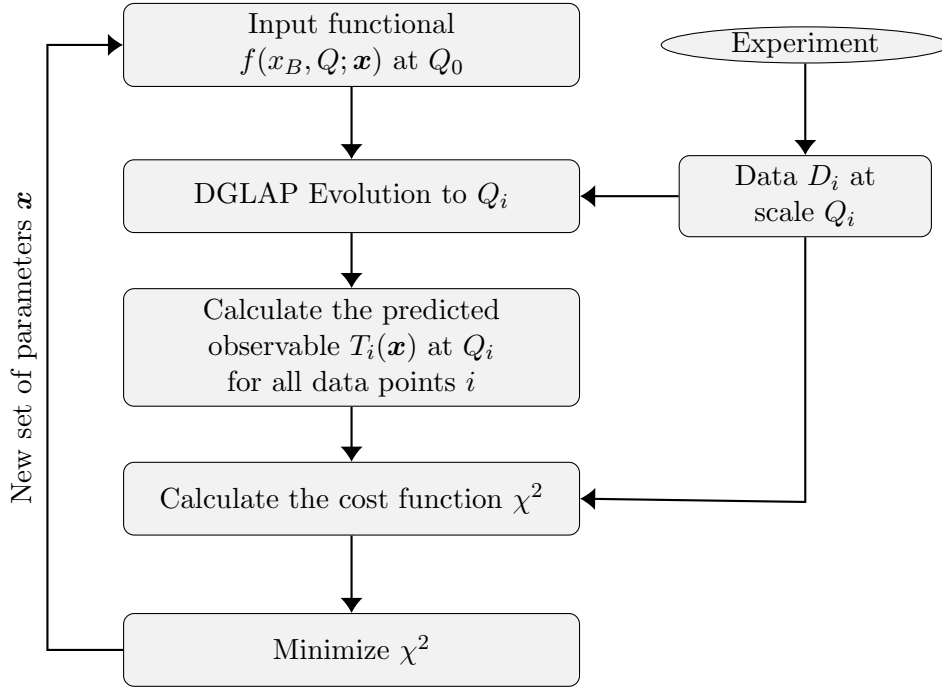


Figure I.14.: Flowchart representation of the PDF optimization.

The cost function χ^2 can then be derived from maximum likelihood estimation by assuming that the parameters that yield the highest likelihood of obtaining the measured data are the optimal choice. Assuming normal-distributed, independent measurement errors σ_i for each data point, the total probability of the data set is just the product of the points' individual probabilities

$$P \propto \prod_{i=1}^N \exp \left[-\frac{1}{2} \left(\frac{D_i - T(\mathbf{x})}{\sigma_i} \right)^2 \right] \frac{1}{\sqrt{2\pi}\sigma_i}. \quad (67)$$

⁷For the remainder of this section, we will denote the Bjorken- x as x_B to avoid confusion with the bold \mathbf{x} used for the vector of parameters.

Instead of maximizing this probability, it is more practical to minimize its negative logarithm. Dropping the constant term and adding a factor of 2 for normalization then gives us

$$\chi^2(\mathbf{x}) \equiv \sum_{i=1}^N \frac{[D_i - T(\mathbf{x})]^2}{\sigma_i^2} \quad (68)$$

as a cost function, which we can minimize to find the optimal parameters for our PDFs. In a global PDF fit, various data sets n of different experiments and physical processes are combined by using the weighted sum of their individual contributions χ_n^2 :

$$\chi^2 = \sum_n w_n \chi_n^2, \quad (69)$$

where the weights w_n can be used to (de)emphasize certain data sets.

Some data sets come with asymmetric uncertainties. In principle, it is straightforward to account for this in the calculation of the χ^2 by simply using the upper or lower uncertainty in Eq. 68 when the theory prediction is higher or lower than the data, respectively. This comes with technical difficulties, however, since the χ^2 function is assumed to be smooth in the minimization and uncertainty estimation. Therefore, the larger of the two uncertainties is generally used symmetrically in our analyses.

Correlated uncertainties

The naive definition of χ^2 given by Eq. (68) is generally insufficient when working with experimental data. To account for systematic uncertainties, we modify the definition χ_n^2 according to the prescription derived in Ref. [135]. Uncorrelated systematic uncertainties can be added in quadrature to the statistical uncertainties $\hat{\sigma}_i^2 \equiv \sigma_i^2 + u_i^2$. Correlated uncertainties, however, require special treatment. If we let β_{ik} denote the correlated error of point $i \in [1, \dots, N_{\text{data}}]$ due to a source $k \in [1, \dots, K]$, the correlated χ^2 for a data set n is given by

$$\chi_n^2(\mathbf{x}) = \sum_{i,j}^{N_{\text{data}}} (D_i - T_i) C_{ij}^{-1} (D_j - T_j), \quad (70)$$

with the covariance matrix

$$C_{ij} = \sigma_i^2 \delta_{ij} + \sum_{k=1}^K \beta_{ik} \beta_{jk}. \quad (71)$$

A special case of correlated uncertainty that requires further treatment is the normalization uncertainty. If there is an uncertainty associated with the normalization of a data set

this normalization can be included in the fit as an additional parameter. Here, we follow the prescription given in Ref. [136], where the fitted normalization factor r for each data set is applied by multiplying the theory prediction with its inverse and the covariance matrix is extended to:

$$\hat{C}_{ij} = \sigma_i^2 \delta_{ij} + \sum_{k=1}^K \beta_{ik} \beta_{jk} + \sigma_{\text{norm}} D_i D_j. \quad (72)$$

A penalty term is then added to keep the fitted normalization factor reasonably close to unity and to avoid d'Agostini bias [136], which causes significantly lower predictions than the data actually suggests. The resulting χ_n^2 for each data set is then given by

$$\chi_n^2(\mathbf{x}) = \sum_{i,j}^{N_{\text{data}}} \left(D_i - \frac{T_i}{r} \right) \hat{C}_{ij}^{-1} \left(D_j - \frac{T_j}{r} \right) + \left(\frac{1-r}{\sigma_{\text{norm}}} \right)^2. \quad (73)$$

A minor improvement of this procedure will be presented later on in Sec. I.2.4.3.

2.3. Minimization

Now that the problem of finding the optimal PDFs given a set of data has been formulated into an optimization problem with a cost function χ^2 , we need a reliable way to minimize this cost function. Considering the high dimensionality of the parameter space this is no simple task. In general, minimization algorithms can be classified into two main groups: global and local minimizers. The former class is generally built on statistical approaches while the latter uses deterministic approaches. As the name implies global minimization algorithms converge towards the global minimum of the function as the number of iterations approaches infinity. Sadly, in practical applications we do not have an infinite amount of time and therefore local minimizers are commonly used because they are faster at converging to a local minimum. In practice, the local minima lie close together in terms of cost function and described physics. Therefore, local optimizers are currently the standard among various (n)PDF groups [9, 18], since they are generally significantly faster and the global minimization property of stochastic methods is rarely realized in practice since the algorithms are stopped once they find any minimum to save time. A simple strategy to mitigate the problem of local minima is to start multiple minimizations from different initial parameters and then compare the resulting minima.

Broyden-Fletcher-Goldfarb-Shanno algorithm

The Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm [59–62] is a method to iteratively solve a nonlinear optimization problem. The BFGS algorithm belongs to the quasi-

Newton subgroup of “hill-climbing” optimization methods. In this context “hill-climbing” describes techniques where the value of the optimized function moves closer to a local optimum in each iteration. The BFGS method is considered the successor of the very similar Davidon-Fletcher-Powell (DFP) method [137], which is the underlying algorithm of the commonly used MIGRAD minimizer [138, 139].

Basic algorithm

Given the task of minimizing a function $f(\mathbf{x})$,⁸ the fundamental strategy of all (quasi-) Newton methods is to iteratively search along directions \mathbf{p}_k , which are obtained by solving

$$H_k \mathbf{p}_k = -\nabla f(\mathbf{x}_k), \quad (74)$$

where H_k is the Hessian matrix at the parameters current parameters \mathbf{x}_k . BFGS then uses a line-search to find the minimum of $f(\mathbf{x}_k + c \cdot \mathbf{p}_k)$. Instead of redoing the computation of the Hessian at the new parameters, it is then updated by adding two symmetric rank-one matrices

$$H_{k+1} = H_k + U_k + V_k. \quad (75)$$

This step places the BFGS algorithm into the quasi-Newton category, where the Hessian matrix is approximated throughout the iterations, as opposed to the full Newton category, where the Hessian is computed explicitly in each iteration. The sum of $U_k = c_1 \mathbf{u} \mathbf{u}^T$ and $V_k = c_2 \mathbf{v} \mathbf{v}^T$ is a rank-two update matrix, which separates the BFGS (and DFP) method from simpler rank-one methods like Broyden’s method [140] and SR1 [141] that do not guarantee positive definiteness of the Hessian. We choose $\mathbf{u} = \mathbf{y}_k \equiv \nabla f(\mathbf{x}_{k+1}) - \nabla f(\mathbf{x}_k)$ and $\mathbf{v} = H_k \mathbf{s}_k \equiv H_k(\mathbf{x}_{k+1} - \mathbf{x}_k)$ and impose the secant condition

$$H_{k+1} \mathbf{s}_k = \mathbf{y}_k. \quad (76)$$

This gives us

$$c_1 = \frac{1}{\mathbf{y}_k^T \mathbf{s}_k} \quad \text{and} \quad c_2 = -\frac{1}{\mathbf{s}_k^T H_k \mathbf{s}_k}. \quad (77)$$

Plugging this into Eq. (75), we obtain the final update equation of the approximate Hessian:

$$H_{k+1} = H_k + \frac{\mathbf{y}_k \mathbf{y}_k^T}{\mathbf{y}_k^T \mathbf{s}_k} - \frac{H_k \mathbf{s}_k \mathbf{s}_k^T H_k^T}{\mathbf{s}_k^T H_k \mathbf{s}_k}. \quad (78)$$

⁸Note that $f(\mathbf{x})$ would not be the parton distribution in case of PDF fitting, but instead it would be the χ^2 function.

It is important to note that this approximate Hessian does not necessarily converge towards the real Hessian. The entire algorithm can be visualized as seen in Fig. I.15. At this point three questions remain unanswered:

1. How do we determine the initial Hessian?
2. How do we perform the line search?
3. How do we know when to stop?

The former two questions are answered in the following two paragraphs. The answer to the last question is simply to count the number N of consecutive iterations that yielded only small improvements $(f(\mathbf{x}_k) - f(\mathbf{x}_{k+1}))/f(\mathbf{x}_{k+1}) < \epsilon$. We then end once this counter reaches a number that makes us confident that any further improvements would not exceed the threshold. Common choices are $N = 5$ and $\epsilon = 0.001$.

Initial Hessian

Probably the most obvious choice here would be to calculate the Hessian via finite differences. Introducing the notation $\mathbf{h}_i = h_i \mathbf{e}_i$, where \mathbf{e}_i is the unit vector in the i -th direction and h_i the step size for the i -th parameter, we can express the finite differences approximation for the Hessian H as

$$H_{ii}(\mathbf{x}) \approx \frac{f(\mathbf{x} + \mathbf{h}_i) - 2f(\mathbf{x}) + f(\mathbf{x} - \mathbf{h}_i)}{h_i^2}, \quad (79)$$

$$H_{ij}(\mathbf{x}) = H_{ji}(\mathbf{x}) \approx \frac{1}{2h_i h_j} [f(\mathbf{x} + \mathbf{h}_i + \mathbf{h}_j) + f(\mathbf{x} - \mathbf{h}_i - \mathbf{h}_j) + 2f(\mathbf{x}) - f(\mathbf{x} + \mathbf{h}_i) - f(\mathbf{x} + \mathbf{h}_j) - f(\mathbf{x} - \mathbf{h}_i) - f(\mathbf{x} - \mathbf{h}_j)]. \quad (80)$$

Even in this leading order, this offers a very good precision, which leads to large improvements in the first iterations of the algorithm. The drawback of this method is that the number of function evaluations is $n^2 + 3n$ for n parameters, which is computationally expensive for large n .

Another possible choice is the identity matrix. This eliminates the overhead of the initialization and causes the first iteration follow direction of the gradient. The Hessian is then slowly approximated as the algorithm iterates.

When minimizing a χ^2 function we have a third option: Approximating the Hessian via the Jacobi matrix as $H \approx J^T J$, which is valid for maximum likelihood estimators [142] and requires only $2n$ function evaluations. In practice, this method requires only a few more iterations to reach the same performance as the first method and the cost of these iterations is easily offset by the saved $n^2 + n$ extra evaluations required for the full finite

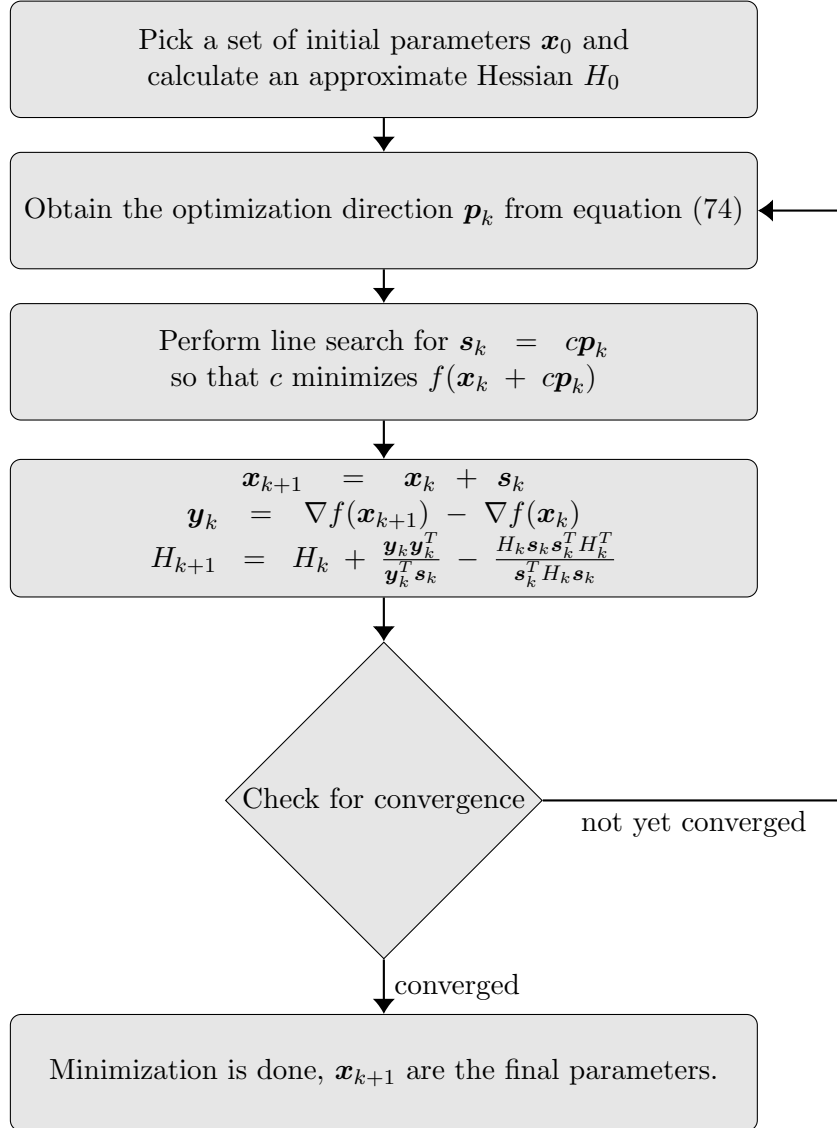


Figure I.15.: Flowchart representation of the BFGS algorithm.

differences Hessian. The only case where the full Hessian might be preferable is for a low number of parameters, where n^2 is still small.

Golden-section search

To perform the line search for the optimal distance along the vector s_k , we use the Golden-section search. This algorithm requires a bound interval and reduces the size of this interval by a factor of $\varphi - 1$ in each iteration, hence the name “Golden-section search”. To achieve this, we first evaluate the function at four points: the boundaries⁹ a and b and two intermediate points

$$c = b - \frac{(b - a)}{\varphi} \quad (81)$$

$$d = a + \frac{(b - a)}{\varphi}. \quad (82)$$

We then replace b by d if $f(d) > f(c)$ or replace a by c , otherwise. A new point is then added such that the relative size of the intervals between the points a and b remain the same as the initial ones. This is repeated until the interval becomes sufficiently small compared to the length of the initial interval $\|c - d\| > \epsilon \|c_0 - d_0\|$. We choose a relative tolerance ϵ because the size of the initial boundary can vary by many orders of magnitude. A simple pseudocode implementation of the algorithm is given in Alg. 1.

Since either point c or d gets updated to the value that the other one previously held, we can perform each iteration with a single function evaluation, if we save the values of $f(c)$ and $f(d)$.

The distance walked along the direction \mathbf{p}_k in the BFGS algorithm is not naturally bounded, but we know that $a = \mathbf{0}$ is the lowest possible value because negative values would go opposite to the direction that we determined in Eq. (74). We can estimate b by starting with a small value (*i.e.* $b = 10^{-8}\mathbf{p}_k$) and increasing it by a factor 10 until $f(b) > f(0)$ so that a minimum necessarily lies between $a = 0$ and b .

2.4. Estimation of uncertainties

Once the optimal PDFs have been determined from the data the natural next question is how large the uncertainties of these PDFs and any derived quantities are. Answering this question, however, is not straightforward. Problems arise mainly from the following four sources:

- Errors in the theory used to predict the measured observables, *e.g.* due to missing higher-order corrections¹⁰ and power suppressed corrections

⁹Note that the boundaries may be vector valued, in which case the search is performed along the straight line spanned by the two points in the vector space.

¹⁰Note that there are techniques to estimate the size of these effects [143]

Algorithm 1 Golden-section search

```

1: procedure GOLDEN-SECTION SEARCH( $f, a, b$ )            $\triangleright f : \mathbb{R}^n \rightarrow \mathbb{R}; \quad a, b \in \mathbb{R}^n$ 
2:    $\phi = \frac{1+\sqrt{5}}{2}$ 
3:    $\epsilon = 0.001$ 
4:    $c \leftarrow b - \frac{(b-a)}{\phi}$ 
5:    $d \leftarrow a + \frac{(b-a)}{\phi}$ 
6:   while  $\|c - d\| > \epsilon \cdot \|c_0 - d_0\|$  do            $\triangleright$  End if interval is small compared
                                                    to the initial one
7:     if  $f(c) > f(d)$  then
8:        $b \leftarrow d$ 
9:        $d \leftarrow c$ 
10:       $c \leftarrow b - \frac{(b-a)}{\phi}$ 
11:    else
12:       $a \leftarrow c$ 
13:       $c \leftarrow d$ 
14:       $d \leftarrow a + \frac{(b-a)}{\phi}$ 
15:    end if
16:  end while
17:  return  $\frac{(a+b)}{2}$             $\triangleright$  The minimum is at  $\frac{(a+b)}{2}$ 
18: end procedure
    
```

- Published data sets often provide their uncertainties only in uncorrelated form with a single normalization uncertainty
- Published errors of some experiments fail standard statistical tests, *e.g.*, having χ^2 differ significantly from 1 per degree of freedom, making them “improbable”
- Different data sets are unproblematic individually but require mutually incompatible PDFs

As such, multiple approaches to determine the uncertainties of PDFs have been proposed. One method of handling the estimation of the uncertainties is the Hessian approach [63] based on a quadratic approximation of χ^2_{global} near its minimum. Other possible approaches are the Lagrange multiplier method [135] and Monte-Carlo replicas [144]. The Lagrange multiplier method has the advantage of not requiring any approximations at the cost of not providing complete information about the minimum’s neighbourhood. Monte-Carlo replicas offer similar information to the Hessian approach, but are generally more computationally expensive, both in their initial generation and in their later use. In this section, we will focus on the Hessian approach, which is used in the nCTEQ framework and is common among other PDF analyses [14, 17, 18].

2.4.1. The Hessian method

The theory is determined by a set of d free parameters $\{x_i\} = \{x_1, \dots, x_d\}$ that model the non-perturbative input and characterize a PDF set $S(\mathbf{x})$. Given the set of parameters \mathbf{x}_0 that minimizes the χ^2 function, we can define the range of acceptable parameters \mathbf{x} by imposing a tolerance T such that

$$\Delta\chi^2(\mathbf{x}) \equiv \chi^2(\mathbf{x}) - \chi^2(\mathbf{x}_0) \leq T^2. \quad (83)$$

The Hessian method then provides an efficient way of exploring the variation of the predictions made using the PDFs in this neighbourhood, as long as the displacement in parameter space $\mathbf{y} = \mathbf{x} - \mathbf{x}_0$ fulfills the quadratic approximation

$$\Delta\chi^2 = \chi^2 - \chi_0^2 \approx \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d H_{ij} y_i y_j, \quad (84)$$

$$\text{with } H_{ij} = \left(\frac{\partial^2 \chi^2}{\partial y_i \partial y_j} \right) \bigg|_{\mathbf{x}=\mathbf{x}_0}. \quad (85)$$

This is done by transforming the parameter space from its original basis into an orthonormal basis spanned by the eigenvectors of the Hessian (hence the name) as illustrated in Fig. I.16 for the case of a two dimensional parameter space. The eigenvectors are rescaled iteratively such that the space of parameters \mathbf{z} , fulfilling $\Delta\chi^2(\mathbf{z}) \leq T^2$, becomes a d -dimensional hypersphere.

Formally, a complete set of orthonormal eigenvectors $\{\mathbf{v}_k\}$ with elements v_{ik} of the Hessian H is defined by

$$\sum_{j=1}^d H_{ij} v_{jk} = \epsilon_k v_{ik} \quad \text{and} \quad \sum_{i=1}^d v_{il} v_{ik} = \delta_{lk}, \quad (86)$$

where ϵ_k is the eigenvalue corresponding to eigenvector \mathbf{v}_k and δ_{lk} is the unit matrix. The displacements in parameter space from the minimum can then be expressed in terms of these eigenvectors as

$$y_i = \sum_{k=1}^d v_{ik} s_k z_k, \quad (87)$$

with the scale factors s_k chosen such that the new parameters z_k fulfill

$$\Delta\chi^2 = \sum_k z_k^2, \quad (88)$$

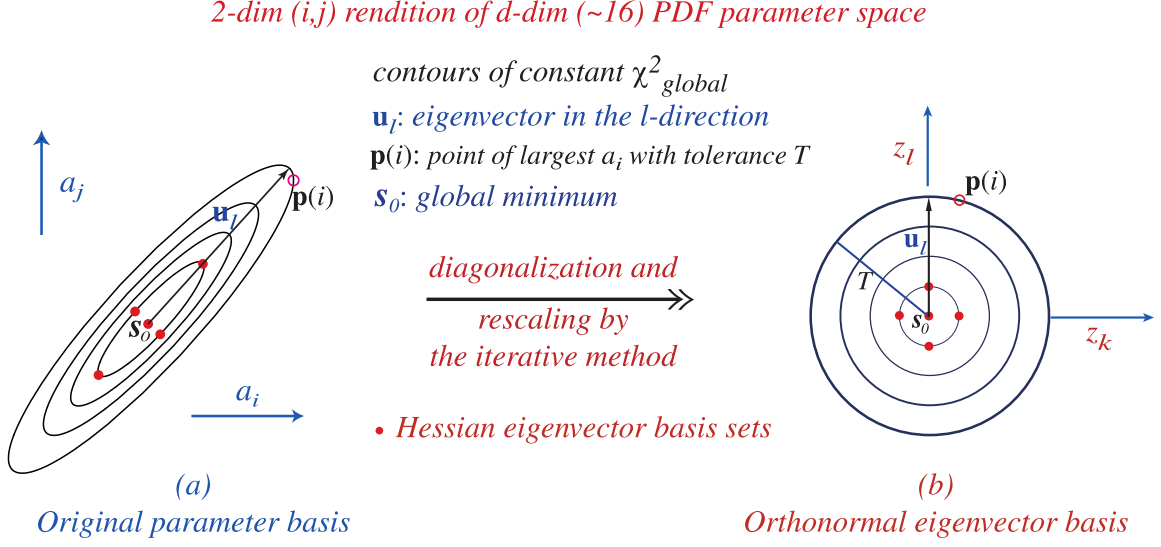


Figure I.16.: Illustration of the basic idea behind the Hessian formalism. The original parameter basis is transformed into an orthonormal basis spanned by the eigenvectors of the Hessian, making the contours of constant χ^2 surfaces of hyperspheres. Figure taken from Ref. [63].

which gives us the hyperspherical property of the constant $\Delta\chi^2$ contours. If the data is following perfect normal distributions, these scaling factors would be $s_k = \sqrt{\frac{1}{\epsilon_k}}$. However, since this is never the case in practice, they are determined by varying them iteratively until Eq. (88) is fulfilled.

Eigenvector basis PDF sets

We define the transformation matrix M_{ik} between the original basis and our new basis as

$$M_{ik} \equiv v_{ik}s_k, \quad (89)$$

such that

$$x_i - x_i^0 = \sum_{k=1}^d M_{ik} z_k. \quad (90)$$

This allows us to define eigenvector PDFs S_i^\pm given by the parameters

$$z_k^{\text{EV basis}}(S_i^\pm) = \pm T \delta_{ki} \quad (91)$$

$$\Rightarrow x_k^{\text{orig. basis}}(S_i^\pm) = x_{0,k}^{\text{orig. basis}} \pm T M_{ki}. \quad (92)$$

These eigenvector PDFs can be used to calculate the uncertainties of any observable $\mathcal{O}(S)$ depending on a PDF set S . For this purpose, it is useful to define the following three quantities:

$$D_k(\mathcal{O}) = \mathcal{O}(S_k^+) - \mathcal{O}(S_k^-), \quad (93)$$

$$D(\mathcal{O}) = \sqrt{\sum_{k=1}^d (D_k(\mathcal{O}))^2}, \quad (94)$$

$$\hat{D}_k(\mathcal{O}) = \frac{D_k(\mathcal{O})}{D(\mathcal{O})}, \quad (95)$$

so that $D_k(\mathcal{O})$ is the gradient vector corresponding to parameter z_k , while $\hat{D}_k(\mathcal{O})$ is the unit vector of the same direction. The largest permissible variation in \mathcal{O} is obtained by varying it along every gradient direction by a distance $\pm T$ from the minimum. The uncertainty $\Delta\mathcal{O}$ of any observable including the PDFs themselves can therefore be expressed in terms of these quantities as

$$\Delta\mathcal{O} = \sum_{k=1}^d (T \hat{D}_k) \frac{\partial \mathcal{O}}{\partial z_k}. \quad (96)$$

By approximating the gradient of the observable through the difference equation

$$\frac{\partial \mathcal{O}}{\partial z_k} = \frac{\mathcal{O}(S_k^+) - \mathcal{O}(S_k^-)}{2T}, \quad (97)$$

the uncertainty can be written as a simple master formula

$$\Delta\mathcal{O} = \frac{1}{2} D(\mathcal{O}). \quad (98)$$

The uncertainties for the original parameters can then be calculated as

$$\Delta x_i = T \sqrt{\sum_k M_{ik}^2} \quad (99)$$

and their values x_i^{ext} resulting in the extremal values of an observable are then given by

$$x_i^{\text{ext}} = x_i^0 \pm \frac{1}{2} \sum_{k=1}^d \hat{D}_k(X) [x_i(S_k^+) - x_i(S_k^-)]. \quad (100)$$

2.4.2. Choosing the tolerance T

In the previous chapter we have conveniently ignored the question of how to choose the tolerance parameter T , which we will rectify now. By including any data set in the fit, we

implicitly assume statistical validity of its uncertainty and compatibility with the other data sets. To make sure the tolerance accounts for these assumptions, we start by investigating the individual χ_k^2 of an experiment k with N_k data points.

The probability distribution of of this χ^2 is given by ¹¹

$$P(\chi_k^2, N_k) = \frac{(\chi_k^2)^{N_k/2-1} e^{-\chi_k^2/2}}{2^{N_k/2} \Gamma(N_k/2)}, \quad (101)$$

allowing us to define the percentile confidence levels ξ_p via

$$\int_0^{\xi_p} P(\chi^2, N) d\chi^2 = p\% \quad \text{where} \quad p = \{50, 90, 99\}. \quad (102)$$

The mean of the χ^2 distribution is then given by ξ_{50} , while ξ_{90} gives the value above which there is only a 10% chance that a fit with such an χ_k^2 properly describes the data. Since the minimum of the fit does not necessarily coincide with the minimum of any given data set due to individual fluctuations or incompatibility with other sets, we rescale the ξ_{90} percentile to account for the position of the minimum

$$\tilde{\xi}_{90} \longrightarrow \xi_{90} \left(\frac{\chi_{0,k}^2}{\xi_{50}} \right). \quad (103)$$

For each eigenvector direction i , this can be used to define an interval

$$z_i^{(k)-} \leq \tilde{z}_i \leq z_i^{(k)-} \quad (104)$$

for the associated parameter \tilde{z}_i , where χ_k^2 stays within the 90% confidence limit given by $\tilde{\xi}_{90}$. The interval where all experiments stay within their respective 90% C.L. is then given by

$$z_i^- \equiv \max_k \left(z_i^{(k)-} \right) \leq \tilde{z}_i \leq z_i^+ \equiv \min_k \left(z_i^{(k)+} \right). \quad (105)$$

For a set of n fit parameters, the global tolerance is then defined as the average of these $2n$ eigenvector tolerances:

$$T^2 = \Delta\chi_{\max}^2 \equiv \sum_i^n \frac{(z_i^-)^2 + (z_i^+)^2}{2n}. \quad (106)$$

¹¹Technically, N_k in should be the number of degrees of freedom, but due to the number of data points typically being significantly larger than the number of parameters in the fit, using the number of data points is a very reasonable approximation.

2.4.3. Treatment of nuisance parameters

Special care has to be given to so called *nuisance parameters*, which may be included in the fit but do not affect the resulting PDFs directly. An example of this are normalization parameters for individual data sets. These parameters only affect the χ^2 function, but the actual PDFs are invariant with regards to them. Including them in the Hessian analysis would, therefore, create an eigenvector basis that is not orthonormal, as there would be more vectors than actual dimension of parameter space. The proper way to treat these nuisance parameters can be found in Ref. [145] and works as follows: Without loss of generality, we can assume that the vector of parameters \mathbf{x} contains n regular parameters first and then m nuisance parameters. The naive treatment that is used in many analyses, *e.g.* nCTEQ15, would then be to just ignore these parameters and consider only the $n \times n$ submatrix of the Hessian. The more proper way, however, is to invert the Hessian first, then prune the last m rows and columns and finally invert it again. This way, the correlation of the true parameters with the nuisance parameters is preserved without inflating the number of eigenvectors.

Equivalent in terms of statistical interpretation, but more elegant numerically is the following method. If we define

$$\chi_n^2(\mathbf{x}) = \sum_{i,j}^{N_{\text{data}}} (D_i - rT_i) \hat{C}_{ij}^{-1} (D_j - rT_j) + \left(\frac{1-r}{\sigma_{\text{norm}}} \right)^2, \quad (107)$$

with the covariance matrix from Eq. (72), we can derive an analytical solution for the optimal normalization parameter r_0 for each data set:

$$r_0 = \frac{1 + \sigma_{\text{norm}}^2 D^T C^{-1} T}{1 + \sigma_{\text{norm}}^2 T^T C^{-1} T}. \quad (108)$$

Note that the covariance matrix C in Eq. (108) is the one from Eq. (71) without the normalization term. Using these normalization parameters, our χ^2 effectively becomes

$$\chi_n^2(\mathbf{x}) = \sum_{i,j}^{N_{\text{data}}} (D_i - T_i) \bar{C}_{ij}^{-1} (D_j - T_j), \quad (109)$$

with

$$\bar{C}_{ij} = \sigma_i^2 \delta_{ij} + \sum_{k=1}^K \beta_{ik} \beta_{jk} + \sigma_{\text{norm}}^2 T_i T_j. \quad (110)$$

This removes the need for explicit fitted normalization parameters, which drastically speeds up the minimization procedure while retaining the impact of the normalization uncertainty on the Hessian.

2.4.4. Monte Carlo replicas

A different method of determining the uncertainties of a PDF fit is the creation of an ensemble of artificial pseudodata sets called *Monte Carlo replicas* [146]. Given a data set of observables \mathcal{O}_i with uncorrelated uncertainties σ_i and K correlated uncertainties β_{ik} , the replicas are generated as

$$\mathcal{O}_i^{(n)} = \mathcal{O}_i \left(1 + r_i^{(n)} \sigma_i + \sum_{k=1}^K r_{i,k}^{(n)} \beta_{ik} \right), \quad (111)$$

where the index i runs over the data points and $r_i, r_{i,k}$ are univariate Gaussian random numbers. The index n denotes the n -th replica. Correlations between data points i and i' can be introduced by setting $r_{ik} = r_{i'k}$. The central values, variances and covariances evaluated from these replicas then reproduce those of the initial data set. The PDF fit is then performed for each set of replicas, resulting in a collection $\boldsymbol{x}^{(n)}$ of optimal parameter sets for each replica. The overall fit quality is then evaluated by calculating the χ^2 using the real data set and the mean of the fit replicas. The PDF uncertainty is calculated by taking the standard deviation of the replica PDFs and similarly, the uncertainty for any observable is obtained by the observable with every replica and then calculating the standard deviation of those.

In contrast to the Hessian method, this procedure does not rely on a quadratic approximation of the χ^2 function around the minimum, but comes at the cost of significantly increased computational cost that originates from performing fits to possibly hundreds of replicas instead of just one data set. As the number of parameters increases, however, this may change as the number of Hessian eigenvectors rises and the cost of evaluating the Hessian matrix rises with the square of this number.

2.5. Optimizations

Performing PDF fits and the various analyses of the resulting PDFs properties can take many hours or even days to complete. To shorten this time requirement, a variety of optimizations are used in the nCTEQ framework. Two important examples are K -factors and grids.

2.5.1. K-factors

To speed up the calculation of higher order theory predictions they can be approximated by combining their leading order computation with a precomputed K -factor. Given a data point of observable \mathcal{O} measured at some kinematics a , we calculate a K -factor for this point as

$$K(a; \mathbf{x}) = \frac{\mathcal{O}^{\text{NLO}}(a; \mathbf{x})}{\mathcal{O}^{\text{LO}}(a; \mathbf{x})}. \quad (112)$$

The NLO result for a different set of PDF parameters $\mathbf{x} + \delta\mathbf{x}$ is then approximated by

$$\mathcal{O}^{\text{NLO}}(a; \mathbf{x} + \delta\mathbf{x}) \approx K(a; \mathbf{x}) \mathcal{O}^{\text{LO}}(a, \mathbf{x} + \delta\mathbf{x}). \quad (113)$$

Naturally, the approximation $K(a; \mathbf{x}) \approx K(a; \mathbf{x} + \delta\mathbf{x})$ is only reasonable as long as $\delta\mathbf{x}$ is sufficiently small, which makes it useful, especially when calculating derivatives via finite differences. Whenever the parameters are changed by a substantial amount, the K -factors need to be updated to avoid sacrificing precision. The previously explained BFGS algorithm is particularly well suited to benefit from K -factors since most function evaluations happen when evaluating the gradients using finite differences, where the approximation is well justified. By performing the line search with the full theory, one can retain optimal precision with minimized cost.

Naturally, K -factors are not limited to NLO, but can be used to approximate arbitrary orders in perturbation theory.

2.5.2. Gridding

Gridding is another way to optimize the runtime of the computation at a minor cost in accuracy. It is applicable whenever integral transforms

$$F(a) = \int_{x_1}^{x_2} f(x, a) g(x) dx \quad (114)$$

are performed with different values of a . If the evaluation of $g(x)$ is computationally expensive, *e.g.*, when it requires evaluation of an integral itself, the time can be severely cut down by replacing the function $g(x)$ with a grid G and an interpolation routine $I(x; G)$. The grid G is just a table of function values $G(x_i) = g(x_i)$, where $x_i \in X$ and X is an ordered set of sampling points with $x_n < x_{n+1}$. The interpolation function $I(x; G)$ is a prescription on how to handle values $x \notin X$. The simplest interpolation that guarantees continuity is linear interpolation:

$$I_{\text{lin}}(x; G) = \frac{G(x_{n+1}) \cdot (x_{n+1} - x) + G(x_n) \cdot (x - x_n)}{x_{n+1} - x_n}, \quad (115)$$

where x_n is chosen such that $x_n \leq x < x_{n+1}$. Note that this interpolation scheme works for any set of sampling points X , as long as its elements fulfill $x_n < x_{n+1}$. The two most common sampling methods are uniformly distributed and logarithmically distributed along the domain. In principle, this can easily be generalized to multidimensional grids by applying the interpolation recursively to one dimension after another. With increasing dimensions, however, the grid size and therefore memory cost and initial computation time rise exponentially.

2.6. Correlation between data sets and PDFs

By looking at the PDFs alone, we cannot judge the impact of each individual new data set on the fit. For this purpose, we can make use of a variety of further methods to study how each data set impacts the PDFs of particular flavours or even individual PDF parameters.

2.6.1. Cosine of correlation angle

The first quantity we may want to analyze is the cosine of the correlation angle between two observables X and Y , as used in Refs. [23, 63]. It is defined as

$$\cos(\phi[X, Y]) = \sum_{i=1}^{N_{\text{EV}}} \frac{(X_i^{(+)} - X_i^{(-)}) (Y_i^{(+)} - Y_i^{(-)})}{\sqrt{\sum_{i'=1}^{N_{\text{EV}}} (X_{i'}^{(+)} - X_{i'}^{(-)})^2} \sqrt{\sum_{i''=1}^{N_{\text{EV}}} (Y_{i''}^{(+)} - Y_{i''}^{(-)})^2}}, \quad (116)$$

where the indices i, i' and i'' of the sums each run over all eigenvector directions of the Hessian and X_i^{\pm} indicates the observable X evaluated with the i -th positive or negative eigenvector parameters. Typically, one of the observables would be the PDF of a certain flavour and the other one the χ^2 value of a given data set. In that case, the cosine of the correlation angle indicates which data sets drive the PDF in a certain direction. A weakness of this quantity is the fact that there is no indication to the strength of the impact that each data set has, as the denominator cancels out the normalization of the numerator.

2.6.2. Effective χ^2 difference

A similar useful quantity is the effective $\Delta\chi^2$ as introduced in Ref. [9]. It is designed to be used complementary with the cosine of the correlation angle and to indicate how strongly a data set j_{exp} affects the observable X

$$\Delta\chi_{\text{eff}}^2[X, j_{\text{exp}}] = \frac{1}{2N_{\text{EV}}} \sum_{i'=1}^{N_{\text{EV}}} \left(X_{i'}^{(+)} - X_{i'}^{(-)} \right)^{-2} \sum_{i=1}^{N_{\text{EV}}} \left[\left(X_i^{(+)} - X_i^{(-)} \right)^2 \times \left(\left| \chi_i^{2(+)}(j_{\text{exp}}) - \chi_i^{2(0)}(j_{\text{exp}}) \right| + \left| \chi_i^{2(-)}(j_{\text{exp}}) - \chi_i^{2(0)}(j_{\text{exp}}) \right| \right) \right]. \quad (117)$$

Again, the indices i and i' of the sums each run over all eigenvector directions of the Hessian and X_i^{\pm} indicates the observable X evaluated with the i -th positive or negative eigenvector parameters. The fact that only the observable X is in the denominator here means that the number of data points and the size of the uncertainties directly affects the value of $\Delta\chi_{\text{eff}}^2$. This means that a high $\Delta\chi_{\text{eff}}^2[X, j_{\text{exp}}]$ indicates a strong impact of a data set j_{exp} on the observable X .

2.6.3. Parameter scans

A very simple, yet powerful tool are parameter scans. Given the vector of parameters \mathbf{x}_0 at the minimum, it is often instructive to plot

$$\Delta\chi^2(j_{\text{exp}}, a, i) \equiv \chi^2(j_{\text{exp}}; \mathbf{x}_0 + a\mathbf{e}_i) - \chi^2(j_{\text{exp}}; \mathbf{x}_0), \quad (118)$$

where \mathbf{e}_i is the unit vector along parameter i for a range $[a_{\text{min}}, a_{\text{max}}]$ such that the total χ^2 deviation

$$\Delta\chi^2(a_{\text{min/max}}, i) = \sum_{j_{\text{exp}}} \Delta\chi^2(j_{\text{exp}}, a_{\text{min/max}}, i) \quad (119)$$

reaches some fixed value, *e.g.*, some multiple of the tolerance T used in the Hessian calculations. Doing this should yield parabolas for each experiment showing exactly how strong they are constraining a parameter and which value of the parameter would be optimal for any given experiment. Plotting these parabolas separated into contributions from the different data sets/processes can give information on which parts of the data sets constrain each parameter and which data sets may be in conflict with each other.

Of course, these scans can be performed along arbitrary directions in parameter space, not just along individual parameters. However, in those cases the physical interpretation of the results might be difficult. A common use for this are scans along the Hessian eigenvectors to validate the quadratic approximation of the χ^2 function.

2.7. Bayesian reweighting

Bayesian reweighting is useful to estimate the impact of a data set on a potential fit without performing an entirely new global analysis. The method was originally derived in Ref. [144],

but we outline the updated version from Ref. [40].

The first step of the procedure is the conversion of the Hessian eigenvector PDFs into a set of equivalent Monte Carlo replicas, since the Hessian eigenvectors do not have a direct probabilistic interpretation. These replicas are given by

$$f_k = f_0 \sum_{i=1}^{N_{\text{ev}}} \frac{f_i^+ - f_i^-}{2} R_{ki}, \quad (120)$$

where f_k are the replicas, f_i the eigenvector PDFs, and R_{ki} normally-distributed random numbers centered around 0 with a standard deviation of 1. The number of replicas N_{rep} is arbitrary, but should be large enough to obtain solid statistics. The average of an observable, which corresponds to the central value of the Hessian set, is given in terms of these replicas by

$$\langle \mathcal{O} \rangle = \frac{1}{N_{\text{rep}}} \sum_{i=1}^{N_{\text{rep}}} \mathcal{O}(f_k) \quad (121)$$

and the variance is given by

$$\delta \langle \mathcal{O} \rangle = \sqrt{\frac{1}{N_{\text{rep}}} \sum_{i=1}^{N_{\text{rep}}} (\mathcal{O}(f_k) - \langle \mathcal{O} \rangle)^2}. \quad (122)$$

The reweighting is then performed by assigning a weight

$$w_k = \frac{e^{-\frac{1}{2} \chi_k^2 / T}}{\frac{1}{N_{\text{rep}}} \sum_{i=1}^{N_{\text{rep}}} e^{-\frac{1}{2} \chi_i^2 / T}} \quad (123)$$

to each replica k , where χ_i^2 is the χ^2 value of the newly added data obtained with PDF replica i and T is the tolerance used for the original Hessian uncertainties.

With these weights, it is straightforward to calculate the updated observables and their variance as

$$\langle \mathcal{O} \rangle = \frac{1}{N_{\text{rep}}} \sum_{i=1}^{N_{\text{rep}}} w_k \mathcal{O}(f_k) \quad (124)$$

$$\text{and } \delta \langle \mathcal{O} \rangle = \sqrt{\frac{1}{N_{\text{rep}}} \sum_{i=1}^{N_{\text{rep}}} w_k (\mathcal{O}(f_k) - \langle \mathcal{O} \rangle)^2}. \quad (125)$$

II. Nuclear PDF analyses with light and heavy meson production data

Before we finally perform the new PDF analyses using the theory and tools developed in the past chapter, we give a final recap of the importance of these analyses and an overview of the assumptions that enter.

PDFs are important as a fundamental quantity, defined in the framework collinear factorization [147] (also called leading twist, or twist-2 factorization), which is derived rigorously in perturbative QCD for a variety of processes in lepton-proton and proton-proton collisions. This framework not only defines the PDFs, but also the hard scattering cross sections at the partonic level.

The collinear factorization theorem states that the cross section σ for a process involving two hadrons A and B in the initial state can be calculated as the convolution integral of the universal PDFs f_i^A and f_j^B with the observable as calculated at parton level $\hat{\sigma}_{i,j}$ summed over the combinations of partons i and j :

$$\sigma = \sum_{i,j} \int \int dx_i dx_j f_i^A(x_i, \mu^2) f_j^B(x_j, \mu^2) \hat{\sigma}_{i,j}(x_i, x_j) + \mathcal{O}\left(\frac{M}{Q}\right), \quad (126)$$

where we integrate over the momentum fractions x_i and x_j and μ^2 is a factorization scale. The error of this calculation is suppressed by powers of $\frac{M}{Q}$, where Q is a characteristic scale for the hard scattering process. At higher orders, the partonic cross section may also depend on a renormalization scale μ_r . Correspondingly, processes with just one hadron in the initial state, *e.g.*, deep inelastic scattering, require just one PDF convolution, while those involving hadronic final states require one additional convolution with the fragmentation function of the final state.

The framework has predictive power due to the fact that the PDFs are universal, *i.e.*, process independent, and also due to the perturbative scale dependence of the PDFs given by the DGLAP evolution equations. As the x dependence of the PDFs is not calculable in perturbation theory, it needs to be determined empirically in a global QCD analysis [8], where a fit to the available experimental data is performed¹. This framework has allowed

¹In principle, the PDFs can be calculated in lattice QCD, but performing such calculations at a useful accuracy is currently not feasible. However, there are proof-of-concept studies where lattice QCD “data” is used to extend global fits [28].

for a determination of the proton PDF with high precision [16, 18–29]. The more general case of nuclear PDFs lacks a rigorous derivation of the twist-2 factorization theorem, but as discussed in Refs. [148, 149], factorization is still a reasonable assumption in both lepton-nucleus and proton-nucleus collisions. However, the higher twist effects may see an enhancement proportional to $A^{1/3}$ in the nuclear case, such that more strict kinematic cuts need to be introduced to mitigate the impact of these effects on the analysis.

Using this assumption, a variety of nuclear PDF analyses have been performed [9–16, 107–109], albeit with significantly larger uncertainties compared to the proton case. In particular, there is still a large uncertainty on the gluon PDF at small x , which is important for the predictions of many processes measured at the LHC, such as vector boson production [40], prompt photon production [41] and production of heavy quarks [42]. Additionally, these uncertainties enter other theoretical analyses, *e.g.*, in the context of the very successful Statistical Hadronization Model describing the general hadronization of particles in a Quark-Gluon Plasma [43]. Finally, precise knowledge of the nuclear PDFs provides a starting point for comparisons with microscopic models predicting the nuclear modification in different x regions as described in Sec. I.1.5.

The reason for the large gluon uncertainties in nuclear PDF analyses is that the bulk of the data included comes from DIS and the Drell-Yan process, which are not sensitive to the gluon PDF at leading order. Additionally, there are less data points and the kinematic range is more limited than it is in proton case. Therefore, the goal of this thesis is extension of the nCTEQ15 nuclear PDF analysis with new data sets to provide stronger constraints on the gluon PDF and to test whether the data can be described purely in terms of twist-2 factorization without higher-twist corrections. In particular, we will begin with the inclusion of single inclusive pion and kaon production in Sec. II.2 and explore the dependence on the fragmentation functions required in the calculation for this process. Afterwards, we continue with similar data for open heavy flavored mesons and heavy quarkonia in Sec. II.3. For this purpose, we introduce a data-driven approach for the theoretical calculation that combines the partonic cross section and fragmentation into one effective scattering-matrix element, which is determined empirically.

1. Overview of relevant nuclear PDF fits

In the following section, we provide a brief overview of relevant nuclear PDF analyses. This includes fits of historical significance, other fits in the nCTEQ framework, as well as fits performed by other groups using different methods.

1.1. Historical nuclear PDF analyses

Following up on an earlier study of the scale dependence of nuclear modification [150], the first nuclear PDF analysis, EKS98 [151], was published by Eskola et al. in 1998. The fit included data structure function ratios from the NMC collaboration and the E665 experiment and the theoretical predictions were performed at leading order. The gluon nPDF in particular was not fitted, but instead is determined only through sum rules and DGLAP evolution. Following this analysis, Hirai et al. [152] were the first to perform a nuclear PDF fit with an estimate of the uncertainties, but included only DIS data in the analysis and the calculations were still performed at leading order. The first fit to use calculations at NLO for DIS and Drell-Yan (DY) process data, was performed by de Florian et al. [153], but this analysis was performed without an estimation of the PDF uncertainties.

1.2. Nuclear PDFs in the nCTEQ framework

The recent nCTEQ analyses are an obvious point of comparison for the work presented in this thesis, as some of them, *i.e.*, nCTEQ15 and nCTEQ15WZ, are direct progenitors of the work presented in Secs. II.2 and II.3. Other analyses like nCTEQ15HIX and nCTEQ15 ν , on the other hand, provide a complementary look focused on other kinematic regions and data sets.

nCTEQ15

The nCTEQ15 global analysis [9] was the original fit in the current nCTEQ framework, providing nuclear PDFs with uncertainties based on an initial analysis by Schienbein et al. [154]. The fit uses NLO calculations for all processes and includes data from deep inelastic scattering (616 data points), the Drell-Yan process (92 points) and neutral pion production (32 points). The total number of data points in these data sets is greater, but cuts are applied to limit the effects of higher twist contributions and target mass corrections:

- DIS: $Q^2 > 4 \text{ GeV}^2$ and $W^2 > 12.25 \text{ GeV}^2$
- pion production: $p_T > 1.7 \text{ GeV}$.

The nPDF parameterization and baseline proton PDF are explained in Sec. I.2.1. The uncertainties are determined using the Hessian method lined out in Sec. I.2.4.1 with tolerance of $T = 35$.

The code used to perform the analysis is written mainly in C++ with some parts using Python or Fortran. The central interface that performs the actual minimization using `Minuit` [138] is written in Python and connects the main `nCTEQ++` code to a modified version of the Fortran package `INCNLO` [155] for the neutral pions². The `nCTEQ++` code itself is mostly written in C++, but the code for DIS and DY cross section calculations is written in Fortran and the DGLAP evolution is handled in a modified version of `HOPPET` [156] that can handle multiple nuclei.

nCTEQ15WZ

The `nCTEQ15WZ` analysis [39] extends `nCTEQ15` by introducing data on W^\pm and Z boson production³ taken in proton-lead collisions at the LHC, which helps both with flavour separation and the gluon PDF. The calculations of the new data use grids stored in `APPLgrid` [157] format, calculated in `MCFM` [158]. Additionally, two changes are made to the fitting procedure: Firstly, three strange quark parameters are opened up, which were previously determined as a fixed fraction of the up and down seaquarks. Secondly, the normalizations of the newly added data sets are included as fitting parameters with a χ^2 penalty as explained in Sec. I.2.2. The analysis results in nPDFs that have an enhanced gluon and strange PDF at low x compared to `nCTEQ15`. The gluon uncertainty is also significantly reduced, while the strange uncertainty rises significantly due to the new flexibility in the parameterization. The up and down quarks remain mostly unchanged.

On the technical side, considerable simplifications of the code have been made, such that the work previously handled by the central Python interface is now integrated directly in the C++ code, and `INCNLO` is compiled directly as a sub-library of the main code. `Minuit` has been replaced with a custom implementation of the BFGS algorithm explained in Sec. I.2.3.

nCTEQ15HIX

Another analysis based on `nCTEQ15` is `nCTEQ15HIX` [119], which mainly aims to include more high x DIS data by applying less restrictive cuts on Q^2 and W^2 . On the theoretical side, corrections for the target mass, higher twist effects and the deuteron structure are introduced to adequately describe the physical effects in this region. These theoretical improvements allow for an adequate description of the recent JLab data that was previously completely excluded by the cuts. Additionally, a rescaled x is used to accomodate for the fact that the nuclear case allows for $x > 1$. Finally, the flexibility of the nPDF parameterization is increased by opening three additional parameters for the valence quarks. The result of the analysis is a set of nPDFs with slightly suppressed low x PDFs for all flavours,

²see Sec. II.2.1 for details on `INCNLO`.

³Note that the W^\pm and Z boson production is technically just a subset of the Drell-Yan process, where a weak boson is exchanged. In this thesis, however, we will use *Drell-Yan process* to refer explicitly to the photon induced process to keep the two subprocesses clearly separated in discussions.

suppressed sea quarks at high x and enhanced valence quarks with significantly decreased uncertainties in the high x region.

nCTEQ15 ν

The nCTEQ15 ν analysis [159] extends nCTEQ15WZ+SIH by taking another look at neutrino DIS data that was previously excluded due to tensions with existing charged lepton DIS data. Various statistical tools are introduced to check both the internal consistency of the neutrino data sets, as well as the consistence with the other data sets in the fit. The analysis identifies the dimuon production and CHORUS (anti-)neutrino scattering data as the only neutrino data that can be fully used in a consistent global fit. Including these data sets leads to a slight reduction in uncertainties for up and down quarks and better constraints on the strange quark.

nCTEQ15WZ+SIH and nCTEQ15HQ

Finally, there are the nCTEQ15WZ+SIH [64] and nCTEQ15HQ [65] analyses, where the former extends nCTEQ15WZ by adding single inclusive production of light mesons and the latter adds heavy mesons on top of that. These two global PDF fits are the main results of this thesis and are presented in detail in Sec. II.2 and II.3, respectively.

1.3. Nuclear PDF analyses by other groups

A variety of other collaborations have also released nuclear PDF fits, each including slightly different data and using different frameworks. The most used PDFs among them come from the EPPS and NNPDF collaborations with their most recent releases being nNNPDF2.0 [12] and EPPS21 [14], respectively, which are briefly explained in the following.

EPPS21

The EPPS21 analysis uses a similar framework as the nCTEQ15 based fits, but instead of parameterizing the absolute nuclear PDFs they fit nuclear correction factors. They use a total of 24 parameters, which are constrained by DIS, DY, WZ, dijet, neutral pion and D meson production data sets. The cuts imposed on the DIS data are less strict than those in nCTEQ15, including all data with $Q^2 > 1.7 \text{ GeV}^2$ and $W^2 > 3.24 \text{ GeV}^2$. With the D meson and neutral pion cuts below $p_T < 3 \text{ GeV}$, which leaves 2077 data points. The analysis uses the same Hessian method as nCTEQ15 with a similar tolerance of $T = 33$, but also provides additional error PDFs encoding the uncertainty introduced by the CT18A baseline proton PDF [18]. The resulting PDFs are generally in good agreement with nCTEQ15WZ, but have larger uncertainties, particularly in the low x down quark.

nNNPDF2.0

The nNNPDF2.0 fit uses a different approach, both for the PDF parameterization and the uncertainties. As the name implies, the PDFs are parameterized using a neural network in an attempt to remove bias introduced by the fixed parameterization of other analyses. Boundary conditions are applied to the network, such that NNPDF3.1 is reproduced in the proton case. This approach yields 256 free parameters, which are constrained by 1467 data points from DIS, neutrino DIS and WZ production. The uncertainties are not determined via the Hessian method used in nCTEQ and EPPS analyses, but via the Monte Carlo replica method explained in Sec. I.2.4.4. The resulting PDFs are also in good agreement with the previously discussed ones, but the large number of parameters results in larger uncertainties, particularly for the gluon.

TUJU21

The TUJU21 [160] is worth highlighting as well, as it uses the open-source framework **xFitter** [161] to perform a nuclear PDF analysis including DIS, Drell-Yan process and WZ production data with calculations at NNLO. The parameterization is similar to the one used in nCTEQ and the uncertainties are also determined in the Hessian formalism with a tolerance $T = 50$. Again, the resulting PDFs agree well with the previously mentioned analyses, but the obtained uncertainties are significantly smaller, particularly for the valence quarks and gluons.

KSASG20

Another nPDF analysis performed at NNLO is KSASG20 [162], which uses the same neutral current DIS and DY data sets as nCTEQ15HIX, with additional charged current DIS data. The analysis uses a cubic polynomial ansatz for the nuclear modification, which is fitted using calculations performed by the open-source framework APFEL [163]. This calculation accounts for target mass corrections and phenomenological higher-twist corrections. Given their Hessian uncertainties with tolerance $T = 20$, they obtain significantly smaller valence quark uncertainties than other fits, but display some disagreement in the gluon PDF above $x > 0.1$.

2. Single inclusive hadron production

In the following we will investigate the impact of single inclusive hadron (SIH) production data on global nuclear PDF analyses. The term *single inclusive* describes processes, where the specified hadron is detected along with an arbitrary number of additional particles. The hadrons we investigate in this are pions, kaons and η mesons.

Contrary to the DIS and DY data that make up the bulk of the data in the nCTEQ15 and nCTEQ15WZ analyses, the SIH data is directly sensitive to the gluon, even at leading order. Compared to DIS and DY data, however, single inclusive hadron production also comes with additional challenges due to the dependence on non-perturbative fragmentation functions (FFs). If these problems can be resolved, the significant contribution of the gluon to the overall cross section gives this data the potential to put new constraints on the nuclear gluon PDF.

The magnitude of the contribution from gluons can be seen in Fig. II.1, which shows the fractional contribution of each parton in the lead nucleus to the overall $p+\text{Pb} \rightarrow \pi^0 + X$ cross section as a function of the transverse momentum p_T . Especially in the low to mid p_T region, where the bulk of the experimental data is taken, the gluon contribution (shown in red) dominates at both ALICE and RHIC energies. At $\sqrt{s} = 200$ GeV, the gluon contribution starts at 70% for $p_T = 1$ GeV, becomes smaller as p_T rises and is overtaken by the up quark around $p_T \approx 10$ GeV. The up quark generally contributes roughly twice as much as the down quark at this energy, while their respective antiquarks and the strange (anti)quark contribute less than 10% combined. At the higher center-of-mass energy of the ALICE data, the gluon contributes around 70% for all $p_T < 10$ GeV and remains at 40% even at $p_T = 100$ GeV. At this energy, the up and down quarks contribute similar amounts to the total cross section, while the up, down and (anti-)strange quarks contribute about 10% across all p_T values. Charm and bottom flavors are omitted in this figure due to their minuscule contributions, but they are fully included in the calculation.

Figure II.2 shows the relative contributions of each parton in the fragmentation function. In particular, the red area shows the contribution from all subprocesses, where the initial partonic scattering produces a gluon, which then fragments into the neutral pion. Qualitatively, the contributions look similar to those of the PDF flavors, but at $\sqrt{s} = 200$ GeV the gluon contribution starts higher and then drops off more quickly. At $\sqrt{s} = 5$ TeV, the gluon contribution peaks with 85% at $p_T = 3$ GeV and then slowly drops as p_T increases. The remaining fraction is split almost evenly between the quark flavours. The large contribution of the gluon seen here makes the accuracy of the gluon FF very important for our PDF fit.

The calculations for both figures are performed using nCTEQ15WZ PDFs [39] and DSS FFs [164], but the qualitative behaviour is the same when other nPDFs or FFs are used. The same computations were carried out for kaon and eta meson production, but no significant

difference in the contributions was observed.

We begin this analysis by providing an overview of the theoretical framework including the general calculation of the cross section and an overview of available FFs. The available data sets are then presented and discussed in Sec. II.2.2. In Sec. II.2.1, we investigate how the calculations depend on the choice of fragmentation function and the scale choices to identify a kinematic region, where reliable predictions can be made. Section II.2.4 then presents the new global PDF fits performed using the SIH data and evaluates the consistency with previous data sets. Finally, the conclusions are summarized in Sec. II.2.5.

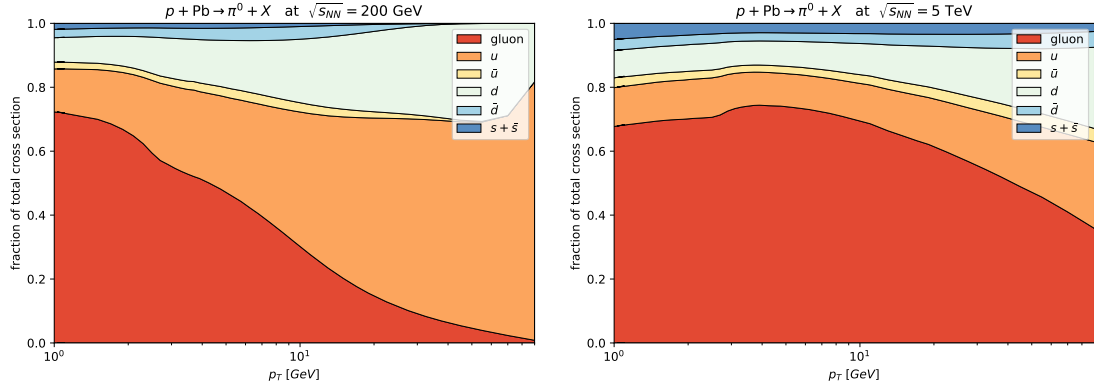


Figure II.1.: Contribution of each PDF flavour to the overall cross section of $pp \rightarrow h + X$ at $\sqrt{s} = 200$ GeV (left panel) and 5 TeV (right panel) for different values of transverse momentum p_T .

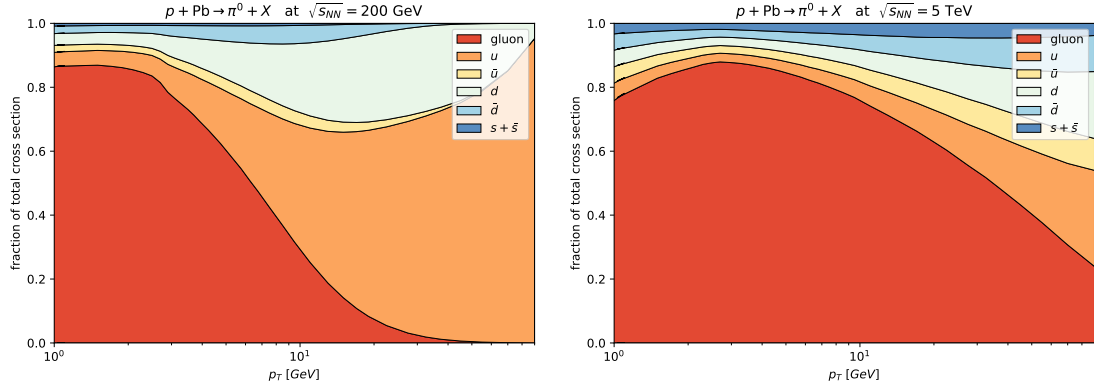
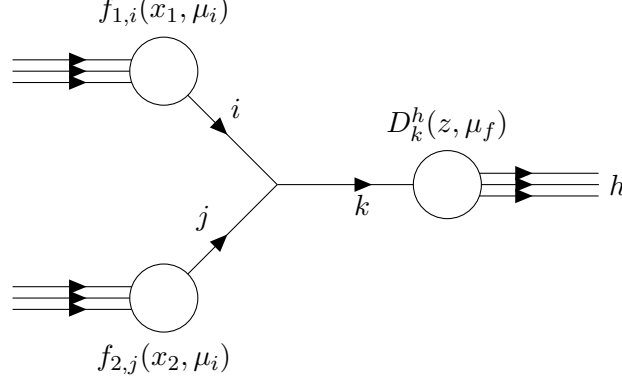


Figure II.2.: Contribution of each FF flavour to the overall cross section of $pp \rightarrow h + X$ at $\sqrt{s} = 200$ GeV (left panel) and 5 TeV (right panel) for different values of transverse momentum p_T .

2.1. Theoretical framework

To obtain the cross section for single inclusive hadron production from the PDFs, the two initial state particles' PDFs are convoluted with the cross section of the partonic subprocess

and the final state fragmentation function. A brief overview of the process at NLO is given in Ref. [165]. In terms of Feynman diagrams, the process can be described by the following graph at leading order:



Going through the diagram from left to right, we have the partons i and j from the two initial hadron's PDFs $f_{1,i}(x, \mu_i)$ and $f_{2,j}(x, \mu_i)$ at the initial state factorization scale μ_i . These partons then interact in a hard scattering process to produce the outgoing parton k , which introduces a renormalization scale μ_r at higher orders. Finally, the fragmentation function $D_k^h(z, \mu_f)$ then gives the probability for the parton k to produce the final state hadron h at the final state factorization scale μ_f . Omitted in the diagram are the additional particles that may be produced in the process. The double differential cross section in pseudorapidity η and transverse momentum p_T can therefore be expressed as

$$\begin{aligned} \frac{d\sigma^h}{dp_T d\eta} = & \sum_{i,j,k} \iiint dx_1 dx_2 f_{1,i}(x_1, \mu_i) f_{2,j}(x_2, \mu_i) \frac{dz}{z^2} D_k^h(z, \mu_f) \\ & \times \left[\left(\frac{\alpha_s(\mu_r)}{2\pi} \right)^2 \frac{d\hat{\sigma}_{ij,k}}{dp_T d\eta} + \left(\frac{\alpha_s(\mu_r)}{2\pi} \right)^3 K_{ij,k}(\mu_i, \mu_r, \mu_f) \right], \end{aligned} \quad (127)$$

where $\frac{d\hat{\sigma}_{ij,k}}{dp_T d\eta}$ is the Born cross section of the partonic subprocess $i + j \rightarrow k + X$ and $K_{ij,k}(\mu_i, \mu_r, \mu_f)$ contains all corresponding higher order corrections. In this analysis we include a total of nine different fragmentation functions, as listed in Tab. II.1. We will give a brief overview of their properties and then compare them both in terms of predictions for proton-proton and proton-nucleus collisions.

2.1.1. Calculating the cross section in practice

Performing all convolutions for each data point in each iteration of a fit would make the single inclusive hadron predictions too computationally expensive. In a PDF fit where hundreds of data points are evaluated thousands of times, this would take days or even weeks. A solution to this problem is performing the convolution of the proton (or deuteron in the case of RHIC data) PDF and the FF ahead of time and saving the results to a grid, so that instead of nested convolutions we only need to perform a single one during

fitting. We use the gridding procedure explained in Sec. I.2.5.2 to produce such grids with 1000 sampling points linearly spaced in x for each parton flavour for every data point. Calculating these grids takes about 48h of computation time per data point on a single CPU core. To complete the calculation in a reasonable amount of time, we used one node of the PALMA II cluster with 36 CPU cores for each data set. Since the grids only need to be computed a single time, the investment quickly pays off because the computation time for of a single SIH data point takes up almost as much time as all the DIS data sets combined, when no grids are used.

The grids produced using this method have been tested to reproduce the full calculation within a margin significantly smaller than the data uncertainty and the FF uncertainty that will be discussed in Sec. II.2.3.3.

2.1.2. Overview of available fragmentation functions

Like PDFs, the FFs describing the hadronization stage of single inclusive hadron production are non-perturbative objects. Over the past decades, various groups have released fragmentation functions that are obtained by fitting them to data in global analyses very similar to those of PDFs. Initially, those fits used only data from single-inclusive hadron production in electron-positron annihilation, but over time groups have added data from semi-inclusive deep inelastic lepton-nucleon scattering and other processes to improve the accuracy and kinematic range of their fits, especially of the gluon initiated fragmentation, which we are interested in. A selection of analyses of FFs for various mesons⁴ is presented in Tab. II.1. Some of those fragmentation functions impose hard limits on their allowed kinematic region in z and Q , resulting in a \sqrt{s} -dependent minimum transverse momentum $p_{T,\min}$ for which predictions can be obtained in our theoretical framework. These limits are shown for $\sqrt{s} = \{200, 5020, 8160\}$ GeV in the fourth column of Tab. II.1. Note that these limitations are also dependent on the program used to calculate the cross section, as they are imposed at the level of z and Q and therefore the p_T limitations depend somewhat on the exact integration routine.

Further fragmentation functions exist for other final states like protons, antiprotons and unidentified charged hadrons (*e.g.*, SGK18 [166], NNFF1.1h [167]), but we exclude those from the analysis due to the comparatively large uncertainties both on the data and the fragmentation functions. There have also been studies on the effect of the nuclear medium on the fragmentation [168], but we exclude the fragmentation functions obtained there from our analysis in order to avoid double counting of the shared data points and because the uncertainties on the nuclear corrections are still very large [169].

We will henceforth denote the combination of DSS14 for pions and DSS17 for kaons

⁴Neutral pions are always treated as the average of positive and negative ones. Some fragmentation functions include further particles, but they are omitted because they are not included in the data used in the current analysis.

Table II.1.: Overview of the various available sets of FFs $D_i^h(z, Q)$. Values of the minimal accessible transverse momentum $p_{T,\min}$, given in the brackets, correspond to $\sqrt{s} = (200, 5020, 8160)$ GeV.

FF	Ref.	Year	$p_{T,\min}$ [GeV]	Available particles	Uncertainties
BKK	[172]	1994	(3, 3, 3)	π_0, π^\pm, K^\pm	-
KKP	[173]	2000	(3, 3, 3)	π_0, π^\pm, K^\pm	-
KRETZER	[174]	2000	(2, 25, 40)	π_0, π^\pm, K^\pm	-
HKNS07	[175]	2007	(2, 25, 40)	π_0, π^\pm, K^\pm	Hessian
AKK	[176]	2008	(5, 125, 200)	π_0, π^\pm, K^\pm	-
NNFF	[177]	2017	(0, 0, 0)	π_0, π^\pm, K^\pm	MC replicas
JAM20	[178]	2021	(0, 0, 0)	π_0, π^\pm, K^\pm	MC replicas
DSS14	[164]	2014	(0, 0, 0)	π_0, π^\pm	Hessian
DSS17	[179]	2017	(0, 0, 0)	K^\pm	Hessian
AESSS	[180]	2011	(0, 0, 0)	η	-

simply as “DSS”⁵. Note that HKNS16 [171] exists as an updated version of HKNS07, but no code could be obtained to use these updated fragmentation functions.

The oldest set of FFs in the list, BKK [172], uses a simple parameterization with 9 parameters each for pions and kaons, that are determined in a fit using NLO calculations and electron-positron annihilation data from the TPC detector at SLAC taken at $\sqrt{s} = 29$ GeV with a kinematic cut excluding data with $x < 0.07$. The KKP [173] and KRETZER FFs [174] are determined in similar analyses, which also include data taken by the ALEPH, DELPHI and OPAL detectors. The AKK analysis [176] extends the data further by including inclusive hadron production data in proton-proton collisions, while the AESSS [180] fit uses similar data to obtain FFs for eta mesons. The proton-proton data in these analyses helps with the determination of the gluon FF, which is only weakly constrained by the electron-proton data. The HKNS07 analysis [175] on the other hand remains limited to electron-positron data, but is the first to provide an estimate of the uncertainties using the Hessian method. In the DSS fits [164, 179], the included data is extended further to include semi-inclusive deep-inelastic scattering (SIDIS) as well, which helps in separating the different flavors’ contributions. Throughout the analyses mentioned thus far, the number of open parameters increases as more data becomes available over a greater range of kinematics. The analysis by the NNPDF collaboration [177] greatly increases the number of parameters by using a neural network to model the FFs. Accordingly, they use Monte Carlo replicas to determine the uncertainties. Additionally, they are the first to use NNLO calculations, but include only electron-positron annihilation data. Finally, the JAM20 analysis [178] performs the FF fit simultaneously with a proton PDF fit,

⁵The name DSS comes from the authors (D. de Florian, R. Sassot, M. Stratmann) of the original DSS fragmentation functions in 2007 [170], but newer versions have kept the abbreviation for clarity even though M. Epele and R.J. Hernandez-Pinto joined as authors.

since these quantities are correlated through the SIDIS data. Naturally, far more detailed explanations of each individual analysis can be found in their respective references.

2.2. Available data

The same deep inelastic scattering (DIS) and Drell-Yan (DY) lepton pair production data sets as used in the nCTEQ15 analysis build the baseline of this analysis. The RHIC pion production data used in nCTEQ15 is extended both in terms of further data points from more recent ALICE analyses and also in terms of further particles, namely charged pions, charged kaons and η mesons. The charged mesons always appear as the average of their positively and negatively charged version, which is also how the neutral pions are calculated in the fragmentation functions⁶. The PHENIX and STAR data is taken at a center of mass energy of 200 GeV while the ALICE data sets are measured at 5020 GeV and 8160 GeV. Table II.2 gives a brief overview of the available data sets, while Fig. II.3 shows the p_T distribution of the available data points for each set. The new ALICE data sets especially provide data in a far wider range of p_T values than the two RHIC neutral pion sets that were used in the nCTEQ15 fit. Between the charged pion data taken at 5 TeV and the neutral pions at 8 TeV, the p_T spectrum spans more than three orders of magnitude.

Table II.2.: Overview of the available data sets including their center of mass energy, observable and number of data points.

Data set	Ref.	ID	\sqrt{s} [GeV]	Observ.	No. points	After cuts
PHENIX π^0	[181]	4003	200	R_{dAu}	21	17
PHENIX η	[181]	4403	200	R_{dAu}	12	11
PHENIX π^\pm	[182]	4103	200	R_{dAu}	20	0
PHENIX K^\pm	[182]	4203	200	R_{dAu}	15	0
STAR π^0	[183]	4002	200	R_{dAu}	13	9
STAR η	[183]	4402	200	R_{dAu}	7	7
STAR π^\pm	[184]	4102	200	R_{dAu}	23	8
ALICE 5 TeV π^0	[185]	4001	5020	R_{pPb}	31	15
ALICE 5 TeV η	[185]	4401	5020	R_{pPb}	16	9
ALICE 5 TeV π^\pm	[186]	4101	5020	R_{pPb}	58	22
ALICE 5 TeV K^\pm	[186]	4201	5020	R_{pPb}	58	22
ALICE 8 TeV π^0	[187]	4004	8160	R_{pPb}	30	19
ALICE 8 TeV η	[187]	4404	8160	R_{pPb}	14	10

Like in other types of experiments, not all available data points can be used for fitting. Kinematic cuts need to be imposed to remove data that cannot be adequately described by the theory. This is in part due to the fact that in the very low p_T region, the process becomes non-perturbative. But even in the perturbative regime, problems may arise since

⁶Most FFs provide this explicitly. For the remainder we calculate this within our code.

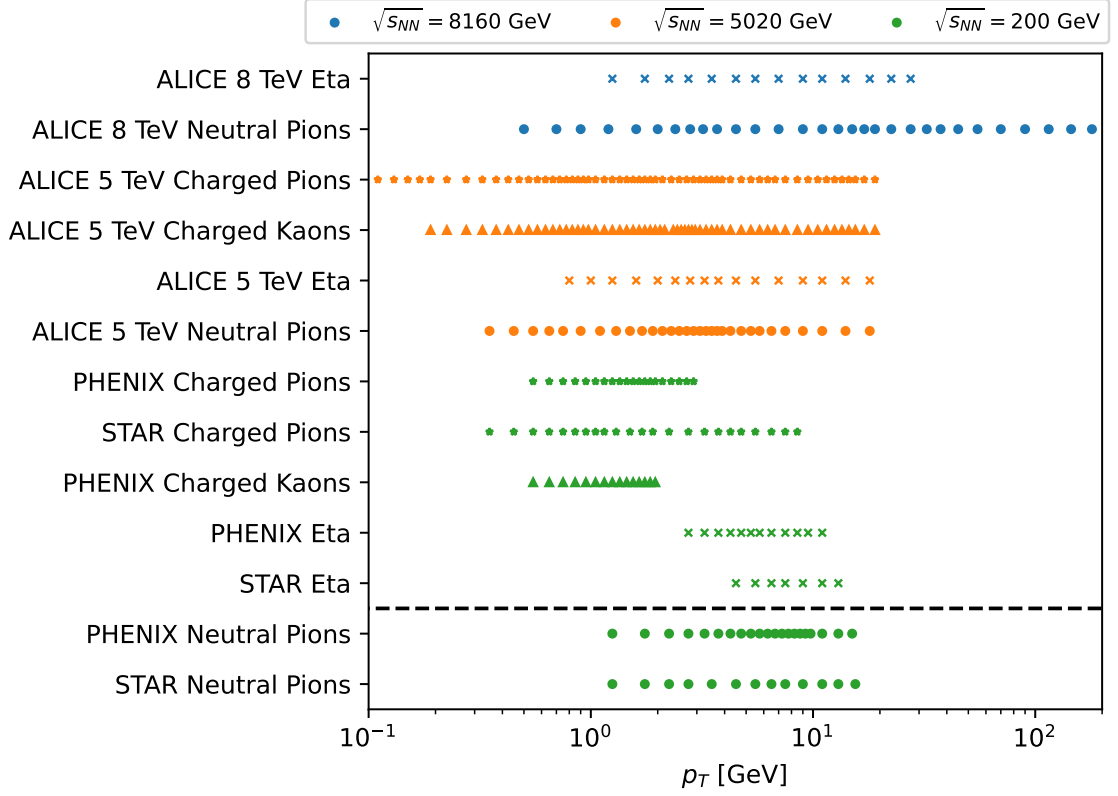


Figure II.3.: Transverse momentum p_T of all data points split by set and colored by \sqrt{s} . The data sets above the dashed line were included in nCTEQ15.

the FFs required in the calculation of the cross sections also have kinematic limitations. These come from the availability of data in the FF fits and perturbativity concerns in the calculations performed during their fits. Therefore, we need to make sure that our cuts also exclude data where the accuracy of the FFs at the given kinematics is insufficient. How to choose those cuts will be discussed in Sec. II.2.4.

All the single inclusive hadron production data is given in terms of ratios

$$R_{dAu}^h = \frac{1}{A_d A_{Au}} \frac{\sigma_{d+Au \rightarrow h+X}}{\sigma_{p+p \rightarrow h+X}} \quad (\text{RHIC}) \quad (128)$$

$$\text{and} \quad R_{pPb}^h = \frac{1}{A_{Pb}} \frac{\sigma_{p+Pb \rightarrow h+X}}{\sigma_{p+p \rightarrow h+X}} \quad (\text{ALICE}) \quad . \quad (129)$$

This is very convenient for PDF fits because many of the uncertainties will be cancelled out to a large extent. On the experimental side, many systematic uncertainties cancel if the proton-proton collisions are measured with the same equipment as the nuclear collisions, leaving mostly statistical uncertainties and an overall normalization factor. From the theory

perspective, the impact of many potentially disruptive factors like the FF uncertainty, the scale choices and missing higher orders will largely cancel out as well, giving us much better precision in our fit than cross sections themselves would.

2.3. Fragmentation function and scale dependence

The upcoming section is dedicated to the determination of the optimal answers for the following three interconnected questions:

- Which FF should be used in the PDF fit and how can we compensate for the introduced theoretical uncertainty?
- What are the best choices for the scales of initial-state factorization, renormalization and final-state factorization?
- Where should the data cuts be applied such that we retain as many data points as possible, while minimizing errors introduced by non-perturbative effects?

2.3.1. Comparing fragmentation functions in proton-proton collisions

Before we go on to the data of nuclear ratios $R_{AA'} \equiv \frac{1}{AA'} \frac{\sigma_{AA' \rightarrow h+X}}{\sigma_{p+p \rightarrow h+X}}$ that are used in the fit, we want to take a look at the $p+p \rightarrow h+X$ baseline to help us understand the limitations of the theory prediction due to the uncertainties associated with the FFs. Figure II.4 shows a comparison⁷ of predictions from various fragmentation functions with data taken in proton-proton collisions by the PHENIX and ALICE experiments at $\sqrt{s} = 200$ GeV and 7000 GeV respectively. The calculations are performed with BKK, KKP, DSS, NNFF and JAM20 FFs and nCTEQ15 proton PDFs. AKK, KRETZER and HKNS FFs cannot be used to make predictions for the ALICE data due to their kinematic restrictions and are therefore omitted here. At 200 GeV all fragmentation functions are able to describe the data above $p_T = 3$ GeV (*i.e.*, yield a $\chi^2/N_{d.o.f.} < 1$), if one allows for a normalization shift. Below this kinematic region all curves show a slope in the theory/data plot which signals qualitative disagreement. There is also a slight upwards slope towards higher p_T for all fragmentation functions besides JAM20, but it is well within the data uncertainties. At ALICE energies the data can be well described by BKK, KKP, DSS and NNFF down to p_T values of 3 GeV if a normalization is introduced, while JAM20 starts deviating slowly from a constant shift below $p_T \approx 5$ GeV and is unable to describe the data below $p_T \approx 4$ GeV. Again, the theory predictions increasingly overshoot the data the further one goes below 3 GeV. Since this effect is not dependent on fragmentation function, it is reasonable to expect the same behavior independent of the produced final state.

⁷The goal of this comparison is *not* a general judgement of the fragmentation functions ability to predict SIH production. For that purpose, one should use JAM20 and MSTW2008 proton PDFs [188] for JAM20 and DSS, respectively. Instead, this comparison is for the explicit purpose of nuclear PDF fitting in the nCTEQ framework, which means that we have to use our proton baseline.

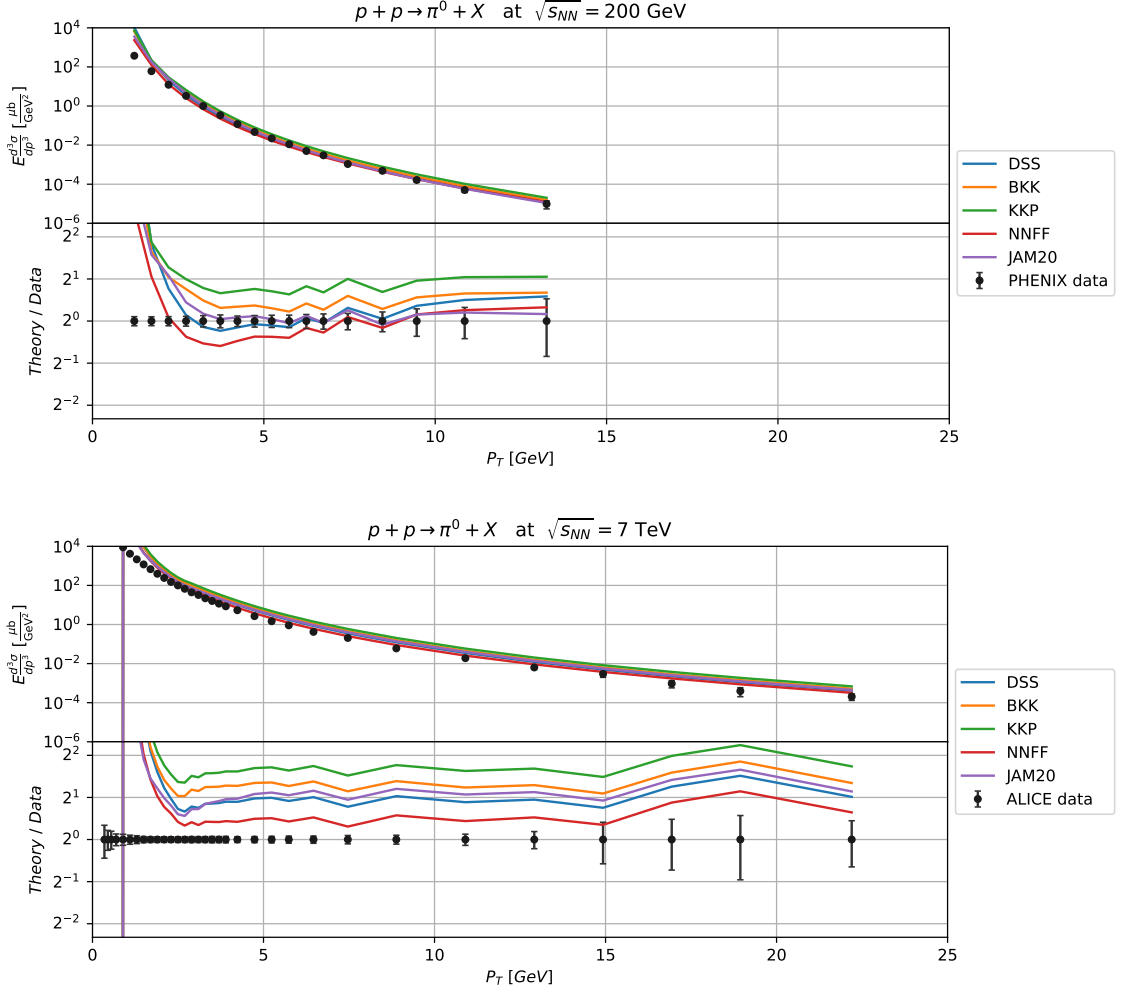


Figure II.4.: Comparison of predictions made with different fragmentation functions for $p + p \rightarrow \pi^0 + X$. The calculations are performed using nCTEQ15 proton PDFs. Both panels show data for neutral pions, with PHENIX data [189] in the upper and ALICE data [190] in the lower one.

2.3.2. Comparing different factorization and renormalization scales

As mentioned in Sec. II.2.1, the prediction of the SIH production cross section at NLO depends on three scale choices: initial state factorization μ_i , final state factorization μ_f and renormalization μ_r . These scales are not determined by the theory as they are merely artifacts of fixed order calculation and would vanish if the cross section could be calculated to all orders in perturbation theory. Frequently, they are chosen to be $\mu_i = \mu_f = \mu_r = c p_T$, where c is a constant that is commonly chosen as either $\frac{1}{2}$ or 1, but there is no general prescription for their choice. Figure II.5 shows the prediction for pion production at 200 GeV and 7 TeV with each scale varied independently between the two aforementioned

choices. The case where all scales are equal to $\frac{1}{2}$ (bold, grey) gives the best description of the 200 GeV data and is also the only scale choice that allows obtaining a $\chi^2/N_{\text{dof}} < 1$ for the 7 TeV data with $p_T > 3 \text{ GeV}$ if the normalization is chosen freely. Therefore, these are the scales we will be using from this point on. The two data sets shown are actually included in the fit of the DSS fragmentation function, where they place a cut above $p_T = 5 \text{ GeV}$. The reason why they need to place the cut higher than we do is their scale choice $\mu_i = \mu_f = \mu_r = p_T$ (blue curve).

Note that this scale choice puts a somewhat hard limit of $p_T > 2.6 \text{ GeV}$ on the kinematic region, because otherwise the initial state factorization scale lies below the initial scale of our PDF evolution. Hence, predictions can only be made by either interpolating our PDFs or by imposing a lower limit on the initial factorization scale

$$\mu_i = \begin{cases} 1.3 \text{ GeV} & \text{for } \frac{1}{2}p_T < 1.3 \text{ GeV}, \\ \frac{1}{2}p_T & \text{otherwise.} \end{cases} \quad (130)$$

The latter option was used for the purpose of these plots, but as we will see, the cuts we choose are above this limit anyway.

It is also interesting to take a look at the impact of the proton PDF that is used to see whether a different baseline PDF might make. Such a comparison is shown in Fig. II.6 for nCTEQ15, JAM20, CT18 and CT14. Contrary to the previous two figures, we interpolate the PDFs below their initial scale for the purpose of this calculation, which leads to the downwards slope instead of the upwards slope seen at low p_T previously. In general, the differences between the proton PDFs are significantly smaller than those between different scales or FF choices were, but the difference in the low p_T region at 7 TeV is not negligible. Interestingly, the nCTEQ15 proton PDF, which is based on relatively old CTEQ6.1m result, shows the best agreement with the data, while newer PDFs show a slight downward slope at low p_T . These results will need to be taken into account once the baseline proton PDF of future nCTEQ releases is updated.

2.3.3. Comparing fragmentation functions in dAu and pPb collisions

In the following section we discuss the differences between the different fragmentation functions in the nuclear ratios $R_{AA'}$. Figure II.7 shows all data sets with predictions using nCTEQ15WZ PDFs for lead, together with each set of FFs. We also include the uncertainties of the DSS FFs to give an estimate of how large the differences between different sets of FFs are compared to their uncertainty. We can see that BKK and KKP, which were used in the nCTEQ15 fit, show very close agreement with DSS. The most notable difference between BKK and KKP is seen in the charged kaon production, where KKP is slightly lower at high p_T values. NNFF and JAM20 also agree very well with

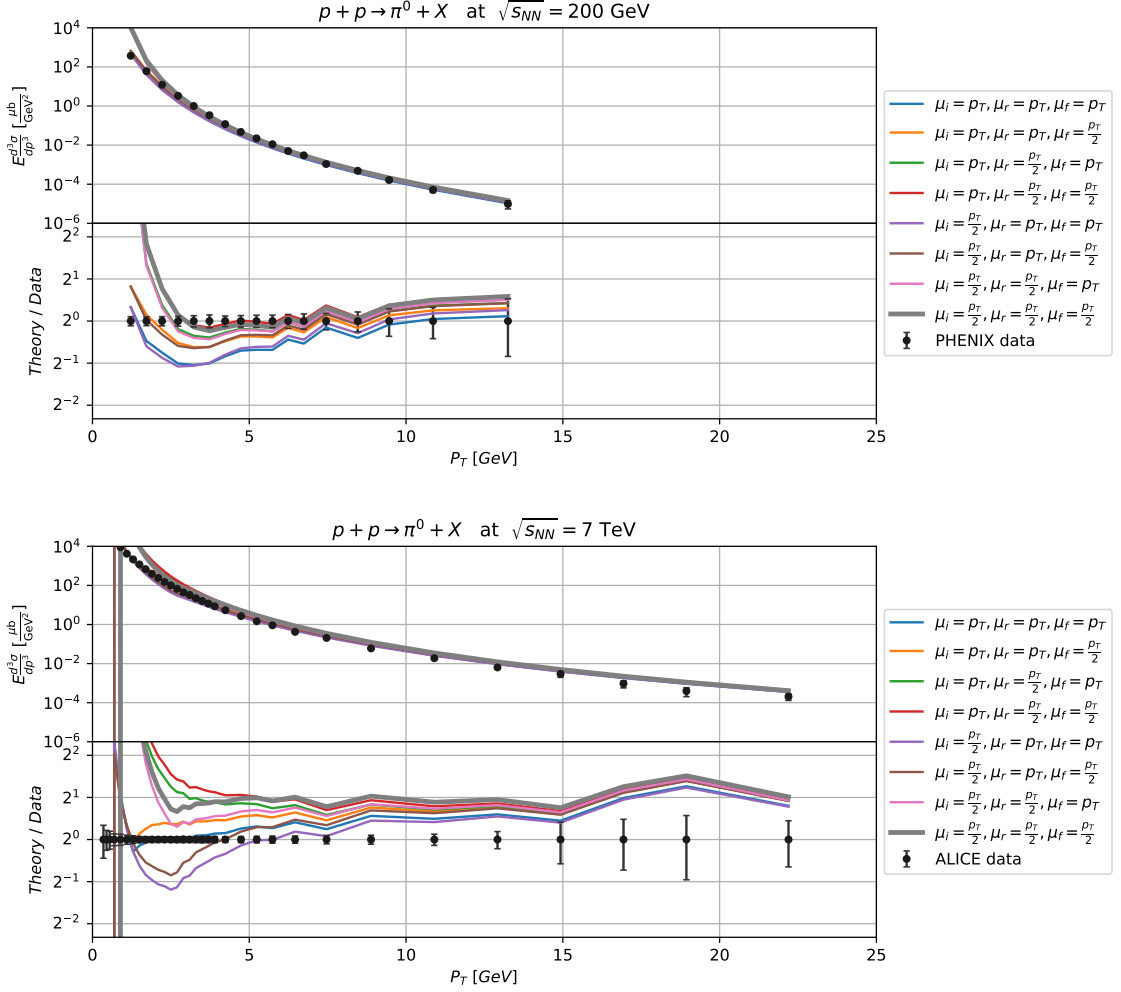


Figure II.5.: Comparison of predictions made with different scale choices for $p + p \rightarrow \pi^0 + X$. The calculations are performed using nCTEQ15 proton PDFs with DSS fragmentation functions. Both panels show data for neutral pions, with PHENIX data in the upper and ALICE data in the lower one.

each other across all data sets and the only case where they lie outside of the uncertainty given by DSS is the high- p_T ALICE pion data. Since the data in this region is sparse and has quite large uncertainties, however, this should not have any significant impact on fits. In the small kinematic region, where AKK allows predictions, they also agree with the previous FFs. KRETZER shows some qualitative differences in the region just above the cut, but lies within the uncertainty of DSS. For HKNS the disagreement is slightly larger but still well below the data uncertainty. Since AESSS is the only available fragmentation function for η mesons, no comparisons can be made, but the data is described reasonably well.

Note also that the predicted ratios are already quite close to the data values. While this

2. Single inclusive hadron production

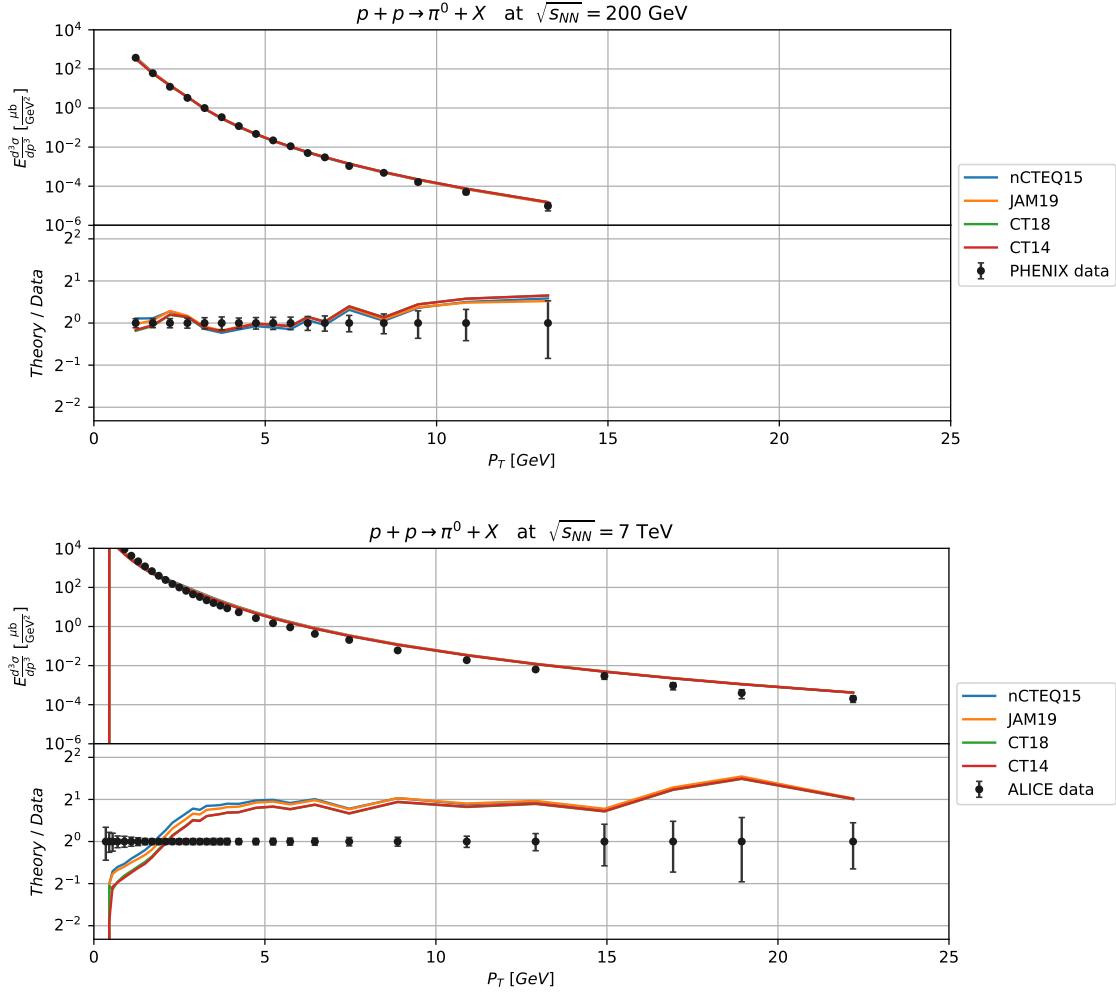


Figure II.6.: Comparison of predictions made with different proton PDFs for $p+p \rightarrow \pi^0 + X$. The calculations are performed using DSS fragmentation functions and $\mu_i = \mu_f = \mu_r = \frac{1}{2}p_T$. Both panels show data for neutral pions, with PHENIX data in the upper and ALICE data in the lower one.

suggests that the data will not significantly alter the central value of the nuclear PDFs, the data will still have an impact on their uncertainties.

Uncertainties of fragmentation functions

In total, four of the available fragmentation functions include uncertainties. HKNS and DSS provide their uncertainties in terms of Hessian eigenvectors, while JAM20 and NNFF provide Monte Carlo replicas. The uncertainties of the DSS fragmentation were already shown in Fig. II.7. Plots showing the uncertainties for NNFF, JAM20 and HKNS are given in Figs. IV.3, IV.2 and IV.3, respectively. The NNFF fragmentation functions yield slightly larger uncertainties than those of DSS shown before. This is likely due to their use

of a neural network instead of a “traditional” parameterization and a slightly smaller data set. The uncertainties of the JAM20 fragmentation on the other hand are so small across the kinematic region with $p_T > 1\text{ GeV}$ that they can be neglected when compared with the data uncertainty. Since their data set is not larger than that used in the DSS fit, the parameterization used by JAM20 might be too restrictive to display the true uncertainties. Finally, the uncertainties from the HKNS eigenvectors are many times as large as the data uncertainty for p_T values below 10 GeV and do not include uncertainties for kaons. These two factors obviously make them unusable for our purposes.

To mitigate the consequences of our FF choice on the fit and to account for the uncertainty from the fit of the fragmentation function, we calculate the uncertainties of the DSS fragmentation function for each data point and add them as a systematic uncertainty onto the data in our fit. These uncertainties are technically PDF dependent, but due to the very weak dependence, this effect can be neglected.

It is important to note that the shown bands do not reflect the full uncertainty of the theory prediction, but rather a lower bound. Further uncertainties come from the following two factors:

- **FF extrapolation:** The theory predictions of low- p_T points may depend on fragmentation functions beyond their fitted kinematic region. This results in an uncertainty due to the extrapolation. Moreover, the accuracy of the Hessian method outside of the region, where data exists, is heavily dependent on the exact parameterization of the FF.
- **Low- p_T corrections:** We know that below some value of p_T , perturbativity starts breaking down, which means that there is a transition region between accurate perturbative predictions and complete non-perturbativity.

As both of these factors have stronger effects with the decreasing p_T , they are the reason why we need to impose cuts on low p_T values.

Accuracy of the grid approximation

Figure II.8 shows a comparison of BKK, KKP and DSS fragmentations with their grid approximations for R_{pPb} and R_{dAu} for a representative selection of data sets. Note that, due to the long computation time of the grids, they are only computed for data not excluded by cuts. It is evident that the deviation between the grid approximation and full calculation is far smaller than the uncertainties of the FFs and especially the data. However, the gridding does not work as well for all FFs. In particular, KRETZER and HKNS yield very different results than their grid approximations.

2.3.4. Fragmentation kinematics

The final check we need to perform on the fragmentation functions is related to the z -dependence of the data at various p_T values. This is due to the fact that most fragmentation functions do not include data in their fits that gives constraints below $z \approx 0.05$ and would therefore rely purely on extrapolation if our calculations would depend on this low- z region. Figure II.9 shows the contribution of different z regions to the total $p + Pb \rightarrow \pi^0 + X$ cross section at 200 GeV and 5 TeV in the upper and lower panel, respectively. At 200 GeV the cross section is dominated by the $0.4 \leq z \leq 0.7$ region across the entire p_T range. The $z \leq 0.3$ region only starts contributing below $p_T < 3$ GeV and $z \leq 0.2$ is barely visible even at $p_T = 1$ GeV. At a center of mass energy of 5 TeV, the contribution of the low z region increases as the $0.1 < z < 0.3$ range is present even at higher p_T values. The $z < 0.1$ region starts contributing below $p_T < 5$ GeV, but stays below 10% even at $p_T = 1$ GeV. Contributions from the region where FF extrapolation effects would become relevant are, therefore, negligible.

2. Single inclusive hadron production

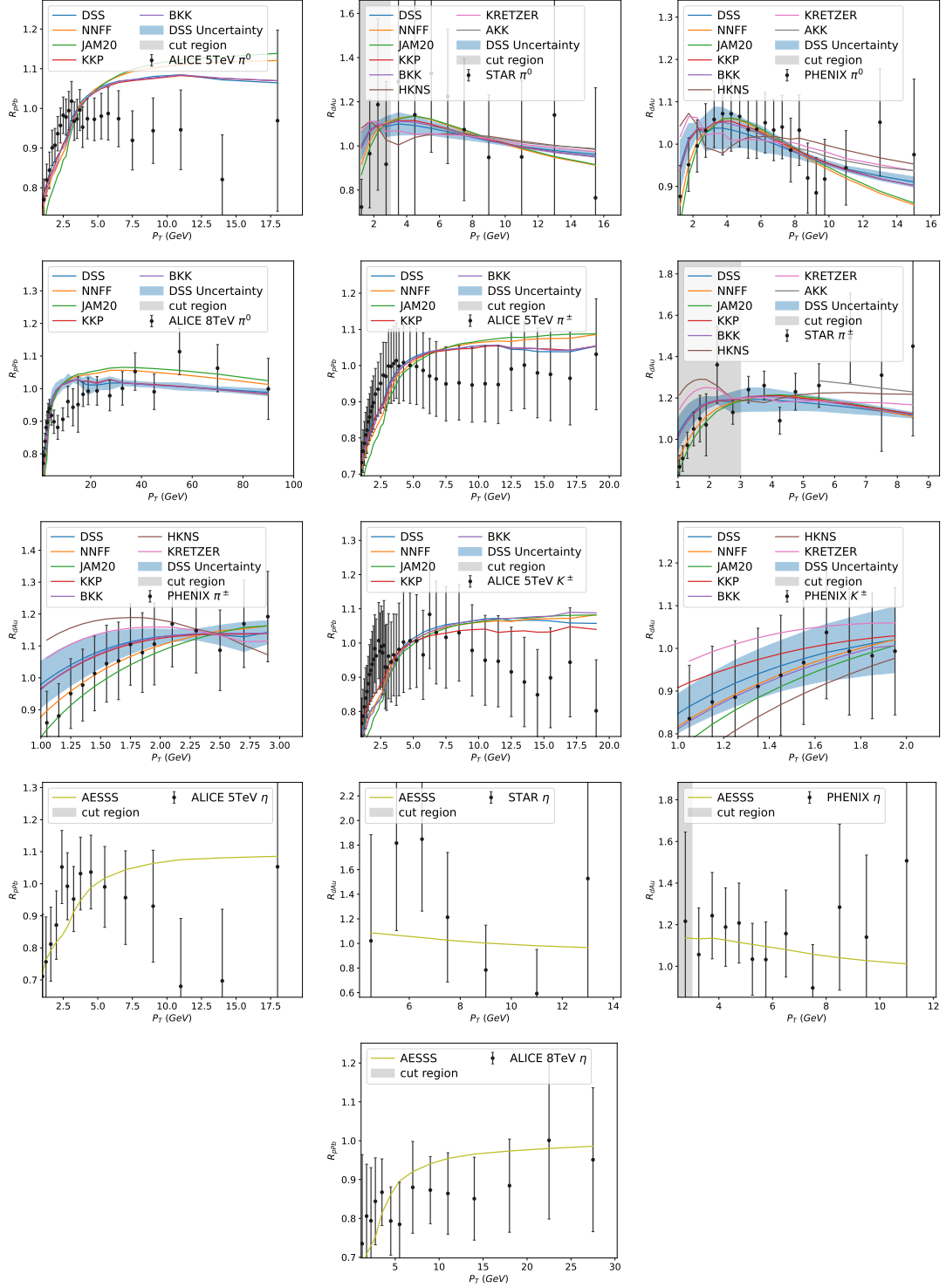


Figure II.7.: Comparison of different FFs with data for nuclear ratios $R_{AA'}$ for pion, kaon, and η production. The theoretical predictions are calculated at NLO in QCD using nCTEQ15WZ nuclear PDFs. The predictions are scaled by the inverse of their fitted normalization. The uncertainties of the DSS fragmentation functions are shown as blue bands. The grey region shows the data that is cut from the fits.

2. Single inclusive hadron production

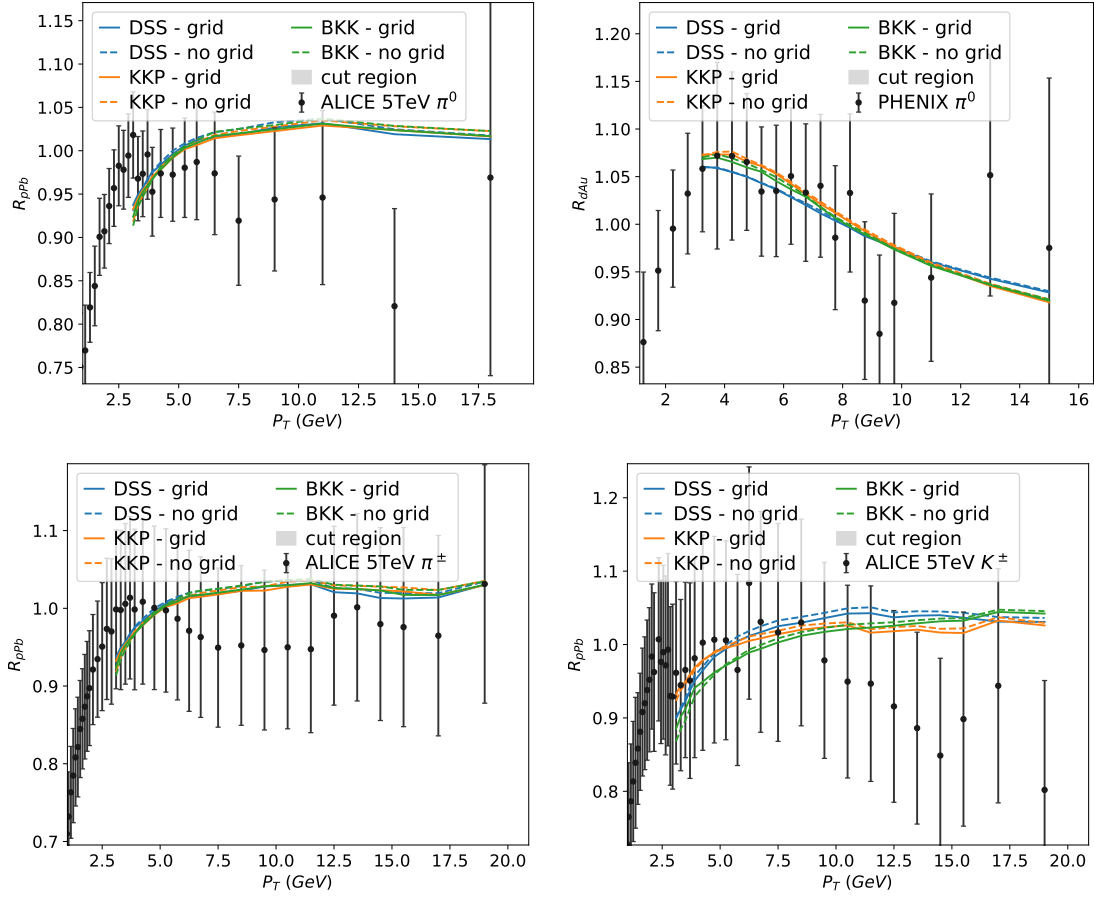


Figure II.8.: Comparison of the predictions for R_{pPb} and R_{dAu} with various FFs (solid lines) and their grid approximations (dashed lines) for a selection of data sets.

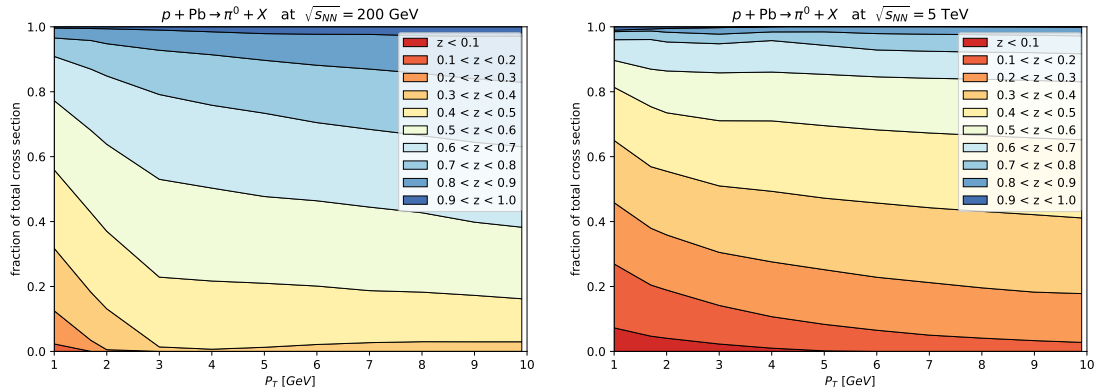


Figure II.9.: Contribution of different z regions of the fragmentation functions $D^{\pi^0}(z, Q)$ to the total $p + Pb \rightarrow \pi^0 + X$ cross section at 200 GeV (left) and 5 TeV (right).

2.4. Fits with SIH data

Before performing the fits, there are a number of decisions that need to be made. The first among those is the selection of data sets that should be included in the fit and the kinematic cuts applied to those. All fits will include the same DIS and DY data sets as nCTEQ15 with the same cuts. An overview of these data sets can be found in Tabs. IV.4 - IV.6. Each fit will then be performed twice: Once with and once without the W^\pm and Z boson production data from nCTEQ15WZ, which is listed in Tab. IV.7. For the new data sets we choose to include only the neutral pions, charged pions and charged kaons. η mesons are excluded due to the lack of known uncertainties for their FFs. Additionally, we cut all SIH data with transverse momenta $p_T < 3 \text{ GeV}$ in accordance with the results of the proton-baseline study in Fig. II.4. This is more restrictive than the cuts used in nCTEQ15(WZ) and EPPS16 [10], which include neutral pions with p_T values as low as 1.7 GeV. As observed in Sec. II.2.3.4, this also ensures that the fragmentation functions are only used in the region where they are well constrained by the data used in their respective fits. This leaves us with a total of 109 new data points; 77 (out of 174) from ALICE and 32 (out of 77) from RHIC. To keep the focus on low- x effects, we exclude the high- x and low- Q data that was introduced in nCTEQ15HIX.

For the fragmentation functions we choose DSS and we evaluate the theory at scales $\mu_i = \mu_r = \mu_f = \frac{1}{2}p_T$. In Fig. II.5, we determined this to be the optimal scale choice and the DSS fragmentation functions come from a relatively recent analysis with a very good description of the data and reasonable estimates for the uncertainties. Recalling the large contribution from the gluon FF to the overall cross section seen in Fig. II.2, the inclusion of $pp \rightarrow H + X$ data in the DSS analysis is also an important feature that is not present in any other FF with the same kinematic range.

Each fit is performed once with these uncertainties added in quadrature to the systematic uncertainties of the data and once without any modification of the data. This should also eliminate concerns over the ambiguity introduced by the choice of one FF over the others. Note that nCTEQ15 and nCTEQ15WZ used older BKK and KKP fragmentation functions for their neutral pion data. This should, however, not be a problem when comparing the old and new fits, as the central value of DSS is very close to the ones of BKK and KKP.

The PDF parameterization uses the 16 open parameters of nCTEQ15 as explained in Sec. I.2.1, but discard the $(\bar{u} + \bar{d})$ -dependend strange parameterization. Instead, we use the same functional as the remaining flavours with the three open parameters $a_0^{s+\bar{s}}, a_1^{s+\bar{s}}$ and $a_2^{s+\bar{s}}$. They describe the overall normalization, the low- x exponent and large- x exponent, respectively. This is the same parameterization that was used in nCTEQ15WZ.

Finally, we need to handle the large normalization uncertainties of the SIH data sets. As in the nCTEQ15WZ fit, we assign an additional normalization parameter to each independent data set and penalize the χ^2 value according to the prescription given in Sec. I.2.2.

We keep the normalizations of the W^\pm and Z boson data fixed at the values determined in the nCTEQ15WZ fit.

2.4.1. Main fits

We are now ready to use the single inclusive hadron production data to extend both the nCTEQ15 and nCTEQ15WZ fits. The reason why we perform a fit without the W^\pm and Z boson data is that this allows a comparison between the impact that each type of data has on the fit individually. Additionally, this gives us another perspective on the compatibility of the two processes by looking at the effect that each one's data has on the predictions for the other. The resulting PDF parameters of these two main fits are given in Tabs. IV.3 and IV.3.

Since the new data sets are taken on gold (RHIC) and lead (ALICE), which are both very heavy, there are no qualitative changes in the A -dependence of the resulting PDF. Therefore, we only present the lead PDFs resulting from the new fits in Figs. II.10 and II.11 for the nCTEQ15 and nCTEQ15WZ baseline, respectively. Each plot shows the respective baseline fit in black, the fit with unmodified data uncertainties in red and the fit with DSS uncertainties added to the data in green.

Taking a look at Fig. II.10, the most striking difference between the baseline and the new fits is the enhancement of the gluon at $x < 0.05$ and the corresponding suppression at higher x values. The central values of the quark PDFs see a similar, yet less pronounced enhancement at low x , driven by the DGLAP contributions from the gluon. Perhaps counterintuitively, the uncertainties of the quarks are larger in the new fits than in the baseline. The reason for this are the newly opened strange quark parameters, which give the parameterization more flexibility to reflect the true uncertainties. The same change was observed when these parameters were opened for the first time in the nCTEQ15WZ fit [39]. The difference observed between the new fits with and without DSS uncertainties is only very minor. The difference in the central value is barely visible, while the error bands for $x < 0.02$ are slightly enlarged when accounting for the FF uncertainty.

Figure II.11 shows the new fits, where nCTEQ15WZ was used as the baseline. As before, the baseline fit is shown in black, while the new fits with and without added uncertainties are shown in green and red, respectively. This time, we see a lesser, yet still significant enhancement in the low- x region and suppression in the high- x region of the gluons central value. This reduction of the impact is due to the fact that the gluon is already constrained more strongly by the inclusion of the WZ data. The up and down quarks see only minor enhancement at low x and no change in the size of their error bands. The strange quark, however, sees a surprisingly significant enhancement. Since Fig. II.1 showed no strong sensitivity to the strange quark in the SIH data, and no similar behavior was observed in the previous fits, the reason for this has to be the preference of the WZ data, given the

changed gluon. The difference in gluon uncertainty between the fits with and without DSS uncertainties are roughly as large as in the previous fit, but this time they are negligible for the quarks.

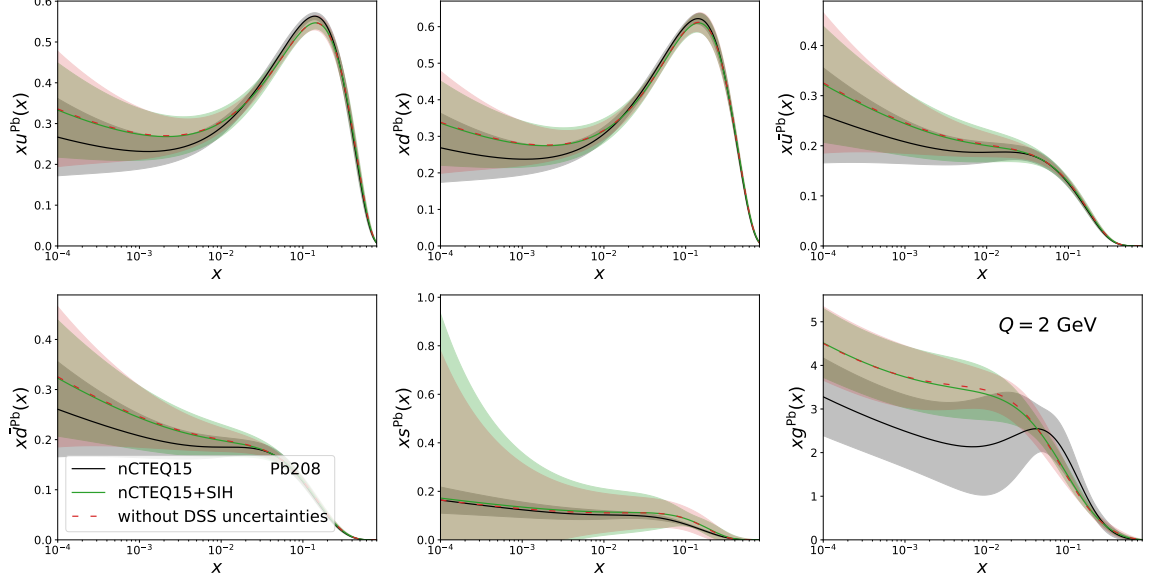


Figure II.10.: Lead PDFs from fits to the nCTEQ15 data + SIH data. The baseline nCTEQ15 fit is shown in black, the fit with unmodified data is shown in red and the fit where the uncertainties from the DSS fragmentation were added as a systematic uncertainty (nCTEQ15+SIH) is shown in green.

Quality and compatibility of the fits

The first quantity to investigate when judging the fit quality is naturally the resulting χ^2 . Figure II.12 shows the χ^2/N_{dof} for each data set involved in the two new main fits that will be denoted as nCTEQ15+SIH and nCTEQ15WZ+SIH from this point on. The DIS and DY data sets remain well described by the new PDFs and, aside from a few exceptions, contribute χ^2/N_{dof} values below 1. The most notable exceptions is the set with ID 5108 (EMC-1988, F_2^{Sn}/F_2^D), which has been noted as an outlier in other analyses [10, 191], as well. The second highest χ^2/N_{dof} comes from the set with ID 5160 (NMC-1997, F_2^D), which is barely affected by the new data, which is taken on very heavy nuclei. The WZ data is not quite as well described, but still well within statistical expectations. Only the data set with ID 6215 (ATLAS Run I, Z production) yields a χ^2/N_{dof} significantly larger than 1, but the same feature is present in the original nCTEQ15WZ analysis. The new SIH data sets yield very similar χ^2/N_{dof} values in both fits, with all of them being well below 1. The largest contributions come from the neutral pions, followed by the charged pions and then, finally, the charged kaons.

2. Single inclusive hadron production

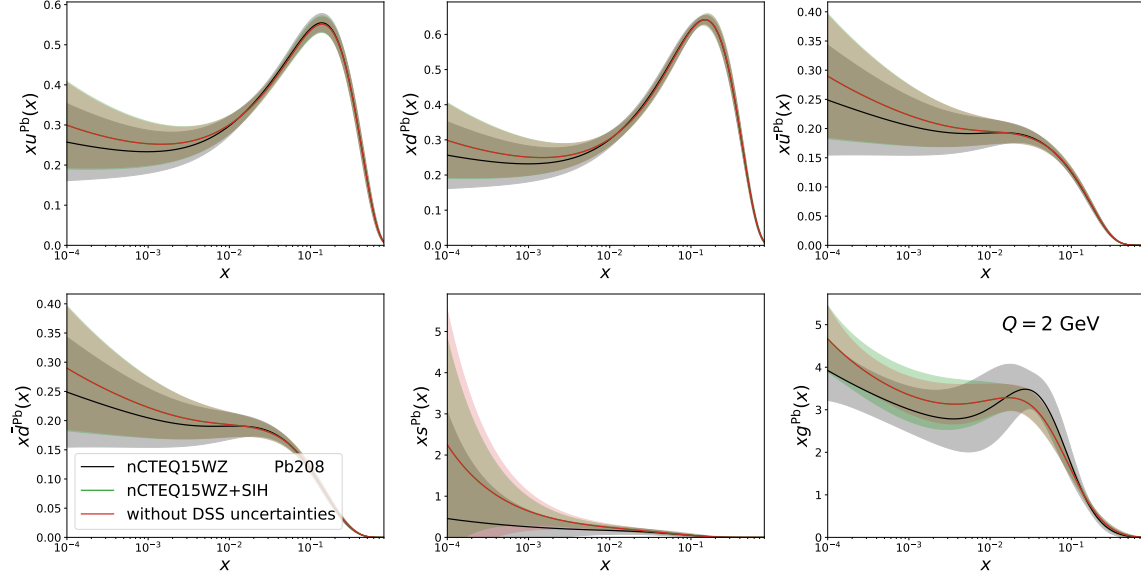


Figure II.11.: Lead PDFs from fits to the nCTEQ15WZ data + SIH data. The baseline nCTEQ15WZ fit is shown in black, the fit with unmodified data is shown in red and the fit where the uncertainties from the DSS fragmentation were added as a systematic uncertainty (nCTEQ15WZ+SIH) is shown in green.

Table ?? gives a more detailed, quantitative breakdown of the χ^2/N_{dof} values of the fits, including values for the data sets excluded in each fit. Going through the table row by row, we see a good description of DIS and DY data from the nCTEQ15 fit, but a very large χ^2/N_{dof} for WZ production. The SIH value of 1.23 looks good at first glance, but the fact that the nCTEQ15+SIH fit improves this value to just 0.38 shows that nCTEQ15 leaves a lot of room for improvement. Adding the SIH data also improves the χ^2/N_{dof} of the excluded WZ production, at a minor cost in terms of DIS fit quality. The χ^2/N_{dof} of the two RHIC neutral pion data sets that are included in nCTEQ15 actually increases slightly in nCTEQ15+SIH, but the significant improvements in all other SIH sets leads to an overall significantly lower χ^2/N_{dof} . Analogous behavior is observed in the nCTEQ15WZ fit, where the χ^2/N_{dof} of the included WZ data decreases significantly at the cost of a minor increase in DIS and the excluded SIH data improves compared to nCTEQ15. Finally, the nCTEQ15WZ+SIH fit including all data sets results in a good compromise between the previous two fits, where the WZ data and SIH data give slightly increased χ^2/N_{dof} values compared to their respective individual fits, but still show good agreement with the data. DIS and DY remain very similar to their value in nCTEQ15WZ.

The fitted normalization parameters of the SIH data sets are given in Tab. II.3. The majority of data sets yields normalization factors very close to one. The only exception are the charged and neutral pions from STAR, which require more significant normalization

shifts, but remain well within their large uncertainties. Accordingly, all the applied χ^2 penalties are small.

Table II.3.: Normalization uncertainties and fitted normalizations of the SIH data sets in the nCTEQ15WZ+SIH fit.

	STAR		PHENIX	ALICE			
	π^0	π^\pm	π^0	5 TeV π^0	5 TeV π^\pm	5 TeV K^\pm	8 TeV π^0
Norm. unc.	17%	17%	10%	6%	6%	6%	3.4%
Fitted norm.	0.942	0.866	1.010	0.995	0.994	1.021	1.021
Penalty	0.12	0.68	0.01	0.01	0.01	0.13	0.41

Correlation between individual data sets and PDFs

Looking at the resulting PDFs and χ^2 values alone does not tell the full story of the new data's impact on the fits. Naturally, we also want to know how each specific data set influences the final PDFs. To investigate this matter, we first take a look at the cosine of

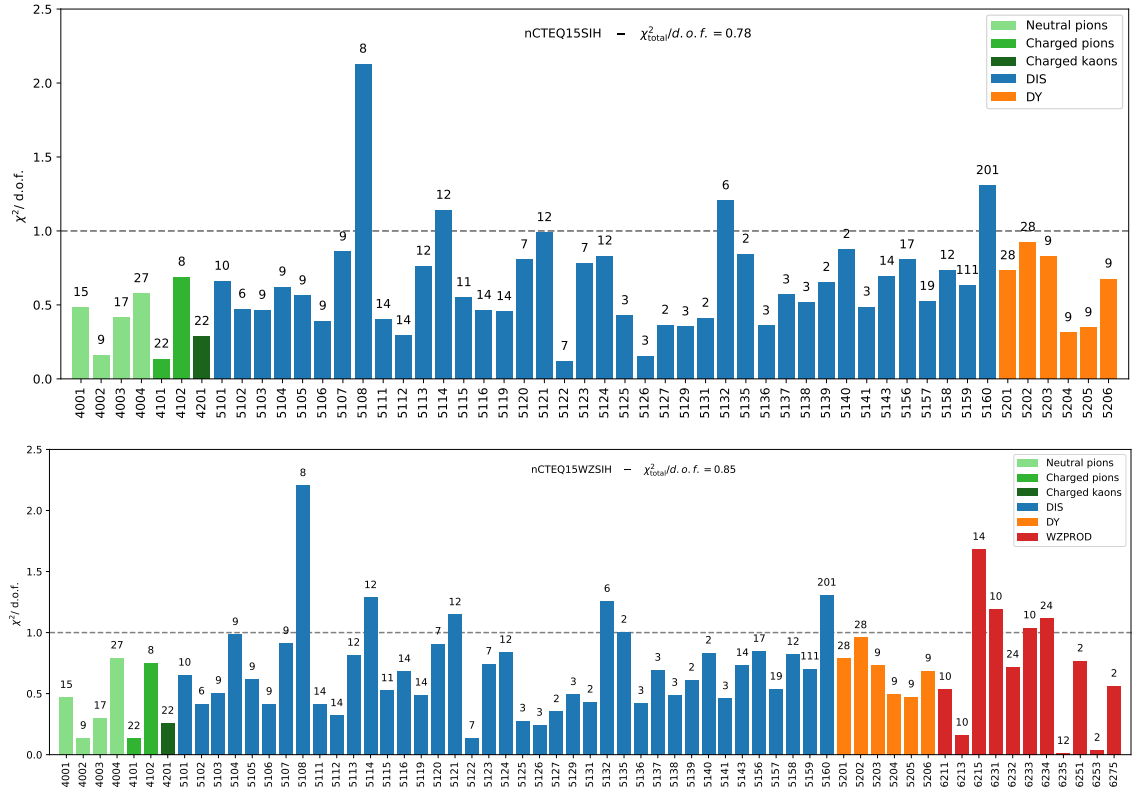


Figure II.12.: The χ^2/N_{dof} of the individual experiments for the nCTEQ15+SIH fit (top panel) and the nCTEQ15WZ+SIH fit (bottom panel). The annotations at the top of each bar indicates the number of data points in the set.

the correlation angle $\cos(\phi[g(x, Q), \chi^2(j_{\text{exp}})])$ between the gluon PDF $g(x, Q)$ in lead and the χ^2 -difference of each data set as defined in Sec. I.2.6.2. Since this quantity is only weakly dependent on the scale Q , we present it only for $Q = 2 \text{ GeV}$, the same scale as the PDFs shown in Fig. II.11. In the cosines shown in Fig. II.13, the most striking feature is the strong anti-correlation between the gluon PDF and the 5 TeV ALICE data sets at low x . This implies that these sets are likely responsible for the change in the gluon PDF's central value, as many of the other data sets with a strong impact on the gluon are positively correlated. The neutral pion data taken at 8 TeV by ALICE shows a very similar, but more pronounced behavior as the NMC96 Sn/C data set, which is the most dominant DIS contribution to the gluon PDF due to its large Q coverage and number of data points. At higher x values the correlations become more mixed which explains the fact that the gluon PDF remains largely unchanged in this region.

To judge the magnitude of the impact that individual data sets have on the final fit, we take a look at the χ^2_{eff} in Fig. II.14. The biggest impact on the gluon PDF still comes from the CMS Run 2 W^\pm data and the large NMC-1996 $F_2^{\text{Sn}}/F_2^{\text{C}}$ data set. Next are the two neutral pion data sets from ALICE with a similar impact as NMC-1995 $F_2^{\text{Ca}}/F_2^{\text{D}}$ and CMS Run 1 W^+ . The impact of the new SIH data is lessened significantly by the fact that more than half of the data is cut to satisfy the p_T constraints, especially since the low p_T data points are generally more precise.

Comparison of data and theory

One of the most important verifications of the PDFs is a direct comparison with the data. Therefore, we show another comparison of the theory predictions with the R_{pPb} data. This time, we compare predictions made with the new PDFs nCTEQ15(WZ)+SIH and their respective progenitors. Figure II.15 shows that the predictions made using nCTEQ15WZ and either of the new fits show very good agreement with the data and with each other. nCTEQ15 also agrees in the region of high transverse momentum but falls off more steeply in the $p_T < 10 \text{ GeV}$ region, which indicates stronger nuclear modification. The fact that the differences between PDFs are relatively small should not come as a surprise, given the small differences in χ^2 that we discussed in Tab. ??.

Verification with other fragmentation functions

As the final check of the fit quality, we want to ensure that our fit is compatible with all of the available FFs. To do this, we take the new nCTEQ15WZ+SIH PDFs to calculate the χ^2/N_{dof} of the included SIH data with each FF. If they are compatible, we should only see a minor change in χ^2 . To estimate the range of acceptable χ^2/N_{dof} values for this, we can

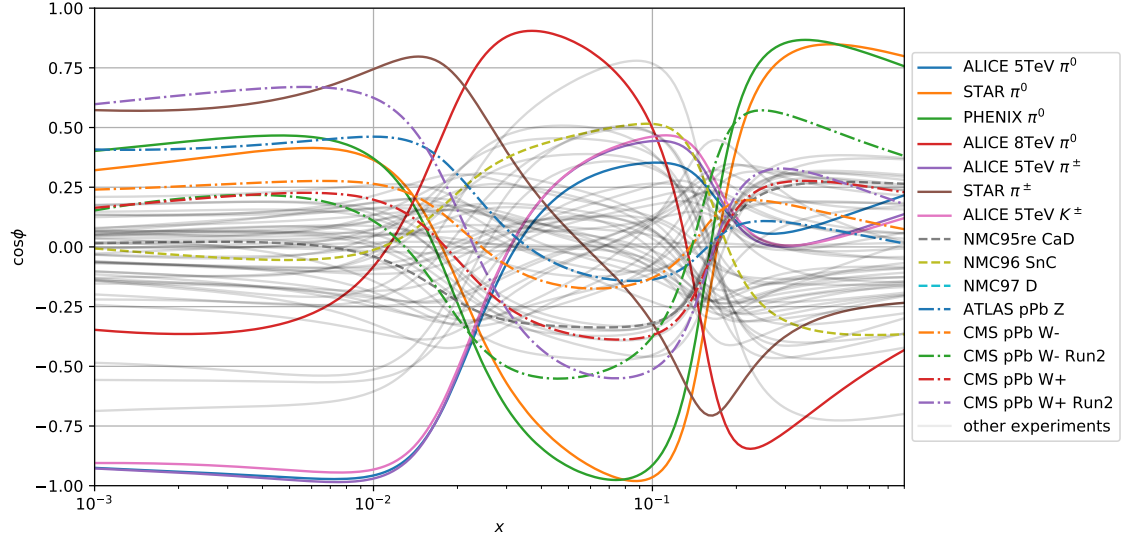


Figure II.13.: Cosine of the correlation angle between gluon PDF at $Q = 2 \text{ GeV}$ and χ^2 of each data set for the nCTEQ15WZ+SIH fit.

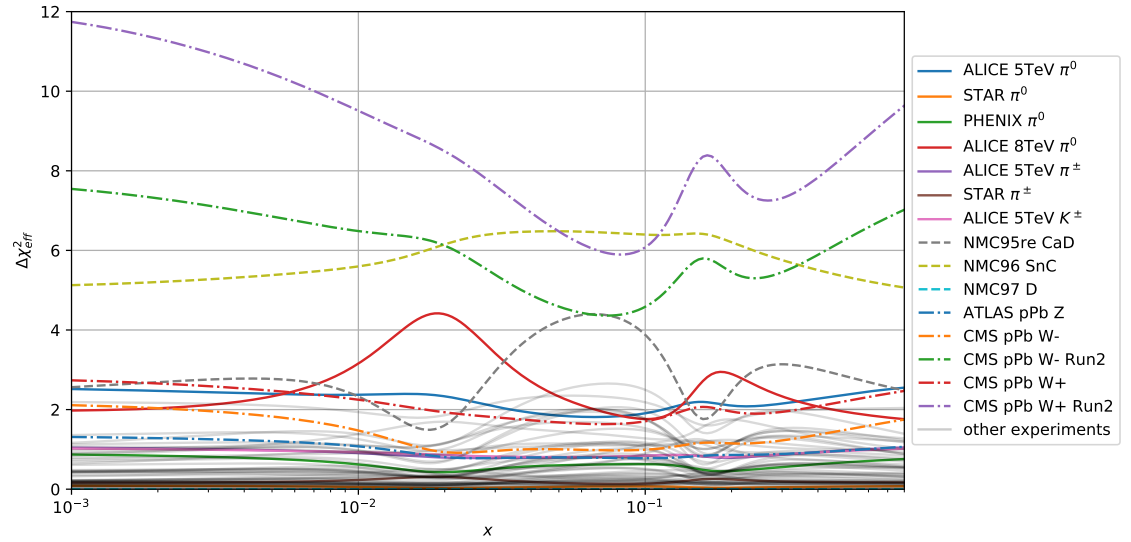


Figure II.14.: $\Delta\chi^2_{\text{eff}}$ for the gluon PDF at $Q = 2 \text{ GeV}$ with each data set for the nCTEQ15WZ+SIH fit.

2. Single inclusive hadron production

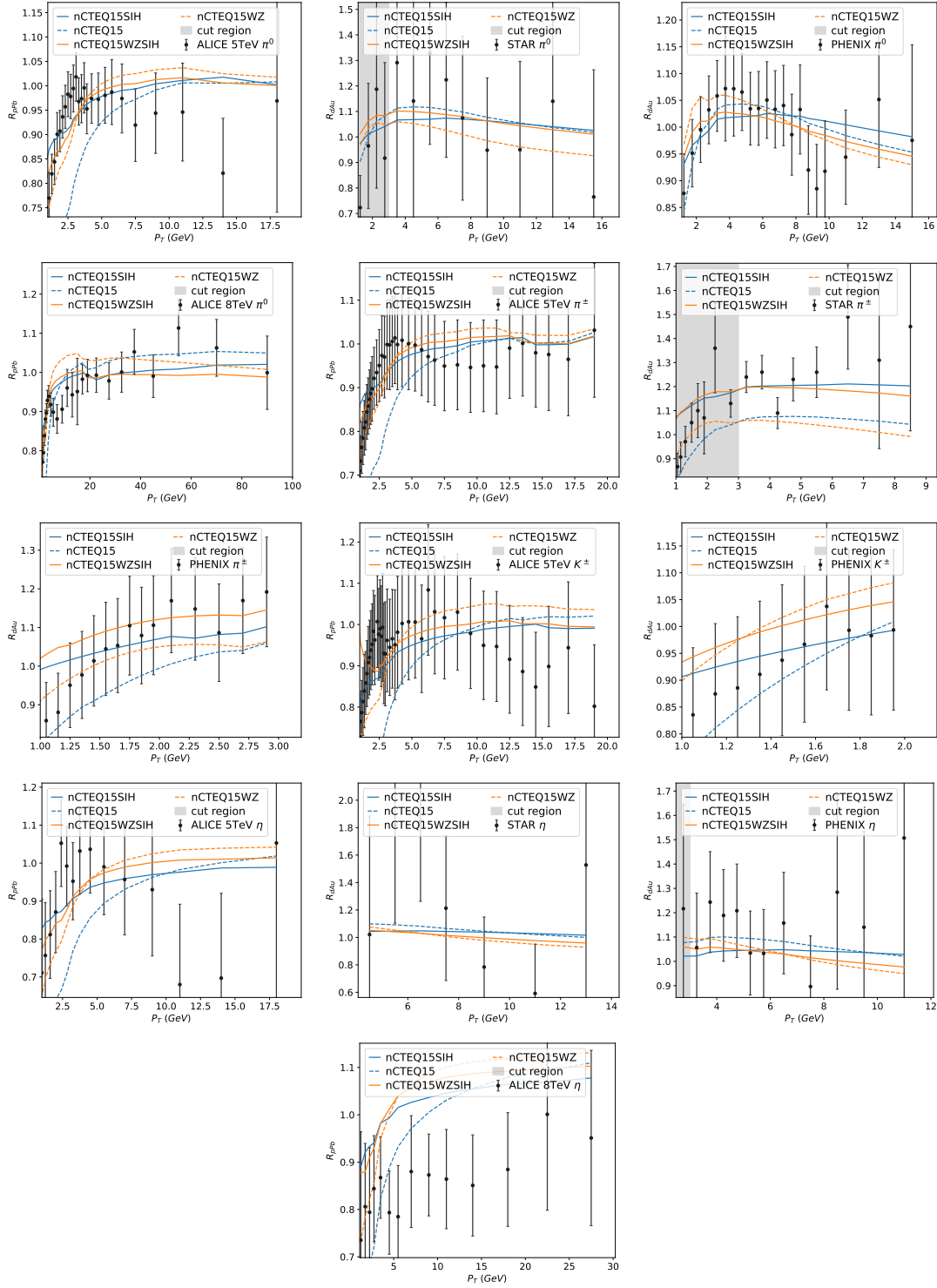


Figure II.15.: Theory predictions for the main fits and their respective baselines. Dashed curves indicate the baseline fits, while solid curves show the fits with SIH data included. The blue curves are based on nCTEQ15 and the orange ones on nCTEQ15WZ.

also calculate it again for DSS, but without the FF uncertainties added to the data. As long as the χ^2/N_{dof} obtained with the other FFs lies below this value, the added uncertainties should compensate for the ambiguity introduced by the FF choice. In Tab. II.4, we can see that this is indeed the case for three of the four other FFs. KKP actually gives a χ^2/N_{dof} slightly lower than DSS, while BKK is just above. NNFF is somewhat higher still, but again below the value of DSS with unmodified data. The sole exception comes from the JAM20 FF, which can easily be explained by remembering Fig. II.4, which showed that JAM20 started to disagree with the proton-proton baseline at higher p_T values than other FFs. The fact that these FFs are determined together with a set of proton PDFs might explain this discrepancy, if the JAM20 PDFs are incompatible with our proton PDF baseline.

Table II.4.: $\chi^2/N_{d.o.f.}$ values of the single inclusive hadron data obtained by using different fragmentation functions. The PDF parameters are taken from the nCTEQ15WZ+SIH fit.

DSS (unmodified data)	DSS	KKP	BKK	NNFF	JAM20
0.461	0.412	0.401	0.420	0.456	0.553

2.4.2. Alternative fits

In the final analysis of this section, we take one more look at the previously excluded η data to find out what the impact of adding these data sets would be. Since there is no single fragmentation function that can describe pions, kaons and η mesons, we use AESSS for the η mesons and keep DSS for the rest. We also keep the uncertainties of the DSS fragmentation functions in the pion and kaon data, but the η meson data remains unmodified as the AESSS fragmentation does not provide any uncertainties. Applying the same p_T cut as for the other data sets, we get an additional 37 data points (18 from RHIC, 19 from ALICE) for these new fits. The resulting PDFs are shown in Figs. II.16 and II.17 for the nCTEQ15 and nCTEQ15WZ baseline, respectively.

In the nCTEQ15 based fit the η meson production data causes an increase in the low x gluon PDF and a similar, but significantly less pronounced effect on the quark flavours. The gluon uncertainties remain mostly unchanged, while the up and down quark uncertainties actually increase. The increase in up and down quark uncertainty is somewhat compensated by a decrease in the strange quark uncertainty, which likely comes from the more general problem that the current data is not sufficient to cleanly separate the different flavours. In the fit with WZ data, the impact of the η meson production is barely visible. The central values of all flavours remain basically unchanged and the uncertainties are only decreased by a small amount. Additionally, the uncertainties of the η meson FFs are certainly larger than those of pions and kaons, which means that this impact would be even further diminished

2. Single inclusive hadron production

if we could add such uncertainties to the systematics of the data.

Table II.5.: χ^2/N_{dof} values for the individual SIH data sets. The shown χ^2 is the sum of regular χ^2 and normalization penalty. Excluded processes are shown in parentheses.

$\chi^2/N_{d.o.f.}$ for selected experiments							
	STAR		PHENIX	ALICE 5 TeV			ALICE 8 TeV
	π^0	π^\pm	π^0	π^0	π^\pm	K^\pm	π^0
nCTEQ15	0.13	(2.68)	0.30	(2.53)	(0.62)	(0.71)	(1.96)
nCTEQ15+SIH	0.16	0.69	0.41	0.48	0.13	0.29	0.58
nCTEQ15WZ	0.17	(3.24)	0.23	(0.67)	(0.21)	(0.41)	(1.58)
nCTEQ15WZ+SIH	0.14	0.75	0.30	0.47	0.13	0.26	0.79

Table II.6.: χ^2/N_{dof} values for the individual processes DIS, DY, SIH, WZ, and their total. The shown χ^2 is the sum of regular χ^2 and normalization penalty. Excluded processes are shown in parentheses (note that both nCTEQ15 and nCTEQ15WZ contain a minor part of the SIH data, namely the neutral pions from STAR and PHENIX).

$\chi^2/N_{d.o.f.}$ for each process					
	DIS	DY	WZ	SIH	Total
nCTEQ15	0.86	0.78	(3.74)	(1.23)	1.28
nCTEQ15+SIH	0.87	0.72	(2.32)	0.38	1.00
nCTEQ15WZ	0.90	0.78	0.90	(0.81)	0.90
nCTEQ15WZ+SIH	0.91	0.77	1.02	0.41	0.85

2.5. Conclusions

Including the single inclusive hadron production data in a new global nPDF fit required a variety of challenges to be overcome. The calculation of the cross section required nested convolution of the PDFs, the partonic subprocess and the FFs and depends on three different scales.

The numerical evaluation of the nested convolutions takes considerable amounts of time, which made the calculation prohibitively expensive in terms of computation time. This problem was solved by computing grids ahead of time that include the partonic process, the proton PDF and the FF, such that only the convolution with the nuclear PDF remains.

The aforementioned FF dependence required additional care as well, since they are non-perturbative objects that are determined in global fits similar to those of PDFs. The impact of choosing a particular set of FFs was controlled by first comparing the predictions to proton-proton data and finding the appropriate p_T region, where the FFs are compatible

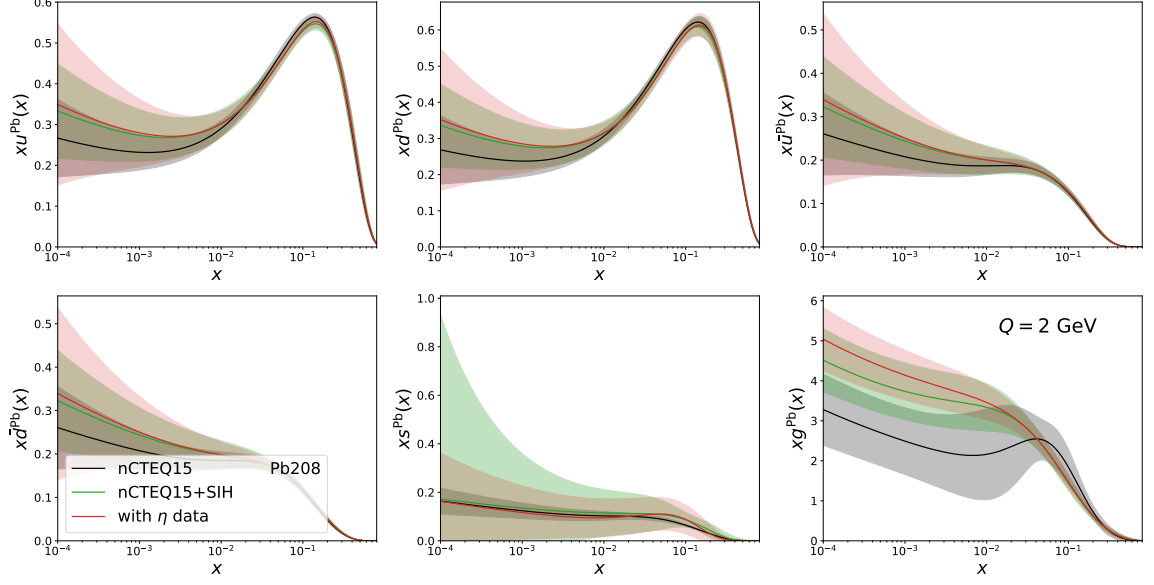


Figure II.16.: Lead PDFs at $Q = 2$ GeV extracted in fits to the nCTEQ15 data + SIH data. The baseline nCTEQ15 fit is shown in black, the fit with η meson data is shown in red and the corresponding main fit is shown in green.

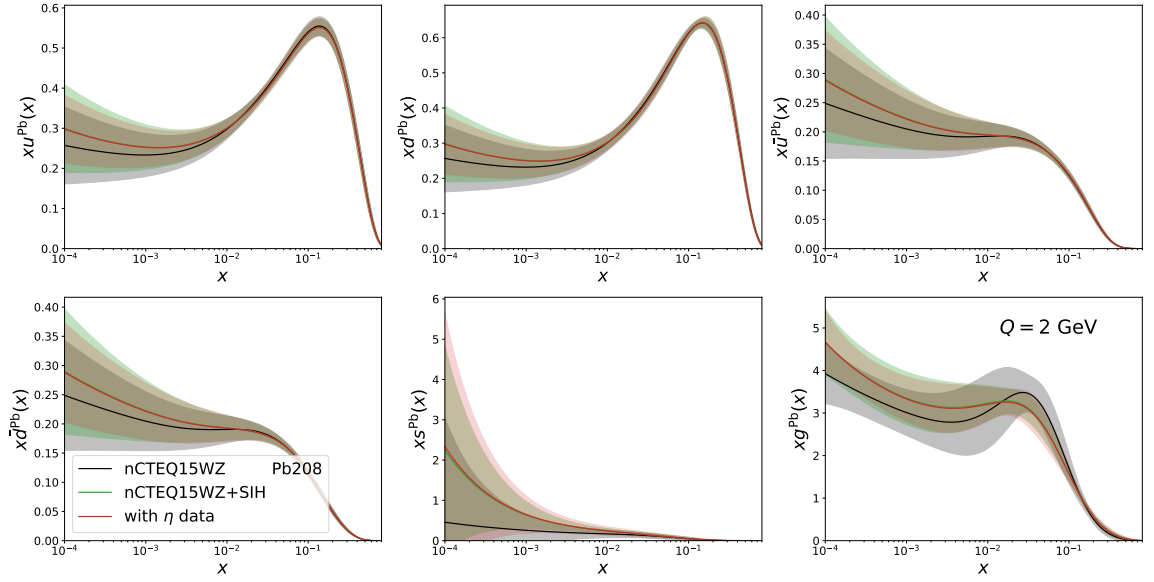


Figure II.17.: Lead PDFs at $Q = 2$ GeV extracted in fits to the nCTEQ15WZ data + SIH data. The baseline nCTEQ15WZ fit is shown in black, the fit with η meson data is shown in red and the corresponding main fit is shown in green.

with the data. Additionally, the uncertainties calculated from the eigenvectors of the DSS FFs were added to the data as a systematic uncertainty to quantitatively account for the ambiguity from the FF choice.

To determine the optimal renormalization and factorization scales another comparison with proton-proton data was made, yielding a unique choice that could describe the data above $p_T > 3$ GeV.

The resulting PDFs show an enhanced gluon at low x compared to the nCTEQ15 and nCTEQ15WZ baselines and have reduced uncertainties in the region around $x \approx 0.01$. The impact on the quark PDFs was minimal and is mainly due to DGLAP contributions from the changed gluon. The PDFs are still well compatible with the previous data sets and yield excellent χ^2 values for the single inclusive hadrons. Correlation analyses show that the main impact comes from the newer ALICE data, while the RHIC data sets only makes minor contributions.

The restrictive cuts removing the data with $p_T < 3$ GeV are a significant hindrance to the potential impact, since this is the region with the most precise data. Future improvements in the theoretical predictions are necessary to allow less restrictive cuts and open up more opportunities for these valuable data sets.

Nevertheless, the nCTEQ15WZ+SIH PDFs present an important step towards the next generation of nuclear PDFs with a deeper understanding of these data sets and the gluon PDF.

3. Heavy quark production

In the previous section, we have shown how data on single inclusive pion and kaon production can help in constraining the nuclear gluon PDF even when compensating for the theoretical uncertainties introduced by the need for final state fragmentation functions. However, the gluon PDF uncertainty still remains significantly larger than those of valence quarks and the $x < 10^{-3}$ region remains unconstrained by the data.

For this reason, we will now focus on a similar process, where significantly more data is available over a wider kinematic range: single inclusive production of *heavy* mesons (in the following called heavy quark production or abbreviated as HQ). This includes both open heavy flavored mesons, *i.e.*, open heavy flavor mesons with exactly one heavy (anti-)quark and quarkonia consisting of a heavy quark and its antiquark. Figures II.19 and II.20 show the kinematic coverage of the data in terms of transverse momentum p_T and center-of-mass rapidity y_{cms} . The black contours give an estimate of the x -dependence of the data, which is given at leading order by

$$x \approx \frac{2p_T}{\sqrt{s}} \exp(-y_{\text{cms}}). \quad (131)$$

This shows that the low- p_T data at very forward rapidity, which is defined to be the proton beam's direction, is sensitive to x values down to 10^{-5} .

The process is very sensitive to the gluon PDF in particular, as the gluon-gluon initiated channel dominates the cross section. This was first shown in a proton PDF analysis in Ref. [192] and followed by Ref. [193], which used a Bayesian reweighting approach to demonstrate the potential use in nuclear PDF extractions. Since then, D^0 meson production data in particular has been used to great effect to reduce the gluon uncertainty in Refs. [14, 194]. The former analysis uses the ratio of double-differential cross sections between proton-proton and proton-lead with theoretical calculations performed in the general-mass variable-flavor-number scheme (GMVFNS) [195–197] in a global fit. In contrast, the latter uses the same data with theoretical predictions from a fixed-flavor number scheme with POWHEG [198] and Pythia8 [199] in a reweighting of their PDFs.

In this thesis, we will include not just D^0 -meson production, but also quarkonium production in a new global nPDF fit using a data-driven approach for the theoretical calculations, that was also used in the original reweighting study [193] after it was initially described in Ref. [200]. This data-driven approach allows us to circumvent the problem that there is currently no universally accepted model for the production of quarkonia in hadronic collisions. In this approach, we fit an effective scattering matrix element to data taken in proton-proton collisions and use that to perform the calculations for the proton-lead data in the nPDF fit. This modifies the nPDF fitting procedure as visualized in Fig. II.18.

The empirical model we use assumes the dominance of the gluon-gluon channel in the

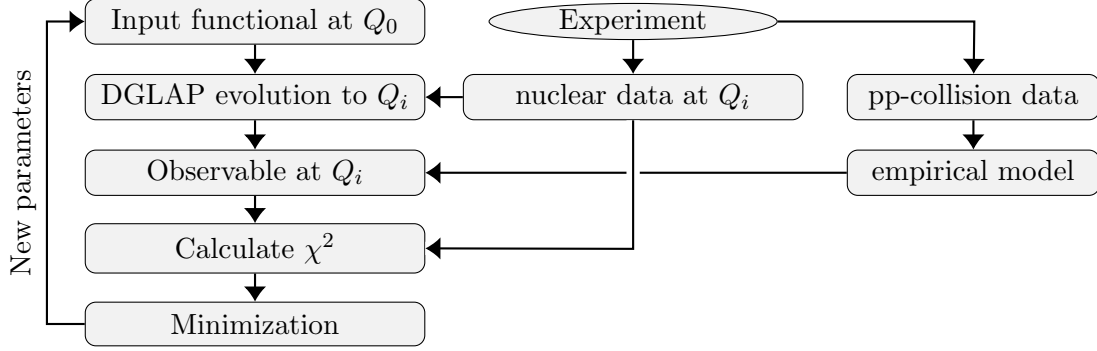


Figure II.18.: Flowchart representation of the PDF optimization with the empirical model for heavy quark production.

total cross section and that it is sufficient to look at partonic subprocesses with $2 \rightarrow 2$ kinematics. As with all other data, we also assume the twist-2 factorization to be valid in pA collisions and that higher twist contributions can be neglected. However, it has been shown that this is not generally the case for heavy quark production, due to fully coherent energy loss effects [201–203], which can account for up to 50% of the nuclear modification of heavy quark production at low p_T . Therefore, we need to apply sufficiently strict kinematic cuts on the data to ensure that we do not absorb such final state effects into the nuclear PDFs.

We begin this section with a brief review of existing models for quarkonium production, before explaining the empirical method we will use for our calculations. In Sec. II.3.2, we then determine the parameters that enter the empirical model from data taken in proton-proton collisions, identify reasonable kinematic cuts and verify the model against calculations performed in the GMVFNS for D^0 production and in non-relativistic QCD (NRQCD) for J/ψ production. The new global nPDF fit with heavy quark production data is then performed in Sec. II.3.3, which also includes a thorough analysis of the compatibility between new and existing data. Section II.4 then provides comparisons of the resulting PDFs with those of other groups and shows the impact on various observables. Finally, Sec. II.3.4 summarizes the conclusions of the analysis and gives an outlook towards possible improvements of the study.

3.1. Theoretical framework

Neither the production or open heavy flavored mesons, nor the production of quarkonia is fully understood from the perspective of perturbative QCD. In former case, multiple schemes for the treatment of the massive quarks in the initial state and for the final state hadronization. In the latter case, there is discussion about the production of the quark antiquark pair and its hadronization. Over the past decades a variety of perturbative calculations have been performed in different frameworks attempting to describe the processes.

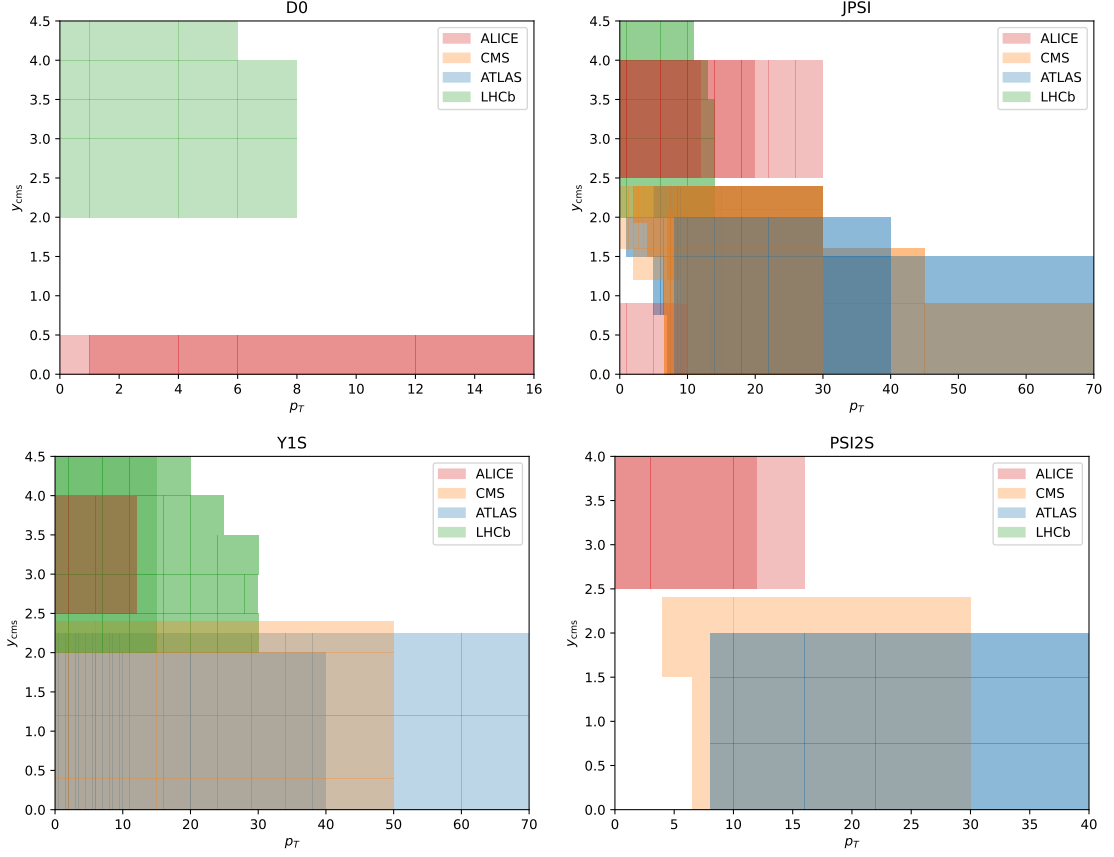


Figure II.19.: Coverage of the kinematic (p_T, y_{cms}) -plane of the open heavy quark and quarkonium production data from proton-proton collisions. ALICE data is shown in red, CMS in orange, ATLAS in blue and LHCb in green.

Different assumptions entering these pQCD calculations lead to a multitude of models and no consensus exists yet on which model is the best.

3.1.1. Different schemes for the production of open heavy flavored mesons

The schemes used to calculate cross sections involving heavy quarks can be broken down into two classes: Fixed-flavor-number schemes (FFNS) and variable-flavor-number schemes (VFNS). As the name suggests, the former considers the same number of flavors to exist at any scale Q , while the latter uses a different number of active flavors depending on the scale. More detailed overviews of heavy flavor production can be found in the two recent reviews Refs. [204, 205].

Fixed-flavor-number schemes

Fixed-flavor-number schemes [206, 207] are conceptually very simple, since the heavy quark Q is neglected in the initial state and the remaining quarks are treated as massless. The

3. Heavy quark production

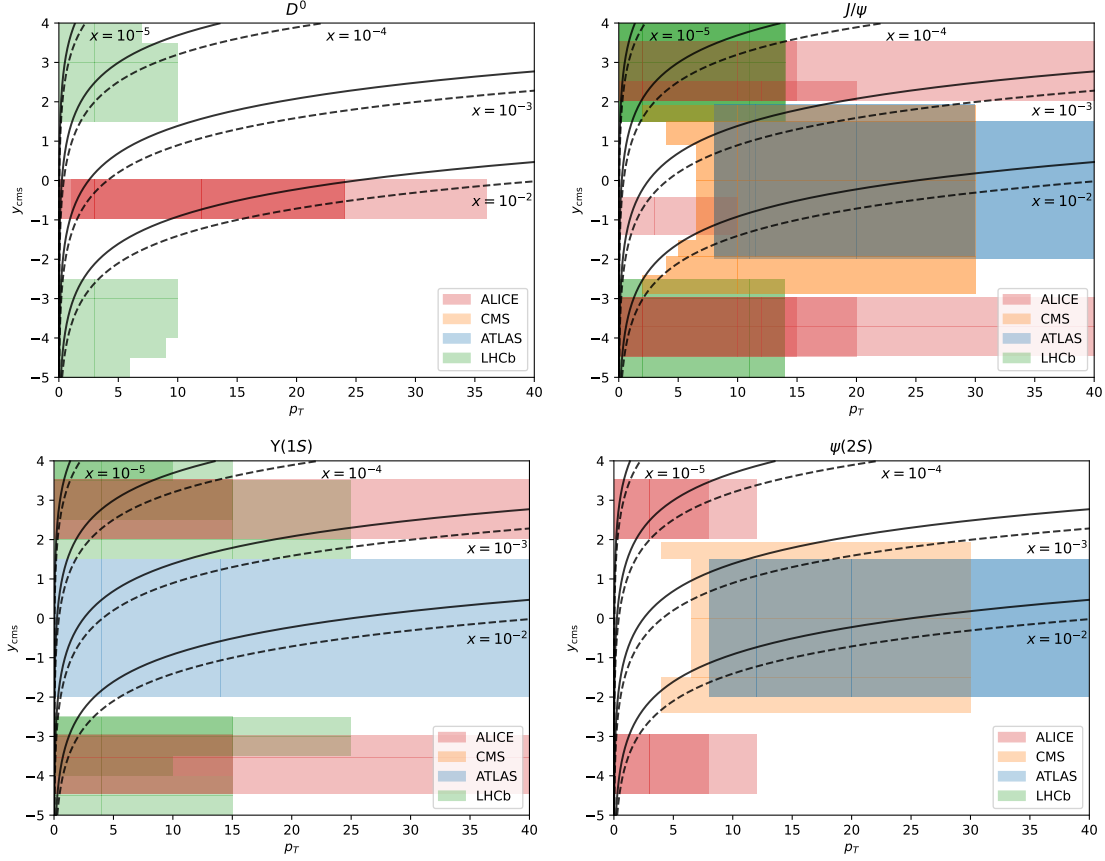


Figure II.20.: Coverage of the kinematic (p_T, y_{cms}) -plane of the open heavy quark and quarkonium production data from proton-lead collisions. ALICE data is shown in red, CMS in orange, ATLAS in blue and LHCb in green. The contours show the estimated x -dependence of the lead PDF for $\sqrt{s} = 5$ TeV (solid) and 8 TeV (dashed) respectively.

cross section for the production of a heavy quark Q can then be expressed as

$$d\sigma^{Q+X} = \sum_{i,j} \iint dx_i dx_j f_i^A(x_i, \mu_i) f_j^B(x_j, \mu_j) d\hat{\sigma}_{ij \rightarrow Q+X}(\mu_i, \mu_R), \quad (132)$$

where f_i^A and f_j^B are the PDFs of flavor i and j inside the hadrons A and B , which depend on the factorization scale μ_F . The partonic cross section $d\hat{\sigma}$ also depends on the factorization scale, as well as the renormalization scale. The sum includes only the active partons in the initial state hadron, *e.g.*, $i, j \in \{u, \bar{u}, d, \bar{d}, s, \bar{s}, g\}$ in the case of three active flavors (3-FFNS), which can be used for both charm and bottom production. The partonic cross section at NLO contains terms proportional to powers of $\left[\alpha_s \log\left(\frac{p_T}{m_Q}\right)\right]$, which become singular in the massless limit and also become large when $p_T \gg m_Q$, such that the perturbative expansion breaks down. Therefore, the calculation is only applicable

in the region $0 \leq p_T \lesssim 5m_H$, as shown in Ref. [208]. A factorization ansatz for the higher p_T region allows for the derivation of perturbative fragmentation functions (PFF), which can be convoluted with the differential cross section [209]. The convolution with this scale-independent FF $D_Q^H(z)$ describes the transition of the heavy quark with momentum p_Q to the observed heavy meson H with momentum $p_H = zp_Q$. It is important to stress that this factorization ansatz is not based on a rigorous derivation like the one for PDFs in Eq. (126), but rather motivated purely by phenomenology. Alternatively, the parton-level calculation can be matched to a parton-shower from a Monte-Carlo event generator, which can then be hadronized using the model of the generator.

Zero-mass variable-flavor-number schemes

As mentioned in the previous section, the logarithms $\left[\alpha_s \log\left(\frac{p_T}{m_Q}\right)\right]$ grow very large for $p_T \gg m_Q$ and need to be resummed in order to obtain reasonable results in this region. This can be done by absorbing them into the PDFs and FFs with scale-dependence governed by the DGLAP evolution. This requires the heavy quark to be considered as a parton for factorization scales $\mu_i > \mu_T$ for some transition scale μ_T , which is usually identified with the heavy-quark mass. This transition between different flavor numbers with changing scales is the defining characteristic of variable-flavor-number schemes (VFNS). The prefix *zero-mass* implies that the heavy-quark mass is assumed to be negligible in the calculation of the short-range cross section. This essentially mirrors the framework used for the light meson production from Sec. II.2.1 and leads to a similar factorization theorem

$$d\sigma^{Q+X} = \sum_{i,j,k} \int \int \int dx_i dx_j dz f_i^A(x_i, \mu_i) f_j^B(x_j, \mu_i) d\tilde{\sigma}_{ij \rightarrow k+X} D_k^H(z, \mu_f) + \mathcal{O}\left(\frac{m_Q^2}{p_T^2}\right), \quad (133)$$

where the error term becomes small only at sufficiently large p_T . In this case, the FFs can be determined in two different ways: The first way, named Binnewies-Kniehl-Kramer (BKK) approach [210–212], is equivalent to the determination of FFs for the light meson production, where the FFs are considered to be one non-perturbative object that is determined from data at some scale μ_0 and $D_k^H(z, \mu_f)$ is obtained through DGLAP evolution. The second method [209] decomposes the full FF into a perturbative FF $D_k^Q(z, \mu_f)$ for the fragmentation of the parton k into the heavy quark Q and a scale-independent FF $D_Q^H(z)$, which describes the non-perturbative hadronization of the heavy quark into the hadron H . In this approach, the PFFs resum the final-state logarithms and receive their scale dependence from DGLAP evolution.

General-mass variable-flavor-number scheme

The previous two schemes are valid only in restricted, but complementary kinematic regions. The general-mass variable-flavor-number scheme (GMVFNS) [196, 197] provides

3. Heavy quark production

a framework that is valid across the entire transverse momentum range at NLO+NLL accuracy. The cross section for inclusive heavy flavor production is given by a similar factorization formula as in the zero-mass case

$$d\sigma^{Q+X} = \sum_{i,j,k} \iiint dx_i dx_j dz f_i^A(x_i, \mu_i) f_j^B(x_j, \mu_i) d\hat{\sigma}_{ij \rightarrow k+X} D_k^H(z, \mu_f), \quad (134)$$

but this time the finite heavy-quark-mass terms proportional to powers of $\frac{m_Q^2}{p_T^2}$ are fully contained in the short-distance cross section for subprocesses with heavy quarks in the final state. They are, however, currently neglected in the subprocesses initiated by heavy quarks. The hard scattering cross sections are defined such that they converge to the massless ones in the limit $\frac{m_Q}{p_T} \rightarrow 0$, which means that the ZMVFNS is approached at large p_T . Additionally, the GMVFNS can recover the FFNS at small p_T . This requires the factorization scale to be chosen such that the heavy-quark PDFs are removed at sufficiently high p_T , because the limit $p_T \rightarrow 0$ yields divergent cross sections from the massless charm or bottom initiated subprocesses.

Fixed-order plus next-to-leading logarithms

The fixed-order plus next-to-leading logarithms (FONLL) approach [213] is another unified framework for the entire kinematic range. The approach matches the massive NLO cross section of the FFNS with that of the massless NLO calculation of the ZMVFNS according to the prescription

$$d\sigma_{\text{FONLL}} = d\sigma_{\text{FFNS}} + (d\sigma_{\text{ZMVFNS}} - d\sigma_{\text{FFNS},0}) \times G(m_Q, p_T), \quad (135)$$

where $d\sigma_{\text{FFNS},0}$ is the cross section in the FFNS in the asymptotic limit $p_T \gg m_Q$, where finite mass terms can be neglected. To recover the FFNS in the low- p_T limit and ZMVFNS in the high- p_T limit, the interpolation function $G(m_Q, p_T)$ needs to fulfill the following two conditions:

$$\lim_{p_T \rightarrow 0} G(m_Q, p_T) \rightarrow 0 \quad \text{and} \quad \lim_{\frac{p_T}{m_Q} \rightarrow 0} G(m_Q, p_T) \rightarrow 1. \quad (136)$$

The second condition is equivalent to the suppression of the heavy quark PDF in the GMVFNS described above, such that the GMVFNS can be recovered in this framework if $G(m_Q, p_T) = 1$. In the FONLL, however, the common choice is

$$G(m_Q, p_T) = p_T^2 / (p_T^2 + a^2 m_Q^2), \quad (137)$$

with $a = 5$ determined from phenomenological studies. Another difference between FONLL and GMVFNS is that the former is usually combined with FFs in the PFF formalism, while the latter is used in conjunction with the BKK approach.

3.1.2. Perturbative models for the production of quarkonia

In this section, we will briefly highlight the color-evaporation model, the color singlet model and non-relativistic QCD which includes the color singlet model, as well as additional color-octet contributions. For an extensive overview of these models and their phenomenology, we refer to a recent review by Lansberg [214].

Color-evaporation model

The color-evaporation model [215, 216] directly links the production cross section of quarkonia \mathcal{Q} to that of quark-antiquark pairs with an invariant mass M that allows hadronization. This region lies between the mass of the quark pair $2m_Q$ and the mass of the lightest open heavy-flavored hadron pair $2m_H$, *e.g.*, $2m_c < M < 2m_{D^0}$ in the case of charmonium. It is then assumed that the pair randomly emits soft particles until it hadronizes, such that the overall production cross section reads

$$d\sigma_{\mathcal{Q}} = F_{\mathcal{Q}} \int_{2m_c}^{2M_D} dM \frac{d\sigma_{c\bar{c}}(M)}{dM}, \quad (138)$$

where $F_{\mathcal{Q}}$ is a universal constant, independent of the kinematics, spin of \mathcal{Q} and $q\bar{q}$ production process. While this simple model can explain J/ψ production data, it has proven incompatible with the experimental measurements of the cross section ratio between different ψ states. In the color-evaporation models these ratios should be independent of the kinematics, but experimental data does show a dependence on the transverse momentum [217, 218]. More recently, an improved color-evaporation model [219] has been put forward to rectify these discrepancies, but the tensions do not fully disappear [220].

Color-singlet model

At the most general level, the production of quarkonia can be divided into two parts: color-singlet and color-octet contributions. The color-singlet contributions include all processes where the intermediate state is a color neutral $q\bar{q}$ pair with the same quantum numbers as the final hadron [221]. The color octet contributions, however, allow for arbitrary color and other quantum numbers in the intermediate state. Figure II.21 provides an illustration of the difference between the two processes.

The color-singlet model assumes that the bound quarks move at non-relativistic speeds in the bound state's rest frame and that their wave function has a sharp peak in momentum space. Using these assumptions, it can be shown that the partonic cross section of the

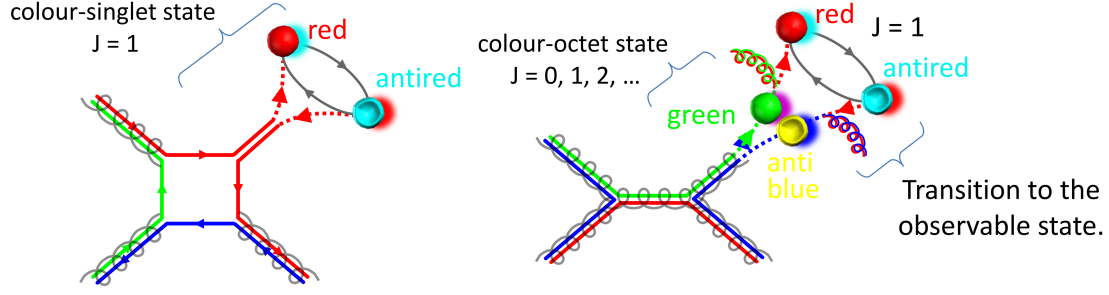


Figure II.21.: Illustrations of quarkonium production in a color singlet (left) and a color octet process (right). Illustrations by Pietro Faccioli (https://idpasc.lip.pt/uploads/talk/file/530/LIP_curso_polarization.pdf).

quarkonium production is equal to the production cross section of a heavy quark pair with zero relative velocity and the same quantum numbers as the final hadron multiplied with the squared Schrödinger wave function at the origin in position space. The cross section for quarkonium production in hadronic collisions is then given by

$$d\sigma_Q = \sum_{i,j} \int dx_i dx_j f_i(x_i, \mu_f) f_j(x_j, \mu_f) d\hat{\sigma}_{i+j \rightarrow (q\bar{q})+X}(\mu_R, \mu_f) |R(0)|^2, \quad (139)$$

with $f_{i,j}(x, \mu_f)$ being the parton densities of the colliding hadrons.

This model has faced numerous phenomenological challenges and, even at NLO, cannot give a reasonable description of some data sets [222, 223]. These discrepancies, however, are reduced when the partially available NNLO corrections are taken into account.

A second problem with this model arises when looking at P -wave decays, where infrared divergences arise unless a binding energy is introduced to regulate these divergences [224]. A different solution to this problem is introduced in the more general framework of non-relativistic QCD in the form of the color-octet mechanism.

Non-relativistic QCD

The basis for NRQCD is the observation that the high mass of the charm and bottom quarks causes them to move slowly compared to speed of light in the meson's rest frame. An expansion in powers of the velocity v can then be used to derive a factorization theorem [225] that separates the perturbative short-distance physics of the $q\bar{q}$ production from the non-perturbative long-distance physics of the hadronization process. At leading order in v , this approach recovers the color-singlet model while higher orders introduce color-octet contributions with quantum numbers n . These contributions cancel the infrared divergences of the color-singlet model, but introduce non-perturbative transitions between the colored intermediate state and the colorless final state that are encoded in long distance matrix elements (LDMEs) $\langle \mathcal{O}_Q^n \rangle$. The cross section formula from the color-singlet model is therefore

modified to

$$d\sigma_{\mathcal{Q}} = \sum_{i,j,n} \int dx_i dx_j f_i(x_i, \mu_f) f_j(x_j, \mu_f) d\hat{\sigma}_{i+j \rightarrow (q\bar{q})_{n+X}}(\mu_R, \mu_f, \mu_\Lambda) \langle \mathcal{O}_{\mathcal{Q}}^n \rangle, \quad (140)$$

where the long distance matrix elements replaced the Schrödinger wave function and a summation over quantum numbers n was added. These LDMEs have been determined in global fits by multiple groups including Ma et al. [226, 227] and Butenschoen et al. [228, 229] using a variety of different data sets. However, the cross sections predicted using the different sets of LDMEs still show qualitative disagreements and not all data can be described in this framework [230, 231].

3.1.3. Empirical determination of effective scattering-matrix elements

The ambiguity between the previously explained models naturally poses a problem for the use of quarkonium production data in our PDF fits, similar to the ambiguity from the fragmentation functions in the case of single inclusive hadrons discussed in the previous section. One way around this problem is the introduction of an empirical method that determines the optimal theoretical description of the process in terms of an effective scattering matrix element from data. Such an approach was first described in Ref. [200] and later used for a reweighting study by Kusina et al. [193, 232]. In the following, we will thoroughly examine the validity and accuracy of this approach, before using it to include both open heavy-flavor meson and quarkonium production data in a new global nPDF fit.

In this data-driven approach, the cross section for two nuclei A and B scattering and producing a quarkonium or open heavy flavored meson \mathcal{Q} is given by the convolution of the two initial state gluon PDFs $f_{1,g}(x_1, \mu)$, $f_{2,g}(x_2, \mu)$, with a fitted effective scattering matrix element $|\overline{\mathcal{A}_{gg \rightarrow \mathcal{Q}+X}}|^2$:

$$\sigma(AB \rightarrow \mathcal{Q} + X) = \int dx_1 dx_2 f_{1,g}(x_1, \mu) f_{2,g}(x_2, \mu) \frac{1}{2\hat{s}} |\overline{\mathcal{A}_{gg \rightarrow \mathcal{Q}+X}}|^2 d\text{PS}, \quad (141)$$

where dPS denotes the integration over the $AB \rightarrow \mathcal{Q}$ phase space. The effective scattering matrix element is parameterized with the Crystal Ball function

$$|\overline{\mathcal{A}_{gg \rightarrow \mathcal{Q}+X}}|^2 = \frac{\lambda^2 \kappa \hat{s}}{M_{\mathcal{Q}}^2} e^{a|y|} \begin{cases} e^{-\kappa \frac{p_T^2}{M_{\mathcal{Q}}^2}} & \text{if } p_T \leq \langle p_T \rangle, \\ e^{-\kappa \frac{\langle p_T \rangle^2}{M_{\mathcal{Q}}^2}} \left(1 + \frac{\kappa}{n} \frac{p_T^2 - \langle p_T \rangle^2}{M_{\mathcal{Q}}^2}\right)^{-n} & \text{if } p_T > \langle p_T \rangle, \end{cases} \quad (142)$$

3. Heavy quark production

where $\kappa, \lambda, \langle p_T \rangle, n$ and a are parameters⁸ determined empirically for each final state \mathcal{Q} . The fifth parameter, a , was not part of the original parameterization [233], but was added to yield a more accurate description of the rapidity dependence, as motivated by the leading order calculation presented in Ref. [234]. The parameters are determined separately for each final state \mathcal{Q} from $pp \rightarrow \mathcal{Q} + X$ data. For some final states, separate data sets are provided for prompt and non-prompt production, meaning production as a direct result of the hadronic collision or as a decay product of the initially produced particle, respectively. Such final states, *i.e.*, J/ψ and $\psi(2S)$, require two sets of parameters to describe the two different production mechanisms. The inclusive production cross section is generally calculated as the sum of prompt and non-prompt production⁹ and therefore creates a correlation between the two sets of parameters, which requires a combined fit of the two. The cross sections for $\Upsilon(1S)$ mesons are only available in terms of inclusive measurements, while all D^0 meson data is for prompt production.

For the particles that were included in the previous reweighting study, we keep the same default scales $\mu = \mu_0$ for the PDFs that were used there, while the $\psi(2S)$ scales are chosen analogously to the ones for J/ψ mesons. A summary of the scales is given in Tab. II.7. To estimate the impact of the scale choice on the final PDF fit, we repeat the entire procedure once with the scales halved and once with the scales doubled.

Table II.7.: Default scale choices μ_0 for the different particles.

μ_{0,D^0}^2	$\mu_{0,J/\psi}^2$	$\mu_{0,B \rightarrow J/\psi}^2$	$\mu_{0,\Upsilon(1S)}^2$	$\mu_{0,\psi(2S)}^2$	$\mu_{0,B \rightarrow \psi(2S)}^2$
$4M_D^2 + p_T^2$	$M_{J/\psi}^2 + p_T^2$	$4M_B^2 + \frac{M_B^2}{M_{J/\psi}^2} p_T^2$	$M_{\Upsilon(1S)}^2 + p_T^2$	$M_{\psi(2S)}^2 + p_T^2$	$4M_B^2 + \frac{M_B^2}{M_{\psi(2S)}^2} p_T^2$

Once the parameters are determined, we can use the same Hessian method to determine their uncertainties as we use in the nPDF analysis. Similar to the FF uncertainties in the previous chapter, we can then add these Crystal Ball uncertainties as an additional systematic uncertainty in the $pPb \rightarrow \mathcal{Q} + X$ data sets.

In principle, this approach can be used for any single inclusive production process in hadron-hadron collisions, if the necessary proton-proton data is available to determine the parameters with sufficient accuracy and the cross section is dominated by the gluon-gluon channel contributions. The lack of accurate baseline data is the reason why final states like

⁸Note that the parameter name “ $\langle p_T \rangle$ ” might be somewhat misleading. The parameterization was initially invented for the purpose of fitting the invariant mass distributions of detected particles, where the parameter carries the physical meaning of the particle’s average transverse momentum [233]. However, this interpretation is lost when the Crystal Ball function is used to parameterize the effective scattering matrix element. We decided to keep the name to avoid confusion when comparing with previous works.

⁹Note the potential ambiguity of the word “inclusive”. Previously, we have used it to denote processes, where the specified hadron is produced along with an arbitrary number of additional particles. From this point on, we will always use “inclusive” to denote the sum of prompt and non-prompt production, as all the data that we studied is “inclusive” in the first sense.

D^\pm mesons and higher Υ excitations are excluded in the present analysis.

3.2. Proton-proton baseline

The first step of the analysis is the determination of the Crystal Ball parameters from the proton-proton collision data. The data sets used for this fit are summarized in Tab. II.8. The majority of the data points come from J/ψ production, followed by $\Upsilon(1S)$, $\psi(2S)$ and finally D^0 mesons. However, not all of the data points can be used in the fit, since kinematic cuts need to be imposed to ensure the validity of the factorization theorem and to exclude data that cannot be described by the data-driven approach.

3.2.1. Kinematic cuts and excluded data

The same kinematic cuts on the rapidity y_{cms} and on the transverse momentum p_T of the produced particles are imposed on all final states. They are chosen such that reasonable χ^2 values are obtained not just for the average of the data set, but also by looking at the χ^2 of individual points. Performing the fits while cutting all data with $p_T < 3 \text{ GeV}$ and outside of the rapidity range $-4 < y_{\text{cms}} < 4$ yields the parameters and corresponding χ^2/N_{dof} values shown in Tab. II.9. Relaxing either of these two cuts would still yield a good χ^2/N_{dof} value for the original data points, but the newly introduced data points contribute significantly more with χ^2 values larger than 5 being common among those points.

In addition to these general cuts, we also exclude some points from specific data sets. First, we exclude the lowest p_T bins of the 2011 CMS J/ψ data sets [239], because the accompanying paper mentions that potentially large detector acceptance effects in this region are not included in the provided uncertainties. Furthermore, the $\Upsilon(1S)$ production data set from the $\sqrt{s} = 7 \text{ TeV}$ LHCb measurement [250] requires special treatment. For this data set, we assign an additional normalization parameter to the lowest rapidity bin ($2.0 < y_{\text{cms}} < 2.5$), which contains 20 p_T bins (17 after cuts). The data in this rapidity bin is in qualitative agreement with all remaining data sets, but the normalization is off by 25%. Without the additional normalization parameter, the χ^2/N_{dof} of the entire $\Upsilon(1S)$ fit is, therefore, increased significantly from 0.92 to 1.6. Other models, based on the color-octet mechanism, observe the same tension with the normalization of this particular rapidity bin [250]. The corresponding paper also includes data for the same measurements performed at $\sqrt{s} = 8 \text{ TeV}$, where this rapidity bin is well described in our fit. Additionally, earlier measurements by LHCb [247] at the same rapidity and $\sqrt{s} = 7 \text{ TeV}$ show no tensions and neither does a similar measurement by CMS [248] with $2.0 < y_{\text{cms}} < 2.4$ and $\sqrt{s} = 7 \text{ TeV}$.

Some newer D^0 meson data sets [252–254] are not part of the fit, but predictions with our fitted parameters yield a good description with a χ^2/N_{dof} value of 0.67. These data sets can be included in future analyses to reduce the uncertainty of the CB parameters for D^0 mesons and thereby strengthen the impact of the proton-lead data on the nuclear

3. Heavy quark production

Table II.8.: Overview of the available $pp \rightarrow Q+X$ production data sets and their number of data points. Note that data sets with multiple different \sqrt{s} values are split for technical reasons.

Experiment	Year	Ref.	ID	Type	Points after / before cuts
ALICE	2012	[235]	3008	Prompt D^0	7 / 9
LHCb	2013	[236]	3014	Prompt D^0	22 / 38
ALICE	2016	[237]	3021	Prompt D^0	7 / 10
ATLAS	2011	[238]	3003	Prompt J/ψ	63 / 64
CMS	2011	[239]	3001	Prompt J/ψ	39 / 44
LHCb	2011	[240]	3002	Prompt J/ψ	50 / 66
CMS	2017	[241]	3020	Prompt J/ψ	50 / 52
ATLAS	2018	[242]	3017	Prompt J/ψ	33 / 33
CMS	2010	[223]	3007	Non-prompt J/ψ	11 / 14
ATLAS	2011	[238]	3006	Non-prompt J/ψ	63 / 64
CMS	2011	[239]	3005	Non-prompt J/ψ	39 / 44
LHCb	2011	[240]	3004	Non-prompt J/ψ	50 / 66
CMS	2017	[241]	3019	Non-prompt J/ψ	50 / 52
ATLAS	2018	[242]	3018	Non-prompt J/ψ	33 / 33
ALICE	2015	[243]	3022	Inclusive J/ψ	10 / 13
ALICE	2017	[244]	3023	Inclusive J/ψ	8 / 11
ALICE	2017	[244]	3024	Inclusive J/ψ	14 / 18
ALICE	2019	[245]	3016	Inclusive J/ψ	4 / 7
ATLAS	2012	[246]	3012	Inclusive $\Upsilon(1S)$	88 / 100
LHCb	2012	[247]	3025	Inclusive $\Upsilon(1S)$	55 / 75
CMS	2013	[248]	3013	Inclusive $\Upsilon(1S)$	30 / 42
ALICE	2014	[249]	3009	Inclusive $\Upsilon(1S)$	3 / 5
LHCb	2015	[250]	3011	Inclusive $\Upsilon(1S)$	89 / 109
LHCb	2015	[250]	3015	Inclusive $\Upsilon(1S)$	89 / 109
ALICE	2015	[243]	3031	Inclusive $\Upsilon(1S)$	3 / 5
ATLAS	2017	[242]	3031	Inclusive $\Upsilon(1S)$	18 / 24
ATLAS	2017	[242]	3026	Prompt $\psi(2S)$	15 / 15
ATLAS	2017	[242]	3027	Non-prompt $\psi(2S)$	15 / 15
ALICE	2015	[243]	3028	Inclusive $\psi(2S)$	6 / 9
CMS	2018	[251]	3029	Inclusive $\psi(2S)$	9 / 10
ALICE	2017	[244]	3030	Inclusive $\psi(2S)$	9 / 12

PDFs.

Finally, we limit the scope of the study to LHC data only. The available RHIC data may not be compatible with the same fit parameters due to the very different collision energies and is unlikely to provide additional constraints due to their large uncertainties.

3.2.2. Baseline fit

The parameters, along with their uncertainties and corresponding χ^2/N_{dof} values for the fits, are shown in Tab. II.9. The parameters for J/ψ and $\psi(2S)$ are obtained in a combined fit with prompt, non-prompt and inclusive data, giving them two sets of parameters each, while $\Upsilon(1S)$ and D^0 require only one set. Overall, the agreement between data and fitted theory is very good with χ^2/N_{dof} values slightly below one for J/ψ , $\psi(2S)$ and $\Upsilon(1S)$. The D^0 data produces a very low χ^2/N_{dof} of 0.25, which might suggest that the experimental uncertainties are overestimated, but due to the low number of data points for D^0 production, no strong conclusions can be drawn.

The uncertainties of the fit are determined with the same Hessian method that is commonly used in PDF fits. Therefore, the uncertainties on the individual parameters do not represent the full picture as they lack the information about correlations. Nevertheless, they clearly show the smaller uncertainties for J/ψ and $\Upsilon(1S)$ mesons, where data is more abundant. The newly added rapidity parameter a is compatible with zero, *i.e.*, no modification, for most particles except for non-prompt J/ψ and especially $\Upsilon(1S)$.

A comparison of the fits with the data is shown in Figs. IV.5 - IV.10. The fits describe the data well across the entire kinematic region that is included in the fits. The uncertainties for J/ψ and $\Upsilon(1S)$ mesons are not visible on the large logarithmic scale, while the D^0 and $\psi(2S)$ uncertainties are visible, but still significantly smaller than those of the data.

Table II.9.: Crystal Ball parameters and χ^2/N_{dof} values for the Crystal Ball function for the different processes.

	D^0	J/ψ	$B \rightarrow J/\psi$	$\Upsilon(1S)$	$\psi(2S)$	$B \rightarrow \psi(2S)$
κ	0.33 ± 0.13	0.48 ± 0.01	0.15 ± 0.01	0.95 ± 0.09	0.22 ± 0.11	0.45 ± 0.04
λ	1.83 ± 0.16	0.30 ± 0.00	0.12 ± 0.00	0.06 ± 0.00	0.07 ± 0.01	0.14 ± 0.01
$\langle p_T \rangle$	2.40 ± 1.53	5.29 ± 0.18	-7.65 ± 0.33	8.64 ± 1.74	8.99 ± 1.43	7.81 ± 0.84
n	2.00 ± 0.23	2.17 ± 0.02	1.56 ± 0.03	1.93 ± 0.18	1.07 ± 0.24	1.65 ± 0.20
a	-0.03 ± 0.03	0.03 ± 0.01	-0.08 ± 0.01	0.22 ± 0.03	-0.11 ± 0.10	0.06 ± 0.04
N_{dof}	34	501		375	55	
$\frac{\chi^2}{N_{\text{dof}}}$	0.25	0.88		0.92	0.77	

3.2.3. Alternative parameterizations of the effective scattering matrix element

Along with the Crystal Ball parameterization, a variety of other parameterizations for the effective scattering matrix element have been tested in an attempt to describe the $p_T < 3 \text{ GeV}$ data that is excluded by the cuts. The most obvious extension of the Crystal Ball function is called CB2 [255] and adds another polynomial tail on the low p_T side of the Gaussian:

$$\overline{|\mathcal{A}_{gg \rightarrow Q+X}|^2} = \frac{\lambda^2 \kappa \hat{s}}{M_Q^2} e^{a|y|} \begin{cases} e^{-\kappa \frac{\langle p_{T,1} \rangle^2}{M_Q^2}} \left(1 + \frac{\kappa}{n_1} \frac{p_T^2 - \langle p_{T,1} \rangle^2}{M_Q^2} \right)^{-n_1} & \text{if } p_T < \langle p_{T,1} \rangle, \\ e^{-\kappa \frac{\langle p_{T,2} \rangle^2}{M_Q^2}} \left(1 + \frac{\kappa}{n_2} \frac{p_T^2 - \langle p_{T,2} \rangle^2}{M_Q^2} \right)^{-n_2} & \text{if } p_T > \langle p_{T,2} \rangle, \\ e^{-\kappa \frac{p_T^2}{M_Q^2}} & \text{else,} \end{cases} \quad (143)$$

which gives a total of seven free parameters $\kappa, \lambda, \langle p_{T,1} \rangle, \langle p_{T,2} \rangle, n_1, n_2, a$ for each final state.

Another parameterization is inspired by the leading-order low- p_T behavior derived in Ref. [234], which predicts a second order polynomial in $\frac{p_T^2}{m_Q^2}$:

$$\overline{|\mathcal{A}_{gg \rightarrow Q+X}|^2} = \frac{\lambda^2 \kappa \hat{s}}{M_Q^2} e^{a|y|} \begin{cases} e^{-\kappa \frac{\langle p_{T,1} \rangle^2}{M_Q^2}} \left(n_1 + n_2 \frac{p_T^2}{M_Q^2} + n_3 \frac{p_T^4}{M_Q^4} \right) & \text{if } p_T < \langle p_{T,1} \rangle, \\ e^{-\kappa \frac{\langle p_{T,2} \rangle^2}{M_Q^2}} \left(1 + \frac{\kappa}{n} \frac{p_T^2 - \langle p_{T,2} \rangle^2}{M_Q^2} \right)^{-n} & \text{if } p_T > \langle p_{T,2} \rangle, \\ e^{-\kappa \frac{p_T^2}{M_Q^2}} & \text{else,} \end{cases} \quad (144)$$

with a total of nine parameters $\kappa, \lambda, \langle p_{T,1} \rangle, \langle p_{T,2} \rangle, n, n_1, n_2, n_3, a$ for each final state.

J/ψ mesons provide the only data set that contains enough data points in the low- p_T region to put reasonable constraints on additional parameters. Fitting the low- p_T part of the two parameterizations described above to the 44 J/ψ data points with $p_T < 3 \text{ GeV}$ yields a χ^2/N_{dof} value for 5.99 for the CB2 parameterization and 2.15 for the polynomial parameterization. Looking at the χ^2 values of individual data points, the largest contributions come the data taken at very forward or backward rapidity.

Introducing an additional fitted exponential rapidity dependence for this p_T region, the χ^2/N_{dof} can be reduced to 4.15 and 1.41 for the CB2 and polynomial parameterizations respectively. However, even if we would consider this an acceptable fit, this number of parameters means that the data points are certainly overfitted, such that the predictive power of the model is lost. Therefore the parameterization given in Eq.(142) remains the best compromise between a good description of the available data with a reasonable number

of free parameters.

3.2.4. Comparison with prompt D^0 production in the GMVFNS

The predictions for D^0 meson production from the data-driven method can be compared to predictions from perturbative QCD in the GMVFNS. The calculations in the GMVFNS are carried out using an implementation of heavy quark production in NLO QCD by Kniehl et al. [256]. In these calculations, we use the nCTEQ15 proton PDF, a charm quark mass of $m_c = 1.3$ GeV, the renormalization and the initial/final factorization scales $\mu_r = \mu_i = \mu_f = \sqrt{p_T^2 + 4m_c^2}$ and KKKS08 fragmentation functions [257]. We estimate the uncertainties by varying the scales individually by factors of two or one-half, but with no two scales separated by a factor four. The resulting $pp \rightarrow D^0 + X$ cross sections are compared to the ones from the Crystal Ball fit, and the corresponding data in Fig. II.22. The uncertainties of the GMVFNS predictions are mostly similar in size to the data uncertainty, but in the low- p_T region they become somewhat larger. The data points mostly lie within these uncertainties, except for the ones at the highest p_T . In general, the GMVFNS predictions slightly overshoot the data, which can be explained by the largely unconstrained gluon fragmentation function, which contributes almost half of the cross section. The gluon FF is not well constrained due to the fact that the fit only includes data from ALEPH, Belle, CLEO and OPAL experiments, which are all based on electron-positron collisions.

The uncertainties of the Crystal Ball fit are significantly smaller than those of the data and GMVFNS at low p_T , but grow to a similar size as the GMVFNS uncertainties for larger p_T values. The central values of the Crystal Ball fit are in close agreement with those of the data, as expected considering the low χ^2/N_{dof} value for D^0 mesons in Tab. II.9. Overall, the data-driven approach is in good agreement with the perturbative QCD calculation. The latter could see significant improvements from a modern FF analysis including data from hadronic collisions, which would put better constraints on the gluon FF.

3.2.5. Comparison with J/ψ production in NRQCD

We can perform a similar comparison between the Crystal Ball fit for prompt J/ψ production and the results obtained in the framework of non-relativistic QCD (NRQCD). The NRQCD results for this comparison have been provided by Butenschoen et al. [228, 229]

3. Heavy quark production

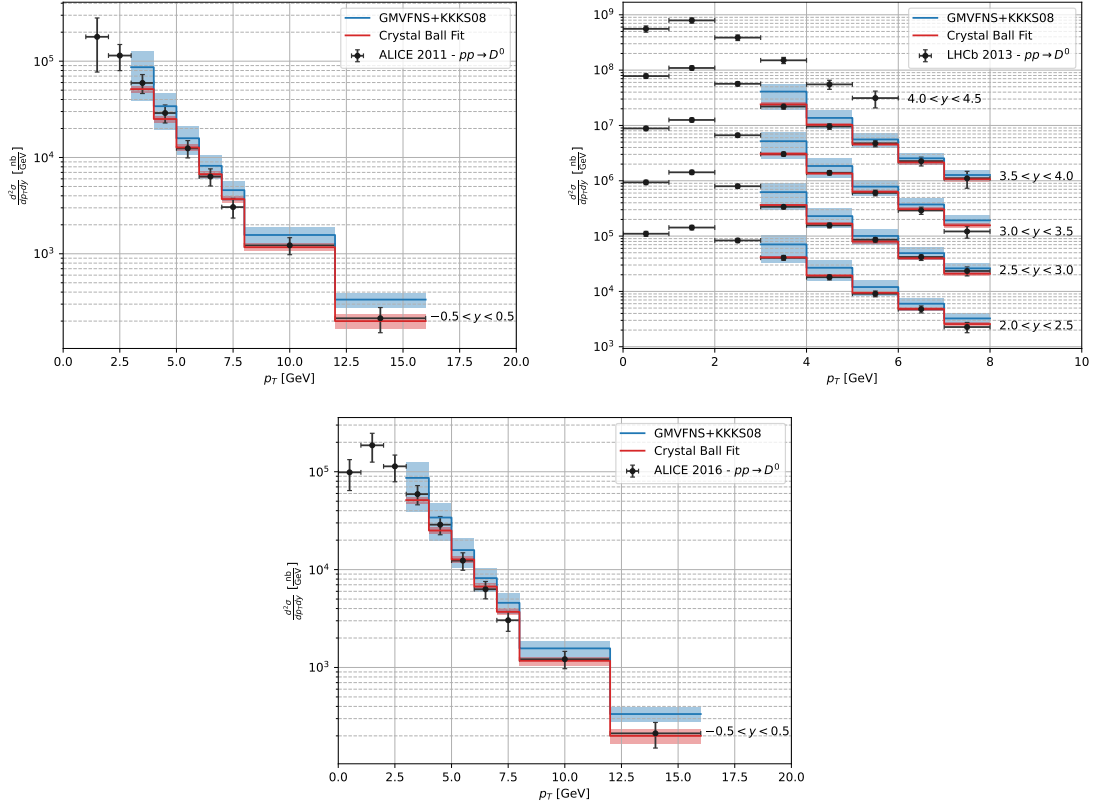


Figure II.22.: Comparison between prompt D^0 production as predicted in the GMVFNS (blue) and with the Crystal Ball approach (red). The uncertainties of the GMVFNS predictions come from an individual variation of the scales by factors of 2 or one half, such that there is never a factor 4 between two scales. Different rapidity bins are shifted on the y-axis by powers of ten for visual clarity.

using the following long-distance matrix elements (LDMEs):

$$\langle \mathcal{O}^{J/\psi}(^3S_1^{[1]}) \rangle = 1.32 \text{ GeV}^3, \quad (145)$$

$$\langle \mathcal{O}^{J/\psi}(^1S_0^{[8]}) \rangle = 0.0497 \text{ GeV}^3, \quad (146)$$

$$\langle \mathcal{O}^{J/\psi}(^3S_1^{[8]}) \rangle = 0.00224 \text{ GeV}^3, \quad (147)$$

$$\langle \mathcal{O}^{J/\psi}(^3P_0^{[8]}) \rangle = -0.0161 \text{ GeV}^5. \quad (148)$$

The LDMEs are obtained in a fit to the data measured by CDF at Tevatron and H1 at HERA. The calculation is performed at NLO in QCD with full relativistic corrections and uses the nCTEQ15 proton PDFs. The uncertainties are obtained by a simultaneous variation of all scales by a factor two around their default values $\mu_{r,0} = \mu_{f,0} = \sqrt{p_T^2 + 4m_c^2}$ and $m_{\text{NRQCD},0} = m_c$ with $m_c = 1.5 \text{ GeV}$.

Figure II.23 shows the comparison of the cross sections obtained by the two different methods for kinematics representative for the LHCb [258], ALICE [259] and ATLAS [260] experiments. Across the entire kinematic region, we see very good agreement between the Crystal Ball fit and the NRQCD prediction. At low p_T values, the uncertainties of the two methods are of similar size, but towards higher p_T the relative uncertainty of the NRQCD prediction rises, while those of the Crystal Ball fit are mostly invariant.

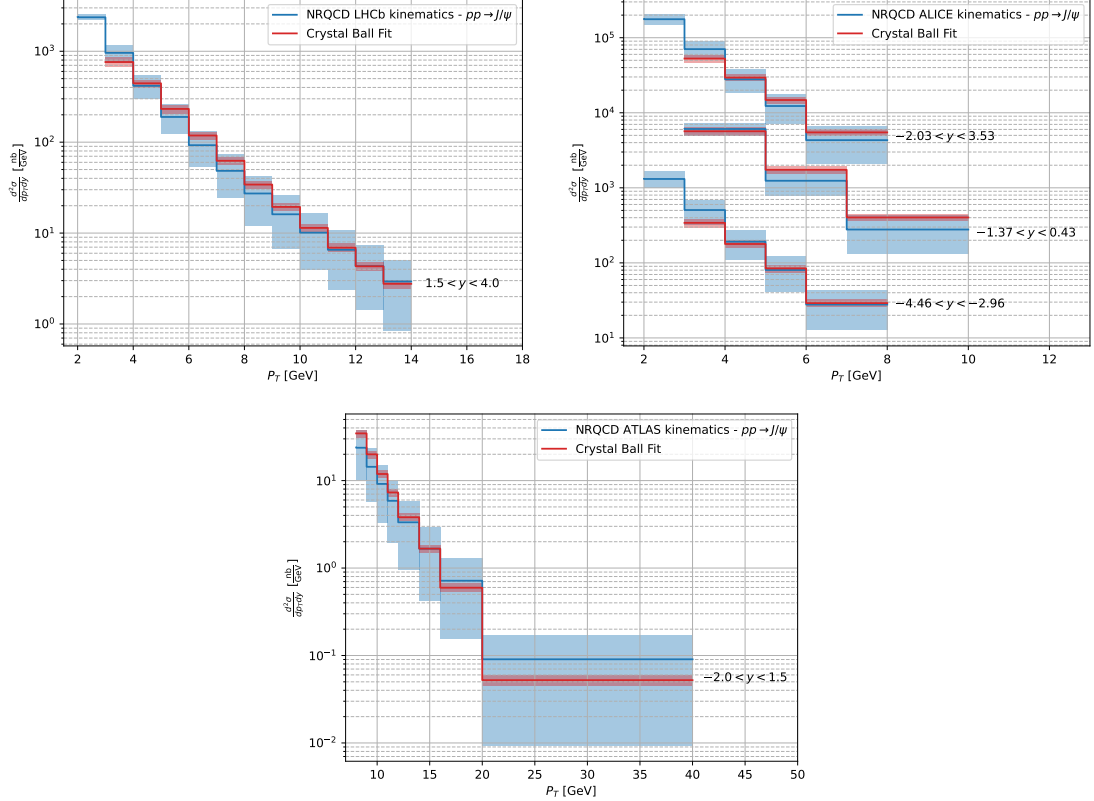


Figure II.23.: Comparison between prompt J/ψ production in pp collisions from NRQCD and with the Crystal Ball approach. The uncertainties of the NRQCD predictions come from a combined scale variation $1/2 < \mu_r/\mu_{r,0} = \mu_f/\mu_{f,0} = \mu_{\text{NRQCD}}/\mu_{\text{NRQCD},0} < 2$ around the base scale $\mu_{r,0} = \mu_{f,0} = \sqrt{p_T^2 + 4m_c^2}$ and $m_{\text{NRQCD},0} = m_c$. Different rapidity bins are shifted on the y-axis by powers of ten for visual clarity.

3.3. Fits with heavy quark data

Using the Crystal Ball parameters that have been obtained and validated with comparisons to pQCD results in the previous section, we can now perform a new global nPDF fit including a vast body of heavy quark data. In most regards, the new fits use the same framework as the preceding nCTEQ15WZ+SIH fit, including the same open parameters

and data sets (see Tab. II.2 for the SIH data sets and Tabs. IV.4-IV.7 for DIS, DY and WZ). Additionally, all settings, like the imposed cuts, pion and kaon fragmentation functions and scales for these data sets, remain unchanged. We do not include the changes and additions made in the nCTEQ15HIX and nCTEQ15 ν analyses, as these do not have a direct connection to the low- x gluon PDF that is the focus of this study. The only change in the general framework is the treatment of normalization parameters, which was previously done by introducing additional fit parameters for each normalization uncertainty. In this study, we use the analytical treatment outlined in Sec. I.2.4.3, which drastically reduces the number of fit parameters due to the large number of new data sets with normalization uncertainty.

3.3.1. Data selection

The new fit extends the nCTEQ15WZ+SIH analysis by adding the heavy-quark data sets listed in Tab. II.10, imposing the same cuts on these new data sets, as we did for the proton-proton baseline. Additionally, we need to exclude the D^0 -production data points with $p_T > 15$ GeV, since there is no baseline data to constrain the Crystal Ball parameters in this region. Finally, we remove two additional data points from the 2018 LHCb measurement of inclusive $\Upsilon(1S)$ production, which are otherwise exceptional outliers with χ^2 values of 26 and 66, respectively. The two data points are taken at the largest p_T values and most forward/backward rapidity of the experiment's kinematic range.

This raises the total number of data points to 1484 (548 new, 936 old). Just like the fragmentation function uncertainties in the previous analysis, we can take the theoretical uncertainties from the Crystal Ball fit into account by adding them in quadrature to the systematic uncertainties of the data sets.

Finally, we will repeat the entire procedure, starting with the proton-proton baseline fit, two more times – once with the scale μ of the heavy quark production doubled and once with the scale halved. This serves to demonstrate that we do not introduce any strong bias into the PDF with our choice of the default scale μ_0 .

3.3.2. The nCTEQ15HQ nuclear PDFs

The new PDFs resulting from this fit, named nCTEQ15HQ¹⁰, are presented in Figs. II.24 and II.25, along with the PDFs obtained in previous nCTEQ analyses. The former figure shows the effective proton PDFs $f_i^{p/Pb}(x, Q^2)$ for lead at $Q = 2$ GeV, while the latter shows the nuclear modification ratios R_i for lead at $Q = 2$ GeV for the flavors $i = u, d, \bar{u}, \bar{d}, s + \bar{s}, g$. A full list of the PDF parameters determined in the fit can be found in Tab. IV.3. The up and down quark PDFs retain very similar central values to the nCTEQ15WZ+SIH fit, with

¹⁰We break the naming convention used for nCTEQ15WZ+SIH and shorten the new name to nCTEQ15HQ. This does *not* imply the removal of the WZ and SIH data.

Table II.10.: Overview of the available $pPb \rightarrow Q + X$ production data sets and their number of data points. Note that data sets with multiple different \sqrt{s} values are split for technical reasons.

Group	Year	Ref.	ID	Type	Points after / before cuts
ALICE	2014	[261]	3101	Prompt D^0	8 / 10
ALICE	2016	[237]	3123	Prompt D^0	8 / 11
LHCb	2017	[262]	3102	Prompt D^0	53 / 92
ALICE	2019	[263]	3122	Prompt D^0	13 / 21
LHCb	2013	[264]	3108	Prompt J/ψ	25 / 40
ATLAS	2015	[260]	3118	Prompt J/ψ	10 / 10
LHCb	2017	[258]	3105	Prompt J/ψ	88 / 140
CMS	2017	[241]	3120	Prompt J/ψ	51 / 53
ATLAS	2018	[242]	3117	Prompt J/ψ	8 / 8
LHCb	2013	[264]	3107	Non-prompt J/ψ	25 / 40
ATLAS	2015	[260]	3119	Non-prompt J/ψ	10 / 10
LHCb	2017	[258]	3106	Non-prompt J/ψ	88 / 140
CMS	2017	[241]	3121	Non-prompt J/ψ	51 / 53
ATLAS	2018	[242]	3116	Non-prompt J/ψ	8 / 8
ALICE	2013	[265]	3103	Inclusive J/ψ	0 / 12
ALICE	2015	[266]	3104	Inclusive J/ψ	10 / 25
ALICE	2018	[259]	3112	Inclusive J/ψ	9 / 24
ALICE	2014	[267]	3110	Inclusive $\Upsilon(1S)$	0 / 4
LHCb	2014	[268]	3111	Inclusive $\Upsilon(1S)$	0 / 2
ATLAS	2018	[242]	3109	Inclusive $\Upsilon(1S)$	6 / 8
LHCb	2018	[269]	3113	Inclusive $\Upsilon(1S)$	36 / 66
ALICE	2019	[270]	3114	Inclusive $\Upsilon(1S)$	3 / 10
ATLAS	2017	[242]	3124	Prompt $\psi(2S)$	8 / 8
CMS	2018	[251]	3115	Prompt $\psi(2S)$	17 / 17
ATLAS	2017	[242]	3125	Non-prompt $\psi(2S)$	8 / 8
ALICE	2014	[271]	3127	Inclusive $\psi(2S)$	2 / 8
ALICE	2020	[272]	3126	Inclusive $\psi(2S)$	3 / 10

a minor suppression at x values below 10^{-4} . The uncertainties of these quarks also remain similar to the previous fit for x values above 10^{-2} , but do not rise as quickly when going towards lower values, which leads them to be reduced by roughly a third at $x = 10^{-5}$. The strange quark PDF takes a very similar shape to the one of the nCTEQ15 fit¹¹, but without the strong rise towards lower x values seen especially in nCTEQ15WZ+SIH. However, the relative uncertainties of the strange quark remain very large and in particular in the low x region they are larger than 100%, which prohibits any strong conclusions on the true behaviour of this flavor. Naturally, the most significant changes are observed in the gluon PDF. The central value remains close to the nCTEQ15WZ+SIH value for $x > 0.02$, but sees a significant suppression below that, where it is very close to the nCTEQ15WZ values. The uncertainties of the gluon are reduced by a remarkable amount for $x < 0.1$ and particularly at $x < 10^{-4}$, where the uncertainties of the previous two fits start to diverge, the new data still imposes strong constraints.

The nuclear PDF ratios in Fig. II.25 show some further details that are not directly visible in the full PDF plot. For once, the peak of the up-quark PDF around $x = 0.1$ is slightly suppressed in the new fit and the peak of the down quark is shifted to a slightly higher x value. While the central value of the strange quark PDF is still very similar to those of the previous two analyses at $x > 0.1$, it shows a more pronounced peak around $x = 0.03$. At lower x , the strange quark in the new fit is suppressed compared to the proton case, as opposed to the enhancement seen in the previous two fits. Due to the large uncertainties in all three cases the nuclear modification factor of the strange quark is still compatible with unity. An exception to this is the high- x region, but the uncertainties there are likely underestimated due to the low number of free strange quark parameters. The nuclear modification factor of the gluon PDF is drawn closer to unity by the new data for $x > 10^{-3}$. Below this, the new gluon modification levels off around $R_g = 0.75$, while the nCTEQ15WZ+SIH gluon begins rising again and eventually becomes enhanced compared to the proton. While the gluon modification of nCTEQ15HQ is still mostly compatible with unity for $x > 10^{-3}$, a suppression of 20 to 30% in the low- x region becomes very clear due to the significantly reduced uncertainties.

Next, we investigate the impact of the scale choice for the heavy quark production on the resulting nPDFs. To do this, we repeat the full procedure, including the fit of the proton-proton baseline with two alternative scale choices, obtained by multiplying the central scale choice with a factor two or one half. Figure II.26 presents a comparison between the PDFs produced with these different scale choices as a ratio over the PDF produced with the central scale. At low x , the reduced scale produces a slightly enhanced gluon PDF, while the larger scale causes a suppression. The three PDFs stay close and cross each other at $x = 0.02$ and $x = 0.15$, before drifting apart in the high- x region. Overall, the central

¹¹It is important to remember that nCTEQ15 did not have any open strange quark parameters, which explains the small uncertainties coming solely from the other flavors' DGLAP contributions.

values of the gluon PDFs see very minor relative shifts, which only rise above 5% in the very high- x region, where the absolute value of the PDFs tend to zero. These shifts are well contained within the uncertainty of the central fit and the uncertainties of the shifted fits are very similar in size to those of the central fit. Contrary to the gluon, the up and down quark PDFs resulting from the shifted scales are not mirrored around the central one. Both alternative scales produce an enhanced up and suppressed down flavor in the high- x region. However, as in the gluon case, these shifts are well contained within the uncertainties. While the enlarged scale causes only a minor increase in uncertainties, the reduced scale leads to a significant increase in the quark uncertainties at low x . The strange quark PDF sees more significant changes than the other flavors. These relative shifts of up to 20% are not surprising, however, due to the large relative uncertainty of this flavor, which is greater than 40% for all values of x .

Finally, we look at the impact of the theoretical uncertainty on the nPDF fit. To do this, the three previous fits are repeated once more, without the uncertainty of the Crystal Ball fit added to the data. The PDFs resulting from those fits are compared to the corresponding fits with added uncertainties in Fig. II.27¹². The most notable changes occur in the up and down quark PDFs at low x , where a slightly stronger suppression is observed and in the gluon around $x = 0.03$, where it is enhanced by up to 5%. The strange quark sees a suppression by up to 20% at all x values if the theoretical uncertainties are removed. In general, the shifts caused by this are of very similar size as the ones caused by the scale variation and therefore also lie well within the PDF uncertainty. These shifts represent the maximal increase in impact that the heavy quark data could have, if the accuracy of the proton-proton baseline is improved, while keeping the same proton-lead data in the PDF fit.

3.3.3. Fit quality

Given the nCTEQ15HQ PDFs discussed in the previous section, we now validate the fit quality against the data sets included in the analysis.

Comparison with data

In the first comparison we look at the predictions for the newly added heavy quark data using the nCTEQ15HQ PDFs. Figures IV.11 - IV.15 show all included proton-lead data sets with the predictions from the new fit, including its uncertainty. Across the entire kinematic region and for all produced final state particles, we see very close agreement with the data and no significant outliers are observed among the data points. The PDF uncertainties are too small to be visible in the vast majority of plots due to the logarithmic y-axis, but where they are visible, they are clearly significantly smaller than the data

¹²Note that the y-axis in Fig. II.27 is zoomed in by a factor two, compared to Fig. II.26.

3. Heavy quark production

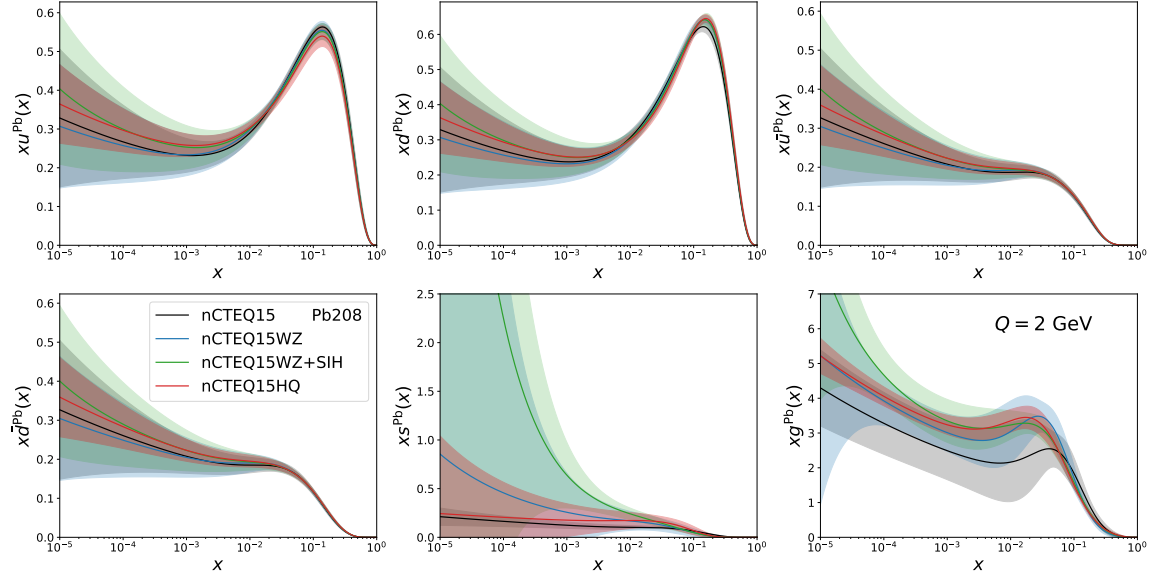


Figure II.24.: Comparison between lead PDFs from different nCTEQ15 versions. The baseline nCTEQ15 fit is shown in black, nCTEQ15WZ in blue, nCTEQ15WZSIH in green, and the new nCTEQ15HQ fit in red.

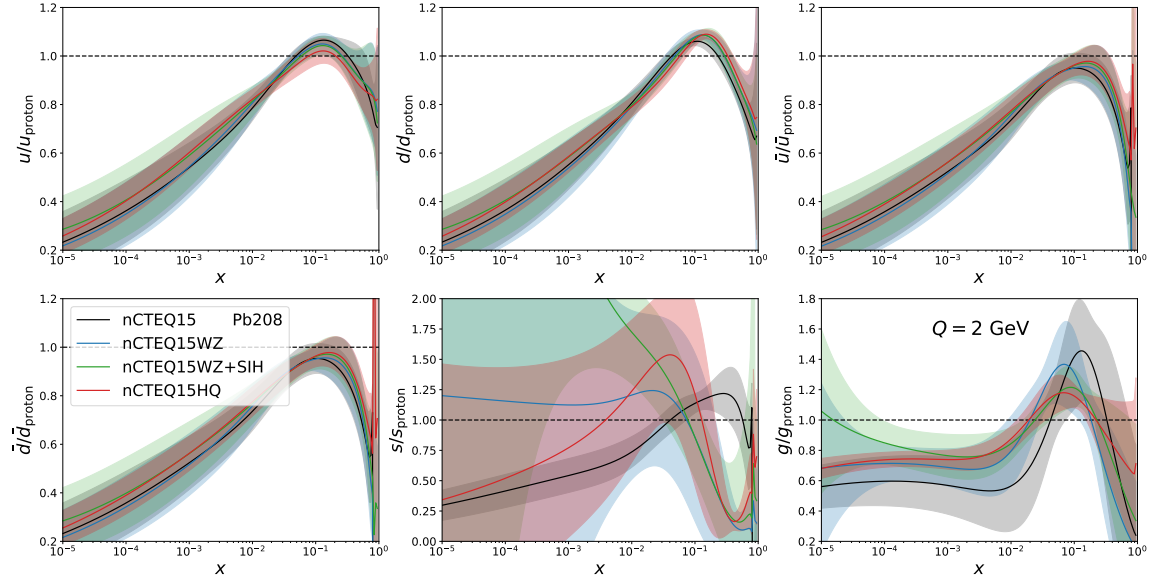


Figure II.25.: Comparison between the ratio of lead over proton PDFs from different nCTEQ15 versions. The baseline nCTEQ15 fit is shown in black, nCTEQ15WZ in blue, nCTEQ15WZSIH in green, and the new nCTEQ15HQ fit in red.

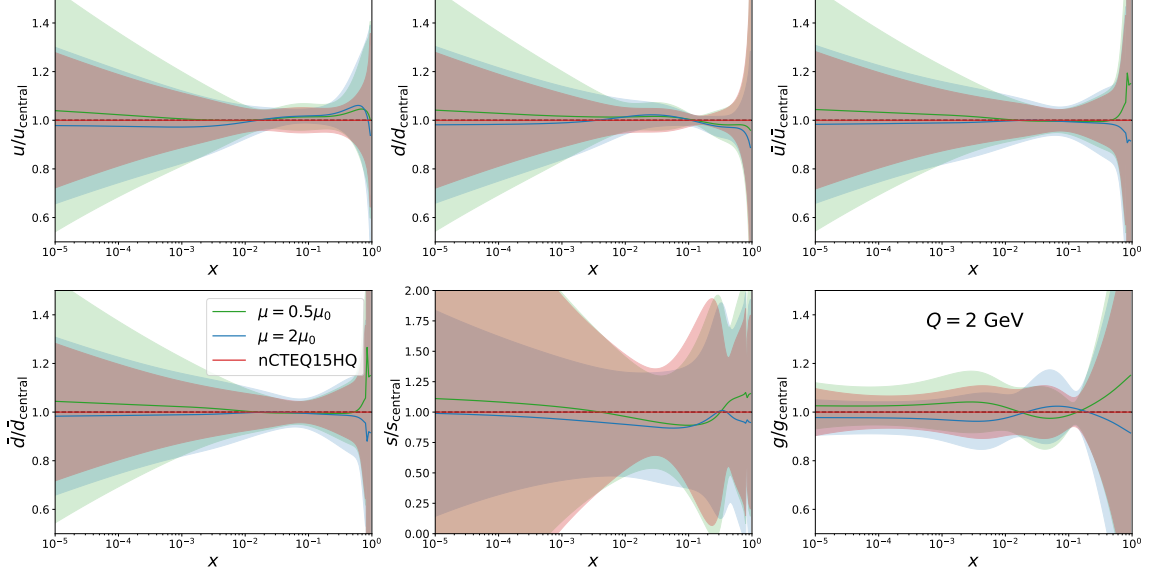


Figure II.26.: Comparisons of fits, where the scale for heavy quark production is varied by a factor two or one half around the central values μ_0 . The PDFs are displayed as a ratio over the central PDF.

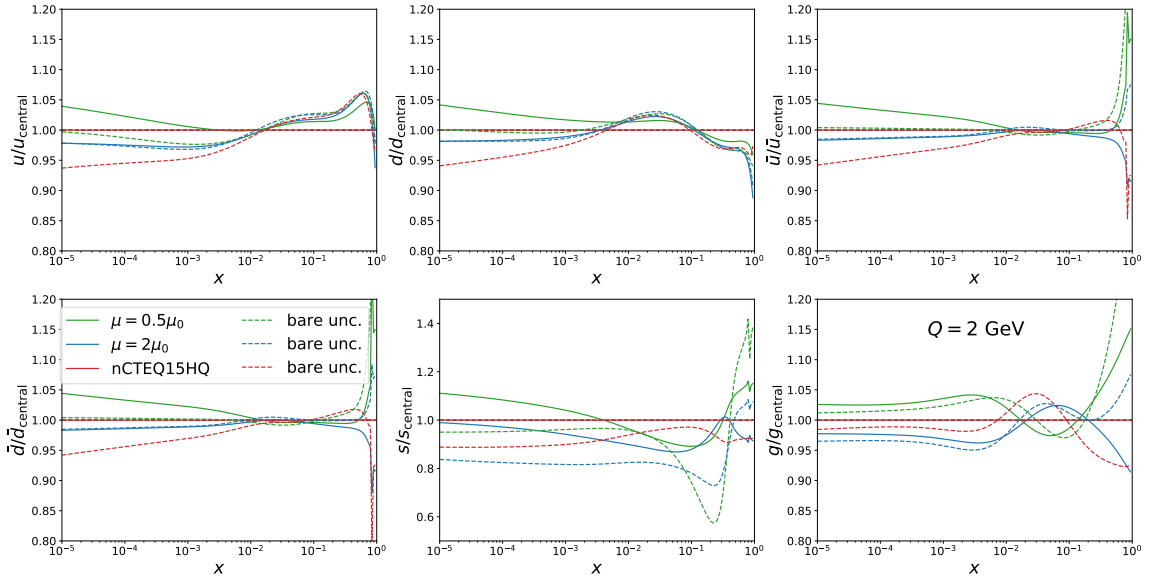


Figure II.27.: Comparison of PDF fits with and without Crystal Ball uncertainties included in the data. The solid and dashed lines show the fit to data with theory uncertainties and pure experimental uncertainties respectively.

uncertainty. It is important to note, however, that these uncertainties do not represent the full uncertainty of the prediction, as the Crystal Ball uncertainty would have to be added, which is significantly larger than the PDF uncertainties as we showed in the corresponding Figs. IV.4 - IV.10 for the proton-proton baseline.

χ^2 values of individual data sets

We compare the distribution of the χ^2/N_{dof} values among the data sets included in the fits for the previous nCTEQ15WZ+SIH fit and the new nCTEQ15 fit in Fig. II.28. The new fit, shown in the lower panel, retains very similar χ^2/N_{dof} values for the data sets shared between both analyses. The only major outlier among those is still the data set with ID 5108 (EMC-1988, F_2^{Sn}/F_2^D) and the outlier among the the WZ production data sets, ID 6215 (ATLAS Run I, Z production), is now better described than before. This indicates no incompatibilities between the new heavy quark data sets and the established data. The new data sets themselves mostly yield χ^2/N_{dof} values below one, with only five data sets producing a χ^2/N_{dof} greater than one. The two largest outliers are the $\psi(2S)$ production data sets with IDs 3125 and 3127 with χ^2/N_{dof} values of 1.29 and 2.05, respectively. The latter data set consists of just two data points, so these may be just natural statistical fluctuations, but it might also hint at problems due to the small number of $\psi(2S)$ data points in the baseline fit.

A more quantitative look at the χ^2/N_{dof} values is given in Tab. II.11, which shows a comparison of the χ^2/N_{dof} values of the new fit and the preceding three nCTEQ15 generations. The χ^2/N_{dof} values are split by process and for heavy quark production they are further separated into the different final state particles. The data for D^0 , $\Upsilon(1S)$ and $\psi(2S)$ production is already well described even by the original nCTEQ15 fit, with D^0 mesons being the only one among those that see further improvement in the nCTEQ15HQ fit. However, the J/ψ meson production, which contains by far the largest number of data points is not well described in the nCTEQ15 fit, with a χ^2/N_{dof} of 2.50. The nCTEQ15WZ fit already improves this significantly, such that the total χ^2/N_{dof} of heavy quark production is reduced to 0.92. The nCTEQ15WZ+SIH fit leads to further improvement of the J/ψ production data, but the other heavy quark data sets become slightly worse described such that the overall χ^2/N_{dof} is slightly increased when compared to nCTEQ15WZ. Finally, the nCTEQ15HQ improves upon all heavy quark data sets compared to the previous fit and gives the best description of the J/ψ data. The improved description of the heavy quark data comes at the cost of a slightly worse description of the DIS and WZ production data, but the total χ^2 is the lowest among all fits. The heavy quark and single inclusive hadron production data agree well with each other, but both show slight tensions with the WZ production.

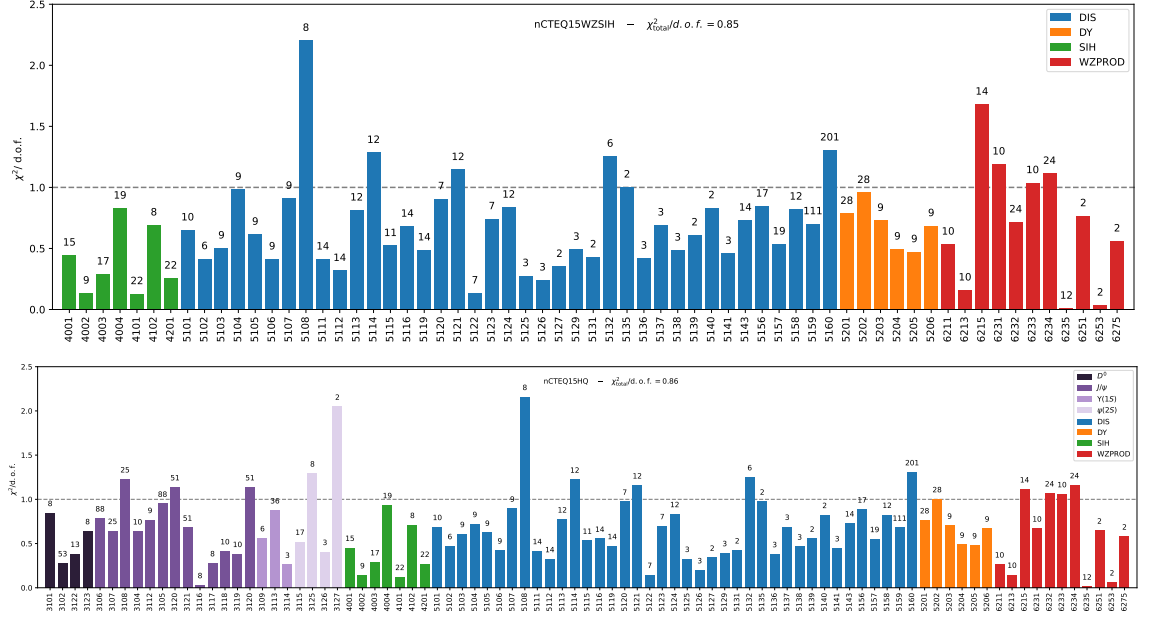


Figure II.28.: χ^2/N_{dof} values for each individual data set in the previous nCTEQ15WZ+SIH fit (upper panel) and the new nCTEQ15HQ fit (lower panel) color coded by process.

Table II.11.: χ^2/N_{dof} values for the the individual processes DIS, DY, WZ, SIH, HQ, individual heavy-quark final states, and their total. For data sets with normalization uncertainty, the values are the sum of regular χ^2 and normalization penalty. Parentheses indicate that a data set is not included in a fit or included only partially in the case of the SIH data in nCTEQ15(WZ). Note that some values are different when compared to Tab. ?? due to the changed normalization treatment.

	D^0	J/ψ	$\Upsilon(1S)$	$\psi(2S)$	DIS	DY	WZ	SIH	HQ	Total
nCTEQ15	(0.56)	(2.50)	(0.82)	(1.06)	0.86	0.78	(2.19)	(0.78)	(1.96)	1.23
nCTEQ15WZ	(0.32)	(1.04)	(0.76)	(1.02)	0.91	0.77	0.63	(0.47)	(0.92)	0.90
nCTEQ15WZ+SIH	(0.46)	(0.84)	(0.90)	(1.07)	0.91	0.77	0.72	0.40	(0.93)	0.92
nCTEQ15HQ	0.35	0.79	0.79	1.06	0.93	0.77	0.78	0.40	0.77	0.86

Parameter scans

To visualize the constraints that the new data puts on the individual nPDF parameters, we perform parameter scans along all of the seven gluon parameters $\{a_1^g, a_4^g, a_5^g, b_0^g, b_1^g, b_4^g, b_5^g\}$. Figure II.29 shows the $\Delta\chi^2$ profiles for these parameters, centered around their nCTEQ15HQ values, split into the different processes. Most parameters are clearly constrained tightly

3. Heavy quark production

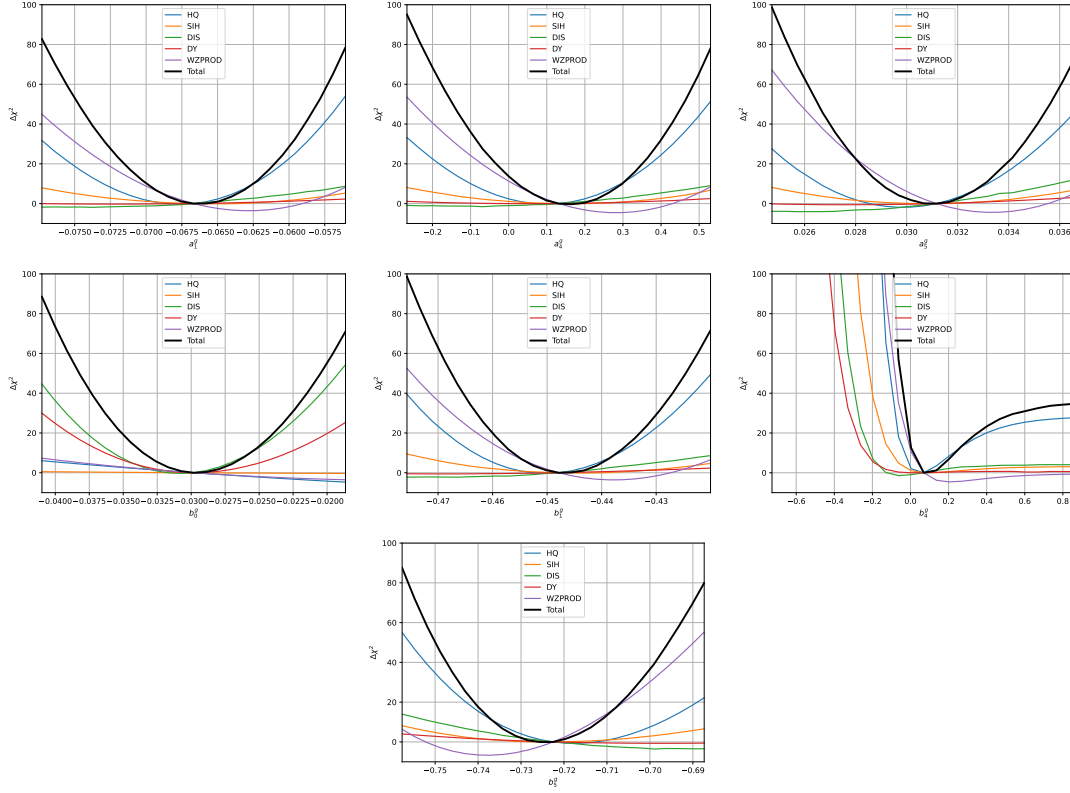


Figure II.29.: Parameter scans along the seven gluon parameters open during the fit, separated into the different processes with the total shown in black.

by the new heavy quark data with the WZ production data further constraining the parameters in one direction, because their minimum is not in the center. This does indicate some tensions between the data sets of the two processes, but the minimum of the WZ production curve is always close enough to the center, that it is well within expectations. The DIS data generally produces more shallow curves, where the minima are also slightly off-center, albeit in the opposite direction of the WZ production data. The single inclusive hadron production data have a similar impact as the DIS data, but are generally well centered. The Drell-Yan data barely affects the parameters. The parameter b_0^g is the sole exception to these observations, as it is mostly constrained by the DIS data with some contributions by the DY data. This parameter controls the nuclear A -dependence of the gluon normalization. It is, therefore, possible that this parameter mostly cancels out, when looking at a singular A , as it is the case for the LHC data taken exclusively in proton-lead collisions¹³. Additionally, the normalization parameters of the data sets can compensate for the gluon normalization for any process dominated by gluon contributions.

¹³Of course, the LHC also takes data in proton-proton and lead-lead collisions, but they are not included in the fit.

3.4. Conclusions

In conclusion, we managed to incorporate a vast new data set of open heavy-flavor and quarkonium production in a new global PDF fit and, thereby, put strong constraints on the gluon PDF.

We implemented a novel data-driven approach to derive an empirical theory for the production of these particles and compared the predictions from this to results from perturbative QCD calculations, *i.e.*, D^0 production in the GMVFNS and J/ψ production in NRQCD. Across the kinematic region included in this fit, the empirical theory produces a good description of the data and the uncertainties are significantly smaller than those estimated for the pQCD methods using scale variations.

The new global PDF analysis, including the heavy quark data, produces good χ^2/N_{dof} values for the new data without introducing tensions with the existing data sets. This is also reflected in the resulting PDFs, which have very similar central values than the previous nCTEQ15WZ(+SIH) fits, but the gluon uncertainties, especially those in the region of $x < 0.01$, are reduced dramatically. Through DGLAP evolution, this also causes a significant reduction in the quark uncertainties at low x . The uncertainty of the strange quark PDF remains very large, but the nCTEQ15 ν analysis [159] shows that this can be remedied through the inclusion of neutrino scattering data.

Fits without the added theoretical uncertainty showed that further improvement of the proton-proton baseline would not result in significant changes to the nPDF fit. The biggest impact of an improved baseline fit would be seen in the $\psi(2S)$ production, but the low number of proton-lead data points would still prevent large effects on the final PDF fit. Further proton-proton data would therefore mostly be useful for particles like D^\pm or higher Υ excitations, which are currently excluded from the fits.

The results also show no strong dependence on the scale choice for the heavy quark production, which stands in contrast to the findings from the previous reweighting analysis [193, 232] using the same data. The scale dependence in the new analysis is greatly diminished due to the cuts imposed on low- p_T data, which is most sensitive to scale changes. Additionally, the inclusion of normalizations as nuisance parameters in the fit mitigates the scale dependence further due to the large normalization uncertainties attached to many of the new data sets.

Currently, the data does not show any indication that higher-twist effects, like fully coherent energy loss, need to be incorporated in the calculation, as that should result in tensions between the heavy quark and vector boson production data. However, the vector boson data is not sensitive to the very low- x region, where tensions might therefore be hidden. This region could be investigated by including prompt photon production in the nPDF fit, which is unaffected by FCEL due to the colorless final state. If a better description of the heavy quark production in proton-proton collisions at $p_T < 3 \text{ GeV}$ is

found, such that the kinematic cut can be lowered, it would be necessary to incorporate FCEL effects in the nPDF fit, due to its large impact in this region.

Finally, one caveat to the analysis has to be mentioned: The proton PDF baseline. Like previous nCTEQ analyses, we use CTEQ6.1m as the proton PDF baseline, which is quite old and lacks many of the modern and precise data sets that are included in newer PDF fits like CT18[18], CJ15[273] and NNPDF4.0[274]. Due to the determination of the Crystal Ball parameters from proton-proton data, the choice proton PDF directly affects the extracted nuclear PDFs. Additionally, at the newly reached level of gluon PDF accuracy, the contribution of the proton PDF uncertainty may no longer be negligible [275]. Changing the proton PDF baseline comes with a variety of technical challenges and also requires a re-evaluation of all established data sets that are given in terms of ratios. The change to an updated proton PDF baseline is therefore left for a dedicated future study.

4. Impact on other observables

In the final section of this chapter, we will take a look at the consequences of the new fits on various observables and present comparisons with the results of nPDF analyses by other groups. We begin by taking a different look at the heavy quark data in terms of the rapidity dependence and continue with heavy quark data from lead-lead collisions, to compare the difference between the full nuclear modification and the modification purely caused by cold nuclear effects. Afterwards, we take a look back at the vector boson and single inclusive hadron production data from the nCTEQ15WZ and nCTEQ15WZ+SIH analyses to see how the descriptions of these sets with strong relation to the heavy quarks has changed. This is followed by a look ahead towards prompt photon production data, which has the potential to deliver complementary constraints on a similar kinematic region as the heavy quark data without being affected by the same higher twist effects. Finally, we conclude the section with a comparison of the new nCTEQ15HQ nuclear PDFs with PDFs obtained in analyses by other groups using different frameworks.

4.1. Rapidity dependence of heavy quark production

First, we take another look at the data used in the fit. Instead of the p_T distributions, we can also look at the rapidity dependence. Some experimental measurements provide the cross section as a differential in rapidity only, but those cannot be included in the fit, because the p_T integration from zero to infinity would violate the kinematic cuts. To gain insight into the rapidity dependence regardless of this, we can look at a specific p_T bin, *e.g.*, $4 \text{ GeV} < p_T < 5 \text{ GeV}$, of the double differential data used in the fit. This is shown for prompt D^0 and J/ψ in the upper and lower panel of Fig. II.30, respectively. The data is compared to predictions calculated with nCTEQ15WZ+SIH and nCTEQ15HQ PDFs and their respective uncertainties. The Crystal Ball uncertainties are not shown. The different PDFs lead to very similar predictions in the backwards rapidity region, where the data is already well described by the older fit. In the forward rapidity region, the nCTEQ15HQ PDFs yield slightly lower cross sections, which correspond to a better description of the data than nCTEQ15WZ+SIH. The uncertainties in this region are also reduced dramatically due to the fact that this is the direction where the low x gluon of the heavy nucleus is probed most strongly.

4.2. Cold nuclear effects in lead-lead collisions

It is also interesting to take a look at some examples of data from lead-lead collisions, which we do not include in nPDF fits. As mentioned before, the twist-2 factorization does not hold in collisions between two heavy ions, due to the strong effects of the nuclear medium created in the collisions. However, this does not mean that our results are useless for

4. Impact on other observables

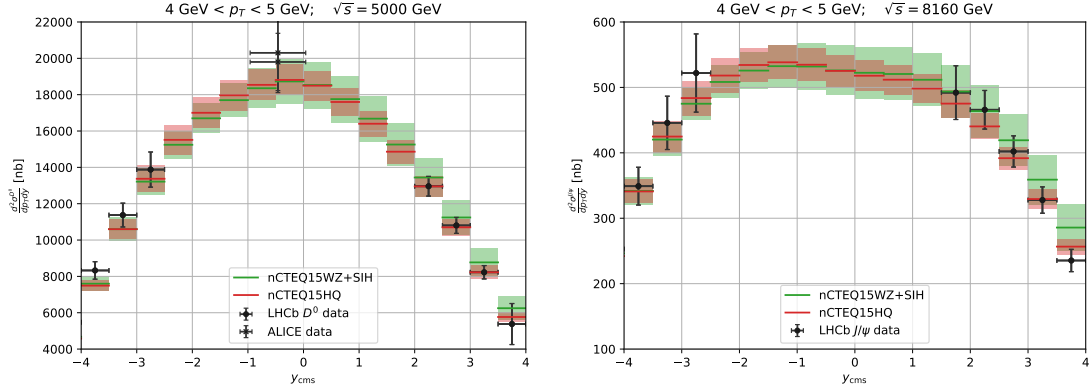


Figure II.30.: Rapidity dependence of the cross section for J/ψ [258] and D^0 [237, 262, 263]. The theory uncertainties do not include the Crystal Ball uncertainty.

observables measured in such collisions. By comparing the total measured cross section in lead-lead collisions, or equivalently the ratio between lead-lead and proton-proton cross section

$$R_{PbPb} \equiv \frac{d\sigma^{PbPb \rightarrow Q}}{dydp_T} \bigg/ \frac{d\sigma^{pp \rightarrow Q}}{dydp_T}, \quad (149)$$

with predictions assuming twist-2 factorization and our nuclear PDFs, we can estimate the magnitude of the hot medium effects from the difference between the two. This is shown in Fig. II.31 for data taken by CMS on prompt D^0 , prompt J/ψ and inclusive $\Upsilon(1S)$ production [276–278]. We see that the cross section in lead-lead collisions is suppressed compared to the proton-proton case for all three particles across the entire p_T range, while the cold nuclear effects only predict a suppression below $p_T < 10$ GeV and an enhancement above. The predictions for $\Upsilon(1S)$ are consistent with unity across the entire p_T range, while the measured R_{PbPb} lies around 0.4. A similar observation holds for J/ψ , but a slight upwards slope is visible in both data and prediction. The prediction for D^0 mesons show a very similar behavior, aside from slightly stronger suppression in the low p_T region. The measured ratio, however, has a significantly different shape than the ones for the quarkonia, such that the data and prediction actually coincide at low p_T . In general, the cold nuclear effects are significantly smaller than those caused by the hot medium.

4.3. Impact on single inclusive hadron and vector boson production

After looking closely at the heavy quark data, we now take a look back at the vector boson production and single inclusive hadron data sets. The goal is to compare the new nCTEQ15HQ PDFs to the nCTEQ15WZ PDFs, which represents the baseline upon which the two analyses of this thesis are built. An overview of the data sets can be found in

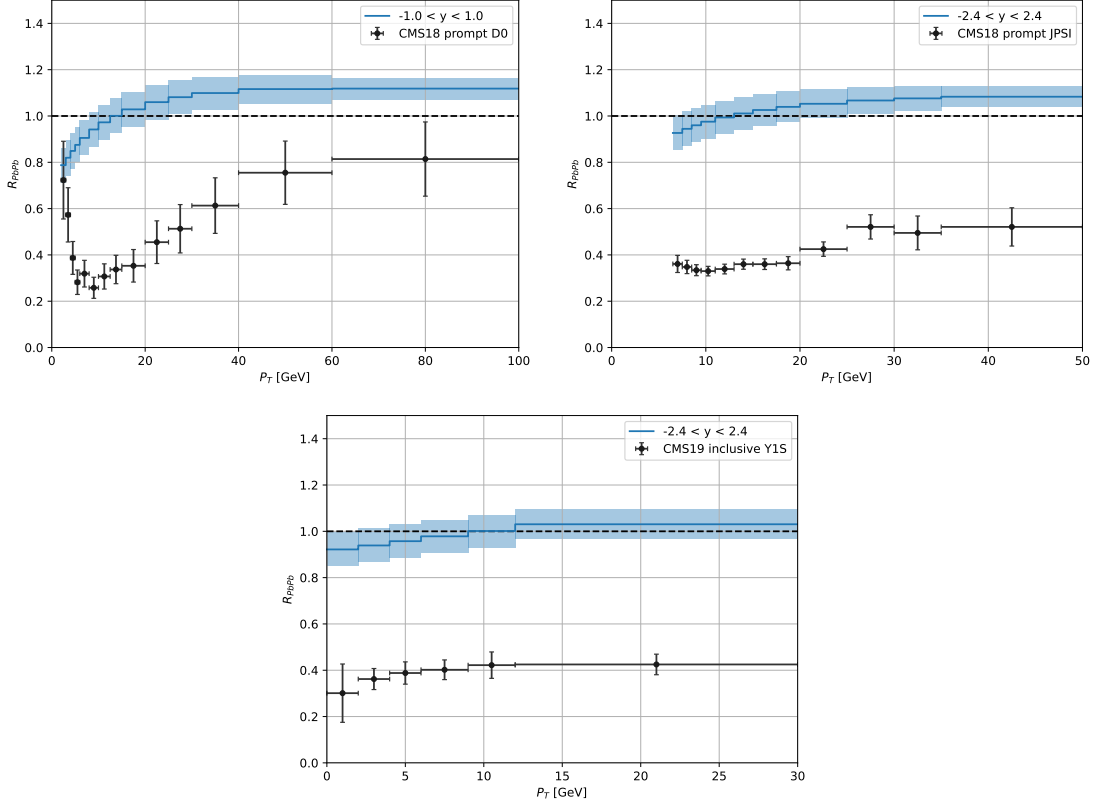


Figure II.31.: Comparison of R_{PbPb} as predicted assuming only cold nuclear effects and measured ratio by CMS for D^0 , J/ψ and $\Upsilon(1S)$ [276–278].

Tabs. IV.7 and II.2 for the WZ and SIH data, respectively. We leave out comparisons to the DIS and DY data included in the fits, since their impact is only weakly linked to the new analyses, which focus on the gluon PDF. Figure II.32 shows a comparison between nCTEQ15WZ and nCTEQ15HQ for the WZ production data, where the calculations for both PDFs are carried out with the same APPLgrids as we use in the fits. Overall, the agreement between the two predictions is very good and both yield a satisfying description of the data across the entire kinematic range. The backward and central rapidity regions show minor differences in the central values, but the two predictions lie well within each others uncertainties. At forward rapidity the two central predictions coincide closely, but nCTEQ15HQ produces significantly smaller uncertainties. Looking back at the PDF comparison in Figs. II.24 and II.25 both of these observations are explained by the difference in PDFs. The change in the gluon's central value around $x \in [0.005, 0.2]$ explains the differences at central and backward rapidity, while the reduced uncertainties at even lower x cause the reduced uncertainties for the predictions at forward rapidity.

Next, we perform an additional comparison to the single inclusive hadron data sets that were introduced in the nCTEQ15WZ+SIH analysis. Again, we compare the predictions of

4. Impact on other observables

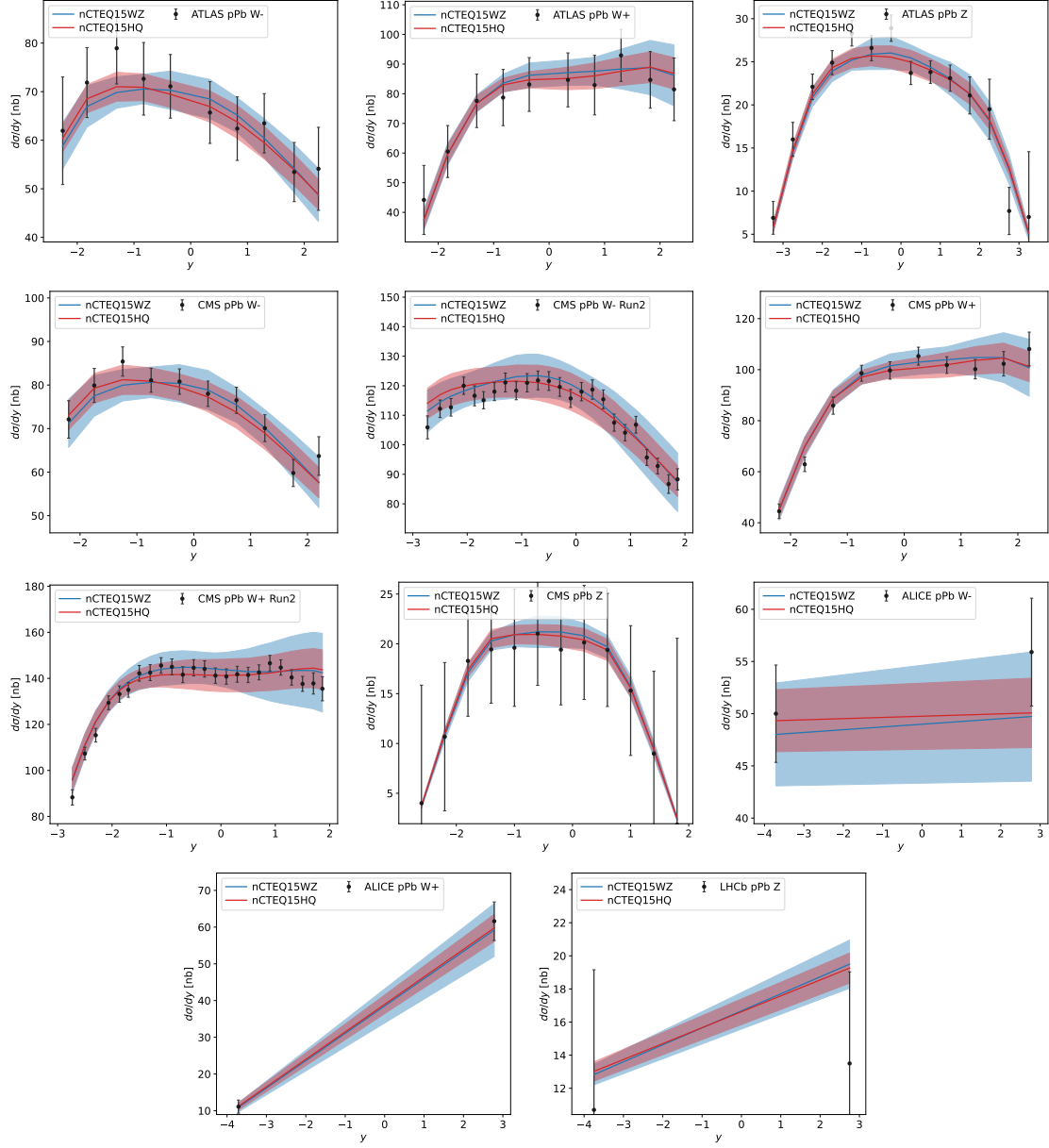


Figure II.32.: Comparison of the vector boson production data sets with predictions from nCTEQ15WZ and nCTEQ15HQ.

nCTEQ15WZ and nCTEQ15HQ PDFs and keep all remaining settings as they were in the PDF fit, *i.e.* the scales are chosen as $\mu_i = \mu_r = \mu_f = \frac{1}{2}p_T$ and the gridded DSS FFs are used for the final state fragmentation. The central values of the two fits are slightly further apart than they were in the case of WZ data, which is unsurprising since the nCTEQ15WZ fit did not include most of the SIH data, except for the neutral pions measured at RHIC. In the case of the ALICE data sets, the two predictions agree closely at low p_T , but nCTEQ15WZ lies a few percent above nCTEQ15HQ at higher p_T . At RHIC energy, the predictions from

nCTEQ15HQ lie below those of nCTEQ15WZ for $p_T \lesssim 7$ GeV before the two cross each other. However, the differences are small enough that the uncertainties of either fit still include the central value of the other one. In general, the uncertainties of the newer fit are smaller and the difference is especially pronounced in the low- p_T region, which corresponds to low x values. While the nCTEQ15WZ predictions have larger uncertainties than the data, the newer fit is generally more precise than the data included by the cuts, but could still profit from the excluded low- p_T data if better theory predictions become available.

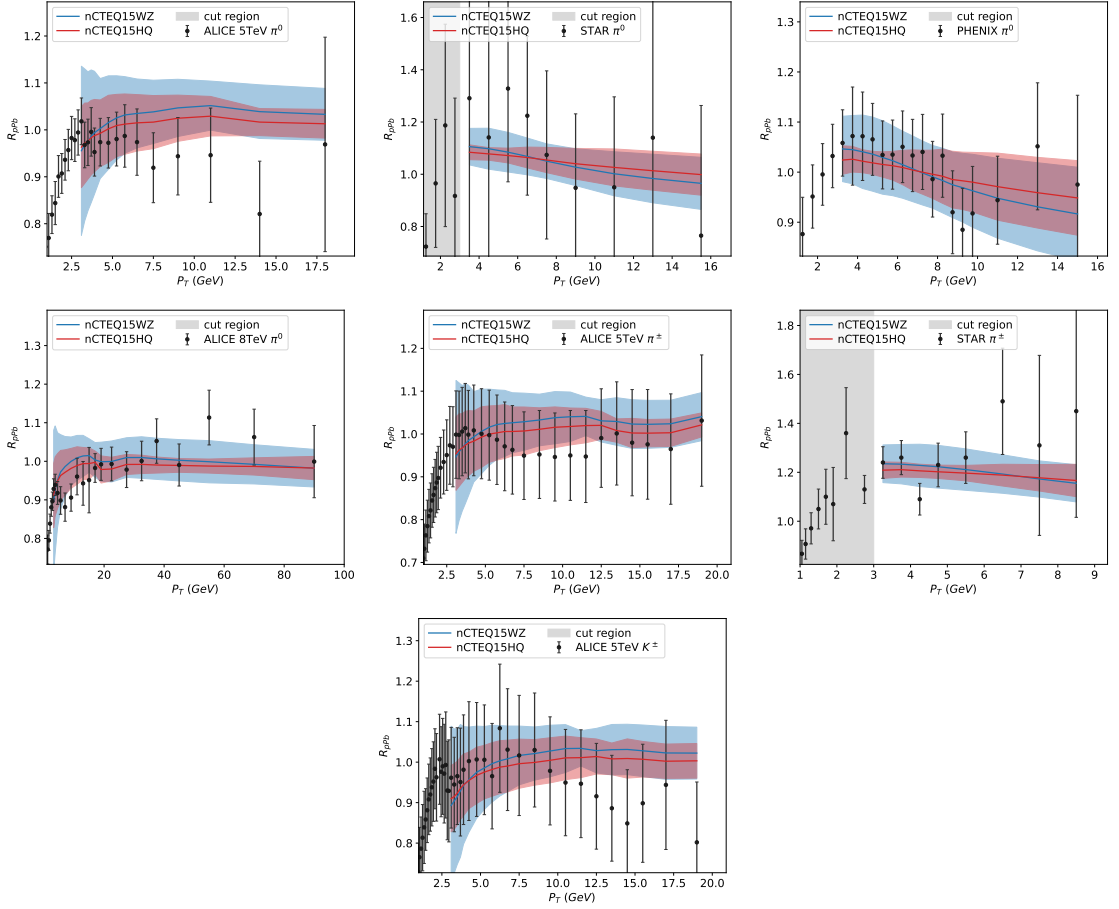


Figure II.33.: Comparison of the single inclusive hadron production data sets with predictions from nCTEQ15WZ and nCTEQ15HQ. Both are calculated with gridded DSS FFs and scales $\mu_i = \mu_r = \mu_f = \frac{1}{2}p_T$.

4.4. Prompt photon production

Another interesting observable is the production of prompt photons, because it can probe a similar kinematic region, but is not affected by the same higher-twist effects as heavy quark production due to the colorless final state. Currently, there is only one data set [41] available for this process in proton-lead collisions, taken by the ATLAS experiment in

collisions at $\sqrt{s} = 8.16$ TeV. The data is split into three pseudorapidity bins

$$\begin{aligned} -2.83 < \eta < -2.02, \\ -1.84 < \eta < 0.91, \\ 1.09 < \eta < 1.9, \end{aligned}$$

with around 15 E_T bins spanning from 20 GeV to 550 GeV. The data is available both as the differential cross section $\frac{d\sigma}{dE_T}$ and as the nuclear modification ratio

$$R_{pPb} \equiv \frac{d\sigma^{pPb \rightarrow \gamma}}{dE_T} \bigg/ \frac{d\sigma^{pp \rightarrow \gamma}}{dE_T}. \quad (150)$$

Some uncertainties, including the normalization uncertainty from the luminosity, cancel fully in this ratio, but the uncertainties are still of similar size as the modification across the kinematic range.

Using MCFM-6.8 [158] and APPLgrid [157], we can create grids to produce theoretical predictions for the cross section with our PDFs. We set the renormalization and factorization scales to be equal to the central value of the photon's transverse energy E_T . The resulting predictions for the cross section and nuclear modification are shown in the left and right panels of Fig. II.34, respectively. The qualitative agreement between the measured and predicted cross section is very good. However, the χ^2/N_{dof} is rather high with a value of 1.66. This disagreement is mostly due to the normalization penalty, which contributes more than 40% of the total χ^2 . The normalization uncertainty of the data is given as 2.4%, while the fitted normalization differs from unity by 13%. This can be partially explained by the missing higher orders, since NNLO corrections are known to enhance the cross section across the entire E_T range [279]. In the nuclear modification ratio, where these missing higher orders mostly cancel out, however, the data is very well described with a χ^2/N_{dof} of 0.53. This is quite noteworthy, since the R_{pPb} data is included in the recent nNNPDF3.0 analysis, where it yields a value of $\chi^2/N_{\text{dof}} = 1.03$ and the excluded cross section data yields $\chi^2/N_{\text{dof}} = 3.3$. Comparing the cross section data to predictions with nCTEQ15 PDFs yields a χ^2/N_{dof} of 2.41, which suggests that the inclusion of heavy quark production vastly improves the description of the prompt photon data, as well. The opposite would be expected if fully coherent energy loss effects would make up a significant part of the nuclear modification.

4.5. Comparison with other nuclear PDF fits

Finally, we want to provide a direct comparison with other nuclear PDF analyses, beginning with the PDFs obtained in the reweighting analysis [193]. In this analysis, data on prompt D^0 , prompt J/ψ , non-prompt J/ψ and inclusive $\Upsilon(1S)$ production was used to perform a

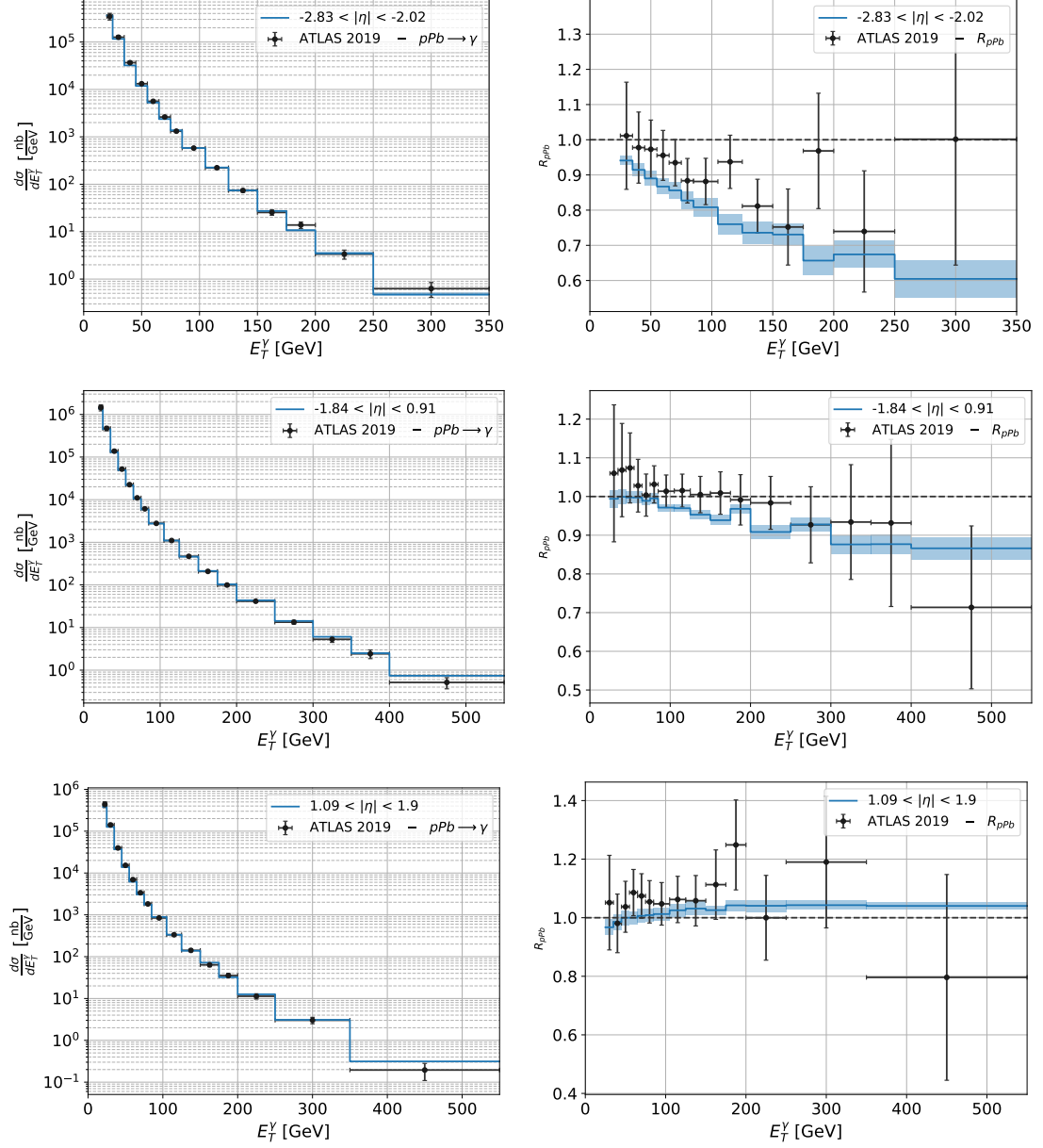


Figure II.34.: Comparison between prompt photon production in proton lead collisions measured by ATLAS [41] and predicted with MCFM-6.8 and nCTEQ15HQ PDFs. The top panels show the cross section, while the bottom ones show the nuclear modification R_{pPb} .

4. Impact on other observables

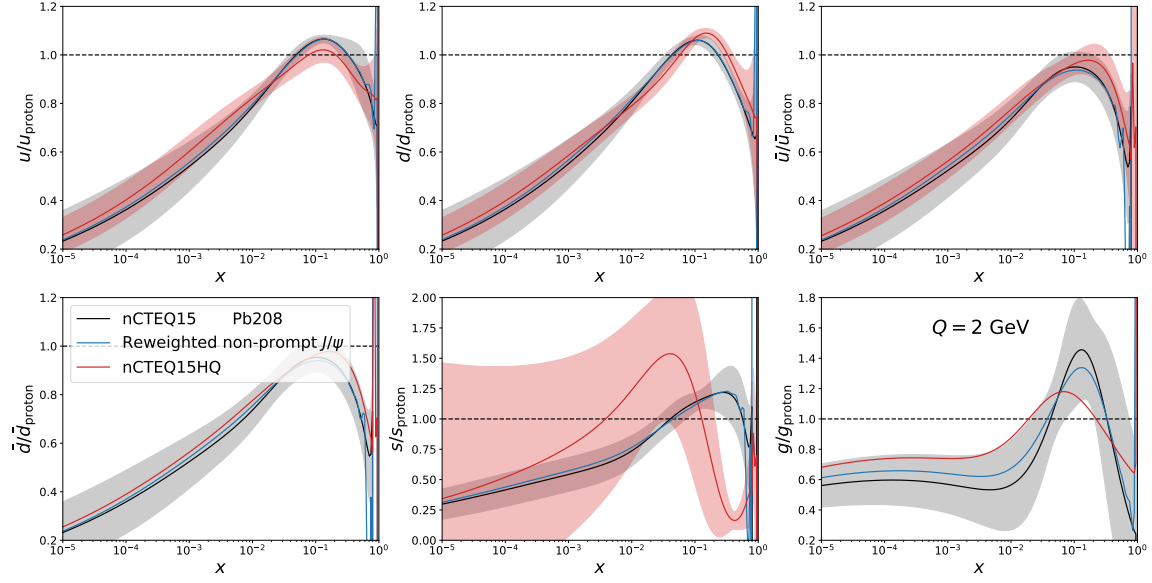


Figure II.35.: Comparison of the nuclear modification of lead between nCTEQ15, nCTEQ15HQ and nCTEQ15 reweighted with non-prompt J/ψ data.

Bayesian reweighting on the eigenvector PDFs of the nCTEQ15 analysis to estimate the impact that these data sets could have in a global analysis. Even though this might seem very similar to the analysis presented here, there are a variety of further improvements implemented that go beyond the step from a reweighting to a global fit. Firstly, the reweighting analysis did not impose any cuts on the data and used a parameterization without explicit y dependence. This lead to a significantly worse χ^2 , particularly in the proton baseline as compared to the present analysis. Additionally, the reweighting was limited to a single final state at a time and many of the newer data sets were excluded, such that significantly less data was involved in each reweighting. Finally, the normalization uncertainties were simply added in quadrature to the statistical and systematic ones in the reweighting study. Figure II.35 shows a comparison of the nuclear PDF modification at $Q = 2$ GeV between nCTEQ15, nCTEQ15HQ and nCTEQ15 reweighted with non-prompt J/ψ data. For visual clarity we omit the other reweighted PDFs in the plot and show only the one where the largest impact on the PDFs was observed. In both new analyses, the up and down quark PDFs remain largely unchanged from their nCTEQ15 values. The quark uncertainty is actually smaller in the reweighted PDFs, which is likely explained by the lack of freedom in the strange PDF, which had no open parameters in nCTEQ15 and accordingly in the reweighting. In the gluon case the reweighted PDFs sit between nCTEQ15 and the global fit across the entire x range, but due to the lower number of data points they remain closer to nCTEQ15. Nevertheless, it can be seen that such reweighting studies can present an accurate proof-of-concept for future global fits.

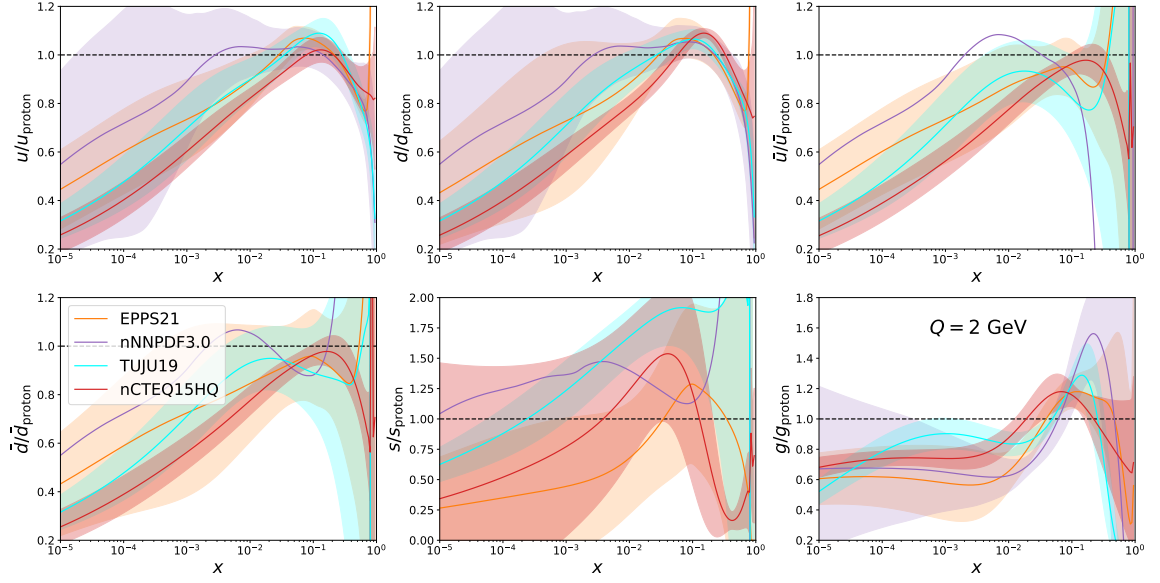


Figure II.36.: Comparison of the nuclear modification of lead between nCTEQ15HQ [65], nNNPDF3.0 [194], EPPS21 [14] and TUJU19 [280].

It is also interesting to compare the obtained nuclear PDFs to those obtained in analyses by other groups using different methods and tools, as described in the overview in Sec. II.1. Such a comparison is presented in Fig. II.36 for the nuclear modification of lead in nCTEQ15HQ, nNNPDF3.0, EPPS21 and TUJU19¹⁴. The up and down quark PDFs show qualitative agreement, but the amount by which the low- x region is suppressed varies between the analyses, with nNNPDF3.0 showing the least and nCTEQ15HQ showing the strongest suppression. However, even those two share overlapping error bands due to the large uncertainty on the nNNPDF3.0 fit. The antiquarks paint a similar picture in the low x region, but there is more disagreement on the position of the maximum in the medium x region. The strange quark PDFs are only weakly constrained in all analyses except for TUJU19, which includes a vast number of neutrino DIS data to constrain this flavor. The gluon PDF shows the same qualitative shape for nCTEQ15HQ, nNNPDF3.0 and EPPS21 with a sharp drop at high- x , a peak in the medium- x region and a suppression that levels off at low x . TUJU19 shows the same behavior with the exception of the $x < 10^{-3}$ region, where the suppression becomes stronger again. Nevertheless, the four analyses are in good agreement with overlapping uncertainties over the vast majority of x values, but the large differences between central values also show significant room for improvement that is to be explored in future analyses.

¹⁴The more recent TUJU21 PDFs are not yet publicly available.

III. Conclusions and Outlook

1. Summary

Tight constraints on the nuclear gluon PDF are important to make precise predictions in a variety of physical processes, but have been elusive due to the lack of the necessary data in global nPDF fits. Addressing this long standing problem by including data on inclusive light and heavy meson production was the main goal of this thesis. Additionally, we include a brief review of the theoretical framework of the parton model and the required statistical tools.

Advancements in the theoretical framework

The first chapter began with a derivation of the parton model from experimental data on deep inelastic scattering to motivate the definition of PDFs as the momentum distributions of a proton's constituents. Following this, we introduced extensions to the theory, *i.e.*, the perturbate Q^2 -dependence from the DGLAP evolution and the first evidence for nuclear modification of PDFs.

Given this physical framework, the next step was the introduction of the required statistical tools to determine the non-perturbative PDFs from experimental data. The statistical tools are based on the χ^2 loss function with modifications for correlated errors and normalization uncertainties. We used this definition to derive the Hessian formalism for the estimation of PDF uncertainties and a variety of other tools to test the fit quality.

In addition to these more general aspects, we have introduced the theoretical approaches to calculate the cross sections of the new inclusive light and heavy meson production processes.

The above mentioned theoretical framework and statistical tools were then used to perform two new global nPDF analyses.

nCTEQ15WZ+SIH

The first new analysis explored the impact of pion, kaon and η meson data in nuclear PDF fits. The calculation of the cross section for these processes depends on a secondary non-perturbative quantity - the final state fragmentation functions. Like PDFs, these have been determined in a variety of global analyses, which required us to find a way of mitigating the

impact of the FF uncertainty on our PDF fit. In the first step of the analysis we therefore benchmark various FFs along with different scale choices against data from proton-proton collisions. These benchmarks lead to the conclusion that the DSS fragmentation functions with scales $\mu_i = \mu_r = \mu_f = \frac{1}{2}p_T$ are the optimal choice for our analysis and that no combination of FFs and scales can properly describe the data with $p_T < 3$ GeV. We then further mitigated the impact of the FF choice by using only R_{pPb} data in the nPDF analysis to cancel out parts of the FF dependence and then added the uncertainties determined from the Hessian eigenvector FFs to the data to account for any residual impact.

The nuclear PDF fits performed with this setup show to a noticeable decrease in gluon uncertainty around $x \approx 10^{-2}$ when compared to nCTEQ15WZ. The central value of the gluon PDF remains similar, but we see a slight enhancement in the low- x region and corresponding suppression at high x .

Looking at detailed breakdowns of the χ^2 , the new fit is in good agreement with both the newly added data, as well as all data from the previous nCTEQ15WZ analysis. These breakdowns indicate that the neutral pion production data from ALICE is the driving force of the observed changes, but the overall impact of the inclusive hadron production remains limited by the kinematic cuts, which remove more than 60% of the data and particularly the low p_T , which has very small uncertainties.

Alternative fits have also been performed including η meson production, but due to the large data uncertainties, the impact is barely noticeable.

nCTEQ15HQ

The second analysis also dealt with inclusive meson production, but this time particularly those involving heavy quarks. The bulk of the data in this case comes from quarkonia, whose production mechanism is still a topic of much debate and different existing theories produce incompatible results.

To get around this, a new data-driven approach was employed, which combines the hadronization and the partonic cross section in the general factorization formula into one parameterized function that is then determined empirically.

The parameters of this theory are determined from heavy quark production data of proton-proton collisions. This approach produced a very good description of the baseline data, when cuts $p_T < 3$ GeV and $|y_{cms}| < 4$ are applied. These cuts are not problematic in this case, since a p_T cut would have been necessary either way for the proton-lead data to mitigate effects of fully coherent energy loss. The obtained theory was also compared with predictions for prompt D^0 production in the general-mass variable-flavor-number scheme and for prompt J/ψ production in non-relativistic QCD. This comparison showed good agreement of the central values, but significantly smaller uncertainties for the data-driven approach.

A total of 548 data points from D^0 , J/ψ , $\psi(2S)$ and $\Upsilon(1S)$ production in proton-lead collisions were then added to a new global PDF analysis. These new data sets showed good compatibility with each other and with the existing data sets of other processes. Particularly the vast body of J/ψ data put strong new constraints on the gluon PDF, which has been determined with unprecedented accuracy to x values as low as $x \approx 10^{-5}$. Due to DGLAP mixing, the reduced gluon uncertainty has also improved the errors on the up and down quarks, but the strange quark uncertainty remains large. The good agreement with previous analyses excluding the heavy quark data means that there are currently no indications of significant higher-twist effects that would need to be considered, but complementary data in the low- x region would be required to fully rule this out.

Both the nCTEQ15WZ+SIH and the subsequent nCTEQ15HQ studies represent important steps towards future nuclear PDF analyses by introducing not just new data but also new tools for future analyses.

2. Prospects for future studies

This final section of the thesis gives an outlook towards possible future improvements of nuclear PDF analyses. We will begin with projects that are already being worked on, or planned for the immediate future. Next, we present some improvements that will take longer time, or represent continuous processes that will be relevant as long as people study parton distributions. Finally, we mention some prospects that may remain out of focus for the current decade, but represent important steps in the long term development of nuclear PDFs.

Short term and work in progress

The first, and most obvious idea for improving nuclear parton distribution functions is the inclusion of further data sets. In particular relevant for the gluon PDF are dijet and prompt photon production. The latter is especially interesting, because it is a color neutral probe for the low- x gluon, which should show tensions with the heavy quark data, if fully coherent energy loss has noticeable effects.

Another important step is the combination of the analyses presented in this thesis with the analyses performed in parallel on different aspects of nuclear PDFs, *i.e.*, nCTEQ15HIX and nCTEQ15 ν . Such a major overhaul of the nCTEQ15 analysis should also include the switch to a more modern proton PDF baseline as these baseline uncertainties become more relevant with every improvement of the nuclear PDFs.

Naturally, a simultaneous fit of the proton and nuclear PDFs would be the optimal solution, but it is certainly wise to take one step after another and begin with a standalone proton analysis.

Additionally, there are projects to use the current framework to test models explaining the nuclear modification, like the leading twist theory of nuclear (anti)-shadowing [49, 91]. There is also ongoing work on nuclear PDFs with a parameterization motivated by the theory of short range correlations, which suggests that the nuclear modification of PDFs comes from the temporary formation of nucleon pairs in the nucleus [58, 115].

Intermediate timescale and continuous improvements

On somewhat longer timescales, possible improvements include the extension of inclusive hadron production mentioned in the conclusion of Sec. II.2, *e.g.*, by improving the theory for light meson production such that the very precise low- p_T data can be used.

The theoretical description of other processes can be improved as well, *e.g.*, by using NNLO calculations or by including higher-twist effects like FCEL in heavy quark production.

Another improvement of the heavy quark study would be the application of the approach to further particles like higher excitations of Υ , or more open heavy flavored mesons like D^\pm and B mesons.

Generally speaking, there has to be a continuous effort to keep all implemented processes up-to-date in terms of newly released data and to implement new processes that open up new kinematic regions or provide complementary test of existing regions.

With increasing amounts of data, the parameterization of the nPDFs can also become more flexible, which allows for a more faithful estimation of the uncertainties by removing potential biases.

Long term goals

Finally, we can take a peek at the long term future of (nuclear) parton distribution functions.

On the experimental side, there are two major developments coming in the next decade: the high-luminosity upgrade for the LHC [281] and the building of the electron-ion collider (EIC) [6, 7]. The former will mostly continue with a similar physics program as the current LHC and improve the accuracy of the data through significantly higher statistics, but there are also proposals for new experiments like the forward physics facility (FPF) [282], which will allow for conceptually new measurements. The electron-ion collider will be even more important for nuclear PDFs, as it explicitly aims to perform measurements on a variety of different nuclei, which will give us a better understanding of the A dependence of nuclear PDFs and extend the kinematic range of the crucial DIS data by an order of magnitude in both x and Q^2 [283].

To anticipate future improvements of the theoretical framework for nuclear parton distribution functions, we can take a look at currently available proton PDFs, which are far

ahead due to their simpler nature and more abundant data. Examples include proton PDFs with transverse structure [31], polarization of partons [32, 33] and multiple scatterings [34, 35]. These improvements can be combined under the concept of generalized parton distributions [30]. Another interesting possibility is the combined fitting of multiple quantities, *i.e.*, proton PDFs, nuclear PDFs and FFs, to gain a deeper understanding of the correlations between these quantities.

Both the experimental and theoretical prospects listed here are far from being an exhaustive list, as the possibilities for improvement are almost endless. Just like they have done for the past four decades, nuclear parton distribution functions will certainly represent a field of abundant research opportunities for many decades to come.

IV. Appendix

1. Feynman rules

All Feynman diagrams shown here and in the rest of this thesis were created using the TikzFeynman L^AT_EX package [284].

1.1. Quantum electrodynamics

From the QED Lagrangian

$$\mathcal{L} = \frac{1}{4}F_{\mu\nu}^2 + i\bar{\psi}\not{D}\psi - m\bar{\psi}\psi \quad \text{with} \quad D_\mu\psi = \partial_\mu\psi + ieA_\mu\psi \quad (151)$$

we get the following Feynman rules:

Photon propagator:

$$\text{~~~~~} = \frac{-i}{p^2 + i\epsilon} \left[g_{\mu\nu} - (1 - \xi) \frac{p_\mu p_\nu}{p^2} \right] \quad \text{Feynman gauge} \quad \frac{-ig_{\mu\nu}}{p^2 + i\epsilon}. \quad (152)$$

Unless otherwise specified, we will always work in Feynman gauge.

Fermion propagator:

$$\text{---}\blacktriangleright\text{---} = \frac{i(\not{p} + m)}{p^2 - m^2 + i\epsilon} \quad (\text{Arrow right for particles, left for antiparticles}). \quad (153)$$

External photons (small circle denotes connection to the rest of the diagram):

$$\text{~~~~~}\circ = \epsilon_\mu(p) \quad (\text{incoming}), \quad (154)$$

$$\circ\text{~~~~~} = \epsilon_\mu^*(p) \quad (\text{outgoing}). \quad (155)$$

External fermions:

$$\longrightarrow \circ = u^s(p), \quad (156)$$

$$\circ \longrightarrow = \bar{u}^s(p), \quad (157)$$

$$\longleftarrow \circ = \bar{v}^s(p), \quad (158)$$

$$\circ \longleftarrow = v^s(p). \quad (159)$$

Interaction vertices:

$$\begin{array}{c} \longrightarrow \\ \bullet \\ \swarrow \nearrow \\ \text{wavy line} \end{array} = \begin{array}{c} \longleftarrow \\ \bullet \\ \swarrow \nearrow \\ \text{wavy line} \end{array} = \begin{array}{c} \text{wavy line} \\ \bullet \\ \swarrow \nearrow \end{array} = \begin{array}{c} \swarrow \nearrow \\ \bullet \\ \text{wavy line} \end{array} = -ie\gamma^\mu. \quad (160)$$

External fermions need to be on-shell according to the LSZ-reduction formula. This means we can use the equations of motion to simplify calculations:

$$(\not{p} - m)u^s(p) = \bar{u}^s(p)(\not{p} - m) = 0, \quad (161)$$

$$(\not{p} + m)v^s(p) = \bar{v}^s(p)(\not{p} + m) = 0. \quad (162)$$

1.2. Quantum chromodynamics

The $SU(N)$ -invariant Lagrangian of QCD for a set of N fermions and N scalars interacting with non-Abelian gauge fields is given by:

$$\begin{aligned} \mathcal{L} = & -\frac{1}{4}(F_{\mu\nu}^a)^2 - \frac{1}{2\xi}(\partial_\mu A_\mu^a)^2 + (\partial_\mu \bar{c}^a)(\delta^{ac}\partial_\mu + gf^{abc}A_\mu^b)c^c \\ & + \bar{\psi}_i(\delta_{ij}i\not{\partial} + gA^a T_{ij}^a - m\delta_{ij})\psi_j, \end{aligned} \quad (163)$$

$$\text{with} \quad F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + gf^{abc}A_\mu^b A_\nu^c. \quad (164)$$

In the Lagrangian c^a and \bar{c}^a are the Fadeev-Popov ghosts and anti-ghosts respectively.

Gluon propagator:

$$\nu; b \text{ wavy line } \mu; a = \frac{-i}{p^2 + i\epsilon} \left[g_{\mu\nu} - (1 - \xi) \frac{p_\mu p_\nu}{p^2} \right] \delta^{ab}. \quad (165)$$

Ghost propagator:

$$b \text{ --- } \overset{p}{\text{---}} \text{ --- } a = \frac{-i}{p^2 + i\epsilon} \delta^{ab}. \quad (166)$$

Colored fermion propagator:

$$b \text{ --- } \overset{p}{\text{---}} \text{ --- } a = \frac{-i}{\not{p} - m + i\epsilon} \delta^{ab}. \quad (167)$$

If a colored propagator appears as an intermediate state, we need to sum over all possible colors.

Fermion vertex:

$$\begin{array}{c} \mu; c \\ | \\ \text{---} \bullet \text{---} \\ / \quad \backslash \\ b \quad a \end{array} = ig\gamma^\mu T_{ab}^c. \quad (168)$$

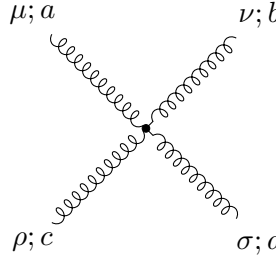
Ghost vertex:

$$\begin{array}{c} \mu; b \\ | \\ \text{---} \bullet \text{---} \\ / \quad \backslash \\ c^c \quad p \quad \bar{c}^a \end{array} = -gf^{abc} p^\mu. \quad (169)$$

3 gluon vertex:

$$\begin{array}{c} \nu; b \\ | \\ \text{---} \bullet \text{---} \\ / \quad \backslash \\ \rho; c \quad k \quad \mu; a \end{array} = gf^{abc} [g^{\mu\nu}(k-p)^\rho + g^{\nu\rho}(p-q)^\mu + g^{\rho\mu}(q-k)^\nu]. \quad (170)$$

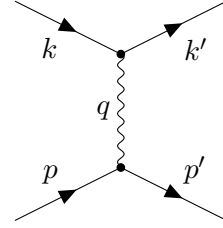
4 gluon vertex:



$$\begin{aligned}
 &= -ig^2 \times [f^{abe} f^{cde} (g^{\mu\rho} g^{\nu\sigma} - g^{\mu\sigma} g^{\nu\rho}) \\
 &\quad + f^{ace} f^{bde} (g^{\mu\nu} g^{\rho\sigma} - g^{\mu\sigma} g^{\nu\rho}) \\
 &\quad + f^{ade} f^{bce} (g^{\mu\nu} g^{\rho\sigma} - g^{\mu\rho} g^{\nu\sigma})].
 \end{aligned} \tag{171}$$

2. Definition of kinematic variables

- p : target momentum
- M : target mass
- k : initial momentum of incoming particle
- k' : momentum of incoming particle after scattering
- m : incoming particle mass
- q : momentum of exchange particle
- E : Energy of incoming particle
- E' : Energy of incoming particle after scattering



Kinematic Variables

$$\nu = \frac{q \cdot p}{M} \tag{172}$$

$$x = \frac{-q^2}{2p \cdot q} = \frac{-q^2}{2M\nu} \tag{173}$$

$$y = \frac{p \cdot q}{p \cdot k} \tag{174}$$

Lab frame kinematics

$$q^2 = -2(k \cdot k' - m^2) = 2k(k - k') = -4EE' \sin^2 \frac{\theta}{2} \quad (175)$$

$$k \cdot k' = 2EE' \sin^2 \frac{\theta}{2} \quad (176)$$

$$k \cdot p = EM \quad (177)$$

$$k' \cdot p = E'M \quad (178)$$

$$k \cdot q = k(k - k') = \frac{q^2}{2} \quad (179)$$

$$k' \cdot q = k(k' - k) = -\frac{q^2}{2} \quad (180)$$

$$q \cdot p = (E - E')M \quad (181)$$

$$\nu = E - E' \quad (182)$$

$$x = \frac{2EE'}{(E - E')M} \sin^2 \frac{\theta}{2} \quad (183)$$

$$y = \frac{E - E'}{E} \quad (184)$$

3. Additional tables

3. Additional tables

Table IV.1.: Values of the PDF parameters of the nCTEQ15WZ+SIH fit at the initial scale $Q_0 = 1.3$ GeV. Free parameters are printed bold, while fixed parameters are in parentheses. Parameters without listed values are fixed through sum rules.

Par.	Value	Par.	Value	Par.	Value	Par.	Value	Par.	Value	Par.	Value
p_0^g	(0.000)	—	—	—	—	—	—	$p_0^{s+\bar{s}}$	(0.500)	—	—
p_1^g	(0.523)	$p_1^{u_v}$	(0.630)	$p_1^{d_v}$	(0.513)	$p_1^{\bar{d}+\bar{u}}$	(-0.324)	$p_1^{s+\bar{s}}$	(-0.324)	$p_1^{\bar{d}/\bar{u}}$	(10.075)
p_2^g	(3.034)	$p_2^{u_v}$	(2.934)	$p_2^{d_v}$	(4.211)	$p_2^{\bar{d}+\bar{u}}$	(8.116)	$p_2^{s+\bar{s}}$	(8.116)	$p_2^{\bar{d}/\bar{u}}$	(4.957)
p_3^g	(4.394)	$p_3^{u_v}$	(-2.369)	$p_3^{d_v}$	(-2.375)	$p_3^{\bar{d}+\bar{u}}$	(0.413)	$p_3^{s+\bar{s}}$	(0.413)	$p_3^{\bar{d}/\bar{u}}$	(15.167)
p_4^g	(2.359)	$p_4^{u_v}$	(1.266)	$p_4^{d_v}$	(0.965)	$p_4^{\bar{d}+\bar{u}}$	(4.754)	$p_4^{s+\bar{s}}$	(4.754)	$p_4^{\bar{d}/\bar{u}}$	(17.000)
p_5^g	(-3.000)	$p_5^{u_v}$	(1.718)	$p_5^{d_v}$	(3.000)	$p_5^{\bar{d}+\bar{u}}$	(0.614)	$p_5^{s+\bar{s}}$	(0.614)	$p_5^{\bar{d}/\bar{u}}$	(9.948)
Par.	Value	Par.	Value	Par.	Value	Par.	Value	Par.	Value	Par.	Value
a_0^g	(-0.256)	—	—	—	—	—	—	$a_0^{s+\bar{s}}$	0.541	—	—
a_1^g	-0.022	$a_1^{u_v}$	-9.003	$a_1^{d_v}$	0.095	$a_1^{\bar{d}+\bar{u}}$	0.348	$a_1^{s+\bar{s}}$	0.707	$a_1^{\bar{d}/\bar{u}}$	(0.000)
a_2^g	(0.000)	$a_2^{u_v}$	-0.127	$a_2^{d_v}$	-0.154	$a_2^{\bar{d}+\bar{u}}$	(0.415)	$a_2^{s+\bar{s}}$	1.633	$a_2^{\bar{d}/\bar{u}}$	(0.000)
a_3^g	(0.383)	$a_3^{u_v}$	(0.018)	$a_3^{d_v}$	(0.085)	$a_3^{\bar{d}+\bar{u}}$	(-0.759)	$a_3^{s+\bar{s}}$	(0.000)	$a_3^{\bar{d}/\bar{u}}$	(0.000)
a_4^g	0.036	$a_4^{u_v}$	16.490	$a_4^{d_v}$	(3.874)	$a_4^{\bar{d}+\bar{u}}$	(-0.203)	$a_4^{s+\bar{s}}$	(0.000)	$a_4^{\bar{d}/\bar{u}}$	(0.000)
a_5^g	0.011	$a_5^{u_v}$	-0.970	$a_5^{d_v}$	-0.494	$a_5^{\bar{d}+\bar{u}}$	-0.044	$a_5^{s+\bar{s}}$	(0.000)	$a_5^{\bar{d}/\bar{u}}$	(0.000)
Par.	Value	Par.	Value	Par.	Value	Par.	Value	Par.	Value	Par.	Value
b_0^g	-0.021	—	—	—	—	—	—	$b_0^{s+\bar{s}}$	(0.104)	—	—
b_1^g	0.215	$b_1^{u_v}$	(0.006)	$b_1^{d_v}$	(0.466)	$b_1^{\bar{d}+\bar{u}}$	(0.172)	$b_1^{s+\bar{s}}$	(0.172)	$b_1^{\bar{d}/\bar{u}}$	(0.000)
b_2^g	(0.000)	$b_2^{u_v}$	(0.524)	$b_2^{d_v}$	(0.440)	$b_2^{\bar{d}+\bar{u}}$	(0.290)	$b_2^{s+\bar{s}}$	(0.290)	$b_2^{\bar{d}/\bar{u}}$	(0.000)
b_3^g	(0.520)	$b_3^{u_v}$	(0.073)	$b_3^{d_v}$	(0.107)	$b_3^{\bar{d}+\bar{u}}$	(0.298)	$b_3^{s+\bar{s}}$	(0.000)	$b_3^{\bar{d}/\bar{u}}$	(0.000)
b_4^g	-0.580	$b_4^{u_v}$	(0.038)	$b_4^{d_v}$	(-0.018)	$b_4^{\bar{d}+\bar{u}}$	(0.888)	$b_4^{s+\bar{s}}$	(0.000)	$b_4^{\bar{d}/\bar{u}}$	(0.000)
b_5^g	-0.891	$b_5^{u_v}$	(0.615)	$b_5^{d_v}$	(-0.236)	$b_5^{\bar{d}+\bar{u}}$	(1.353)	$b_5^{s+\bar{s}}$	(0.000)	$b_5^{\bar{d}/\bar{u}}$	(0.000)

Table IV.2.: Values of the PDF parameters of the nCTEQ15WZ+SIH fit at the initial scale $Q_0 = 1.3$ GeV. Free parameters are printed bold, while fixed parameters are in parentheses. Parameters without listed values are fixed through sum rules.

Par.	Value	Par.	Value	Par.	Value	Par.	Value	Par.	Value	Par.	Value
p_0^g	(0.000)	—	—	—	—	—	—	$p_0^{s+\bar{s}}$	(0.500)	—	—
p_1^g	(0.523)	$p_1^{u_v}$	(0.630)	$p_1^{d_v}$	(0.513)	$p_1^{\bar{d}+\bar{u}}$	(-0.324)	$p_1^{s+\bar{s}}$	(-0.324)	$p_1^{\bar{d}/\bar{u}}$	(10.075)
p_2^g	(3.034)	$p_2^{u_v}$	(2.934)	$p_2^{d_v}$	(4.211)	$p_2^{\bar{d}+\bar{u}}$	(8.116)	$p_2^{s+\bar{s}}$	(8.116)	$p_2^{\bar{d}/\bar{u}}$	(4.957)
p_3^g	(4.394)	$p_3^{u_v}$	(-2.369)	$p_3^{d_v}$	(-2.375)	$p_3^{\bar{d}+\bar{u}}$	(0.413)	$p_3^{s+\bar{s}}$	(0.413)	$p_3^{\bar{d}/\bar{u}}$	(15.167)
p_4^g	(2.359)	$p_4^{u_v}$	(1.266)	$p_4^{d_v}$	(0.965)	$p_4^{\bar{d}+\bar{u}}$	(4.754)	$p_4^{s+\bar{s}}$	(4.754)	$p_4^{\bar{d}/\bar{u}}$	(17.000)
p_5^g	(-3.000)	$p_5^{u_v}$	(1.718)	$p_5^{d_v}$	(3.000)	$p_5^{\bar{d}+\bar{u}}$	(0.614)	$p_5^{s+\bar{s}}$	(0.614)	$p_5^{\bar{d}/\bar{u}}$	(9.948)
Par.	Value	Par.	Value	Par.	Value	Par.	Value	Par.	Value	Par.	Value
a_0^g	(-0.256)	—	—	—	—	—	—	$a_0^{s+\bar{s}}$	0.281	—	—
a_1^g	-0.011	$a_1^{u_v}$	-3.325	$a_1^{d_v}$	-0.042	$a_1^{\bar{d}+\bar{u}}$	0.400	$a_1^{s+\bar{s}}$	-0.483	$a_1^{\bar{d}/\bar{u}}$	(0.000)
a_2^g	(0.000)	$a_2^{u_v}$	-0.003	$a_2^{d_v}$	-0.715	$a_2^{\bar{d}+\bar{u}}$	(0.415)	$a_2^{s+\bar{s}}$	5.383	$a_2^{\bar{d}/\bar{u}}$	(0.000)
a_3^g	(0.383)	$a_3^{u_v}$	(0.018)	$a_3^{d_v}$	(0.085)	$a_3^{\bar{d}+\bar{u}}$	(-0.759)	$a_3^{s+\bar{s}}$	(0.000)	$a_3^{\bar{d}/\bar{u}}$	(0.000)
a_4^g	0.032	$a_4^{u_v}$	18.050	$a_4^{d_v}$	(3.874)	$a_4^{\bar{d}+\bar{u}}$	(-0.203)	$a_4^{s+\bar{s}}$	(0.000)	$a_4^{\bar{d}/\bar{u}}$	(0.000)
a_5^g	0.008	$a_5^{u_v}$	-1.061	$a_5^{d_v}$	-0.155	$a_5^{\bar{d}+\bar{u}}$	0.055	$a_5^{s+\bar{s}}$	(0.000)	$a_5^{\bar{d}/\bar{u}}$	(0.000)
Par.	Value	Par.	Value	Par.	Value	Par.	Value	Par.	Value	Par.	Value
b_0^g	-0.025	—	—	—	—	—	—	$b_0^{s+\bar{s}}$	(0.104)	—	—
b_1^g	-0.680	$b_1^{u_v}$	(0.006)	$b_1^{d_v}$	(0.466)	$b_1^{\bar{d}+\bar{u}}$	(0.172)	$b_1^{s+\bar{s}}$	(0.172)	$b_1^{\bar{d}/\bar{u}}$	(0.000)
b_2^g	(0.000)	$b_2^{u_v}$	(0.524)	$b_2^{d_v}$	(0.440)	$b_2^{\bar{d}+\bar{u}}$	(0.290)	$b_2^{s+\bar{s}}$	(0.290)	$b_2^{\bar{d}/\bar{u}}$	(0.000)
b_3^g	(0.520)	$b_3^{u_v}$	(0.073)	$b_3^{d_v}$	(0.107)	$b_3^{\bar{d}+\bar{u}}$	(0.298)	$b_3^{s+\bar{s}}$	(0.000)	$b_3^{\bar{d}/\bar{u}}$	(0.000)
b_4^g	-0.544	$b_4^{u_v}$	(0.038)	$b_4^{d_v}$	(-0.018)	$b_4^{\bar{d}+\bar{u}}$	(0.888)	$b_4^{s+\bar{s}}$	(0.000)	$b_4^{\bar{d}/\bar{u}}$	(0.000)
b_5^g	-1.051	$b_5^{u_v}$	(0.615)	$b_5^{d_v}$	(-0.236)	$b_5^{\bar{d}+\bar{u}}$	(1.353)	$b_5^{s+\bar{s}}$	(0.000)	$b_5^{\bar{d}/\bar{u}}$	(0.000)

3. Additional tables

Table IV.3.: Values of the PDF parameters of the nCTEQ15HQ fit at the initial scale $Q_0 = 1.3$ GeV. Free parameters are printed bold, while fixed parameters are in parentheses. Parameters without listed values are fixed through sum rules.

Par.	Value	Par.	Value	Par.	Value	Par.	Value	Par.	Value	Par.	Value
p_0^g	(0.000)	—	—	—	—	—	—	$p_0^{s+\bar{s}}$	(0.500)	—	—
p_1^g	(0.523)	$p_1^{u_v}$	(0.630)	$p_1^{d_v}$	(0.513)	$p_1^{\bar{d}+\bar{u}}$	(-0.324)	$p_1^{s+\bar{s}}$	(-0.324)	$p_1^{\bar{d}/\bar{u}}$	(10.075)
p_2^g	(3.034)	$p_2^{u_v}$	(2.934)	$p_2^{d_v}$	(4.211)	$p_2^{\bar{d}+\bar{u}}$	(8.116)	$p_2^{s+\bar{s}}$	(8.116)	$p_2^{\bar{d}/\bar{u}}$	(4.957)
p_3^g	(4.394)	$p_3^{u_v}$	(-2.369)	$p_3^{d_v}$	(-2.375)	$p_3^{\bar{d}+\bar{u}}$	(0.413)	$p_3^{s+\bar{s}}$	(0.413)	$p_3^{\bar{d}/\bar{u}}$	(15.167)
p_4^g	(2.359)	$p_4^{u_v}$	(1.266)	$p_4^{d_v}$	(0.965)	$p_4^{\bar{d}+\bar{u}}$	(4.754)	$p_4^{s+\bar{s}}$	(4.754)	$p_4^{\bar{d}/\bar{u}}$	(17.000)
p_5^g	(-3.000)	$p_5^{u_v}$	(1.718)	$p_5^{d_v}$	(3.000)	$p_5^{\bar{d}+\bar{u}}$	(0.614)	$p_5^{s+\bar{s}}$	(0.614)	$p_5^{\bar{d}/\bar{u}}$	(9.948)
Par.	Value	Par.	Value	Par.	Value	Par.	Value	Par.	Value	Par.	Value
a_0^g	(-0.256)	—	—	—	—	—	—	$a_0^{s+\bar{s}}$	0.237	—	—
a_1^g	-0.066	$a_1^{u_v}$	-0.958	$a_1^{d_v}$	-0.139	$a_1^{\bar{d}+\bar{u}}$	0.395	$a_1^{s+\bar{s}}$	0.612	$a_1^{\bar{d}/\bar{u}}$	(0.000)
a_2^g	(0.000)	$a_2^{u_v}$	-0.002	$a_2^{d_v}$	-0.858	$a_2^{\bar{d}+\bar{u}}$	(0.415)	$a_2^{s+\bar{s}}$	12.236	$a_2^{\bar{d}/\bar{u}}$	(0.000)
a_3^g	(0.383)	$a_3^{u_v}$	(0.018)	$a_3^{d_v}$	(0.085)	$a_3^{\bar{d}+\bar{u}}$	(-0.759)	$a_3^{s+\bar{s}}$	(0.000)	$a_3^{\bar{d}/\bar{u}}$	(0.000)
a_4^g	0.131	$a_4^{u_v}$	18.977	$a_4^{d_v}$	(3.874)	$a_4^{\bar{d}+\bar{u}}$	(-0.203)	$a_4^{s+\bar{s}}$	(0.000)	$a_4^{\bar{d}/\bar{u}}$	(0.000)
a_5^g	0.030	$a_5^{u_v}$	-1.081	$a_5^{d_v}$	-0.045	$a_5^{\bar{d}+\bar{u}}$	-0.047	$a_5^{s+\bar{s}}$	(0.000)	$a_5^{\bar{d}/\bar{u}}$	(0.000)
Par.	Value	Par.	Value	Par.	Value	Par.	Value	Par.	Value	Par.	Value
b_0^g	-0.029	—	—	—	—	—	—	$b_0^{s+\bar{s}}$	(0.104)	—	—
b_1^g	-0.447	$b_1^{u_v}$	(0.006)	$b_1^{d_v}$	(0.466)	$b_1^{\bar{d}+\bar{u}}$	(0.172)	$b_1^{s+\bar{s}}$	(0.172)	$b_1^{\bar{d}/\bar{u}}$	(0.000)
b_2^g	(0.000)	$b_2^{u_v}$	(0.524)	$b_2^{d_v}$	(0.440)	$b_2^{\bar{d}+\bar{u}}$	(0.290)	$b_2^{s+\bar{s}}$	(0.290)	$b_2^{\bar{d}/\bar{u}}$	(0.000)
b_3^g	(0.520)	$b_3^{u_v}$	(0.073)	$b_3^{d_v}$	(0.107)	$b_3^{\bar{d}+\bar{u}}$	(0.298)	$b_3^{s+\bar{s}}$	(0.000)	$b_3^{\bar{d}/\bar{u}}$	(0.000)
b_4^g	0.069	$b_4^{u_v}$	(0.038)	$b_4^{d_v}$	(-0.018)	$b_4^{\bar{d}+\bar{u}}$	(0.888)	$b_4^{s+\bar{s}}$	(0.000)	$b_4^{\bar{d}/\bar{u}}$	(0.000)
b_5^g	-0.722	$b_5^{u_v}$	(0.615)	$b_5^{d_v}$	(-0.236)	$b_5^{\bar{d}+\bar{u}}$	(1.353)	$b_5^{s+\bar{s}}$	(0.000)	$b_5^{\bar{d}/\bar{u}}$	(0.000)

Table IV.4.: Overview of the available data sets for F_2^D or ratios thereof.

Experiment	Ref.	ID	Observable	Points after / before cuts
NMC-97	[285]	5160	F_2^D	201 / 292
Hermes	[286]	5156	F_2^{He}/F_2^D	17 / 182
NMC-95,re	[287]	5124	F_2^{He}/F_2^D	12 / 18
SLAC-E139	[76]	5141	F_2^{He}/F_2^D	3 / 18
NMC-95	[288]	5115	F_2^{Li}/F_2^D	11 / 24
SLAC-E139	[76]	5138	F_2^{Be}/F_2^D	3 / 17
FNAL-E665-95	[289]	5125	F_2^C/F_2^D	3 / 11
SLAC-E139	[76]	5139	F_2^C/F_2^D	2 / 7
EMC-88	[290]	5107	F_2^C/F_2^D	9 / 9
EMC-90	[291]	5110	F_2^C/F_2^D	0 / 9
NMC-95	[288]	5113	F_2^C/F_2^D	12 / 24
NMC-95,re	[287]	5114	F_2^C/F_2^D	12 / 18
Hermes	[286]	5157	F_2^N/F_2^D	19 / 175
BCDMS-85	[77]	5103	F_2^N/F_2^D	9 / 9
SLAC-E049	[79]	5134	F_2^{Al}/F_2^D	0 / 18
SLAC-E139	[76]	5136	F_2^{Al}/F_2^D	3 / 17
NMC-95,re	[287]	5121	F_2^{Ca}/F_2^D	12 / 18
FNAL-E665-95	[289]	5126	F_2^{Ca}/F_2^D	3 / 11
SLAC-E139	[76]	5140	F_2^{Ca}/F_2^D	2 / 7
EMC-90	[291]	5109	F_2^{Ca}/F_2^D	0 / 9
SLAC-E049	[292]	5131	F_2^{Fe}/F_2^D	2 / 14
SLAC-E139	[76]	5132	F_2^{Fe}/F_2^D	6 / 23
SLAC-E140	[80]	5133	F_2^{Fe}/F_2^D	0 / 10
BCDMS-87	[293]	5101	F_2^{Fe}/F_2^D	10 / 10
BCDMS-85	[77]	5102	F_2^{Fe}/F_2^D	6 / 6
EMC-93	[294]	5104	F_2^{Cu}/F_2^D	9 / 10
EMC-93(chariot)	[294]	5105	F_2^{Cu}/F_2^D	9 / 9
EMC-88	[290]	5106	F_2^{Cu}/F_2^D	9 / 9
Hermes	[286]	5158	F_2^{Kr}/F_2^D	12 / 167
SLAC-E139	[76]	5135	F_2^{Ag}/F_2^D	2 / 7
EMC-88	[290]	5108	F_2^{Sn}/F_2^D	8 / 8
FNAL-E665-92	[295]	5127	F_2^{Xe}/F_2^D	2 / 10
SLAC-E139	[76]	5137	F_2^{Au}/F_2^D	3 / 18
FNAL-E665-95	[289]	5129	F_2^{Pb}/F_2^D	3 / 11

3. Additional tables

Table IV.5.: Overview of the available data sets for $F_2^A/F_2^{A'}$.

Experiment	Ref.	ID	Observable	Points after / before cuts
NMC-95,re	[287]	5123	F_2^C/F_2^{Li}	7 / 25
NMC-95,re	[287]	5122	F_2^{Ca}/F_2^{Li}	7 / 25
NMC-96	[296]	5112	F_2^{Be}/F_2^C	14 / 15
NMC-96	[296]	5111	F_2^{Al}/F_2^C	14 / 15
NMC-95,re	[287]	5120	F_2^{Ca}/F_2^C	7 / 25
NMC-96	[296]	5119	F_2^{Ca}/F_2^C	14 / 15
NMC-96	[296]	5143	F_2^{Fe}/F_2^C	14 / 15
NMC-96	[297]	5159	F_2^{Sn}/F_2^C	111 / 146
NMC-96	[296]	5116	F_2^{Pb}/F_2^C	14 / 15

Table IV.6.: Overview of the available Drell-Yan process data sets.

Experiment	Ref.	ID	Observable	Points after / before cuts
FNAL-E772-90	[104]	5203	σ^{pC}/σ^{pH2}	9 / 9
FNAL-E772-90	[104]	5204	$\sigma^{pCa}/\sigma^{pH2}$	9 / 9
FNAL-E772-90	[104]	5205	$\sigma^{pFe}/\sigma^{pH2}$	9 / 9
FNAL-E772-90	[104]	5206	σ^{pW}/σ^{pH2}	9 / 9
FNAL-E886-99	[298]	5201	$\sigma^{pFe}/\sigma^{pBe}$	28 / 28
FNAL-E886-99	[298]	5202	σ^{pW}/σ^{pBe}	28 / 28

Table IV.7.: Overview of the available W^\pm and Z boson production data sets. The measured observable is always $\frac{d\sigma}{dy}$ and no cuts are applied.

Experiment	Ref.	ID	Particle	Nb. of Points
ATLAS	[299]	6213	W^+	10
CMS	[300]	6233	W^+	10
CMS (Run II)	[301]	6234	W^+	24
ALICE	[302, 303]	6253	W^+	2
ATLAS	[299]	6211	W^-	10
CMS	[300]	6231	W^-	10
CMS (Run II)	[301]	6232	W^-	24
ALICE	[302, 303]	6251	W^-	2
ATLAS	[304]	6215	Z	14
CMS	[305]	6235	Z	12
LHCb	[306]	6275	Z	2

4. Additional figures

This appendix contains a variety of figures that are not shown directly in the main text to improve readability. The main text contains notes whenever a figure from this section is relevant and each figure in this section has a link to the relevant section in it's caption.

4.1. Single inclusive hadron production

Figs. IV.1 - IV.3 show the uncertainties of NNFF, JAM20 and HKNS fragmentation functions compared to the data sets that can be predicted using these FFs. NNFF and JAM20 both use Monte Carlo replicas, but yield very different results. The uncertainties for NNFF are smaller, but on the same order of magnitude as the data uncertainties, while JAM20 has extremely narrow error bands. Since the two fits use similar data sets, the difference must come from the parameterizations used. As the name implies, NNFF uses a neural network with many more degrees of freedom than the "traditional" parameterization of JAM20. Contrary to these two FFs, HKNS uses Hessian eigenvectors and yields uncertainties significantly larger than those of the data sets we can compare it with. The reason for these large uncertainties is the lack of high-precision data at the time when the HKNS fit was performed.

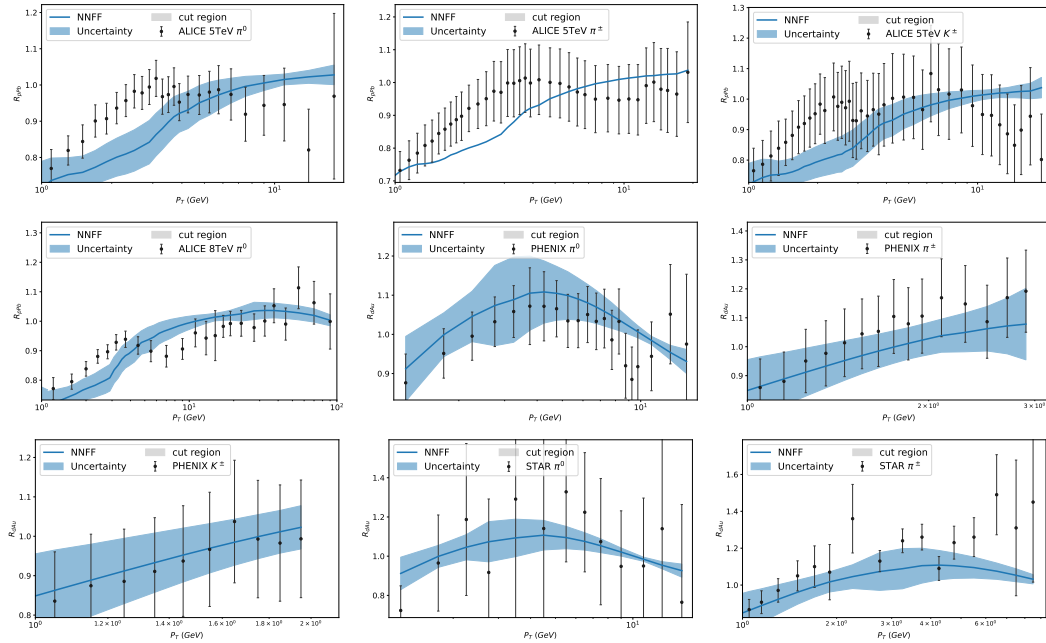


Figure IV.1.: Uncertainties calculated from the NNFF replicas using nCTEQ15WZ PDFs. The computed uncertainties use 83 of the 101 provided replicas because the remaining 18 lead to unphysical behavior such as large jumps from one p_T value to another or negative cross sections due to numerical problems. Figure referenced in Sec. II.2.3.3

4. Additional figures

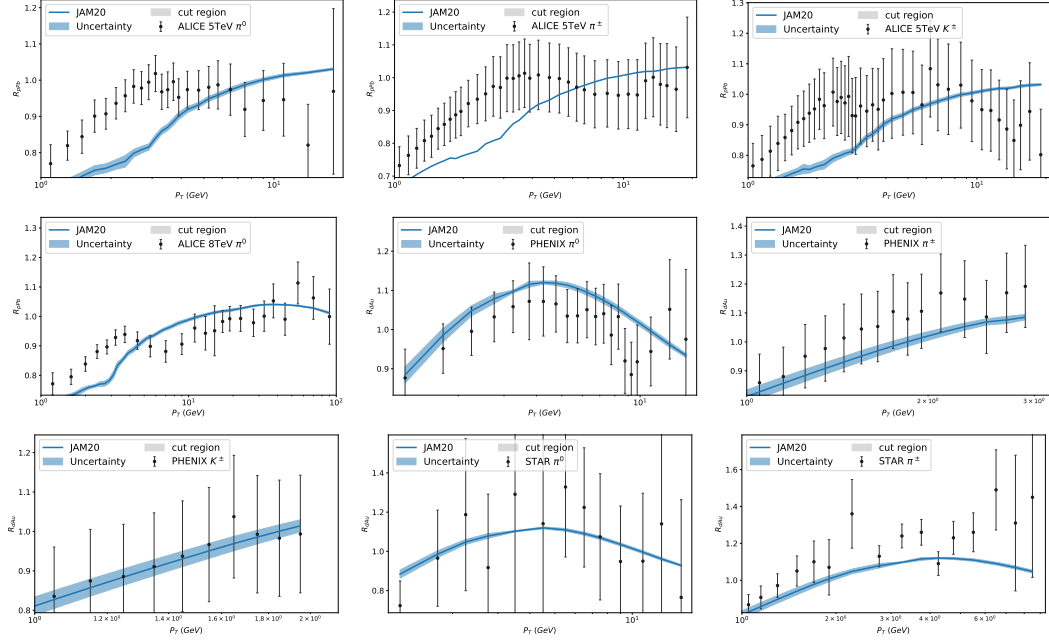


Figure IV.2.: Uncertainties calculated from the JAM20 replicas using nCTEQ15WZ PDFs.
Figure referenced in Sec. II.2.3.3

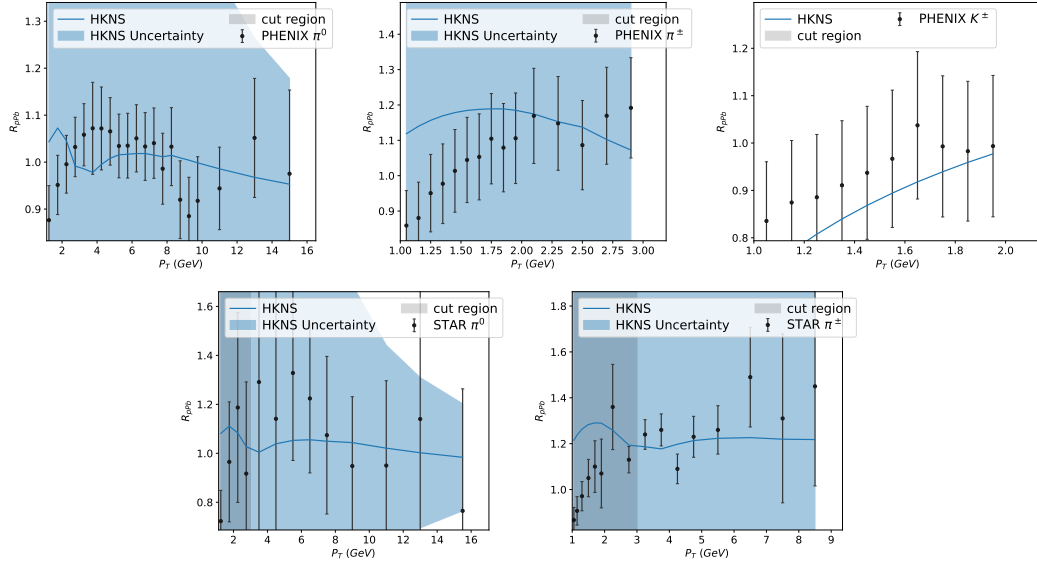


Figure IV.3.: Uncertainties calculated from the HKNS eigenvectors using nCTEQ15WZ PDFs. Note that there are no uncertainties provided for kaon production.
Figure referenced in Sec. II.2.3.3

4.2. Heavy quark production

This section contains comparisons between data and theory prediction for the heavy quark production. Figs. IV.4 - IV.10 show the data from proton-proton collisions with predictions

from the Crystal Ball fit using nCTEQ15 proton PDFs. The theoretical uncertainties shown come from the Hessian eigenvectors of the Crystal Ball fit. Figs. IV.11 - IV.15 show the data from proton-lead collisions with predictions from the Crystal Ball fit using nCTEQ15HQ lead PDFs. The theoretical uncertainties in this case are just the PDF uncertainties without the Crystal Ball uncertainty.

4.2.1. Heavy quark production in proton-proton collisions

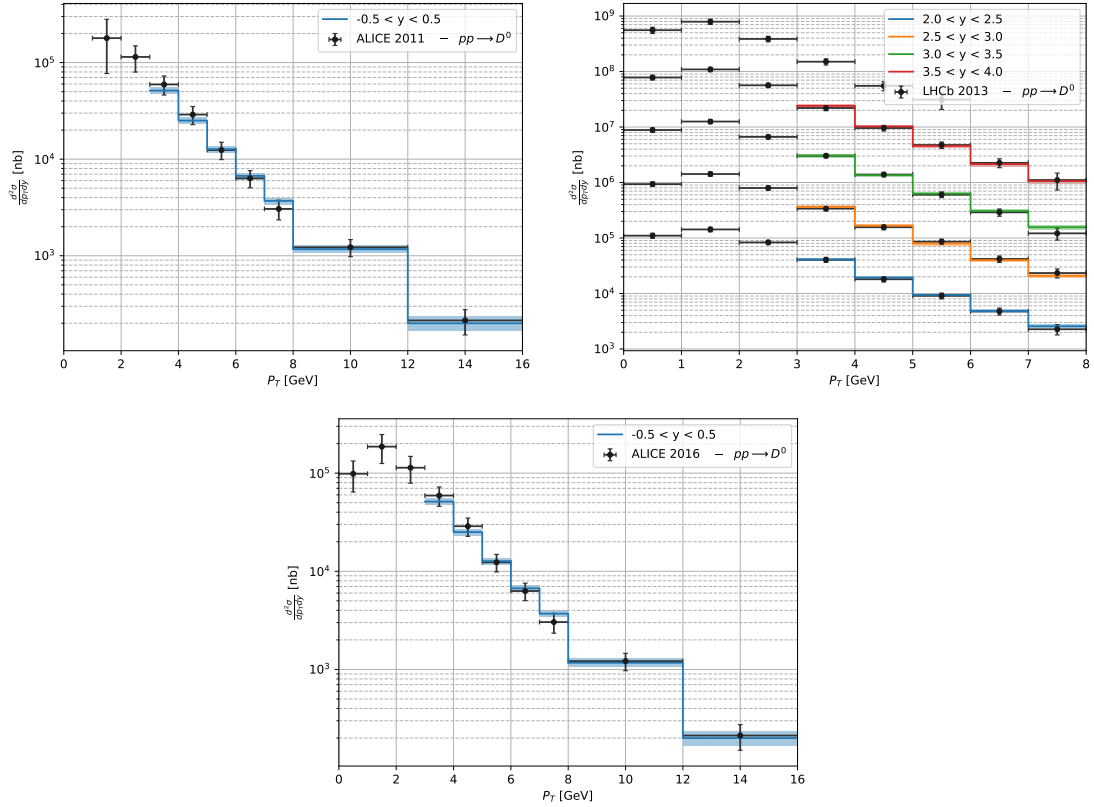


Figure IV.4.: Predictions for D^0 production in proton-proton collisions with uncertainties from the Crystal Ball fit. Different rapidity bins are shifted on the y-axis by powers of ten for visual clarity. Figure referenced in Sec. II.3.2.

4. Additional figures

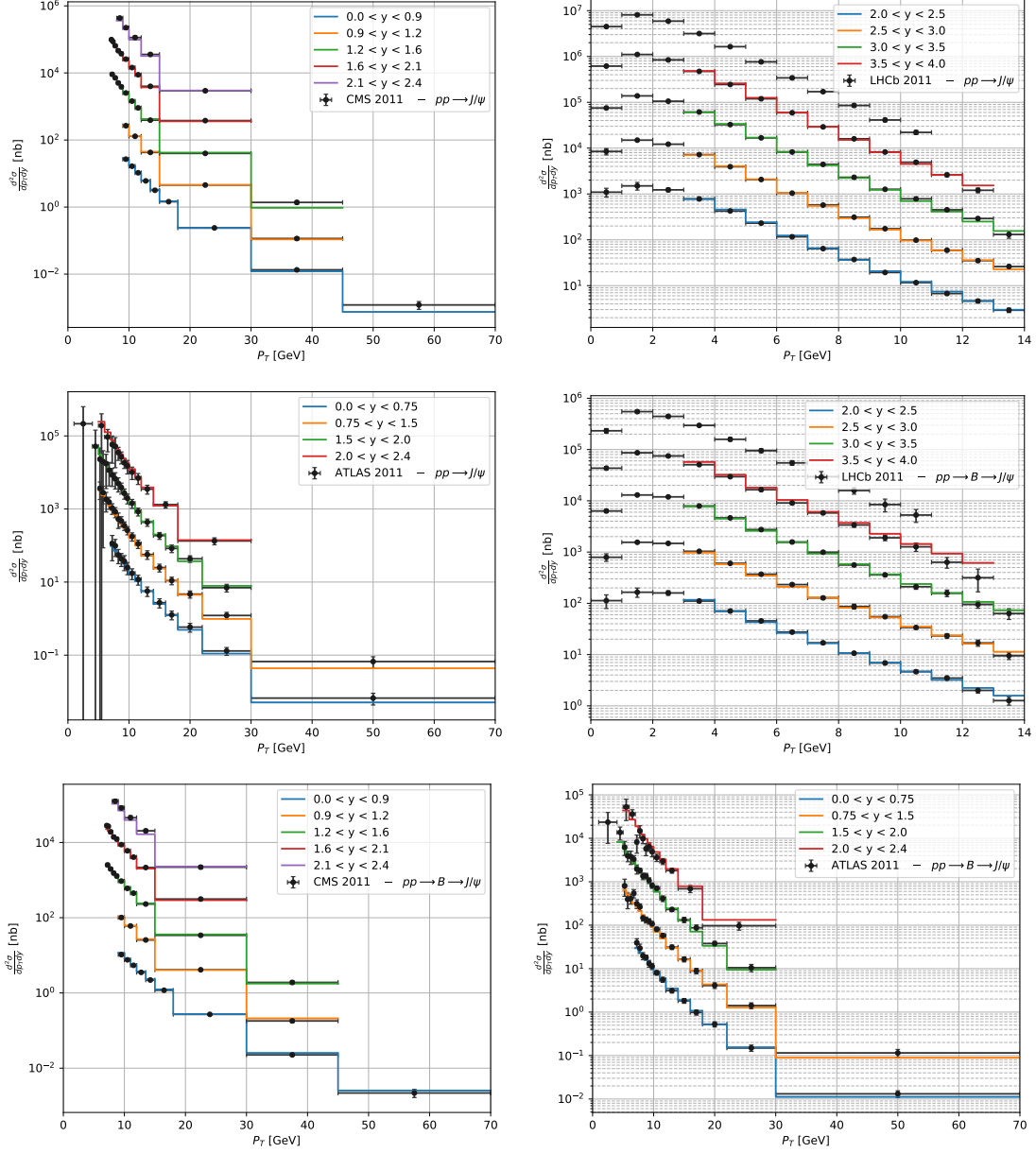


Figure IV.5.: Predictions for J/ψ production in proton-proton collisions with uncertainties from the Crystal Ball fit. Different rapidity bins are shifted on the y-axis by powers of ten for visual clarity. Figure referenced in Sec. II.3.2.

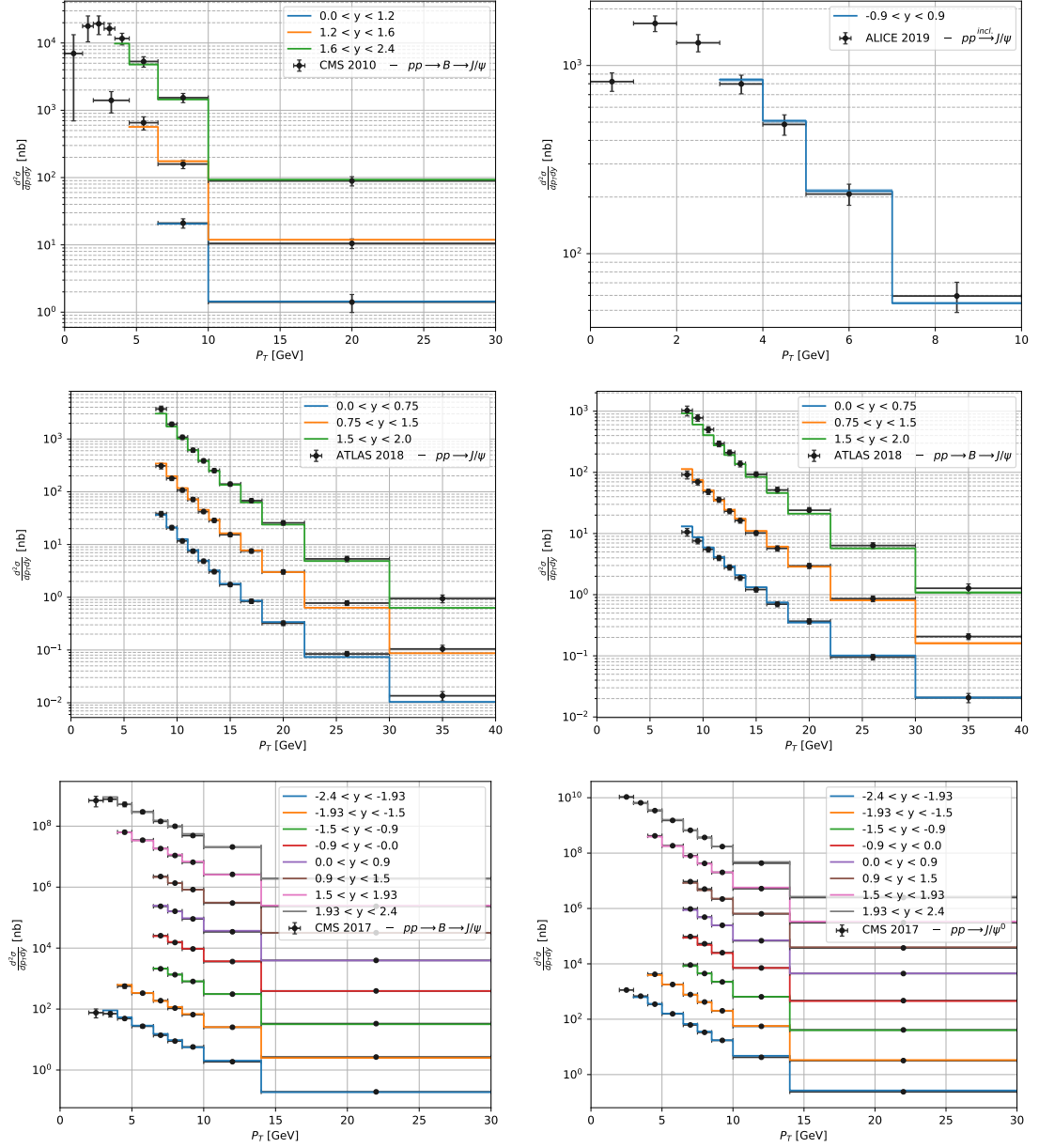


Figure IV.6.: Predictions for J/ψ production in proton-proton collisions with uncertainties from the Crystal Ball fit. Different rapidity bins are shifted on the y-axis by powers of ten for visual clarity. Figure referenced in Sec. II.3.2.

4. Additional figures

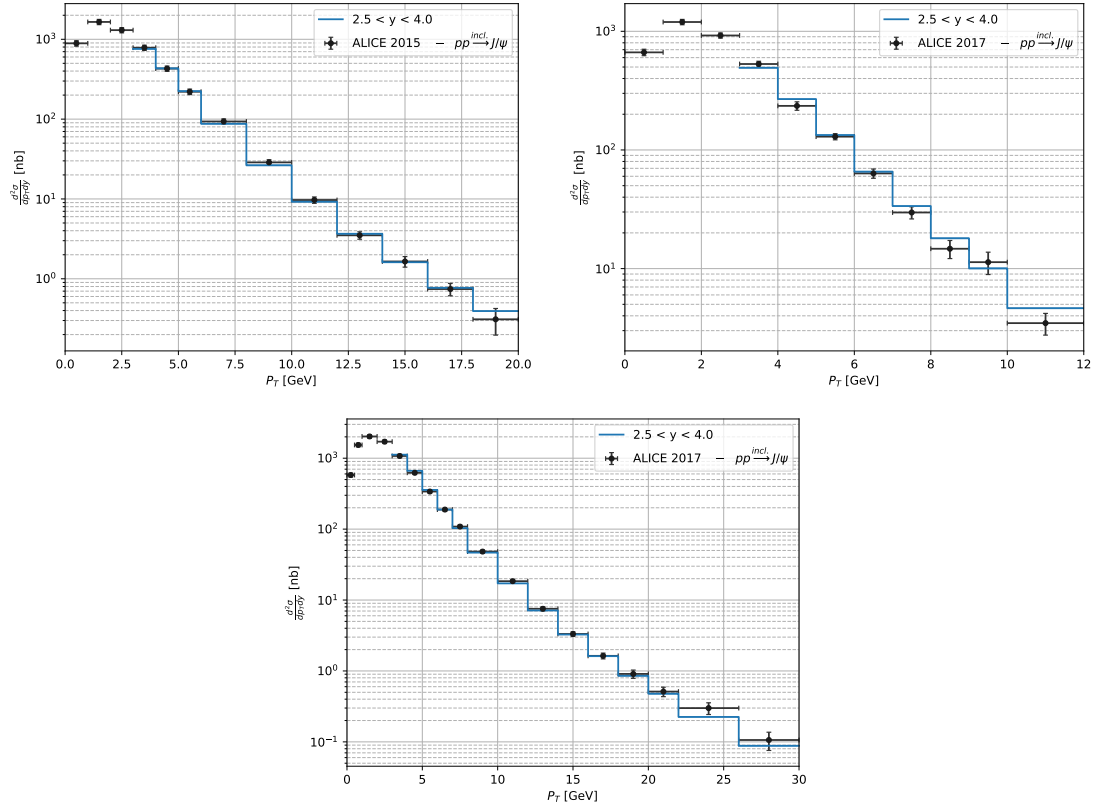


Figure IV.7.: Predictions for J/ψ production in proton-proton collisions with uncertainties from the Crystal Ball fit. Different rapidity bins are shifted on the y-axis by powers of ten for visual clarity. Figure referenced in Sec. II.3.2.

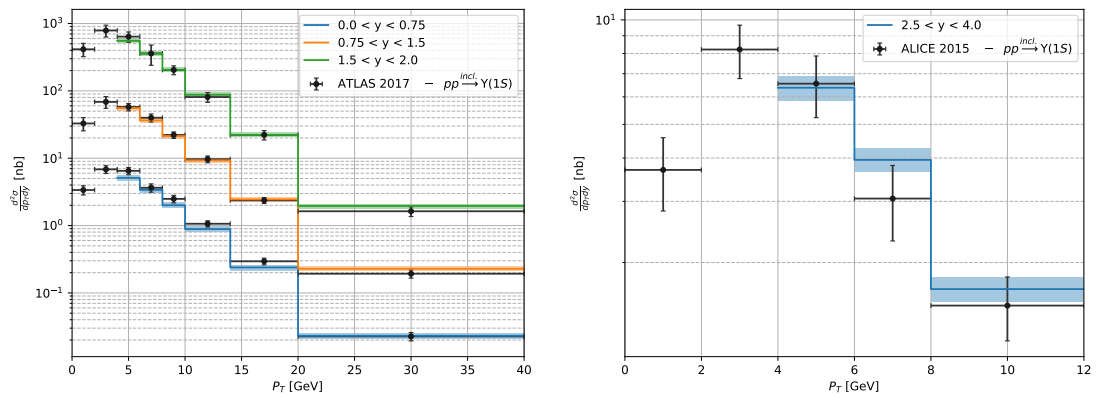


Figure IV.8.: Predictions for $\Upsilon(1S)$ production in proton-proton collisions with uncertainties from the Crystal Ball fit. Different rapidity bins are shifted on the y-axis by powers of ten for visual clarity. Figure referenced in Sec. II.3.2.

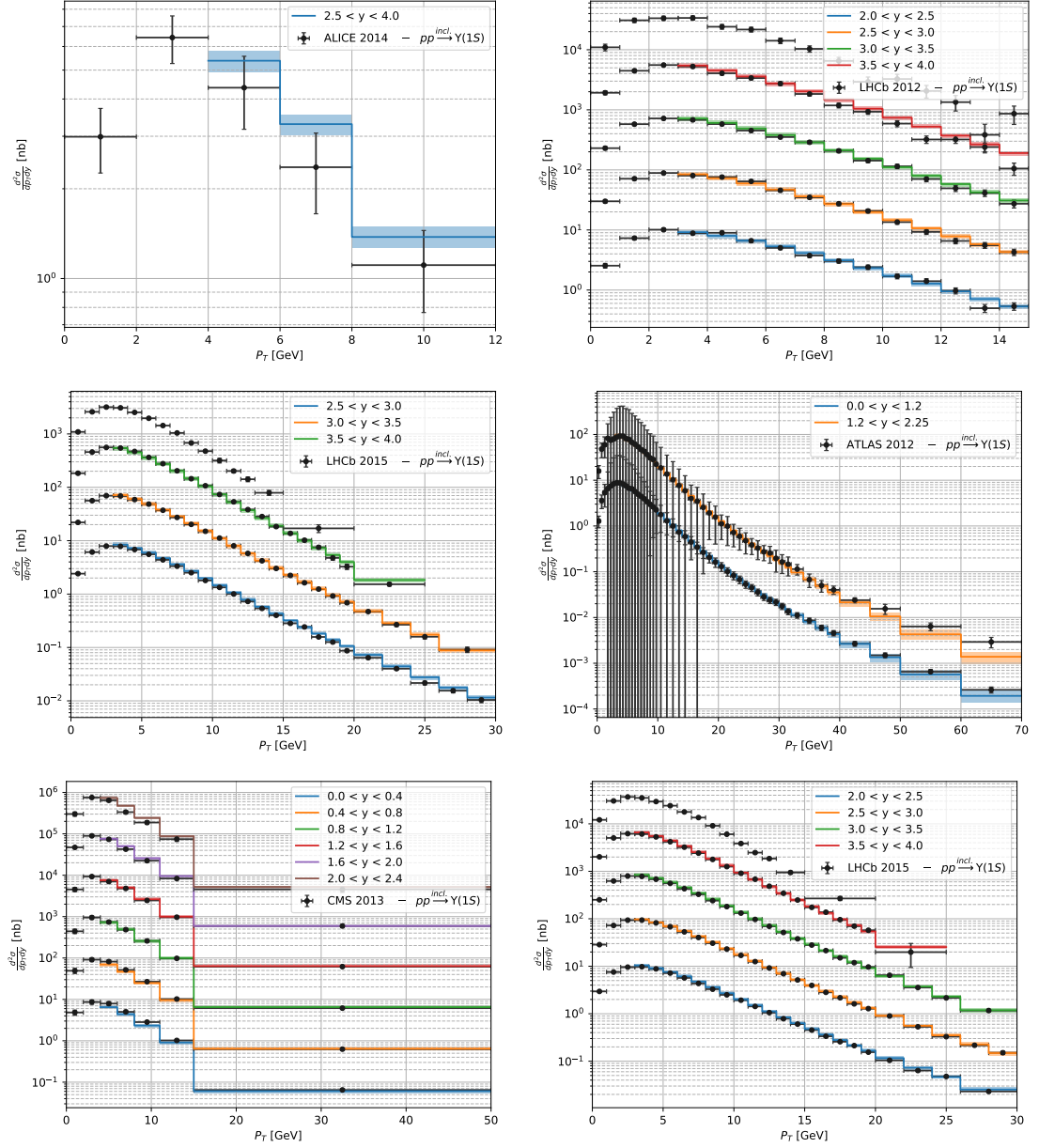


Figure IV.9.: Predictions for $\Upsilon(1S)$ production in proton-proton collisions with uncertainties from the Crystal Ball fit. Different rapidity bins are shifted on the y-axis by powers of ten for visual clarity. Figure referenced in Sec. II.3.2.

4. Additional figures

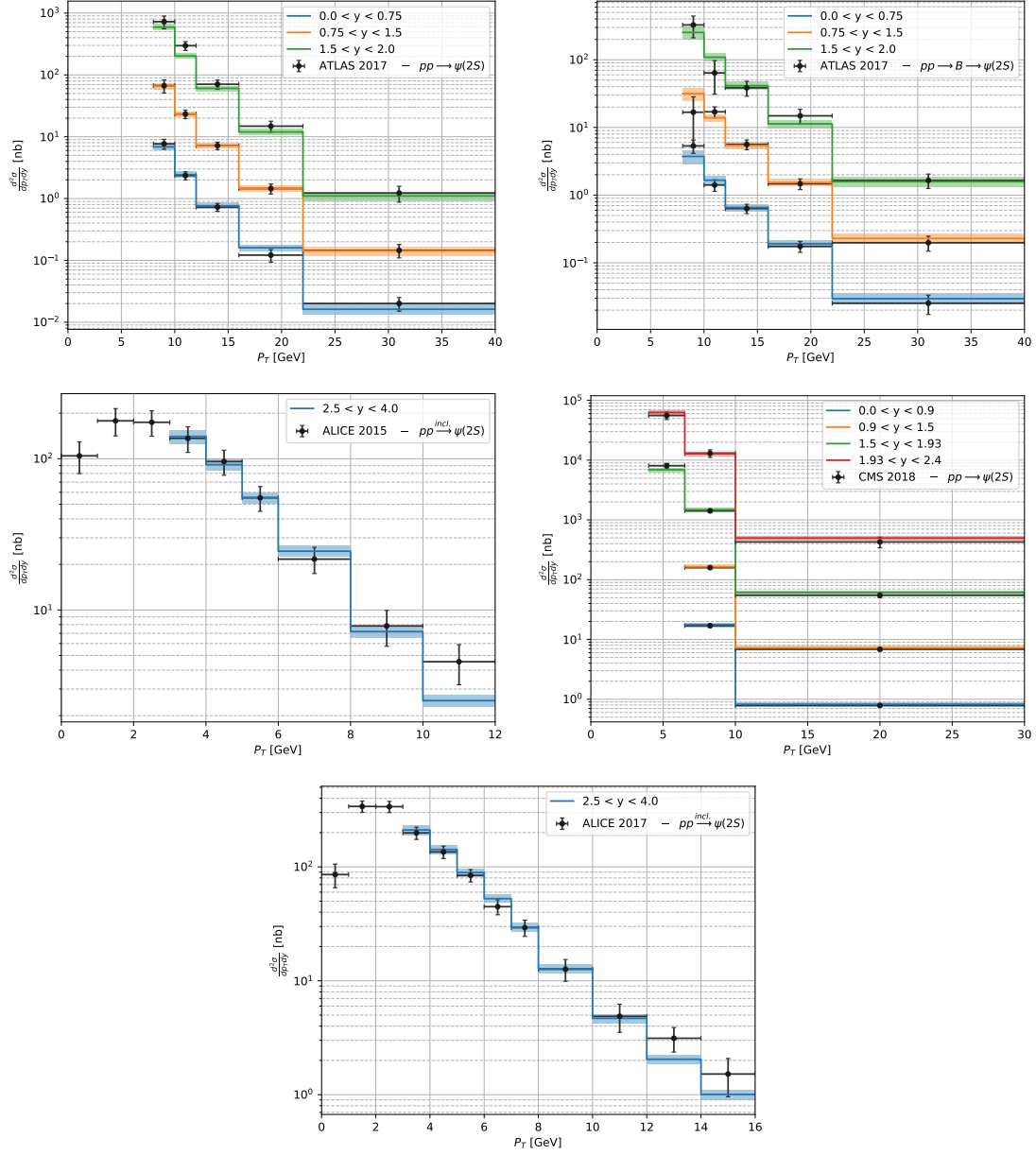


Figure IV.10.: Predictions for $\psi(2S)$ production in proton-proton collisions with uncertainties from the Crystal Ball fit. Different rapidity bins are shifted on the y-axis by powers of ten for visual clarity. Figure referenced in Sec. II.3.2.

4.2.2. Heavy quark production in proton-lead collisions

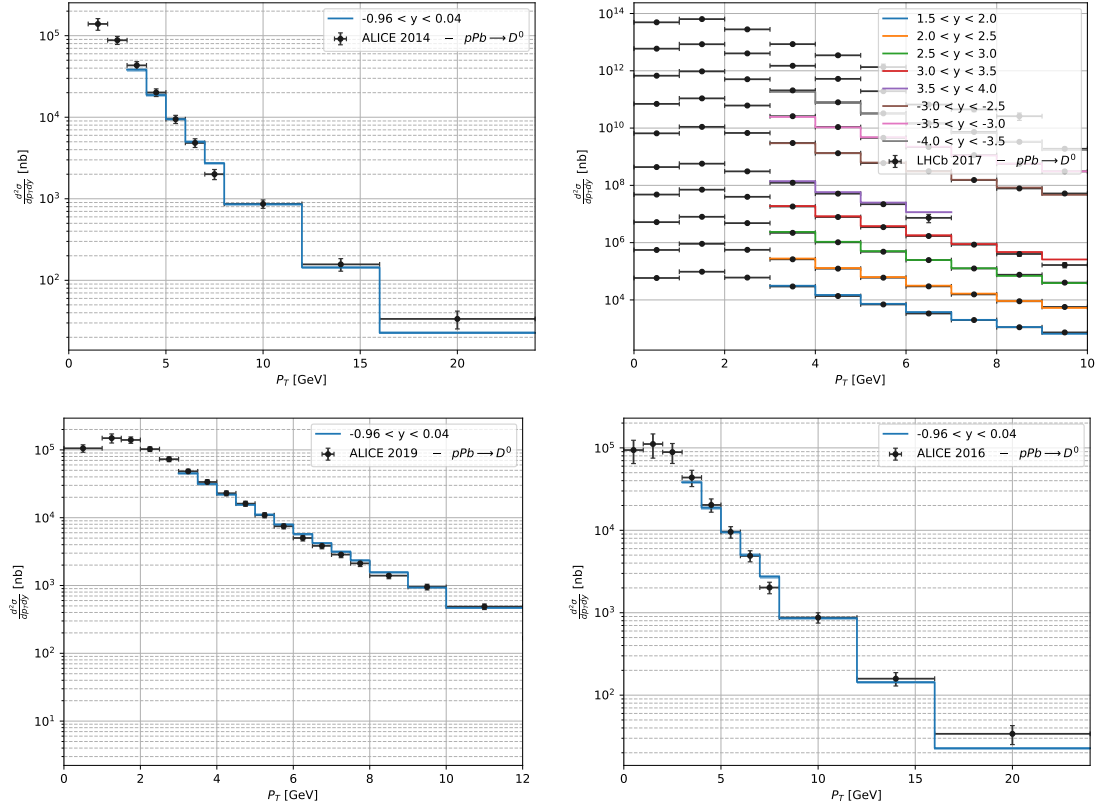


Figure IV.11.: Predictions for D^0 production in proton-lead collisions with PDF uncertainties of the nCTEQ15HQ fit. Different rapidity bins are shifted on the y-axis by powers of ten for visual clarity. Figure referenced in Sec. II.3.3.3.

4. Additional figures

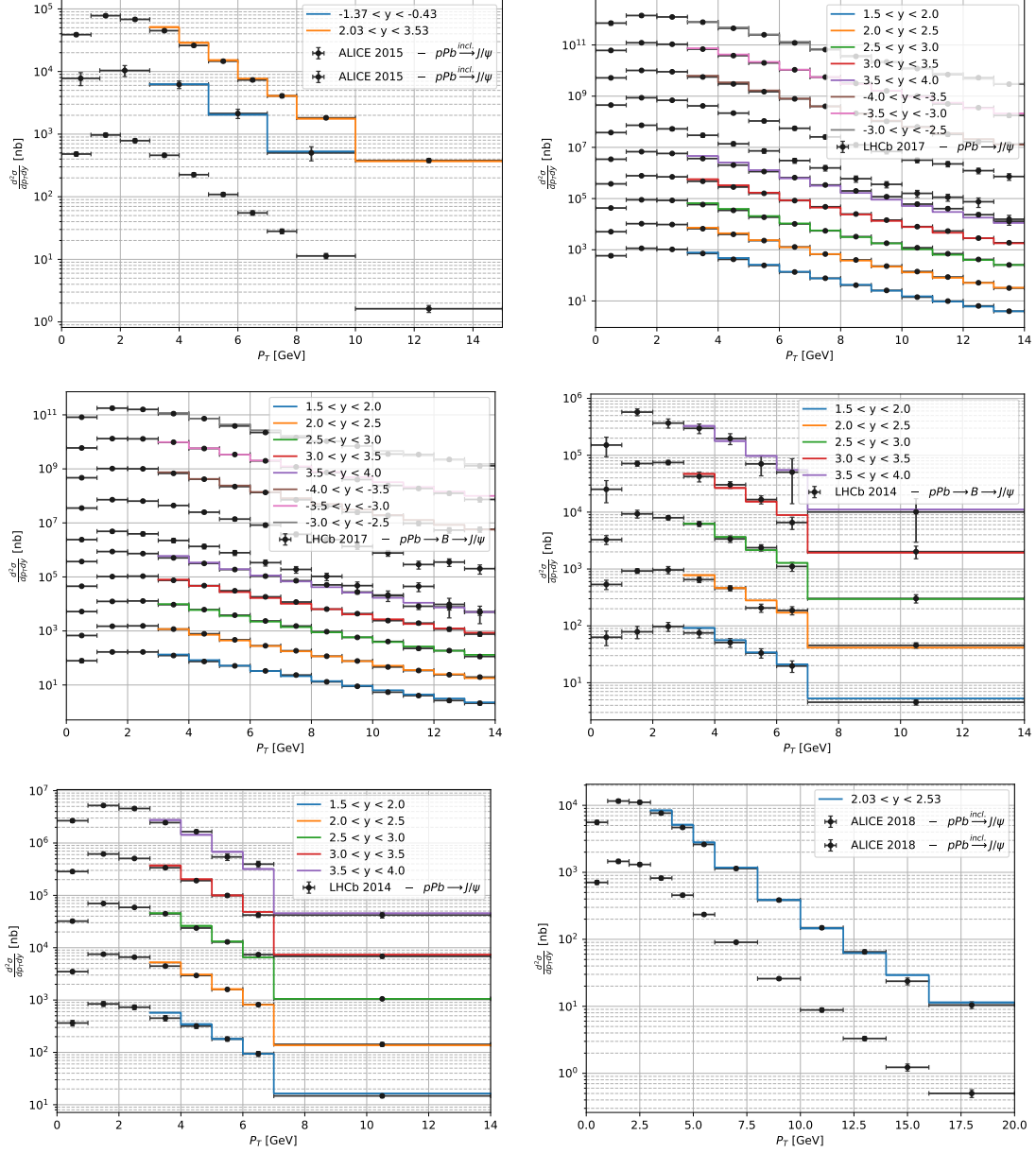


Figure IV.12.: Predictions for J/ψ production in proton-lead collisions with PDF uncertainties of the nCTEQ15HQ fit. Different rapidity bins are shifted on the y-axis by powers of ten for visual clarity. Figure referenced in Sec. II.3.3.3.

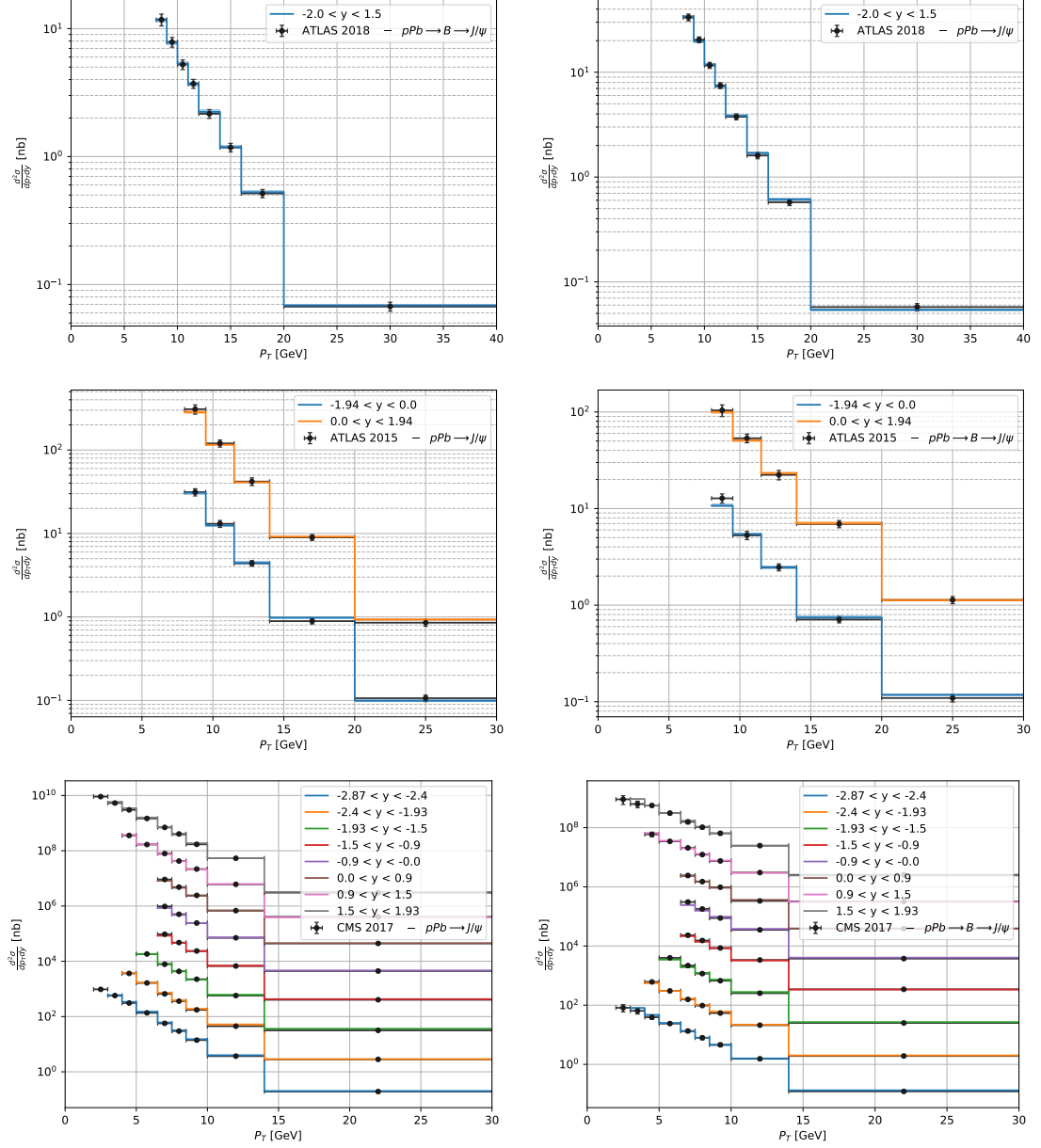


Figure IV.13.: Predictions for J/ψ production in proton-lead collisions with PDF uncertainties of the nCTEQ15HQ fit. Different rapidity bins are shifted on the y-axis by powers of ten for visual clarity. Figure referenced in Sec. II.3.3.3.

4. Additional figures

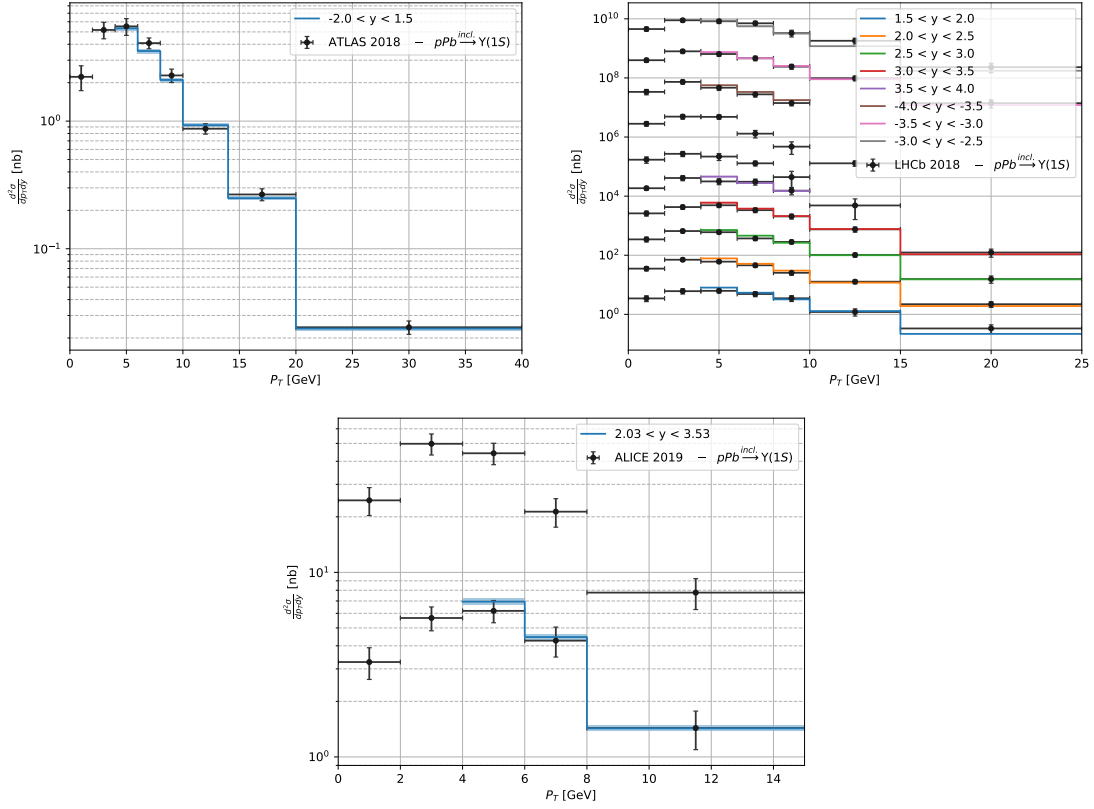


Figure IV.14.: Predictions for $\Upsilon(1S)$ production in proton-lead collisions with PDF uncertainties of the nCTEQ15HQ fit. Different rapidity bins are shifted on the y-axis by powers of ten for visual clarity. Figure referenced in Sec. II.3.3.3.

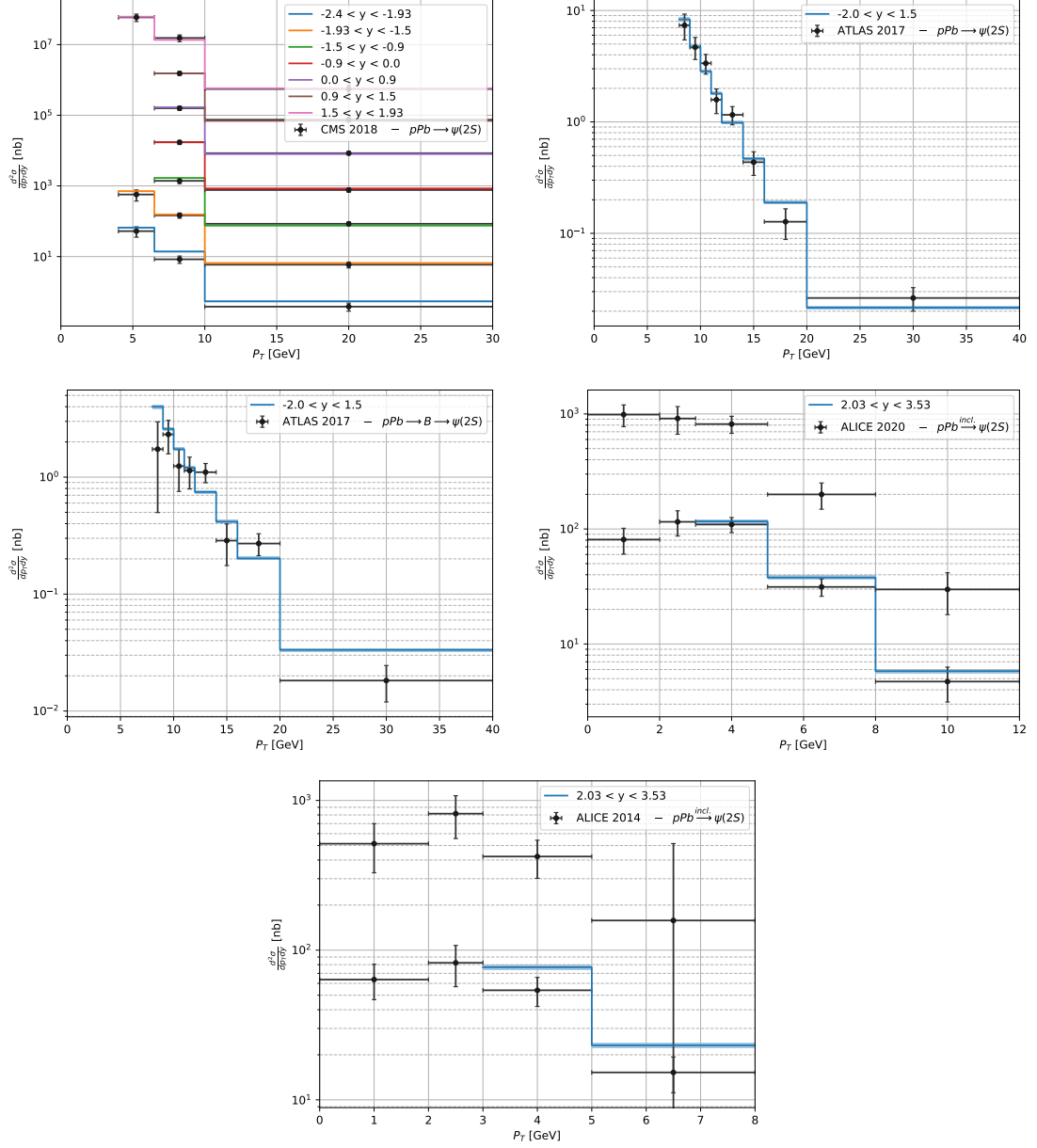


Figure IV.15.: Predictions for $\psi(2S)$ production in proton-lead collisions with PDF uncertainties of the nCTEQ15HQ fit. Different rapidity bins are shifted on the y-axis by powers of ten for visual clarity. Figure referenced in Sec. II.3.3.3.

Bibliography

- [1] E. Rutherford. “The scattering of alpha and beta particles by matter and the structure of the atom”. In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 21 (1911). DOI: 10.1080/14786440508637080.
- [2] F. Abe et al. “Observation of top quark production in $\bar{p}p$ collisions”. In: *Phys. Rev. Lett.* 74 (1995), pp. 2626–2631. DOI: 10.1103/PhysRevLett.74.2626. arXiv: hep-ex/9503002.
- [3] S. Abachi et al. “Observation of the top quark”. In: *Phys. Rev. Lett.* 74 (1995), pp. 2632–2637. DOI: 10.1103/PhysRevLett.74.2632. arXiv: hep-ex/9503003.
- [4] S. Chatrchyan et al. “Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC”. In: *Phys. Lett. B* 716 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex].
- [5] G. Aad et al. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. In: *Phys. Lett. B* 716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].
- [6] A. Accardi et al. “Electron Ion Collider: The Next QCD Frontier: Understanding the glue that binds us all”. 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. arXiv: 1212.1701 [nucl-ex].
- [7] E. C. Aschenauer et al. “The electron-ion collider: assessing the energy dependence of key measurements”. In: *Rept. Prog. Phys.* 82.2 (2019), p. 024301. DOI: 10.1088/1361-6633/aaf216. arXiv: 1708.01527 [nucl-ex].
- [8] Jorge G. Morfin and Wu-Ki Tung. “Parton distributions from a global QCD analysis of deep inelastic scattering and lepton pair production”. In: *Z. Phys. C* 52 (1991), pp. 13–30. DOI: 10.1007/BF01412323.
- [9] K. Kovarik et al. “nCTEQ15 - Global analysis of nuclear parton distributions with uncertainties in the CTEQ framework”. In: *Phys. Rev. D* 93.8 (2016), p. 085037. DOI: 10.1103/PhysRevD.93.085037. arXiv: 1509.00792 [hep-ph].
- [10] K. J. Eskola et al. “EPPS16: Nuclear parton distributions with LHC data”. In: *Eur. Phys. J. C* 77.3 (2017), p. 163. DOI: 10.1140/epjc/s10052-017-4725-9. arXiv: 1612.05741 [hep-ph].

-
- [11] R. Abdul Khalek, J. J. Ethier, and J. Rojo. “Nuclear parton distributions from lepton-nucleus scattering and the impact of an electron-ion collider”. In: *Eur. Phys. J. C* 79.6 (2019), p. 471. DOI: 10.1140/epjc/s10052-019-6983-1. arXiv: 1904.00018 [hep-ph].
- [12] R. Abdul Khalek et al. “nNNPDF2.0: Quark Flavor Separation in Nuclei from LHC Data”. In: (2020). arXiv: 2006.14629 [hep-ph].
- [13] Néstor Armesto et al. “An analysis of the impact of LHC Run I proton–lead data on nuclear parton densities”. In: *Eur. Phys. J. C* 76.4 (2016), p. 218. DOI: 10.1140/epjc/s10052-016-4078-9. arXiv: 1512.01528 [hep-ph].
- [14] Kari J. Eskola et al. “EPPS21: A global QCD analysis of nuclear PDFs”. In: (2021). arXiv: 2112.12462 [hep-ph].
- [15] J. J. Ethier and E. R. Nocera. “Parton Distributions in Nucleons and Nuclei”. In: *Ann. Rev. Nucl. Part. Sci.* 70 (2020), pp. 1–34. DOI: 10.1146/annurev-nucl-011720-042725. arXiv: 2001.07722 [hep-ph].
- [16] K. Kovarik, P. M. Nadolsky, and D. E. Soper. “Hadron structure in high-energy collisions”. In: *Rev.Mod.Phys.* 92 (2020), p. 045003. arXiv: 1905.06957 [hep-ph].
- [17] A. Accardi et al. “Constraints on large- x parton distributions from new weak boson production and deep-inelastic scattering data”. In: *Phys. Rev. D* 93.11 (2016), p. 114017. DOI: 10.1103/PhysRevD.93.114017. arXiv: 1602.03154 [hep-ph].
- [18] Tie-Jiun Hou et al. “New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC”. In: *Phys. Rev. D* 103.1 (2021), p. 014013. DOI: 10.1103/PhysRevD.103.014013. arXiv: 1912.10053 [hep-ph].
- [19] R. D. Ball et al. “Parton distributions from high-precision collider data”. In: *Eur. Phys. J. C* 77.10 (2017), p. 663. DOI: 10.1140/epjc/s10052-017-5199-5. arXiv: 1706.00428 [hep-ph].
- [20] R. Abdul Khalek et al. “Towards Ultimate Parton Distributions at the High-Luminosity LHC”. In: *Eur. Phys. J. C* 78.11 (2018), p. 962. DOI: 10.1140/epjc/s10052-018-6448-y. arXiv: 1810.03639 [hep-ph].
- [21] J. Gao, L. Harland-Lang, and J. Rojo. “The Structure of the Proton in the LHC Precision Era”. In: *Phys. Rept.* 742 (2018), pp. 1–121. DOI: 10.1016/j.physrep.2018.03.002. arXiv: 1709.04922 [hep-ph].
- [22] S. Alekhin, J. Blümlein, and S. Moch. “Strange sea determination from collider data”. In: *Phys. Lett. B* 777 (2018), pp. 134–140. DOI: 10.1016/j.physletb.2017.12.024. arXiv: 1708.01067 [hep-ph].

-
- [23] P. M. Nadolsky et al. “Implications of CTEQ global analysis for collider observables”. In: *Phys. Rev. D* 78 (2008), p. 013004. DOI: 10.1103/PhysRevD.78.013004. arXiv: 0802.0007 [hep-ph].
- [24] N. Sato et al. “Strange quark suppression from a simultaneous Monte Carlo analysis of parton distributions and fragmentation functions”. In: *Phys. Rev. D* 101.7 (2020), p. 074020. DOI: 10.1103/PhysRevD.101.074020. arXiv: 1905.03788 [hep-ph].
- [25] L.A. Harland-Lang et al. “Parton distributions in the LHC era: MMHT 2014 PDFs”. In: *Eur. Phys. J. C* 75.5 (2015), p. 204. DOI: 10.1140/epjc/s10052-015-3397-6. arXiv: 1412.3989 [hep-ph].
- [26] R. S. Thorne et al. “Updates of PDFs using the MMHT framework”. In: *PoS DIS2019* (2019), p. 036. DOI: 10.22323/1.352.0036. arXiv: 1907.08147 [hep-ph].
- [27] R. D. Ball et al. “Precision determination of electroweak parameters and the strange content of the proton from neutrino deep-inelastic scattering”. In: *Nucl. Phys. B* 823 (2009), pp. 195–233. DOI: 10.1016/j.nuclphysb.2009.08.003. arXiv: 0906.1958 [hep-ph].
- [28] H.-W. Lin et al. “Parton distributions and lattice QCD calculations: a community white paper”. In: *Prog. Part. Nucl. Phys.* 100 (2018), pp. 107–160. DOI: 10.1016/j.ppnp.2018.01.007. arXiv: 1711.07916 [hep-ph].
- [29] H.-W. Lin et al. “Parton distributions and lattice QCD calculations: toward 3D structure”. In: (2020). arXiv: 2006.08636 [hep-ph].
- [30] M. Diehl. “Generalized parton distributions”. In: *Phys. Rept.* 388 (2003), pp. 41–277. DOI: 10.1016/j.physrep.2003.08.002. arXiv: hep-ph/0307382.
- [31] G. Calucci and D. Treleani. “Proton structure in transverse space and the effective cross-section”. In: *Phys. Rev. D* 60 (1999), p. 054023. DOI: 10.1103/PhysRevD.60.054023. arXiv: hep-ph/9902479.
- [32] Alexandre Deur, Stanley J. Brodsky, and Guy F. De Téramond. “The Spin Structure of the Nucleon”. In: (2018). DOI: 10.1088/1361-6633/ab0b8f. arXiv: 1807.05250 [hep-ph].
- [33] Y. Goto et al. “Polarized parton distribution functions in the nucleon”. In: *Phys. Rev. D* 62 (2000), p. 034017. DOI: 10.1103/PhysRevD.62.034017. arXiv: hep-ph/0001046.
- [34] N. Paver and D. Treleani. “Multi - Quark Scattering and Large p_T Jet Production in Hadronic Collisions”. In: *Nuovo Cim. A* 70 (1982), p. 215. DOI: 10.1007/BF02814035.

-
- [35] Markus Diehl and Andreas Schafer. “Theoretical considerations on multiparton interactions in QCD”. In: *Phys. Lett. B* 698 (2011), pp. 389–402. DOI: 10.1016/j.physletb.2011.03.024. arXiv: 1102.3081 [hep-ph].
 - [36] Richard D. Ball et al. “Parton distributions with QED corrections”. In: *Nucl. Phys. B* 877 (2013), pp. 290–320. DOI: 10.1016/j.nuclphysb.2013.10.010. arXiv: 1308.0598 [hep-ph].
 - [37] A. D. Martin et al. “Parton distributions incorporating QED contributions”. In: *Eur. Phys. J. C* 39 (2005), pp. 155–161. DOI: 10.1140/epjc/s2004-02088-7. arXiv: hep-ph/0411040.
 - [38] S. Alekhin. “Parton distribution functions from the precise NNLO QCD fit”. In: *JETP Lett.* 82 (2005), pp. 628–631. DOI: 10.1134/1.2166909. arXiv: hep-ph/0508248.
 - [39] A. Kusina et al. “Impact of LHC vector boson production in heavy ion collisions on strange PDFs”. In: *Eur. Phys. J. C* 80.10 (2020), p. 968. DOI: 10.1140/epjc/s10052-020-08532-4. arXiv: 2007.09100 [hep-ph].
 - [40] A. Kusina et al. “Vector boson production in pPb and PbPb collisions at the LHC and its impact on nCTEQ15 PDFs”. In: *Eur. Phys. J. C* 77.7 (2017), p. 488. DOI: 10.1140/epjc/s10052-017-5036-x. arXiv: 1610.02925 [nucl-th].
 - [41] Morad Aaboud et al. “Measurement of prompt photon production in $\sqrt{s_{\text{NN}}} = 8.16$ TeV p +Pb collisions with ATLAS”. In: *Phys. Lett. B* 796 (2019), pp. 230–252. DOI: 10.1016/j.physletb.2019.07.031. arXiv: 1903.02209 [nucl-ex].
 - [42] T. Stavreva et al. “Probing gluon and heavy-quark nuclear PDFs with gamma + Q production in pA collisions”. In: *JHEP* 01 (2011), p. 152. DOI: 10.1007/JHEP01(2011)152. arXiv: 1012.1178 [hep-ph].
 - [43] Anton Andronic et al. “Decoding the phase structure of QCD via particle production at high energy”. In: *Nature* 561.7723 (2018), pp. 321–330. DOI: 10.1038/s41586-018-0491-6. arXiv: 1710.09425 [nucl-th].
 - [44] G. Altarelli and G. Parisi. “Asymptotic freedom in parton language”. In: *Nuclear Physics B* 126 (1977). DOI: 10.1016/0550-3213(77)90384-4.
 - [45] Yuri L. Dokshitzer. “Calculation of the Structure Functions for Deep Inelastic Scattering and e^+e^- Annihilation by Perturbation Theory in Quantum Chromodynamics.” In: *Sov. Phys. JETP* 46 (1977).
 - [46] V.N. Gribov and L.N. Lipatov. “Deep inelastic electron scattering in perturbation theory”. In: *Physics Letters B* 37 (1971). DOI: 10.1016/0370-2693(71)90576-4.
 - [47] L. N. Lipatov. “The parton model and perturbation theory”. In: *Sov. J. Nucl. Phys.* 20 (1975).

-
- [48] Nestor Armesto. “Nuclear shadowing”. In: *J. Phys. G* 32 (2006), R367–R394. DOI: 10.1088/0954-3899/32/11/R01. arXiv: hep-ph/0604108.
 - [49] L. Frankfurt, V. Guzey, and M. Strikman. “Leading Twist Nuclear Shadowing Phenomena in Hard Processes with Nuclei”. In: *Phys. Rept.* 512 (2012), pp. 255–393. DOI: 10.1016/j.physrep.2011.12.002. arXiv: 1106.2091 [hep-ph].
 - [50] B. Z. Kopeliovich, J. G. Morfin, and Ivan Schmidt. “Nuclear Shadowing in Electro-Weak Interactions”. In: *Prog. Part. Nucl. Phys.* 68 (2013), pp. 314–372. DOI: 10.1016/j.ppnp.2012.09.004. arXiv: 1208.6541 [hep-ph].
 - [51] Sergey A. Kulagin and R. Petti. “Global study of nuclear structure functions”. In: *Nucl. Phys. A* 765 (2006), pp. 126–187. DOI: 10.1016/j.nuclphysa.2005.10.011. arXiv: hep-ph/0412425.
 - [52] Stanley J. Brodsky and Hung Jung Lu. “Shadowing and Antishadowing of Nuclear Structure Functions”. In: *Phys. Rev. Lett.* 64 (1990), p. 1342. DOI: 10.1103/PhysRevLett.64.1342.
 - [53] Stanley J. Brodsky, Ivan Schmidt, and Jian-Jun Yang. “Nuclear antishadowing in neutrino deep inelastic scattering”. In: *Phys. Rev. D* 70 (2004), p. 116003. DOI: 10.1103/PhysRevD.70.116003. arXiv: hep-ph/0409279.
 - [54] Donald F. Geesaman, K. Saito, and Anthony William Thomas. “The nuclear EMC effect”. In: *Ann. Rev. Nucl. Part. Sci.* 45 (1995), pp. 337–390. DOI: 10.1146/annurev.ns.45.120195.002005.
 - [55] P. R. Norton. “The EMC effect”. In: *Rept. Prog. Phys.* 66 (2003), pp. 1253–1297. DOI: 10.1088/0034-4885/66/8/201.
 - [56] Or Hen et al. “The EMC Effect and High Momentum Nucleons in Nuclei”. In: *Int. J. Mod. Phys. E* 22 (2013), p. 1330017. DOI: 10.1142/S0218301313300178. arXiv: 1304.2813 [nucl-th].
 - [57] Simona Malace et al. “The Challenge of the EMC Effect: existing data and future directions”. In: *Int. J. Mod. Phys. E* 23.08 (2014), p. 1430013. DOI: 10.1142/S0218301314300136. arXiv: 1405.1270 [nucl-ex].
 - [58] O. Hen et al. “Nucleon-Nucleon Correlations, Short-lived Excitations, and the Quarks Within”. In: *Rev. Mod. Phys.* 89.4 (2017), p. 045002. DOI: 10.1103/RevModPhys.89.045002. arXiv: 1611.09748 [nucl-ex].
 - [59] C. G. Broyden. “The Convergence of a Class of Double-rank Minimization Algorithms 1. General Considerations”. In: *IMA Journal of Applied Mathematics* 6.1 (1970), 76–90. eprint: <https://academic.oup.com/imamat/article-abstract/6/1/76/746016?redirectedFrom=fulltext>. URL: <https://doi.org/10.1093/imamat/6.1.76>.

-
- [60] R. Fletcher. “A new approach to variable metric algorithms”. In: *The Computer Journal* 13.3 (1970), pp. 317–322. ISSN: 0010-4620. DOI: 10.1093/comjnl/13.3.317. eprint: <https://academic.oup.com/comjnl/article-pdf/13/3/317/988678/130317.pdf>. URL: <https://doi.org/10.1093/comjnl/13.3.317>.
 - [61] D. Goldfarb. “A family of variable-metric methods derived by variational means”. In: *Mathematics of Computation* 24 (1970), pp. 23–26. DOI: 10.1090/S0025-5718-1970-0258249-6. eprint: <https://www.ams.org/journals/mcom/1970-24-109/S0025-5718-1970-0258249-6/home.html>.
 - [62] D. F. Shanno. “Conditioning of quasi-Newton methods for function minimization”. In: *Mathematics of Computation* 24 (1970), pp. 647–656. ISSN: 0010-4620. DOI: 10.1090/S0025-5718-1970-0274029-X. eprint: <https://www.ams.org/journals/mcom/1970-24-111/S0025-5718-1970-0274029-X/home.html>.
 - [63] J. Pumplin et al. “Uncertainties of predictions from parton distribution functions. 2. The Hessian method”. In: *Phys. Rev. D* 65 (2001), p. 014013. DOI: 10.1103/PhysRevD.65.014013. arXiv: [hep-ph/0101032](https://arxiv.org/abs/hep-ph/0101032).
 - [64] P. Duwentäster et al. “Impact of inclusive hadron production data on nuclear gluon PDFs”. In: (2021). arXiv: 2105.09873 [hep-ph].
 - [65] P. Duwentäster et al. “Impact of heavy quark and quarkonium data on nuclear gluon PDFs”. In: (2022). arXiv: 2204.09982 [hep-ph].
 - [66] R. P. Feynman. *High Energy Collisions: International Conference Proceedings*. Gordon & Breach, 1970. ISBN: 978-0-677-13950-0.
 - [67] J. D. Bjorken. “Current algebra at small distances”. In: *Conf. Proc.* C670717 (1967).
 - [68] J. D. Bjorken. “Asymptotic Sum Rules at Infinite Momentum”. In: *Physical Review* 179 (1969). DOI: 10.1103/physrev.179.1547.
 - [69] F. Halzen and A. Martin. *Quarks and Leptons*. John Wiley & Sons, 1984. 416 pp. ISBN: 0471887412.
 - [70] M. Gell-Mann. “A schematic model of baryons and mesons”. In: *Physics Letters* 8 (1964). DOI: 10.1016/s0031-9163(64)92001-3.
 - [71] A.C. Benvenuti et al. “A high statistics measurement of the proton structure functions $F_2(x, Q^2)$ and R from deep inelastic muon scattering at high Q^2 ”. In: *Phys. Lett B* 223 (1989). DOI: 10.1016/0370-2693(89)91637-7.
 - [72] P. Duwentäster. “Parton Distribution Functions from Deep Inelastic Scattering”. In: *Masters thesis* (2019), p. 88. URL: https://www.uni-muenster.de/imperia/md/content/physik_tp/theses/klasen/duwentaester_msc.pdf.
 - [73] J. D. Hunter. “Matplotlib: A 2D graphics environment”. In: *Computing in Science & Engineering* 9.3 (2007), pp. 90–95. DOI: 10.1109/MCSE.2007.55.

-
- [74] M. D. Schwartz. *Quantum Field Theory and the Standard Model*. Cambridge University Pr., 2014. ISBN: 1107034736.
 - [75] Michele Arneodo. “Nuclear effects in structure functions”. In: *Phys. Rept.* 240 (1994), pp. 301–393. DOI: 10.1016/0370-1573(94)90048-5.
 - [76] J. Gomez et al. “Measurement of the A-dependence of deep inelastic electron scattering”. In: *Phys. Rev. D* 49 (1994), pp. 4348–4372. DOI: 10.1103/PhysRevD.49.4348.
 - [77] G. Bari et al. “A Measurement of Nuclear Effects in Deep Inelastic Muon Scattering on Deuterium, Nitrogen and Iron Targets”. In: *Phys. Lett. B* 163 (1985), p. 282. DOI: 10.1016/0370-2693(85)90238-2.
 - [78] A. C. Benvenuti et al. “Nuclear Effects in Deep Inelastic Muon Scattering on Deuterium and Iron Targets”. In: *Phys. Lett. B* 189 (1987), pp. 483–487. DOI: 10.1016/0370-2693(87)90664-2.
 - [79] A. Bodek et al. “A Comparison of the Deep Inelastic Structure Functions of Deuterium and Aluminum Nuclei”. In: *Phys. Rev. Lett.* 51 (1983), p. 534. DOI: 10.1103/PhysRevLett.51.534.
 - [80] S. Dasu et al. “Measurement of kinematic and nuclear dependence of $R = \sigma_L / \sigma_T$ in deep inelastic electron scattering”. In: *Phys. Rev. D* 49 (1994), pp. 5641–5670. DOI: 10.1103/PhysRevD.49.5641.
 - [81] U. Landgraf. “Nuclear effects in deep inelastic scattering”. In: *Nucl. Phys. A* 527 (1991). Ed. by J. L. Matthews et al., pp. 123C–135C. DOI: 10.1016/0375-9474(91)90110-R.
 - [82] I. Schienbein et al. “Nuclear parton distribution functions from neutrino deep inelastic scattering”. In: *Phys. Rev. D* 77 (2008), p. 054013. DOI: 10.1103/PhysRevD.77.054013. arXiv: 0710.4897 [hep-ph].
 - [83] A. Bodek and J. L. Ritchie. “Fermi Motion Effects in Deep Inelastic Lepton Scattering from Nuclear Targets”. In: *Phys. Rev. D* 23 (1981), p. 1070. DOI: 10.1103/PhysRevD.23.1070.
 - [84] Edmond Iancu, Andrei Leonidov, and Larry D. McLerran. “Nonlinear gluon evolution in the color glass condensate. 1.” In: *Nucl. Phys. A* 692 (2001), pp. 583–645. DOI: 10.1016/S0375-9474(01)00642-X. arXiv: hep-ph/0011241.
 - [85] Francois Gelis et al. “The Color Glass Condensate”. In: *Ann. Rev. Nucl. Part. Sci.* 60 (2010), pp. 463–489. DOI: 10.1146/annurev.nucl.010909.083629. arXiv: 1002.0333 [hep-ph].
 - [86] Jamal Jalilian-Marian. “From small to large x : toward a unified description of high energy collisions”. In: *J. Phys. Conf. Ser.* 1602.1 (2020). Ed. by Rene Bellwied et al., p. 012024. DOI: 10.1088/1742-6596/1602/1/012024.

-
- [87] R. J. Glauber. “Cross-sections in deuterium at high-energies”. In: *Phys. Rev.* 100 (1955), pp. 242–248. DOI: 10.1103/PhysRev.100.242.
 - [88] S. Mandelstam. “Cuts in the Angular Momentum Plane. 2”. In: *Nuovo Cim.* 30 (1963), pp. 1148–1162. DOI: 10.1007/BF02828822.
 - [89] L. L. Frankfurt and M. I. Strikman. “Hard Nuclear Processes and Microscopic Nuclear Structure”. In: *Phys. Rept.* 160 (1988), pp. 235–427. DOI: 10.1016/0370-1573(88)90179-2.
 - [90] John C. Collins. “Proof of factorization for diffractive hard scattering”. In: *Phys. Rev. D* 57 (1998). [Erratum: *Phys.Rev.D* 61, 019902 (2000)], pp. 3051–3056. DOI: 10.1103/PhysRevD.61.019902. arXiv: hep-ph/9709499.
 - [91] L. Frankfurt, V. Guzey, and M. Strikman. “Dynamical model of antishadowing of the nuclear gluon distribution”. In: *Phys. Rev. C* 95.5 (2017), p. 055208. DOI: 10.1103/PhysRevC.95.055208. arXiv: 1612.08273 [hep-ph].
 - [92] S. V. Akulinichev, Sergey A. Kulagin, and G. M. Vagradov. “The Role of Nuclear Binding in Deep Inelastic Lepton Nucleon Scattering”. In: *Phys. Lett. B* 158 (1985), pp. 485–488. DOI: 10.1016/0370-2693(85)90799-3.
 - [93] S. V. Akulinichev et al. “LEPTON NUCLEUS DEEP INELASTIC SCATTERING”. In: *Phys. Rev. Lett.* 55 (1985), pp. 2239–2241. DOI: 10.1103/PhysRevLett.55.2239.
 - [94] Edmond L. Berger, F. Coester, and Robert B. Wiringa. “Pion Density in Nuclei and Deep Inelastic Lepton Scattering”. In: *Phys. Rev. D* 29 (1984), p. 398. DOI: 10.1103/PhysRevD.29.398.
 - [95] Magda Ericson and Anthony William Thomas. “Pionic Corrections and the EMC Enhancement of the Sea in Iron”. In: *Phys. Lett. B* 128 (1983), pp. 112–116. DOI: 10.1016/0370-2693(83)90085-0.
 - [96] C. E. Carlson and T. J. Havens. “Quark Distributions in Nuclei”. In: *Phys. Rev. Lett.* 51 (1983), p. 261. DOI: 10.1103/PhysRevLett.51.261.
 - [97] R. L. Jaffe. “Quark Distributions in Nuclei”. In: *Phys. Rev. Lett.* 50 (1983), p. 228. DOI: 10.1103/PhysRevLett.50.228.
 - [98] F. E. Close, R. G. Roberts, and Graham G. Ross. “The Effect of Confinement Size on Nuclear Structure Functions”. In: *Phys. Lett. B* 129 (1983), pp. 346–350. DOI: 10.1016/0370-2693(83)90679-2.
 - [99] O. Nachtmann and H. J. Pirner. “Color Conductivity in Nuclei and the Emc Effect”. In: *Z. Phys. C* 21 (1984), p. 277. DOI: 10.1007/BF01577042.
 - [100] Wolfgang Bentz and Anthony William Thomas. “The Stability of nuclear matter in the Nambu-Jona-Lasinio model”. In: *Nucl. Phys. A* 696 (2001), pp. 138–172. DOI: 10.1016/S0375-9474(01)01119-8. arXiv: nucl-th/0105022.

-
- [101] Jason Robert Smith and Gerald A. Miller. “Chiral solitons in nuclei: Saturation, EMC effect and Drell-Yan experiments”. In: *Phys. Rev. Lett.* 91 (2003). [Erratum: *Phys.Rev.Lett.* 98, 099902 (2007)], p. 212301. DOI: 10.1103/PhysRevLett.91.212301. arXiv: nucl-th/0308048.
 - [102] Claudio Ciofi Degli Atti and S. Liuti. “On the Effects of Nucleon Binding and Correlations in Deep Inelastic Electron Scattering by Nuclei”. In: *Phys. Lett. B* 225 (1989), pp. 215–221. DOI: 10.1016/0370-2693(89)90808-3.
 - [103] Claudio Ciofi degli Atti and S. Liuti. “Realistic microscopic approach to deep inelastic scattering of electrons off few nucleon systems”. In: *Phys. Rev. C* 41 (1990), pp. 1100–1114. DOI: 10.1103/PhysRevC.41.1100.
 - [104] D. M. Alde et al. “Nuclear dependence of dimuon production at 800-GeV. FNAL-772 experiment”. In: *Phys. Rev. Lett.* 64 (1990), pp. 2479–2482. DOI: 10.1103/PhysRevLett.64.2479.
 - [105] N. Fomin et al. “Scaling of the F_2 structure function in nuclei and quark distributions at $x > 1$ ”. In: *Phys. Rev. Lett.* 105 (2010), p. 212502. DOI: 10.1103/PhysRevLett.105.212502. arXiv: 1008.2713 [nucl-ex].
 - [106] J. Arrington, D.F. Geesaman, K. Hafidi, et al. “Inclusive Scattering from Nuclei at $x > 1$ in the quasielastic and deeply inelastic regimes”. In: *JLab Proposal* PR12-06-105 (2006). URL: https://www.jlab.org/exp_prog/proposals/06/PR12-06-105.pdf.
 - [107] V. Guzey and M. Klasen. “Constraints on nuclear parton distributions from dijet photoproduction at the LHC”. In: *Eur. Phys. J. C* 79.5 (2019), p. 396. DOI: 10.1140/epjc/s10052-019-6905-2. arXiv: 1902.05126 [hep-ph].
 - [108] M. Klasen, K. Kovarik, and J. Potthoff. “Nuclear parton density functions from jet production in DIS at an EIC”. In: *Phys. Rev. D* 95.9 (2017), p. 094013. DOI: 10.1103/PhysRevD.95.094013. arXiv: 1703.02864 [hep-ph].
 - [109] M. Klasen and K. Kovařík. “Nuclear parton density functions from dijet photoproduction at the EIC”. In: *Phys. Rev. D* 97.11 (2018), p. 114013. DOI: 10.1103/PhysRevD.97.114013. arXiv: 1803.10985 [hep-ph].
 - [110] Robert Jastrow. “Many-Body Problem with Strong Forces”. In: *Phys. Rev.* 98 (1955), pp. 1479–1484. DOI: 10.1103/PhysRev.98.1479.
 - [111] D. B. Day et al. “Y Scaling in Electron Nucleus Scattering”. In: *Phys. Rev. Lett.* 59 (1987), pp. 427–430. DOI: 10.1103/PhysRevLett.59.427.
 - [112] L. L. Frankfurt et al. “Evidence for short range correlations from high Q^2 (e, e-prime) reactions”. In: *Phys. Rev. C* 48 (1993), pp. 2451–2461. DOI: 10.1103/PhysRevC.48.2451.

-
- [113] E. Piasetzky et al. “Evidence for the strong dominance of proton-neutron correlations in nuclei”. In: *Phys. Rev. Lett.* 97 (2006), p. 162504. DOI: 10.1103/PhysRevLett.97.162504. arXiv: nucl-th/0604012.
 - [114] Nadia Fomin et al. “New Results on Short-Range Correlations in Nuclei”. In: *Ann. Rev. Nucl. Part. Sci.* 67 (2017), pp. 129–159. DOI: 10.1146/annurev-nucl-102115-044939. arXiv: 1708.08581 [nucl-th].
 - [115] L. B. Weinstein et al. “Short Range Correlations and the EMC Effect”. In: *Phys. Rev. Lett.* 106 (2011), p. 052301. DOI: 10.1103/PhysRevLett.106.052301. arXiv: 1009.5666 [hep-ph].
 - [116] A. Kusina, A.W. Denniston, T. Jezo, et al. “Short-range correlations and the phenomenological determination of nuclear PDFs”. In: *XXIX International Workshop on Deep-Inelastic Scattering and Related Subjects* (2022). URL: https://indico.cern.ch/event/1072533/contributions/4776239/attachments/2435944/4171910/kusina_DIS2022.pdf.
 - [117] Ingo Schienbein et al. “A Review of Target Mass Corrections”. In: *J. Phys. G* 35 (2008), p. 053101. DOI: 10.1088/0954-3899/35/5/053101. arXiv: 0709.1775 [hep-ph].
 - [118] Ingo Schienbein et al. “Target mass corrections in lepton-nucleus DIS revisited”. In: *Unpublished* (). In preparation.
 - [119] E. P. Segarra et al. “nCTEQ15HIX – Extending nPDF Analyses into the High- x , Low Q^2 Region”. In: (2020). arXiv: 2012.11566 [hep-ph].
 - [120] P. Braun-Munzinger et al. “Hadron production in Au - Au collisions at RHIC”. In: *Phys. Lett. B* 518 (2001), pp. 41–46. DOI: 10.1016/S0370-2693(01)01069-3. arXiv: hep-ph/0105229.
 - [121] P. Braun-Munzinger et al. “Thermal and hadrochemical equilibration in nucleus-nucleus collisions at the SPS”. In: *Phys. Lett. B* 365 (1996), pp. 1–6. DOI: 10.1016/0370-2693(95)01258-3. arXiv: nucl-th/9508020.
 - [122] Peter Braun-Munzinger, Krzysztof Redlich, and Johanna Stachel. “Particle production in heavy ion collisions”. In: (2003). Ed. by Rudolph C. Hwa and Xin-Nian Wang, pp. 491–599. DOI: 10.1142/9789812795533_0008. arXiv: nucl-th/0304013.
 - [123] J. Stachel et al. “Confronting LHC data with the statistical hadronization model”. In: *J. Phys. Conf. Ser.* 509 (2014). Ed. by David Evans et al., p. 012019. DOI: 10.1088/1742-6596/509/1/012019. arXiv: 1311.4662 [nucl-th].
 - [124] Claude Amsler et al. “Review of Particle Physics”. In: *Phys. Lett. B* 667 (2008), pp. 1–1340. DOI: 10.1016/j.physletb.2008.07.018.

-
- [125] J. Stachel. “Statistical hadronization at the QGP phase boundary”. 2016. URL: https://indico.cern.ch/event/491202/contributions/2249893/attachments/1327632/1993671/pbm70_stachel.pdf.
 - [126] P. Braun-Munzinger and J. Stachel. “(Non)thermal aspects of charmonium production and a new look at J/ψ suppression”. In: *Phys. Lett. B* 490 (2000), pp. 196–202. DOI: 10.1016/S0370-2693(00)00991-6. arXiv: nucl-th/0007059.
 - [127] Peter Braun-Munzinger and Johanna Stachel. “On charm production near the phase boundary”. In: *Nucl. Phys. A* 690 (2001). Ed. by P. Braun-Munzinger and A. Zilges, pp. 119–126. DOI: 10.1016/S0375-9474(01)00936-8. arXiv: nucl-th/0012064.
 - [128] T. Matsui and H. Satz. “ J/ψ Suppression by Quark-Gluon Plasma Formation”. In: *Phys. Lett. B* 178 (1986), pp. 416–422. DOI: 10.1016/0370-2693(86)91404-8.
 - [129] A. Andronic et al. “Testing charm quark thermalisation within the Statistical Hadronisation Model”. In: *Nucl. Phys. A* 982 (2019). Ed. by Federico Antinori et al., pp. 759–762. DOI: 10.1016/j.nuclphysa.2018.09.004. arXiv: 1807.01236 [nucl-th].
 - [130] Anton Andronic et al. “Transverse momentum distributions of charmonium states with the statistical hadronization model”. In: *Phys. Lett. B* 797 (2019), p. 134836. DOI: 10.1016/j.physletb.2019.134836. arXiv: 1901.09200 [nucl-th].
 - [131] D. Stump et al. “Inclusive jet production, parton distributions, and the search for new physics”. In: *JHEP* 10 (2003), p. 046. DOI: 10.1088/1126-6708/2003/10/046. arXiv: hep-ph/0303013.
 - [132] J. Pumplin et al. “New generation of parton distributions with uncertainties from global QCD analysis”. In: *JHEP* 07 (2002), p. 012. DOI: 10.1088/1126-6708/2002/07/012. arXiv: hep-ph/0201195.
 - [133] J. F. Owens et al. “The Impact of new neutrino DIS and Drell-Yan data on large- x parton distributions”. In: *Phys. Rev. D* 75 (2007), p. 054030. DOI: 10.1103/PhysRevD.75.054030. arXiv: hep-ph/0702159.
 - [134] I. Niculescu et al. “Moments of nuclear and nucleon structure functions at low Q^2 and the momentum sum rule”. In: *Phys. Rev. C* 73 (2006), p. 045206. DOI: 10.1103/PhysRevC.73.045206. arXiv: hep-ph/0509241.
 - [135] D. Stump et al. “Uncertainties of predictions from parton distribution functions. 1. The Lagrange multiplier method”. In: *Phys. Rev. D* 65 (2001), p. 014012. DOI: 10.1103/PhysRevD.65.014012. arXiv: hep-ph/0101051.
 - [136] G. D’Agostini. “On the use of the covariance matrix to fit correlated data”. In: *Nucl. Instrum. Meth. A* 346 (1994), pp. 306–311. DOI: 10.1016/0168-9002(94)90719-6.

-
- [137] W. C. Davidon. “VARIABLE METRIC METHOD FOR MINIMIZATION”. In: *UNT Digital Library* ANL-5990 (1959). DOI: 10.2172/4252678. URL: <https://www.osti.gov/biblio/4252678>.
 - [138] F. James and M. Winkler. “C++ MINUIT User’s Guide”. Part of the CERN ROOT package. URL: <https://root.cern.ch/root/html/doc/guides/minuit2/Minuit2.pdf>.
 - [139] I. Antcheva et al. “ROOT A C++ framework for petabyte data storage, statistical analysis and visualization”. In: *Computer Physics Communications* 180 (2009). DOI: 10.1016/j.cpc.2009.08.005.
 - [140] Charles G Broyden. “A class of methods for solving nonlinear simultaneous equations”. In: *Mathematics of computation* 19.92 (1965), pp. 577–593.
 - [141] Stephen Wright, Jorge Nocedal, et al. “Numerical optimization”. In: *Springer Science* 35.67-68 (1999), p. 7.
 - [142] White. “Maximum Likelihood Estimation of Misspecified Models”. In: *Econometrica* 50.1 (1982), pp. 1–25. ISSN: 00129682, 14680262. URL: <http://www.jstor.org/stable/1912526>.
 - [143] Rabah Abdul Khalek et al. “A first determination of parton distributions with theoretical uncertainties”. In: *Eur. Phys. J. C* (2019), 79:838. DOI: 10.1140/epjc/s10052-019-7364-5. arXiv: 1905.04311 [hep-ph].
 - [144] Walter T. Giele and Stephane Keller. “Implications of hadron collider observables on parton distribution function uncertainties”. In: *Phys. Rev. D* 58 (1998), p. 094023. DOI: 10.1103/PhysRevD.58.094023. arXiv: hep-ph/9803393.
 - [145] D. R. Cox. *Principles of Statistical Inference*. Cambridge University Press, 2006. DOI: 10.1017/CB09780511813559.
 - [146] G. Watt and R. S. Thorne. “Study of Monte Carlo approach to experimental uncertainty propagation with MSTW 2008 PDFs”. In: *JHEP* 08 (2012), p. 052. DOI: 10.1007/JHEP08(2012)052. arXiv: 1205.4024 [hep-ph].
 - [147] John C. Collins, Davison E. Soper, and George F. Sterman. “Factorization of Hard Processes in QCD”. In: *Adv. Ser. Direct. High Energy Phys.* 5 (1989), pp. 1–91. DOI: 10.1142/9789814503266_0001. arXiv: hep-ph/0409313.
 - [148] Alberto Accardi et al. “Hard probes in heavy ion collisions at the lhc: pdfs, shadowing and pa collisions”. In: *3rd Workshop on Hard Probes in Heavy Ion Collisions: 3rd Plenary Meeting*. 2004. arXiv: hep-ph/0308248.
 - [149] Jian-wei Qiu. “QCD factorization and rescattering in proton nucleus collisions”. In: (2003). arXiv: hep-ph/0305161.

-
- [150] K. J. Eskola. “Scale dependence of nuclear gluon structure”. In: *Nucl. Phys. B* 400 (1993), pp. 240–266. DOI: 10.1016/0550-3213(93)90406-F.
 - [151] K. J. Eskola, V. J. Kolhinen, and P. V. Ruuskanen. “Scale evolution of nuclear parton distributions”. In: *Nucl. Phys. B* 535 (1998), pp. 351–371. DOI: 10.1016/S0550-3213(98)00589-6. arXiv: hep-ph/9802350.
 - [152] M. Hirai, S. Kumano, and M. Miyama. “Determination of nuclear parton distributions”. In: *Phys. Rev. D* 64 (2001), p. 034003. DOI: 10.1103/PhysRevD.64.034003. arXiv: hep-ph/0103208.
 - [153] D. de Florian and R. Sassot. “Nuclear parton distributions at next-to-leading order”. In: *Phys. Rev. D* 69 (2004), p. 074028. DOI: 10.1103/PhysRevD.69.074028. arXiv: hep-ph/0311227.
 - [154] I. Schienbein et al. “PDF Nuclear Corrections for Charged and Neutral Current Processes”. In: *Phys. Rev. D* 80 (2009), p. 094004. DOI: 10.1103/PhysRevD.80.094004. arXiv: 0907.2357 [hep-ph].
 - [155] M. Werlen. “INCNLO-direct photon and inclusive hadron production code Website”. Version 1.4. URL: http://lapth.cnrs.fr/PHOX_FAMILY.
 - [156] Gavin P. Salam and Juan Rojo. “A Higher Order Perturbative Parton Evolution Toolkit (HOPPET)”. In: *Comput. Phys. Commun.* 180 (2009), pp. 120–156. DOI: 10.1016/j.cpc.2008.08.010. arXiv: 0804.3755 [hep-ph].
 - [157] T. Carli et al. “A posteriori inclusion of parton density functions in NLO QCD final-state calculations at hadron colliders: The APPLGRID Project”. In: *Eur. Phys. J. C* 66 (2010), pp. 503–524. DOI: 10.1140/epjc/s10052-010-1255-0. arXiv: 0911.2985 [hep-ph].
 - [158] John M. Campbell, R. Keith Ellis, and Walter T. Giele. “A Multi-Threaded Version of MCFM”. In: *Eur. Phys. J. C* 75.6 (2015), p. 246. DOI: 10.1140/epjc/s10052-015-3461-2. arXiv: 1503.06182 [physics.comp-ph].
 - [159] K. F. Muzakka et al. “Compatibility of Neutrino DIS Data and Its Impact on Nuclear Parton Distribution Functions”. In: (2022). arXiv: 2204.13157 [hep-ph].
 - [160] Ilkka Helenius, Marina Walt, and Werner Vogelsang. “TUJU21: NNLO nuclear parton distribution functions with electroweak-boson production data from the LHC”. In: (2021). arXiv: 2112.11904 [hep-ph].
 - [161] V. Bertone et al. “xFitter 2.0.0: An Open Source QCD Fit Framework”. In: *PoS DIS2017* (2018). Ed. by Uta Klein, p. 203. DOI: 10.22323/1.297.0203. arXiv: 1709.01151 [hep-ph].

-
- [162] Hamzeh Khanpour et al. “Nuclear parton distribution functions with uncertainties in a general mass variable flavor number scheme”. In: *Phys. Rev. D* 104.3 (2021), p. 034010. DOI: 10.1103/PhysRevD.104.034010. arXiv: 2010.00555 [hep-ph].
 - [163] Valerio Bertone, Stefano Carrazza, and Juan Rojo. “APFEL: A PDF Evolution Library with QED corrections”. In: *Comput. Phys. Commun.* 185 (2014), pp. 1647–1668. DOI: 10.1016/j.cpc.2014.03.007. arXiv: 1310.1394 [hep-ph].
 - [164] D. de Florian et al. “Parton-to-Pion Fragmentation Reloaded”. In: *Phys. Rev. D* 91.1 (2015), p. 014035. DOI: 10.1103/PhysRevD.91.014035. arXiv: 1410.6027 [hep-ph].
 - [165] P. Aurenche et al. “Large $p(T)$ inclusive π^0 cross-sections and next-to-leading-order QCD predictions”. In: *Eur. Phys. J. C* 13 (2000), pp. 347–355. DOI: 10.1007/s100520000309. arXiv: hep-ph/9910252.
 - [166] M. Soleymaninia, M. Goharipour, and H. Khanpour. “First QCD analysis of charged hadron fragmentation functions and their uncertainties at next-to-next-to-leading order”. In: *Phys. Rev. D* 98.7 (2018), p. 074002. DOI: 10.1103/PhysRevD.98.074002. arXiv: 1805.04847 [hep-ph].
 - [167] V. Bertone et al. “Charged hadron fragmentation functions from collider data”. In: *Eur. Phys. J. C* 78.8 (2018), p. 651. DOI: 10.1140/epjc/s10052-018-6130-4. arXiv: 1807.03310 [hep-ph].
 - [168] R. Sassot, M. Stratmann, and P. Zurita. “Fragmentations Functions in Nuclear Media”. In: *Phys. Rev. D* 81 (2010), p. 054001. DOI: 10.1103/PhysRevD.81.054001. arXiv: 0912.1311 [hep-ph].
 - [169] Pía Zurita. “Medium modified Fragmentation Functions with open source xFitter”. In: (2021). arXiv: 2101.01088 [hep-ph].
 - [170] Daniel de Florian, Rodolfo Sassot, and Marco Stratmann. “Global analysis of fragmentation functions for protons and charged hadrons”. In: *Phys. Rev. D* 76 (2007), p. 074033. DOI: 10.1103/PhysRevD.76.074033. arXiv: 0707.1506 [hep-ph].
 - [171] M. Hirai et al. “Impacts of B-factory measurements on determination of fragmentation functions from electron-positron annihilation data”. In: *PTEP* 2016.11 (2016), 113B04. DOI: 10.1093/ptep/ptw154. arXiv: 1608.04067 [hep-ph].
 - [172] J. Binnewies, B. A. Kniehl, and G. Kramer. “Next-to-leading order fragmentation functions for pions and kaons”. In: *Z. Phys. C* 65 (1995), pp. 471–480. DOI: 10.1007/BF01556135. arXiv: hep-ph/9407347.
 - [173] B. A. Kniehl, G. Kramer, and B. Potter. “Fragmentation functions for pions, kaons, and protons at next-to-leading order”. In: *Nucl. Phys. B* 582 (2000), pp. 514–536. DOI: 10.1016/S0550-3213(00)00303-5. arXiv: hep-ph/0010289.

-
- [174] S. Kretzer. “Fragmentation functions from flavor inclusive and flavor tagged e^+e^- annihilations”. In: *Phys. Rev. D* 62 (2000), p. 054001. DOI: 10.1103/PhysRevD.62.054001. arXiv: hep-ph/0003177.
 - [175] M. Hirai et al. “Determination of fragmentation functions and their uncertainties”. In: *Phys. Rev. D* 75 (2007), p. 094009. DOI: 10.1103/PhysRevD.75.094009. arXiv: hep-ph/0702250.
 - [176] S. Albino, B. A. Kniehl, and G. Kramer. “AKK Update: Improvements from New Theoretical Input and Experimental Data”. In: *Nucl. Phys. B* 803 (2008), pp. 42–104. DOI: 10.1016/j.nuclphysb.2008.05.017. arXiv: 0803.2768 [hep-ph].
 - [177] V. Bertone et al. “A determination of the fragmentation functions of pions, kaons, and protons with faithful uncertainties”. In: *Eur. Phys. J. C* 77.8 (2017), p. 516. DOI: 10.1140/epjc/s10052-017-5088-y. arXiv: 1706.07049 [hep-ph].
 - [178] E. Moffat et al. “Simultaneous Monte Carlo analysis of parton densities and fragmentation functions”. In: (2021). arXiv: 2101.04664 [hep-ph].
 - [179] D. de Florian et al. “Parton-to-Kaon Fragmentation Revisited”. In: *Phys. Rev. D* 95.9 (2017), p. 094019. DOI: 10.1103/PhysRevD.95.094019. arXiv: 1702.06353 [hep-ph].
 - [180] C. A. Aidala et al. “Global Analysis of Fragmentation Functions for Eta Mesons”. In: *Phys. Rev. D* 83 (2011), p. 034002. DOI: 10.1103/PhysRevD.83.034002. arXiv: 1009.6145 [hep-ph].
 - [181] S. S. Adler et al. “Centrality dependence of π^0 and η production at large transverse momentum in $\sqrt{s_{NN}} = 200$ GeV d+Au collisions”. In: *Phys. Rev. Lett.* 98 (2007), p. 172302. DOI: 10.1103/PhysRevLett.98.172302. arXiv: nucl-ex/0610036.
 - [182] A. Adare et al. “Spectra and ratios of identified particles in Au+Au and d+Au collisions at $\sqrt{s_{NN}} = 200$ GeV”. In: *Phys. Rev. C* 88.2 (2013), p. 024906. DOI: 10.1103/PhysRevC.88.024906. arXiv: 1304.3410 [nucl-ex].
 - [183] B. I. Abelev et al. “Inclusive π^0 , η , and direct photon production at high transverse momentum in $p + p$ and d+Au collisions at $\sqrt{s_{NN}} = 200$ GeV”. In: *Phys. Rev. C* 81 (2010), p. 064904. DOI: 10.1103/PhysRevC.81.064904. arXiv: 0912.3838 [hep-ex].
 - [184] J. Adams et al. “Identified hadron spectra at large transverse momentum in p+p and d+Au collisions at $\sqrt{s_{NN}} = 200$ GeV”. In: *Phys. Lett. B* 637 (2006), pp. 161–169. DOI: 10.1016/j.physletb.2006.04.032. arXiv: nucl-ex/0601033.
 - [185] S. Acharya et al. “Neutral pion and η meson production in p-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV”. In: *Eur. Phys. J. C* 78.8 (2018), p. 624. DOI: 10.1140/epjc/s10052-018-6013-8. arXiv: 1801.07051 [nucl-ex].

-
- [186] J. Adam et al. “Multiplicity dependence of charged pion, kaon, and (anti)proton production at large transverse momentum in p-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV”. In: *Phys. Lett. B* 760 (2016), pp. 720–735. DOI: 10.1016/j.physletb.2016.07.050. arXiv: 1601.03658 [nucl-ex].
 - [187] S. Acharya et al. “Nuclear modification factor of light neutral-meson spectra up to high transverse momentum in p-Pb collisions at $\sqrt{s_{NN}} = 8.16$ TeV”. In: (2021). arXiv: 2104.03116 [nucl-ex].
 - [188] A. D. Martin et al. “Parton distributions for the LHC”. In: *Eur. Phys. J. C* 63 (2009), pp. 189–285. DOI: 10.1140/epjc/s10052-009-1072-5. arXiv: 0901.0002 [hep-ph].
 - [189] S. S. Adler et al. “Mid-rapidity neutral pion production in proton proton collisions at $\sqrt{s} = 200$ -GeV”. In: *Phys. Rev. Lett.* 91 (2003), p. 241803. DOI: 10.1103/PhysRevLett.91.241803. arXiv: hep-ex/0304038.
 - [190] B. Abelev et al. “Neutral pion and η meson production in proton-proton collisions at $\sqrt{s} = 0.9$ TeV and $\sqrt{s} = 7$ TeV”. In: *Phys. Lett. B* 717 (2012), pp. 162–172. DOI: 10.1016/j.physletb.2012.09.015. arXiv: 1205.5724 [hep-ex].
 - [191] D. de Florian et al. “Global Analysis of Nuclear Parton Distributions”. In: *Phys. Rev. D* 85 (2012), p. 074028. DOI: 10.1103/PhysRevD.85.074028. arXiv: 1112.6324 [hep-ph].
 - [192] Oleksandr Zenaiev et al. “Impact of heavy-flavour production cross sections measured by the LHCb experiment on parton distribution functions at low x ”. In: *Eur. Phys. J. C* 75.8 (2015), p. 396. DOI: 10.1140/epjc/s10052-015-3618-z. arXiv: 1503.04581 [hep-ph].
 - [193] Aleksander Kusina et al. “Gluon Shadowing in Heavy-Flavor Production at the LHC”. In: *Phys. Rev. Lett.* 121.5 (2018), p. 052004. DOI: 10.1103/PhysRevLett.121.052004. arXiv: 1712.07024 [hep-ph].
 - [194] Rabah Abdul Khalek et al. “nNNPDF3.0: Evidence for a modified partonic structure in heavy nuclei”. In: (2022). arXiv: 2201.12363 [hep-ph].
 - [195] Ilkka Helenius and Hannu Paukkunen. “Revisiting the D-meson hadroproduction in general-mass variable flavour number scheme”. In: *JHEP* 05 (2018), p. 196. DOI: 10.1007/JHEP05(2018)196. arXiv: 1804.03557 [hep-ph].
 - [196] B. A. Kniehl et al. “Inclusive D^{*+} production in p anti-p collisions with massive charm quarks”. In: *Phys. Rev. D* 71 (2005), p. 014018. DOI: 10.1103/PhysRevD.71.014018. arXiv: hep-ph/0410289.

-
- [197] B. A. Kniehl et al. “Collinear subtractions in hadroproduction of heavy quarks”. In: *Eur. Phys. J. C* 41 (2005), pp. 199–212. DOI: 10.1140/epjc/s2005-02200-7. arXiv: hep-ph/0502194.
 - [198] Simone Alioli et al. “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX”. In: *JHEP* 06 (2010), p. 043. DOI: 10.1007/JHEP06(2010)043. arXiv: 1002.2581 [hep-ph].
 - [199] Torbjörn Sjöstrand et al. “An introduction to PYTHIA 8.2”. In: *Comput. Phys. Commun.* 191 (2015), pp. 159–177. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph].
 - [200] C. H. Kom, A. Kulesza, and W. J. Stirling. “Pair Production of J/ψ as a Probe of Double Parton Scattering at LHCb”. In: *Phys. Rev. Lett.* 107 (2011), p. 082002. DOI: 10.1103/PhysRevLett.107.082002. arXiv: 1105.4186 [hep-ph].
 - [201] Francois Arleo and Stephane Peigne. “ J/ψ suppression in p-A collisions from parton energy loss in cold QCD matter”. In: *Phys. Rev. Lett.* 109 (2012), p. 122301. DOI: 10.1103/PhysRevLett.109.122301. arXiv: 1204.4609 [hep-ph].
 - [202] Francois Arleo and Stephane Peigne. “Heavy-quarkonium suppression in p-A collisions from parton energy loss in cold QCD matter”. In: *JHEP* 03 (2013), p. 122. DOI: 10.1007/JHEP03(2013)122. arXiv: 1212.0434 [hep-ph].
 - [203] François Arleo, Greg Jackson, and Stéphane Peigné. “Impact of fully coherent energy loss on heavy meson production in pA collisions”. In: *JHEP* 01 (2022), p. 164. DOI: 10.1007/JHEP01(2022)164. arXiv: 2107.05871 [hep-ph].
 - [204] A. Andronic et al. “Heavy-flavour and quarkonium production in the LHC era: from proton–proton to heavy-ion collisions”. In: *Eur. Phys. J. C* 76.3 (2016), p. 107. DOI: 10.1140/epjc/s10052-015-3819-5. arXiv: 1506.03981 [nucl-ex].
 - [205] Oleksandr Zenaiev. “Charm Production and QCD Analysis at HERA and LHC”. In: *Eur. Phys. J. C* 77.3 (2017), p. 151. DOI: 10.1140/epjc/s10052-017-4620-4. arXiv: 1612.02371 [hep-ex].
 - [206] W. Beenakker et al. “QCD corrections to heavy quark production in hadron hadron collisions”. In: *Nucl. Phys. B* 351 (1991), pp. 507–560. DOI: 10.1016/S0550-3213(05)80032-X.
 - [207] P. Nason, S. Dawson, and R. Keith Ellis. “The One Particle Inclusive Differential Cross-Section for Heavy Quark Production in Hadronic Collisions”. In: *Nucl. Phys. B* 327 (1989). [Erratum: Nucl.Phys.B 335, 260–260 (1990)], pp. 49–92. DOI: 10.1016/0550-3213(89)90286-1.

-
- [208] B. A. Kniehl et al. “Inclusive B -meson production at small p_T in the general-mass variable-flavor-number scheme”. In: *Eur. Phys. J. C* 75.3 (2015), p. 140. DOI: 10.1140/epjc/s10052-015-3360-6. arXiv: 1502.01001 [hep-ph].
 - [209] Matteo Cacciari and Mario Greco. “Large p_T hadroproduction of heavy quarks”. In: *Nucl. Phys. B* 421 (1994), pp. 530–544. DOI: 10.1016/0550-3213(94)90515-0. arXiv: hep-ph/9311260.
 - [210] J. Binnewies, Bernd A. Kniehl, and G. Kramer. “Inclusive B meson production in e^+e^- and $p\bar{p}$ collisions”. In: *Phys. Rev. D* 58 (1998), p. 034016. DOI: 10.1103/PhysRevD.58.034016. arXiv: hep-ph/9802231.
 - [211] J. Binnewies, Bernd A. Kniehl, and G. Kramer. “Coherent description of D^{*+} -production in e^+e^- and low Q^2 $e p$ collisions”. In: *Z. Phys. C* 76 (1997), pp. 677–688. DOI: 10.1007/s002880050590. arXiv: hep-ph/9702408.
 - [212] J. Binnewies, Bernd A. Kniehl, and G. Kramer. “Predictions for D^{*+} - photoproduction at HERA with new fragmentation functions from LEP-1”. In: *Phys. Rev. D* 58 (1998), p. 014014. DOI: 10.1103/PhysRevD.58.014014. arXiv: hep-ph/9712482.
 - [213] Matteo Cacciari, Mario Greco, and Paolo Nason. “The $P(T)$ spectrum in heavy flavor hadroproduction”. In: *JHEP* 05 (1998), p. 007. DOI: 10.1088/1126-6708/1998/05/007. arXiv: hep-ph/9803400.
 - [214] Jean-Philippe Lansberg. “New Observables in Inclusive Production of Quarkonia”. In: *Phys. Rept.* 889 (2020), pp. 1–106. DOI: 10.1016/j.physrep.2020.08.007. arXiv: 1903.09185 [hep-ph].
 - [215] Harald Fritzsch. “Producing Heavy Quark Flavors in Hadronic Collisions: A Test of Quantum Chromodynamics”. In: *Phys. Lett. B* 67 (1977), pp. 217–221. DOI: 10.1016/0370-2693(77)90108-3.
 - [216] F. Halzen. “Cvc for Gluons and Hadroproduction of Quark Flavors”. In: *Phys. Lett. B* 69 (1977), pp. 105–108. DOI: 10.1016/0370-2693(77)90144-7.
 - [217] R Aaij et al. “Measurement of $\psi(2S)$ meson production in pp collisions at $\sqrt{s}=7$ TeV”. In: *Eur. Phys. J. C* 72 (2012). [Erratum: *Eur.Phys.J.C* 80, 49 (2020)], p. 2100. DOI: 10.1140/epjc/s10052-012-2100-4. arXiv: 1204.1258 [hep-ex].
 - [218] A. Adare et al. “Ground and excited charmonium state production in $p+p$ collisions at $\sqrt{s} = 200$ GeV”. In: *Phys. Rev. D* 85 (2012), p. 092004. DOI: 10.1103/PhysRevD.85.092004. arXiv: 1105.1966 [hep-ex].
 - [219] Yan-Qing Ma and Ramona Vogt. “Quarkonium Production in an Improved Color Evaporation Model”. In: *Phys. Rev. D* 94.11 (2016), p. 114029. DOI: 10.1103/PhysRevD.94.114029. arXiv: 1609.06042 [hep-ph].

-
- [220] Vincent Cheung and Ramona Vogt. “Production and polarization of prompt J/ψ in the improved color evaporation model using the k_T -factorization approach”. In: *Phys. Rev. D* 98.11 (2018), p. 114029. DOI: 10.1103/PhysRevD.98.114029. arXiv: 1808.02909 [hep-ph].
 - [221] Edmond L. Berger and D. Jones. “Inelastic photoproduction of $\frac{J}{\psi}$ and Υ by gluons”. In: *Phys. Rev. D* 23 (7 1981), pp. 1521–1530. DOI: 10.1103/PhysRevD.23.1521. URL: <https://link.aps.org/doi/10.1103/PhysRevD.23.1521>.
 - [222] D. Acosta et al. “Measurement of the J/ψ meson and b -hadron production cross sections in $p\bar{p}$ collisions at $\sqrt{s} = 1960$ GeV”. In: *Phys. Rev. D* 71 (2005), p. 032001. DOI: 10.1103/PhysRevD.71.032001. arXiv: hep-ex/0412071.
 - [223] Vardan Khachatryan et al. “Prompt and Non-Prompt J/ψ Production in pp Collisions at $\sqrt{s} = 7$ TeV”. In: *Eur. Phys. J. C* 71 (2011), p. 1575. DOI: 10.1140/epjc/s10052-011-1575-8. arXiv: 1011.4193 [hep-ex].
 - [224] Riccardo Barbieri, Raoul Gatto, and E. Remiddi. “Singular Binding Dependence in the Hadronic Widths of 1^{++} and 1^{+-} Heavy Quark anti-Quark Bound States”. In: *Phys. Lett. B* 61 (1976), pp. 465–468. DOI: 10.1016/0370-2693(76)90729-2.
 - [225] Geoffrey T. Bodwin, Eric Braaten, and G. Peter Lepage. “Rigorous QCD analysis of inclusive annihilation and production of heavy quarkonium”. In: *Phys. Rev. D* 51 (1995). [Erratum: Phys.Rev.D 55, 5853 (1997)], pp. 1125–1171. DOI: 10.1103/PhysRevD.55.5853. arXiv: hep-ph/9407339.
 - [226] Yan-Qing Ma, Kai Wang, and Kuang-Ta Chao. “ $J/\psi(\psi')$ production at the Tevatron and LHC at $\mathcal{O}(\alpha_s^4 v^4)$ in nonrelativistic QCD”. In: *Phys. Rev. Lett.* 106 (2011), p. 042002. DOI: 10.1103/PhysRevLett.106.042002. arXiv: 1009.3655 [hep-ph].
 - [227] Yan-Qing Ma, Kai Wang, and Kuang-Ta Chao. “A complete NLO calculation of the J/ψ and ψ' production at hadron colliders”. In: *Phys. Rev. D* 84 (2011), p. 114001. DOI: 10.1103/PhysRevD.84.114001. arXiv: 1012.1030 [hep-ph].
 - [228] Mathias Butenschoen and Bernd A. Kniehl. “Reconciling J/ψ production at HERA, RHIC, Tevatron, and LHC with NRQCD factorization at next-to-leading order”. In: *Phys. Rev. Lett.* 106 (2011), p. 022003. DOI: 10.1103/PhysRevLett.106.022003. arXiv: 1009.5662 [hep-ph].
 - [229] Mathias Butenschoen and Bernd A. Kniehl. “World data of J/ψ production consolidate NRQCD factorization at NLO”. In: *Phys. Rev. D* 84 (2011), p. 051501. DOI: 10.1103/PhysRevD.84.051501. arXiv: 1105.0820 [hep-ph].
 - [230] Pietro Faccioli et al. “Quarkonium production in the LHC era: a polarized perspective”. In: *Phys. Lett. B* 736 (2014), pp. 98–109. DOI: 10.1016/j.physletb.2014.07.006. arXiv: 1403.3970 [hep-ph].

-
- [231] Geoffrey T. Bodwin et al. “Fragmentation contributions to J/ψ production at the Tevatron and the LHC”. In: *Phys. Rev. Lett.* 113.2 (2014), p. 022001. DOI: 10.1103/PhysRevLett.113.022001. arXiv: 1403.3612 [hep-ph].
 - [232] Aleksander Kusina et al. “Reweighted nuclear PDFs using heavy-flavor production data at the LHC”. In: *Phys. Rev. D* 104.1 (2021), p. 014010. DOI: 10.1103/PhysRevD.104.014010. arXiv: 2012.11462 [hep-ph].
 - [233] John Gaiser. “Charmonium Spectroscopy From Radiative Decays of the J/ψ and $\psi(2S)$ ”. In: (). URL: <https://www.slac.stanford.edu/cgi-bin/getdoc/slac-r-255.pdf>.
 - [234] R.K. Ellis, W.J. Stirling, and B.R. Webber. *QCD and Collider Physics*. Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology. Cambridge University Press, 2003. ISBN: 9780521545891. URL: <https://books.google.me/books?id=TqrPVoS6s0UC>.
 - [235] B. Abelev et al. “Measurement of charm production at central rapidity in proton-proton collisions at $\sqrt{s} = 7$ TeV”. In: *JHEP* 01 (2012), p. 128. DOI: 10.1007/JHEP01(2012)128. arXiv: 1111.1553 [hep-ex].
 - [236] R Aaij et al. “Prompt charm production in pp collisions at $\sqrt{s}=7$ TeV”. In: *Nucl. Phys. B* 871 (2013), pp. 1–20. DOI: 10.1016/j.nuclphysb.2013.02.010. arXiv: 1302.2864 [hep-ex].
 - [237] Jaroslav Adam et al. “ D -meson production in p -Pb collisions at $\sqrt{s_{NN}}=5.02$ TeV and in pp collisions at $\sqrt{s}=7$ TeV”. In: *Phys. Rev. C* 94.5 (2016), p. 054908. DOI: 10.1103/PhysRevC.94.054908. arXiv: 1605.07569 [nucl-ex].
 - [238] Georges Aad et al. “Measurement of the differential cross-sections of inclusive, prompt and non-prompt J/ψ production in proton-proton collisions at $\sqrt{s} = 7$ TeV”. In: *Nucl. Phys. B* 850 (2011), pp. 387–444. DOI: 10.1016/j.nuclphysb.2011.05.015. arXiv: 1104.3038 [hep-ex].
 - [239] Serguei Chatrchyan et al. “ J/ψ and ψ_{2S} production in pp collisions at $\sqrt{s} = 7$ TeV”. In: *JHEP* 02 (2012), p. 011. DOI: 10.1007/JHEP02(2012)011. arXiv: 1111.1557 [hep-ex].
 - [240] R. Aaij et al. “Measurement of J/ψ production in pp collisions at $\sqrt{s} = 7$ TeV”. In: *Eur. Phys. J. C* 71 (2011), p. 1645. DOI: 10.1140/epjc/s10052-011-1645-y. arXiv: 1103.0423 [hep-ex].
 - [241] Albert M Sirunyan et al. “Measurement of prompt and nonprompt J/ψ production in pp and pPb collisions at $\sqrt{s_{NN}} = 5.02$ TeV”. In: *Eur. Phys. J. C* 77.4 (2017), p. 269. DOI: 10.1140/epjc/s10052-017-4828-3. arXiv: 1702.01462 [nucl-ex].

-
- [242] Morad Aaboud et al. “Measurement of quarkonium production in proton–lead and proton–proton collisions at 5.02 TeV with the ATLAS detector”. In: *Eur. Phys. J. C* 78.3 (2018), p. 171. DOI: 10.1140/epjc/s10052-018-5624-4. arXiv: 1709.03089 [nucl-ex].
 - [243] Jaroslav Adam et al. “Inclusive quarkonium production at forward rapidity in pp collisions at $\sqrt{s} = 8$ TeV”. In: *Eur. Phys. J. C* 76.4 (2016), p. 184. DOI: 10.1140/epjc/s10052-016-3987-y. arXiv: 1509.08258 [hep-ex].
 - [244] Shreyasi Acharya et al. “Energy dependence of forward-rapidity J/ψ and $\psi(2S)$ production in pp collisions at the LHC”. In: *Eur. Phys. J. C* 77.6 (2017), p. 392. DOI: 10.1140/epjc/s10052-017-4940-4. arXiv: 1702.00557 [hep-ex].
 - [245] Shreyasi Acharya et al. “Inclusive J/ψ production at mid-rapidity in pp collisions at $\sqrt{s} = 5.02$ TeV”. In: *JHEP* 10 (2019), p. 084. DOI: 10.1007/JHEP10(2019)084. arXiv: 1905.07211 [nucl-ex].
 - [246] Georges Aad et al. “Measurement of Upsilon production in 7 TeV pp collisions at ATLAS”. In: *Phys. Rev. D* 87.5 (2013), p. 052004. DOI: 10.1103/PhysRevD.87.052004. arXiv: 1211.7255 [hep-ex].
 - [247] R. Aaij et al. “Measurement of Upsilon production in pp collisions at $\sqrt{s} = 7$ TeV”. In: *Eur. Phys. J. C* 72 (2012), p. 2025. DOI: 10.1140/epjc/s10052-012-2025-y. arXiv: 1202.6579 [hep-ex].
 - [248] Serguei Chatrchyan et al. “Measurement of the $\Upsilon(1S)$, $\Upsilon(2S)$, and $\Upsilon(3S)$ Cross Sections in pp Collisions at $\sqrt{s} = 7$ TeV”. In: *Phys. Lett. B* 727 (2013), pp. 101–125. DOI: 10.1016/j.physletb.2013.10.033. arXiv: 1303.5900 [hep-ex].
 - [249] Betty Bezverkhny Abelev et al. “Measurement of quarkonium production at forward rapidity in pp collisions at $\sqrt{s} = 7$ TeV”. In: *Eur. Phys. J. C* 74.8 (2014), p. 2974. DOI: 10.1140/epjc/s10052-014-2974-4. arXiv: 1403.3648 [nucl-ex].
 - [250] Roel Aaij et al. “Forward production of Υ mesons in pp collisions at $\sqrt{s} = 7$ and 8 TeV”. In: *JHEP* 11 (2015), p. 103. DOI: 10.1007/JHEP11(2015)103. arXiv: 1509.02372 [hep-ex].
 - [251] Albert M Sirunyan et al. “Measurement of prompt $\psi(2S)$ production cross sections in proton-lead and proton-proton collisions at $\sqrt{s_{NN}} = 5.02$ TeV”. In: *Phys. Lett. B* 790 (2019), pp. 509–532. DOI: 10.1016/j.physletb.2019.01.058. arXiv: 1805.02248 [hep-ex].
 - [252] Shreyasi Acharya et al. “Measurement of D^0 , D^+ , D^{*+} and D_s^+ production in pp collisions at $\sqrt{s} = 5.02$ TeV with ALICE”. In: *Eur. Phys. J. C* 79.5 (2019), p. 388. DOI: 10.1140/epjc/s10052-019-6873-6. arXiv: 1901.07979 [nucl-ex].

-
- [253] Shreyasi Acharya et al. “Measurement of D-meson production at mid-rapidity in pp collisions at $\sqrt{s} = 7$ TeV”. In: *Eur. Phys. J. C* 77.8 (2017), p. 550. DOI: 10.1140/epjc/s10052-017-5090-4. arXiv: 1702.00766 [hep-ex].
 - [254] Roel Aaij et al. “Measurements of prompt charm production cross-sections in pp collisions at $\sqrt{s} = 5$ TeV”. In: *JHEP* 06 (2017), p. 147. DOI: 10.1007/JHEP06(2017)147. arXiv: 1610.02230 [hep-ex].
 - [255] “Quarkonium signal extraction in ALICE”. URL: https://cds.cern.ch/record/2060096/files/QuarkoniumSignalExtraction_v5.pdf.
 - [256] Bernd A. Kniehl et al. “Open charm hadroproduction and the charm content of the proton”. In: *Phys. Rev. D* 79 (2009), p. 094009. DOI: 10.1103/PhysRevD.79.094009. arXiv: 0901.4130 [hep-ph].
 - [257] T. Kneesch et al. “Charmed-meson fragmentation functions with finite-mass corrections”. In: *Nucl. Phys. B* 799 (2008), pp. 34–59. DOI: 10.1016/j.nuclphysb.2008.02.015. arXiv: 0712.0481 [hep-ph].
 - [258] R. Aaij et al. “Prompt and nonprompt J/ψ production and nuclear modification in pPb collisions at $\sqrt{s_{NN}} = 8.16$ TeV”. In: *Phys. Lett. B* 774 (2017), pp. 159–178. DOI: 10.1016/j.physletb.2017.09.058. arXiv: 1706.07122 [hep-ex].
 - [259] Shreyasi Acharya et al. “Inclusive J/ψ production at forward and backward rapidity in p-Pb collisions at $\sqrt{s_{NN}} = 8.16$ TeV”. In: *JHEP* 07 (2018), p. 160. DOI: 10.1007/JHEP07(2018)160. arXiv: 1805.04381 [nucl-ex].
 - [260] Georges Aad et al. “Measurement of differential J/ψ production cross sections and forward-backward ratios in p + Pb collisions with the ATLAS detector”. In: *Phys. Rev. C* 92.3 (2015), p. 034904. DOI: 10.1103/PhysRevC.92.034904. arXiv: 1505.08141 [hep-ex].
 - [261] Betty Bezverkhny Abelev et al. “Measurement of prompt D-meson production in p–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV”. In: *Phys. Rev. Lett.* 113.23 (2014), p. 232301. DOI: 10.1103/PhysRevLett.113.232301. arXiv: 1405.3452 [nucl-ex].
 - [262] Roel Aaij et al. “Study of prompt D^0 meson production in pPb collisions at $\sqrt{s_{NN}} = 5$ TeV”. In: *JHEP* 10 (2017), p. 090. DOI: 10.1007/JHEP10(2017)090. arXiv: 1707.02750 [hep-ex].
 - [263] Shreyasi Acharya et al. “Measurement of prompt D^0 , D^+ , D^{*+} , and D_s^+ production in p–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV”. In: *JHEP* 12 (2019), p. 092. DOI: 10.1007/JHEP12(2019)092. arXiv: 1906.03425 [nucl-ex].
 - [264] R Aaij et al. “Study of J/ψ production and cold nuclear matter effects in pPb collisions at $\sqrt{s_{NN}} = 5$ TeV”. In: *JHEP* 02 (2014), p. 072. DOI: 10.1007/JHEP02(2014)072. arXiv: 1308.6729 [nucl-ex].

-
- [265] Betty Bezverkhny Abelev et al. “ J/ψ production and nuclear effects in p-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV”. In: *JHEP* 02 (2014), p. 073. DOI: 10.1007/JHEP02(2014)073. arXiv: 1308.6726 [nucl-ex].
 - [266] Jaroslav Adam et al. “Rapidity and transverse-momentum dependence of the inclusive J/ψ nuclear modification factor in p-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV”. In: *JHEP* 06 (2015), p. 055. DOI: 10.1007/JHEP06(2015)055. arXiv: 1503.07179 [nucl-ex].
 - [267] Betty Bezverkhny Abelev et al. “Production of inclusive $\Upsilon(1S)$ and $\Upsilon(2S)$ in p-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV”. In: *Phys. Lett. B* 740 (2015), pp. 105–117. DOI: 10.1016/j.physletb.2014.11.041. arXiv: 1410.2234 [nucl-ex].
 - [268] Roel Aaij et al. “Study of Υ production and cold nuclear matter effects in pPb collisions at $\sqrt{s_{NN}}=5$ TeV”. In: *JHEP* 07 (2014), p. 094. DOI: 10.1007/JHEP07(2014)094. arXiv: 1405.5152 [nucl-ex].
 - [269] Roel Aaij et al. “Study of Υ production in pPb collisions at $\sqrt{s_{NN}} = 8.16$ TeV”. In: *JHEP* 11 (2018). [Erratum: *JHEP* 02, 093 (2020)], p. 194. DOI: 10.1007/JHEP11(2018)194. arXiv: 1810.07655 [hep-ex].
 - [270] Shreyasi Acharya et al. “ Υ production in p-Pb collisions at $\sqrt{s_{NN}}=8.16$ TeV”. In: *Phys. Lett. B* 806 (2020), p. 135486. DOI: 10.1016/j.physletb.2020.135486. arXiv: 1910.14405 [nucl-ex].
 - [271] Betty Bezverkhny Abelev et al. “Suppression of $\psi(2S)$ production in p-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV”. In: *JHEP* 12 (2014), p. 073. DOI: 10.1007/JHEP12(2014)073. arXiv: 1405.3796 [nucl-ex].
 - [272] Shreyasi Acharya et al. “Measurement of nuclear effects on $\psi(2S)$ production in p-Pb collisions at $\sqrt{s_{NN}} = 8.16$ TeV”. In: *JHEP* 07 (2020), p. 237. DOI: 10.1007/JHEP07(2020)237. arXiv: 2003.06053 [nucl-ex].
 - [273] Sanghwa Park et al. “CJ15 global PDF analysis with new electroweak data from the STAR and SeaQuest experiments”. In: *28th International Workshop on Deep Inelastic Scattering and Related Subjects*. 2021. arXiv: 2108.05786 [hep-ph].
 - [274] Richard D. Ball et al. “The Path to Proton Structure at One-Percent Accuracy”. In: (2021). arXiv: 2109.02653 [hep-ph].
 - [275] Kari J. Eskola et al. “Proton-PDF uncertainties in extracting nuclear PDFs from W^\pm production in p+Pb collisions”. In: *Eur. Phys. J. C* 82.3 (2022), p. 271. DOI: 10.1140/epjc/s10052-022-10179-2. arXiv: 2202.01074 [hep-ph].

-
- [276] Albert M Sirunyan et al. “Measurement of nuclear modification factors of $\Upsilon(1S)$, $\Upsilon(2S)$, and $\Upsilon(3S)$ mesons in PbPb collisions at $\sqrt{s_{NN}} = 5.02$ TeV”. In: *Phys. Lett. B* 790 (2019), pp. 270–293. DOI: 10.1016/j.physletb.2019.01.006. arXiv: 1805.09215 [hep-ex].
 - [277] Albert M Sirunyan et al. “Measurement of prompt and nonprompt charmonium suppression in PbPb collisions at 5.02 TeV”. In: *Eur. Phys. J. C* 78.6 (2018), p. 509. DOI: 10.1140/epjc/s10052-018-5950-6. arXiv: 1712.08959 [nucl-ex].
 - [278] Albert M Sirunyan et al. “Nuclear modification factor of D^0 mesons in PbPb collisions at $\sqrt{s_{NN}} = 5.02$ TeV”. In: *Phys. Lett. B* 782 (2018), pp. 474–496. DOI: 10.1016/j.physletb.2018.05.074. arXiv: 1708.04962 [nucl-ex].
 - [279] John M. Campbell et al. “Direct photon production and PDF fits reloaded”. In: *Eur. Phys. J. C* 78.6 (2018), p. 470. DOI: 10.1140/epjc/s10052-018-5944-4. arXiv: 1802.03021 [hep-ph].
 - [280] M. Walt, I. Helenius, and W. Vogelsang. “Open-source QCD analysis of nuclear parton distribution functions at NLO and NNLO”. In: *Phys. Rev. D* 100.9 (2019), p. 096015. DOI: 10.1103/PhysRevD.100.096015. arXiv: 1908.03355 [hep-ph].
 - [281] G. Apollinari et al. “High Luminosity Large Hadron Collider HL-LHC”. In: *CERN Yellow Rep.* 5 (2015). Ed. by G Apollinari et al., pp. 1–19. DOI: 10.5170/CERN-2015-005.1. arXiv: 1705.08830 [physics.acc-ph].
 - [282] Jonathan L. Feng et al. “The Forward Physics Facility at the High-Luminosity LHC”. In: (2022). arXiv: 2203.05090 [hep-ex].
 - [283] R. Abdul Khalek et al. “Science Requirements and Detector Concepts for the Electron-Ion Collider: EIC Yellow Report”. In: (2021). arXiv: 2103.05419 [physics.ins-det].
 - [284] J. P. Ellis. “TikZ-Feynman: Feynman diagrams with TikZ”. In: *Comput. Phys. Commun* 210 (2017). DOI: 10.1016/j.cpc.2016.08.019.
 - [285] M. Arneodo et al. “Measurement of the proton and deuteron structure functions, $F_2(p)$ and $F_2(d)$, and of the ratio σ_L / σ_T ”. In: *Nucl. Phys. B* 483 (1997), pp. 3–43. DOI: 10.1016/S0550-3213(96)00538-X. arXiv: hep-ph/9610231.
 - [286] A. Airapetian et al. “Measurement of $R = \sigma_L / \sigma_T$ in deep inelastic scattering on nuclei”. In: (2002). arXiv: hep-ex/0210068.
 - [287] P. Amaudruz et al. “A Reevaluation of the nuclear structure function ratios for D, He, Li-6, C and Ca”. In: *Nucl. Phys. B* 441 (1995), pp. 3–11. DOI: 10.1016/0550-3213(94)00023-9. arXiv: hep-ph/9503291.
 - [288] M. Arneodo et al. “The Structure Function ratios $F_2(\text{li}) / F_2(D)$ and $F_2(C) / F_2(D)$ at small x ”. In: *Nucl. Phys. B* 441 (1995), pp. 12–30. DOI: 10.1016/0550-3213(95)00023-2. arXiv: hep-ex/9504002.

-
- [289] M. R. Adams et al. “Shadowing in inelastic scattering of muons on carbon, calcium and lead at low $x(\text{Bj})$ ”. In: *Z. Phys. C* 67 (1995), pp. 403–410. DOI: 10.1007/BF01624583. arXiv: hep-ex/9505006.
 - [290] J. Ashman et al. “Measurement of the Ratios of Deep Inelastic Muon - Nucleus Cross-Sections on Various Nuclei Compared to Deuterium”. In: *Phys. Lett. B* 202 (1988), pp. 603–610. DOI: 10.1016/0370-2693(88)91872-2.
 - [291] M. Arneodo et al. “Measurements of the nucleon structure function in the range $0.002 - \text{GeV}^2 < x < 0.17 - \text{GeV}^2$ and $0.2 - \text{GeV}^2 < q^2 < 8 - \text{GeV}^2$ in deuterium, carbon and calcium”. In: *Nucl. Phys. B* 333 (1990), pp. 1–47. DOI: 10.1016/0550-3213(90)90221-X.
 - [292] A. Bodek et al. “Electron Scattering from Nuclear Targets and Quark Distributions in Nuclei”. In: *Phys. Rev. Lett.* 50 (1983), p. 1431. DOI: 10.1103/PhysRevLett.50.1431.
 - [293] A. C. Benvenuti et al. “Nuclear Effects in Deep Inelastic Muon Scattering on Deuterium and Iron Targets”. In: *Phys. Lett. B* 189 (1987), pp. 483–487. DOI: 10.1016/0370-2693(87)90664-2.
 - [294] J. Ashman et al. “A Measurement of the ratio of the nucleon structure function in copper and deuterium”. In: *Z. Phys. C* 57 (1993), pp. 211–218. DOI: 10.1007/BF01565050.
 - [295] M. R. Adams et al. “Saturation of shadowing at very low x_{BJ} ”. In: *Phys. Rev. Lett.* 68 (1992), pp. 3266–3269. DOI: 10.1103/PhysRevLett.68.3266.
 - [296] M. Arneodo et al. “The A dependence of the nuclear structure function ratios”. In: *Nucl. Phys. B* 481 (1996), pp. 3–22. DOI: 10.1016/S0550-3213(96)90117-0.
 - [297] M. Arneodo et al. “The Q^{*2} dependence of the structure function ratio $F_2^{\text{Sn}} / F_2^{\text{C}}$ and the difference $R^{\text{Sn}} - R^{\text{C}}$ in deep inelastic muon scattering”. In: *Nucl. Phys. B* 481 (1996), pp. 23–39. DOI: 10.1016/S0550-3213(96)90119-4.
 - [298] M. A. Vasilev et al. “Parton energy loss limits and shadowing in Drell-Yan dimuon production”. In: *Phys. Rev. Lett.* 83 (1999), pp. 2304–2307. DOI: 10.1103/PhysRevLett.83.2304. arXiv: hep-ex/9906010.
 - [299] Georges Aad et al. “Measurement of W^\pm boson production in Pb+Pb collisions at $\sqrt{s_{\text{NN}}} = 5.02 \text{ TeV}$ with the ATLAS detector”. In: *Eur. Phys. J. C* 79.11 (2019), p. 935. DOI: 10.1140/epjc/s10052-019-7439-3. arXiv: 1907.10414 [nucl-ex].
 - [300] Vardan Khachatryan et al. “Study of W boson production in pPb collisions at $\sqrt{s_{\text{NN}}} = 5.02 \text{ TeV}$ ”. In: *Phys. Lett. B* 750 (2015), pp. 565–586. DOI: 10.1016/j.physletb.2015.09.057. arXiv: 1503.05825 [nucl-ex].

-
- [301] Albert M Sirunyan et al. “Observation of nuclear modifications in W^\pm boson production in pPb collisions at $\sqrt{s_{NN}} = 8.16$ TeV”. In: *Phys. Lett. B* 800 (2020), p. 135048. DOI: 10.1016/j.physletb.2019.135048. arXiv: 1905.01486 [hep-ex].
- [302] Jaroslav Adam et al. “W and Z boson production in p-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV”. In: *JHEP* 02 (2017), p. 077. DOI: 10.1007/JHEP02(2017)077. arXiv: 1611.03002 [nucl-ex].
- [303] Kgotlaesele Senosi. “Measurement of W-boson production in p-Pb collisions at the LHC with ALICE”. In: *PoS Bormio2015* (2015), p. 042. DOI: 10.22323/1.238.0042. arXiv: 1511.06398 [hep-ex].
- [304] Georges Aad et al. “Z boson production in p +Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV measured with the ATLAS detector”. In: *Phys. Rev. C* 92.4 (2015), p. 044915. DOI: 10.1103/PhysRevC.92.044915. arXiv: 1507.06232 [hep-ex].
- [305] Vardan Khachatryan et al. “Study of Z boson production in pPb collisions at $\sqrt{s_{NN}} = 5.02$ TeV”. In: *Phys. Lett. B* 759 (2016), pp. 36–57. DOI: 10.1016/j.physletb.2016.05.044. arXiv: 1512.06461 [hep-ex].
- [306] R. Aaij et al. “Observation of Z production in proton-lead collisions at LHCb”. In: *JHEP* 09 (2014), p. 030. DOI: 10.1007/JHEP09(2014)030. arXiv: 1406.2885 [hep-ex].