THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : PHYS - Physique
Spécialité : Physique Théorique
Unité de recherche : Laboratoire de Physique Subatomique et Cosmologie

Amélioration de la précision des PDF nucléaires avec nCTEQ : corrections liées à la masse de la cible en DIS et production de photons prompts

Improving the precision of nuclear PDFs in the nCTEQ framework: target mass corrections in DIS and the prompt photon production

Présentée par :

Chloé LEGER

Direction de thèse :

Ingo SCHIENBEIN — Directeur de thèse
MAITRE DE CONFERENCES HDR, UNIVERSITE GRENOBLE ALPES

Rapporteurs :

FRANÇOIS ARLEO
CHARGE DE RECHERCHE HDR, CNRS BRETAGNE ET PAYS DE LA LOIRE
JEAN-PHILIPPE LANSBERG
DIRECTEUR DE RECHERCHE, CNRS DELEGATION ILE-DE-FRANCE SUD

Thèse soutenue publiquement le **4 décembre 2025**, devant le jury composé de :

CHRISTOPHE FURGET, PROFESSEUR DES UNIVERSITES, UNIVERSITE GRENOBLE ALPES	Président
FRANÇOIS ARLEO, CHARGE DE RECHERCHE HDR, CNRS BRETAGNE ET PAYS DE LA LOIRE	Rapporteur
JEAN-PHILIPPE LANSBERG, DIRECTEUR DE RECHERCHE, CNRS DELEGATION ILE-DE-FRANCE SUD	Rapporteur
CYNTHIA HADJIDAKIS, DIRECTRICE DE RECHERCHE, CNRS DELEGATION ILE-DE-FRANCE SUD	Examinatrice
MARIANE MANGIN-BRINET, DIRECTRICE DE RECHERCHE, CNRS DELEGATION ALPES	Examinatrice

Contents

Introduction and Survey **1**

1 Theoretical framework/background **4**

1.1 A short introduction to QCD 5

1.1.1 Lagrangian of QCD 6

1.1.2 Strong coupling 8

1.2 Basics of Deep Inelastic Scattering and the Parton Model 11

1.2.1 Deeply Inelastic Scattering 11

1.2.2 The Naive Parton Model 15

1.2.3 Limits and expectations 16

1.2.4 Operator Product Expansion approach 17

1.3 Perturbative QCD 18

1.3.1 Factorization Theorems for hard processes 18

1.3.2 Hard part 20

1.3.3 DGLAP evolution 22

1.3.4 Sum rules and other properties of PDFs 22

1.4 QCD Global Analysis of PDFs 23

1.4.1 General framework 23

1.4.2 Nuclear PDFs 24

1.4.3 nCTEQ framework 26

2 Target Mass Corrections to Deep Inelastic Scattering **29**

2.1 Introduction 29

2.2 Kinematics of lepton-nucleus DIS 36

2.2.1 Definition of kinematic variables 36

2.2.2 DIS cross sections 39

2.3 Nuclear structure functions in the OPE 42

2.3.1 Light-cone dominance in the nuclear case. 42

2.3.2 Structure functions in the OPE in a nutshell . . . 45
2.3.3 Master formulas . . . 50
2.4 Rescaling . . . 52
2.4.1 Kinematics . . . 52
2.4.2 Re-scaled Structure Functions . . . 53
2.4.3 Impact of quark masses . . . 55
2.5 Focus on Parton Model in the (massless) nuclear case . . . 57
2.5.1 Nuclear DGLAP evolution . . . 58
2.5.2 Relation to the OPE . . . 59
2.5.3 Rescaling . . . 59
2.5.4 Kinematic $\boldsymbol{W}$ cut . . . 63
2.5.5 nPDFs for $x_N \geq 1$ . . . 66
2.5.6 Threshold problem and higher twist contributions . . . 70
2.6 Relations with parton model . . . 70
2.6.1 Helicity formulation of the parton model . . . 71
2.6.2 Helicity decomposition . . . 72
2.6.3 Boost-invariant polarizations . . . 75
2.6.4 Relationship between TMCs on the light-front and in the OPE . . . 76
2.6.5 Relation of the parton model to the OPE . . . 78
2.7 Numerical results . . . 79
2.7.1 Proton structure functions with TMCs . . . 80
2.7.2 Nuclear structure functions with TMCs . . . 83
2.7.3 Parameterizing $F_i^{\mathrm{TMC}}/F_i^{\mathrm{Leading-TMC}}$ . . . 92
2.7.4 Reduced cross sections with TMCs for nuclei . . . 96
2.8 Conclusions . . . 99

3 Inclusive prompt photon production 102
3.1 Theoretical framework . . . 103
3.1.1 Definitions and formalism of the prompt photon production process 103
3.1.2 Kinematics . . . 105
3.1.3 Isolated cross section . . . 108
3.1.4 Born Level . . . 111
3.1.5 Fragmentation . . . 113
3.1.6 Higher-orders terms . . . 114
3.1.7 Conclusion . . . 115
3.2 Theoretical computations . . . 115

3.2.1 Jetphox . . . 116
3.2.1.1 Presentation of the code . . . 116
3.2.1.2 Technical features . . . 117
3.2.1.3 Calculation of the observables . . . 118
3.2.2 MCFM . . . 121
3.2.2.1 Presentation of the code . . . 121
3.2.2.2 Calculations . . . 122
3.2.3 Other calculations in the literature . . . 122
3.3 Available data and validation of theoretical calculations . . . 125
3.3.1 Proton-proton collisions at different center-of-mass energies . . . 126
3.3.2 Proton-lead collisions . . . 132
3.4 Conclusions . . . 134

4 Fast calculation of prompt photon production cross-sections with APPLGRID 135
4.1 General purpose of APPLGRID . . . 136
4.1.1 ***What is a grid?*** . . . 136
4.1.2 Simple case . . . 136
4.1.3 Application to direct photon production . . . 139
4.2 Grids with MCFM . . . 140
4.2.1 MCFM interface . . . 140
4.2.2 Grids with MCFM for *p-Pb* collisions . . . 143
4.2.3 Comparison of MCFM with JETPHOX . . . 144
4.3 Grids with Jetphox . . . 146
4.3.1 Introduction . . . 146
4.3.2 Interface and installation . . . 147
4.3.3 Using the interface . . . 148
4.3.4 Modifications in JETPHOX . . . 149
4.3.5 Results for the Born term . . . 150
4.3.6 Outlook for NLO grids and fragmentation . . . 151

Conclusions and Outlook 153

A Jetphox materials 156

B Applgrid 171

Introduction and Survey

The twentieth century witnessed profound revolutions in our understanding of nature, both at the smallest scales of matter and at the largest scales of the cosmos. These revolutions did more than expand existing ideas; they changed how we understand reality, requiring us to move beyond classical intuitions and adopt new frameworks centered on probability, symmetry and its breaking, and quantum structure. This shift was most evident in the subatomic world, where the development of quantum mechanics in the 1920s offered the first consistent way to describe phenomena that classical physics could not explain. The combination of quantum mechanics with special relativity then led to quantum field theory, the framework in which particle physics is studied.

As the ambition to probe deeper into the laws of matter grew, experimentalists pushed technology to its limits by constructing increasingly powerful particle accelerators. The Stanford Linear Accelerator (SLAC) was one such facility that played a crucial role in advancing the study of particle physics, enabling investigations of the internal structure of matter through high-energy collisions. These experiments revealed that hadrons were not elementary but instead exhibited an inner composition, pointing toward more fundamental building block. The theoretical response to these discoveries coalesced into the Standard Model (SM) of particle physics, a gauge-theoretic framework that unifies the electromagnetic, weak, and strong interactions. Within this structure, matter is organized into quarks and leptons, while forces are mediated by gauge bosons.

Among these interactions, the strong force remained the most enigmatic. Its peculiar property of confining quarks within hadrons still resists understanding even with the advent of Quantum Chromodynamics (QCD) in the 1970s. QCD, a non-Abelian gauge theory based on color charge, does not (yet?) provide a proof of confinement, but predicted a rich dynamical structure that continues to challenge both theory and computations. Conversely, the asymptotic freedom regime in QCD, in a way, made it possible to overcome this problem by allowing us to learn about quarks and gluons, even though they cannot be observed individually.

A crucial window into QCD dynamics is provided by Deep Inelastic Scattering (DIS).

With the enhanced energies achieved at SLAC and subsequent facilities, physicists could probe the nucleon at sufficiently short distances to expose its partonic substructure. Through DIS, the interplay of quarks and gluons was revealed with increasing clarity, leading to the formulation of Parton Distribution Functions (PDFs) as indispensable tools, even today, for connecting the underlying theory with measurable cross sections. The extension of this framework to nuclei, involving nuclear PDFs (nPDFs) and their refinements, remains a central pursuit in contemporary high-energy physics. Even more importantly, the accuracy of predictions is fundamental. At the Large Hadron Collider (LHC), precision predictions for processes such as Higgs production, electroweak boson production, and jets depend critically on PDFs. In the nuclear case, proton–lead and heavy-ion collisions provide unique information about the modification of partonic distributions in nuclei, with implications for quark–gluon plasma studies and background estimates in searches for new physics. At lower energies, Jefferson Lab continues to refine our knowledge of nuclear structure through precision electron scattering, while the future Electron–Ion Collider (EIC) promises unprecedented opportunities to map the partonic content of nuclei across wide kinematic ranges. The consistent treatment of nuclear effects, including Target Mass Corrections (TMCs) is indispensable for extracting reliable nuclear PDFs.

Another powerful probe of QCD dynamics is the prompt photon production in hadronic collisions. Photons, being colorless, escape the strongly interacting medium with little further modification, thereby providing direct access to the dynamics of the hard scattering. At the LHC, prompt photons are especially sensitive to the gluon distribution in the proton and in the nucleus, making them valuable inputs to global PDF and nPDF analyses. However, the theoretical description of photon production involves subtle issues such as direct versus fragmentation contributions, isolation criteria, and higher-order corrections. In practice, tools like JETPHOX and MCFM exist and are able to compute this process at next-to-leading order, and their validation against data is essential for reliable phenomenology.

Finally, the growing precision of collider data and the breadth of processes included in global fits have made computational efficiency a key challenge. Direct evaluation of higher-order QCD cross sections is too costly for iterative fits of PDFs. Grid-based methods such as APPLGRID have therefore become indispensable, allowing for the storage of perturbative coefficients in a form that can be quickly convolved with PDFs during the fitting procedure. Extending these methods to processes like prompt photon production, and ensuring their compatibility with modern frameworks such as nCTEQ++, represents an important step

in advancing global nuclear PDF analyses.

This thesis is organized as follows: Chapter 1 introduces the theoretical framework, emphasizing the role of Parton Distribution Functions and their importance in precision phenomenology. Chapter 2 then clarifies the differences between nucleon and nuclear structure, examining the formulation of nuclear PDFs and developing a new treatment of Target Mass Corrections in lepton-nucleus DIS for global analysis. Chapter 3 turns to the phenomenology of prompt photon production in hadronic collision (proton-proton and proton-lead) at the LHC, with an emphasis on including these data into the nCTEQ global fit. Finally, Chapter 4 presents a grid method (APPLGRID) that allows for fast calculation of cross-sections for the prompt photon process. We introduce the method for MCFM and present a new interface with JETPHOX. In addition, Appendix A contains some code elements from JETPHOX, used in Chapter 3, as well as tables of results corresponding to plots. Appendix B contains code elements related to Chapter 4.

List of publications/contributions:

- R. Ruiz, K. F. Muzakka, C. Léger, P. Risse, A. Accardi, P. Duwentäster, T. J. Hobbs, T. Ježo, C. Keppel and M. Klasen, *et al.* **"Target mass corrections in lepton–nucleus DIS: Theory and applications to nuclear PDFs,"** Prog. Part. Nucl. Phys.(2024), doi:10.1016/j.ppnp.2023.104096 [arXiv:2301.07715 [hep-ph]]. (Chapter 2)

- P. Risse, N. Derakhshanian, P. Duwentäster, T. Ježo, C. Keppel, M. Klasen, K. Kovařík, A. Kusina, C. Léger and J. G. Morfín, *et al.* **"Towards a New nCTEQ global nPDF release,"** [arXiv:2307.07814 [hep-ph]] .

Chapter 1

Theoretical framework/background

Introduction

The purpose of this chapter is to provide the theoretical background necessary to understand the rest of this thesis which aims at improving our understanding of the parton content of hadrons, especially in the proton and nuclei. A particular focus is on the gluon content and the question how much it can be constrained by LHC data on prompt photon production in proton-lead collisions.

The notion of a parton has been first coined by Richard Feynman in 1969 when he introduced his parton model [1] as a means to analyze high-energy hadron collisions. At high energies, any hadron (e.g., a proton or neutron) can be considered as a collection of point-like, essentially non-interacting, constituents called **partons**. This model was then successfully applied to explain SLAC data on deep-inelastic scattering (DIS) by James D. Bjorken and Emmanuel A. Paschos [2]. It should be noted that the parton model was introduced before the formulation of Quantum-Chromo-Dynamics (QCD) as the Quantum Field Theory of strong interaction dynamics in 1973 [3–6]. It was then soon realized that the partons could be identified with the quarks and gluons of QCD.

Furthermore, QCD corrections were applied to Feynman's *naive* parton model leading to the QCD-improved parton model which was later validated by rigorous factorization theorems [7–9] derived from first principles in QCD. These factorization theorems constitute the underlying theoretical framework for the work in this thesis and for the calculation of a wide class of hadron collider observables in general. In particular, they provide field theoretical definitions of parton densities and precise formulae how to calculate scattering processes in hadron-hadron or lepton-hadron collisions. They are therefore of crucial importance.

This chapter is inspired by the CTEQ handbook of QCD [10] and the lecture notes

by P.Aurenche et al. [11] and the references therein, and is organized as follows: Sec. 1.1 is devoted to an introduction to QCD, with the description of the QCD Lagrangian in Sec. 1.1.1, followed by a discussion of the strong coupling and asymptotic freedom in Sec. 1.1.2.

Next, in Sec. 1.2 we present, through a rather historical development, the basis of deep inelastic scattering (Sec. 1.2.1) and the original idea proposed by Feynman with the (naive) Parton Model (Sec. 1.2.2). We also present the limitations of the last two sections (Sec. 1.2.3) as well as the contribution of QCD. We finally outline the Operator Product Expansion (Sec. 1.2.4) to help linking all the concepts.

In Sec. 1.3, we outline the perturbative QCD framework. We first introduce the concepts of factorization (Sec. 1.3.1) and we then explore the two main ingredients of the theory, the hard coefficients (Sec. 1.3.2) and the parton distributions functions together with the DGLAP evolution equations (Sec. 1.3.3). Finally, some of their other properties are addressed in Sec. 1.3.4.

In Sec. 1.4, we will present the standard aspects of a QCD global analysis of PDFs, first with a general overview (Sec. 1.4.1), while in Sec. 1.4.2 we introduce the nuclear corrections for PDFs and then we continue with the specific nCTEQ framework for nuclear PDFs (Sec. 1.4.3) which is used throughout this thesis.

1.1 A short introduction to QCD

Quantum chromodynamics (QCD) [3–5] is a quantum field theory of the strong interaction which is one of the three fundamental interactions contained in the Standard Model (SM) of particle physics. The SM summarizes our current understanding of Nature at the most microscopic level and is based on a few (symmetry) principles.

The interactions between the particles in the SM are (almost) entirely fixed by a fundamental principle known as local gauge invariance. More specifically, the action of the SM is invariant under transformations of the SM gauge group:

$$SU(3)_c \times SU(2)_L \times U_Y(1)\,.$$

The $SU(2)_L \times U_Y(1)$ group describes the electroweak (EW) interactions where $SU(2)_L$ is the weak isospin group and $U_Y(1)$ is the hypercharge group. This EW group is spontaneously broken at an energy $E \sim 250$ GeV and the residual symmetry group, $U_{EM}(1)$, describes the Quantum Electro-Dynamics (QED). Its generator, $Q = T_3 + Y/2$, is a combination of the generators T_3 of $SU(2)_L$ and Y of $U_Y(1)$. The corresponding gauge boson

is the **photon**. The characteristic coupling, at zero energy, is $\alpha_{em} \simeq 1/137 \ll 1$ allowing us to perform perturbative calculations. Finally, the Abelian nature of QED simplifies very much the calculations and that's why this theory is very precisely tested and experimentally validated. This is a major difference with Quantum Chromo-Dynamics (QCD).

QCD is governed by the basic principle of Poincaré invariance, the $SU(3)$-color ($SU(3)_c$) gauge invariance with its associated gauge bosons, the **gluons**, and renormalizability[1]. More generally, any relativistic quantum field theory can be defined by providing information on (i) the field content including the quantum numbers (spin, electric charge, color, etc.), (ii) a Lagrangian describing the local interactions, and (iii) the values of the input parameters (masses, couplings, etc.) of the Lagrangian. The values of the input parameters require experimental information. Despite the greater complexity, largely due to its non-Abelian nature, this theory is validated by many experiments. We describe in detail the QCD Lagrangian in the next section.

1.1.1 Lagrangian of QCD

The Lagrangian of QCD is given by

$$\mathcal{L}_{QCD} = \mathcal{L}_F + \mathcal{L}_G + \mathcal{L}_{GF} + \mathcal{L}_{FP} \,, \tag{1.1}$$

where the first two terms ($\mathcal{L}_F + \mathcal{L}_G$) are known as the Yang-Mills Lagrangian while the last two terms are introduced during quantization. $\mathcal{L}_F$ is the fermionic part and provides the interaction between quarks and gluons:

$$\mathcal{L}_F = \bar{\psi}(i\not{D} - m_q)\psi \,, \tag{1.2}$$

where $\psi = (\psi_i)$ are the quark fields with $i = 1, 2, 3$ the color index. A sum over different quark flavours is implicit ($\sum_{\psi=u,d,s,c,b,t}$) and m_q is the quark mass. Next, $\not{D} = \gamma^\mu D_\mu$ is the contraction of the Dirac gamma matrices $\gamma_{\mu=0,1,2,3}$ and the covariant derivative D_μ, where

$$D_\mu = \partial_\mu \mathbf{1} + igT_a G^a_\mu \tag{1.3}$$

is given in the 3-dimensional fundamental representation. G^a_μ are the eight ($a = 1, \ldots, 8$) gluon fields and T_a are the eight ($a = 1, \ldots, 8$) generators of the $SU(3)_c$ gauge symmetry in the fundamental representation. $\mathcal{L}_G$ is the gauge kinetic term for the gluons and contains

[1]Note that these three principles (Poincaré invariance, gauge invariance, and renormalizability) do not forbid a CP violating term in the Lagrangian which is phenomenologically known to be absent or extremely small.

the self coupling of the gluons, with three and four gluon vertices:

$$\mathcal{L}_G = -\frac{1}{4} G^a_{\mu\nu} G^{\mu\nu}_a . \tag{1.4}$$

Here, $G^a_{\mu\nu}$ is the gluonic field strength tensor and is defined as

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu - g f_{abc} G^b_\mu G^c_\nu \,, \tag{1.5}$$

where f_{abc} are the structure constants of the $SU(3)_c$ algebra and g the gauge coupling. For completeness, the gauge fixing term, $\mathcal{L}_{GF}$ is defined as

$$\mathcal{L}_{GF} = \begin{cases} \dfrac{-1}{2\xi} (\partial^\mu G^a_\mu)^2, & \text{(covariant gauge)} \\ \\ \dfrac{-1}{2\xi} \left(n^\mu G^a_\mu \right)^2, & \text{(axial gauge)} \end{cases} \tag{1.6}$$

and comes from the quantization process, and ξ is the gauge parameter. $\mathcal{L}_{FP}$ is the Faddeev-Popov term, introducing ghost fields and is given by

$$\mathcal{L}_{FP} = \begin{cases} -\chi^{*a} \partial^\mu (\mathcal{D}^{ab}_\mu \chi^b) = (\partial^\mu \chi^{*a}) \mathcal{D}^{ab}_\mu \chi^b, & \text{(covariant gauge)} \\ \\ -\chi^{*a} (n^\mu \mathcal{D}_\mu)^{ab} \chi^b = -\chi^{*a} (n \cdot \partial - g f_{abc}\, n \cdot G^c) \chi^b, & \text{(axial gauge)} \end{cases} \tag{1.7}$$

where χ are ghost fields. In an axial gauge, $n \cdot G^c = 0$ and the interaction between ghosts with gluons and other physical degrees of freedom is canceled, such that ghosts can be ignored. These terms are needed to define QCD consistently at the quantum level.

Finally, $\mathcal{L}_{QCD}$ depends on eight parameters: the six quark masses ($m_{u,d,s,c,b,t}$), the gauge fixing parameter (ξ) and the gauge coupling (g). Obviously, the observables should not depend on the chosen gauge and hence on ξ. The gauge coupling g will be described in detail in Sec. 1.1.2.

In order to make predictions, and ultimately test the theory, it is necessary to provide values for these parameters. Currently there is no established theory beyond the standard model allowing to predict these parameters and for this reason, they have to be determined by experiment. However we have to deal with these two issues:

- The degrees of freedom associated with the QCD Lagrangian are not directly observable in experiments, as quarks and gluons are confined within color-neutral hadrons.

When quarks and gluons are produced in high-energy collisions, they rapidly undergo hadronization, forming non-colored hadrons[2]. These hadrons are subsequently detected and measured by experimental apparatus.

- The strong coupling is much bigger than the one of the electromagnetic interaction, when evaluated at the scale of the nucleon mass. This implies that perturbative techniques are not systematically applicable for the strong interaction calculations, in contrast to Quantum ElectroDynamics (QED), where perturbative methods are effective due to the smaller coupling constant ($\alpha \simeq 1/137$).

In addition, due to the non-abelian nature of QCD, calculations are generally much more involved than in QED. Despite the inherent challenges in this theory, we can test systematically QCD due to two crucial features; the **asymptotic freedom** (see 1.1.2) and the **factorization theorems** (see 1.3.1).

1.1.2 Strong coupling

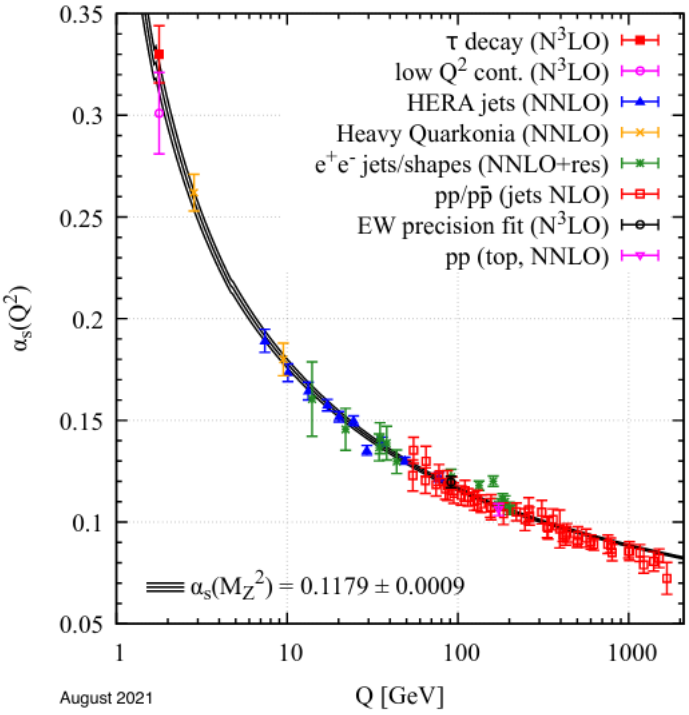


Figure 1.1: The coupling $\alpha_s(Q)$ as a function of energy Q. Figure from the PDG [12].

The QCD successes in describing strong interactions are attributed to asymptotic freedom (weak short-distance interaction) and confinement (strong long-distance interac-

[2]With the exception of the top quark which decays before hadronization.

tion). Despite the fact that hadrons are well-described by the quark model[3], **free** quarks have never been observed in scattering experiments. Paradoxically, some high-energy cross-sections are successfully described by non-interacting point-like particles. Historically, this behavior has been observed before the formulation of QCD, and was the basic concept that lead to the Parton Model, which we shall describe below in Sec. 1.2.2.
In the definition of the Lagrangian in Sec. 1.1.1, the gauge coupling g was presented as a constant, but is actually a running coupling constant. The relation between g and α_s is defined as (in natural units)

$$\alpha_s = \frac{g^2}{4\pi} \,. \tag{1.8}$$

Multiple experiments made it possible to extract α_s and its behavior can be seen in Fig. 1.1. The Q-dependence emerges from the renormalization of the Lagrangian parameters and the strong coupling is then governed by Renormalization-Group-Equations (RGEs). The coupling follows

$$\beta(\alpha_s) = \frac{d\alpha_s}{d\ln\mu^2} \,, \tag{1.9}$$

where μ is an energy scale, called the renormalization scale (μ_R). A particularity of QCD (in contrast to QED) is the fact that gluons carry a color charge and can interact between them. The consequence of this self-coupling is that the β function is **negative** in QCD, in the small coupling region. Asymptotically, in the limit where the energy tends to infinity, the coupling tends to zero and QCD becomes a free theory. This is what we call the **asymptotic freedom**. The β function can then be expanded in a perturbative series as

$$\beta(\alpha_s) = -\frac{\beta_0}{4\pi^2}\alpha_s^2 - \frac{\beta_1}{(4\pi)^3}\alpha_s^3 - \ldots \quad = -\sum_{n=0}^{\infty}\left(\frac{\alpha_s}{4\pi}\right)^{n+2}\beta_n \,. \tag{1.10}$$

The β_n functions are currently known up to five loops ($n = 4$) [13] and

$$\beta_0 = \frac{11N - 2N_f}{12\pi} \,. \tag{1.11}$$

For example, at LO, the coupling at two arbitrary scales (μ, μ_0) is given by

$$\alpha_s(\mu^2) = \frac{\alpha_s(\mu_0^2)}{1 + \frac{\alpha_s(\mu_0^2)}{12\pi}(11N - 2N_f)\ln(\frac{\mu^2}{\mu_0^2})} \,, \tag{1.12}$$

where N and N_f are respectively the color number ($N = 3$) and number of active quark flavors ($3 \leq N_f \leq 6$). More precisely (with quark masses in the $\overline{MS}$ scheme),

- $N_f = 3$ for $Q < m_c$ where $m_c \sim 1.27$ GeV is the c- quark mass,

[3]Ignoring recent observations such as tetra- and penta-quarks.

- $N_f = 4$ for $m_c \leq Q < m_b$ where $m_b \sim 4.18$ GeV is the b-quark mass,
- $N_f = 5$ for $m_b \leq Q < m_t$ where $m_t \sim 163$ GeV is the t-quark mass,
- $N_f = 6$ for $m_t \leq Q$.

Renormalization gives rise to a characteristic scale of the theory, Λ_{QCD}. Still at LO, we can rearrange Eq. (1.12) as

$$\alpha_s(\mu^2) = \frac{12\pi}{(11N - 2N_f)\ln(\frac{\mu^2}{\Lambda^2_{QCD}})}\,, \tag{1.13}$$

where

$$\Lambda^2_{QCD} = \mu_0^2 \exp\left(\frac{-12\pi}{(11N - 2N_f)\alpha_s(\mu_0^2)}\right). \tag{1.14}$$

At NNLO with $N_f = 5$, $\Lambda_{QCD} = 210 \pm 11$ MeV and is determined experimentally [14][4]. As can be seen in Eq. (1.13), when $\mu \to \Lambda_{QCD}$, α_s diverges. This is called the Landau pole of QCD and the coupling tends to infinity; obviously the perturbative techniques can't be employed in this kinematical domain. Evaluated at the Z-boson mass ($M_z \approx 91$ GeV),

$$\alpha_s^{\overline{MS}}(\mu^2 = M_z^2) = 0.1181 \pm 0.0011 \ll 1\,, \tag{1.15}$$

and therefore allows perturbative techniques. The transition between the non-perturbative and perturbative regime is far from clear-cut. It is commonly accepted that the perturbative regime starts to break down below $1 \sim 2$ GeV.

The aforementioned points can be summarized as follows:

- at short distance or high energy, the interaction weakens and we can use a perturbative approach. Asymptotically, in the limit $Q \to \infty$, QCD becomes/behaves like a *free* theory, without interactions,
- at long distance or low energy, the perturbative evolution of the strong coupling constant breaks down and other approaches are needed to study the interaction strength in the Infra-Red region (IR). The IR behavior is probably linked to the **confinement** of quarks within hadrons but the mechanism is not yet fully understood. In any case, it is an experimental fact that free quarks have not been observed so far.

[4] In fact, the value of Λ_{QCD} depends on the renormalization scheme.

1.2 Basics of Deep Inelastic Scattering and the Parton Model

1.2.1 Deeply Inelastic Scattering

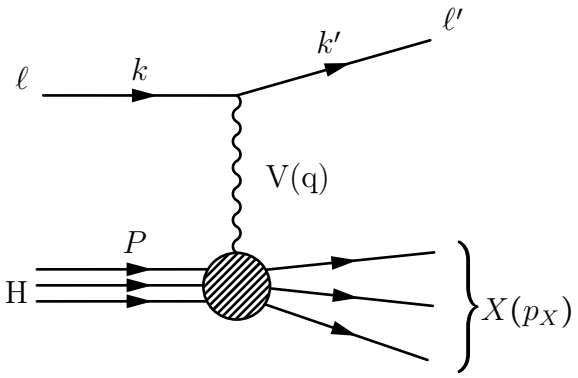


Figure 1.2: Diagrammatic representation of DIS of a lepton (ℓ) off a hadron (H) in the one-boson approximation : $\ell(k) + H(P) \longrightarrow \ell'(k') + X(p_X)$.

Deeply inelastic scattering (DIS) is a foundational process that has played a pivotal role in shaping the landscape of particle physics. In this section, a highly general form is established in order to lay the foundation for the subsequent paragraphs and chapters. This development also serves to underscore the historical interconnection between the core concepts used in this thesis. Considering a collision between a lepton and a hadron, in the single exchange boson (V) approximation, we have

$$\ell(k) + H(P) \longrightarrow \ell'(k') + X(p_X)\,, \tag{1.16}$$

where ℓ is an incoming lepton with momentum k, H is a hadron (typically a nucleon or a nucleus)[5] with momentum P and a mass M. In the final state, the outgoing lepton ℓ' has a momentum k' and X is an arbitrary hadronic state, with a momentum p_X and an invariant mass M_X. The interaction takes place by the exchange of a vector boson V that can be a γ, Z or $W^{\pm}$, with a momentum $q = k - k'$. We shall now describe the kinematics of this process. An inclusive sum over all hadronic final states is performed. Therefore, only the initial momentum k and P, as well as the final momentum k', are available and need to be measured. The momentum transfer between hadron and lepton is spacelike ($q^2 < 0$) and we define the square of the momentum transfer as

$$Q^2 = -q^2 > 0\,. \tag{1.17a}$$

[5]The clear distinction between nucleon and nucleus is not explicitly considered here, but will be later.

A crucial (dimensionless) quantity x, the Bjorken scaling variable, is defined as

$$0 \leq \ x = \frac{Q^2}{2P \cdot q} = \frac{Q^2}{2M\nu} \ \leq 1 \,, \tag{1.18}$$

where ν is the energy transferred from the lepton to the hadron:

$$\nu = \frac{P \cdot q}{M} \,. \tag{1.19}$$

It is also customary to introduce the inelastic parameter

$$0 \leq y = \frac{P \cdot q}{P \cdot k} \leq 1 \,. \tag{1.20}$$

In the target rest frame $P = (M, \overrightarrow{0})$, we have

$$\nu = E_\ell - E_{\ell'} \,, \tag{1.21}$$

$$y = \left(\frac{E_\ell - E_{\ell'}}{E_\ell} \right) . \tag{1.22}$$

The center-of-mass (c.o.m) energy squared (s) is described by

$$s = (P + k)^2 \,, \qquad S = 2P \cdot k = s - M^2 - m_\ell^2 \,, \tag{1.23}$$

and the invariant mass of the hadronic final state by

$$M_X^2 = p_X^2 = (P + q)^2 \,. \tag{1.24}$$

The variables x, y and S are not independent and it is easy to verify the relation $Q^2 = Sxy$ which is useful to estimate x given S and Q^2. Finally, the DIS regime is valid for

$$Q^2 \gg M^2 \approx 1 \text{ GeV}^2 \qquad \text{(deep)} \,, \tag{1.25}$$

$$W^2 \equiv M_X^2 = M^2 + Q^2 (\frac{1 - x}{x}) \geq 4 \text{ GeV}^2 \qquad \text{(deeply inelastic)} \,. \tag{1.26}$$

Here we note that the kinematics are general and are valid regardless of the formalism used (e.g. OPE or Parton Model). The inclusive cross section for the DIS process is obtained by

$$d\sigma(l + H \longrightarrow l' + X) = \frac{1}{F} \left(\sum_n \int |\mathcal{M}|^2 d\mathrm{Q}_h^{(n)} \right) \frac{d^3 l'}{(2\pi)^3 2E_l'} \,, \tag{1.27}$$

where

- $F = 4\sqrt{(p\cdot k)^2 - p^2k^2}$ is the Moller flux,
- $\frac{d^3l'}{(2\pi)^3 2E'_l}$ is the leptonic phase space,
- $|\mathcal{M}|^2$ is the square of the invariant transition matrix element,
- $d\mathrm{Q}_h^{(n)}$ is the n-particle phase space of the hadronic system.

In the single exchange boson approximation, one can write :

$$\sum_n \int |\mathcal{M}|^2 d\mathrm{Q}_h^{(n)} \sim L_{\mu\nu} W^{\mu\nu} \,, \tag{1.28}$$

where $L_{\mu\nu}$ is the **leptonic tensor** and $W^{\mu\nu}$ the **hadronic tensor**.

The hadronic tensor is defined as follows :

$$W_{\mu\nu} = \frac{1}{4\pi}\frac{1}{2}\sum_{S_p}\sum_{n=1}\sum_{S'_{p_1},\ldots,S'_{p_n}} \int dQ_h^{(n)} \left\langle p|J_\nu^\dagger(0)|X\right\rangle \langle X|J_\mu(0)|p\rangle \,. \tag{1.29}$$

We shall now expand the hadronic tensor in the most general way respecting the symmetries of the theory. The hadronic tensor is hence expanded in terms of all allowed Lorentz tensors each multiplied by a **scalar function**, the latter being the **Structure Functions** (SFs), which depend on all possible Lorentz invariants only, i.e., q^2 and $p\cdot q$ or alternatively Q^2 and x:

$$\begin{aligned} W^{\mu\nu}(p,q) = &-g_{\mu\nu}W_1(x,Q^2) + \frac{p_\mu p_\nu}{M^2}W_2(x,Q^2) - i\epsilon_{\mu\nu\rho\sigma}\frac{p^\rho q^\sigma}{M^2}W_3(x,Q^2) \\ &+\frac{q_\mu q_\nu}{M^2}W_4(x,Q^2) + \frac{p_\mu q_\nu + p_\nu q_\mu}{M^2}W_5(x,Q^2) + \frac{p_\mu q_\nu - p_\nu q_\mu}{M^2}W_6(x,Q^2)\,. \end{aligned} \tag{1.30}$$

The current conservation at the hadronic vertex requires $q^\mu W_{\mu\nu} = q^\nu W_{\mu\nu} = 0$, and then

$$W_5 = -\frac{p\cdot q}{q^2}W_2\,, \qquad W_4 = \left(\frac{p\cdot q}{q^2}\right)^2 W_2 + \frac{M^2}{q^2}W_1\,. \tag{1.31}$$

Finally, adding parity symmetry ($W_3 = 0$ and $W_6 = 0$), the hadronic tensor can be expressed as

$$W^{\mu\nu} = -\left(g^{\mu\nu} - \frac{q^\mu q^\nu}{q^2}\right)W_1 + \frac{1}{M^2}\left(P^\mu - q^\mu\frac{P\cdot q}{q^2}\right)\left(P^\nu - q^\nu\frac{P\cdot q}{q^2}\right)W_2\,. \tag{1.32}$$

When contracting the hadronic tensor with a general leptonic tensor, the W_4 and W_5 SFs are strongly suppressed in the cross-section by terms of the lepton mass squared (m_ℓ^2/s). Therefore, they are usually neglected except a tau lepton is involved. The SF W_6 does not contribute at all in unpolarized DIS. We will therefore only consider the SFs W_1, W_2, W_3

SFs here, as well as in the next chapter for convenience. The modern convention is to denote the SFs by the *scaling* dimensionless F_i rather than W_i with

$$\{F_1, F_2, F_3\} = \left\{W_1, \frac{Q^2}{2xM^2}W_2, \frac{Q^2}{xM^2}W_3\right\} . \tag{1.33}$$

The **Structure Functions** describe the internal structure of a hadron; they describe the long distance physics and therefore cannot be calculated perturbatively. However they can be extracted from experimental data.
Considering the easiest example of $ep \to eX$, where the exchanged boson is a photon and neglecting the lepton mass m_ℓ, the inclusive cross-section reads

$$\frac{d^2\sigma}{dxdy} = \frac{4\pi\alpha_{em}^2 S}{Q^4}\left(xy^2F_1(x,Q^2) + (1-y-xy\frac{M^2}{S})F_2(x,Q^2)\right) , \tag{1.34}$$

where x, y are Lorentz-invariants defined in Eq. (1.18), (1.20) and $Q^2 = Sxy$.

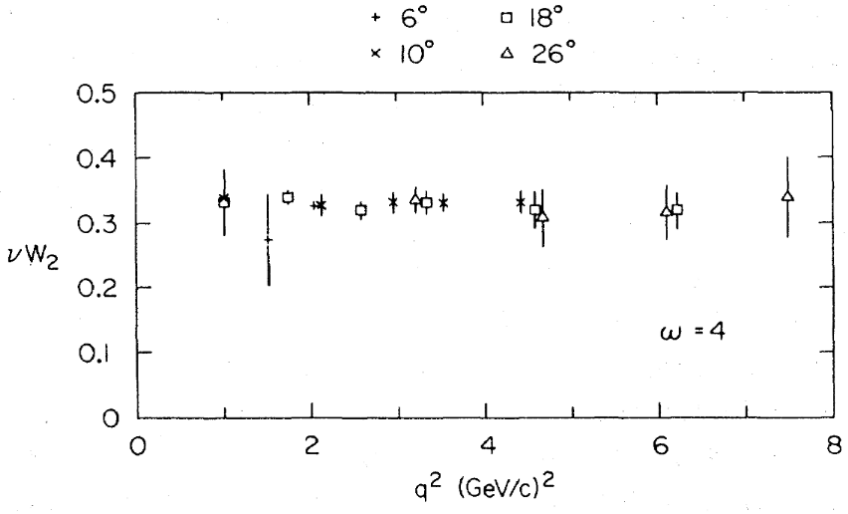


Figure 1.3: An observation of scaling for the structure function W_2 in the proton, as a function of q^2, at $x = 0.25$ ($x = 1/\omega$). Figure taken from [15].

Bjorken predicted that, in the infinite momentum limit ($Q^2 \to \infty$ and $\nu \to \infty$), such that the Bjorken variable x in Eq. (1.18) is fixed, and if the nucleon contains point-like constituents, the SFs would no longer depend on Q. As can be seen in Fig. 1.3, the experiment carried out at SLAC in the late 60's, has confirmed this behavior [16]. These results supported/reinforced Feynman's idea of nucleons being constituted of partons, which will be discussed in the next subsection.

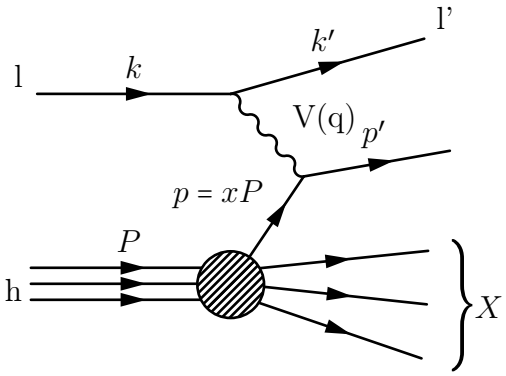


Figure 1.4: Graphical representation of DIS in the parton model.

1.2.2 The Naive Parton Model

The Parton Model was first proposed by Feynman [1], and phenomenologically applied by Bjorken and Paschos [2] in 1969, as an intuitive model to explain the experimental results (as in Fig. 1.3) in lepton-hadron DIS at high energy conducted at SLAC in the late sixties [16, 17]. The model is based on the following fundamental assumptions: the nucleon ($P = (E_N, 0, 0, E_N)$) is composed of point-like constituents, **partons**, which do not interact with each other at sufficiently high energy, in an infinite momentum frame ($E_N \gg M_N$). It is then supposed that all the nucleon's momentum P is carried by the partons p of type i, which means

$$p_i = x_i P \qquad \text{and} \qquad \sum_i x_i = 1 \,. \tag{1.35}$$

In the infinite momentum frame, the internal time scales of interactions between different partons in the nucleon are slowed down by the time-dilatation. For this reason, a hard probe sees essentially *free*, non interacting partons. The total cross section will become an incoherent sum of individual (*elastic*) partonic cross sections

$$d\sigma = \sum_i \int_0^1 dx f_i(x) d\hat{\sigma}(p_i) \,, \tag{1.36}$$

where $d\hat{\sigma}$ is the partonic cross-section and $f_i(x)$ is the probability to find a parton of type i with momentum fraction x in the nucleon[6].

In the naive approach of the parton model, where the exchanged boson is a photon, we can express the hadronic tensor of Eq. (1.30) as

$$W^{\mu\nu}(P, q) = \sum_i \int_x^1 \frac{dx}{x} f_i(x)\, \hat{\omega}_i^{\mu\nu}(p_i, q) \,, \tag{1.37}$$

[6]Thus, $f_i(x)dx$ is the number of partons of type i carrying a momentum fraction of the nucleon momentum in $[x, x + dx]$.

where $\hat{\omega}_i^{\mu\nu}$ is the partonic tensor and follows (in the naive Parton Model)

$$4\pi\,\hat{\omega}_i^{\mu\nu} = \int \frac{d^4p'}{(2\pi)^4} 2\pi\delta(p'^2)(2\pi)^4\delta^{(4)}(p+q-p')\left\langle\langle p|J^{\mu\dagger}(0)|p'\rangle\langle p'|J^{\nu}(0)|p\rangle\right\rangle_{Spin} \tag{1.38}$$

$$= (2\pi)\,x\,e_i^2\left[-\left(g^{\mu\nu}-\frac{q^\mu q^\nu}{q^2}\right)+\frac{2x}{P\cdot q}\left(P^\mu - q^\mu\frac{P\cdot q}{q^2}\right)\left(P^\nu - q^\nu\frac{P\cdot q}{q^2}\right)\right], \tag{1.39}$$

if the parton is massless, free, a spin-1/2 fermion and carrying an electro-magnetic charge e_i. Inserting Eq. (1.39) in Eq. (1.37) and comparing with Eq. (1.32), we finally found that

$$F_1(x) = \frac{1}{2}\sum_i e_i^2\, f_i(x) \quad \text{and} \quad F_2(x) = \sum_i e_i^2\, x f_i(x) \quad \Rightarrow \quad F_2(x) = 2xF_1(x)\,, \tag{1.40}$$

where the last formula is the Callan-Gross relation [18]. This relation was experimentally observed and thus was an additional argument for identifying partons with quarks, those that were postulated earlier by Gell-Mann [19]. Actually, in the parton model, quarks are of two different types: valence quarks (q_v) and sea quarks ($\overline{q}$). The first are identifiable with Gell-Mann quarks, while the second have no equivalent. As a result, a proton is made up of 3 valence quarks ($u_v u_v d_v$) where a valence quark is defined as

$$q_v = q - \overline{q}\,, \tag{1.41}$$

where $q = u, d, s, ...$ are the different quark flavors. Finally, for a proton, this leads to the quark number sum rule

$$\int_0^1 dx\, u_v(x) = 2, \qquad \int_0^1 dx\, d_v(x) = 1, \qquad \int_0^1 dx\,(s_v = s - \overline{s})(x) = 0\,. \tag{1.42}$$

1.2.3 Limits and expectations

While the DIS experiments and the resulting *naive* Parton Model constituted a major conceptual development within the field of strong interactions, several challenges also emerged. In reality, Bjorken scaling is only an approximation; a violation of scale invariance is observed when approaching the $x \in [0,1]$ intervals upper and lower bounds ($x \gtrsim 0.6$ and $x \lesssim 0.1$), as can be seen in Fig. 1.5. Another important experimental result showed shortcomings in the naive parton model; one might have expected, in the case of a proton composed exclusively of $u_v u_v d_v$ quarks, that the momentum sum rule $\int_0^1 dx\, x(u(x) + \overline{u} + d(x) + \overline{d})$ is equal to 1. In fact, it was experimentally found that the sum tends to ~ 0.45, which meant that the momentum of the nucleon contains other contributions, the **gluon** distribution $g(x)$. Finally, the naive parton model did not explain

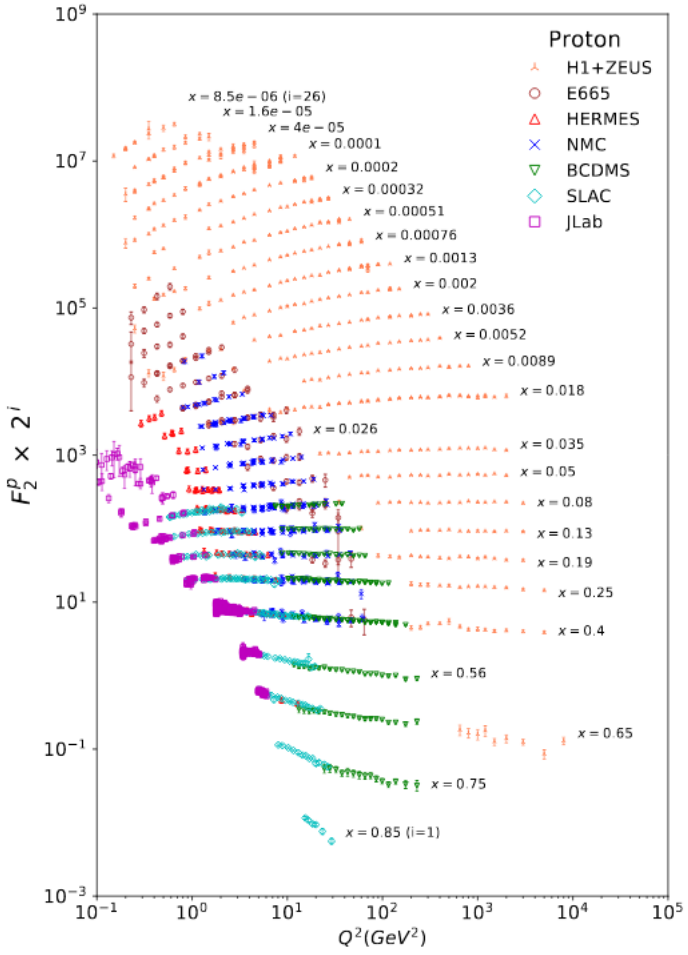


Figure 1.5: The scaling violation of structure function F_2 in a proton as a function of Q^2. Figure taken from PDG [12].

why, at high energy, partons behave like free particles, even though these particles are strongly bound in hadrons. Furthermore, with the successes of QED a few decades earlier, there was a desire to describe strong interactions using renormalizable gauge field theory. This became particularly important when it was realized in 1973 that non-Abelian gauge theories such as QCD explain asymptotic freedom, as already mentioned in Sec. 1.1.2.

1.2.4 Operator Product Expansion approach

We briefly outline here the concept of OPE; since a complete description is done in Chap. 2. At the same time, in the late 1960s, a formal field theoretic approach for the SFs and the Parton Model was proposed using Wilson Operator Product Expansion (OPE) [20]. Callan and Gross used the OPE and the renormalization group and showed that an approximate Bjorken scaling in DIS requires a 'weak' strong interaction at short distance [21]. Then the necessity of asymptotically free theory for strong interaction became imperative and the researches led to the demonstration that only non-Abelian

theory have this property [4].

The original formulation of the OPE, as introduced by Wilson [20] states that when the 4-vector y is close to x, one can express

$$A(x)B(y) = \sum_n C_n(x-y)O_n(x)\,, \tag{1.43}$$

where $A(x), B(x)$ are two local fields, O_n a set of local operators and $C_n(x-y)$ are *"Wilson coefficient"* functions. Furthermore, these operators can be ordered by dimension or **twist**. The twist is defined as τ = dimension - spin (of the operator). We can express the hadronic tensor using the optical theorem $W_{\mu\nu}$ = Im $T_{\mu\nu}$, where

$$T_{\mu\nu} = i \int d^4x e^{iqx} \left\langle N|T\left[J_\mu^\dagger(x)J_\nu(0)\right]|N\right\rangle . \tag{1.44}$$

Using a generalization of OPE and the condition of light cone dominance of DIS ($x^2 \sim 0$), we have

$$T\left[J_\mu(x)J_\nu(0)\right] \simeq \sum_{i,\tau,n} c_\tau^{i,\mu_1\cdots\mu_n}(x^2)\, O^{i,\tau}_{\mu_1\cdots\mu_n}(0)\,, \tag{1.45}$$

where $c_\tau^{i,\mu_1\cdots\mu_n}$ are Wilson coefficients and are computable in perturbation theory, encoding the short distance physics and $O^{i,\tau}_{\mu_1\cdots\mu_n}$ are local operators, of twist-τ, built from quark and gluons fields, encoding the long distance physics. At large-Q^2, the leading twist ($\tau = 2$) operator largely dominate and higher twist are suppressed by power of $1/Q^2$.

This section has set up the foundational ideas needed to connecting experimental observables (with the example of DIS) to the theoretical description using QCD. Finally, these two formalisms; OPE and Parton Model, are two manifestations of **factorization**. In the next section, we will see how QCD associated with the Parton Model/OPE has led to the perturbative QCD formalism and its applications.

1.3 Perturbative QCD

This section introduces the essential elements of perturbative QCD (pQCD) as applied to high-energy scattering processes.

1.3.1 Factorization Theorems for hard processes

> «A general cross section is a combination of short- and long-distance behavior, and is hence not computable directly in perturbation theory for QCD. (...) **Factorization** theorems allow

US TO DERIVE PREDICTIONS FOR THESE CROSS SECTIONS, BY SEPARATING (FACTORIZING) LONG-DISTANCE FROM SHORT-DISTANCE BEHAVIOR IN A SYSTEMATIC FASHION.» Collins, Soper, Sterman [8]

The philosophy behind the factorization theorems is the separation of **long** and **short** distance physics, respectively **soft** and **hard** scale behaviors in the aim of calculating a high energy cross-section. This can be generally expressed as

QCD Factorization theorems

$$\sigma = \begin{cases} f \otimes \widehat{\sigma} \\ \\ f_1 \otimes \widehat{\sigma} \otimes f_2 \end{cases} + \mathcal{O}\left(\frac{\Lambda^n}{Q^n}\right). \tag{1.46}$$

where

- f, f_1, f_2 are the partonic densities inside the entering hadron(s). These functions contain the long distance effects and are not perturbatively calculable, they must be determined by experimental data. On the other hand, they are **universal**, that is to say the same for all processes, which makes the formalism predictive. PDFs will be treated specifically in Sec. 1.3.3.
- $\hat{\sigma}$ is the hard part. This only contains the small-distance effects and is calculable order by order in perturbation theory, using Feynman rules of QCD. The hard part will be treated specifically in the next section (see Sec. 1.3.2).
- $\mathcal{O}\left(\frac{\Lambda^n}{Q^n}\right)$ is the error of the factorization approximation where typically $n = 1$ or 2 depending on the process ($n = 2$ for DIS). As can be seen, the larger the hard scale (Q) compared to the hadronic scale ($\Lambda \sim 0.5 \ldots 1$ GeV), the more the error gets suppressed.

The factorization theorems hold for different processes involving hadrons as inclusive DIS, Drell-Yan lepton pair production, single inclusive hadron production and many more and can be found in [8, 22] and reference therein. Moreover, it exists different kind (in the sense of expression/manifestation) of factorization, depending on processes or sets of processes, each with their specific characteristics adapted to them. Collinear factorization, for example is adapted for inclusive DIS, Drell-Yan or inclusive hadron production whereas Transverse-Momentum Dependent (TMD) factorization is more suitable for Drell-Yan at

small q_T or Semi Inclusive DIS where a low p_t hadron is detected, and some more (see Chapter 13 in [22]). Note here that the original proofs of factorization considered massless partons[7].

As an example of a factorization theorem, we can rewrite the hadronic tensor of Eq. (1.37) in the QCD-improved parton model framework

$$W^{\mu\nu} = \sum_i \int_x^1 \frac{dx}{x} f_i(x, \mu_{Fact}, \mu_R)\, \hat{\omega}_i^{\mu\nu}\left(x, Q^2, \mu_{Fact}^2/Q^2, \mu_R^2/Q^2, \alpha_s(\mu_R^2)\right) + \mathcal{O}\left(\frac{\Lambda^2}{Q^2}\right). \quad (1.47)$$

where μ_F is the factorization scale and μ_R the renormalization scale. As a consequence, the DIS structure functions are obtained in factorized form as convolutions

$$F_2(x, Q^2) = \sum_i x\, f_i \otimes C_{2,i}\,, \quad F_1(x, Q^2) = \sum_i f_i \otimes C_{1,i}\,, \quad (1.48)$$

where i runs over partonic flavors, f_i are the PDFs of flavor i and $C_{1,i}$ and $C_{2,i}$ are Wilson coefficients.

1.3.2 Hard part

The hard part is calculable in perturbation theory using Feynman rules and is process dependent. It is typically expanded in powers of α_s, as

$$\hat{\sigma} = \hat{\sigma}^{(0)} + \alpha_s(\mu_R)\hat{\sigma}^{(1)} + \alpha_s^2(\mu_R)\hat{\sigma}^{(2)} + \cdots. \quad (1.49)$$

At leading order in perturbation theory, the hard cross-section ($\hat{\sigma}^{(0)}$) corresponds exactly to the partonic cross-section.

As an example in DIS, we can expand the Wilson coefficients of Eq. (1.48) as

$$C_{j,i}(z, Q^2) = C_{j,i}^{(0)}(z) + C_{j,i}^{(1)}(z, Q^2) + \cdots, \quad (1.50)$$

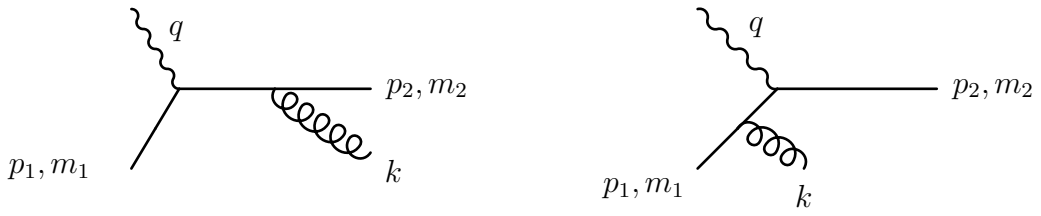
where i is still the partonic flavor and j is the SF index (1, 2, 3, etc.) The Wilson coefficients are known up to next-to-next-to-next-to-leading order (N^3LO). Note that at LO, we have

$$C_{2,q}^{(0)}(z) = e_q^2\delta(1-z) \quad \text{and} \quad C_{2,g}^{(0)}(z) = 0\,, \quad (1.51)$$

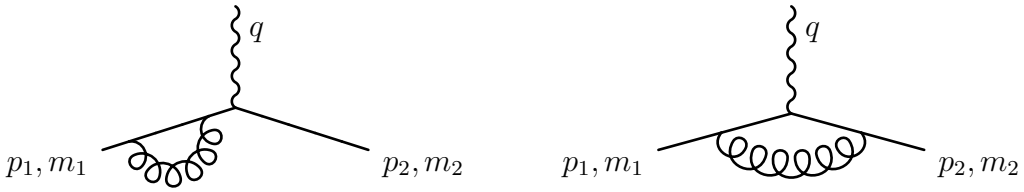
and then the results of the parton model in Eq. (1.40) are recovered. At next-to-leading

[7]This assumption is valid for the three light quarks (u, d, s), since $m_{u,d,s} \ll \Lambda_{QCD}$, but some complication arrive when taking into account heavy quarks $(c, b, (t))$. A solution, was proposed by Appelquist and Carazone [23, 24], the *decoupling theorem*, where two regimes are introduced; $\mu \ll m_Q$ and $\mu \gg m_Q$.

order (NLO) in QCD, the hard part ($\hat{\sigma}^{(n)}$ with $n > 0$), contains a renormalized partonic cross-section plus a regularization term (subtraction term) that comes from the procedure of mass factorization. The divergences that appear at NLO are of three types and can be treated in different ways. The **Ultra-Violet** (UV) divergences appear in loop diagram,



a) Real corrections.



b) Virtual corrections.

Figure 1.6: QCD corrections to the process $V + q \to q$ in DIS. See also Fig. 1.4. The diagrams in a) are real corrections, the ones in b) are virtual corrections.

since the momentum can tends to ∞ (as in Fig. 1.6 b) if $k \to \infty$, using the DIS example) and are treated by renormalization. This explains the introduction of a dependence on the renormalization scale (μ_R) of the perturbative development and reflects the high energy behavior. The **soft** divergences appear when a (real) low-momentum gluon (or a photon in QED) is emitted. Due to the KLN theorem [25, 26], the soft real (Fig. 1.6 a) for $k \to 0$) and soft virtual (Fig. 1.6 b) for $k \to 0$) contributions cancel, order by order. The **collinear** divergences appear when a (real) gluon or photon is emitted by a *massless* fermion (as u, d, s quarks) and their both momenta are collinear ($k // p_{1,2}$). These kinds of divergences are factorized and absorbed in non-perturbative functions as Fragmentation Functions (FFs) and PDFs. In contrast to UV divergences, the soft and collinear divergences reflect the low energy/large distance behavior of the theory and are both denoted as Infra-Red (IR) divergences.

Finally, different methods are available to treat the different divergences, leading to different renormalization and factorization schemes. The choice of scheme often depends on which processes we work on. More importantly, the observables (like cross-sections) do not depend on the scheme and scales, up to missing higher orders due to the cut in the

perturbative expansion. The factorization scheme could be thought of as the manner in which the hard part is separated from the PDFs, beyond LO. The scheme must therefore be the same for the hard part and the PDFs. Two common factorization schemes are the DIS scheme [27] and the $\overline{MS}$ scheme [28].

1.3.3 DGLAP evolution

Within the pQCD framework, the PDFs also depend on the factorization and renormalization scales. The factorization scale (μ_F) has come from the absorption of collinear singularities (as mention before in Sec. 1.3.2). The evolution across μ_F of the PDFs is governed by the DGLAP (Dokshitzer-Gribov-Lipatov-Altarelli-Parisi) [29–31] evolution equations:

$$\frac{df_{i/h}(x,\mu_{Fact})}{d\log\mu_{Fact}} = \sum_j \int_x^1 \frac{dy}{y} P_{ij}(x/y,\alpha_s(\mu_{Fact}))f_{j/h}(y,\mu_{Fact})\,, \tag{1.52}$$

where the P_{ij} are splitting functions, usually called Altarelli-Parisi kernels. They are given by perturbative expansions in α_s as

$$P_{ij}(x/y,\alpha_s(\mu_{Fact})) = \quad P_{ij}^{(0)}(x/y,\alpha_s(\mu_{Fact}))\frac{\alpha_s(\mu_{Fact})}{2\pi} \tag{1.53}$$

$$+ P_{ij}^{(1)}(x/y,\alpha_s(\mu_{Fact}))\left(\frac{\alpha_s(\mu_{Fact})}{2\pi}\right)^2 + \cdots\,. \tag{1.54}$$

Note here that the x-dependence of the PDFs is still not calculable. However, due to the universality of the PDFs combined with DGLAP equations and experimental data over a sufficiently wide x-range, the PDFs can be determined very precisely.

1.3.4 Sum rules and other properties of PDFs

Another aspect of the naive parton model remains valid despite the QCD corrections; the sum rules. These rules allow for better constraints of the (x-)parametrization of the PDFs in a global analysis, which is the subject of the following section. The obvious momentum sum rule follows

$$\int_0^1 dx\, x\left(f_g(x,Q^2) + \sum_{i=u,d,s,c,..}(f_{q_i}(x,Q^2) + f_{\bar{q}_i}(x,Q^2))\right) = 1\,, \tag{1.55}$$

and is valid at any scale Q^2. For a proton, the (valence) quark number sum rules is given by

$$\int_0^1 dx\,(f_u(x,Q^2) - f_{\overline{u}}(x,Q^2)) = 2, \qquad \int_0^1 dx\,(f_d(x,Q^2) - f_{\overline{d}}(x,Q^2)) = 1\,, \tag{1.56}$$

again valid at any scale Q^2. Due to (strong) isospin invariance, the density of the u-quark in the proton is equal [8] to the density of d-quark in the neutron:

$$f_u = f_u^p = f_d^n\,, \qquad f_d = f_d^p = f_u^n\,, \qquad f_{\overline{u}} = f_{\overline{u}}^p = f_{\overline{d}}^n\,, \qquad f_{\overline{d}} = f_{\overline{d}}^p = f_{\overline{u}}^n\,. \tag{1.57}$$

Summary: Using Sec. 1.1 and Sec. 1.2 as foundation, Sec. 1.3 provides the framework for applying pQCD theory to high-energy scattering by building from general principles (factorization) to the practical implementation of the hard and soft part, concluding with some theoretical constraints (as sum rules). We will then be able to combine all these ingredients to determine the PDFs in a global analysis, which is the aim of the following section.

1.4 QCD Global Analysis of PDFs

1.4.1 General framework

We have shown previously that through QCD factorization theorems, the cross sections can be separated into two parts: the hard part, calculable in perturbation theory and the non-perturbative part, the PDFs. The latter are not calculable but are universal, i.e. independent of the process under consideration. It will then be possible to determine these PDFs using experimental data, via a global QCD analysis. Many collaborations exist and perform these analyses; some of them are: CTEQ [32], CJ [33] or NNPDF [34] for the proton PDFs and nCTEQ [35], EPPS [36] or nNNPDF [37] for nuclear PDFs. The basis of the global QCD analysis is common to all these different groups and is described in Fig. 1.7. It essentially consists of selecting a set of experimental data, calculating theoretical predictions for these data with PDFs as input, then fitting these PDFs by minimizing the χ^2 between theory and experimental data. What differs from one collaboration to another is essentially the choice of input data and cuts, the parametrization of the PDFs, or even the minimization algorithm. In the end, the final results between different groups should be in agreement, taking into account the uncertainties.

[8] In reality, the strong isospin is not an exact but a very good symmetry, remaining valid at the percent level.

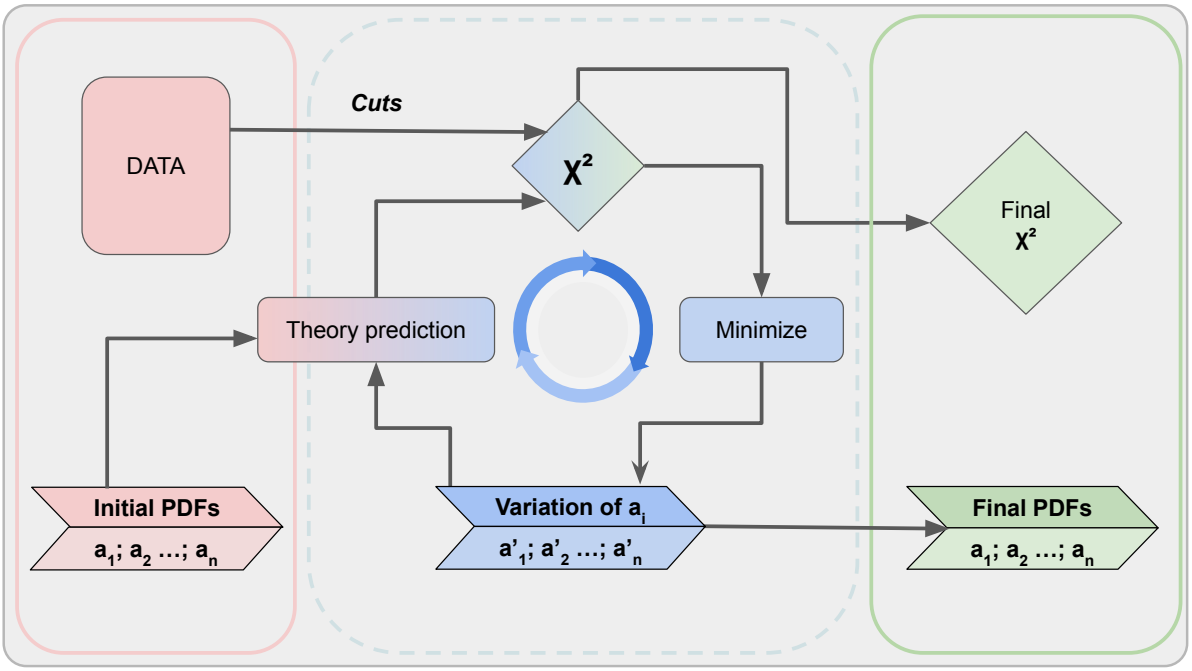


Figure 1.7: General flowchart of a QCD global analysis.

This is a rather heavy calculation, because it is necessary to redo the theoretical computations at each stage of the fit for thousands of data points; evolving the 'newly adjusted' PDFs once again through DGLAP and redoing the theoretical calculations with this 'new' PDFs in order to compare them again with the data, and this, in a minimization loop until we obtain the best possible PDFs. This is why, if we consider a large number of data points, the theoretical predictions need to be fast and flexible. Providing a fast calculation for one of the processes, prompt photon production, is the aim of Chap. 4.

From a second perspective, a global QCD analysis serves as a valuable tool for testing QCD and the factorization theorems, especially when their formal proof is not yet established for all processes, and hence the universality of the PDFs. Any tensions observed within the fitting process could 'potentially' indicate a breakdown of factorization.

1.4.2 Nuclear PDFs

In the previous sections, we discussed PDFs in a proton (f_i^p) or a neutron (f_i^n). One might think that for the dynamics of nuclear collisions, if considering that nucleons do not interact with each other in the nucleus, the proton and neutron PDFs would be sufficient. The consequence would be that the cross sections for DIS would then be only a sum of individual nucleon SFs that compose the nucleus under consideration. But experimentally, as can be seen in Fig. 1.8, this naive assumption does not hold; this suggests that nuclear effects are present. One refers to the different regions such as:

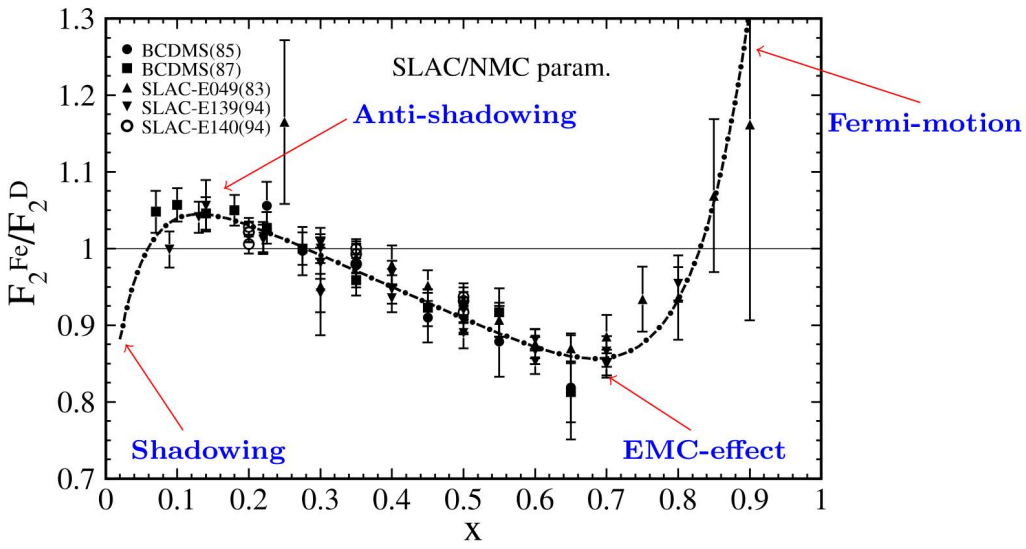


Figure 1.8: A characteristic nuclear correction ratio F_2^{Fe}/F_2^D for iron over deuterium. We observe the shadowing region ($x \lesssim 0.1$), the anti-shadowing region ($x \sim 0.15$), the EMC region ($x \sim 0.6$), and the Fermi region ($x \gtrsim 0.8$). The figure is taken from Ref. [38] which contains the details and references for the data sets.

- $x \gtrsim 0.8$ is the Fermi region,
- $0.3 \lesssim x \lesssim 0.8$ is the EMC region,
- $0.1 \lesssim x \lesssim 0.3$ is the anti-shadowing region and
- $x \lesssim 0.1$ is the shadowing region.

The question then becomes how to describe these effects. Furthermore, since the modifications differ depending on the x-region, are these different physical effects or a single effect applicable everywhere? Are the effects more important as nuclei are bigger? Are these nuclear effects a modifications of the hard part or are they non-perturbative effects that can be absorbed by the PDFs, or maybe both? In our approach, we assume the factorization theorems hold for nuclear collisions and nuclear effects are absorbed in nuclear PDFs. This then leads to the nPDFs parametrization for a nucleus

$$f_i^A = \frac{Z}{A} f_i^{p/A} + \frac{A-Z}{A} f_i^{n/A}\,, \tag{1.58}$$

where $f_i^{p,n/A}$ are **effective** nucleon PDFs **bounded** in a nucleus A (in contrast to the *free* nucleon PDFs $f_i^{p,n}$). In this thesis, we will examine two different x-ranges: $x \gtrsim 0.5$ in Chapter 2, and the gluon PDFs at low-x (in the shadowing region) in Chapters 3 and 4. We show in Fig 1.9 the lead ($A = 208, Z = 82$) PDFs from the different versions of nCTEQ.

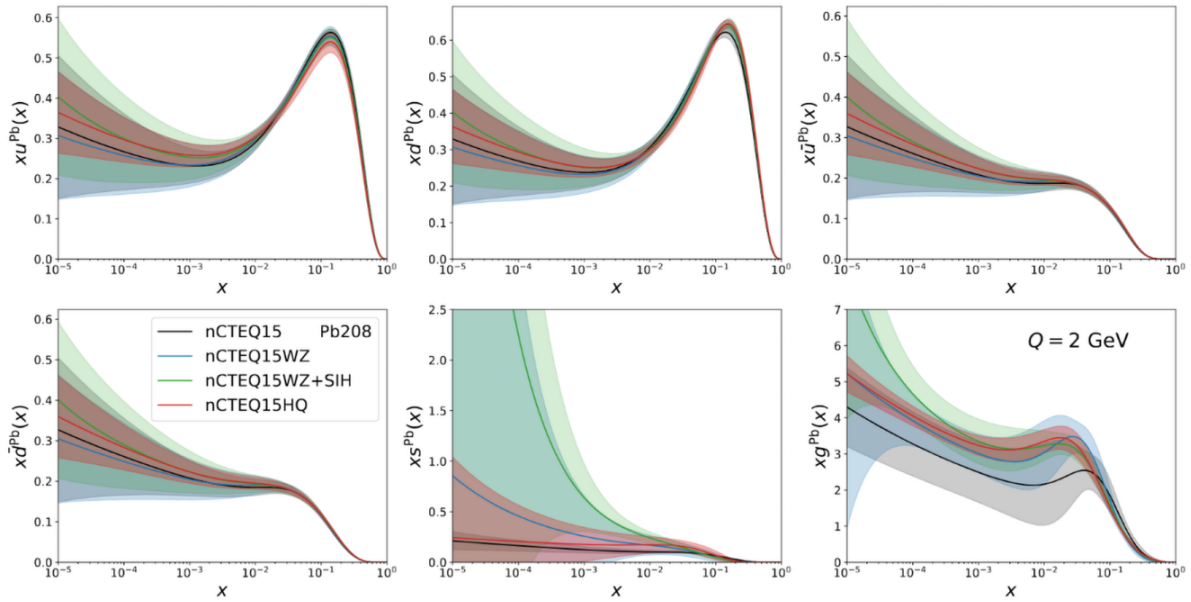


Figure 1.9: The lead (Pb) PDFs from different nCTEQ15 versions. The baseline fit (nCTEQ15) is shown in black, while nCTEQ15WZ is in blue, nCTEQ15WZSIH in green and nCTEQ15HQ in red. Figure taken from [39].

1.4.3 nCTEQ framework

The first generation of nCTEQ PDF fits was introduced in 2009 [40]. Then another generation of nuclear global fits of the nCTEQ collaboration was presented in 2015 [35] and included neutral current DIS, Drell-Yan, and RHIC pion production data. As time went by, new data and fast theoretical calculations became available, multiple analyses were added [39, 41–45], giving rise to multiple new nPDFs, each with its own specifications. Many different elements have been modified through the various analyses, such as the number of parameters, or even the parameterization itself. The code has also undergone extensive modifications, starting with the switch from FORTRAN to C++, and more recently in its architecture. The new version ncteqpp2.0 incorporates all these improvements and defines the new basis.

Parametrization:

Here, we show the parametrization used in the Heavy Quark analysis [39], whose result is the nCTEQ15HQ PDF set. The PDFs are parametrized by

$$x f_i^{p/A}(x, Q_0) = c_0 x^{c_1} (1-x)^{c_2} e^{c_3 x} (1 + e^{c_4 x})^{c_5}\,, \tag{1.59}$$

for $i = u_v, d_v, g, (\overline{u} + \overline{d}), (s + \overline{s}), (s - \overline{s})$ and

$$\frac{\overline{d}(x, Q_0)}{\overline{u}(x, Q_0)} = c_0 x^{c_1} (1-x)^{c_2} + (1 + c_3 x)(1-x)^{c_4}. \tag{1.60}$$

The nuclear dependence is parametrized directly in the c-coefficient of Eq. (1.59) as

$$c_k \to c_k(A) \equiv p_k + a_k(1 - A^{-b_k}) \,, \tag{1.61}$$

where $k = 1, \cdots, 5$. This parametrization is designed to allow a recovery of the *free* proton PDFs for A=1, where $c_k^p = p_k$. Moreover, the input scale is chosen as $Q_0 = 1.3$ GeV to match the proton PDFs. Note that the heuristic form proposed here is subjective, and can be adapted if desired.

In fact in the new release nCTEQ25, currently underway, the parametrization has evolved towards

$$x f_i^{p/A}(x, \mu_0) = c_0^k x^{c_1^k} (1-x)^{c_2^k} (1 + c_3^k \sqrt{x} + c_4^k x) \,, \tag{1.62}$$

for $k = u_v, g, (\overline{u} + \overline{d}), (\overline{d} - \overline{u}), (s + \overline{s})$ and the d_v distribution has extra parameters as

$$x d_v^{p/A}(x, \mu_0) = c_0^{d_v} x^{c_1^{d_v}} (1-x)^{c_2^{d_v}} (1 + c_3^{d_v} \sqrt{x} + c_4^{d_v} x) + c_5^{d_v} x^{c_6^{d_v}} \, x u_v^{p/A}(x, \mu_0) \,. \tag{1.63}$$

Furthermore, the $\bar{d}/\bar{u}$ is parametrized as follows:

$$\frac{\bar{d}}{\bar{u}}(x, \mu_0) = c_0^k x^{c_1^k} (1-x)^{c_2^k} + 1 + c_3^k x (1-x)^{c_4^k}, \tag{1.64}$$

for $k = \bar{d}/\bar{u}$. The nuclear dependence is parameterized as follows,

$$c_j(A) \equiv p_j + a_j \ln A + b_j \ln^2 A \,, \tag{1.65}$$

where $j = 0, \cdots, 6$. In this analysis, $s - \overline{s} = 0$ is taken.

Fitting:

As mentioned before, the next step is to minimize the $\chi^2(\mathbf{a})$. The theoretical predictions depend on the set of parameters $\mathbf{a}$, which define the PDFs at the input scale (the c_i^k). The simplest definition of the χ^2 function, assuming normally distributed, uncorrelated data points is

$$\chi^2(\mathbf{a}) = \sum_i \frac{[D_i - T_i(\mathbf{a})]^2}{\sigma_i^2}, \tag{1.66}$$

where D_i are the data, T_i are the corresponding theoretical predictions and $\sigma_i^2 \equiv \hat{\sigma}_i^2 + \bar{\sigma}_i^2$ is the sum of the statistical uncertainty $\hat{\sigma}_i^2$ and systematic uncertainties $\bar{\sigma}_i^2$ added in quadrature. Note here that we do not take into account the theoretical uncertainties [9].

[9] There are some efforts to include theoretical uncertainties in PDF fits.

Actually, we do not use this naive form of the χ^2 in the nCTEQ analyses, neither in nCTEQ15 nor in the new one, because we must take into account the normalization of the data and correlations of its uncertainties. A full coverage of these subjects can be found in all nCTEQ papers and related theses.

Uncertainties:

Once the minimization of the χ^2 is done, we obtain a parametrization of the *central* value of the nPDFs, but we need to estimate the fitting error around this. In the nCTEQ analyses, we are using the Hessian method [46, 47], where the χ^2 near its minimum, can be approximated by

$$\chi^2 \approx \chi_0^2 + \sum_{i,j} H_{ij} y_i y_j, \tag{1.67}$$

where $y_i \equiv a_i - a_{i,0}$ are the parameter shifts from the minimum (the central value mentioned before) and H_{ij} is the Hessian matrix defined as

$$H_{ij} = \frac{1}{2}\left(\frac{\partial^2 \chi^2}{\partial y_i \partial y_j}\right)_{a_i = a_{i,0}}. \tag{1.68}$$

The uncertainties are then obtained with the numerical calculation of this matrix and can be expressed as

$$\Delta X^+ = \sqrt{\sum_k^{N_{\rm par}} \left[\max\left(X_k^+ - X_0, X_k^- - X_0, 0\right)\right]^2}, \tag{1.69}$$

$$\Delta X^- = \sqrt{\sum_k^{N_{\rm par}} \left[\min\left(X_k^+ - X_0, X_k^- - X_0, 0\right)\right]^2}. \tag{1.70}$$

where $X_k^\pm = X(f^\pm)$, $X_0 = X(f_0)$, and N_{par} is the number of parameters, or the number of eigenvectors of the Hessian matrix.

Finally, one of the main goals of this work is to work on accuracy of the theoretical predictions ($T_i(\mathbf{a})$) and trying to include new data (D_i) with the aim of obtaining more precise PDFs, i.e. with less uncertainties.

Chapter 2

Target Mass Corrections to Deep Inelastic Scattering

This chapter is devoted to the Target Mass Corrections (TMCs) to deep-inelastic scattering. It is widely inspired from the recent review [48].

2.1 Introduction

Deep-inelastic scattering (DIS) of high-energy leptons on nucleons and nuclei is a crucial process for investigating hadron structure in terms of their quark and gluon constituents [15, 49–51]. Both nucleons [32, 33, 52–55] and nuclei [35, 36, 39–44, 56–65] provide essential data from electron and neutrino DIS experiments, which form the foundation for contemporary determinations of parton distribution functions (PDFs). These PDFs are fundamental for describing and understanding the internal structure of hadrons. This process will once again be central to research at the future Electron-Ion Collider (EIC) [66,67], where DIS on nucleons and various nuclear targets will be investigated with high precision across a broad kinematic range. Similar efforts are also being considered for the proposed Forward Physics Facility at CERN [68]. Furthermore, high-precision fixed-target measurements from JLab offer valuable complementary information [69]. In addition to these facilities, DIS is also crucial for current and future accelerator-based neutrino oscillation experiments, such as DUNE and the Short-Baseline Neutrino Program at Fermilab [70–72]. These experiments, which operate at somewhat lower energies compared to the EIC, must also account for (quasi)elastic and resonant processes. Precise control over the contributions from each of these processes is essential, and duality-based arguments for the scaling of DIS cross sections into the resonance region at higher energies can be significant [73, 74]. However, because the effective neutrino-nucleon interaction is

relatively weak, neutrino experiments often employ massive nuclear targets to increase event rates.

To achieve the ambitious physics goals of future experiments, a precise understanding of both charged lepton-nucleus and neutrino-nucleus interactions is essential, encompassing both experimental and theoretical aspects. On the theoretical front, there are broad efforts to improve the description of DIS in Monte Carlo simulation tools [75–81], increase the perturbative accuracy of matrix elements, particularly for more exclusive DIS channels, and of course to refine determinations of parton distribution functions in nuclei [35–37, 39, 65, 82]. Nevertheless, the distinct properties of charged lepton-nucleus and neutrino-nucleus DIS—especially in extreme kinematics—are not yet fully understood and remain a subject of ongoing research [44]. This work specifically addresses the case in DIS where momentum transfers are small and the scaling variable approaches unity.

There are two major theoretical approaches to lepton-*nucleon* DIS that are standard material in textbooks and review articles; both have been employed to include the effects of target nucleons having nonzero masses. One approach is based on the operator product expansion (OPE) [20, 83, 84]. The other approach is the QCD-improved parton model, which is rigorously derived from the first principles of QCD within the framework of collinear factorization theorems [7–9, 22, 85–88].

Operator Product Expansion (OPE): The OPE was initially used to derive target mass corrections (TMCs) to structure functions in DIS at leading order (LO) in QCD in the seminal paper by Georgi & Politzer in 1976 [89, 90]. While the application of the OPE to DIS is widely accepted, proofs of the OPE have only been provided for simple scalar models [91, 92]. To the best of our knowledge, extending these results to QCD in general remains an open question.

Target Mass Corrections (TMCs): Qualitatively, TMCs in DIS account for the mass of the target nucleon (or target nucleus) in the kinematical variables and structure functions, which are usually derived in the limit of a massless target. In the work by Georgi and Politzer it was argued that for each power, or *twist*, in the OPE expansion of hadronic matrix elements in DIS there is a series of (kinematical [93, 94]) corrections of the form $x_A^j(M_A^2/Q^2)^{j'}$, with $(x_A M_A^2/Q^2)^j$ representing the leading power corrections. Here, x_A is the Bjorken scaling variable, M_A is the mass of the target hadron, and $Q^2 > 0$ is the squared virtuality of the intermediate exchange boson, with j, j' being positive powers. Remarkably, the complete series of $(x_A M_A^2/Q^2)^k$ corrections at leading power can be summed, resulting in a closed-form solution for TMCs at this power [89, 90]. In the same year, Barbieri, et al., derived TMCs at LO incorporating effects arising from

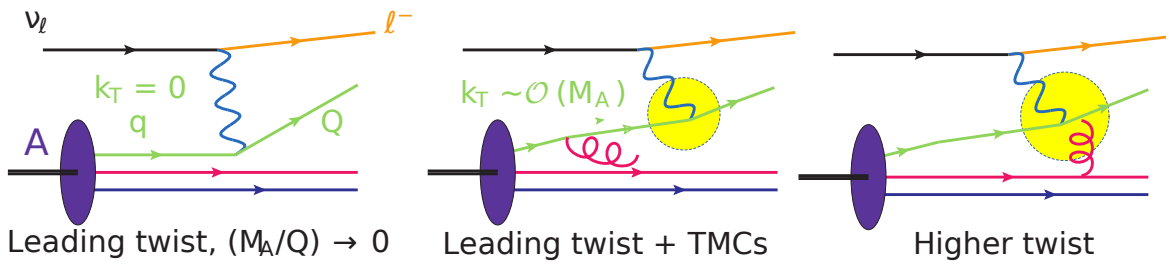


Figure 2.1: Schematic view of $\nu_\ell A \to \ell X$ for a lowest-order heavy quark production, $qW^* \to Q$. **[left]** Leading twist $\{M_A = 0, k_T = 0\}$. **[center]** Leading twist with TMCs {finite M_A, $k_T \sim \mathcal{O}(M_A)$}. The "hard scattering" is indicated by the yellow circle. The hadron dynamics gives the parton (green) a finite k_T, illustrated by the (red) gluon radiation; this is not a next-to-leading order (NLO) correction as the soft (low energy) gluon radiation is outside (before) the hard scattering process (yellow circle). **[right]** Higher twist. A second parton (gluon) is exchanged between the hadron and the hard scattering (yellow circle).

non-zero quark masses [95, 96]. Subsequently, the original work of Georgi & Politzer was extended to next-to-leading order (NLO) in QCD for the structure functions $W_{1,2}$ within the framework of local duality in electro-production and using off-shell regularization [97]. More recently, Kretzer and Reno presented TMCs for charged current (CC) and weak neutral current (NC) neutrino-nucleon DIS, including NLO QCD corrections and heavy quark mass effects using modern conventions [98].

Parton Model: Within the context of the parton model, Ellis, Furmanski, and Petronzio derived target mass corrections (TMCs) at leading order (LO) in QCD using a non-collinear version of the parton model that incorporates the effects of partonic transverse momentum (k_T) [93, 94]. They found agreement with the results based on the Operator Product Expansion (OPE) [89, 90, 95, 96] at leading power, demonstrating the equivalence of the OPE at LO in QCD to a non-collinear parton approach in which the incident parton is on-shell but not collinear with its parent nucleon.

In these works, the term "kinematical" operators was introduced to describe TMCs in the OPE, distinguishing them from "dynamical" higher twist operators that remain in the OPE as $(x_A M_A^2/Q^2) \to 0$ [93, 94].

Higher Twist (HT): At this point, we emphasize the distinction between TMCs and (genuine) higher-twist corrections. TMCs originate from the modification of kinematics due to the presence of a hadron's mass M_A. The behavior of amplitudes in the massless limit implies that mass terms are relatively suppressed by the hard scale of the process, resulting in corrections that scale as powers of (M_A^2/Q^2). In contrast, higher-twist corrections can arise from the exchange of additional bosons (gluons and pions) between the

hard process and the hadronic remnant. Consequently, these corrections are *dynamical*. Examples include double parton scattering and color (re)connection between active and spectator quarks. The presence of an *extra* boson exchange in higher-twist processes introduces additional propagators, resulting in amplitudes that are also suppressed by the hard scale of the process.[1] This leads to corrections in powers of $(\Lambda^2_{\rm NP}/Q^2)$, where $\Lambda_{\rm NP}$ is the characteristic non-perturbative scale of a hadron. Since, in practice, $\Lambda_{\rm NP}$ is often set to M_A, the two types of corrections can be confused despite their distinct origins. In some sense, TMCs in DIS *accidentally* have the same characteristic power-suppression as in higher-twist corrections.

We illustrate the differences between leading twist, TMCs, and higher twist in Fig. 2.1 for leading-order heavy quark Q production in the charged current process $\nu_\ell A \to \ell X$. In the left panel (a) is a depiction of the leading-twist process in the limit of vanishing TMCs, i.e., $k_T = 0$ and $(M_A/Q) \to 0$. In the center panel (b) is the leading-twist process with finite (M_A/Q) which yields TMCs and generates $k_T \sim \mathcal{O}(M_A)$ via a gluon emission that occurs at a time $\tau_{\rm TMC} \sim 1/k_T$ **before** the hard process (indicated with a yellow circle). In the right panel (c) is the same process but with a typical higher-twist correction, i.e., an additional parton (gluon) exchange between the hadron and the hard scattering. The characteristic time of the interactions is inversely proportional to the energy scale, (time) ~ 1/(Energy). For large Q, the hard scattering time scale $(1/Q)$ is short, while the characteristic time scale of the hadron dynamics $(1/\Lambda_{NP})$ is long. The probability of a second parton (gluon) participating in the hard interaction is proportional to the ratio of the time scales; thus, we expect higher twist contributions to be suppressed by powers of (Λ_{NP}/Q).

ACOT Formalism: Returning to the literature, we also observe that TMC prescriptions based on the OPE and the factorization approach of Ellis, Furmanski, and Petronzio were similarly compared for semi-inclusive processes in Ref. [100]. In the collinear parton model ($k_T = 0$), TMCs have been incorporated into the Aivazis-Collins-Olness-Tung (ACOT) formalism [101, 102], which is rigorously based on the factorization theorem including heavy quark masses [9]. Additionally, Ref. [103] presents a comprehensive QCD framework for evaluating tau-neutrino deep-inelastic charged-current cross sections, including next-to-leading order (NLO) corrections, charm production, tau-mass threshold effects, and target mass effects within the collinear approximation.

Proton TMCs: An earlier review [104] of TMCs for nucleons in unpolarized DIS culminated in a set of so-called master formulae [98]. We extend these formulae and

[1] For sufficiently inclusive observables, the contributions from soft gluons are subject to cancellations [22, 99].

present them below in Eq. (2.38) for *nuclei*. This set of equations is quite remarkable, organizing the rather complicated expressions in the aforementioned OPE-based papers into a simple, easy-to-use, and modular form. It is valid at all orders in perturbative QCD while still accounting for quark masses. In principle, TMCs derived for DIS off nucleons [89,90,98,104,105] apply also to DIS off nuclear targets since the OPE is (theoretically) independent of target states. However, previous discussions [104] do not address the subtle distinctions between a nucleon and nuclear targets. Moreover, the lack of consistent distinction between nucleons and nuclei in their respective kinematic notations, the potential for the spin of a generic nucleus to differ from 1/2, and the necessity of master equations such as those in Eq. (2.38) have prompted the question of whether nucleon-like expressions can be valid for nuclear targets. The ability to express these quantities conveniently in terms of *averaged nucleon kinematics* (as seen in Eq. (2.53)) has prompted the question of whether nucleon-like expressions can be valid for nuclear targets.

Nuclear TMCs: With an abundance of recent nuclear data from both electron and neutrino beams, and with the EIC and third-generation long-baseline oscillation experiments on the horizon, there is a strong need to rigorously revisit the derivation of TMCs to structure functions in charged-lepton and neutrino DIS with a particular focus on the nuclear case. One of the main goals for this work is to demonstrate the validity of the TMC master equations, as given in Eq. (2.38), for the case of massive nuclear targets within the framework of the OPE. In this thesis, we present the main results and refer to the review [48] for a detailed pedagogical derivation of TMCs from the OPE.

Comparing OPE and Parton Model: Another main goal of this work is to discuss and contrast the results for TMCs obtained in both the OPE and the collinear parton model; throughout this work, we consider them on equal footing. As mentioned earlier, the collinear parton model is rigorously based on factorization theorems. These theorems provide field-theoretic definitions of PDFs and make statements about the error of the factorization approximation, which is generally inversely proportional to a positive power of a hard scale of the process. It is widely believed that collinear factorization remains valid in lepton-nucleus and proton-nucleus collisions, possibly with nuclear-enhanced higher twist terms [106]. However, the literature on factorization in the nuclear case is sparse and we consider this a working assumption. Moreover, the non-collinear parton model (where the TMCs were shown to be equivalent to the OPE results at leading power) is not covered by the factorization proofs mentioned above. While exploring these more theoretical questions is interesting and relevant, it is beyond the scope of this thesis. Here, we take the collinear parton model and the OPE for granted, and explore extensively what

happens when transitioning from nucleons to nuclei in DIS.

Organization: The starting point of our analysis is to consider the full nucleus as our target and apply only general symmetry principles, e.g., Lorentz invariance, to derive nuclear structure functions and their TMCs. This approach ensures that, until Sec. 2.4, there is no reference to or dependence on nucleon degrees of freedom. Furthermore, until Sec. 2.5 there is no reference to (or dependence on) the individual partonic degrees of freedom[2].

To do this, we first outline in Sec. 2.2 key kinematic relations and definitions in DIS of a lepton off a nucleus, $\ell_1(k_1) + A(p_A) \to \ell_2(k_2) + X(p_X)$. As depicted in Fig. 2.2, A is a nuclear target with mass number A, ℓ denotes either a charged lepton ($\ell^\pm$) or a neutrino (ν), and X represents all possible final-state hadrons.

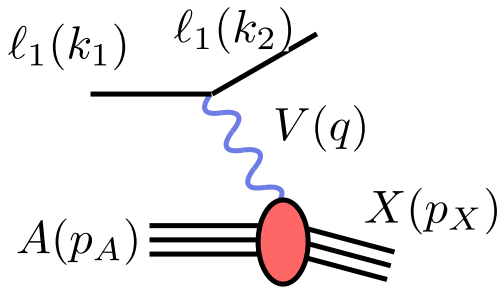
Next, in Sec. 2.3, we turn to nuclear structure functions in the OPE. More specifically, in Sec. 2.3.1, we examine the criteria for light-cone dominance in **nuclear** DIS, give a brief introduction to structure functions in the OPE in Sec. 2.3.2 and then present in Sec. 2.3.3 a master formula for nuclear TMCs. The result is analogous to the one in the nucleon case [104], but expressed in terms of the nuclear scaling variable x_A and the mass of the nucleus M_A.

In Sec. 2.4, we go on to perform a rescaling in order to express our formula for nuclear TMCs in terms of the more familiar averaged nucleon quantities $x_N = Ax_A$ and $M_N = M_A/A$. This is the key step that allows us to compare structure functions across different nuclei, including the proton, in a meaningful way. As an important consequence of rescaling, we obtain an alternative formula for nuclear TMCs that is universally applicable to all nuclei.

In Sec. 2.5, we introduce the QCD-improved parton model, both before and after rescaling. We show that rescaling at the hadronic level and the partonic level are consistent, and we comprehensively discuss the relationship between nuclear structure functions and nuclear PDFs. We believe that the discussions in Secs. 2.4 and 2.5, although relatively straightforward, are original and have not been fully detailed in the existing literature. Our motivation for these discussions stems from the need to establish proper theoretical definitions of nuclear structure functions and nuclear parton distribution functions as they are commonly and intuitively used in the literature.

In Sec. 2.6, we derive nuclear TMCs in the ACOT light cone formalism, which has

[2]Explicitly, until Sec. 2.5 we consider operators and matrix elements derived from quark, antiquark, and gluon fields but do not identify these as partons, nor identify structure functions as combinations of PDFs, i.e., the parton model.



Kinematic variable	Description
$\nu_A = \frac{q\cdot p_A}{M_A} \overset{lab}{=} E_{\ell_1} - E_{\ell_2}$	Lepton energy loss in the nucleon rest frame (laboratory frame)
$y_A = \frac{q\cdot p_A}{k_1\cdot p_A} \overset{lab}{=} \frac{\nu_A}{E_{\ell_1}}$	Inelasticity $y_A \in [0,1]$
$Q^2 = -q^2 > 0$	Squared boson momentum transfer
$x_A = \frac{Q^2}{2p_A\cdot q} = \frac{Q^2}{2M_A\nu_A}$	Bjorken x_A with $x_A \in [0,1]$
$W_A^2 = (p_A+q)^2 = M_A^2 + Q^2\frac{1-x_A}{x_A}$	Squared mass of the recoil system
$s = (k_1+p_A)^2 = \frac{Q^2}{x_A\, y_A} + M_A^2 + m_\ell^2$	Squared center-of-mass System (CMS) energy

Figure 2.2: We consider the basic charged current (V=$W^\pm$) or neutral current (V=$\{\gamma, Z\}$) lepton-nucleus DIS process $\ell_1(k_1) + A(p_A) \to \ell_2(k_2) + X(p_X)$ where the incoming lepton can be a charged lepton (e, μ) or a neutrino (ν_e, ν_μ, ν_τ). Note, the energy loss symbol ν_A should not be confused with the neutrino symbols. The 4-momentum of the exchanged boson is denoted $q = k_1 - k_2$, and $p_A^2 = M_A^2$. Here, the "lab frame" denotes the rest frame of A, meaning that $p_A = (M_A, 0, 0, 0)$.

been used in past nCTEQ analyses [35, 38, 40–43, 107], and compare them to the OPE results.

Having provided rigorous theoretical definitions for the physical observables, we perform detailed numerical studies in Sec. 2.7. This includes a comparison to a selection of data, as well as cross section predictions for current and future DIS experiments. We provide a parameterization of the TMCs that is accurate at the sub-percent level and can be used with any underlying set of nuclear structure functions available in the massless parton model to obtain the full TMCs for any nucleus in a simple way. This parameterization

will hopefully be useful for the community investigating hadron structure.

Finally, in Sec. 2.8, we highlight the key observations of this analysis and present our conclusions.

2.2 Kinematics of lepton-nucleus DIS

In this section, we layout the definitions, notations, and baseline assumptions used throughout the rest of this chapter. While much of the following is standard material for modern textbooks, many different conventions are used in the literature. We start with definitions of kinematic variables in Sec. 2.2.1. Then, in Sec. 2.2.2 we move to the DIS cross section expressed in terms of leptonic and hadronic tensors.

2.2.1 Definition of kinematic variables

Consider a basic charged-current (CC) or neutral-current (NC) lepton-nucleus DIS process, as shown in figure 2.2:

$$\ell_1(k_1) + A(p_A) \to \ell_2(k_2) + X(p_X) \ . \tag{2.1}$$

Here, $\ell_1(k_1)$ is a charged ($\{e^{\pm}, \mu^{\pm}\}$) or neutral ($\{\nu_e, \nu_\mu, \nu_\tau\}, \{\overline{\nu_e}, \overline{\nu_\mu}, \overline{\nu_\tau}\}$) lepton with momentum k_1. In the following, we will neglect the lepton mass[3]. $A(p_A)$ is the nuclear target, considered to be unpolarized, with the atomic number Z and the mass number A, carrying the momentum p_A. Furthermore, $\ell_2(k_2)$ is the outgoing lepton, with momentum k_2 and $X(p_X)$ is a n-body final state, a fragment of the initial nucleus A, with the momentum $p_X = \sum_{i=1}^{n} p_{Xi}$. We work at LO in the electroweak theory, in the so-called *one-boson-exchange approximation*, which means that the leptonic system ($\ell_1\ell_2$) probes the hadronic system via the exchange of a single (off-mass-shell), electroweak boson $V = \gamma$, Z for NC or $W^{\pm}$ for CC interactions with (time-like) momentum $q = k_1 - k_2 = p_X - p_A$.

The kinematic terms defined in the theoretical chapter 1 section 1.2.1 are still valid but we will identify the hadron $h(p)$ of the equation (1.16) as a nucleus A. We rewrite the variables for the nuclear case in the figure 2.2 and add new useful quantities for the

[3]However, we allow for the possibility of a massive outgoing lepton ($\tau^{\pm}$).

upcoming development (ξ_A and r_A) which we remind below:

$$k_1^2 = m_{\ell_1}^2, \quad k_2^2 = m_{\ell_2}^2, \quad p_A^2 = M_A^2 > 0, \quad Q^2 = -q^2 > 0, \tag{2.2a}$$

$$s = (k_1 + p_A)^2 = m_{\ell_1}^2 + M_A^2 + 2p_A \cdot k_1, \tag{2.2b}$$

$$x_A = \frac{Q^2}{2p_A \cdot q}, \quad \nu_A = \frac{p_A \cdot q}{M_A}, \quad y_A = \frac{p_A \cdot q}{p_A \cdot k_1}, \quad Q^2 = (s - m_{\ell_1}^2 - M_A^2) x_A y_A, \tag{2.2c}$$

$$W_A^2 = p_X^2 = (p_A + q)^2 = M_A^2 - Q^2 + 2\nu_A M_A = M_A^2 + Q^2 \frac{1 - x_A}{x_A} \geq M_A^2 . \tag{2.2d}$$

Importantly, the above are defined at the level of nuclei.Here, $\sqrt{s}$ is the hadron-level center-of-mass (cm) energy. Moreover, the (dimensionless) Bjorken scaling variable $x_A \in [0, 1]$ is defined such that in the elastic scattering limit, i.e., where X remains as an intact and on-shell A, the condition $W_A^2 = M_A^2$ leads to $2(p_A \cdot q) = Q^2$ such that $x_A = 1$. In the inelastic scattering case, i.e., where A disassociates, the condition $W_A^2 > M_A^2$ leads to $2(p_A \cdot q) > Q^2$, or that $x_A < 1$. The quantity ν_A, which has mass dimension 1, is a measure of the energy carried by the virtual state V. In the rest frame of A, it simplifies to the energy transfer from the leptons to the hadronic system: $\nu_A|_{lab} = k_1^0 - k_2^0$. Here and below, the *lab frame* denotes the rest frame of A, meaning that $p_A = (M_A, 0, 0, 0)$. The (dimensionless) inelasticity variable spans $y_A \in [0, 1]$, and in the lab frame we have $y_A = \frac{\nu_A}{k_1^0}$. In the rest frame of A, it is a measure of ℓ_2's energy; in other frames, y_A is a measure of ℓ_2's outgoing angle. Note, the inelasticity variable $y_A \in [0, 1]$ should not be confused with the rapidity or pseudorapidity variables $y, \eta \in [-\infty, \infty]$ which measure the particle's direction.

We will now define the Nachtmann variable [108] ξ_A, a very important variable used throughout this chapter:

$$\xi_A = \frac{2}{1 + r_A} x_A \equiv R_M x_A , \tag{2.3}$$

where

$$R_M = \frac{2}{1 + r_A}, \tag{2.4}$$

and

$$r_A = \sqrt{1 + 4x_A^2 M_A^2 / Q^2} = 1 + \frac{2x_A^2 M_A^2}{Q^2} + \mathcal{O}\left(\frac{4x_A^4 M_A^4}{Q^4}\right). \tag{2.5}$$

R_M is the target-mass-dependent factor relating the Bjorken scaling variable x_A to the Nachtmann scaling variable ξ_A. In other words, ξ_A is the true scaling variable in DIS when one accounts for the mass of the hadronic target [89, 90]. The "M" subscript of R_M indicates this quantity depends on the hadron mass M_A. In the limit that $(M_A/Q) \to 0$

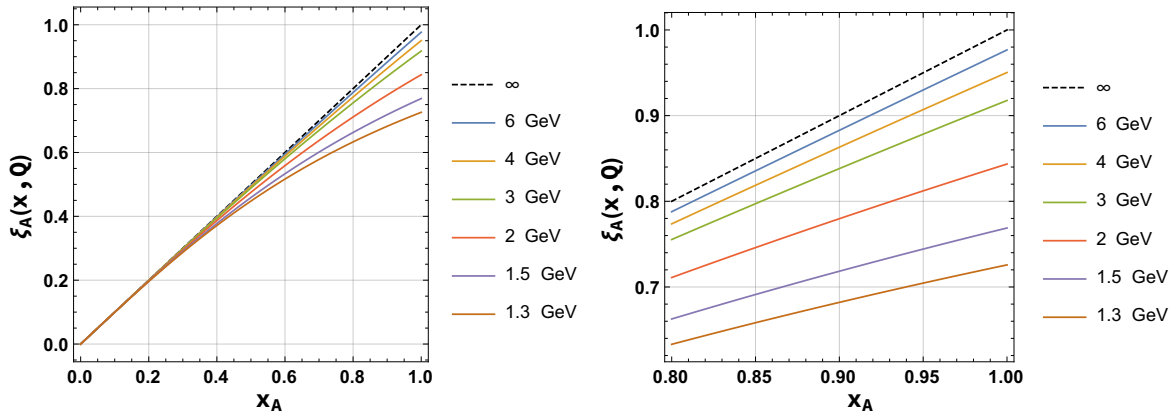


Figure 2.3: (a) The Nachtmann scaling variable $\xi_A(x_A, Q)$ (see Eqs. (2.3)-(2.5)) as a function of the Bjorken scaling variable x_A for a selection of Q values (in GeV), as indicated in the legend, with $M_A = M_{\text{proton}}$. (b) Same as (a) but for large x_A. We observe that for large x_A and small Q values, the target mass M_A modifies the true (Nachtmann) scaling variable ξ_A relative to the usual Bjorken scaling variable x_A, *c.f.*, Ref. [104].

or that $x_A \to 0$, we have $r_A \to 1$, $R_M \to 1$, and $\xi_A \to x_A$ since:

$$\xi_A = x_A \left[1 - \frac{x_A^2 M_A^2}{Q^2} + \mathcal{O}\left(\frac{x_A^4 M_A^4}{Q^4}\right)\right] . \tag{2.6}$$

We see in Eq. (2.6) that the Nachtmann variable ξ_A is essentially the Bjorken scaling x_A modified by the target mass M_A. The r_A and R_M factors are ultimately kinematical in origin.

Qualitatively, the R_M factor indicates that (slightly) less *3-momentum* is available than suggested by x_A and simplified arguments based purely on energy conservation. In other words, the leading kinematics in DIS are modified by $\mathcal{O}(x_A M_A^2/Q^2)$ terms that can be neglected at high momentum-transfers. As discussed and demonstrated in later sections of this work, this is often remedied in practice by the appropriate substitution of x_A with ξ_A. The difference between ξ_A and x_A is illustrated in Fig. 2.3, where ξ_A is shown as a function of x_A for a selection of exchange-boson virtualities Q with $M_A = M_{\text{proton}}$. Fig. 2.3(a) displays the full (ξ_A, x_A) plane while Fig. 2.3(b) highlights the large x region, where the difference between ξ_A and x_A is more pronounced. Overall, x_A and ξ_A are mostly indistinguishable for $x_A \lesssim 0.4 - 0.5$ over a large range of Q and M_A. At around $x_A = 0.8$, one has approximately $\xi_A \approx 0.63$ (0.71) [0.76] for $Q = 1.3$ (2) [3] GeV. For the same Q values, one has approximately $\xi_A \approx 0.76$ (0.78) [0.84] at around $x_A = 0.9$.

2.2.2 DIS cross sections

As we learned in Sec. 1.2.1, under strong but general assumptions, we can write any (sufficiently inclusive) DIS cross-section $\ell_1 + A \xrightarrow{V} \ell_2 + X$ as a combination of kinematical factors and hadronic structure functions, which parameterize the internal dynamics of A.

The following discussion relies exclusively on kinematics and symmetries, including Lorentz symmetry. The only approximation involved in constructing structure functions *from experimentally observable kinematics* is the assumption that DIS is mediated solely by the exchange of a single electroweak boson[4]. Except this assumption, the expressions derived from kinematics remain entirely general, encompassing all twist contributions, at all orders in the strong coupling constant, and all effects of hadron and quark masses. Notably, the parton model is not a requisite for these considerations.

We also learned in Sec. 1.2.1 that the differential DIS cross section ($d\sigma$) is proportional to the product of a leptonic tensor with a hadronic tensor. Here we just rewrite in the nuclear case

$$d\sigma \;\propto\; \tilde{W}^{A}_{\mu\nu}\, L^{\mu\nu}\,. \tag{2.7}$$

Ambiguities frequently arise in the literature regarding whether a given quantity in lepton-nucleus scattering is defined at the level of the *nucleus* or the *nucleon*. A central objective of this work is the elucidation of the relationship between the nuclear and nucleon levels in the DIS limit. Therefore, the tilde (~) notation is introduced here, the use of it will be more clear later in Sec. 2.4, when we perform a rescaling of *nuclear* momenta to *average nucleon* momenta. This rescaling is a key-point of the demonstration.

The leptonic tensor $L^{\mu\nu}$ is the square of the $\ell_1 \to \ell_2 V$ splitting matrix element (meaning the matrix element of the leptonic current). It can be constructed in a standard way (e.g. trace technology or helicity amplitudes) and its final expression only depends on which current is implied (and also if we take into account the lepton mass). The general expression is

[4]Investigations relaxing this assumption can be found in [109–112].

Vertex	Coupling strength	g_R^ℓ	g_L^ℓ	g_V^ℓ	g_A^ℓ
$V-\ell_1-\ell_2$	$\tilde{g}$	$(g_V^\ell+g_A^\ell)$	$(g_V^\ell-g_A^\ell)$	$\frac{(g_R+g_L)}{2}$	$\frac{(g_R-g_L)}{2}$
$\gamma-\ell-\ell$	eQ^ℓ	1	1	1	0
$Z-\ell-\ell$	$\frac{g_W}{\cos\theta_W}$	$-Q^\ell\sin^2\theta_W$	$(T_3^\ell)_L-Q^\ell\sin^2\theta_W$	$\frac{1}{2}(T_3^\ell)_L-Q^\ell\sin^2\theta_W$	$-\frac{1}{2}(T_3^\ell)_L$
$W-\ell_1-\ell_2$	$\frac{g_W}{\sqrt{2}}$	0	1	$\frac{1}{2}$	$-\frac{1}{2}$

Table 2.1: Electroweak chiral couplings and coupling strength normalizations used for fermions ℓ_1,ℓ_2 with weak isospin charge $(T_3^\ell)_L=\pm1/2$ and electric charge Q^ℓ, with normalization $Q^e=-1$.

$$\sum_{\{\lambda\}} L^{\mu\nu} = \sum_{\{\lambda\}} \left(L^\dagger\right)^\mu L^\nu = \sum_{\{\lambda\}} \langle\ell_1|\left(J^\dagger_{\ell_2\ell_1}(0)\right)^\mu|\ell_2\rangle\langle\ell_2|J^\nu_{\ell_2\ell_1}(0)|\ell_1\rangle \tag{2.8}$$

$$\begin{aligned} &= 4\tilde{g}^2\Big\{(g_V^2+g_A^2)(k_1^\mu k_2^\nu+k_1^\nu k_2^\mu)-\left[(g_V^2+g_A^2)(k_1\cdot k_2)-(g_V^2-g_A^2)m_1m_2\right]g^{\mu\nu} \\ &\quad -2i(g_Vg_A)k_{1\alpha}k_{2\beta}\epsilon^{\mu\nu\alpha\beta}\Big\} \end{aligned} \tag{2.9}$$

$$\begin{aligned} &\overset{m_1,m_2\to0}{=} 4\tilde{g}^2\Big\{(g_V^2+g_A^2)(k_1^\mu k_2^\nu+k_1^\nu k_2^\mu)-(g_V^2+g_A^2)(k_1\cdot k_2)g^{\mu\nu} \\ &\quad -2i(g_Vg_A)k_{1\alpha}k_{2\beta}\epsilon^{\mu\nu\alpha\beta}\Big\}, \end{aligned} \tag{2.10}$$

where Eq. (2.9) is the general case and Eq. (2.10) is the massless limit. Table 2.1 shows the different possibilities, depending on the current. Note that in the simple case of massless charged leptons exchanging a photon with the nucleus A, we retrieve the well-known expression

$$\sum_{\{\lambda\}} L^{\mu\nu}\Big|_{\text{QED}} = 4e^2\Big\{k_1^\mu k_2^\nu+k_1^\nu k_2^\mu-(k_1\cdot k_2)g^{\mu\nu}\Big\}, \tag{2.11}$$

where $\{\lambda\}$ represents the sum over spins of external particles; spin-averaging is not yet performed here.

We can decompose the hadronic tensor $\tilde{W}^A_{\mu\nu}(p_A,q)$ into a sum over tensor-valued coefficient functions multiplied by dimensionless, scalar-valued functions $\tilde{W}_i$. For unpolarized targets, $\tilde{W}^A_{\mu\nu}(p_A,q)$ can be a function of external momenta p_A and q only. As already mentioned in Chap. 1, the $\tilde{W}_i$ are known in the literature as **structure functions** since they parameterize the structure of nucleons and nuclei.

However, unlike Eq. (1.29), the state to which the current $J(z)$ is applied is now the full state of the nucleus $|A\rangle$. One obtains for an unpolarized (spin-averaged) nuclear

target, the hadronic tensor

$$\tilde{W}^A_{\mu\nu}(p_A,q) \equiv \frac{1}{4\pi}\sum\!\!\!\!\!\!\int d^4z\; e^{iq\cdot z}\langle A(p_A)|\, J_\mu(z)|X(p_X)\rangle\langle X(p_X)|J_\nu(0)\,|A(p_A)\rangle \tag{2.12}$$

$$= -g_{\mu\nu}\tilde{W}_1 + \frac{p_{A\mu}p_{A\nu}}{M_A^2}\tilde{W}_2 - i\epsilon_{\mu\nu\rho\sigma}\frac{p_A^\rho q^\sigma}{M_A^2}\tilde{W}_3 + \frac{q_\mu q_\nu}{M_A^2}\tilde{W}_4 + \frac{p_{A\mu}q_\nu + p_{A\nu}q_\mu}{M_A^2}\tilde{W}_5 + \frac{p_{A\mu}q_\nu - p_{A\nu}q_\mu}{M_A^2}\tilde{W}_6\ . \tag{2.13}$$

In Eq. (2.12), the normalization factor $1/(4\pi)$ is conventional and the sum and integral ($\sum\!\!\!\!\!\!\int$) run over all discrete and continuous configurations of X, implying that $\tilde{W}^A_{\mu\nu}(p_A,q)$ is inclusive with respect to X.With further strong but reasonable assumptions, the number of structure functions appearing in Eq. (2.13) can be reduced in real-life calculations [113–115]. For instance:

- $\tilde{W}_3$ is only nonzero if parity symmetry is violated; it therefore only appears for W and Z boson exchanges.
- $\tilde{W}_4$ and $\tilde{W}_5$ contain factors of q_μ and q_ν. This implies that the contractions with the leptonic tensor $L^{\mu\nu}\cdot q_\mu$ and $L^{\mu\nu}\cdot q_\nu$ are proportional to the masses of ℓ_1 and ℓ_2, entering in the cross section at $\mathcal{O}(\frac{m_l^2}{s})$ and are therefore negligible.
- $\tilde{W}_6$ does not appear in the cross section and is therefore irrelevant. In addition, $\tilde{W}_6$ would only be nonzero if QCD violated time-reversal (or charge-parity) symmetry.

In a modern notation, one uses $\tilde{F}_i$ rather than $\tilde{W}_i$ for SFs like

$$\left\{\tilde{F}_1,\ \tilde{F}_2,\ \tilde{F}_3,\ \tilde{F}_4,\ \tilde{F}_{5,6}\right\} = \left\{\tilde{W}_1,\ \frac{Q^2}{2x_AM_A^2}\tilde{W}_2,\ \frac{Q^2}{x_AM_A^2}\tilde{W}_3,\ \frac{Q^2}{2M_A^2}\tilde{W}_4,\ \frac{Q^2}{2x_AM_A^2}\tilde{W}_{5,6}\right\}\ . \tag{2.14}$$

These two notations are equivalent; however, the newer one highlights the Q^2 dependence and the scaling by extracting/showing that $\tilde{F}_i$ depend only on x_A.

In the following, we will focus on $\tilde{F}_1$, $\tilde{F}_2$ and $\tilde{F}_3$ only. A full discussion of the other SFs can be found in the appendix of [48].

Spin 1 and greater: An important remark; contrary to the nucleon target with spin-$\frac{1}{2}$, nuclear targets can have a spin 1 or greater and in general implies some additional structure functions. However, in our analysis we study the scattering of *unpolarized* leptons off an *unpolarized* nucleus of spin greater or equal to $\frac{1}{2}$ and in this case the

hadronic tensor keeps the same decomposition as in the spin-$\frac{1}{2}$ case and we will not address explicitly these additional functions.

2.3 Nuclear structure functions in the OPE

In this section, we discuss nuclear structure functions and their TMCs in the context of the OPE. In Sec. 2.3.1, we define the criteria for light-cone dominance specifically for nuclear targets, whose masses can readily exceed $\mathcal{O}(50-100)$ GeV. Sec. 2.3.2 outlines the construction of nuclear structure functions and their corresponding TMCs, with a more detailed derivation available in [48]. Finally, in Sec. 2.3.3, we present the main formulae for nuclear TMCs expressed in terms of the nuclear quantities x_A and M_A.

2.3.1 Light-cone dominance in the nuclear case.

A crucial step in deriving structure functions for nucleons and nuclei within the OPE involves expanding the product of currents that comprise the hadronic tensor of Eq. (2.12). This expansion, however, requires that one is kinematically in the limit of light-cone dominance, i.e., $z^2 \sim 0$, where z is the Fourier conjugate of the DIS momentum q. While this is well established in the case of a nucleon target,[5] demonstrating light-cone dominance for nuclear targets requires care due to the relative sizes of $Q^2 = -q^2 > 0$ and M_A^2.

In many constructions of PDFs and related quantities in DIS, light cone-dominance is described as corresponding to the momentum configuration

$$Q^2 \sim |p_N \cdot q| \gg M_N^2 \,, \tag{2.15}$$

where $p_N = p_A/A$ and $M_N = M_A/A$ are, respectively, the four-momentum and mass of a single nucleon, *c.f.*, Table 2.2. Under this condition, naively replacing p_N and M_N by the nuclear momentum p_A and nuclear mass M_A would imply (incorrectly) that light-cone dominance is only satisfied when

$$Q^2 \sim |p_A \cdot q| \gg M_A^2 \,. \qquad \text{[Incorrect]} \tag{2.16}$$

This wrong condition suggests increasingly large Q^2 are needed for increasing A, and downplays the validity of both perturbative QCD and the parton model. For example: Eq. (2.16) implies that $Q \gg 50$ GeV, which is more than half the Z boson's mass, would

[5] A demonstration of the light cone dominance of DIS can be found, for example, in the textbook by Muta [91, 92], see (2nd edition) pages 228-229 and 262.

be needed to describe DIS data for ^{56}Fe, or that $Q \gg 180$ GeV, which is beyond the top quark's mass, would be needed for ^{197}Au. Fortunately, this is not the case as Eq. (2.15) is the correct condition even in the nuclear case, as we argue in the following.

To show this, we apply the arguments of Refs. [91,92] for light-cone dominance in DIS with a nucleon to the case of an arbitrary *nucleus*. Our goal is to identify the dominant contribution to the Fourier integral in Eq. (2.12) in the deeply inelastic limit:

$$Q^2 \to \infty, \quad \nu_A \to \infty, \quad \text{such that} \quad \frac{Q^2}{\nu_A} = 2M_A x_A \quad \text{is fixed.} \tag{2.17}$$

Note that in terms of averaged nuclear quantities, one has $Q^2/\nu_A = Q^2/\nu_N = 2M_A x_A = 2M_N x_N$. That is to say, a bigger nuclear mass M_A is compensated by a smaller Bjorken variable x_A such that the fixed quantity Q^2/ν is independent of the atomic number.

Following Refs. [91,92], we examine $\tilde{W}^A_{\mu\nu}(p_A, q)$ in Eq. (2.12), and identify the integration regions that give rise to the dominant contributions. In the DIS limit of Eq. (2.17), as $|q \cdot z| \to \infty$ the exponential of the Fourier integral oscillates without bound and thus makes a vanishing contribution to the integral. Therefore, we need only to consider the region with finite $|q \cdot z|$ in the deeply inelastic regime.

In the target's rest frame, $\nu_A = q_0$. Defining the quantity $\mathscr{z} = \vec{q} \cdot \vec{z}/|\vec{q}|$, we have

$$q \cdot z \;=\; q_0 z_0 - \vec{q} \cdot \vec{z} \;=\; \nu_A \left(z_0 - \frac{|\vec{q}|}{\nu_A} \mathscr{z} \right) \;=\; \nu_A \left(z_0 - \sqrt{1 + Q^2/\nu_A^2}\ \mathscr{z} \right) , \tag{2.18}$$

since $|\vec{q}|^2 = \nu_A^2 + Q^2$. We note again that this equation is independent of A, as can be seen from the left side of the equation.

In the DIS limit, $Q^2/\nu_A^2 = (2M_A x_A)/\nu_A$ is small and we can expand the square root:

$$q \cdot z \;=\; \nu_A (z_0 - \mathscr{z}) \; - M_A x_A \mathscr{z} \; + \; \mathcal{O}\left(\frac{M_A^2 \mathscr{z}}{\nu_A} \right) . \tag{2.19}$$

Here, the target mass M_A appears but only in combination with x_A and always satisfies $M_A x_A = M_N x_N$. In order to keep $|q \cdot z|$ finite in the deeply inelastic limit, each term on the right hand side of Eq. (2.19) must separately be finite. Being separately infinite requires the scaling $(z_0 - \mathscr{z}) \sim (\mathscr{z}/\nu_A) \to \infty$, which cannot be consistently satisfied. Since $M_A x_A$ is fixed, $\mathscr{z}$ itself must be finite. Therefore, for some constants $c > 0$ and $d > 0$, one has

$$|z_0 - \mathscr{z}| < c/\nu_A \quad \text{and} \quad |\mathscr{z}| < d/(x_A M_A) . \tag{2.20}$$

The first inequality implies that $|z_0| < |\mathscr{z}| + c/\nu_A$. After squaring and using $\mathscr{z}^2 = \vec{z}^2 - z_\perp^2 < \vec{z}^2$,

where $z_\perp$ is the component of z orthogonal to $\vec{q}$, we obtain the inequality chain

$$z_0^2 \;<\; (|\vec{z}| + c/\nu_A)^2 = \vec{z}^2 + \frac{2c|\vec{z}|}{\nu_A} + \mathcal{O}\left(\frac{1}{\nu_A^2}\right) \;<\; \vec{z}^2 + \frac{2cd}{(x_A M_A \nu_A)} + \mathcal{O}\left(\frac{1}{\nu_A^2}\right). \tag{2.21}$$

To obtain the rightmost bound, we used the second inequality in Eq. (2.20). Using the rightmost equality in Eq (2.17), we obtain the final result:

$$z^2 \;=\; z_0^2 - \vec{z}^2 \;<\; \frac{2cd}{(x_A M_A \nu_A)} + \mathcal{O}\left(\frac{1}{\nu_A^2}\right) \;=\; \frac{4cd}{Q^2} + \mathcal{O}\left(\frac{1}{\nu_A^2}\right). \tag{2.22}$$

We therefore find that the dominant region remains $0 \le z^2 \le (4cd/Q^2)$, and is independent of the target. This assumes, of course, that no extra A dependence is hidden in the constants c and d. Such an assumption, however, is safe upon inspecting the above derivation. For example: suppose there were separate (minimal) c_A and d_A for each A that satisfied the inequality of Eq. (2.20). Since there are finite many A, we can simply take the largest c_A and d_A, and then define $c = \max\{c_A\}$ and $d = \max\{d_A\}$. This implies that there is a large enough c and d such that Eq. (2.20) holds independent of A.

In light of Eq. (2.22), the requirement of Eq. (2.16) is obviously too harsh. Instead, we advocate that Eq. (2.15) should be understood as requiring $Q^2 \gtrsim$ a few [6] GeV^2, independent of the hadronic target, in order to justify applying perturbative QCD and the parton model. Then, given a non-perturbative (NP) scale $\Lambda_{\rm NP} \lesssim 1$ GeV, which parametrizes the onset of hadronization, light-cone dominance in DIS for a nucleus A occurs when

$$Q^2 \;\sim\; |p_A \cdot q| \;\gg\; \Lambda_{\rm NP}^2. \tag{2.23}$$

Under this criterion, Eq. (2.15) is automatically satisfied since nucleon masses scale as $M_N \sim \Lambda_{\rm NP}$. At smaller Q^2, but still satisfying $Q^2 \gtrsim \Lambda_{\rm NP}^2$, perturbative techniques and the parton model can still be employed as evident by the phenomenon of "precocious scaling" [89,90,97]. This denotes the onset of Bjorken scaling at moderate energies despite $\alpha_s(\mu_r = Q) \lesssim \mathcal{O}(1)$ being sizable, and follows from the fact that light quark masses m_q are comparable to the scale of the QCD Landau pole, $\Lambda_{\rm QCD} \sim \mathcal{O}(0.1)$ GeV.

Such a separation of scales implies the inequality

$$Q^2 \;\gtrsim\; \Lambda_{\rm NP}^2 \;\gg\; \Lambda_{\rm QCD}^2 \;\sim\; m_q^2\,, \tag{2.24}$$

where we take m_q to be the light-quark constituent mass at soft quark momenta. This

[6] What *a few* precisely means has to be answered phenomenologically.

demonstrates that, up to corrections of order $\mathcal{O}(\Lambda^2_{\rm QCD}/Q^2)$, light quarks can be approximated as free and massless partons, even for low Q^2. However, when such Q^2 is probing a nuclear target A at large x_A, it is paramount to stress that $x_A^2 M_A^2/Q^2$ is not guaranteed to be small or below unity. In this case, TMCs are important.

2.3.2 Structure functions in the OPE in a nutshell

In the following, we briefly outline the key elements of the OPE required to obtain expressions for structure functions with TMCs at leading twist $\tau = 2$. Formally, the (twist 2) TMCs are $\mathcal{O}(x_A M_A^2/Q^2)$ corrections to (twist 2) structure functions that are generated by the *same* (twist 2) operators. The only difference is that kinematic factors involving the mass M_A^2 are taken into account, as can be seen in Eq. (2.34) when j is taken higher than 0. For this reason they are sometimes called *kinematical* power corrections [93, 94]. Such power corrections can be isolated from *dynamical* power corrections, by setting powers of (M_A^2/Q^2) in the OPE to zero (or equivalently including only the $j = 0$ terms in Eq.(2.34)). We generally follow the early literature of Refs. [89, 90, 97, 108, 114, 116], which shed light on the underlying dynamics of hadrons, but note that TMCs can also be derived using diagrammatic methods and finite parton k_T [93, 94]. While we only outline the derivation here, a fuller derivation of leading twist nuclear structure functions and their TMCs in the OPE is provided in the Appendices of [48].

Constructing TMCs for structure functions in the OPE provides a systematic organization of short and long distance physics that does not manifestly rely on the perturbativity of QCD such that it is possible to obtain results that are all orders in α_s. Inspired by the seminal work of Refs. [89, 90], the construction of structure functions follows by first recognizing that the inclusive hadronic tensor describing $\gamma^* A \to X$ scattering in the deeply inelastic limit,[7]

$$\tilde{W}^A_{\mu\nu}(p_A, q) \;=\; \frac{1}{4\pi} \int d^4z \; e^{iq\cdot z} \langle A | J^\dagger_{XA\mu}(z) \; J_{XA\nu}(0) | A \rangle \;, \tag{2.25}$$

which is equivalent to the expression in Eq. (2.12), and the time-ordered amplitude for the virtual Compton scattering process $\gamma^* A \to \gamma^* A$ in the short-distance limit

[7] The subscripts XA on the vector currents J_μ, J_ν are simply labels to denote the hadronic current that takes A to X. The current is independent of states A, X but depends on whether the current is electromagnetic or weak.

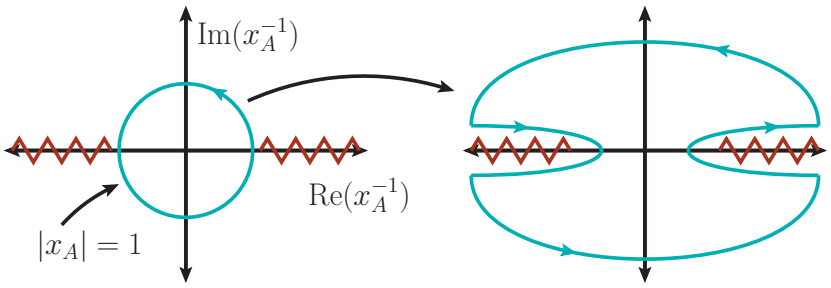


Figure 2.4: **[Left]** A contour (circle) along the radius of convergence of the time-ordered matrix element $\tilde{T}^A_{\mu\nu}(p_A, q)$ in the complex $(1/x_A)$ plane for $\gamma^* A \to \gamma^* A$ scattering. Branch cuts (sawtooth lines) are along the real axis at $x_A^{-1} < -1$ and $x_A^{-1} > 1$. **[Right]** A deformation of the contour along the branch cuts with vanishing arcs.

$$\tilde{T}^A_{\mu\nu}(p_A, q) \equiv \int d^4z\, e^{iq\cdot z}\, \langle A|\mathcal{T} J^\dagger_{A\mu}(z)\; J_{A\nu}(0)|A\rangle\;, \tag{2.26}$$

$$\begin{aligned} &= -\, g_{\mu\nu}\Delta\tilde{T}^A_1 + \frac{p_{A\mu}p_{A\nu}}{M_A^2}\Delta\tilde{T}^A_2 - i\epsilon_{\mu\nu\alpha\beta}\frac{p_A^\alpha q^\beta}{M_A^2}\Delta\tilde{T}^A_3 \\ &\quad + \frac{q_\mu q_\nu}{M_A^2}\Delta\tilde{T}^A_4 + \frac{(p_{A\mu}q_\nu \pm p_{A\nu}q_\mu)}{M_A^2}\Delta\tilde{T}^A_{5,6}\;, \end{aligned} \tag{2.27}$$

are related by the following dispersion relations [84, 114]

$$\tilde{T}^A_{\mu\nu}(p_A, q)\Big|^{(1/x_A)+i\varepsilon}_{(1/x_A)-i\varepsilon} = 4\pi\; \tilde{W}^A_{\mu\nu}(p_A, q), \quad \text{for} \quad x_A > 0\;, \tag{2.28a}$$

$$\tilde{T}^A_{\mu\nu}(p_A, q)\Big|^{(1/x_A)-i\varepsilon}_{(1/x_A)+i\varepsilon} = 4\pi\; \left[\tilde{W}^A_{\mu\nu}(p_A, -q)\right]^\dagger, \quad \text{for} \quad x_A < 0\;. \tag{2.28b}$$

In Eq. (2.27), $\tilde{T}^A_{\mu\nu}(p_A, q)$ is decomposed into coefficients of Lorentz structures in the same manner as $\tilde{W}^A_{\mu\nu}(p_A, q)$ in Eq. (2.13). As $\tilde{T}^A_{\mu\nu}(p_A, q)$ is time ordered $(\mathcal{T})$, the first and second relation in Eq. (2.28) are for DIS with respectively particles and anti-particles. We assume here and below that the hadronic currents J^μ_A and J^ν_A are always renormalized objects in QCD. The functions $\Delta\tilde{T}^A_i$ are obtained by contracting the tensor $\tilde{T}^A_{\mu\nu}(p_A, q)$ with suitable projection operators $P^{\mu\nu}_i$ such that

$$P^{\mu\nu}_i \tilde{T}^A_{\mu\nu} = \Delta\tilde{T}^A_i. \tag{2.29}$$

The distinction between the short-distance limit and the DIS limit is important. In the DIS limit, $(Q^2/M_A^2) \to \infty$ while $x_A = (Q^2/2p_A\cdot q)$ is fixed. In the short-distance limit, $(Q^2/M_A^2) \to \infty$ while x_A/Q is fixed, meaning that x_A grows with (Q/M_A); equivalently, the quantity $(1/x_A)$ is small with increasing (Q/M_A) in the short-distance limit. Formally, $\tilde{T}^A_{\mu\nu}(p_A, q)$ and $\tilde{W}^A_{\mu\nu}(p_A, q)$ are defined in two different limits but can be related through

analytic continuation. When analytically continued to complex values of x_A, $\tilde{T}^A_{\mu\nu}$ has desirable analytic properties for $|(1/x_A)| < 1$ (the short-distance limit), whereas $\tilde{W}^A_{\mu\nu}$ is meaningful for $|(1/x_A)| > 1$ (the DIS limit). Hence the discontinuities (branch cuts) of $\tilde{T}^A_{\mu\nu}$ correspond to the kinematics of DIS. Therefore, one can take $\tilde{T}^A_{\mu\nu}(p_A, q)$, decompose it into a contour integral over x_A^{-1} using Cauchy's integral formula, and deform the contour around the discontinuities at $x_A^{-1} < 1$ and $x_A^{-1} > 1$ as shown in Fig. 2.4. The saw-tooth lines along $|x_A^{-1}| > 1$ are the branch cuts of $\tilde{T}^A_{\mu\nu}$. Taking the difference of $\tilde{T}^A_{\mu\nu}$ when evaluated just below and above the branch cut, i.e., at the points $x_A^{-1} \pm i\varepsilon$ for $x_A^{-1} > 1$, gives $\tilde{W}^A_{\mu\nu}(p_A, q)$. A fuller derivation, can be found in the Appendix A.3 of [48] (see above Eq. (A.45)).

An important consequence of Eq. (2.28) is establishing a link between individual $\Delta\tilde{T}^A_i$ in Eq. (2.27) and the Mellin moments of the structure functions $\tilde{F}^A_i$ defined in Eq. (2.14). Following [114], we can write for $\tilde{F}_2$ [8]

$$\left(\frac{Q^2}{2x_A M_A^2}\right) \Delta\tilde{T}^A_2(x_A, Q^2) = -4i \sum_{N=0}^{\infty} \tilde{F}^{AN}_2(Q^2)\, x_A^{-N} \ . \tag{2.30a}$$

Here, $\tilde{F}^{AN}_i(Q^2)$ is the N^{th} Mellin moment of the structure function $\tilde{F}^A_i(x_A, Q^2)$. The normalizations of the Mellin transformation and its inverse are set by the definitions

$$F^N = \int_0^1 dz\ z^{N-1}\ F(z) \quad \text{with} \quad F(z) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} dN\ z^{-N}\ F^N \ , \tag{2.31}$$

where c has to be to the right of any poles of F^N in the complex N-plane. As a brief comment, the Mellin transform is an integral over the positive reals, while the inverse transform is an integral over the complex plane. We therefore assume that structure functions can be analytically continued. In practice, Mellin and inverse-Mellin transforms are implemented in the many computer codes that perform the DGLAP evolution equations in Mellin space, see e.g. Ref. [117].

The next step is to take the operator-product expansion of $\tilde{T}^A_{\mu\nu}$ in the short-distance limit [84]:

[8] A treatment of all six SFs $\tilde{F}^A_i$ is available in the appendices of [48].

$$\lim_{z\to 0}\tilde{T}^A_{\mu\nu}(p_A,q) \overset{\text{OPE}}{=} -2i\sum_{\iota,n} c^{\tau=2,\iota}_{\mu\nu\mu_1\ldots\mu_n}(q)\,\langle A|\mathcal{O}^{\mu_1\ldots\mu_n}_{\iota,\tau=2}|A\rangle \quad + \quad \text{power corrections at } \mathcal{O}(\tau>2) \tag{2.32}$$

$$\begin{aligned} &= -2i\sum_{k=k_{\min}}^{\infty}\Big[-2g^{\mu\nu}q_{\mu_1}q_{\mu_2}C_1^{2k} + g^{\mu}_{\mu_1}g^{\nu}_{\mu_2}Q^2C_2^{2k} - i\epsilon^{\mu\nu\alpha\beta}g_{\alpha\mu_1}q_\beta q_{\mu_2}C_3^{2k} \\ &\quad + 4\frac{q^\mu q^\nu}{Q^2}q_{\mu_1}q_{\mu_2}C_4^{2k} + 2(g^\mu_{\mu_1}q^\nu q_{\mu_2} \pm g^\nu_{\mu_1}q^\mu q_{\mu_2})C_{5,6}^{2k}\Big]\cdot\begin{Bmatrix}1\,; & k=1\\ \prod_{m=3}^{2k}q_{\mu_m}\,; & k>1\end{Bmatrix} \\ &\quad \times\frac{2^{2k}}{Q^{4k}}A^{2k}_{\tau=2}\tilde{\Pi}^{\mu_1\ldots\mu_{2k}} \quad + \quad \mathcal{O}(\tau>2). \end{aligned} \tag{2.33}$$

In Eq. (2.33), the summation over n is set to $n = 2k$. The starting point for the summation over k depends on the particular Wilson coefficient; specifically, $k_{\min} = 2$ for C_2^{2k}. while $k_{\min} = 1$ for all other C_ι^{2k}'s. In Eq. (2.32), the $\mathcal{O}^{\mu_1\ldots\mu_n}_{\iota,\tau}$ are composite local operators, consisting of quark fields, QCD covariant derivatives, and/or gluon field strengths. Operators with n un-contracted Lorentz indices carry a spin of n. Operators and their coefficients are organized according to their twist $\tau = d - n$, where d is the dimension of $\mathcal{O}^{\mu_1\ldots\mu_n}_{\iota,\tau}$ in the standard sense of dimension power counting in an effective field theory. In this study, we focus exclusively on the OPE at leading twist, i.e., $\tau = 2$. The index ι catalogs all operators $\mathcal{O}$ at a fixed spin and twist. All numerical prefactors are sequestered into effective Wilson coefficients $c^{\tau,\iota}_{\mu\nu\mu_1\ldots\mu_n}(q)$, which are only functions of the external momentum q. Finally, the functions $\Delta\tilde{T}^A_i(x_A,Q^2)$ are obtained by applying the projection operators (defined in Eq. (2.29)) to the right-hand side of Eq. (2.33).

The Wilson coefficients[9] C_ι^{2k}, which are the same for all possible nuclei, can be identified and matched to quantities calculated in perturbative QCD. The reduced matrix element $A^{2k}_{\tau=2}$ is the scalar-valued coefficient of the expectation value $\langle A|\mathcal{O}^{\mu_1\ldots\mu_{2k}}_{\iota,\tau=2}|A\rangle = A^{2k}_{\tau=2}\times\tilde{\Pi}^{\mu_1\ldots\mu_{2k}}$. Importantly, the Lorentz structure of this expectation value ($\tilde{\Pi}^{\mu_1\ldots\mu_{2k}}$) can only be a function of p_A as the q dependence is separated off elsewhere. Consequentially, $\tilde{\Pi}^{\mu_1\ldots\mu_{2k}}$ can be organized according to permutations and powers of momenta $p_A^{\mu_m}$

[9]In principle, the Wilson coefficients C_ι^{2k} also carry a species index f, which runs over the gluon and quark flavors. In this case, C_ι^{2k} should be replaced by $\sum_{f=u,g,\bar{d},\ldots}C_{if}^{2k}$, but this dependence is implicit here for simplicity.

and metrics $g^{\mu_m\mu_n}$, and scales as [89, 90]

$$\tilde{\Pi}^{\mu_1\cdots\mu_{2k}} = \sum_{j=0}^{k}(-1)^j \frac{(2k-j)!}{2^j(2k)!} \underbrace{\{g...g\}}_{j\ g^{\mu_n\mu_m}\text{'s}} \underbrace{\{p_A...p_A\}}_{(2k-2j)\ p_A^{\mu_n}\text{'s}} (p_A^2)^j \ + \text{ power corrections} \tag{2.34}$$

$$\sim \underbrace{\{p_A...p_A\}}_{2k\ p_A^{\mu_m}} + \underbrace{\{p_A...p_A\}}_{(2k-2)\ p_A^{\mu_m}} \underbrace{\{g...g\}}_{1\ g^{\mu_m\mu_n}} (p_A^2)^1 + \underbrace{\{p_A...p_A\}}_{(2k-4)\ p_A^{\mu_m}} \underbrace{\{g...g\}}_{2\ g^{\mu_m\mu_n}} (p_A^2)^2 + \ldots . \tag{2.35}$$

Focusing on Eq. (2.35), one sees that the momentum factor $\tilde{\Pi}^{\mu_1\cdots\mu_{2k}}$ scales as a tower of momentum and mass terms. For a fixed index k, the series ranges from the product of $2k$ instances of the nucleus' momentum $p_A^{\mu_m}$ and no instance of the nucleus' mass $p_A^2 = M_A^2$ (as shown in the leftmost term), to no instances of $p_A^{\mu_m}$ but k instances of $p_A^2 = M_A^2$. The brackets $\{...\}$ represent the collection of all possible permutations of Lorentz indices. The dimension of each term is $[\text{mass}]^{2k}$. Again, we neglect higher-order power corrections. Whether one retains $(M_A^2)^j$ terms or neglects them outright, keeping only $j = 0$, is the distinction between obtaining structure functions with TMCs or without TMCs (which we shall later label as **No-TMC** in Sec. 2.7). For example: after contracting all Lorentz indices in Eq. (2.33), one can neglect all factors of (M_A^2/Q^2). This is equivalent to truncating the summation in Eq. (2.34) at $j = 0$. Doing so and comparing the result to Eq. (2.30) recovers (see Appendices of [48] for details):

$$\tilde{F}_i^{AN} = \int_0^1 dy\ y^{N-1}\ \tilde{F}_i^A(y,Q^2) = C_i^N(Q^2)A_{\tau=2}^N + \text{ power corrections}, \quad \text{for} \quad i = 1,3,4,5, \tag{2.36a}$$

$$\tilde{F}_2^{A(N-1)} = \int_0^1 dy\ y^{N-2}\ \tilde{F}_2^A(y,Q^2) = C_2^N(Q^2)A_{\tau=2}^N + \text{ power corrections}. \tag{2.36b}$$

Non-trivially, Eq. (2.36) states that one can identify the product of the Wilson coefficient and reduced hadronic matrix element, $(C_\iota^{2k}A_{\tau=2}^{2k})$, as integer Mellin moments of structure functions, up to power corrections. Note that the remark in footnote 9 still holds. If one does not truncate the momentum factor $\tilde{\Pi}^{\mu_1\cdots\mu_{2k}}$ in Eq. (2.34) at $j = 0$ and instead retains the sums over all j, and hence retains the sum over powers of $p_A^2 = M_A^2$, then one obtains expressions that are analogous to Eq. (2.36).

To continue with the example started above, one can identify the Mellin transformation of the structure function F_2 with a nonzero target mass as the product $(C_{\iota=2}^{2k}A_{\tau=2}^{2k})$ with a coefficient proportional to $(M_A^2/Q^2)^j$:

$$\int_0^1 dx_A\, x_A^{N-2}\tilde{F}_2^{A,\text{TMC}}(x_A,Q^2) = \sum_{j=0}^{\infty}\left(\frac{M_A^2}{Q^2}\right)^j \frac{(N+j)!}{j!\ (N-2)!}\ \frac{C_2^{N+2j}A_{\tau=2}^{N+2j}}{(N+2j)(N+2j-1)}\ . \tag{2.37}$$

Intuitively, the right hand side of Eq. (2.37) states that structure functions with TMCs can be thought simply as the product (in moment space) of a structure function for a massless target, i.e., the $(C_i A_{\tau=2})$ factor, and a kinematical factor, i.e., everything else. For this reason, as already mentioned, the TMCs under discussion are sometimes called *purely kinematical* [93,94,118]. It is not obvious but, with the use of generating functions, the inverse-Mellin transform of Eq. (2.37) has a closed form [89, 90, 119]. That is to say, one can obtain concise expressions for structure functions with TMCs in terms of structure functions for a massless target in x-space. These results are summarized in the following section for F_1^A, F_2^A, and F_3^A, and in [48] for all F_i.

2.3.3 Master formulas

Following the procedure of Sec. 2.3.2, one obtains a set of master formulae for (twist-2) target mass corrections to SFs for **nuclei**. Using the notations of [98, 104], the general equation is

$$\tilde{F}_j^{A,\mathrm{TMC}}(x_A, Q^2) = \sum_{i=1}^{6} A_j^i \tilde{F}_i^{A,(0)}(\xi_A, Q^2) + B_j^i \tilde{h}_i^A(\xi_A, Q^2) + C_j \tilde{g}_2^A(\xi_A, Q^2)\,. \tag{2.38}$$

Three important remarks:

- These formulae are similar to those previously established for the nucleon case (Eq.(23) of [104]).
- The arguments on the left and right sides of the equation are different; $\tilde{F}_j^{A,\mathrm{TMC}}$ takes the Bjorken scaling variable x_A when each individual terms takes the Nachtmann variable ξ_A.
- No invocation of perturbative QCD; constructing the SFs in the OPE means the Wilson coefficient C_ι^{2k} are defined at any order of α_s. It implies that coefficients A_j^i, B_j^i and C_j, and the variable ξ_A in eq. (2.38) are the same, whatever the perturbative order of $\tilde{F}_i^{A,(0)}$.

More specifically, the target mass corrected structure functions $\tilde{F}_{1,2,3}^{A,\mathrm{TMC}}(x_A, Q^2)$ are given

by

$$\tilde{F}_1^{A,\mathrm{TMC}}(x_A) = \left(\frac{x_A}{\xi_A r_A}\right)\tilde{F}_1^{A,(0)}(\xi_A,Q^2) + \left(\frac{M_A^2 x_A^2}{Q^2 r_A^2}\right)\tilde{h}_2^A(\xi_A,Q^2) + \left(\frac{2M_A^4 x_A^3}{Q^4 r_A^3}\right)\tilde{g}_2^A(\xi_A,Q^2)\,, \tag{2.39a}$$

$$\tilde{F}_2^{A,\mathrm{TMC}}(x_A) = \left(\frac{x_A^2}{\xi_A^2 r_A^3}\right)\tilde{F}_2^{A,(0)}(\xi_A,Q^2) + \left(\frac{6M_A^2 x_A^3}{Q^2 r_A^4}\right)\tilde{h}_2^A(\xi_A,Q^2) + \left(\frac{12M_A^4 x_A^4}{Q^4 r_A^5}\right)\tilde{g}_2^A(\xi_A,Q^2)\,, \tag{2.39b}$$

$$\tilde{F}_3^{A,\mathrm{TMC}}(x_A) = \left(\frac{x_A}{\xi_A r_A^2}\right)\tilde{F}_3^{A,(0)}(\xi_A,Q^2) + \left(\frac{2M_A^2 x_A^2}{Q^2 r_A^3}\right)\tilde{h}_3^A(\xi_A,Q^2)\,, \tag{2.39c}$$

where

$$\tilde{h}_2^A(\xi_A,Q^2) = \int_{\xi_A}^1 du_A\ \frac{\tilde{F}_2^{A,(0)}(u_A,Q^2)}{u_A^2}, \tag{2.40a}$$

$$\tilde{h}_3^A(\xi_A,Q^2) = \int_{\xi_A}^1 du_A\ \frac{\tilde{F}_3^{A,(0)}(u_A,Q^2)}{u_A}, \tag{2.40b}$$

$$\tilde{g}_2^A(\xi_A,Q^2) = \int_{\xi_A}^1 du_A\ \tilde{h}_2^A(u_A,Q^2). \tag{2.40c}$$

The equation Eq. (2.38) does not assume or impose any Callan-Gross relation [18] between F_1 and F_2. This relation, typically expressed as $2xF_1 = F_2$, arises as leading order prediction of the massless ($m_q = 0$ and $M_N = 0$) quark-parton model. At this same order and without any mass contributions (either hadronic or partonic), the longitudinal structure function is given by $F_L = F_2 - 2xF_1$. Therefore, the Callan-Gross relation can be viewed as a statement about F_L. This relation is modified when mass effects (whether hadronic or partonic) are included, for example, through helicity inversion, or when contributions beyond leading order in QCD are considered. Furthermore, if there were elementary constituents within hadrons with spin zero, such as scalar quarks or scalar gluons, these would contribute to the longitudinal structure function F_L at leading order. Thus, the Callan-Gross relation also implies the absence of elementary scalar (spin zero) constituents in a hadron.

For completeness, the longitudinal structure function with TMCs is given by:

$$\begin{aligned}\tilde{F}_L^{A,\mathrm{TMC}}(x_A) &= r_A^2\tilde{F}_2^{A,\mathrm{TMC}}(x_A) - 2x_A\tilde{F}_1^{A,\mathrm{TMC}}(x_A)\\ &= \frac{x_A^2}{\xi_A^2 r_A}[\tilde{F}_2^{A,(0)}(\xi_A) - 2\xi_A\tilde{F}_1^{A,(0)}(\xi_A)] + \frac{4M_A^2 x_A^3}{Q^2 r_A^2}\,\tilde{h}_2^A(\xi_A) + \frac{8M_A^4 x_A^4}{Q^4 r_A^3}\,\tilde{g}_2^A(\xi_A)\\ &= \frac{x_A^2}{\xi_A^2 r_A}\tilde{F}_L^{A,(0)}(\xi_A) + \frac{4M_A^2 x_A^3}{Q^2 r_A^2}\,\tilde{h}_2^A(\xi_A) + \frac{8M_A^4 x_A^4}{Q^4 r_A^3}\,\tilde{g}_2^A(\xi_A)\ .\end{aligned} \tag{2.41}$$

This general result is non-zero and thus violates the Callan-Gross relation, as it should. The leading term $\tilde{F}_L^{A,(0)}$ will be non-zero for finite quark masses, and the sub-leading terms $\tilde{h}_i^A, \tilde{g}_2^A$ contribute for finite hadron mass M_A. Note that effects from heavy quark masses are separate from the M_A-modifications of the Nachtmann variable ξ_A, and are taken into account in the $\tilde{F}_i^{A,(0)}$ as outlined in Sec. 2.4 and Ref. [104]; we address this in Sec. 2.4.3.

2.4 Rescaling

In this section, we perform a rescaling between the nuclear (variables with A subscript) and the averaged nucleon (variables with N subscript) kinematics. The no re-scaled functions are still identified with a tilde (~) when those re-scaled are without.

2.4.1 Kinematics

From Eqs. (2.39) and (2.41), it can be observed, with some minor algebraic manipulation, that M_A^{2j} terms are always accompanied by x_A^{2j} factors in x-space. Furthermore, since structure functions $\tilde{F}_2^{A,(0)}(x_A)$ have their main support at $x_A = x_N/A \sim 1/A$,[10] we have effectively

$$\frac{M_A^{2j} x_A^{2j}}{Q^{2j}} = \frac{(M_N^{2j} A^{2j})\, x_A^{2j}}{Q^{2j}} = \frac{M_N^{2j} x_N^{2j}}{Q^{2j}} \sim \frac{M_N^{2j}}{Q^{2j}} \,. \tag{2.42}$$

We define the *average nucleon mass* $M_N = \frac{M_A}{A}$, such that the terms with $j > 0$ in Eq. (2.35) are suppressed for $Q^2 > M_N^2$ independent of the nuclear target. Subsequently, we define the *average nucleon momentum* $p_N = \frac{p_A}{A}$ inside a nucleus. Then the nucleon scaling variable becomes

$$x_N = \frac{Q^2}{2p_N \cdot q} = \frac{Q^2}{2\frac{p_A}{A} \cdot q} = A x_A, \qquad \text{where} \quad x_N \in [0, A] \,, \tag{2.43}$$

in contrast to the original Bjorken variable

$$x_A = \frac{Q^2}{2p_A \cdot q} = x_N/A \;, \qquad \text{where} \quad x_A \in [0, 1] \;. \tag{2.44}$$

Note that the original x_A variable can be constructed directly from the external momenta of the particles, whereas the *averaged quantities* x_N and p_N are not directly observable for nuclear targets with $A > 1$.

[10] Note that we have introduced the nucleon scaling variable $x_N = A x_A$; this will be discussed in detail throughout this section. Although x_N can in principle extend to $x_N = A$, the dominant range of the kinematics is $0 \leq x_N \lesssim 1$. Discussion on the $x_N > 1$ region are given in Sec. 2.5.5.

Using $M_A = AM_N$, the Nachtmann scaling variable becomes

$$\xi_A = \left(\frac{2}{1+\sqrt{1+4x_A^2M_A^2/Q^2}}\right)x_A = \left(\frac{2}{1+\sqrt{1+4x_N^2M_N^2/Q^2}}\right)\frac{x_N}{A} = \frac{1}{A}\xi_N. \tag{2.45}$$

On the other hand, the quantity R_M and r, are the same for both the nucleus and (averaged) nucleon case:

$$R_M = \frac{2}{1+r_A} = \frac{2}{1+r_N}\ . \tag{2.46}$$

and

$$r_A = \sqrt{1+4x_A^2M_A^2/Q^2} = \sqrt{1+4x_N^2M_N^2/Q^2} = r_N\,. \tag{2.47}$$

It is convenient to summarize all the kinematic terms and their transformation in the table 2.2.

2.4.2 Re-scaled Structure Functions

In order to get the rescaled structure fonctions, we first have to define the rescaled hadronic tensor (from Eq. (2.13)):

$$W_{\mu\nu}^A(p_N,q) \equiv \frac{1}{A}\tilde{W}_{\mu\nu}^A(p_A,q)\ , \tag{2.48}$$

and then the rescaled structure functions (we omit here the Q^2 dependence for brevity):

$$F_1^A(x_N) \coloneqq \frac{x_A}{x_N}\tilde{F}_1^A(x_A), \quad F_2^A(x_N) \coloneqq \tilde{F}_2^A(x_A), \quad F_3^A(x_N) \coloneqq \frac{x_A}{x_N}\tilde{F}_3^A(x_A).^{11} \tag{2.49}$$

With a particular attention to integration variables one finds

$$\tilde{h}_2^A(\xi_A) = \int_{\xi_A}^1 du_A \frac{\tilde{F}_2^{A(0)}(u_A)}{u_A^2} = A\int_{\xi_N}^A du_N \frac{F_2^{A(0)}(u_N)}{u_N^2} =: Ah_2^A(\xi_N)\,, \tag{2.50}$$

$$\tilde{h}_3^A(\xi_A) = \int_{\xi_A}^1 du_A \frac{\tilde{F}_3^{A(0)}(u_A)}{u_A} = A\int_{\xi_N}^A du_N \frac{F_3^{A(0)}(u_N)}{u_N} =: Ah_3^A(\xi_N)\,, \tag{2.51}$$

$$\tilde{g}_2^A(\xi_A) = \int_{\xi_A}^1 du_A \tilde{h}_2^A(u_A) = \frac{1}{A}\int_{\xi_N}^A du_N Ah_2^A(u_N) =: g_2^A(\xi_N)\,. \tag{2.52}$$

Using $x_N = Ax_A$, $\xi_N = A\xi_A$, $r_A = r_N$, $M_A = AM_N$, we finally obtain the master TMC formula for the **rescaled nuclear structure functions**:

[11] $F_{4-6}^A(x_N) \coloneqq \frac{x_A}{x_N}\tilde{F}_{4-6}^A(x_A)$.

Summary table of key relations

Nucleus A		Nucleon N
$M_A = A\,M_N$		$M_N = M_A/A$
$p_A = A\,p_N$		$p_N = p_A/A$
$x_A = \dfrac{Q^2}{2p_A\cdot q} \equiv x_N/A$ $x_A \in [0,1]$		$x_N = \dfrac{Q^2}{2p_N\cdot q} \equiv A\,x_A$ $x_N \in [0,A]$
$W_A^2 = (p_A+q)^2$		$W_N^2 = (p_N+q)^2$
$\nu_A = \dfrac{(q\cdot p_A)}{M_A} = \dfrac{Q^2}{2M_A x_A} \equiv \nu_N$		$\nu_N = \dfrac{(q\cdot p_N)}{M_N} = \dfrac{Q^2}{2M_N x_N} \equiv \nu_A$
$y_A = \dfrac{\nu_A}{E} \equiv y_N$		$y_N = \dfrac{\nu_N}{E} \equiv y_A$

Nachtmann variable & hadronic mass		
$r_A = \sqrt{1+\dfrac{4x_A^2M_A^2}{Q^2}} \equiv r_N$		$r_N = \sqrt{1+\dfrac{4x_N^2M_N^2}{Q^2}} \equiv r_A$
$\xi_A = R_M x_A \equiv \xi_N/A$ $\xi_A \in [0,1]$		$\xi_N = R_M x_N \equiv A\,\xi_A$ $\xi_N \in [0,A]$
Since $r_A = r_N$ We define $r \equiv r_A = r_N$, such that $R_M = \dfrac{2}{1+r}$		
Also, $\xi_A/x_A = \xi_N/x_N = R_M = \dfrac{2}{1+r}$		
We also define $\varepsilon = (xM/Q)$.		

Table 2.2: We summarize the key relations for a nucleus (A) and nucleon (N). We also find it convenient to define $\varepsilon = (xM/Q)$; we omit the subscripts on ε for brevity as $(x_A M_A/Q) = (x_N M_N/Q)$. We caution that W_A and W_N are **not** simply related, *c.f.*, Sec. 2.5.4. Note, the target mass modifies the scaling variable via $\xi_A = R_M\,x_A$. Additionally, we introduce the shorthand notation $r = r_A = r_N$.

$$F_1^{A,\mathrm{TMC}}(x_N,Q^2) = \left(\frac{x_N}{\xi_N r_N}\right) F_1^{A,(0)}(\xi_N,Q^2) + \left(\frac{M_N^2 x_N^2}{Q^2 r_N^2}\right) h_2^A(\xi_N,Q^2) + \left(\frac{2M_N^4 x_N^3}{Q^4 r_N^3}\right) g_2^A(\xi_N,Q^2)\,, \tag{2.53a}$$

$$F_2^{A,\mathrm{TMC}}(x_N,Q^2) = \left(\frac{x_N^2}{\xi_N^2 r_N^3}\right) F_2^{A,(0)}(\xi_N,Q^2) + \left(\frac{6M_N^2 x_N^3}{Q^2 r_N^4}\right) h_2^A(\xi_N,Q^2) + \left(\frac{12M_N^4 x_N^4}{Q^4 r_N^5}\right) g_2^A(\xi_N,Q^2)\,, \tag{2.53b}$$

$$F_3^{A,\mathrm{TMC}}(x_N,Q^2) = \left(\frac{x_N}{\xi_N r_N^2}\right) F_3^{A,(0)}(\xi_N,Q^2) + \left(\frac{2M_N^2 x_N^2}{Q^2 r_N^3}\right) h_3^A(\xi_N,Q^2)\,. \tag{2.53c}$$

The above rescaled equations for the **nucleon** structure functions F_i (without the "~" notation) take the same form as the master equations for the **nuclear** structure functions $\tilde{F}_i$ in Eq. (2.38) but are entirely expressed in terms of averaged nucleon variables [12]. This demonstration is one of the main results of this chapter.

2.4.3 Impact of quark masses

We emphasize that all of the above derivations do not explicitly rely on the parton model. This is because the quark and gluon substructure of nuclei enters only implicitly during the intermediate steps of deriving the master and rescaled formulae. The hadronic currents appearing in the hadronic tensor are composed of quark and gluon fields, and the derivation relies on the validity of the OPE for the products of these currents. As a result, due to the nature of factorization, all details concerning quark masses are fully contained within the structure functions $F_i^{A,(0)}$, and the forms of the rescaled formulae in Eq. (2.53) remain unchanged.

The treatment of the quark masses was reviewed in detail in Ref. [104], but it is instructive to briefly mention how the quark masses affect the rescaling of the kinematic variables as summarized in Table 2.3. Similar to how the hadron mass modifies the scaling variable through the relation $\xi = R_M x$ due to momentum conservation, the inclusion of quark masses further modify the scaling variable via the relation [101, 102, 120]

$$\bar{\xi} \;=\; R_{ij}\;\xi \;=\; R_{ij}\;R_M\;x\;. \tag{2.54}$$

Here, ξ contains modifications due to the hadron mass M_A, while $\bar{\xi}$ (with the overline)

[12] We refer to Ref. [48] for the corresponding result for the SFs $F_{4,5,6}$.

Partonic masses in scaling variables
$\bar{\xi}_A = R_{ij}\xi_A = R_{ij}R_M x_A$ Nucleus $\bar{\xi}_N = R_{ij}\xi_N = R_{ij}R_M x_N$ Nucleon
$\xi_{A,N}$: TMC corrections (no overbar) $\bar{\xi}_{A,N}$: TMC + parton mass corrections
For $q_i + V \to q_j$ $R_{ij} = \frac{1}{2Q^2}[Q^2 - m_i^2 + m_j^2 + \Delta(Q^2, m_i^2, m_j^2)]$ $\Delta(a,b,c) = \sqrt{a^2 + b^2 + c^2 - 2(ab + bc + ca)}$
For $g + V \to q_j + q_k$ $R_{jk} = 1 + \frac{(m_j + m_k)^2}{Q^2}$

Table 2.3: Summary of partonic mass relations in scaling variables. The target mass modifies the Nachtmann scaling variable via $\xi_A = R_M\, x_A$, where R_M is defined in Table 2.2. This is further modified by the parton masses $\bar{\xi}_A = R_{ij}\xi_A = R_{ij}R_M x_A$; hence these multiplicative corrections factorize between the hadron and parton masses [104]. In the limit where the initial-state parton mass $m_i \to 0$, we obtain the "slow rescaling" result: $R_{ij} = \left(1 + m_j^2/Q^2\right)$. Thus, ξ_A contains modifications due to the hadron mass M_A, while $\bar{\xi}_A$ (with the over-bar) contains modifications due to **both** the hadron mass M_A and the parton masses $\{m_i, m_j\}$.

contains modifications due to **both** the hadron mass M_A and the parton masses m_i, m_j. The R_M factor, which is given in Eq. (2.46), depends on the hadron (nucleus) mass, and the R_{ij} factor depends on the incoming and outgoing quark masses $\{m_i, m_j\}$, respectively (*c.f.*, Table 2.3). Note that the hadronic and partonic correction factors factorize and are subsequently applied multiplicatively. Intuitively, this factorization means that hadronic kinematics can be considered separately from partonic kinematics, up to momentum conservation. Additionally, in the limit where the mass of the initial-state quark vanishes $m_i = 0$, we have the familiar *slow rescaling* limit $R_{ij} \approx (1 + m_j^2/Q^2)$.[13]

[13]For further details of this section, see Refs. [101, 102, 104, 120].

2.5 Focus on Parton Model in the (massless) nuclear case

We now turn to a discussion of structure functions in the context of the parton model. Here, we summarize the main elements of the conventional *massless*-QCD-improved parton model, also known as Zero Mass Variable Flavor Number Scheme (ZM-VFNS), where quark-mass effects in the hard scattering cross sections and target-mass effects are neglected. There is an extensive body of literature on computations with heavy partons; Collins, for example, extended the factorization proof [86] to the case of heavy partons in Ref. [8], and implementations can be found in Refs. [102, 121–124]. A discussion of the parton model including such mass effects will be given in Sec. 2.6.

In the QCD-improved parton model the nuclear structure functions ($\tilde{\mathcal{F}}_k^A$) are given as convolutions of target-independent, short-distance, Wilson coefficients ($C_{k,i}$) with universal nuclear parton distribution functions ($\tilde{f}_i^A$):

$$\tilde{\mathcal{F}}_k^A(x_A, Q^2) = \int_{x_A}^1 \frac{dy_A}{y_A}\ \tilde{f}_i^A(y_A, Q^2)\ C_{k,i}(x_A/y_A)\ + \dots\,, \tag{2.55}$$

where the ellipses represent higher twist contributions $\tilde{\mathcal{F}}_k^{A,\tau\geq 4}(x_A, Q^2)$, and a sum over $i = q, g$ species is understood. As in the previous section, we use the tilde notation to identify the nuclear PDFs $\tilde{f}_i^A(x_A, Q^2)$, and the rescaled PDFs are without the tilde $f_i^A(x_N, Q^2)$. Here, we use the shorthand

$$\tilde{\mathcal{F}}_k^A(x_A) = [\tilde{F}_1(x_A), \tilde{F}_2(x_A)/x_A, \tilde{F}_3(x_A)] \qquad \text{for} \qquad k = 1, 2, 3\,. \tag{2.56}$$

The nuclear PDFs ($\tilde{f}_i^A$) are defined as Fourier transforms of matrix elements of local, twist-2 operators, which we denote[14] as $\mathcal{O}_i(z)$ and which are defined initially in coordinate space. These operators are composed of quark and gluon fields, and are acted upon by the hadronic, i.e., nuclear, state $A(p_A)$:

$$\tilde{f}_i^A(x_A, Q^2) \sim \langle A(p_A)|\mathcal{O}_i|A(p_A)\rangle\,, \tag{2.57}$$

where $x_A \in [0, 1]$ and the factorization and renormalization scales have been identified with the exchange boson virtuality $\mu_F = \mu_R = Q$.

In this work we focus on unpolarized DIS. However, the application to polarized DIS

[14]The notation for field operators, which typically carry indices throughout this work, should not be confused with the notation $\mathcal{O}(x)$, which we also use in this work (without indices) to denote an uncertainty or expansion to order x.

are relatively straightforward. For a spin-j target there are $2(2j+1)$ independent helicity-dependent quark distributions $\tilde{q}_+^m(x)$ and $\tilde{q}_-^m(x)$ with $-j \le m \le j$. Here $\tilde{q}_+^m(x)$ $[\tilde{q}_-^m(x)]$ describes the number density for finding a quark with momentum fraction x and spin m parallel [anti-parallel] to the the z-axis of a hadron moving with infinite momentum along the z-axis. Due to parity invariance of the strong interaction, we have a relation between the up (parallel) and down (anti-parallel) spin ($\tilde{q}_+^m = \tilde{q}_-^{-m}$) such that there remain only $2j+1$ independent distributions. However, this symmetry is broken once weak interaction effects are taken into account. For a spin-j target, the unpolarized quark densities are related to the helicity-dependent distributions as follows:

$$\tilde{q}_i^A(x_A, Q^2) := \frac{1}{2j+1} \sum_{m=-j}^{j} \left(\tilde{q}_+^m + \tilde{q}_-^m\right). \tag{2.58}$$

For a spin-1/2 target (and using $\tilde{q}_+^{1/2} = \tilde{q}_-^{-1/2}$), this reduces to the well-known expression:

$$\tilde{q}_i^A(x_A, Q^2) := \tilde{q}_+^{1/2} + \tilde{q}_-^{1/2}. \tag{2.59}$$

Similar definitions can be written down for the unpolarized gluon distribution.

2.5.1 Nuclear DGLAP evolution

The unpolarized nuclear PDFs (for any target spin) satisfy the usual Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) evolution equations [29–31]:

$$\frac{d\tilde{f}_i^A(x_A, Q^2)}{d\ln Q^2} = \frac{\alpha_s(Q^2)}{2\pi} \int_{x_A}^1 \frac{dy_A}{y_A} P_{ij}(y_A)\, \tilde{f}_j^A(x_A/y_A, Q^2) \tag{2.60}$$

$$= \frac{\alpha_s(Q^2)}{2\pi} \int_{x_A}^1 \frac{dy_A}{y_A} P_{ij}(x_A/y_A)\, \tilde{f}_j^A(y_A, Q^2). \tag{2.61}$$

Furthermore, the following sum rules due to charge, baryon number, and momentum conservation are satisfied:

$$\int_0^1 dx_A\ \tilde{u}_v^A(x_A, Q^2) = 2Z + N, \tag{2.62a}$$

$$\int_0^1 dx_A\ \tilde{d}_v^A(x_A, Q^2) = Z + 2N, \tag{2.62b}$$

$$\int_0^1 dx_A\ x_A \sum_i \tilde{f}_i^A(x_A, Q^2) = 1, \tag{2.62c}$$

where Z is the electric charge of the nucleus with baryon number $A = Z + N$.

2.5.2 Relation to the OPE

A Mellin transformation can turn a convolution integral $(A \otimes B)(x) = \int_x^1 dy/y A(y) B(x/y)$ into an ordinary product of Mellin moments:

$$\int_0^1 dx x^{n-1} (A \otimes B)(x) \equiv (A \otimes B)_n = A_n \cdot B_n \ , \tag{2.63}$$

where

$$A_n \equiv \int_0^1 dx x^{n-1} A(x) \qquad \text{and} \qquad B_n \equiv \int_0^1 dx x^{n-1} B(x) \ . \tag{2.64}$$

Applying this to structure functions of the QCD-improved parton model in Eq. (2.55), we obtain:

$$\int_0^1 dx_A x_A^{n-1} \tilde{\mathcal{F}}_k^A(x_A, Q^2) = \sum_n \tilde{f}_i^{A,n}(Q^2) \, C_{k,i}^n \, . \tag{2.65}$$

where on the RHS the Mellin transformation has yielded a simple sum of Mellin moments. (Note that we indicate the Mellin moment by the "n" superscript, as the tilde-notation has already been used for the pre-scaled PDFs, which we discuss in the following section, and N already denotes averaged nucleon quantities.)

The relation between the structure functions in the parton model and the OPE can now be easily seen by comparing the Mellin-moment of Eq. (2.65) with the j=0 term in Eq. (2.37):

$$\tilde{\mathcal{F}}_k^{(0),n} = C_{k,i}^n \, \tilde{A}_i^n = C_{k,i}^n \, \tilde{f}_i^{A,n}(Q^2) \ , \quad \text{for} \quad k = 1, 2, 3 \, , \tag{2.66}$$

where a sum over $i = q, g$ is understood. From this equation we see that the matrix elements $\tilde{A}^n$ of Eq. (2.37) are simply the Mellin moments of the parton distribution functions:

$$\tilde{A}_i^n(Q^2) = \tilde{f}_i^{A,n}(Q^2) \equiv \int_0^1 dx_A x_A^{n-1} \, \tilde{f}_i(x_A, Q^2) \ . \tag{2.67}$$

Similarly, the Wilson coefficients $C_{k,i}^n$ of Eq. (2.66) are directly identified with those of Eq. (2.37) in the OPE.

2.5.3 Rescaling

The parton distributions $\tilde{f}_i^A(x_A)$ are number densities defined on the interval $x_A \in [0, 1]$ such that $\tilde{f}_i^A(x_A) dx_A$ can be interpreted as the number of partons i carrying a fraction of the parent hadron momentum in the interval $[x_A, x_A + dx_A]$. We define rescaled parton densities $f_i^A(x_N, Q^2)$ in the variable $x_N \in [0, A]$ with $x_N = A x_A$ by imposing that the

number of partons remains unchanged in the corresponding momentum intervals:

$$f_i^A(x_N, Q^2) dx_N := \tilde{f}_i^A(x_A, Q^2) dx_A \ . \tag{2.68}$$

This equation may appear a bit odd as $x_N \in [0, A]$ on the LHS, while $x_A \in [0, 1]$ on the RHS; we will provide additional clarification in the following discussion. Additionally, it should be noted that the rescaled nuclear PDF $f_i^A(x_N, Q^2)$ are the ones usually used in the literature as $f_i^A(x_N, Q^2)$ can be meaningfully compared across different nuclei, in contrast to $\tilde{f}_i^A(x_A, Q^2)$.

It can be easily checked that the rescaling defined here at the level of PDFs is consistent with the rescaling introduced above at the hadronic level. Correspondingly, the rescaled objects satisfy analogous DGLAP evolution equations:

$$\frac{df_i^A(x_N, Q^2)}{d \ln Q^2} \quad = \quad \frac{\alpha_s(Q^2)}{2\pi} \int_{x_N}^{A} \frac{dy_N}{y_N} \ P_{ij}\left(\frac{x_N}{y_N}\right) \ f_j^A(y_N, Q^2) \, . \tag{2.69}$$

Furthermore, the sum rules take the following form after the rescaling:

$$\int_0^A dx_N \quad u_v^A(x_N, Q^2) = 2Z + N \, , \tag{2.70a}$$

$$\int_0^A dx_N \quad d_v^A(x_N, Q^2) = Z + 2N \, , \tag{2.70b}$$

$$\int_0^A dx_N \quad x_N \sum_i f_i^A(x_N, Q^2) = A \, , \tag{2.70c}$$

which match those in Eq. (2.62). However, note the above integrals extend to A since $x_N \in [0, A]$.

While the rescaled nuclear PDFs are formally defined on the interval $x_N \in [0, A]$, it is important to identify how the bulk of the PDFs are distributed. For a proton with three valence quarks, we generally expect the valence PDFs to be peaked in the region of $x_A = x_N \sim 1/3$. In a similar manner, for nuclei of mass A, which carries A nucleons, we generally expect the valence PDFs to be peaked in the region of $x_A \sim 1/(3A)$, or equivalently $x_N \sim 1/3$, which is independent of A and thus facilitates a meaningful comparison between different nuclei. Therefore, even though x_N can in principle span the range $[0, A]$, we expect the dominant support of the PDFs to be in the region $x_A \leq 1/A$, or $x_N \leq 1$. For x_N to exceed unity, one *parton* alone would essentially need to acquire more momentum than an average *nucleon*; this is highly unlikely. Thus, it is common to assume the probability for $x_N > 1$ to be small, and global analyses of nPDFs typically implement the condition that $f_i^A(x_N) = 0$ for $x_N > 1$. This induces only a very small

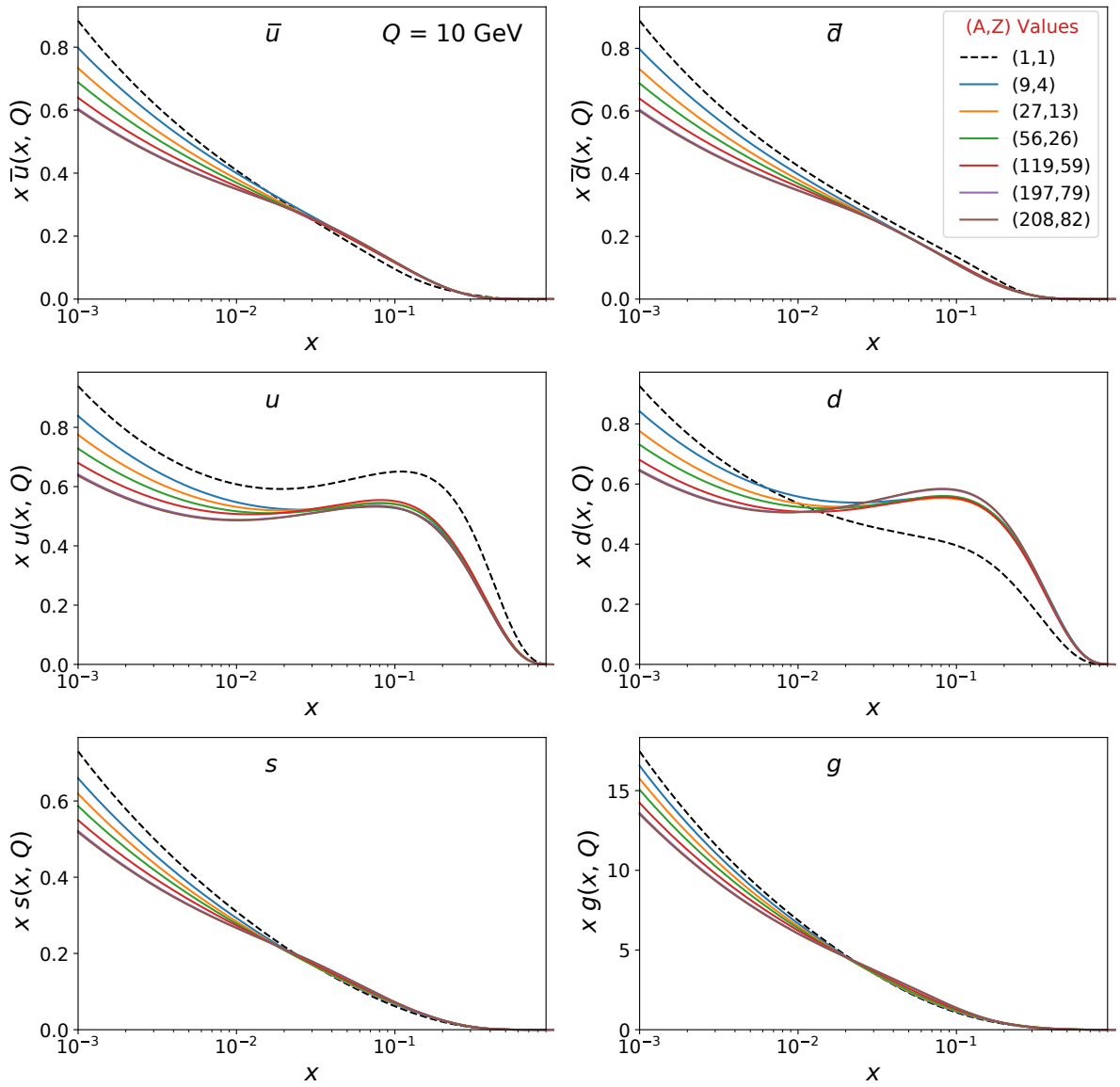


Figure 2.5: We display the nCTEQ15 nPDFs [35] as a function of x at $Q = 10$ GeV for selected nuclear (A, Z) values as indicated in the legend. These are the scaled nPDFs of Eq. (2.72), and this feature is most evident when comparing the $xu(x, Q)$ and $xd(x, Q)$ of the proton $(A = 1)$.

error [41], and we discuss the possibility of $x_N > 1$ further in Sec. 2.5.5.

Setting $f_i^A(x_N, Q_0^2) = 0$ for $x_N \geq 1$ at the initial scale Q_0 yields an important computational benefit since the DGLAP evolution framework consistently leads to $f_i^A(x_N, Q^2) = 0$ for $x_N \geq 1$ at all larger scales Q. Consequently, the advantage of setting $f_i^A(x_N \geq 1, Q^2) = 0$ is that the same evolution equations can be used for all nuclei, including nucleons:

$$\frac{df_i^A(x_N, Q^2)}{d\ln Q^2} = \begin{cases} \frac{\alpha_s(Q^2)}{2\pi} \int_{x_N}^1 \frac{dy_N}{y_N} \, P_{ij}\left(\frac{x_N}{y_N}\right) \, f_j^A(y_N, Q^2) & : 0 < x_N \leq 1 \,, \\ 0 & : 1 < x_N \leq A \,. \end{cases} \tag{2.71}$$

With the above framework for nuclear PDFs, we can perform DGLAP evolution on the interval $x \in [0,1]$ in a manner similar to proton evolution. However, the sum rules of Eq. (2.70a), (2.70b), and (2.70c) differ from the proton case. Thus, it is common (but not necessary) to further decompose the nuclear PDF f_i^A as

$$f_i^A(x,Q) = \frac{Z}{A} f_i^{p/A}(x,Q) + \frac{A-Z}{A} f_i^{n/A}(x,Q)\,, \tag{2.72}$$

where $f_i^{p/A}$ and $f_i^{n/A}$ represent *effective bound proton and neutron* PDFs for finding a parton i inside a nucleon. The *bound proton* PDF $f_i^{p/A}$ satisfies identical sum rules as the (free) proton PDF, so this quantity can be computed using standard DGLAP programs. The *bound neutron* PDF is commonly obtained from the *bound proton* one by assuming isospin symmetry.[15]

For comparison of nPDFs across different nuclei, it is typically the scaled nPDF $f_i^{(A,Z)}(x,Q)$ of Eq. (2.72) which can be meaningfully compared, as we illustrate in Fig. 2.5. This quantity is defined on $x \in [0,1]$ and normalized so that the integrated momentum is unity. Conversely, although the effective *bound proton and neutron* PDFs of Eq. (2.72) are convenient for computing the DGLAP evolution, these are <u>not</u> physically measurable objects. Caution is required when comparing them.

In particular, if we try to invert the relations of Eq. (2.72) for the up and down quarks, we find:

$$u^{p/A} = \frac{Z}{2Z-A}\,u^A - \frac{A-Z}{2Z-A}\,d^A \qquad \text{and} \qquad d^{p/A} = \frac{Z}{2Z-A}\,d^A - \frac{A-Z}{2Z-A}\,u^A\,. \tag{2.73}$$

This relation has the expected limits that for a proton: with $A = p$ and $A = Z = 1$ we find $\{u^{p/A}{=}u^{A=p};\ d^{p/A}{=}d^{A=p}\}$. Also, for a neutron with $A = n$ and $\{A,Z\} = \{1,0\}$ we find $\{d^{p/A}{=}u^{A=n};\ u^{p/A}{=}d^{A=n}\}$. However, for an isoscalar target ($A = 2Z$), Eq. (2.73) is singular because in this limit it is impossible to differentiate the u^A and d^A distributions and the decomposition into $u^{p/A}$ and $d^{p/A}$ becomes arbitrary. Since the majority of the nuclear data is taken on nuclear targets which lie along the nuclear stability line in the $\{A,Z\}$ plane where $A \sim 2Z$ (except for very heavy nuclei), our ability to separately determine the nuclear u^A and d^A distributions is limited.

[15] While isospin symmetry is typically used to relate $f_i^{n/A}$ to $f_i^{p/A}$, in principle one could add isospin violating contributions. The recent Marathon data [125] suggests there may be isospin violations in the very large x region [126].

2.5.4 Kinematic W cut

Determinations of PDFs are performed by global analyses using a wide array of data sets that can be described by the parton model within the factorization framework. It is therefore important to exclude data in the kinematic region that are not reliably predicted within this model.

In lepton-nucleon scattering, there are different types of scattering scenarios depending on the value of the hadronic invariant mass, $W^2 = (p+q)^2$, and the virtuality of the exchanged boson Q [127]. We can divide these cases into four categories:

1. elastic scattering ($W = M_N$),
2. shallowly inelastic scattering (SIS) ($W \leq 2$ GeV),
3. soft DIS ($W > 2$ GeV, $Q < 1$ GeV), and
4. DIS ($W > 2$ GeV and $Q \geq 1$ GeV).

Among these scenarios, DIS and soft DIS (dominated by non-resonant pion production) are the most reliably described by the parton model. Therefore, a W cut is traditionally imposed on DIS data to extract PDFs.

While imposing a W cut is straightforward for the case of lepton-proton scattering, it is more subtle for lepton-nucleus scattering when we compute in terms of the rescaled variables. For $\ell_1 + A \to \ell_2 + X$ scattering, we have:

$$W_A^2 = (p_A + q)^2 = M_A^2 + Q^2 \, \frac{1 - x_A}{x_A} = A^2 M_N^2 + Q^2 \, \frac{1 - x_N/A}{x_N/A} \,. \tag{2.74}$$

The average W_A per nucleon, $W_{\text{aver}} = W_A/A$, is then given by

$$W_{\text{aver}}^2 \;=\; \left(\frac{W_A}{A}\right)^2 = M_N^2 + Q^2 \, \frac{A - x_N}{A^2 x_N} \,. \tag{2.75}$$

Additionally, we observe that W_{aver}^2 and W_N^2 are not the same:

$$W_N^2 \;=\; (p_N + q)^2 = M_N^2 + Q^2 \, \frac{1 - x_N}{x_N} \,. \tag{2.76}$$

For $x_N \in [0, A]$, we find W_{aver}^2 is always positive while W_N^2 can be negative for $x_N > 1$. On the other hand, W_N is independent of A and can be used to compare nuclei as long as x_N is restricted to values below unity. If we impose a lower bound on W_A, this translates to

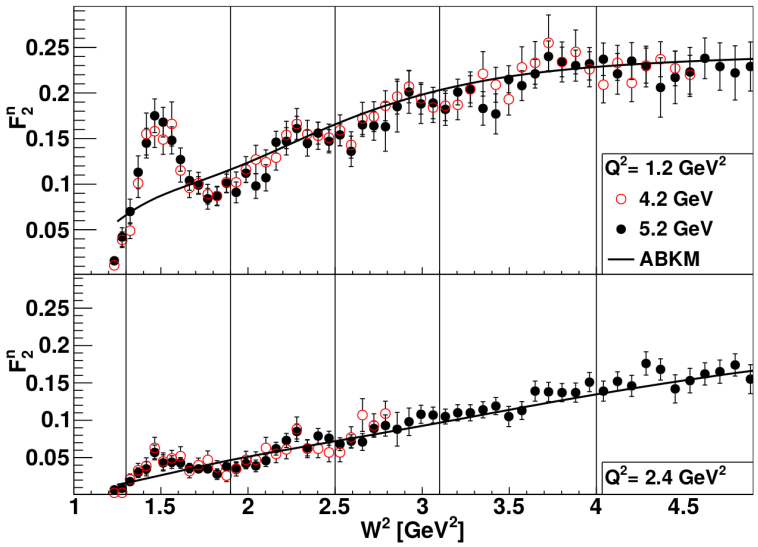


Figure 2.6: We display the neutron structure function F_2^n as measured by the BONuS data [128] at selected Q^2 values. The open (filled) circles represent data for a beam energy of $E = 4.223$ (5.262) GeV. This is compared with ABKM PDF predictions [124] including higher twist effects and target mass corrections. We observe the resonance structure at low W^2 values. Figure taken from Ref. [128].

the well-known upper bound on x_A of

$$0 \le x_A \le \left[1 + \frac{W_A^2 - M_A^2}{Q^2}\right]^{-1} \le 1\,. \tag{2.77}$$

In terms of the rescaled variable $x_N = Ax_A \in [0, A]$ this translates to

$$0 \le x_N \le A\left[1 + \frac{W_A^2 - M_A^2}{Q^2}\right]^{-1} \le A\,. \tag{2.78}$$

Consequently, when we impose a W kinematic cut for the nuclear case ($A > 1$), we need to be cognizant of these issues. One could try to limit the resonance region with a mass cut ($m_{\rm cut}$) by setting $W_N^2 = M_N^2 + m_{\rm cut}^2/A^2$. For $A > 1$, the resulting high-x cut turns out to be more relaxed than the usual W cut. Thus, using the traditional W cut here is much safer and hence, despite its pathological nature at $x > 1$, this cut can still be used. While cutting on W_A or (W_A/A) is more natural, the value of W_N, which is A dependent, needs further investigations.

Structure function data for nucleons and nuclei

In Fig. 2.6 we display F_2^n as measured by the BONuS experiment and compared with the ABKM global fit [124]. BONuS studied electron-deuteron scattering with two electron

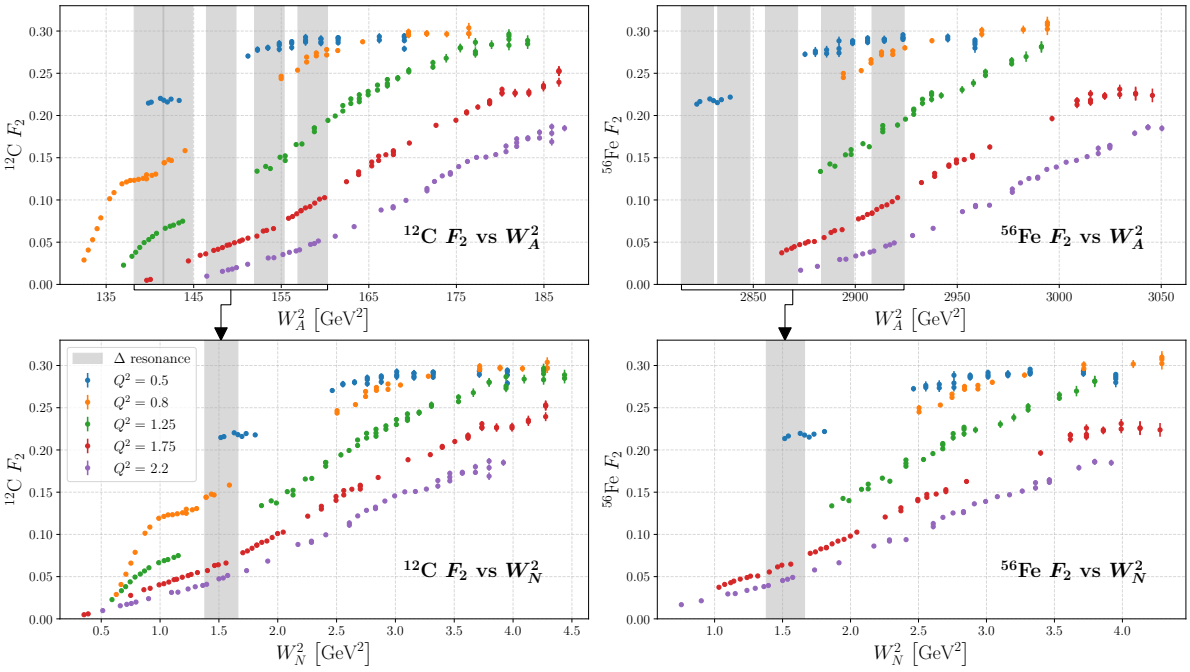


Figure 2.7: We plot F_2^A for carbon and iron for selected Q^2 values (in GeV^2) as a function of both W_A^2 (upper panels) and W_N^2 (lower panels). The first (Δ) resonance region is highlighted with the vertical shaded bands. For W_N^2, this region is $\sim M_\Delta^2$. For W_A^2, there is a more pronounced Q^2 dependence so that the resonance region for the lowest Q^2 is at the lowest W_A^2, and the one for the highest Q^2 is at the highest W_A^2 values. The data are taken from Ref. [129].

energies (E = 4.2 and 5.2 GeV) and used a novel spectator tagging technique to extract the free neutron cross section. The resonance structure is evident in the upper panel with Q^2 = 1.2 GeV^2 where we observe the first (Δ) resonance in the region of $W^2 \sim 1.5$ GeV^2. As we go to larger Q^2 values, the effect of the resonances are reduced as seen in the lower panel (Q^2 = 2.4 GeV^2).

Changing our focus from a nucleon to a nucleus, in Fig. 2.7 we display the F_2^A structure functions for carbon and iron at selected Q^2 values as measured in Ref. [129]. The region of the first resonance is highlighted by the shaded vertical bands. We immediately note that the characteristic resonance structure is more difficult to observe for the nuclear case as compared to the nucleon (neutron) case above. Furthermore, this is complicated by the (comparably) limited data in the low W^2 region.

We start by considering the bottom panels of Fig. 2.7, which plot F_2^A as a function of W_N^2. Here, we have highlighted the resonance region $m_\Delta^2 \sim W_N^2$. Due to the limited data in the low W^2 region, it is difficult to identify the resonances; however, at larger Q^2 values it is clear that any resonance structure again appears to be reduced. We now draw attention to the upper panels of Fig. 2.7, which are plotted versus W_A^2. The W_A^2-axes of the neutron, carbon and iron all differ as they are now proportional to A. Additionally,

there is a relative shift of the curves due to the different Q^2 dependence as compared with W_N^2. Specifically, we have converted using: $W_A^2 = AW_N^2 + A(A-1)M_N^2 + (A-1)Q^2$. This conversion also shifts the resonance region in a manner dependent on Q^2; in the upper panels, the resonance region for the lowest Q^2 is at the lowest W_A^2, and the one for the highest Q^2 is at the the highest W_A^2 values.
We now summarize our main observations in the following.

- The resonance structure is most evident at lower Q^2 values. As Q^2 increases, the resonance peaks are reduced as the phase space for the DIS continuum is growing and becomes dominant.
- As we move to heavier nuclei, the resonance structure is also comparably reduced.
- The W_N^2 variable has the advantage that this identifies a uniform kinematic region largely independent of Q^2 or A. This simplifies the task of placing cuts on the resonance region.
- The W_A^2 variable has the disadvantage that the kinematic region is dependent on A and Q^2, and this complicates the task of placing cuts on the resonance region.

2.5.5 nPDFs for $x_N \geq 1$

The rescaled PDFs allow us to compare distributions from different nuclei in a meaningful manner as illustrated in Fig. 2.5. Operationally, all current global nuclear PDF analyses work on the interval $x_N \in [0,1]$ even though the technically allowed range is $x_N \in [0,A]$. While details of the $x_N > 1$ region are beyond the scope of this study, we briefly mention experimental measurements that provide insights into this extreme kinematic region.

The Fermi region

The effects of nuclear binding can modify the partonic momentum, and thus dramatically impact the resulting structure functions [130]. As already mentioned in the Chapter 1, Fig. 1.8 shows the characteristic form of the nuclear correction factor F_2^{Fe}/F_2^D with[16] shadowing at small x_N ($x_N < 0.1$), anti-shadowing just beyond the shadowing region

[16]The nomenclature is historical: The term "shadowing" refers to the front of a large-A nucleus casting a metaphoric shadow over its back since a leptonic probe is likelier to scatter off the front (and not "see" the back) when partons have larger wavelengths (smaller x_N). This leads to a suppression of the F_2^A/F_2^D ratio at $x_N \lesssim 0.1$; for further details of this geometric picture, see Ref. [105, 130]. The term "anti-shadowing" refers to the experimental observation that the ratio is enhanced at $x_N \sim 0.1$. The term "EMC effect" refers to the suppression of the F_2^A/F_2^D ratio at moderate x_N discovered by the EMC collaboration [131]. The term "Fermi region" refers to the high-x_N region when Fermi motion becomes relevant enabling $x_N \geq 1.0$ for bound nucleons.

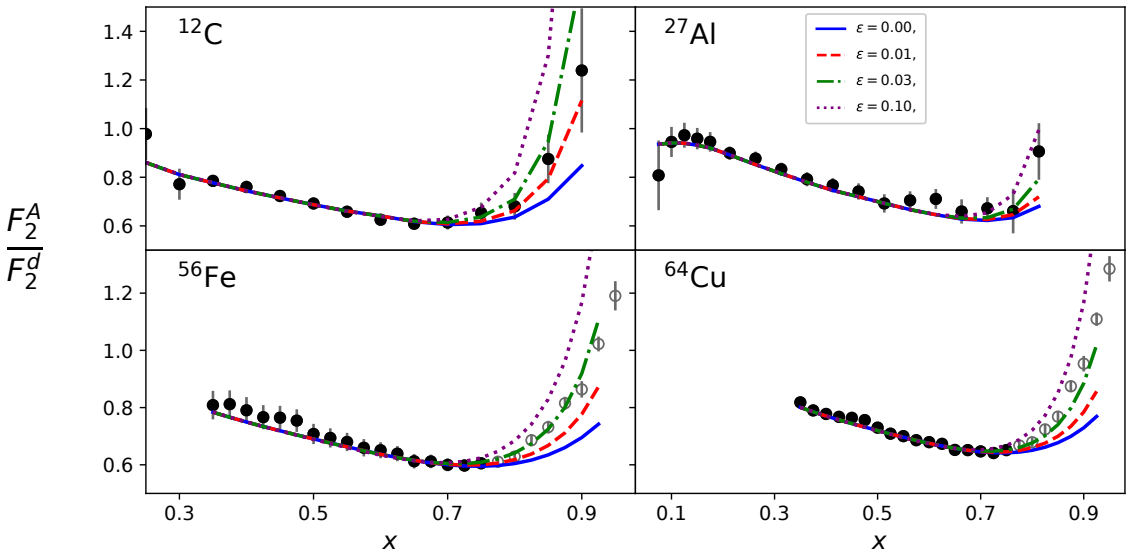


Figure 2.8: We display the ratio F_2^A/F_2^d for selected nuclei data sets in the large x region. The different curves show the impact of the ε parameter for the rescaling $x'_N \to [x_N - \varepsilon x_N^\kappa \log_{10} A]$ with $\kappa = 10$, *c.f.*, Ref. [41] for details. The data with solid circles satisfy the cuts $Q > 1.3$ GeV and $W > 1.7$ GeV; the open circles do not satisfy the cuts and are not included in the fit. This figure is taken from Ref. [41].

($x_N \sim 0.1$), the EMC region for intermediate x, and the Fermi region for large x. The nuclear binding effects are especially evident at larger momentum fractions (large x) in the Fermi region where the PDFs are steeply falling. There are a variety of theoretical approaches to describe this region, as well as experimental measurements of these nuclear binding effects. We provide only a brief overview below and refer to the reference for additional details.

Challenges of the Fermi region

The limitations of constraining the PDFs in the $x_N \sim 1$ region are evident when examining the nuclear ratio F_2^A/F_2^p in the large x_N region. As x increases, the $\mathcal{R} = F_2^A/F_2^p$ ratio transitions from $\mathcal{R} \lesssim 1$ in the EMC region, to $\mathcal{R} \gtrsim 1$ as $x_N \to 1$ in the Fermi region. We find that $F_2^p(x)$ with $A = 1$ for the proton must vanish at $x_N = 1$, whereas $F_2^A(x)$ can be finite since $x_N \in [0, A]$. This means that the denominator of $\mathcal{R}$ is vanishing while the numerator is finite. Hence, $\mathcal{R}$ will rapidly increase at $x \sim 1$, and is consistent with experimental measurements. More generally, we expect $F_2^{A_1} > F_2^{A_2}$ for $A_1 > A_2$, and this is the case for $F_2^{^{12}C}/F_2^D$ as shown in Fig. 2.8.

Theoretical implementations

On the theoretical side, there are a number of approaches that have been proposed to model nPDFs in the $x_N > 1$ region, and we point to a selection of references [56, 74, 132–140]. There is also a LO solution of DGLAP equation for quarks taking into account the full support (including $x_N > 1$ region) which is detailed in Ref. [141].

One technique to describe the $x_N > 1$ region is to use a convolution of the nucleon PDFs $f(x, Q)$ defined on the interval $x \in [0, 1]$ with a smearing kernel S_A which will shift the PDFs out to larger x values:

$$f_A(x, Q) = \int_x^A \frac{dy}{y} S_A(y, Q)\, f(x/y, Q)\,. \tag{2.79}$$

Here, the smearing kernel S_A is typically a Gaussian with an A-dependent width; larger nuclei have a larger smearing width, and hence a larger proportion of the f_A distribution will populate the $x > 1$ region. The behavior of the PDF convolution can also be approximated using an A-dependent rescaling: $x'_N \to [x_N - \varepsilon x_N^\kappa \log_{10} A]$, *c.f.*, Ref. [41]. This mimics the PDF convolution method described above, but the rescaling avoids the convolution integral, so it is computationally fast to evaluate.

In Fig. 2.8 we display F_2^A/F_2^d for selected nuclei and data sets, and indicate various levels of rescaling (controlled by the ε parameter). The data points are described in Ref. [41], and the solid curve (blue) with ε=0 is the default theory with no x-rescaling. The solid points are within the kinematic cuts of the global fit ($Q > 1.3$ GeV, $W_N > 1.7$ GeV), and the hollow points are outside these cuts. As we increase the ε parameter, we see this increases the ratio in the large x region, and can improve the comparison of the data and theory.

Judging from this sample of data, it appears that a detailed modeling of the Fermi region can improve the description of the data in this extreme kinematic region (low W_N, low Q, large x_N). These corrections could be a combination of TMCs, $x_N > 1$ effects, higher twist, or other non-perturbative corrections. Thus, a complete analysis must include and balance all these contributions.

Experimental investigations

There are a number of studies that have explored the structure functions in the Fermi region at very large x_N, and we mention a few examples below.

For example: Refs. [142, 143] examined SLAC DIS data from a variety of nuclear targets and found that in the larger x_N region ($x_N \gtrsim 0.5$) the Fermi motion effects are

similar to those in the deuteron, and increase for heavier nuclei. Measurements in this region are particularly challenging, and are subject to a variety of uncertainties as outlined in the Appendix of Ref. [142].

For the CC neutrino DIS process, CCFR [144] measured ~2000 events in the region $x_N > 0.75$, and used this to study F_2 at very large x_N. They fit the high x_N behavior to a decaying exponential $\propto \exp(-sx_N)$ normalized at $x_N = 0.65$, and find a slope of $s = 8.3 \pm 0.7$. They note this compares favorably with E133 at SLAC [145], which observed $s \sim [7, 8]$.

Hall C at Jefferson Lab measured the NC structure function F_2 on a variety of nuclei ranging from ^{2}H and ^{3}He up to ^{197}Au in the large x_N region [146, 147] out to $x_N \sim 1.4$. This work extends an earlier study of Ref. [148], and related measurements performed at SLAC [149]. Taking account of the kinematic pre-factors and the $\{h_2, g_2\}$ contributions, they extract $F_2^{(0)}(\xi_N, Q^2)$, where ξ_N is the Nachtmann variable, and fit this to a decaying exponential of the form $\exp(-s\xi_N)$ in the region $\xi_N \geq 0.75$. The large ξ_N fits finds the slope is in the general range $s \sim 15 \pm 2$. The slope has a mild increase with increasing Q^2, and a mild decrease with increasing nuclear A. These results also compare favorably with the BCDMS muon scattering data [150] which finds $s = 16.5 \pm 0.6$ for $x_N \sim [0.75, 1.05]$.

Parton momentum in the Fermi region

Based on the above large x_N measurements, we can ask the question: How much parton momentum is in the region $x_N > 1$?

The experimental measurements discussed above parameterize F_2 at large x_N as $\exp[-sx_N]$, where s can range from about ~8.3 to ~14. With s=8.3, the fraction of F_2 in the range $x_N \in [1, \infty]$ is less than ~6% than that in the range $x_N \in [0.65, 1]$. Given that the average momentum fraction in the interval $x_N \in [0.65, 1]$ is $\lesssim 2\%$ of the total nucleon momentum, this gives a rough estimate of the potential momentum fraction beyond $x_N > 1$ of $6\% \times 2\% \lesssim 0.1\%$ of the total momentum fraction. For $s \sim 14$, the decrease at large x_N is even steeper. Here, the F_2 ratio of $x_N \in [0.65, 1]$ compared to $x_N \in [1, \infty]$ drops to $\lesssim 0.1\%$, and the potential momentum fraction beyond $x_N > 1$ estimate becomes $1\% \times 2\% \lesssim 0.02\%$ of the total. In both of the above cases, the integrated parton momentum is very small in the $x_N > 1$ region, but the impact on the $F_2^{A_1}/F_2^{A_2}$ ratio still yields the characteristic increase in the $x_N \sim 1$ Fermi region.

2.5.6 Threshold problem and higher twist contributions

On the basis of the correspondence between the limits $x_N \to 1$ and $W_N^2 \to M_A^2$, the very high-x_N behavior of DIS structure functions at fixed Q^2 is sensitive to the presence of power-suppressed corrections beyond leading twist (twist-4 for unpolarized processes). It has been conjectured that the problematic threshold behavior in target-mass corrected structure functions in the free nucleon case are at least partly due to the absence of higher-twist contributions. As an approximate test of the inclusion of twist-4 contributions, Ref. [151] proposed a $1/Q^2$ expansion of the OPE mass-corrected structure functions, which indeed suppresses the problematic threshold behavior. These calculations were done for F_2, and later, for other tensor structure functions in Ref. [152]. While this was done for the free nucleon, a similar exercise might be carried out for nuclear structure functions as another TMC prescription.

In particular, Kulagin and Petti [151] showed that by expanding the target mass corrected structure functions to leading order in $1/Q^2$, the resulting functions have the correct $x_N \to 1$ limits. While avoiding the threshold problem, this prescription, however, raises the question of whether the $1/Q^2$ approximation is sufficiently accurate for structure functions near $x_N \approx 1$ at moderate Q^2.

To test the convergence of the $1/Q^2$ expansion at large x, Ref. [152] further expands the OPE result(s) to include $\mathcal{O}(1/Q^4)$ corrections. In fact, one can demonstrate that for a structure function that behaves at large x as $(1-x)^n$, the target mass corrected result will vanish in the $x_N \to 1$ limit up to order $1/Q^{2n-2}$ in the expansion. For $n \approx 3$, as is typical for nucleon structure functions, the threshold problem will therefore appear only at order $1/Q^6$.

2.6 Relations with parton model

We now turn to a discussion of the parton model, including the effects of quark and hadron mass. Such mass effects can be rigorously included in the QCD-improved parton model, as shown in a general factorization proof with quark masses by Collins [9]. This work extended the factorization proof to the case of massive partons and demonstrated that all contributions proportional to the parton mass can be fully factorized. The underlying scheme is also known in the literature as the Aivazis-Collins-Olness-Tung (ACOT) scheme [101, 102, 153], which is an example of what is called a General Mass Variable Flavor Number Scheme (GM-VFNS) in the modern literature [102,121–124,154]. In the following subsections we summarize DIS structure functions in the ACOT scheme, including quark

and hadron (nuclear) mass effects, and compare and contrast these expressions to the structure functions in the OPE. As discussed in the introduction (Sec. 2.1), we consider the theoretical status of both approaches, the OPE and the collinear-factorization-based QCD-improved parton model, on equal footing.

In the naïve, massless parton model, we have a hadron A of momentum p_A that emits a collinear parton of momentum $k = x_A p_A$. If we try and extend this simple picture to the case of massive partons, the presence of the parton's mass m_i violates collinear kinematics and we encounter ambiguities proportional to the parton mass. We work in the ACOT formalism [101, 102, 153] to compute the structure functions in the helicity basis. This has the advantage that the polarization vectors are boost invariant between the hadron and parton frame, and provides advantages for incorporating both the hadron and parton masses, as we illustrate below. We observed in Sec. 2.4 that hadron masses modify the scaling variable via the relation $\xi_A = R_M x_A$, where the ξ_A is the Nachtmann variable. Within the context of the OPE and the parton model, the effect of the quark masses factorizes from the hadronic modifications. The rescaling variable is modified by an additional factor R_{ij}, which is a function of the quark masses $\{m_i, m_j\}$ and is defined in Table 2.3. Thus, we have $\bar{\xi}_A = R_{ij} R_M x_A$. Recall the Nachtmann variable ξ_A includes the TMCs due to M_A, and the variable $\bar{\xi}_A$ (with the overline) includes both the TMCs and the parton mass corrections.

In the following, we also review the structure functions in the light-front formalism, and compare with the TMC expressions obtained from the OPE. There has been extensive discussion in the literature about computations with massive partons [102, 121–124]. The present discussion is focused on the detailed organization of the massive parton contributions in the formalism outlined by Collins [9]. Thus, we include all terms of order $\mathcal{O}(m_i/Q)^2$, and the non-factorizable terms are suppressed by powers of $(\Lambda_{\rm NP}/Q)^2$.

2.6.1 Helicity formulation of the parton model

When we generalize the massless QCD parton model to the massive case, there are potential ambiguities proportional to the parton mass m_i, which may enter if one is not careful about organizing the perturbative expansion. This organization is more easily dealt with in the helicity basis[17] as we outline in this section; however, helicity and more common tensor formulation lead to identical results.

The original ACOT formalism for massive DIS was derived in the helicity basis, and

[17]For clarity, this is the same technology used to compute helicity amplitudes and can be found in popular textbooks [155, 156].

this provides some advantages in working with both hadron and parton masses. For example: in the helicity formalism, we can define polarization vectors (Sec. 2.6.2) in a boost-invariant manner so that they are the same for both the hadron and parton reference frames (Sec. 2.6.3). This has the advantage that there are no mixing of terms between the partonic and hadronic structure functions[18] in the helicity basis that could otherwise originate from making a boost (helicity eigenstates, e.g., polarization vectors and spinors, are not boost-invariant quantities). Additionally, by working with light-front momenta, $\{p^{\pm}, k^{\pm}\}$, we can take advantage of the fact that $p^{+} \gg p^{-}$ and $k^{+} \gg k^{-}$, and define our scaling variable ξ as $k^{+} = \xi p^{+}$, which is valid for both massless and massive partons. In this Section, we will initially use ξ (un-subscripted) as a generic scaling variable, and we will connect this to ξ_A and $\bar{\xi}_A$ in Sec. 2.6.4.

For the present discussion we focus on the target mass corrections that enter the hadronic structure functions. We show that the helicity formalism matches the leading terms of the OPE master formula. We compare these result both analytically in Sec. 2.6.4 and numerically in Sec. 2.6.5 (and Sec. 2.7).

2.6.2 Helicity decomposition

In analogy to what was done in Sec. 2.2, one can also decompose the hadronic tensor $\tilde{W}^{A}_{\mu\nu}$ to obtain the six structure functions $\tilde{F}^{A}_{mn}$ in a basis of helicity polarizations by projecting out the Lorentz tensors [10, 101, 102]. Conventionally, this basis is defined using the polarization vectors $\epsilon_{\mu}(q, \lambda)$ of the intermediate exchange boson V, with momentum q and helicity λ. Explicitly, the decomposition for nucleus A is given by

$$\tilde{W}^{A}_{\mu\nu}(p_A, q) \;=\; \sum_{m,n} \varepsilon^{*}_{\mu}(q, m)\; \tilde{F}^{A}_{mn}(p_A, q)\; \varepsilon_{\nu}(q, n)\;, \tag{2.80}$$

where polarization vectors according to the ACOT convention are

$$\begin{aligned} \varepsilon_{\mu}(q, m = \pm) &= \frac{1}{\sqrt{2}}\,(0, \mp 1, -i, 0)\;, \quad \varepsilon_{\mu}(q, m = 0) = \frac{(-q^2)p_{\mu} + (p\cdot q)q_{\mu}}{\sqrt{(-q^2)[(p\cdot q)^2 - q^2 p^2]}}\;, \\ \varepsilon_{\mu}(q, m = s) &= \frac{q_{\mu}}{\sqrt{Q^2}}\;. \end{aligned} \tag{2.81}$$

(Note the $\mp$ Condon-Shortley convention for the transverse polarizations.) Here, the transverse polarization vectors are defined in a collinear frame where $q_z > 0$. The longitudinal polarization vector is defined with respect to a reference vector p (usually taken to be

[18] The partonic structure functions describe the theoretically calculable interaction of the partons with a vector boson.

Tensor $\Leftrightarrow$ Helicity	$\tilde{F}_1^A$	$\tilde{F}_2^A$	$\tilde{F}_3^A$	$\tilde{F}_4^A$	$\tilde{F}_5^A$	$\tilde{F}_6^A$
$\tilde{F}_+^A \equiv \tilde{F}_{++}^A$	1		$-r$			
$\tilde{F}_-^A \equiv \tilde{F}_{--}^A$	1		$+r$			
$\tilde{F}_0^A \equiv \tilde{F}_{00}^A$	-1	$\frac{r^2}{2x_A}$				
$\tilde{F}_S^A \equiv \tilde{F}_{SS}^A$	1	$\frac{1}{2x_A}$		$2\,\frac{Q^2}{M_A^2}$	$\frac{-1}{x_A}\,\frac{Q^2}{M_A^2}$	
$\tilde{F}_{0S}^A + \tilde{F}_{0S}^A$		$\frac{r}{x_A}$			$\frac{-r}{x_A}\,\frac{Q^2}{M_A^2}$	
$\tilde{F}_{0S}^A - \tilde{F}_{0S}^A$						$\frac{-r}{x_A}\,\frac{Q^2}{M_A^2}$

Table 2.4: The transformation between the tensor and helicity basis. The dotted lines highlight the scalar-based modes which are suppressed by the lepton mass. This table demonstrates there is a one-to-one relation between the helicity and tensor structure functions. This table matches the ACOT convention of Table III in Ref. [101] with appropriate conversions: (i) we replace $r \to \rho$; (ii) Ref. [101] includes an additional factor of $\frac{1}{2}$ in the definitions of $\{W_3, W_5, W_6\}$, so this table includes a relative factor of 2 in the columns for $\{\tilde{F}_3^A, \tilde{F}_5^A, \tilde{F}_6^A\}$. Additionally, note there is a relative factor of $2x$ between the helicity F_0 and longitudinal F_L structure function definitions $F_L = 2xF_0$ as noted in Eq. (2.92).

p_A). The scalar polarization vector $\varepsilon_\mu(q, m = s)$, which is sometimes called the "auxiliary polarization vector" [155,157], ensures the completeness relationship for off-shell and t-channel exchanges of the photon:

$$-g_{\mu\nu} \;=\; \varepsilon_\mu(q,+)\varepsilon_\nu^*(q,+) \;+\; \varepsilon_\mu(q,-)\varepsilon_\nu^*(q,-) \;-\; \varepsilon_\mu(q,0)\varepsilon_\nu^*(q,0) \;+\; \varepsilon_\mu^\gamma(q,s)\varepsilon_\nu^{\gamma *}(q,s)\;. \tag{2.82}$$

The indices m, n in Eq. (2.80) span the helicities plus (+), minus (−), longitudinal (0), and scalar (s). Angular momentum conservation reduces these to six combination $\{++, --, 00, ss, 0s, s0\}$, just as we have six hadronic structure functions in the tensor basis F_i with $\{i = 1...6\}$. Additionally, the three combinations $\{F_{ss}, F_{s0}, F_{0s}\}$ are suppressed by the lepton mass in a manner analogous to $\{F_4, F_5, F_6\}$, so typically these are ignored. We indicate the remaining combinations, $\{F_{++}, F_{00}, F_{--}\}$, by the short-hand notation F_λ with $\lambda = \{+, 0, -\}$.

Using the orthogonality and normalization conditions of the polarization vectors, the decomposition of Eq. (2.80) can be inverted, leading to the expression

$$\tilde{F}_{mn}^A(p_A, q) \;=\; \varepsilon^\mu(q, m)\; \tilde{W}_{\mu\nu}^A(p_A, q)\; \varepsilon^{\nu *}(q, n)\;, \tag{2.83}$$

which can be used to construct all six structure functions in the helicity basis. Explicit calculation gives

$$\tilde{F}^A_{mn}(x_A) = \sum_{i=1}^{6} R_i^{A(m,n)}(x_A, Q^2) \times \tilde{F}_i^A(x_A) \ , \tag{2.84}$$

where the coefficient functions $R_i^{A(m,n)}(x_A, Q^2)$ are given in Table 2.4. In the DIS limit, the $R_i^{A(m,n)}$ coefficients simplify since $\sqrt{1+Q^2/(q^0)^2} = \sqrt{1+(2x_A M_A)/q^0} \to 1$ in the high energy limit. Alternative helicity bases can lead to relative sign differences in Table 2.4, particularly for the scalar polarizations. For example: in the HELAS basis [158], Eq. (2.83) is augmented by factors of $1/(\varepsilon\cdot\varepsilon^*)$ to account for the normalization differences. Moreover, defining Eq. (2.83) using $\tilde{W}_i$, as done in the original ACOT works [101, 102] can also lead to normalization differences.

We obtain the hadron-level cross section by contracting the hadronic and leptonic tensors:

$$\begin{aligned} d\sigma &\sim L^{\mu\nu}(q,k)\ W_{\mu\nu}(p,q) \\ &\equiv L^{m'n'}(q,k)\ d^m_{m'}(\psi)^{-1}\ d^n_{n'}(\psi)\ F_{mn}(p,q) \ , \end{aligned} \tag{2.85}$$

where we show the corresponding relations in both the tensor and helicity basis. Here the Wigner rotation matrix $d^n_{n'}(\psi)$ is the $SO(2,1)$ analog of the familiar 3×3 "rotation" matrix (since the exchanged particle is spin-1 with 3 polarization states), and ψ is the boost angle that transforms between the hadronic Breit ("brick wall") frame, where $p\|q$, and the leptonic frame where $k\|k'$, with

$$\cosh\psi = \frac{2p\cdot(k+k')}{\Delta[-Q^2, p^2, p_x^2]} = \frac{\xi^2 M^2 - Q^2 + 2\xi(s-M^2)}{\xi^2 M^2 + Q^2} \quad \xrightarrow[M\to 0]{} \quad \frac{2-y}{y} \, . \tag{2.86}$$

Here, $\Delta^2[a,b,c] = a^2+b^2+c^2-2(ab+bc+ca)$, and p_x is the the 4-momentum of the outgoing proton remnant (X), *c.f.*, Ref. [102]. An elegant feature of Eq. (2.85) is that the leptonic (L^{mn}) and hadronic (F_{mn}) helicity structure functions are simple and all the kinematic complications due to the target mass are contained in the $d(\psi)$ rotation matrices. This simplicity is a consequence of the underlying group-theoretic approach to the factorized structure.[19]

[19] While the helicity formalism elegantly exhibits the underlying symmetries of the interaction, as we go to higher-order the computation of the Wigner rotations for multiple intermediate particles can become more complex [102].

2.6.3 Boost-invariant polarizations

If we work in light-front coordinates $\{x^+, \vec{x}, x^-\}$ with $x^\pm = (x^0 \pm x^3)/\sqrt{2}$ and $\vec{x} = \{x^1, x^2\}$, then

$$p_A^\mu = \left\{p_A^+, \vec{0}, \frac{M_A^2}{2p_A^+}\right\} \qquad \text{and} \qquad q^\mu = \left\{-\xi p_A^+, \vec{0}, \frac{Q^2}{2\xi p_A^+}\right\}, \tag{2.87}$$

such that $p_A^2 = M_A^2$ and $q^2 = -Q^2$. Additionally, $2(p_A{\cdot}q) = Q^2/x_A = Q^2/\xi - \xi M_A^2$. We now choose to define the parton momentum for a massive state as

$$k^\mu = \left\{\xi p_A^+, \vec{0}, \frac{m^2}{2\xi p_A^+}\right\}, \tag{2.88}$$

such that $k^2 = m^2$ and $k^+ = \xi p_A^+$, with parton mass m. This scaling relation holds for both massless and massive partons.

We require two reference vectors to define the polarizations. For the hadron polarizations $\epsilon_n^\nu(p,q)$ we choose the vectors $\{p,q\}$, while for the parton polarizations $\epsilon_n^\nu(k,q)$ we choose the vectors $\{k,q\}$. For both the hadron and parton polarizations, the scalar polarization is given by $\epsilon_0^\mu = q^\mu/Q$, and the transverse polarizations are defined as usual to be $\epsilon^\pm = (0, \mp 1, -i, 0)/\sqrt{2}$. The longitudinal polarizations are constructed from the reference vectors using a completeness relation [101]. The polarization vectors only depend on the reference vectors p^μ and k^μ to the extent that they define the $t-z$ plane in conjunction with q^μ; see Ref. [101] for details.

Thus, the polarization vectors are invariant between the hadron and parton reference frame. Additionally, there is a direct relation between the hadronic helicity structure functions F_λ and the partonic helicity structure functions ω_λ:

$$F_\lambda = \delta^{\lambda\lambda'} f \otimes \omega_{\lambda'}\,. \tag{2.89}$$

Here, f is the PDF, $\otimes$ represents a convolution, $\delta^{\lambda\lambda'}$ is the usual Kronecker δ-function, and λ sums over the six helicity configurations. Specifically, we note there is no mixing as indicated by the $\delta^{\lambda\lambda'}$ function. This is in contrast to the structure functions in the tensor basis:

$$F_i = c^{ij}\, f \otimes \omega_j\,, \tag{2.90}$$

where the mixing coefficients c^{ij} contain off-diagonal elements, with indices i, j summing over the six tensor structure functions [101].

	(OPE) TMC	ACOT-TMC	$\frac{\text{ACOT–TMC}}{\text{(OPE) TMC}}$
F_1	$\frac{x}{\xi r}$	1	$\frac{\xi r}{x} \sim 1+\varepsilon^2$
F_2	$\frac{x^2}{\xi^2 r^3}$	$\frac{x}{\xi r^2}$	$\frac{\xi r}{x} \sim 1+\varepsilon^2$
F_3	$\frac{x}{\xi r^2}$	$\frac{1}{r}$	$\frac{\xi r}{x} \sim 1+\varepsilon^2$

Table 2.5: Tabulation of prefactors from the formulae for the TMC (obtained using the OPE) of Eq. (2.39) and ACOT-TMC light-front derivations of Eq. (2.93), where $\varepsilon = (x_A M_A/Q) = (x_N M_N/Q)$, and $r = r_A = r_N = \sqrt{1+4\varepsilon^2}$. While the F_i's are generally evaluated at x_N, using Table 2.2 we have $x_N/\xi_N = x_A/\xi_A$ and these ratios are equivalent. The TMC and ACOT-TMC prefactors differ by a uniform factor $\xi r/x$ and match to $\mathcal{O}(\varepsilon^2)$, and this is plotted in Fig. 2.9. The source of this factor is discussed in the text.

2.6.4 Relationship between TMCs on the light-front and in the OPE

Finally, coming back to the hadronic level, we can relate the hadronic helicity structure functions F_λ to the tensor structure functions F_i using Eq. (2.84) and obtain:[20]

$$F_1 = \frac{1}{2}\{F_+ + F_-\} , \qquad F_2 = \frac{x}{r^2}\{F_+ + F_- + 2F_0\} , \qquad F_3 = \frac{1}{2r}\{-F_+ + F_-\} , \tag{2.91}$$

where $r = r_A = r_N = \sqrt{1+4\varepsilon^2}$ and $\varepsilon = (xM/Q)$. The F_0 helicity structure function is related to the longitudinal structure F_L of Eq. (2.41) via

$$F_L = 2x\,F_0 = r^2 F_2 - 2xF_1 . \tag{2.92}$$

Note there is a relative factor of $2x$ connecting the helicity F_0 and the longitudinal F_L structure functions.

Translating the above relations into the formulae for the TMCs in the light-front approach yields:

$$F_1^{\text{ACOT–TMC}}(x) = F_1^{(0)}(\xi) , \tag{2.93a}$$

$$F_2^{\text{ACOT–TMC}}(x) = \frac{x}{\xi r^2} F_2^{(0)}(\xi) , \tag{2.93b}$$

$$F_3^{\text{ACOT–TMC}}(x) = \frac{1}{r} F_3^{(0)}(\xi) , \tag{2.93c}$$

and $F_L^{\text{ACOT–TMC}}(x)$ can be constructed using Eq. (2.41). The proof of collinear factoriza-

[20]Note that while the F_i structure functions of Eq. (2.53) are evaluated at x_N, the relations of this section hold at the nucleus and averaged nucleon level; hence, we omit here the super/subscripts on F_i and x_N.

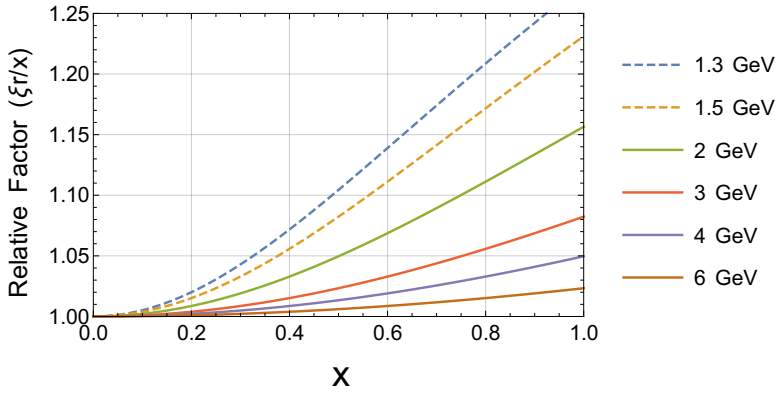


Figure 2.9: We display the relative factor $(\xi r/x)$ between the ACOT and leading OPE terms as tabulated in Table 2.5 for $Q = \{1.3, 1.5, 2, 3, 4, 6\}$ GeV as a function of x with $M = M_{proton}$. The lower two Q values $\{1.3, 1.5\}$ are dashed to indicate these values are below typical kinematic cuts ($Q_{cut} \sim 2$ GeV) for global PDF fits. Recall $(\xi_A r_A/x_A) = (\xi_N r_N/x_N)$.

tion for heavy quarks is detailed in Ref. [9].

The correspondence between the ACOT-TMC and (OPE) TMC expressions are summarized in Table 2.5. In the right-most column, we display the relative conversion factor $(\xi r/x)$, which is the same for all three $F_i^{\rm ACOT-TMC}(x)$ results. Further manipulation of the helicity structure functions in Eq. (2.84) reveals that $F_{4,5,6}$ also follow this pattern. For F_i, we see that the TMC and ACOT-TMC approach align to $\mathcal{O}(\varepsilon^2)$, where $\varepsilon = (xM/Q)$ vanishes outside the dominant TMC region of large x and small Q. Note, these comparisons are for the leading term $F_i^{\rm Leading-TMC}$ only (see below and Sec. 2.7), as the higher order terms in $(xM/Q)^2$ with the h_i and g_i functions are not included in the ACOT expressions.

In the above Eq. 2.93, we have used generic (un-subscripted) variables $\{x, \xi\}$, and this equation is valid for both **nuclear** variables $\{x_A, \xi_A\}$, and **nucleon** variables $\{x_N, \xi_N\}$, (recall, $r_A = r_N$). On the LHS of Eq. 2.93 we have the Bjorken scaling variable x and on the RHS this is modified by the presence of the target mass to be the Nachtmann variable ξ. Note that the nuclear structure function $F_i^{(0)}$ was derived without knowing any details of the partonic structure; this is the essence of the factorization between the partonic (R_{ij}) and hadronic (R_M) corrections. Therefore, the correct scaling variable in Eq. 2.93 is ξ, and not $\bar{\xi} = R_{ij}\,\xi$.

Separately, when we compute $F_i^{(0)}$ in the parton model, we will have individual partonic contributions $f(\bar{\xi}, m_i, m_j)$ that will depend on the parton masses $\{m_i, m_j\}$. We can incorporate the parton masses by using the full rescaling variable $\bar{\xi} = R_{ij}\,\xi = R_{ij} R_M\, x$ (with overline). This scaling of course cannot be done at the hadron level as the scaling for the light quarks $\{u, d, s\}$ will be different from the heavy quarks $\{c, b, t\}$.

2.6.5 Relation of the parton model to the OPE

The parton model ACOT-TMC results differ from the TMC results obtained with OPE in two respects: i) there is an overall factor of $(\xi r/x)$, and ii) the OPE result contains additional $(xM/Q)^2$ and $(xM/Q)^4$ terms. We will discuss the origin of these differences below.

Correspondence of the leading factor $(\xi r/x)$: The TMC (OPE) results of Eq. (2.53) contain an overall factor of $(\xi r/x)$ as compared with the ACOT-TMC results of Eq. (2.93). The source of this factor comes from fully including the hadron mass in the calculation as outlined in the appendix of [48] and is summarized in Table 2.5. Additionally, this factor is also present in Ref. [93]; *c.f.*, Eq. (1.24) where, matching notation, $(1+\mu\xi^2) = (r\xi/x)$ with $\mu = (M/Q)^2$. In Fig. 2.9 we show the size of this relative factor as a function of x for selected Q values. For small x and large Q this factor is close to unity, but grows for large x and small Q values where the TMCs are typically substantial.

Correspondence of the (M/Q) powers: The OPE master equation, Eq. (2.39), includes higher powers of (xM^2/Q^2), i.e., the $\tilde{h}_i^A$ and $\tilde{g}_2^A$ terms, which are formally twist-2 (leading twist). The source of these is the intrinsic transverse momentum of partons (k_T) and manifests through the conservation of momentum [93, 94]. For such configurations, partonic cross sections are evaluated with (initial) parton momenta being on-shell but not collinear with the parent hadron. Details are provided in Refs. [93, 94], which include the transverse momentum in the description of the partons, [*c.f.*, $f(x, k_T^2)$ in Eq. (1.22) of Ref. [93]]. As the transverse momentum is of order M, k_T-dependent contributions generate the (M^2/Q^2) and (M^4/Q^4) terms of Eq. (2.39). Thus, if the transverse momentum is accounted for in the parton model, one matches the OPE result [93, 94].

Recap: We have derived the structure function results for the parton model using the ACOT formalism in a helicity basis as summarized in Eq. (2.93). We can obtain a complete correspondence between the parton model and OPE structure function results if i) we fully account for the hadron mass in the delta function (c.f., Eq.A.124 of [48]), which ensures the relationship between the Bjorken and Nachtmann variables, and ii) we account for the transverse momentum of the parton (c.f., Ref. [93]).

2.7 Numerical results

Having discussed various theoretical and phenomenological aspects of TMCs for nuclei from an analytic perspective, we now turn our focus to the quantitative aspect of TMCs. In this section, we explore the numerical impact of TMCs on first the proton structure functions (Sec. 2.7.1), and then the nuclear structure functions (Sec. 2.7.2). We will examine both the leading-TMC and sub-leading-TMC contributions, and explore the variation across the nuclear A range. These observations will allow us to design a simple parameterization of the TMCs which are computationally efficient (Sec. 2.7.3). Finally, we will compute the impact of the TMCs for DIS reduced cross sections with sample kinematics from HERA, JLab, EIC, and LBNF (Sec. 2.7.4).

In order to build intuition for numerical results throughout this section, it is useful to recall several relations between structure functions and quark and antiquark PDFs, but which hold only at LO in the absence of CKM mixing and masses. In particular, for neutrino and antineutrino DIS on an arbitrary nuclear target (A), where A may also be just a proton, the charged current structure functions ($F_i^{W^\pm}$) and PDFs are related by:

$$F_1^{\nu A} = (d + s + \bar{u} + \bar{c}) \,, \qquad F_1^{\bar{\nu} A} = \left(u + c + \bar{d} + \bar{s}\right) \,, \tag{2.94a}$$

$$F_2^{\nu A} = 2x \left(d + s + \bar{u} + \bar{c}\right) \,, \qquad F_2^{\bar{\nu} A} = 2x \left(u + c + \bar{d} + \bar{s}\right) \,, \tag{2.94b}$$

$$F_3^{\nu A} = +2 \left(d + s - \bar{u} - \bar{c}\right) \,, \qquad F_3^{\bar{\nu} A} = -2 \left(u + c - \bar{d} - \bar{s}\right) \,, \tag{2.94c}$$

in the limit of four quarks. Here $\{u, d, ...\}$ are the PDFs of a full nucleus A. Likewise, when the exchange of Z bosons can be neglected in charged lepton DIS, the neutral current structure functions (F_i^γ) are related to PDFs at LO by

$$F_2^{l^\pm A} = x \frac{1}{9} \left[4(u + \bar{u}) + (d + \bar{d}) + 4(c + \bar{c}) + (s + \bar{s})\right]. \tag{2.95}$$

While the LO relations shown above are useful for building intuition, our calculations are performed at full NLO in QCD including the quark mass contributions; specifically, we use the S-ACOT(χ) scheme [159].

Since we will compare separate components of the TMC contributions, to reduce ambiguity we introduce the nomenclature which we will use throughout our presentation.

Schematically, the OPE TMC terms are related as follows:

$$\overbrace{F_2^{A,\text{TMC}}(x_N,Q^2)}^{\textbf{TMC}} = \underbrace{\left(\frac{x_N^2}{\xi_N^2 r_N^3}\right)\overbrace{F_2^{A,(0)}(\xi_N,Q^2)}^{\substack{\textbf{No-TMC}\\ \textbf{with } \xi\to x}}}_{\textbf{Leading-TMC}} + \underbrace{\left(\frac{6M_N^2x_N^3}{Q^2r_N^4}\right)h_2^A(\xi_N,Q^2)}_{\textbf{h-term}} + \underbrace{\left(\frac{12M_N^4x_N^4}{Q^4r_N^5}\right)g_2^A(\xi_N,Q^2)}_{\textbf{g-term}}. \tag{2.96}$$

We provide additional details below.

TMC: We identify the full set of contributions to the structure function, given by the rescaled OPE equation of Eq. (2.53), as the "TMC" result; this label is without additional qualifiers.

No-TMC: As we take the $(xM^2/Q^2) \to 0$ limit, the pre-factors of $F_i^{A,(0)}$ become unity, and the higher-order $(xM/Q)^2$ and $(xM/Q)^4$ terms containing the $\{h_i, g_i\}$ functions vanish. Additionally, we have $\xi \to x$ so that the expressions in Eq. (2.53) reduces to $F_i^{A,(0)}(x_N, Q^2)$ alone. We refer to this as the "No-TMC" result.[21]

Leading-TMC: The "Leading-TMC" structure functions are obtained from Eq. (2.53) by keeping only the first term on the RHS which is proportional to $(xM/Q)^0$. Specifically, we are neglecting the terms proportional to $(xM/Q)^2$ and $(xM/Q)^4$ which contain the $\{h_i, g_i\}$ functions.

ACOT-TMC: We obtain the "ACOT-TMC" structure functions from the ACOT TMC equation of Eq. (2.93). This result is similar to the "Leading-TMC," but the pre-factors differ by $(\xi r/x)$ as detailed in Table 2.5. Importantly, the "ACOT-TMC" result does not include the higher-order $(xM/Q)^2$ and $(xM/Q)^4$ terms containing the $\{h_i, g_i\}$ functions.

h- and g-terms: The "h" and "g" terms are those terms in the full (OPE) TMC result that are proportional to the h_i and g_i functions. We observe that the "h" contributions are proportional to (M_N^2/Q^2), while the "g" contributions are proportional to (M_N^4/Q^4).

2.7.1 Proton structure functions with TMCs

In comparison to structure functions for massless protons, structure functions with TMCs for nuclei contain two additional layers of complexity. The first, of course, is the larger

[21] Note, we always retain the full quark mass dependence in all calculations as this is factorized from the hadron-level kinematics. *C.f.* Ref. [104] for details. For this reason, we **do not** refer to the **No-TMC** term as a *massless* result.

nucleon content; the second are the TMCs themselves. Therefore, in order to establish a baseline intuition of TMCs for nuclear structure functions, we consider briefly TMCs for proton structure functions. For more comprehensive studies of TMCs for protons, see Ref. [104] and references therein.

We will show results for a range of Q values. For a typical global analysis, the PDF evolution uses an initial scale Q_0 in the range of 1.3 to 1.5 GeV, and a typical kinematic cut is $Q \gtrsim 2$ GeV. For example, in Fig. 2.10 we display $Q = 1.5$ GeV, which is near the initial evolution scale but below the typical kinematic cut. We also display $Q = 10$ GeV, which is above the kinematic Q cut (the W cut is more complex).

We begin with Fig. 2.10, where in the upper panel we plot the absolute structure functions F_i for charged current (W^-) and neutral current (γ, Z) exchange showing full "TMC" (solid), "Leading-TMC" (dash), "No-TMC" (dash-dot), and "ACOT-TMC" (dotted), at $Q = 1.5$ GeV and 10 GeV as a function of the Bjorken scaling variable x. In the lower panel we show the ratio of the full "TMC" to the "No-TMC" results.

F_i structure functions: (Fig. 2.10 Upper Panels) We focus first on the upper panels of Fig. 2.10 which shows the proton structure functions vs. x. All the structure functions are steeply decreasing for large x. This reflects the underlying PDF structure of Eqs. (2.94) and (2.95): PDFs, and hence structure functions, must decrease at large x due to momentum conservation.

For both $Q = 1.5$ GeV and $Q = 10$ GeV, we observe that the full "TMC," "Leading-TMC" and "ACOT-TMC" results roughly coincide. The good agreement between the full "TMC" and "Leading-TMC" allows us to infer that the sub-leading contributions containing the $\{h_i, g_i\}$ terms are small; this is to be expected as they are suppressed by powers of $(M/Q)^2$ and $(M/Q)^4$. We will further examine the impact of these sub-leading terms in Sec. 2.7.2. The approximate agreement of the "Leading-TMC" and the "ACOT-TMC" reflects the impact of the relative factor of $(\xi r/x)$ discussed in Sec. 2.6.4.

In contrast, at $Q = 1.5$ GeV, the "No-TMC" results are dramatically different from the other three curves. The "No-TMC" expression is most similar to the "Leading-TMC" as illustrated below for the case of F_1:

$$F_1^{\text{TMC-leading}}(x_N) = \frac{x_N}{\xi_N r_N} F_1^{(0)}(\xi_N)\,, \tag{2.97a}$$

$$F_1^{\text{No-TMC}}(x_N) = F_1^{(0)}(x_N)\,, \tag{2.97b}$$

where $r_N = \sqrt{1 + 4\varepsilon^2}$. There are two differences between these expressions: i) for No-TMC

the pre-factor is set to unity, and ii) the argument of $F_i^{(0)}$ is set to x instead of ξ. Given that $\xi = 2x/(1+r) \simeq x(1-x^2M^2/Q^2)$, the expression for $F_1^{\text{No-TMC}}$ is neglecting the (M/Q) mass effects.

We can separately explore the impact of these two components and determine which is dominant. For F_1 the relative factor $x/(\xi r)$ is precisely the inverse of what is displayed in Fig. 2.9. Thus, we observe that for $Q = 1.5$ GeV and $x = 0.9$, this factor yields a shift of ~20%. Comparing with Fig. 2.11, we see this is clearly not sufficient to explain the observed large differences between the No-TMC and Leading-TMC results. While F_2 and F_3 have slightly different prefactors, we obtain the same conclusion that these prefactors do not generate the sizable differences observed between the "No-TMC" and the other results.

Thus, the dominant variation of $F_i^{\text{No-TMC}}$ and $F_i^{\text{TMC-Leading}}$ must be due to the different arguments: x vs. ξ. This correspondence of x and ξ is displayed in Fig. 2.3 where we observe that $x \sim \xi$ for small x and large Q, but for large x and small Q we find $\xi < x$. Because the PDFs are steeply falling at large x (where x and ξ have the largest deviation), it is the difference of ξ vs. x in the arguments on the RHS of Eq. (2.97) that is driving the variation observed in Fig. 2.11. While we have detailed the comparison of F_1 in Fig. 2.10, the same rationale applies to the other structure functions displayed.

In contrast to the sizable differences between the "No-TMC" result and the other three curves for the lower Q value (1.5 GeV), the "No-TMC" result is roughly comparable to the others for $Q = 10$ GeV. In total, these comparisons affirm that TMCs are negligible for small-x or for intermediate to large Q.

F_i structure function ratios: (Fig. 2.10 Lower Panels) To highlight the importance of including the TMCs, we now examine the lower panels of Fig. 2.10 which display the ratio of F_i^{TMC} to $F_i^{\text{No-TMC}}$. The ratio plots accentuate subtle features that are not evident on the log-scale plots of the upper panels.

If we first focus on the results at $Q = 1.5$ GeV, we observe the deviation between F_i^{TMC} and $F_i^{\text{No-TMC}}$ differs markedly for the separate $F_{1,2,3}$ structure functions. These differences arise from: i) the different prefactors, see, e.g. Eq. (2.53), ii) the different argument of $F_i^{(0)}$ (as noted previously), and also iii) the sub-leading $\{h_i, g_i\}$ contributions. It is the combination of these components that causes the variation among the Fig. 2.10 sub-figures. For example, we see the F_1 curves deviate at smaller x values as compared with the F_2 and F_3 results. Additionally, the F_2 and F_3 ratios dip below unity for intermediate

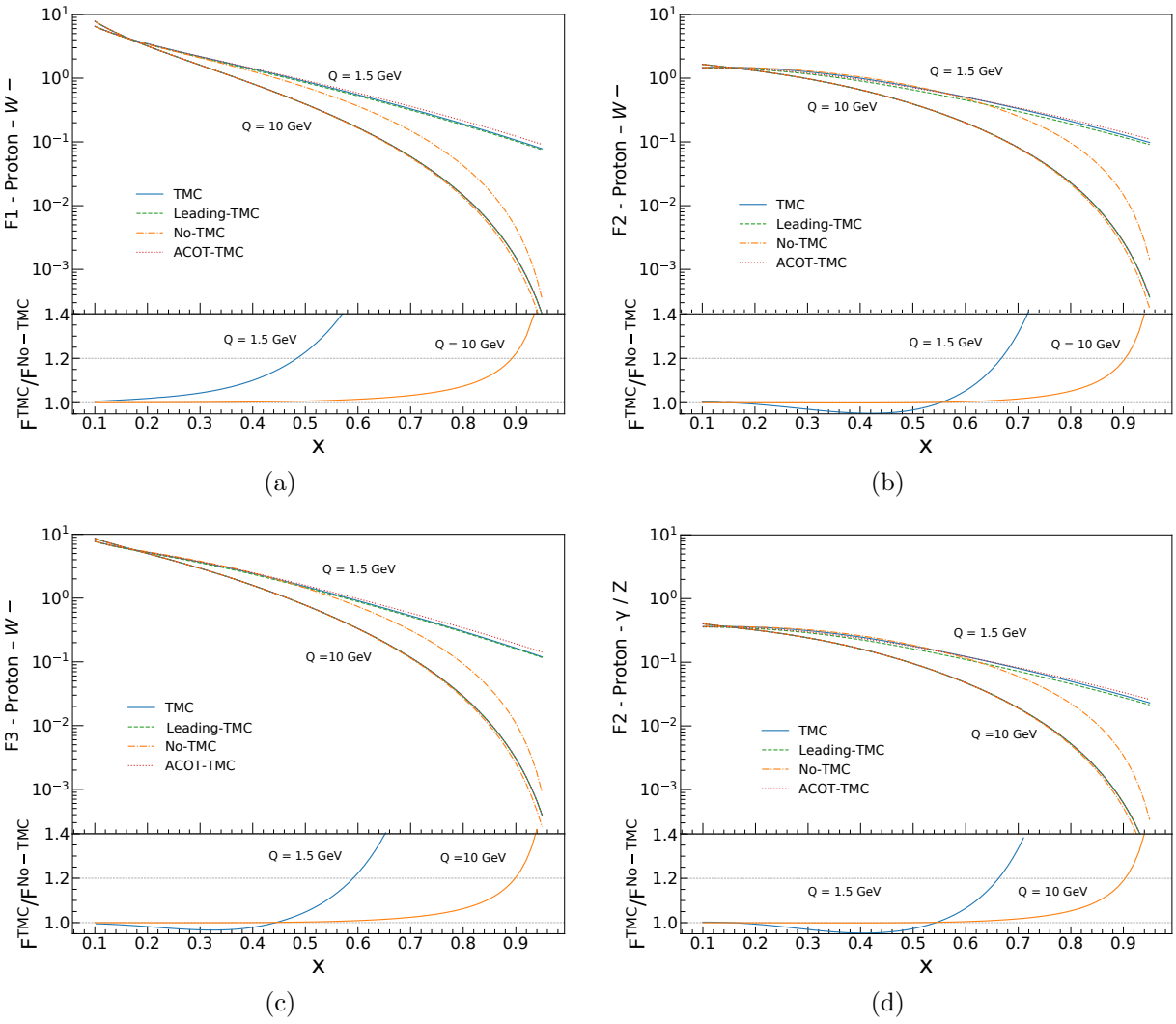


Figure 2.10: **Upper panels:** Proton structure functions for charged current W^- (a) F_1, (b) F_2, and (c) F_3 and (d) neutral current γ/Z F_2 as a function of momentum fraction x_N. We display the results for the full "TMC" (solid blue), "Leading-TMC" (dashed green), "No-TMC" (dot-dash orange), and "ACOT-TMC" (dotted red) at Q = 1.5 GeV (upper lines) and 10 GeV (lower lines). **Lower panels:** Ratio of F_i with full "TMC" to "No-TMC" at Q = 1.5 GeV and 10 GeV.

x values ($x \sim 0.3 - 0.5$) while the F_1 ratio is above unity.

2.7.2 Nuclear structure functions with TMCs

Having examined the proton structure functions in Fig. 2.10, we now consider the corresponding nuclear structure functions of Fig. 2.11. The list of nuclear isotopes that we consider are summarized in Tab. 2.6.

Symbol	A	Z	J^P	Symbol	A	Z	J^P	Symbol	A	Z	J^P
H	1	1	$\frac{1}{2}^+$	N	14	7	1^+	$\mathrm{Ag}_{\mathrm{iso}}$	108	54	-
D	2	1	1^+	Ne	20	10	0^+	$\mathrm{Sn}_{\mathrm{iso}}$	119	59.5	-
^{3}He	3	2	$\frac{1}{2}^+$	Al	27	13	$\frac{5}{2}^+$	Xe	131	54	$\frac{3}{2}^+$
He	4	2	0^+	Ar	40	18	0^+	W	184	74	0^+
Li	6	3	1^+	Ca	40	20	0^+	Au	197	79	$\frac{3}{2}^+$
Li	7	3	$\frac{3}{2}^-$	Fe	56	26	0^+	$\mathrm{Au}_{\mathrm{iso}}$	197	98.5	-
Be	9	4	$\frac{3}{2}^-$	$\mathrm{Cu}_{\mathrm{iso}}$	64	32	-	$\mathrm{Pb}_{\mathrm{iso}}$	207	103.5	-
C	12	6	0^+	$\mathrm{Kr}_{\mathrm{iso}}$	84	42	-	Pb	208	82	0^+

Table 2.6: List of nuclear PDFs considered in this work. We use the nCTEQ15 nPDF set [35]. The nuclei indicated with "iso" subscript are isoscalar corrected; thus, the Z value can be half-integer. We will also show results with a neutron PDF computed using isospin symmetry.

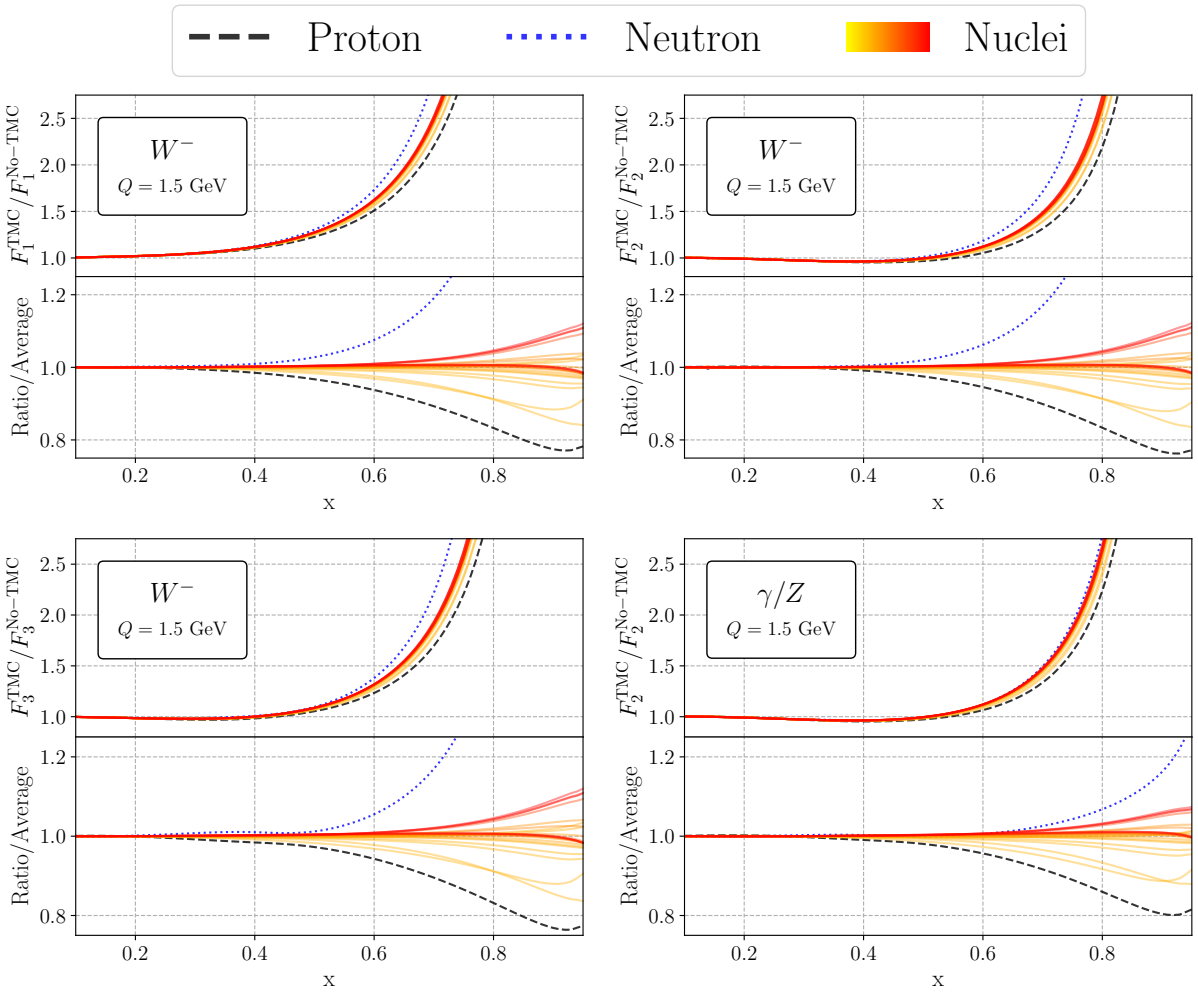


Figure 2.11: Ratio of nuclear structure functions F_i for the full "TMC" over "No-TMC" as a function of x_N at Q = 1.5 GeV. We display ratios for charged current W^- (a) F_1, (b) F_2, and (c) F_3 and (d) neutral current γ/Z F_2 for the nuclei listed in Table 2.6. The proton is indicated with the black dashed line, and the neutron with the blue dotted line. The narrow band of colored lines shows the various nuclear results.

Full TMCs for nuclei

In Fig. 2.11 we plot the ratio of the full TMC structure functions $F_i^{\rm TMC}$ to the No-TMC structure functions $F_i^{\rm No-TMC}$ for charged current (W^-) and neutral current (γ, Z) exchange for the nuclei of Table 2.6. This ratio highlights the impact of the TMCs.

We observe the general behavior of the nuclear ratios displayed in Fig. 2.11 are very similar to the corresponding proton results in the lower panels of Fig. 2.10. Because the nuclear and proton results are so similar, we only plot a single Q value (1.5 GeV) in Fig. 2.11 as other Q values (e.g., Q = 10 GeV) will be similar to the proton results. We will study the detailed Q dependence in the following.

The results of Fig. 2.11 clearly demonstrate that the TMCs for the nuclei are effectively independent of A. We attribute this curious finding to the re-scaling property ($x_A M_A = x_N M_N$) as shown in Sec. 2.4. This re-scaling allows us to identically rewrite the TMCs for structure functions in Eq. (2.38), which are functions of nuclear-level quantities x_A and M_A, in terms of nucleon-level quantities x_N and M_N. As a result, the variable nucleon content of nuclei, which is the defining characteristic of nuclei, is averaged out, resulting in near universal behavior.

To illustrate the emergence of this universality, let us Taylor-expand the nuclear structure function ratio $F_1^{\rm TMC-Leading}/F_1^{\rm No-TMC}$ to show this is independent of the nuclear A value up to corrections $\varepsilon^2 = (xM/Q)^2$. We use the expressions of Eq. (2.97) and computing the derivative as $F_i^{(0)}(\xi_N) \approx F_i(x_N) + \delta x_N [dF_i(y)/dy]_{y=x_N}$, where $\delta x_N = (\xi_N - x_N) \approx (x^3 M_p^2/Q^2) \ll 1$, If we apply this approximation to the F_1 ratio, for example, we then obtain:

$$\frac{F_1^{\rm TMC-Leading}(x_N)}{F_1^{\rm No-TMC}(x_N)} \sim \left(1-\varepsilon^2\right) \frac{F_1^{(0)}(\xi_N)}{F_1^{(0)}(x_N)} \sim \left(1-\varepsilon^2\right) \frac{F_1^{(0)}(x_N) + \mathcal{O}(\varepsilon^2)}{F_1^{(0)}(x_N)} \sim \left(1-\varepsilon^2\right) + \mathcal{O}\left(\varepsilon^4\right) \ , \tag{2.98}$$

where we have defined $\varepsilon = (xM/Q)$.

Remarkably, we find that the ratio $F_1^{\rm TMC-Leading}(x_N)/F_1^{\rm No-TMC}(x_N)$ is independent of A up to ε^2. Recall the full TMC ratio $F_1^{\rm TMC}(x_N)/F_1^{\rm No-TMC}(x_N)$ differs only by the additional $\{h_i, g_i\}$ terms which are suppressed by powers of $(M/Q)^2$. This implies that nuclear TMCs themselves, when defined in terms of averaged quantities, are essentially universal.

We now explore this approximate universal property of the TMCs, and determine how we can exploit this property to simplify certain nuclear TMC calculations.

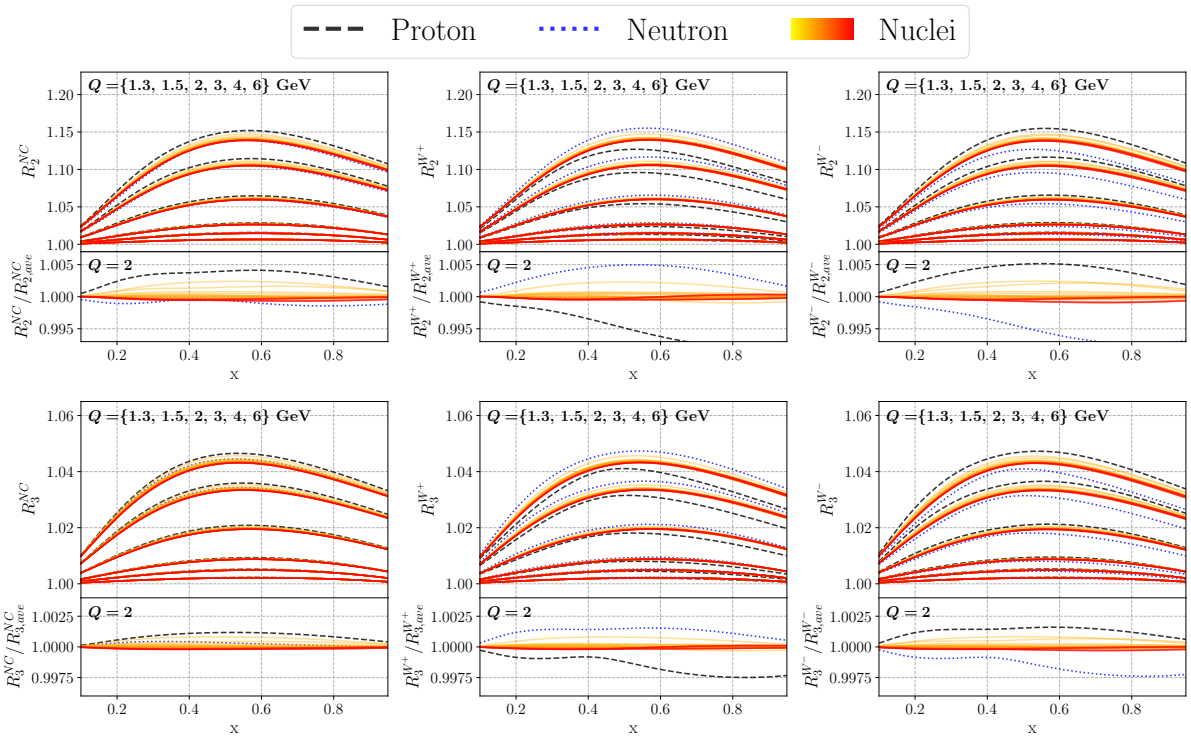


Figure 2.12: **Upper panels:** We display the nuclear structure function ratios of $F_i^{\text{TMC}}/F_i^{\text{TMC-Leading}}$ for selected CC and NC processes for $Q = \{1.3, 1.5, 2, 3, 4, 6\}$ GeV (from top to bottom) as a function of x_N. **Lower panels:** We display the variation of this ratio compared to the average variation as computed in Eq. (2.100) for $Q = 2$ GeV. In the top row we show F_2 results for (a) NC γ/Z exchange, (b) CC W^+ exchange, and (c) CC W^- exchange, and the bottom row shows F_3 results for the same processes. In all panels, the proton is indicated with the black dashed line, and the neutron with the blue dotted line. The narrow band of colored lines shows the various nuclear results.

Comparison of full vs. leading-TMCs for nuclei

In Fig. 2.10 we observed that the "Leading-TMC" yielded a reasonable approximation to the full "TMC" result, even at low Q and large x values. Computationally, the "Leading-TMC" result is much simpler to compute as the full "TMC" result include contributions from the $\{h_i, g_i\}$ terms, each of which includes an integral. Additionally, if we can take advantage of the approximate A independence observed in the previous section, this may allow us to greatly simplify the calculation of TMCs for the many different nuclei present in a typical nPDF fit.

Magnitude of ratios: (Fig. 2.12 upper panels) We start by comparing the ratio of the "Leading-TMC" to the full "TMC" result as shown in the upper panels of Fig. 2.12. The results are shown for selected CC and NC processes with representative values $Q = \{1.3, 1.5, 2, 3, 4, 6\}$ GeV.

We display results for all the nuclei of Table 2.6 In these plots, the individual nuclear A

values are not labeled, but the ratios coalesce into clearly defined bands for each Q value. This coalescence holds for all permutations of structure functions and boson exchanges. The different nuclei are shown as colored solid lines in the figure. We also include results for the proton (dashed black) and neutron (dotted blue).

Focusing first on the bands of the top panels, we see that for F_2^A the differences between the full TMCs and Leading-TMCs are as large as $\mathcal{O}(10\% - 15\%)$ at $x \sim 0.5$ for $Q = 1.3$ GeV, and $\mathcal{O}(4\% - 5\%)$ for F_3^A. For larger Q values we expect these differences to be reduced as they are proportional to powers of $(M/Q)^2$. At $Q = 2$ GeV, which is a typical Q_{cut} value for many global PDF analyses, the ratio for F_2^A is $\lesssim 5\%$, and the ratio of F_3^A is $\lesssim 2\%$. For $Q \gtrsim 3-4$ GeV, we find that these differences reduce to the sub-percent level, and effectively vanish at $Q = 6$ GeV.

Nuclear A dependence of ratios: Having discussed the magnitude of the ratios $F_i^{\rm TMC}/F_i^{\rm TMC-Leading}$, we now examine the A dependence of these ratios. A distinctive feature of Fig. 2.12 is the coalescence of the results into individual bands, and this suggests that the nuclear A dependence of this ratio is minimal. [The proton and neutron ratios (shown as dashed and dotted black lines, respectively), do not necessarily lie within the bands, and we will discuss these separately.] This apparent universality of the full TMC/Leading-TMC ratios can be traced back to: i) the fact that this ratio has only mild dependence on the underlying PDF, and ii) the fact that for nuclei, in the large x region it is the average $u + d$ (isoscalar) PDF that dominates this result.

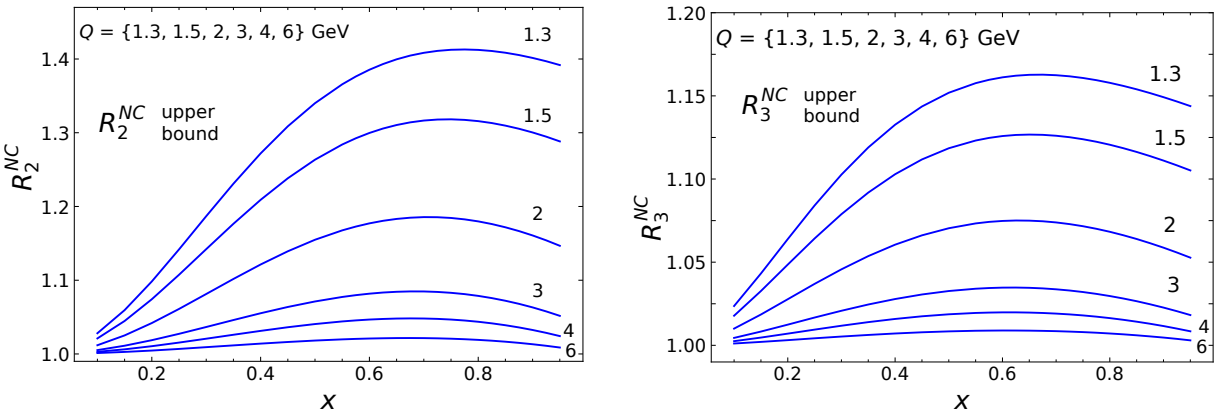
One approach to understand these features is to consider the analytically computed upper bound for these ratios. If we assume the structure functions are monotonically decreasing (an entirely reasonable assumption in the large x region), it is possible to obtain the following constraints on these ratios [104]:

$$\frac{F_2^{\rm TMC}}{F_2^{\rm TMC-leading}}(x,Q^2) \;\le\; 1 + \left(\frac{M}{Q}\right)^2 \frac{6x\xi}{r}(1-\xi) + \left(\frac{M}{Q}\right)^4 \frac{12x^2\xi^2}{r^2}(-\ln\xi - 1 + \xi)\,, \tag{2.99a}$$

$$\frac{F_3^{\rm TMC}}{F_3^{\rm TMC-leading}}(x,Q^2) \;\le\; 1 - \left(\frac{M}{Q}\right)^2 \frac{2x\xi}{r}\ln\xi\,. \tag{2.99b}$$

Note, these bounds have absolutely no dependence on the PDF. Here, we also explicitly see the powers of $(M/Q)^2$ which drive the ratios to unity for large Q.

In Fig. 2.13, we plot these bounds as a function of x for selected Q values. Comparing the magnitudes of the bounds with the values of Fig. 2.12, we see the bounds are quite conservative. Note that these bounds are valid for any nPDFs in the large x region (where F_i^A is decreasing monotonically), and that is the relevant region of interest for the TMC



(a) (b)

Figure 2.13: We show the upper bounds of $R_i^{\rm NC} = F_i^{\rm TMC}/F_i^{\rm TMC-Leading}$ for a) F_2 and b) F_3 with selected $Q = \{1.3, 1.5, 2, 3, 4, 6\}$ GeV, (from top to bottom) as a function of x_N, as computed with Eqs. (2.99a) and (2.99b).

effects. Consequently, the independence of the bound on the PDF helps us understand the comparable nuclear A independence of the $F_i^{\rm TMC}/F_i^{\rm TMC-Leading}$ ratios observed in Fig. 2.12.

Nuclear A variation of ratios: The observation that these ratios are relatively insensitive to the nuclear A value suggests a "short cut" which can be used to efficiently implement TMCs into numerical calculations.

More specifically, computing the full TMCs for each nucleus A requires a calculation of the $h_i^A(\xi_N, Q^2)$ and $g_i^A(\xi_N, Q^2)$ terms in Eq. (2.53). Each term requires a separate integration over a nuclear structure function. Clearly, such steps would require significant CPU time if they were included inside the fitting loop of a global PDF analysis. If it is true that the size of TMCs to nuclear structure functions are relatively insensitive to the nuclear A value, as our works suggests, then TMCs can be estimated through the application of a "universal" correction factor, $R_i^V(x, Q^2)$. Such a factor can be constructed by taking an average TMC summed over the various n_A nuclei:

$$R_i^V(x_N, Q^2) \;=\; \frac{1}{n_A}\sum_k^{n_A} \frac{F_i^{A_k,\rm TMC}(x_N, Q^2)}{F_i^{A_k,\rm TMC-Leading}(x_N, Q^2)} \;=\; \frac{1}{n_A}\sum_k^{n_A} R_i^{k,V}\,, \tag{2.100}$$

where "V" labels the type of the exchanged gauge boson ($\gamma, Z, W^\pm$), and the sum over k excludes the proton and neutron.

The utility of the "universal" correction factor $R_i^V(x, Q^2)$ is that we can compute this once at the beginning of a fit, and then apply this correction factor inside the fitting loop without the need for recomputing additional integrals over h_i and g_i functions.

To determine the potential precision of such an approximation, we show in the lower

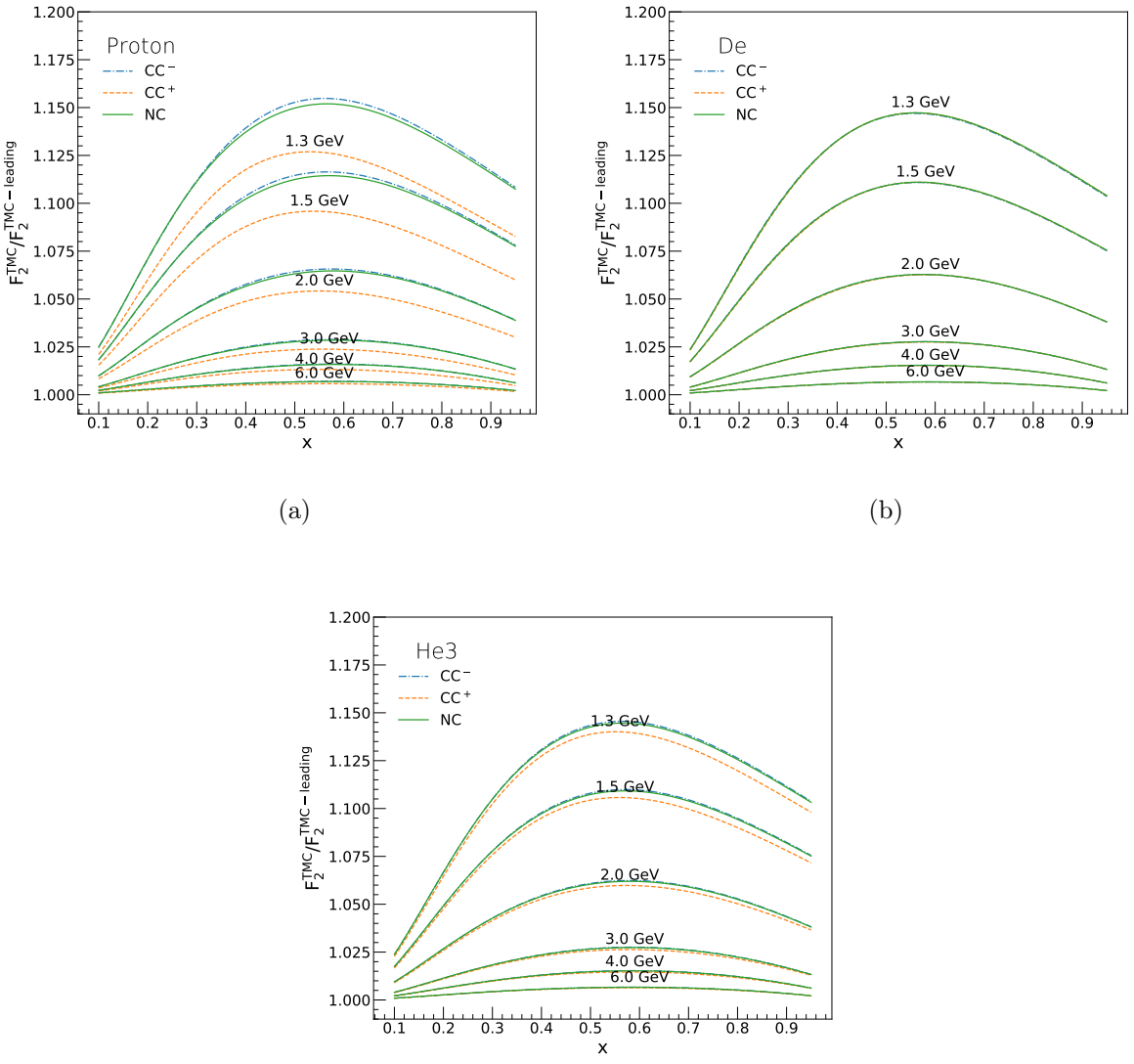
panels of Fig. 2.12 a ratio of ratios constructed by taking $(F_i^{\rm TMC}/F_i^{\rm TMC-Leading})$ over R_i^V for $Q = 2$ GeV, which is a typical $Q_{\rm cut}$ value for global fits. It is remarkable how narrow the band of nuclear PDFs lie in comparison to the universal R_i^V function. For F_2 we see the nuclear bands lie within 0.2% of the universal curve, while for F_3 the bands lie within 0.1%. For larger Q values, the precision is further increased. This is clearly sufficient accuracy for any TMC calculation given the typical nPDF uncertainty in the large x region.

Variation of proton & neutron ratios: In Fig. 2.12 we observed that all the nuclear PDFs were uniformly within a narrow band. In contrast, the proton and neutron results displayed a wider variation lying outside these nuclear bands. We will explain the source of this variation, and understand why the universal correction factor $R_i^V(x, Q^2)$ previously introduced, still works well in the case of nuclei.

Specifically, in Fig. 2.12 we see that the proton ratios lie above the nuclear band for the NC and CC W^- plots, and below for the CC W^+ plots; furthermore, the behavior of the neutron ratios is exactly the opposite of the proton. We can understand these features by identifying the dominant contribution for each process. As $x \to 1$, we neglect the gluon and sea quarks to obtain: $F_2(\gamma/Z) \sim \frac{x}{9}[4u(x) + d(x)]$ and $F_2(W^-) \sim 2xu(x)$, while $F_2(W^+) \sim 2xd(x)$. Recall that for a proton PDF, the d/u ratio in the large x region can typically be as small as ~ 0.1; that is, the up quark is an order of magnitude larger than the down quark.

We now understand the pattern of Fig. 2.12. The structure function for the NC and CC W^- processes are driven by the up quark, while the CC W^+ is driven by the down quark. Thus, the proton results are above the others for NC and CC W^- but below for CC W^+ in plots of Fig. 2.12. Since we obtain the neutron PDFs by isospin symmetry $(u \leftrightarrow d)$, so we also understand why the neutron results are the opposite of the proton.

This exercise also demonstrates why the results for the nuclear PDFs lie together in a uniform band. Most of the nuclei are closer to the isoscalar state $(A \sim 2Z)$ than either the proton $(Z = A)$ or neutron state $(Z = 0)$. For an isoscalar nuclei we have $u = d$, and it is effectively the average of the up and down PDFs $(u+d)/2$ which provides the dominant contributions to Fig. 2.12. Consequently, we find that the nuclear band is effectively an average of the proton and neutron result, and thus lies between the two in Figs. 2.12. The one slight variation of this pattern is the parity-violating F_3 NC structure function; this effect must come from the Z boson contribution because the parity-conserving photon contribution vanishes.



(a) (b)

(c)

Figure 2.14: The nuclear structure function ratio $F_2^{\mathrm{TMC}}/F_2^{\mathrm{TMC-Leading}}$ as a function of x_N for Q = $\{1.3, 1.5, 2, 3, 4, 6\}$ GeV for (a) the proton, (b) deuteron, and (c) ^{3}He. For each nucleus, we overlay the result for NC exchange (solid green), CC W^+ exchange (dotted orange), and CC W^- exchange (dot-dashed blue). Note that for the deuteron in Fig. (b), the curves for the three CC/NC processes coincide.

Approximate nuclear A independence: We illustrate this further in Fig. 2.14 which shows the ratios for the NC and CC processes for the a) proton, b) deuteron, and c) ^{3}He. We choose these examples because the deuteron is an isoscalar, and ^{3}He is one of the most "non-isoscalar" nuclei (with the exception of hydrogen/proton).

The pattern in Fig. 2.14 is reminiscent of Fig. 2.12. In Fig. 2.14(a) for the proton, we see the (u-quark dominated) NC and CC W^- processes lie above the (d-quark dominated) CC W^+ process. In Fig. 2.14(b) for the isoscalar deuteron, we see all the curves coincide since $u = d$ for this nuclei. Finally, in Fig. 2.14(c) for the ^{3}He, we see the NC and CC W^- processes lie slightly above the CC W^+ process, but the difference is not as dramatic as for the proton case. Note in these cases, the $u : d$ ratio is $2 : 1$ in the proton, $1 : 1$ in the deuteron, $5 : 4$ in ^{3}He. For any heavier nuclei, the $u : d$ ratio will be comparatively closer to unity, and the results of the individual NC and CC processes will tend to coincide.

The outcome of these findings is that given the typical uncertainty of nuclear PDFs in the large x region, applying an average, A-independent set of TMCs yields precision which is sufficient for a wide variety of real-world calculations. We shall exploit this observation in Sec. 2.7.3 and derive an approximate parameterization for these TMCs which can be applied to the general nuclear case.

Setting aside the proton and neutron cases, at $Q = 2$ GeV we find the deviation of individual nuclei from the average to be less than $\sim 0.2\%$ for F_2, and less than $\sim 0.1\%$ for F_3. And of course, these differences decrease for larger Q.

Sub-leading TMCs for nuclei

As a final study, we want to demonstrate the individual size of the sub-leading h_2^A and g_2^A components of the full TMCs in Eq. (2.53). In Fig. 2.15 we present the results for NC (γ/Z) and CC ($W^\pm$) processes, and we display a pair of ratios $F_2^{\text{TMC+h2}}/F_2^{\text{TMC-Leading}}$ and $F_2^{\text{TMC}}/F_2^{\text{TMC-Leading}}$, for each Q value.

Recall that $F_2^{\text{TMC-Leading}}$ neglects the sub-leading h_2^A and g_2^A contributions, where h_2^A is suppressed by $(M/Q)^2$, and g_2^A is suppressed by $(M/Q)^4$. The $F_2^{\text{TMC+h2}}$ expression includes the h_2^A contribution, but neglects g_2^A. Finally, the full F_2^{TMC} expression includes both h_2^A and g_2^A; this could equivalently be called $F_2^{\text{TMC+h2+g2}}$.

Since the g_2^A term is suppressed by $(M/Q)^4$, we expect this will only contribute for very large x at low Q^2 values. Therefore, we expect $F_2^{\text{TMC+h2}}$ and F_2^{TMC} to coincide throughout most of the parameter space. This expectation is validated in Fig. 2.15 where we do see that $F_2^{\text{TMC+h2}}$ and F_2^{TMC} differ only for the lowest few Q values in the large x region. For example, at $Q = 2$ GeV the differences are barely discernible in the figure, and

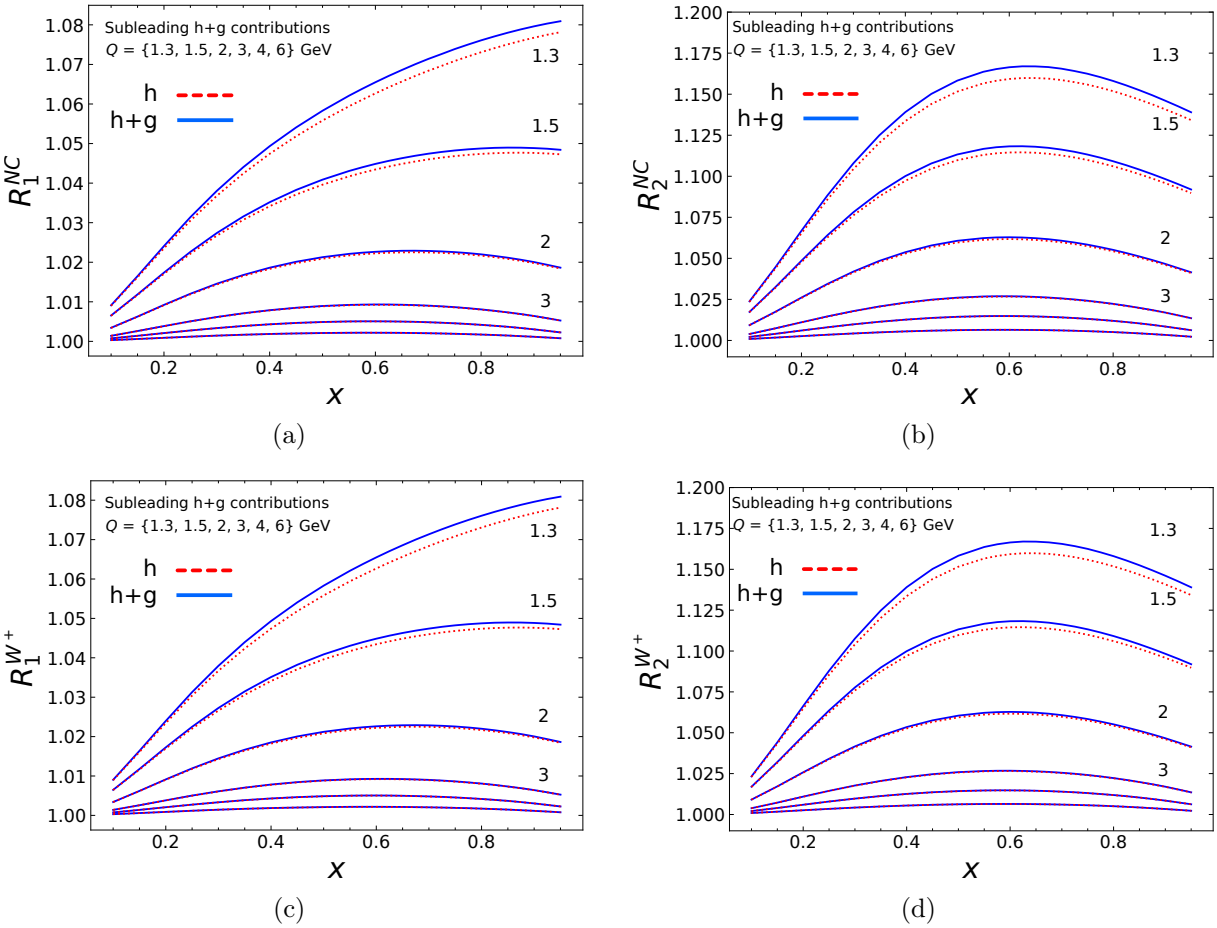


Figure 2.15: To highlight the effect of the sub-leading terms $\{h_i, g_i\}$, we display the ratios $R_i^{Z,W^\pm}$. Specifically, the "h" curve (dashed red) is the result for $[F_i^{\text{TMC–Leading}} + F_i^{\text{h–term}}]$, and the "h+g" curve (solid blue) is $[F_i^{\text{TMC–Leading}} + F_i^{\text{h–term}} + F_i^{\text{g–term}}]$, both compared to $[F_i^{\text{TMC–Leading}}]$. The notation is detailed in Eq. (2.96). We display selected $Q = \{1.3, 1.5, 2, 3, 4, 6\}$ GeV (from top to bottom) as a function of x_N using the nCTEQ15 PDF for ^{12}C. We observe for $Q \geq 2$ GeV (a typical kinematic cut), the "h+g" contributions are less than ~2%.

the effect is even smaller as Q increases.

Thus, h_2^A provides the dominant contribution throughout the kinematic range, but g_2^A does contribute for low Q values. The pattern for the other structure functions $F_{1,3}^A$ is similar.

2.7.3 Parameterizing $F_i^{\text{TMC}}/F_i^{\text{Leading–TMC}}$

In Sec. 2.7.2 we demonstrated that the full TMC/Leading-TMC ratios were effectively insensitive to the nuclear A value. In this section, we provide a parameterization for these ratios which allows us to efficiently compute the full TMCs given the Leading-TMC structure functions.

	$F_1^{TMC}/F_1^{leading}$		$F_2^{TMC}/F_2^{leading}$		$F_3^{TMC}/F_3^{leading}$	
nPDFs	λ_1	δ_1	λ_2	δ_2	λ_3	δ_3
nCTEQ15	2.275	-0.014	2.144	0.100	2.183	0.029
EPPS16	2.226	-0.026	2.086	0.095	2.142	0.022
nNNPDF2.0	2.226	-0.025	2.086	0.103	2.197	0.023
TUJU19	2.290	-0.0170	2.167	0.099	2.200	0.031

Table 2.7: We approximate the $\{h_i(\xi), g_i(\xi)\}$ contributions to $F_i^{\rm TMC}$ using $F_i^{\rm TMC-Leading}$ together with a 2-parameter function $\gamma_a(Q)$. Here, $\gamma_a(Q)$ implicitly depends on the parameters $\{\lambda_a, \delta_a\}$ which are fitted for each structure function $\{F_{1,2,3}\}$ and for selected PDFs displayed in the table. The TMC corrections are accurate to within $\sim 0.3\%$ for $Q >$ 1.3 GeV. The $\{\lambda_i, \delta_i\}$ parameters are independent of the exchanged boson $(\gamma, Z, W^{\pm})$, and are relatively insensitive to the specific underlying nPDF. Once the $\{\lambda_i, \delta_i\}$ parameters are determined for a particular nPDF, the full TMC corrections can be efficiently computed (without convolution integrations) to high precision as illustrated in Fig. 2.16.

Starting from Eq. (2.53), we divide $F_i^{\rm TMC}(x)$ by the leading term $F_i^{\rm Leading-TMC}(x)$:

$$\frac{F_1^{\rm TMC}(x)}{F_1^{\rm Leading-TMC}(x)} = 1 + \frac{M^2}{Q^2}\frac{x\xi}{r}\frac{h_2(\xi)}{F_1^{(0)}(\xi)} + \frac{M^4}{Q^4}\frac{2x^2\xi}{r^2}\frac{g_2(\xi)}{F_1^{(0)}(\xi)}\,, \tag{2.101a}$$

$$\frac{F_2^{\rm TMC}(x)}{F_2^{\rm Leading-TMC}(x)} = 1 + \frac{M^2}{Q^2}\frac{6x\xi^2}{r}\frac{h_2(\xi)}{F_2^{(0)}(\xi)} + \frac{M^4}{Q^4}\frac{12x^2\xi^2}{r^2}\frac{g_2(\xi)}{F_2^{(0)}(\xi)}\,, \tag{2.101b}$$

$$\frac{F_3^{\rm TMC}(x)}{F_3^{\rm Leading-TMC}(x)} = 1 + \frac{M^2}{Q^2}\frac{2\xi x}{r}\frac{h_3(\xi)}{F_3^{(0)}(\xi)}\,. \tag{2.101c}$$

The goal of this section is to parameterize these ratios.

As explained in detail in Appendix B of [48] , if we assume the structure functions vanish at $x = 1$ and expand them in a Taylor series, we can obtain simple expressions for the ratios in terms of the parameter function $\gamma_i(Q)$, which characterizes the first derivative of the structure function. This allows us to obtain the approximate expressions for the ratios $h_i/F_j^{(0)}$ and $g_2/F_j^{(0)}$:

$$\frac{h_2(\xi)}{F_2^{(0)}(\xi)} = \frac{1-\xi}{\xi} + \gamma_2(Q)\frac{1-\xi}{\xi^2}\sum_{j=1}^{j_{\rm max}}\frac{(-1)^j}{j!(j+1)}\,{}_2F_1\left(2, j+1, j+2, 1-\frac{1}{\xi}\right)\,, \tag{2.102a}$$

$$\frac{h_3(\xi)}{F_3^{(0)}(\xi)} = -\ln(\xi) + \gamma_3(Q)\frac{1-\xi}{\xi}\sum_{j=1}^{j_{\rm max}}\frac{(-1)^j}{j!(j+1)}\,{}_2F_1\left(1, j+1, j+2, 1-\frac{1}{\xi}\right)\,, \tag{2.102b}$$

$$\frac{h_2(\xi)}{F_1^{(0)}(\xi)} = 2\xi r^2\left[\frac{1-\xi}{\xi} + \gamma_1(Q)\frac{1-\xi}{\xi^2}\sum_{j=1}^{j_{\rm max}}\frac{(-1)^j}{j!(j+1)}\,{}_2F_1\left(2, j+1, j+2, 1-\frac{1}{\xi}\right)\right]\,, \tag{2.102c}$$

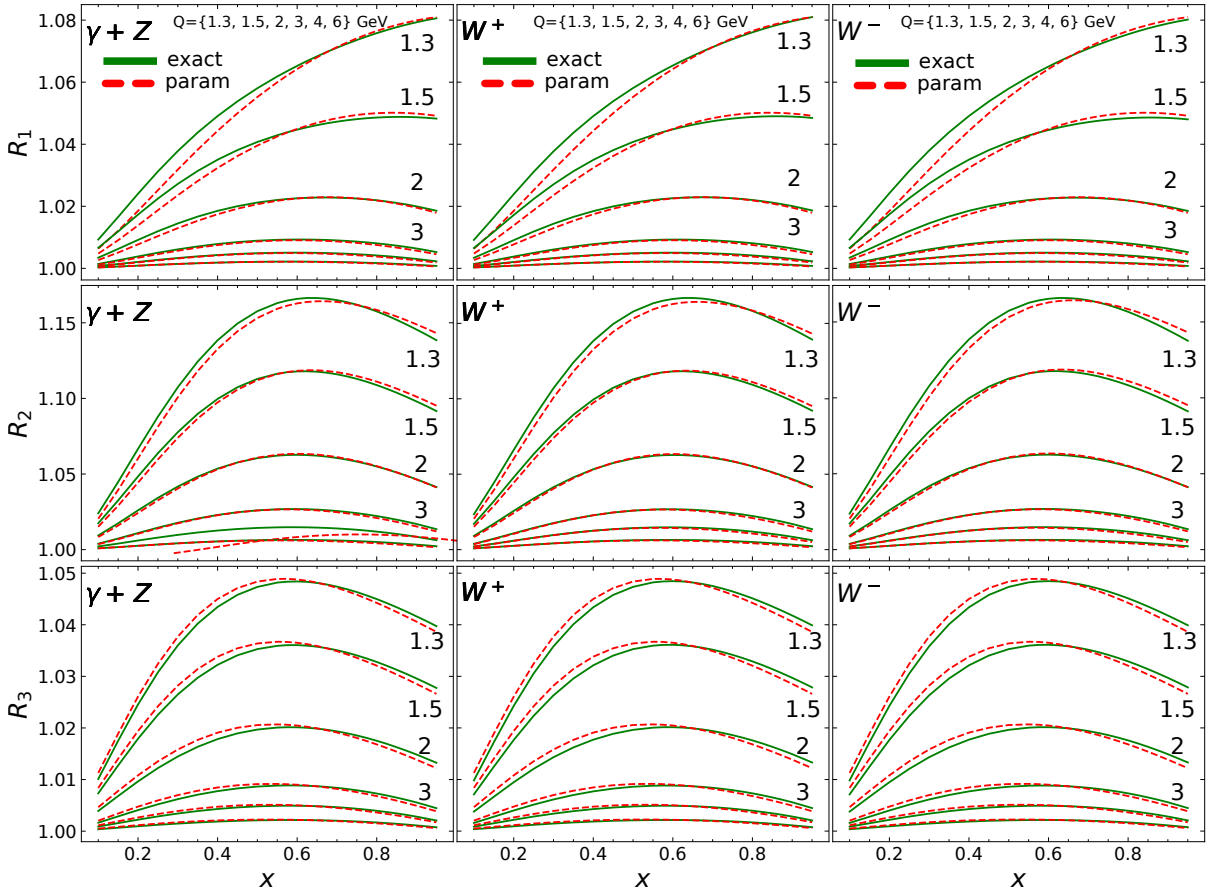


Figure 2.16: We present the nuclear structure function ratio $R_i = F_i^{\rm TMC}/F_i^{\rm TMC-Leading}$ averaged over the nuclei of Table 2.6 as defined in Eq. (2.100). The ratios are shown as a function of x_N for $Q = \{1.3, 1.5, 2, 3, 4, 6\}$ GeV (from top to bottom) using the nCTEQ15 nPDFs. We compare the exact results (green solid) of Eq. (2.53) to the approximate parameterizations (red dashed) based on Eq. (2.102) with parameters given in Table 2.7. For $Q > 2$ GeV, we observe the parameterization can match the exact results to better than 0.2%.

$$\frac{g_2(\xi)}{F_2^{(0)}(\xi)} = -\ln(\xi) - (1-\xi) + \gamma_2(Q)\frac{(1-\xi)^2}{\xi^2}\sum_{j=1}^{j_{\rm max}}\frac{(-1)^j}{j!(j+2)}\,{}_2F_1\left(2, j+2, j+3, 1-\frac{1}{\xi}\right)\,, \tag{2.102d}$$

$$\frac{g_2(\xi)}{F_1^{(0)}(\xi)} = 2\xi r^2\left[-\ln(\xi) - (1-\xi) + \gamma_1(Q)\frac{(1-\xi)^2}{\xi^2}\sum_{j=1}^{j_{\rm max}}\frac{(-1)^j}{j!(j+2)}\,{}_2F_1\left(2, j+2, j+3, 1-\frac{1}{\xi}\right)\right]\,, \tag{2.102e}$$

where ${}_2F_1(a,b,c,z)$ is a hypergeometric function. Although the summation over the index j can, in principle, go to infinity, we truncate the series at $j = j_{\rm max}$. The $1/(j!)$ prefactor of the hypergeometric function helps ensure the series converges quickly, and we find $j_{\rm max} = 4$ yields results that are $\lesssim 1\%$ accuracy.

The function $\gamma_a(Q)$ is given by

$$\gamma_a(Q) = \lambda_a \ln(Q)^{\delta_a}\,. \tag{2.103}$$

The values of $\{\lambda_a, \delta_a\}$ are obtained by fitting the parameterization to the exact results for each $F_i^{\rm TMC}$ structure function. Note that the values of $\{\lambda_a, \delta_a\}$ are independent of the type of the exchanged bosons; they have only a mild dependence on the specific structure function F_i^A as displayed in Table 2.7.

Additionally, while we have computed these results assuming an NLO DGLAP evolution, these results should be similar for an NNLO evolution. This is because: i) we are working at relatively low Q values where the effects of the DGLAP evolution from Q_0 to Q will be minimal, and ii) since this quantity is a ratio of structure functions, the modification of the ratio is expected to be less than the impact on the individual structure functions.

We have also computed the results for a variety of different PDF sets to ensure this parameterization is largely insensitive to the specific PDF. This also is displayed in Table 2.7, where we observe a mild variation of the parameters between PDF sets.

In Figure 2.16, we show the comparison between our parameterization with the exact results obtained using nCTEQ15 nPDFs. We can see that our parameterization works very well to reproduce the exact results. The parameterization has the advantage that it can be computed efficiently whereas the full $F_i^{\rm TMC}$ results require additional convolution integrations for the $\{h_i(\xi), g_i(\xi)\}$ functions. For values of $Q \geq 2$ GeV, which is a typical cut for the global nPDF fits, we see the parameterization describes the full TMC results to better than 0.2%.

Summary: In the previous sections we established several results which will facilitate efficient calculation of TMCs for nuclear structure functions. We briefly summarize the key observations below.

- TMCs can yield large corrections in the double limit of large x and small Q.
- The "Leading-TMC" result provides an excellent approximation to the full "TMC" result as the $h_{2,3}$ contributions are suppressed by additional $(M/Q)^2$, and the g_2 contributions by additional $(M/Q)^4$ powers.
- The nuclear dependence of the "TMC" result is approximately independent of A. This observation suggests we can determine an A-independent "universal" correction factor which can be used to quickly calculate the nuclear structure function TMCs for any nuclei. As illustrated in the bottom panels of Fig. 2.12, this "universal" approximation holds to $\lesssim 0.1\%$ for $Q \geq 2$ GeV.

- As demonstrated in Fig. 2.16, we can parameterize this "universal" correction factor with Eq. (2.102) into a 2-parameter form that can be easily incorporated in a efficient nuclear PDF fitting program. This parameterization matches the exact results to better than 0.2% for $Q \geq 2$ GeV.

Given the "Leading-TMC" result, the above allows us to efficiently approximate the full "TMC" result for any nuclear structure function by applying "universal" correction factor. This avoids the computation of the numerical integrals contained in the $\{h_2, h_3, g_2\}$ functions.

2.7.4 Reduced cross sections with TMCs for nuclei

When extracting structure functions from DIS data or comparing theoretical expectations for DIS to data, one often works with the "reduced cross section" $\sigma_{\text{Red.}}$. These quantities are derived from double differential cross sections scaled by kinematic factors, and provide a direct connection between DIS kinematics and structure functions.

The general cross section for neutral and charged current DIS can be expressed as: [14]

$$\frac{d^2\sigma^i}{dx\,dy} = x(s-M^2)\,\frac{d^2\sigma^i}{dxdQ^2} = \frac{4\pi\alpha^2}{xyQ^2}\,\eta^i \left\{ \left(1 - y - \frac{x^2y^2M^2}{Q^2}\right) F_2^i + \left(y^2 x\right) F_1^i \mp \left(y - \frac{y^2}{2}\right) x F_3^i \right\} , \tag{2.104}$$

where i=NC, CC corresponds to neutral-current or charged-current processes. Additionally, $\eta^{NC} = 1$ and $\eta^{CC} = 4\eta_W$, with $\eta_W = \frac{1}{2}\left[\frac{G_F M_W^2}{4\pi\alpha}\frac{Q^2}{Q^2+M_W^2}\right]^2$. In the last term of Eq. (2.104), the "–" sign is taken for an incoming anti-lepton $\{e^+, \bar{\nu}\}$ and the "+" sign for an incoming lepton $\{e^-, \nu\}$.

Neutral current cross section: We find it convenient to define the reduced cross section by dividing out the leading factors.

$$\frac{d^2\sigma^{\text{NC}}}{dx\,dy} = x(s-M^2)\,\frac{d^2\sigma^{\text{NC}}}{dxdQ^2} = \frac{4\pi\alpha^2}{xyQ^2}\left[\frac{Y_+}{2}\,\sigma_{\text{Red.}}^{NC}\right] , \tag{2.105}$$

$$\sigma_{\text{Red.}}^{NC} = \left(1 + \frac{2y^2\varepsilon^2}{Y_+}\right) F_2^{\text{NC}} \mp \frac{Y_-}{Y_+}\, x F_3^{\text{NC}} - \frac{y^2}{Y_+} F_L^{\text{NC}} , \tag{2.106}$$

with $F_L = r^2F_2 - 2xF_1$, $r = \sqrt{1+4\varepsilon^2}$, $\varepsilon = (xM/Q)$ and $Y_\pm = 1 \pm (1-y)^2$; the r^2 term in F_L is essential to obtain the correct sign for the ε^2 term of Eq. (2.106). Note that at LO

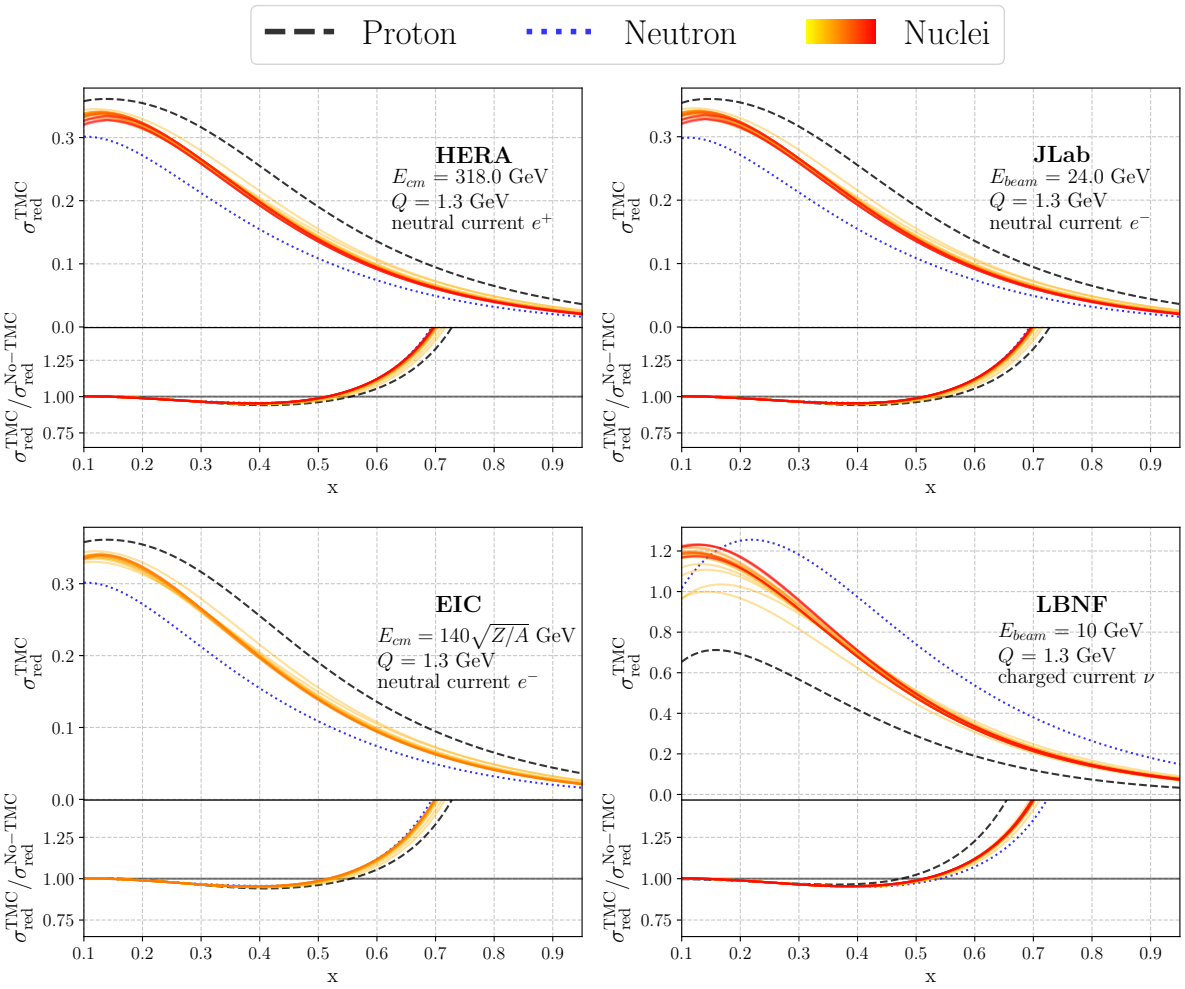


Figure 2.17: We present reduced cross sections (See Eqs. (2.106) and (2.108)) for the following scenarios: (a) HERA inspired e^+p neutral current scattering with E_{cm} = 318 GeV for Q = 1.3 GeV; (b) JLab inspired e^-A charged current scattering with E_{beam} = 24 GeV for Q = 1.3 GeV; (c) EIC inspired e^-A neutral current scattering with $E_{cm} = 140\sqrt{Z/A}$ GeV for Q = 1.3 GeV; (d) LBNF inspired νA charged current scattering with E_{beam} = 10 GeV for Q = 1.3 GeV. **Upper panels:** The upper panels display the full "TMC" results. **Lower panels:** The lower panels display the ratio of the full "TMC" to "No-TMC". The proton is indicated with the black dashed line, and the neutron with the blue dotted line. The narrow band of colored lines shows the various nuclear results.

($F_L \sim 0$) for parity-conserving photon exchange ($F_3 = 0$), so that we have $\sigma_{\text{Red.}}^{NC}$ simply reduces to F_2^{NC}.

Charged current cross section: For the charged current process, we have:

$$\frac{d^2\sigma^{\text{CC}}}{dx\,dy} = x(s-M^2)\frac{d^2\sigma^{\text{CC}}}{dxdQ^2} = \frac{G_F^2}{2\pi x}\frac{Q^2}{y}\left(\frac{M_W^2}{Q^2+M_W^2}\right)^2\left[\sigma_{\text{Red.}}^{CC}\right], \tag{2.107}$$

$$\sigma_{\text{Red.}}^{CC} = \frac{Y_+}{2}\left(1+\frac{2y^2\varepsilon^2}{Y_+}\right)F_2^{\text{CC}} \mp \frac{Y_-}{2}xF_3^{\text{CC}} - \frac{y^2}{2}F_L^{\text{CC}}. \tag{2.108}$$

Here we have followed the HERA convention [160] where the CC differs from NC by a factor of $Y_+/2$.

For both the NC and CC cases, the factor of $2y^2\varepsilon^2/Y_+$ with $\varepsilon = (xM/Q)$ multiplying F_2 arises directly from the DIS kinematics; however, this is often neglected as it is small for typical HERA kinematics.

TMCs for nuclear reduced cross sections

As described in the preceding subsections, TMCs can have a large numerical impact on nuclear structure functions in the low Q^2 and large x double limit. To illustrate the TMC impact, in Fig. 2.17 we present results for experimental scenarios including HERA, JLab, EIC, and LBNF. The upper panels of Fig. 2.17 display the reduced cross sections, and the lower panels show the ratio of the full "TMC" to the "No-TMC" result. As before, all the nuclear isotopes of Table 2.6 are within the colored band, and the proton and neutron are displayed separately.

In Fig. 2.17-(a) we show the HERA-inspired e^+p neutral current σ^{NC}_{Red} at $\sqrt{s} = 318$ GeV at $Q = 1.3$ GeV for the full "TMC" results. Although HERA ran proton beams, we also show the nuclear results which may be useful for a future LHeC/FCCeh facility. The TMCs yield a slight suppression at intermediate x values (~ 0.4), and an enhancement at larger x values ($\gtrsim 0.6$) which can be significant. The behavior of $\sigma^{NC}_{Red.}$ strongly resembles that of F_2 in Fig. 2.10; this is expected as given our choices of Q and $E_{\rm CM}$ we find $y \ll 1$ and $Y_- \sim 0$, so we see from Eq. (2.106) that the F_2 contribution dominates. For the higher Q values (not shown) TMCs are reduced as for F_2 in Fig. 2.10, and are generally negligible except in the very large x region.

In Fig. 2.17-(b) we show a JLab-inspired e^-A neutral-current σ^{NC}_{Red} with $E_{\rm Beam} = 24$ GeV at $Q = 1.3$ GeV. The character of the curves for $E_{\rm Beam} = 12$ GeV (not shown) is quite similar to the displayed results. The results here are qualitatively similar to the F_2 result of Fig. 2.10 and the HERA case above with a reduction for intermediate x (~ 0.4), and an enhancement at larger x.

In Fig. 2.17-(c) we show an EIC-inspired (e^-A) σ^{NC}_{Red} with $E_{cm} = 140\sqrt{Z/A}$ GeV at $Q = 1.3$ GeV. As with the above HERA and JLab cases, these results largely follow the F_2 structure function.

In Fig. 2.17-(d) we show an LBNF-inspired $\nu - A$ charged-current scattering with $E_{\rm beam} = 10$ GeV at $Q = 1.3$ GeV. In the low x region, we see the various nuclear curves separate due to the differing nuclear mass values.

From this exercise, one can see that for large $\sqrt{s}$, we generally have small y, and σ_{Red}

closely follows F_2. For smaller $\sqrt{s}$, we have larger y, and the F_3 and F_L structure functions can now contribute.

2.8 Conclusions

The structure of the proton has been extensively investigated, and recent work has extended the precision to next-to-next-to-leading-order (NNLO) and beyond. With the upcoming Electron Ion Collider (EIC) and other experiments discussed in the introduction, we have the opportunity to build upon the proton analyses and also determine the nuclear structure with extreme precision.

To exploit the current and future nuclear measurements, and to obtain the desired accuracy of theory calculations, it is essential to include TMCs in the analyses of structure functions. However, extending the TMCs from the proton to nuclei requires a rigorous theoretical foundation. This foundation is provided by the OPE. Additionally, as we have shown, the equivalence of the OPE with the parton model (with appropriate M_A and k_T contributions) provides a familiar framework to implement TMC in analyses. Some of the tools from the proton analyses can be applied directly to the nuclear case, while other elements (such as the nuclear DGLAP equations with $x_N \in [0, A]$ and $W_A \neq W_N$) must be adapted. Ultimately, the answer to the question, *Why nucleon-like expressions could possibly be valid for nuclear targets?*, is simply that the consequences of asymptotic freedom, e.g., light-cone dominance and the validity of perturbative methods, begin to emerge when momentum transfers Q surpass the non-perturbative scale $\Lambda_{\rm NP} \lesssim 1$ GeV (see Sec. 2.3.1). While this is (accidentally) comparable to the scale of free nucleons, masses of target hadrons in DIS (albeit nucleons or nuclei) are a sub-leading ingredient and contribute as powers of $\mathcal{O}(x_A^2 M_A^2/Q^2)$. The goal of this presentation is to explicitly delineate the proper ingredients and adaptations required for a precision nuclear analysis.

The present investigation provides both pedagogical material and a practical reference for computing and investigating TMCs for nuclear processes. We briefly outline the key components and results in the following:

- **OPE Formalism:** We extend the OPE formalism to the case of interacting nuclei and explicitly provide the correspondence between TMC expressions for a nucleon and a nucleus, given in Eq. (2.39). To derive these relations (see Sec. 2.3.2 and Appendix A of [48]), we use only fundamental symmetry principles of the nuclei, and make no assumptions about the parton or other models. This is the essence of the factorization between the hadron and parton level processes.

- **TMC Power Expansion:** We investigate analytically and numerically the magnitude of TMCs and show it is the quantity $(x_A M_A/Q) = (x_N M_N/Q)$ that enters the TMC formula for **nuclei** (Eq. (2.38)) and **nucleons** (Eq. (2.53)). This result is important for heavy nuclear A targets such as lead. While M_A can be large ($\sim A\, M_N$), it is always combined with $x_A \lesssim 1/A$ (since an individual parton momentum fraction will typically be less than $1/A$ of the total **nuclear** momentum). Thus, the OPE expansion for TMCs is effectively in powers of (M_N/Q) and not (M_A/Q), so it is equally accurate for both protons and nuclei.

- **Nuclear TMCs:** In Sec. 2.3.3, we extend the proton results of Ref. [104] and derive a set of master equations (Eq. (2.39)) for nuclear structure functions with TMCs that are the same for all A, including the proton, neutron, and nuclear cases. This allows us to easily apply the tools of the OPE to nuclei as well as protons.

- **Nucleon TMCs:** In Sec. 2.4.2, we rescale our **nuclear** structure functions of Eq. (2.39) to construct **nucleon** structure functions in Eq. (2.53). These rescaled structure functions are the ones commonly studied and presented in comparisons between different nuclear data. This facilitates extraction of the *nuclear correction factors* by taking ratios of rescaled structure functions, as illustrated in Fig. 1.8. Additionally, this allows for investigations of collective nuclear effects (*e.g.*, shadowing, anti-shadowing, EMC, Fermi, etc.).

- **The Parton Model:** In Sec. 2.5.1, we extend the rescaled results to the parton model and obtain the corresponding set of DGLAP equations and associated sum rules, including isospin symmetry. We also identify the complication of separately extracting $u(x)$ and $d(x)$ nPDFs. This correspondence allows us (with the reasonable assumption $x_N \leq 1$) to re-use our proton DGLAP evolution code for the case of nucleons.

- **Nuclear Kinematics:** We establish the relations between the kinematic variables including the **nuclear** scaling variables $\{x_A, \xi_A\}$ and the **nucleon** scaling variables $\{x_N, \xi_N\}$. Furthermore, we indicate in Sec. 2.5.4 some subtle complications we encounter such as defining an appropriate W^2 cut for the nucleon case. Specifically, as demonstrated in Fig. 2.7, the dependence of W_A^2 on A and Q^2, complicate the task of placing cuts on the resonance region.

- **The OPE and the Parton Model:** We obtain a complete correspondence between the parton model and OPE structure function results, and demonstrate these are

equivalent **if** we include i) target mass effects and ii) k_T transverse momentum contributions in the parton model. Additionally, in Sec. 2.6.1 we show the utility of the ACOT helicity approach with light-cone kinematics, and demonstrate the correspondence to the traditional tensor basis as shown in Table 2.4.

- **Numerical Comparisons:** In Sec. 2.7, we provide a variety of numerical comparisons to demonstrate the size of the TMCs in various kinematic regions. These comparisons are shown for both structure functions, and the reduced cross section for some realistic machine kinematics. We found that for typical kinematic cuts used in global analyses $\{Q \gtrsim 2 \text{ GeV}, W \gtrsim 1.7 \text{ GeV}, x \lesssim 0.7\}$, the $F_i^{\text{TMC–Leading}}$ was within $\sim 5\%$ or better of the full F_i^{TMC} result, and improves quickly to $\sim 1\%$ or better for $Q \gtrsim 6$ GeV.

- **TMC Computation:** We also present numerically the dependence of the TMCs on the nuclear A value and the exchanged vector boson $\{\gamma, Z, W^{\pm}\}$. We find the relative magnitude of the TMCs is weakly dependent on the exchanged boson, and **very** weakly dependent on the nuclear A. This latter observation allows us to derive an approximate (and computationally expedient) parameterization for the nuclear TMCs suitable for including inside a global fit function (see Sec. 2.7.3). For $Q > 2$ GeV, the parameterized result matches the exact results to better than 0.2%, which is well below the typical nPDF precision.

In the near future, there will be ambitious new facilities and experiments that will investigate lepton-nucleus and neutrino-nucleus interactions in extreme kinematic regions with high precision. Just as Ref. [104] provided the details for the proton case, this report offers foundational and reference material that can facilitate incisive investigations of nuclear phenomena.

We can use the tools from the precision proton analyses as an archetype, and extend them into the nuclear dimension. Furthermore, the nuclear interactions open up a whole new class of phenomena (saturation, shadowing) and new states of matter (quark-gluon plasma, Color Glass Condensates) to explore. Discovering and analyzing these phenomena will require precision on both the experimental and theoretical fronts; a rigorous foundation for theoretical calculations of structure functions with TMCs is an essential step toward this goal.

Chapter 3

Inclusive prompt photon production

Introduction

The production of high-transverse-momentum prompt photons in hadronic collisions, despite using an electroweak probe, is a key process in QCD. Since the 1970s, both theoretical and experimental studies have investigated this channel due to its fundamental importance [161–180]. At leading order (LO), there are only two contributing subprocesses (the QCD-Compton process and the $q\bar{q}$ annihilation), with a direct point-like coupling of the photon to the quarks. Moreover, since the dominant channel is $qg \longrightarrow q\gamma$, prompt photons provide direct access to the gluon parton distribution inside the hadron and to α_s. Historically, the relatively small electromagnetic coupling ($\alpha_{em} \simeq 1/137 = 0.0072$) led to small cross-sections, making precise measurements challenging. At the LHC, however, even sub-nanobarn cross-sections are measurable with high precision. An additional motivation for the extensive study of this process is to verify whether the conditions for factorization are satisfied and allow us to match the experimental results. Discrepancies between experimental prompt photon data and theoretical predictions have often been reported in the literature. These have led to several efforts and alternative approaches beyond the *collinear* factorization theorem, such as accounting for k_T-effects of the incoming partons [181], or using Soft-Collinear Effective Theory (SCET) [182]. In this work, we investigate how inclusive prompt photon production can constrain the gluon PDF, in the framework of the QCD-improved parton model. Specifically, we aim to constrain the gluon PDF at low momentum fraction x in the context of a global analysis of nuclear PDFs within the nCTEQ framework.

This chapter is organized as follows: Sec. 3.1 presents the theoretical framework of prompt photon production, beginning with a definition of the process and the formalism for its theoretical description in Sec. 3.1.1 and continuing with an exploration of the

kinematics for this process in Sec 3.1.2. A general presentation of the prompt photon cross-section is provided in Sec. 3.1.3. Next, the Born level is discussed in Sec. 3.1.4, followed by the role of fragmentation in the process in Sec. 3.1.5. The inclusion of higher-order terms is considered in Sec. 3.1.6, before concluding with a summary of the theoretical considerations in Sec. 3.1.7.

Next, the Sec. 3.2 focuses on available theoretical computations, starting with an overview of the JETPHOX code in Sec. 3.2.1. A presentation of the code is done in Sec. 3.2.1.1, while its technical features are presented in Sec. 3.2.1.2, and the calculation of observables is detailed in Sec. 3.2.1.3. This is followed by a description of MCFM in Sec. 3.2.2, discussing its generalities in Sec. 3.2.2.1 and the evolution across different versions is outlined in Sec. 3.2.2.2. Other relevant calculations available in the literature are summarized in Sec. 3.2.3.

Finally, in Sec. 3.3, we validate the theoretical calculations against experimental data, first for proton-proton collisions using JETPHOX and MCFM in Sec. 3.3.1, and then for proton-nucleus collisions using JETPHOX in Sec. 3.3.2.

Lastly, Sec. 3.4 provides concluding remarks that summarize the chapter and its findings.

3.1 Theoretical framework

3.1.1 Definitions and formalism of the prompt photon production process

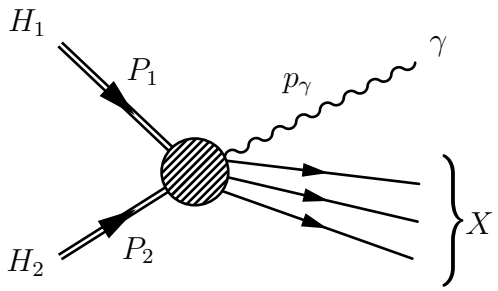


Figure 3.1: Graphical representation of a collision of two hadrons H_1 and H_2 with 4-momenta P_1 and P_2, producing one high-p_T photon (γ) with 4-momentum p_γ plus an unspecified hadronic state (X) in the final state.

We consider the inclusive production of a single photon in a hadronic collision:

$$H_1(P_1) + H_2(P_2) \longrightarrow \gamma(p_\gamma) + X\,, \tag{3.1}$$

where H_1 and H_2 are the two incoming hadrons with momenta P_1 and P_2 respectively, γ a high-p_T photon with momentum p_γ and X denotes all other final state particles which are not observed. We work with photons produced at *large* transverse momentum p_T such that the strong coupling constant α_s evaluated at an energy scale ($\mu \sim p_T$) is sufficiently small for using perturbative techniques, thanks to asymptotic freedom [173]. To begin with, it is important to properly define the *photon* under consideration, both theoretically and experimentally. The terminology can vary in the literature depending on the time period, the physics community (e.g. heavy-ion or pQCD) and most importantly, the production mechanism involved. Experimentally, hadronic collisions produce photons in several ways. They may originate: i) directly from the hard partonic interaction, ii) from fragmentation of a parton, iii) from the decay of a hadron (e.g. $\pi^0 \to \gamma\gamma$ which is the most important background), or iv) from thermal radiation (e.g. from the Quark-Gluon-Plasma (QGP), if it is present).

Trying to be as unambiguous as possible, we will use the following definitions (partly based on [183]):

◦ **Theoretically**

- **Direct photons** are photons produced directly in the hard subprocesses (at partonic level), well before the hadronization. At Born level, the direct contribution formally consists only of two subprocesses: the QCD-Compton mechanism $qg \longrightarrow q\gamma$ and the quark-antiquark annihilation $q\bar{q} \longrightarrow \gamma g$. For a Feynman diagram representation of direct photon production at LO, see Fig. 3.5. This contribution is fully described in Sec. 3.1.4. This excludes photons from hadron decays and photons from parton fragmentation.

- **Photons from fragmentation** are photons emitted quasi-collinearly by a high-p_T parton during the hard subprocess. These photons have often been referred to as *bremsstrahlung* or even as *anomalous* contribution in older literature. This contribution is fully described in Sec. 3.1.5.

- Decay photons are photon produced from the decay of hadrons (e.g., $\pi^0 \to \gamma\gamma$). They do not originate from the hard partonic subprocess and are therefore excluded from the definition of prompt photons. Decay photons constitute a dominant background in inclusive photon measurements and are typically reduced using photon isolation techniques (See definitions below and Sec. 3.1.3).

- **Prompt photons**, in the context of this work, are defined as the sum of **direct** and **fragmentation** photons. They can be **isolated** (see below) or not.

- *Inclusive* refers to processes in which a single photon is observed, while summing inclusively over all other possible final state particles (X).

◦ **Experimentally**

- **Prompt photons** are the detected photons, excluding those originating from hadron decays (e.g., $\pi^0 \to \gamma\gamma$).
- **Direct photons** is a term also used in experimental papers (as in [184] for the ALICE experiment), where it usually refers to **prompt photons** as defined in the first point above.
- Although the signature of a photon can be easily detected experimentally, it is quite difficult to distinguish its origin (e.g., do they come from the hard collision or the decay of hadrons?). **Isolated photons** are the photons that remain after applying an isolation criterion, which consists of excluding most of the hadronic activity around/near the photon, in order to exclude decay photons. Note that, it is *impossible* to experimentally distinguish direct and fragmentation photons, since their separation does not correspond to a measurable observable and is theoretically scheme-dependent. That being said, the application of isolation criteria has the effect of significantly reducing the fragmentation contribution [185, 186].

Since we rely on both experimental results and theoretical predictions to determine PDFs, it is necessary to carefully take into account both, experimental constraints and theoretical considerations, in the calculations.

3.1.2 Kinematics

We begin by defining the useful quantities that will be used throughout the chapter. As can be seen in Fig. 3.2, P_1 and P_2 are the 4-momenta of the incoming hadrons H_1 and H_2. Choosing the z-axis as the beam axis, neglecting hadron masses, the 4-momenta can be written as follows:

$$P_1 = (E_1, 0, 0, E_1)\,, \tag{3.2}$$

$$P_2 = (E_2, 0, 0, -E_2)\,, \tag{3.3}$$

$$p_\gamma = (E_\gamma, \vec{p}_T^{\,\gamma}, E_\gamma \cos\theta)\,, \tag{3.4}$$

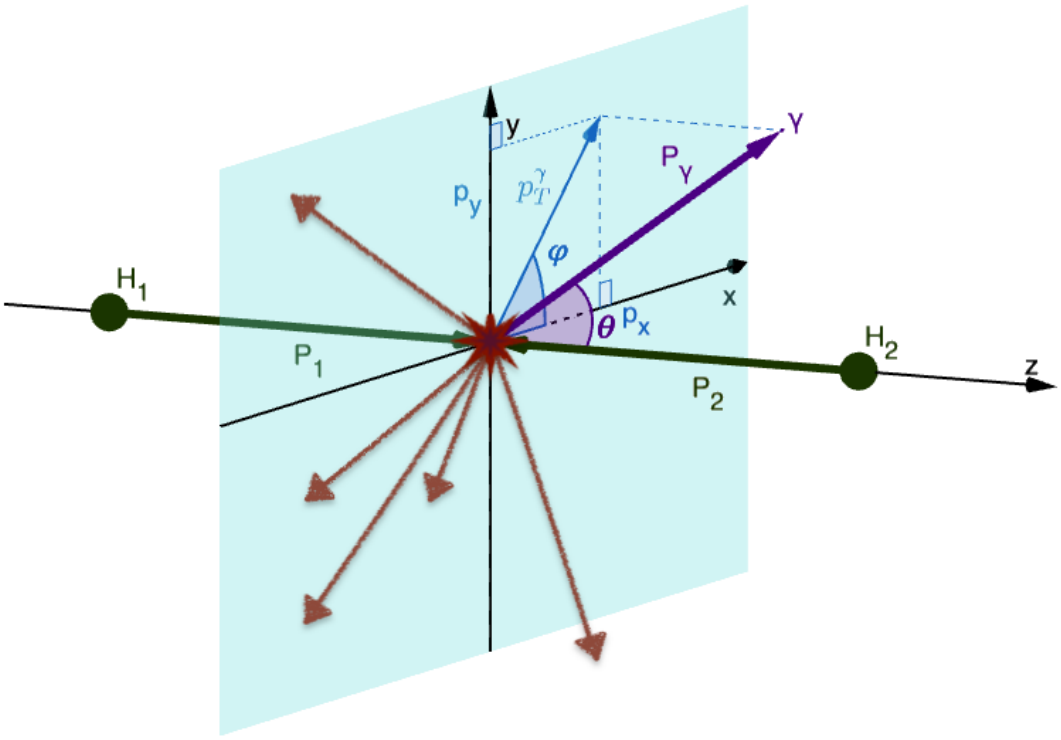


Figure 3.2: This graph shows all the kinematic variables used in the following section. H_1 and H_2 are the two incoming hadrons with momentum P_1 and P_2, respectively. The photon, shown in violet, has a momentum p_γ, and the polar angle θ is formed between the beam axis z and the γ direction. The transverse momentum p_T^γ is the projection of $\vec{p}_\gamma$ onto the transversal plane (x,y), shown in blue, where φ is the azimuthal angle.

with $E_\gamma = p_\gamma^0 = \mid \vec{p}_\gamma \mid$ and $\vec{p}_T^\gamma = (p_x, p_y)$. We also define the photon rapidity y_γ:

$$y_\gamma = \frac{1}{2} \ln \left(\frac{p_\gamma^0 + p_\gamma^z}{p_\gamma^0 - p_\gamma^z} \right) , \tag{3.5}$$

and the transverse momentum of the photon

$$p_T^\gamma = \sqrt{p_x^2 + p_y^2} \, . \tag{3.6}$$

In terms of the rapidity and the transverse momentum, one can easily see that[1]

$$p_\gamma = (p_T^\gamma \cosh y_\gamma, \vec{p}_T^\gamma, p_T^\gamma \sinh y_\gamma) \, . \tag{3.7}$$

We also define a widely used experimental quantity: the pseudorapidity of the photon,

$$\eta^\gamma = -\ln \left(\tan \frac{\theta}{2} \right) = \frac{1}{2} \ln \left(\frac{|\vec{p_\gamma}| + p_\gamma^z}{|\vec{p_\gamma}| - p_\gamma^z} \right) , \tag{3.8}$$

[1]This can be seen by using $p_\gamma^z = |\vec{p_\gamma}| \cos(\theta)$ and the identity $\tan\left(\frac{\theta}{2}\right) = \sqrt{\frac{1-\cos\theta}{1+\cos\theta}}$.

where θ is the polar angle between the beam axis (z) and the photon (γ), see Fig. 3.2. For a massless particle such as the photon, $p_\gamma^0 = |\,\overrightarrow{p_\gamma}\,|$ and thus Eq. (3.5) and Eq. (3.8) are equal.

Note that in the following we will suppress the index γ using p_T, y, η instead of $p_T^\gamma, y_\gamma, \eta^\gamma$, for the sake of simplicity. Finally it is customary to define the associated hadronic Mandelstam variables

$$s = (P_1 + P_2)^2 =^* 2P_1 \cdot P_2\,, \tag{3.9a}$$

$$u = (P_2 - p_\gamma)^2 =^* -2P_2 \cdot p_\gamma\,, \qquad (^* \text{ neglecting masses}) \tag{3.9b}$$

$$t = (P_1 - p_\gamma)^2 =^* -2P_1 \cdot p_\gamma\,. \tag{3.9c}$$

In the hadronic center of mass system, we have[2] $E_1 = E_2 = \frac{\sqrt{s}}{2}$.

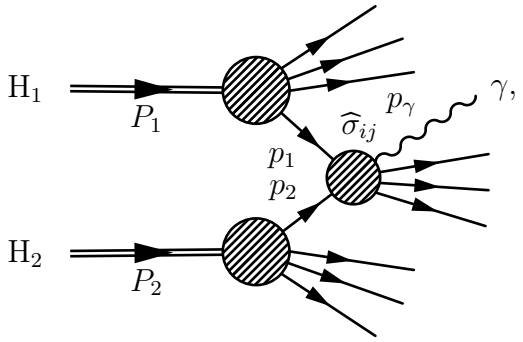


Figure 3.3: Hadro-production of a single high-p_T photon (γ) with 4-momentum p_γ plus an unspecified hadronic state (X) in the final state, within the framework of the Parton Model. The two incoming hadrons H_1 and H_2 have 4-momenta P_1 and P_2, and their associated partonic momentum p_1 and p_2. The partonic cross-section $\widehat{\sigma}_{ij}$, with i, j the flavor of the two incoming partons, is represented by the shaded circle.

Within the framework of the Parton Model (see Fig. 3.3), we have

$$p_1 = x_1 P_1, \qquad\qquad p_2 = x_2 P_2, \tag{3.10}$$

with momentum fractions $0 < x_1, x_2 < 1$. Combining Eq. (3.9) and Eq. (3.10), the partonic Mandelstam variables, for a $2 \to 2$-process, are then defined as

$$\widehat{s} \;=\; (p_1 + p_2)^2 = (p_\gamma + p_4)^2 = x_1 x_2 s \geq 4p_T^2, \tag{3.11}$$

$$\widehat{u} \;=\; (p_1 - p_4)^2 = (p_2 - p_\gamma)^2 = -2p_2 \cdot p_\gamma = -x_2\sqrt{s}\, p_T\, e^{y}, \tag{3.12}$$

$$\widehat{t} \;=\; (p_2 - p_4)^2 = (p_1 - p_\gamma)^2 = -2p_1 \cdot p_\gamma = -x_1\sqrt{s}\, p_T\, e^{-y}, \tag{3.13}$$

[2]Note that for pp collisions, the lab frame coincides with the center-of-mass frame. This is not the case in pA collisions.

where p_γ is the momentum of the final photon, p_T its transverse momentum and p_4 the momentum of the other outgoing parton. In the hadronic center of mass (c.o.m) frame, the parton momenta can be written as

$$p_1 = \frac{\sqrt{s}}{2}(x_1, 0, 0, x_1), \tag{3.14}$$

$$p_2 = \frac{\sqrt{s}}{2}(x_2, 0, 0, -x_2). \tag{3.15}$$

Another important relation arises from momentum conservation in the partonic cross-section at LO (see next section). We obtain a connection between x, p_T and η

$$x_1 \propto \frac{p_T}{\sqrt{s}} e^{\eta} \quad , \qquad x_2 \propto \frac{p_T}{\sqrt{s}} e^{-\eta}. \tag{3.16}$$

This relation shows how to reach the small-x region; either by increasing the pseudorapidity η (in the backward region) and $\sqrt{s}$ and/or decreasing p_T. In order to achieve small x values, we need to attain a balance between p_T small enough to reduce x but still large enough (a crucial requirement) to use perturbative techniques, $p_T \gg \Lambda_{QCD}$. Similarly for η, we have to work within experimental limits, as the energy of the center of mass we can reach and also the geometry of the detectors. As an example, in *p-Pb* collisions studied by ATLAS [187], the smallest value reached is *only* $x_2 \simeq 10^{-2}$. More recent ALICE measurements in *p-Pb* collisions (at the same c.o.m energy as in ATLAS) reports values as low as $x \simeq 2.9 \times 10^{-3}$ [188], which is a slight improvement over the ATLAS configuration. However, we are interested in probing even smaller x values in the future. To this end, the FoCaL [189] detector is currently being constructed by the ALICE collaboration. The detector has been designed with the initial objective of reaching these small x values (down to $\sim 10^{-6}$) with the inclusive photon production and jet production processes, through the implementation of an exceptionally high-rapidity coverage ($3.4 < \eta < 5.8$).

3.1.3 Isolated cross section

As already mentioned in Chap. 1, processes involving large momentum transfer imply a sufficiently small value of α_s to justify the use of perturbative techniques [173].

There is no absolute consensus in the literature regarding the inclusion of additional phenomena beyond the 'usual' factorization theorem used for this process. Therefore, we adopt the following widely accepted framework: the prompt photon production is considered as the sum of two contributions; the direct part (denoted DIR in the following)

and the fragmentation part (FRAG), according to the following formula[3]

$$\sigma^{\gamma} = \sigma^{(\mathrm{DIR})}(\mu_R, \mu_{Fact}, \mu_{Frag}) + \sum_{k=q,\bar{q},g} \sigma^{(\mathrm{FRAG})}(\mu_R, \mu_{Fact}, \mu_{Frag}) \otimes D_{k\to\gamma}(\mu_{Frag}) \,, \tag{3.17}$$

where μ_R is the renormalization scale, μ_F the initial-state factorization scale and μ_{Frag} the fragmentation scale. The function $D_{k\to\gamma}(\mu_{Frag})$ is the fragmentation function of a parton k to a photon γ [190]. Both contributions have been known up to NLO for several decades [174–177, 191–194]. The direct part has also been computed at NNLO in recent years [195, 196] and more recently, NNLO results for both contributions have become available [197–199]. Since our objective is to implement this process in a full NLO global analysis, we will restrict the cross-section to NLO.

A key aspect in this process is the distinction between prompt photons and those originating from other sources. Experimentally, this is achieved by applying an isolation criterion in the calorimeter. A cone is defined by

$$(\eta - \eta_\gamma)^2 + (\phi - \phi_\gamma)^2 \leq R^2 \tag{3.18}$$

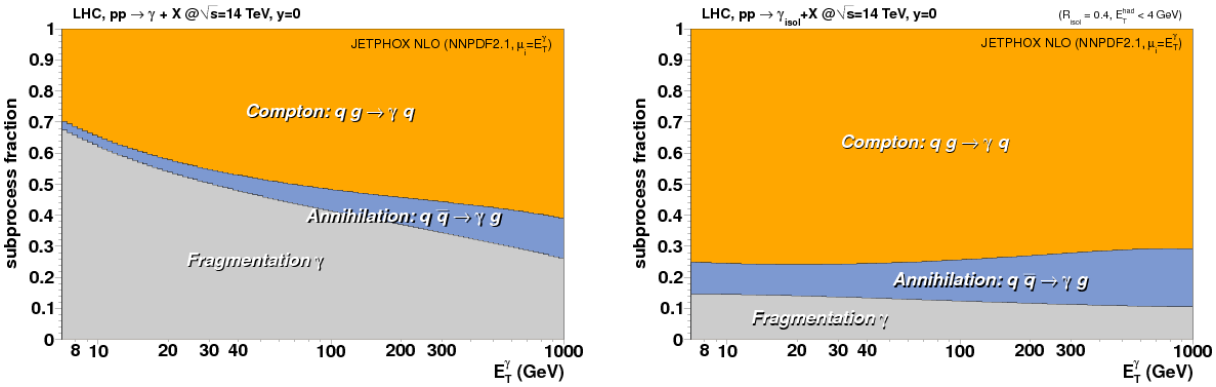
around the photon, within which the hadronic transverse energy must remain below a specified threshold:

$$E_{T(had)} \leq E_{T(max)} \,. \tag{3.19}$$

Here, R is a fixed value and $E_{T(max)}$ can either be a constant or a fraction (ϵ_h) of the photon's transverse momentum (p_T) and both are experiment-specific. This type of isolation criterion is also very convenient from a theoretical point of view, as it can be implemented at the partonic level [200]. However, it does affect the calculation of the cross-section. Indeed, as a consequence of the restricted phase space imposed by this criterion, the process is no longer strictly inclusive. This may introduce a limitation to the applicability of the factorization theorem, particularly concerning the cancellation of singularities. This issue has been discussed in [178, 201], and it has been demonstrated that the factorization remains valid [200], and that the final cross-section is finite at all orders in perturbation theory, as long as R and $E_{T(max)}$ are different from 0 [202].

Another significant consequence of applying an isolation criterion is the substantial reduction of the fragmentation contribution as illustrated in Fig. 3.4. However, it is not completely canceled out. This partial suppression sometimes leads to an underestimation

[3]Here, the photon is not yet isolated.



(a) **Non-isolated** (b) **Isolated**

Figure 3.4: Figure a) shows the ratio of the different contributions involved in prompt photon production; Compton QCD, annihilation and fragmentation. This simulation has been made with JETPHOX for a pp-collision at 14 TeV at central rapidity (η=0). Figure b) is the same simulation but with an isolation criterion applied. The consequence of the isolation is a modification of the subprocess fraction; the fragmentation one is considerably reduced, even at small x. Both plots were taken from Ref. [186].

or even neglect of the fragmentation contribution in the literature. It is important to emphasize that the separation between direct and fragmentation contributions is not absolute. Rather, it depends on the chosen factorization scale and the factorization scheme, which define how these contributions are partitioned. And then, since both the direct and fragmentation parts depend on the μ_{Frag}, only their sum leads to theoretically consistent results. This ensures the cancellation of the μ_{Frag} dependence in the final physical observable.

More specifically, during this thesis we will work with the differential cross-section in transverse momentum p_T and η

$$\frac{d\sigma}{d\overrightarrow{p_T}d\eta} = \frac{d\sigma^{\mathbf{DIR}}}{d\overrightarrow{p_T}d\eta} + \frac{d\sigma^{\mathbf{FRAG}}}{d\overrightarrow{p_T}d\eta}\,, \tag{3.20}$$

where

$$\frac{d\sigma^{\mathbf{DIR}}}{d\overrightarrow{p_T}d\eta} = \sum_{i,j=q,\bar{q},g} \int dx_1 dx_2\; F_{i/H_1}(x_1,\mu_{Fact})\; F_{j/H_2}(x_2,\mu_{Fact})\; \frac{\alpha_s(\mu_R)}{2\pi} \times \left(\frac{d\widehat{\sigma}_{ij}}{d\overrightarrow{p_T}d\eta} + \frac{\alpha_s(\mu_R)}{2\pi} K_{ij}^{\mathbf{DIR}}(\mu_R,\mu_{Fact},\mu_{Frag}) \right)\,, \tag{3.21}$$

and

$$\frac{d\sigma^{\textbf{FRAG}}}{d\overrightarrow{p_T}d\eta} = \sum_{i,j,k=q,\bar{q},g} \int dx_1 dx_2 \, \frac{dz}{z^2} F_{i/H_1}(x_1, \mu_{Fact}) \, F_{j/H_2}(x_2, \mu_{Fact}) \, D_{\gamma/k}(z, \mu_{Frag}) \left(\frac{\alpha_s(\mu_R)}{2\pi}\right)^2 \times \left(\frac{d\widehat{\sigma}_{ij\to k}}{d\overrightarrow{p_T}d\eta} + \frac{\alpha_s(\mu_R)}{2\pi} K^{\textbf{FRAG}}_{ij,k}(\mu_R, \mu_{Fact}, \mu_{Frag})\right). \quad (3.22)$$

F_{i/H_1} and F_{j/H_2} are respectively the PDF of parton i in hadron H_1 and the PDF of parton j in hadron H_2. The function $D_{\gamma/k}(z, \mu_{Frag})$ is the fragmentation function describing the probability for a parton k to fragment into a γ carrying a momentum fraction z, such that $p_T = z\, p_{T,k}$ evaluated at the fragmentation scale μ_{Frag}. $\widehat{\sigma}_{ij\to k}$ are the partonic cross-section and the sum $\sum_{i,j,(k)}$ runs over all possible subprocesses involving the initial-state partons and, in the case of fragmentation, the final-state parton k. α_s is the strong coupling, evaluated at the scale μ_R. The terms $K^{\textbf{DIR},(\textbf{FRAG})}_{ij,(k)}$ represent the NLO corrections to the DIR (FRAG) contributions, respectively [172, 174, 177] ([194, 203]). Here the $\overline{\text{MS}}$ scheme is used.

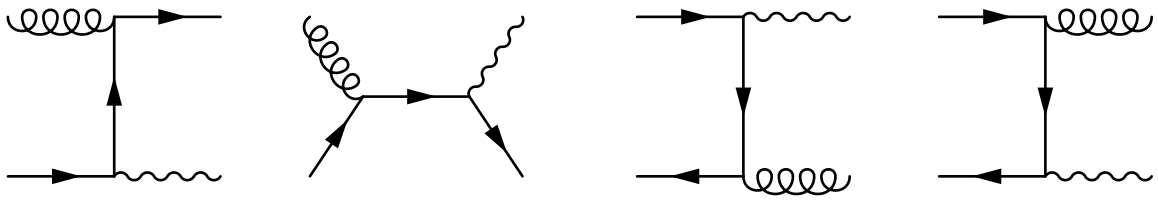
In the following, we describe the process at LO for both the direct and fragmentation contributions, and we discuss the implications of the corresponding perturbative calculations.

3.1.4 Born Level

At LO, the inclusive hadronic cross-section can be expressed by:

$$p_\gamma^0 \frac{d\sigma}{dp_\gamma^3} \overset{4}{=} \frac{d\sigma}{dp_T^2 dy} = \sum_{i,j} \int dx_1 dx_2 F_{i/H_1}(x_1, \mu_{Fact}) F_{j/H_2}(x_2, \mu_{Fact}) \frac{d\widehat{\sigma}^{ij}}{dp_T^2 dy}, \quad (3.23)$$

where $F_{i,j/H_{1,2}}$ are the PDFs, μ_F is the factorization scale, which is basically taken equal to the order of p_T and $\frac{d\widehat{\sigma}^{ij}}{dp_T^2 dy}$ is the differential partonic cross section.



a) QCD-Compton scattering. b) $q\bar{q}$-annihilation.

Figure 3.5: LO Feynman Diagrams for direct photon production.

The only two subprocesses at LO are the QCD-Compton (depicted in Fig. 3.5a) and

[4] Using the azimuthal symmetry: $\frac{dp_\gamma^3}{p_\gamma^0} = dp_T^2\, dy = 2\pi\, p_T\, dp_T\, dy$.

the $q\bar{q}$-annihilation (depicted in Fig. 3.5b). The partonic cross section, which is equivalent to the one given in equation (3.21), $\frac{d\widehat{\sigma}_{ij}}{d\overrightarrow{p_T}d\eta}$, is

$$\begin{aligned}\frac{d\widehat{\sigma}^{ij}}{dp_T^2 dy} &= \frac{1}{2\widehat{s}}\frac{1}{2}\frac{1}{(2\pi)^2}\int\frac{d^3p}{2p^0}\delta^{(4)}(p_1+p_2-p_\gamma-p)\overline{\sum}|\mathcal{M}|^2_{ij} \\ &= \frac{1}{\widehat{s}}\delta(\widehat{s}+\widehat{t}+\widehat{u})\frac{1}{(4\pi)^2}\overline{\sum}|\mathcal{M}|^2_{ij}\ ,\end{aligned} \tag{3.24}$$

with

$$\delta(\widehat{s}+\widehat{t}+\widehat{u}) = \frac{1}{s}\delta(x_1x_2 - x_1x_Te^{-y^\gamma} - x_2x_Te^{y^\gamma})\ , \tag{3.25}$$

and $\overline{\sum}|\mathcal{M}|^2_{ij}$ is the sum of matrix elements corresponding to the Feynman diagrams in Fig. 3.5. The step from Eq. (3.24) to (3.25) for the delta function is valid only for massless partons in a 2-body scattering. The corresponding matrix elements are

$$\begin{aligned}\frac{1}{(4\pi)^2}\overline{\sum}|\mathcal{M}|^2_{q\bar{q}} &= e_q^2\,\alpha\,\alpha_s(\mu_R)\,2\,\frac{c_F}{N}\left(\frac{\widehat{t}}{\widehat{u}}+\frac{\widehat{u}}{\widehat{t}}\right) \\ &= e_q^2\,\alpha\,\alpha_s(\mu_R)\,\frac{8}{9}\left(\frac{x_1e^{-y}-x_T}{x_T}+\frac{x_T}{x_1e^{-y}-x_T}\right)\ ,\end{aligned} \tag{3.26}$$

and

$$\begin{aligned}\frac{1}{(4\pi)^2}\overline{\sum}|\mathcal{M}|^2_{g(x_1)q(x_2)} &= e_q^2\,\alpha\,\alpha_s(\mu_R)\,\frac{1}{N}\left(\frac{\widehat{s}}{-\widehat{u}}+\frac{-\widehat{u}}{\widehat{s}}\right) \\ &= e_q^2\,\alpha\,\alpha_s(\mu_R)\,\frac{1}{3}\left(\frac{x_1e^{y}}{x_T}+\frac{x_T}{x_1e^{y}}\right)\ ,\end{aligned} \tag{3.27}$$

where μ_R is the renormalization scale. Substituting Eq. (3.23) and Eq. (3.25) in Eq. (3.24), the inclusive cross-section is done by

$$\frac{d\sigma}{dp_T^2dy} = \frac{1}{s(-t)}\sum_{i,j}\int_{x_1^{min}}^1\frac{dx_1}{x_1^2}F_{i/H_1}(x_1,\mu_{Fact})F_{j/H_2}(x_2,\mu_{Fact})\frac{1}{(4\pi)^2}\overline{\sum}|\mathcal{M}|^2_{ij}\ , \tag{3.28}$$

where

$$x_2 = \frac{x_1x_Te^{-y}}{x_1-x_Te^{y}}\,. \tag{3.29}$$

The limit conditions are then

$$x_2 \le 1 \quad\Rightarrow\quad x_1^{\min} = \frac{x_Te^{y}}{1-x_Te^{-y}}\ , \tag{3.30}$$

$$x_1^{\min} \le 1 \quad\Rightarrow\quad y_{\text{ext}} = \ln\frac{1}{2}\left(\frac{1}{x_T}\pm\sqrt{\frac{1}{x_T^2}-4}\right). \tag{3.31}$$

As an example, if we consider an $x_T \simeq 10^{-2}$, which is the order of magnitude currently

achieved at the LHC, $y_{\text{ext}} \simeq \pm 4.6$.

Finally, when we include Eq.3.27 and Eq.3.26 in Eq.3.24 and then in Eq.3.23, the differential cross-section can be expressed as

$$\frac{d\sigma}{dp_T^2 dy} = \alpha\alpha_s(\mu_R)\frac{e^{-y}}{s^2 x_T}\sum_q e_q^2 \int \frac{dx_1}{x_1^2}\left\{F_{q/H_1}(x_1,\mu_{Fact})F_{\bar{q}/H_2}(x_2,\mu_{Fact})\frac{8}{9}\left(\frac{x_1 e^{-y}-x_T}{x_T}+\frac{x_T}{x_1 e^{-y}-x_T}\right)\right.$$
$$+ F_{g/H_1}(x_1,\mu_{Fact})\frac{1}{3}\left(F_{q/H_2}(x_2,p_T^2)+F_{\bar{q}/H_2}(x_2,\mu_{Fact})\right)\left(\frac{x_1 e^{y}}{x_T}+\frac{x_T}{x_1 e^{y}}\right)$$
$$\left. + [H_1, x_1 \leftrightarrow H_2, x_2]\right\} . \tag{3.32}$$

One important remark is that the LO expression for prompt photon production is proportional to α_s, in contrast to DIS, where α_s appears only at NLO. This feature implies that prompt photon production can also potentially provide an additional constraint on the strong coupling.

3.1.5 Fragmentation

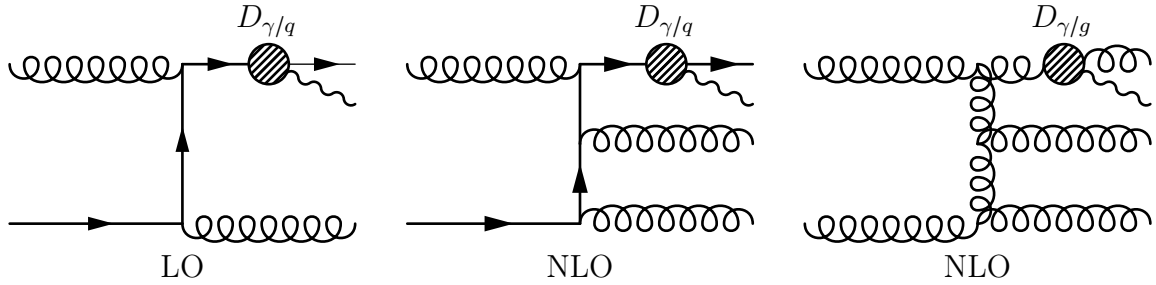


Figure 3.6: A few examples of fragmentation diagrams at LO and NLO.

From a topological point of view, a direct photon is more likely to be produced separately from the hadronic environment, whereas fragmentation photons are more likely to be accompanied by hadrons, except when these photons carry the major part of the momentum of the hard parton from which they originate [185]. The fragmentation function $D_{\gamma/i}$ follows the inhomogeneous DGLAP evolution equation [180]

$$\frac{dD_{\gamma/i}(\mu_{Frag})}{d\ln(\mu^2_{Frag})} = \mathbf{P}_{\gamma i} + \frac{\alpha_s(\mu_{Frag})}{2\pi}\sum_{j=q,g} P_{ji} \otimes D_{\gamma/j} \ , \tag{3.33}$$

where $\mathbf{P}_{\gamma i}$ and P_{ji} are Altarelli-Parisi splitting functions, which have a perturbative expansion in α_s.

The LO approximation of Eq. (3.22) turns to

$$\frac{d\sigma^{\mathbf{FRAG}}}{dp_T d\eta} = \frac{1}{8\pi s^2} \sum_{i,j,a,b} \int \frac{F_{i/H_1}(x_1, \mu_{Fact})}{x_1} \frac{F_{j/H_2}(x_2, \mu_{Fact})}{x_2} \times \left|\overline{\mathcal{M}}_{Born}[i+j \to a+b]\right|^2 D_a^{\gamma}(z, \mu_{Frag}) \times \Theta(z - z_{min}) dz d\eta_b d\eta_a p_{T_a} dp_{T_a}, \quad (3.34)$$

where a is the parent parton of the photon (with rapidity η_a) and b the second parton (with rapidity η_b) produced in the $2 \to 2$ partonic process and it implies

$$p_T^{\gamma} = z p_{T_a} \qquad \text{and} \qquad \eta_{\gamma} = \eta_a \,, \quad (3.35)$$

where z is the momentum fraction of parton a carried by the fragmented γ. In the hadronic center of mass, we have

$$x_{1,2} = \frac{p_{T_a}}{\sqrt{s}} \left(e^{\pm\eta_{\gamma}} + e^{\pm\eta_b}\right) \,, \quad (3.36)$$

and

$$z_{min} = \max\left\{\frac{p_T^{\gamma}}{\sqrt{s}} \left(e^{\pm\eta_{\gamma}} + e^{\pm\eta_b}\right)\right\} \,. \quad (3.37)$$

3.1.6 Higher-orders terms

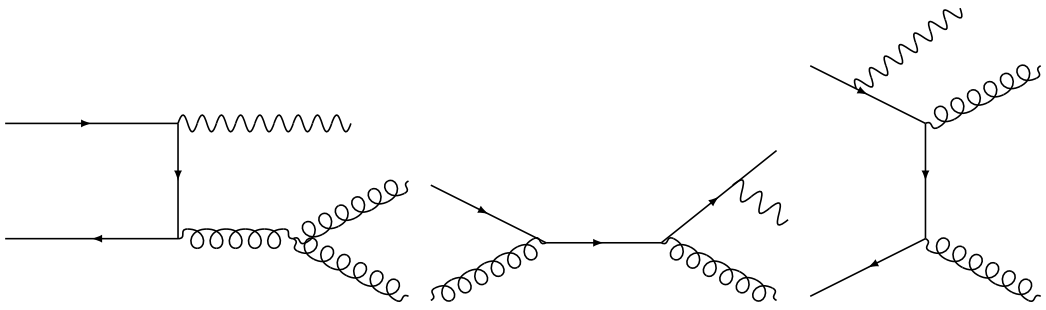
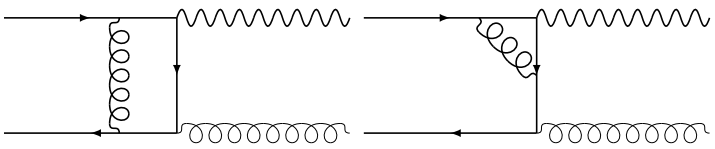
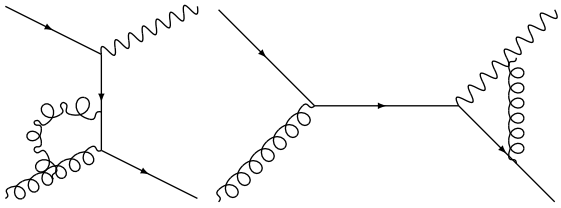


Figure 3.7: A few examples of direct real NLO processes where one new vertex is created.

In order to improve predictions, NLO contributions need to be added, both for the direct part and the fragmentation part. These contributions appear as an additional particle; a real correction (one more vertex) as in Fig. 3.7 or a virtual correction, a loop (two more vertices) as in Fig. 3.8.



(a) Annihilation.



(b) Compton.

Figure 3.8: A few examples of virtual NLO processes where two new vertices are created.

3.1.7 Conclusion

To conclude, we decompose the partonic terms into α_s orders

$$\widehat{\sigma}^{\mathbf{DIR}}(p;\mu_R,\mu_{Fact},\mu_{Frag}) = \left(\frac{\alpha_s(\mu_R)}{\pi}\right)\sigma^{\mathbf{DIR}}_{Born}(p;\mu_{Fact}) + \left(\frac{\alpha_s(\mu_R)}{\pi}\right)^2\sigma^{\mathbf{DIR}}_{NLO}(p;\mu_R,\mu_{Fact},\mu_{Frag})\,, \tag{3.38}$$

$$\widehat{\sigma}^{\mathbf{FRAG}}(p;\mu_R,\mu_{Fact},\mu_{Frag}) = \left(\frac{\alpha_s(\mu_R)}{\pi}\right)^2\sigma^{\mathbf{FRAG}}_{Born}(p;\mu_{Fact}) + \left(\frac{\alpha_s(\mu_R)}{\pi}\right)^3\sigma^{\mathbf{FRAG}}_{NLO}(p;\mu_R,\mu_{Fact},\mu_{Frag})\,, \tag{3.39}$$

where $\sigma^{\mathbf{DIR}}_{Born}(p;\mu_{Fact})$ is fully described in Sec. 3.1.4, while the others can be found in [175–177, 191–194, 204].

Then, taking into account the α_{em}/α_s factor contained in the fragmentation part, which comes from integrating the inhomogeneous term ($\mathbf{P}_{\gamma i}$) of the fragmentation function $D_{\gamma/k}(z,\mu_{Frag})$ in Eq. (3.22), we end up with the same order in α_s for the DIR and FRAG contributions at LO and then at NLO [180]. We can rearrange Eqs.(3.38) and (3.39) to make this clearer in the following analysis, when we discuss the different contributions [5]:

$$\sigma^{tot}_{LO} = \sigma^{\mathbf{DIR}}_{LO} + \sigma^{\mathbf{FRAG}}_{LO}\,, \tag{3.40}$$

$$\sigma^{tot}_{NLO} = \sigma^{\mathbf{DIR}}_{LO} + \sigma^{\mathbf{FRAG}}_{LO} + \sigma^{\mathbf{DIR}}_{NLO} + \sigma^{\mathbf{FRAG}}_{NLO}\,. \tag{3.41}$$

3.2 Theoretical computations

This section is devoted to the study of direct photon production using the JETPHOX [185] code. We will present detailed simulations for differential cross-sections, covering a wide

[5] Note here that this is not the same choice made in Ref. [198].

range of kinematics, and compare these results with available LHC data from proton-proton and proton-lead collisions. Even though our primary interest lies in the $p-Pb$ data for constraining the nuclear gluon density, the comparison with the $p-p$ data remains important for the following reasons: i) to ensure that JETPHOX is being used correctly, ii) to gauge to which degree JETPHOX is able to describe the $p-p$ data, and iii) because $p-p$ simulations will be used to produce fast grids.

We will also extensively compare JETPHOX to MCFM [205–207] in Sec. 3.2.2.

3.2.1 Jetphox

3.2.1.1 Presentation of the code

JETPHOX is a program, combining *analytical* and *Monte Carlo approaches*, which can compute differential and integrated cross-sections for processes including at least one photon in the final state (e.g. $\gamma + X$, or $\gamma + \text{jet}$)[6] and can generate events at parton level. More precisely, JETPHOX[7] is designed to compute the double-inclusive cross-section:

$$E\frac{d\sigma(H_1 + H_2 \to \gamma + jet + X)}{d\overrightarrow{p_1}d\eta_2}, \tag{3.42}$$

where the γ and the jet are high-p_T particles, with $\overrightarrow{p_1}$ the momentum of the photon and η_2 the rapidity of the recoiling jet. This form of differential expression allows modularity in obtaining the final observable; the code is able to compute cross-sections (σ) as well as differential cross-sections ($d\sigma/dp_T^\gamma$ or $d\sigma/d\eta^\gamma$). It is also possible to implement different experimental cuts (e.g., kinematic cuts or isolation cuts) at partonic level. In the case of single-inclusive photon production, the cross-section is obtained by integrating over all other final-state particles. Note that the approximation of massless quarks is used in this code[8].

Currently, JETPHOX is the only *public* code capable of computing the isolated prompt photon cross-section at NLO in QCD for both contributions; the **direct** and **fragmentation** part, as defined in Eq. (3.41).

[6] Note that the photon can be replaced by any hadron if the appropriate fragmentation function is used.

[7] This code is part of the PHOX family, and is directly derived from DIPHOX [208], a code implemented to compute the double cross section $E\frac{d\sigma(A+B\to C+D+X)}{d\overrightarrow{p_1}d\eta_2}$ (A, B are hadrons and C, D are photons or hadrons) and was used for the di-photon final state process, an irreducible background to the search of Higgs bosons at LHC [209].

[8] It means that the code cannot be used as it is for heavy quark production.

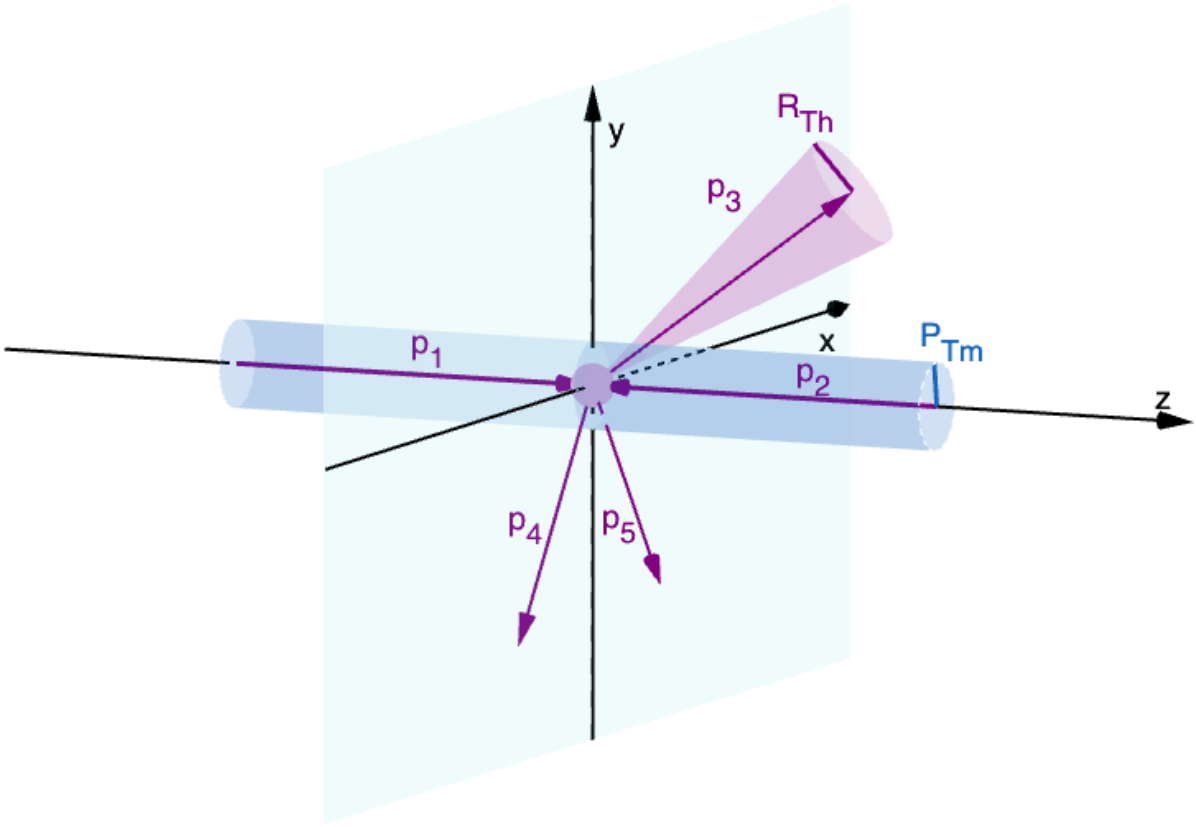


Figure 3.9: Schematic representation of p_{Tm} and R_{th} parameters used in JETPHOX. See Eq. (3.43).

3.2.1.2 Technical features

As already mentioned in Sec. 3.1.6, singularities arise and must be treated. JETPHOX combines the phase space slicing method [210, 211] and the subtraction method [212, 213] to treat the soft and collinear singularities present in the perturbative matrix elements.

At the partonic level, let us consider the general process $1 + 2 \to 3 + 4 + 5$, as illustrated in Fig.3.9, where 3 and 4 are high-p_T particles and thus well separated in phase space. The final state parton 5, however, can be soft or collinear to one of the other four partons (i.e 1, 2, 3 or 4). To regulate the divergences, two arbitrary and unphysical parameters are introduced, p_{Tm} and R_{Th}, satisfying

$$p_{Tm} \ll |\overrightarrow{p_{T3,4}}| \qquad \text{and} \qquad R_{Th} \ll 1\,. \tag{3.43}$$

According to these two parameters, the phase-space is then sliced in the following way:

- **Part 1** : $|\vec{p}_{T5}| < p_{Tm}$. This part contains the IR and initial-state collinear singularities and a small fraction of the final-state collinear singularities from parton 5, when it is very soft.

- **Part 2a** :$|\overrightarrow{p}_{T5}| > p_{Tm}$ and parton 5 is inside a cone C_3 which corresponds to the particle 3 direction where $(y_5 - y_3)^2 + (\phi_5 - \phi_3)^2 \leq R_{th}^2$. This part yields (most of) the final-state collinear singularities, i.e. 5 collinear to 3, those that are not contained in **part 1**.

- **Part 2b** : same as **2a** but with $3 \leftrightarrow 4$.

- **Part 2c**: $|\overrightarrow{p}_{T5}| \geq p_{Tm}$ and parton 5 is outside both cones C_3 and C_4. There is no divergence here, and this contribution can be directly treated in 4 space-time dimensions.

The divergences of part **1**, **2a**, **2b** are regularized using dimensional regularization from 4 to $n = 4 - 2\epsilon$ with $\epsilon < 0$ and finally the n-dimensional integration of these three regions is performed (analytically) over the following kinematic variables of the parton 5: the transverse momentum (p_T), the rapidity (η) and the azimuthal angles (ϕ). This leads to poles in $1/\epsilon$ as well as dependencies on p_{Tm} and/or R_{th}^2. In the part **2a** and **2b**, a subtraction method is also applied to keep the exact R_{th} dependence. With a careful[9] choice of the slicing parameters R_{th} and p_{Tm}, the final result does not depend on these two parameters [185]. Since the integration is carried out analytically by neglecting terms that are proportional to (power of) p_{Tm}, this parameter has to be chosen small enough; 0.5% of $p_{T3,4}^{min}$ works [209] and will be set in the following analysis (see line 220 in Appendix A.1).

Then, after combining with the corresponding virtual contributions, the soft singularities cancel out. The remaining contributions are then factorized and absorbed in PDFs and FFs. The results of these calculations are pseudo cross-sections, in the sense that they are not true physical cross-sections since they can be (and are generally) negative. There are no divergences in **2c**, so the integration is performed numerically, without approximation.

3.2.1.3 Calculation of the observables

Once all singularities have been canceled, all contributions can be integrated. JETPHOX can provide the integrated cross-section (σ^{tot}) as well as generate events, allowing the user to reconstruct the desired observable onto a histogram. As the code is relatively lengthy, it is beneficial to have the option of reusing these events for different observables. Additionally, this approach allows for only limited modularity: it permits the computation

[9] A sufficiently small p_{Tm} and R_{th} means, here, that they give, respectively, vanishing terms and finite terms when $p_{Tm} \to 0$ and $R_{th} \to 0$.

of different observables but with constrained kinematics, PDFs, or FFs. This is not an optimal solution for global analyses, which is why we will introduce, in Chap.4, a solution enabling these events to be utilized in more flexible and efficient way.

The integration is performed using BASES [214], a Monte Carlo integrator based on the VEGAS algorithm. Although the integrated cross-section σ^γ can be calculated relatively quickly, even when all contributions are included, this is not the observable we aim to compare with experimental data. The experiments that we are interested in, such as ATLAS or ALICE at CERN, provide differential distributions such as $\frac{d\sigma^\gamma}{dp_T^\gamma}$ or $\frac{d\sigma^\gamma}{d\eta^\gamma}$. JETPHOX is capable of generating partonic events, thanks to SPRING module [214] and stores the 4-momenta of the final particles, along with the required normalization factors, in a ROOT [215] file. This file contains a TTree (referred to as a *ntuple*[10] in JETPHOX and in this work) which contains the events and a TVectorT that stores the total cross-section, the number of event and the center of mass energy. These two objects contain the necessary information to calculate the final observable(s), which can then be projected onto a histogram to obtain the specified distribution. A brief analysis code is provided in Appendix 1 (l.75→l.103). Additionally, JETPHOX offers the possibility of calculating PDFs uncertainties in a single run. The resulting ntuple is structured as follows, for each event you have:

- `ntrack`: is the number of final partons (partons 3,4,5 in Fig.3.2) which is 2 or 3 in our cases.
- `e[ntrack],px[ntrack],py[ntrack],pz[ntrack]`: are the 4-momentum vector of each particle in the final state. Components 0,1 and 2 refer respectively to the photon(3); the hard parton(4) and soft or collinear parton(5) of Fig.3.2).
- `iprov`: is the index of the phase space configuration of the event, where the full description is in Table 3.1.
- `x3`: is the fragmentation variable. It corresponds to the momentum fraction of the parent parton, carried by the fragmented photon. It is set to 1, when there is no fragmentation.
- `nb_member`: is the PDF member used in the calculation; equal to 1 if no error is calculated, else equal to *nb_member*.
- `pdf_weight[nb_member]`: is the vector that contains the weight ratio for each member divided by the central value.

[10] A ntuple is an object inherited from a TTree, just restricted to being filled by float.

ntrack	iprov	Phase-Space type	JETPHOX function
3	34	$2 \longrightarrow 3$ (parton 5 // 3)	f4dimd3
3	44	$2 \longrightarrow 3$ (parton 5 // 4)	f4dimd4
2	21	quasi-2 $\longrightarrow$ 2 (parton 5 // initial state)	f2dimd1
2	23	quasi-2$\longrightarrow$ 2 (parton 5 // γ (3))	f2dimd3
2	10	$2 \longrightarrow 2$ (virtual)	f1dimd
2	11	$2 \longrightarrow 2$ (born)	f1dimd

Table 3.1: Correspondence table between partonic configurations and index in JETPHOX analysis.

- `nb_evt` : is just the number of generated event, needed for event normalization.
- `xsec` : is the total pseudo-cross section of the contribution(s) taken into account, also needed for event normalization.
- `sqrt_s`: is the center of mass energy.

Generate events.

There are several steps involved in moving from different partonic configurations to a final ROOT file containing a defined number of events :

- Select the number of events to generate in the parameter.indat input file (set `TRUE` at line 272 and define the desired number of events at line 411 in Appendix.A.1).
- JETPHOX computes the integrated cross-section and stores it in a file/variable. In practice, it calculates an 'integrated cross section' for each type of phase-space configuration (as $2 \longrightarrow 2, \text{quasi}-2 \longrightarrow 2, 2 \longrightarrow 3$ etc.). Each configuration will receive a fraction of the total number of generated events, proportional to its contribution to the total cross-section.
- SPRING generates events using the Hit and Miss method, which is represented schematically in Fig. 3.10.
- The result is the probability distribution :

$$\frac{dN}{dx} \,. \tag{3.44}$$

and if we want a cross-section distribution, we use:

$$\frac{d\sigma}{dx} = \frac{\sigma_{tot}}{N}\frac{dN}{dx} \,, \tag{3.45}$$

which explains why we store the effective cross-sections for proper normalization.

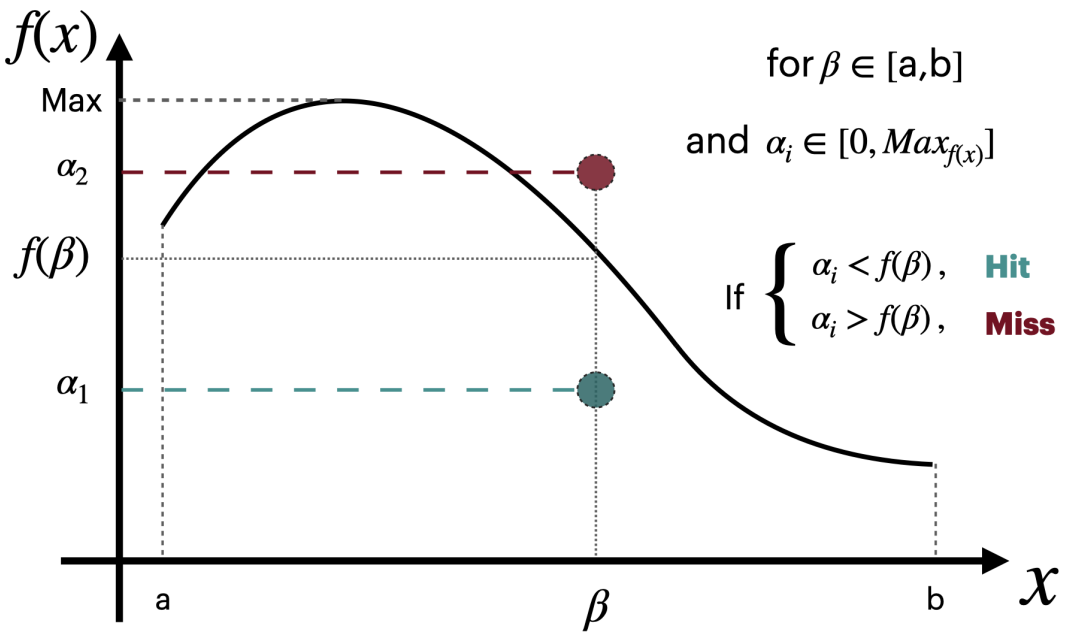


Figure 3.10: Diagrammatic representation of the *Hit and Miss Method*.

This method allows for flexibility in terms of the desired final observable. With proper normalization, we can easily obtain a p_T or η distribution. Additionally, the hit-and-miss method enables faster convergence to the desired result; fewer events are needed to achieve the same precision as with a basic event generator. However, it tends to populate the high-values bins more intensively, which may lead to an underestimation in the lower-value bins. This issue can be fixed by using bins that are not too wide.

3.2.2 MCFM

3.2.2.1 Presentation of the code

MCFM (A Monte Carlo for FeMtobarn processes) [205,206,216] is a program designed for hadron colliders. A large number of processes (currently more than 300) are available at NLO in QCD, with some also computed at NNLO and even N3LO in QCD. Additionally, certain processes include NLO electroweak corrections. MCFM is a highly versatile tool, capable of supporting the computation of a wide range of processes, thereby making it an invaluable resource for global analyses. Furthermore, MCFM provides gridding capabilities through an existing interface, thereby enhancing its usability for a range of applications. This is particularly relevant for global analyses requiring a large number of different processes. In fact, several theoretical calculations within the nCTEQ framework are produced using MCFM (e.g. $W^{\pm}, Z$ production [217]).

However, several challenges are associated with using MCFM. Firstly, the code base

is extensive and intricate, making it challenging to identify the specific components that are pertinent to a given process. Secondly, the lack of comprehensive documentation is a notable shortcoming, since not all processes or updates are accompanied by dedicated publications. This can present a significant obstacle when attempting to obtain detailed information about particular functionalities or recent developments within the code. Although such a resource did not exist at the start of this work, the MCFM website [218] has since become a comprehensive platform, offering a wealth of information and related papers, which we will use below.

We have begun to work with the latest version of MCFM available at the moment, MCFMv9.1 [207]. While this version reproduces the data reasonably well, as you can see later in Fig. 3.15, it does not support grids generation. For this reason, we switched to an earlier version, v6.8, which does allow for grid production. Note here that the implementation of direct photon production contains an error in this version of the code and a patch [11] was required to compute correct calculations [219].

3.2.2.2 Calculations

There is no dedicated publication detailing the implementation of the single photon production process in MCFM, but it was introduced in version 6.1 (id process=280). The process is briefly mentioned in [206], in the context of the di-bosons processes, since the latter can contain final states with photons. Collinear singularities are treated using the Catani-Seymour subtraction method, in contrast to JETPHOX where phase space slicing is used. As for the fragmentation contribution, it is included at LO, but is neglected at NLO due to its small size, and largely suppressed by the isolation criteria [206].

MCFM also relies on a Monte Carlo approach for integration, using QCDLoop [220] library, and is fully parallelized, allowing faster code execution.

Sec.3.3 will aim to demonstrate the agreement between JETPHOX and MCFM calculations, to take advantage of MCFM's ability to produce grids.

3.2.3 Other calculations in the literature

In this section, we present other existing calculations in the literature. We will not use them in the following but we discuss them here for completeness.

NNLOjet[1]

NNLOJET is a fixed-order parton-level event generator. It was designed for jet

[11] This patch has been provided by one of the author of Ref. [219], J.Rojo.

[1] http://nnlojet.hepforge.org

cross-sections at NNLO. The Antenna Subtraction Method [221] has been used to cancel infrared singularities. Using this code, a paper has been published focusing on a realistic photon isolation and the importance of the fragmentation contribution [198]. This latter contribution has been calculated at NNLO, as a convolution of NLO hard-scattering processes convoluted with fragmentation functions of the final state partons into the photon. In this sense, the fragmentation contribution has the same accuracy as the one in JETPHOX whereas the direct contribution is one order higher. Comparisons between the two codes for the γ + jet final state can be found in (Table 1 of [198]) and it would be interesting to perform a similar comparison for γ + X. This code has been used in a number of phenomenological studies, see, e.g., Ref. [198]. Unfortunately the NNLOJET code was not public at the beginning of this thesis but it is now public [222].

Sherpa[2] [223, 224]

Sherpa is a Monte Carlo event generator, designed for $\ell\ell$, $\ell\gamma$, $\gamma\gamma$, ℓh and hh collisions (where ℓ are leptons and h are hadrons). There are several final states available; including prompt photon production. This code uses a Parton Shower (PS) formalism merging QCD-matrix elements to the PS. A Phase Space Slicing method is employed by introducing a Q_{ij} cut to separate the tree-level matrix element (ME) region and the (initial and final) PS region. Sherpa was used in the ATLAS analysis for $p - p \longrightarrow \gamma + X$ at 13 TeV [225].

PeTeR[3] [226, 227]

PeTeR is a code designed to compute the transverse momentum spectrum of the production of an (on-shell) electroweak boson (γ, W, Z, Higgs). It is based on the Soft-Collinear Effective Theory (SCET) and performs a threshold resummation for large p_T up to Next-to-Next-to-Next-to-Leading-Logarithmic (N3LL) accuracy, matched to full NLO in α_s.

POWHEG[4] [228–230]

The POWHEG BOX code for prompt photon production [231, 232] consistently combines the NLO calculation with a PS by avoiding double counting, using the POWHEG method. In particular, the photon FF can be modeled by an interleaved QCD+QED PS.

[2] https://sherpa-team.gitlab.io/
[3] https://peter.hepforge.org/
[4] https://powhegbox.mib.infn.it/

PYTHIA[5] [80]

PYTHIA is a general-purpose Monte-Carlo generator of high-energy physics processes. It contains a lot of different tools, including computation of the partonic cross-section (at LO), initial/final PS, matching and merging of ME and PS, multi-parton interaction (MPI), hadronization/fragmentation (with the Lund string model of hadronization) and decay. As such, it is widely used by the experimental collaborations, alone or in association with other codes, in the analysis of experimental data.

For the prompt photon production (paragraph 3.2.1 of [80]), only the $2 \longrightarrow 2$ processes with massless partons at partonic level with a $p_\perp$ (= p_T) cut is available. Furthermore, PS and/or Hadronization can be added for a more realistic description of events.

MadGraph5_aMC@NLO[6] [76, 233]

MadGraph5_aMC@NLO is another framework for the automated computation of tree-level and NLO differential cross sections, and their matching to parton shower simulations. In particular, it is conveniently possible to match the fixed order calculations to both, the Herwig++ [234] and the Pythia6 [235] PS.

In summary, there is a large choice of codes available, including different terms of the perturbation theory:

- LO+LL accuracy (Sherpa, PYTHIA)
- NLO = NLOdir + NLOfrag (Jetphox)
- NNLOdir[7] + LOfrag (MCFM)
- NNLOdir + NNLOfrag[8] (NNLOJet)
- NLO + LL (POWHEG, MadGraph5aMC@NLO)
- NLO + N3LL threshold (PeTeR)

Unfortunately, not all of these codes have been benchmarked against each other and experimental data. Nevertheless, many of them have been used in phenomenological studies. For instance, the ATLAS analysis for prompt photon production at 8 TeV (in

[5] https://pythia.org/
[6] http://amcatnlo.web.cern.ch/amcatnlo/
[7] Note that the NNLO direct part is obtained by multiplying the NLO direct part with a K-factor.
[8] The fragmentation part is the same as the NLO fragmentation part in JETPHOX.

$p-p$ collisions) [236] used most of these codes (PYTHIA, Sherpa, JETPHOX, PeTeR, and MCFM) and compared their results.

3.3 Available data and validation of theoretical calculations

This section presents an analysis of prompt photon production in proton-proton (p-p) and proton-nucleus (p-A) collisions at the Large Hadron Collider (LHC), using experimental data from ATLAS and ALICE[9]. We perform theoretical simulations using two well-established codes, JETPHOX and MCFM, to compare with p-p collision data at center-of-mass energies of 7 [241,242], 8 [243], and 13 [225] TeV from ATLAS, incorporating both scale and, in some cases, PDF uncertainties. Additionally, these simulations are extended to p-A collisions at 8.16 TeV [187], and the results are compared to data from ATLAS (and ALICE).

The motivations for this study are multifaceted: i) to assess the ability of JETPHOX and MCFM to describe p-p data, ii) to select the default code for subsequent studies within the nCTEQ global analysis framework, iii) to use the p-p results as a baseline for generating grids using APPLGRID for fast computations, and iv) to apply these grids, convoluted with proton and lead PDFs, to evaluate theoretical predictions against p-Pb data.

In Sec. 3.3.1, we focus on the p-p case, while Sec. 3.3.2 addresses the p-Pb analysis. To ensure the validity of our results, we introduce a modification to JETPHOX, enabling the convolution with two distinct hadronic PDFs to accommodate the p-Pb scenario. Additionally, minor modifications have been made to the publicly available version 1.3.1_3 of JETPHOX. Specifically, the masses of the c and b quarks have been updated according to recent values reported by the PDG [12]: 1.27 GeV and 4.18 GeV, respectively.

The overarching goals of this work are to qualitatively assess whether p-Pb direct photon data can provide meaningful constraints on the nuclear gluon PDF and to lay the groundwork for incorporating these data, along with the corresponding grids, into the nCTEQ global analysis framework.

[9]There are also p-p data from CMS [237–239] which are not included in our analysis since there are no corresponding p-A data. Moreover, there are preliminary p-p and p-Pb data from LHCb shown in a conference proceeding from 2019 [240].

3.3.1 Proton-proton collisions at different center-of-mass energies

In the following, we present numerical results for prompt photon production at the LHC at c.m.s energies of 7, 8 and 13 TeV.

7 TeV

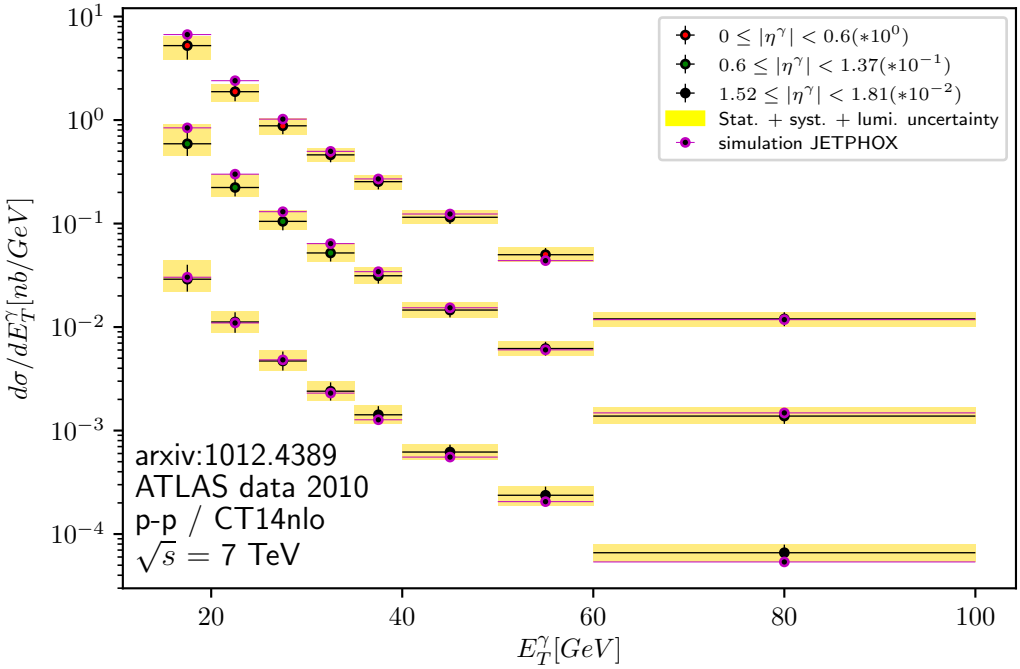


Figure 3.11: ATLAS data [241], with experimental uncertainties in yellow bands, for the differential cross-section of prompt photon production, as a function of photon transverse momentum p_T^γ, at $\sqrt{s}$ = 7 TeV, compared to Jetphox predictions at NLO for the three rapidity ranges available : $0 \leq |\eta^\gamma| \leq 0.6$, $0.6 \leq |\eta^\gamma| \leq 1.37$, $1.52 \leq |\eta^\gamma| \leq 1.81$. Here, the CT14nlo pdf was used for the theoretical predictions (in comparison in [241], CTEQ6.6 was used). The isolation criterion is taken as the same as for data ; $E_T^{iso} < 3$ GeV. The Jetphox predictions are taken at $\mu_R = \mu_F = \mu_{Frag} = p_T^\gamma$ and no theoretical uncertainties (i.e., scale variations or PDF errors) are shown.

Fig. 3.11 presents data on direct photon production from the ATLAS experiment at 7 TeV [241]. The data are divided into three distinct rapidity ranges ($0 < |\eta^\gamma| < 0.6$, $0.6 < |\eta^\gamma| < 1.37$ and $1.52 < |\eta^\gamma| < 1.81$), covering a transverse momentum (p_T) range from 15 GeV to 100 GeV. The plot shows the differential cross section ($d\sigma/dp_T$) as a function of transverse momentum (p_T). The theoretical predictions were generated using the CT14nlo PDFs, with the isolation criterion set to match the ATLAS paper specifications: $E_T^{iso} < 3$ GeV and R= 0.4. While theoretical uncertainties are not included in this plot and will be discussed later, the experimental uncertainties are indicated by yellow bands. As shown

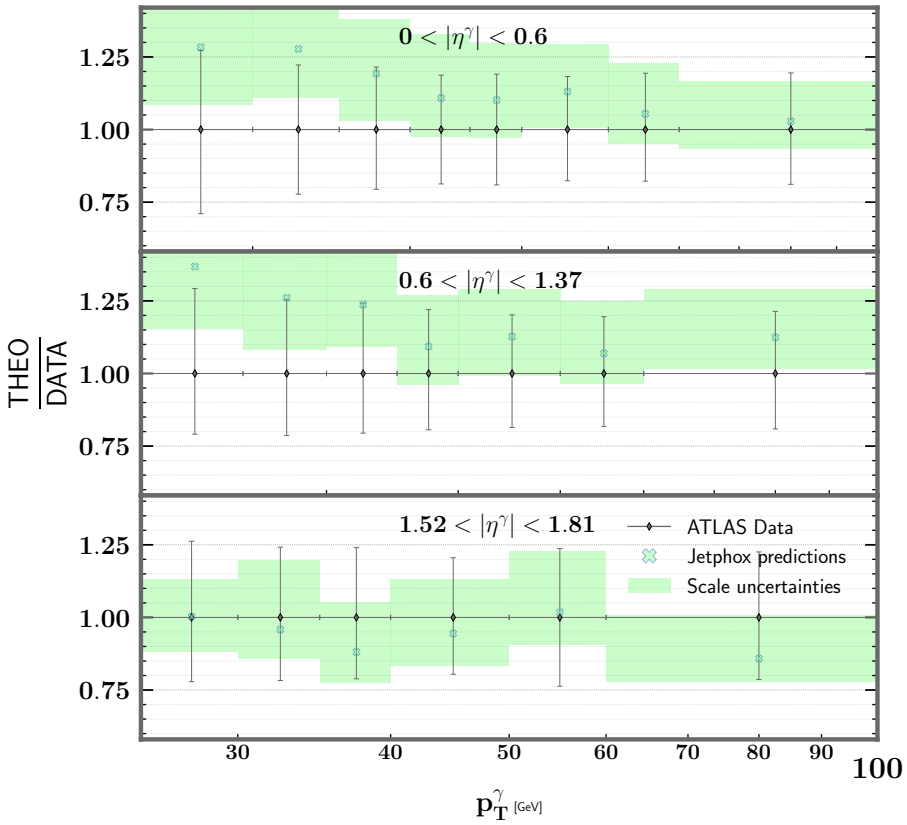


Figure 3.12: The ratio of theory over data is shown for the three rapidity ranges and with the scale uncertainties. Same data as Fig. 3.11 are used.

in the logarithmic plot, the theoretical predictions exhibit very good agreement with the experimental data over six orders of magnitude.

Fig. 3.12 provides a more detailed view of the three rapidity regions through ratio plots of data over theory. The scale uncertainties were calculated using the following prescription:

$$\{C_1\ \mu_R, C_2\ \mu_{Fact}, C_3\ \mu_{Frag}\}\,, \tag{3.46}$$

where $\forall i \in \{1,2,3\}$, the coefficients C_i are chosen from $\left\{\frac{1}{2}, 1, 2\right\}$, subject to the constraint:

$$\forall j \neq i; \quad \frac{C_i}{C_j} \neq 4 \ \text{ and } \ \frac{C_j}{C_i} \neq 4\ . \tag{3.47}$$

This results in 15 scale variation combinations. For each bin, the *min* and *max* values across these combinations define the uncertainty envelope, with $(1,1,1)$ taken as the central value. The corresponding combinations used to determine the upper and lower bounds are listed in Table 3.2.

	$(\mu_R, \mu_F, \mu_{Frag}) = (x_1, x_2, x_3) \times p_T^\gamma$	
p_T^γ (GeV)	Max	Min
15-20	$\left(1, \frac{1}{2}, 1\right)$	$\left(\frac{1}{2}, 1, \frac{1}{2}\right)$
20-25	$\left(1, \frac{1}{2}, 1\right)$	$(1, 2, 1)$
25-30	$\left(1, \frac{1}{2}, 1\right)$	$(1, 2, 1)$
30-35	$\left(1, \frac{1}{2}, 1\right)$	$(1, 2, 1)$
35-40	$\left(1, \frac{1}{2}, 1\right)$	$(1, 2, 1)$
40-50	$\left(1, \frac{1}{2}, 1\right)$	$(1, 2, 1)$
50-60	$\left(1, \frac{1}{2}, 1\right)$	$(1, 2, 1)$
60-100	$\left(1, \frac{1}{2}, 1\right)$	$(1, 2, 1)$

Table 3.2: Values of $x_{i=1,2,3}$ for the uncertainty envelope, as defined in Eq.(3.47), for JETPHOX predictions of Fig. 3.12 (upper panel).

This approach confirms that, for all rapidity regions, each data point of the theoretical predictions is in perfect agreement with the NLO theoretical predictions, within the uncertainties. The aforementioned prescription will be used consistently throughout this thesis whenever scale uncertainties are calculated with JETPHOX[10].

Next, Fig. 3.13 compares JETPHOX and MCFM predictions with the 7 TeV ATLAS data from [241], focusing on the central photon rapidity region $|\eta^\gamma| < 0.6$. The isolation criterion applied in both simulations is $E_T^{iso} < 3$ GeV, with a cone radius of R = 0.4. The simulation shown in red corresponds to JETPHOX at NLO accuracy, which includes both the direct and fragmentation contributions. In contrast, the violet points represent MCFM, which includes the direct contribution at NLO and the fragmentation contribution at LO. Both simulations use the CT14nlo PDFs [244] for the protons and the NLO photon FFs BFGII [190] in the $\overline{MS}$ scheme. The renormalization, factorization and fragmentation scales are set to $\mu_R = \mu_F = \mu_{Frag} = p_T$. The top panel of the figure displays the p_T-differential distribution, while the bottom panel shows the ratio of THEORY/DATA, providing a clearer visual comparison of the simulation's agreement with the experimental data. In this case, both codes demonstrate similarly good agreement with the experimental data.

These data were later extended in [242] to cover a higher p_T range, with photon rapidity bins $|\eta^\gamma| < 1.37$ and $1.52 < |\eta^\gamma| < 2.37$ for $100 \leq E_T^\gamma \leq 1000$ GeV and $100 \leq E_T^\gamma \leq 600$ respectively.

[10]For MCFM, the calculation of the scale uncertainties can be done automatically, however it takes into account only a variation of μ_R and μ_F.

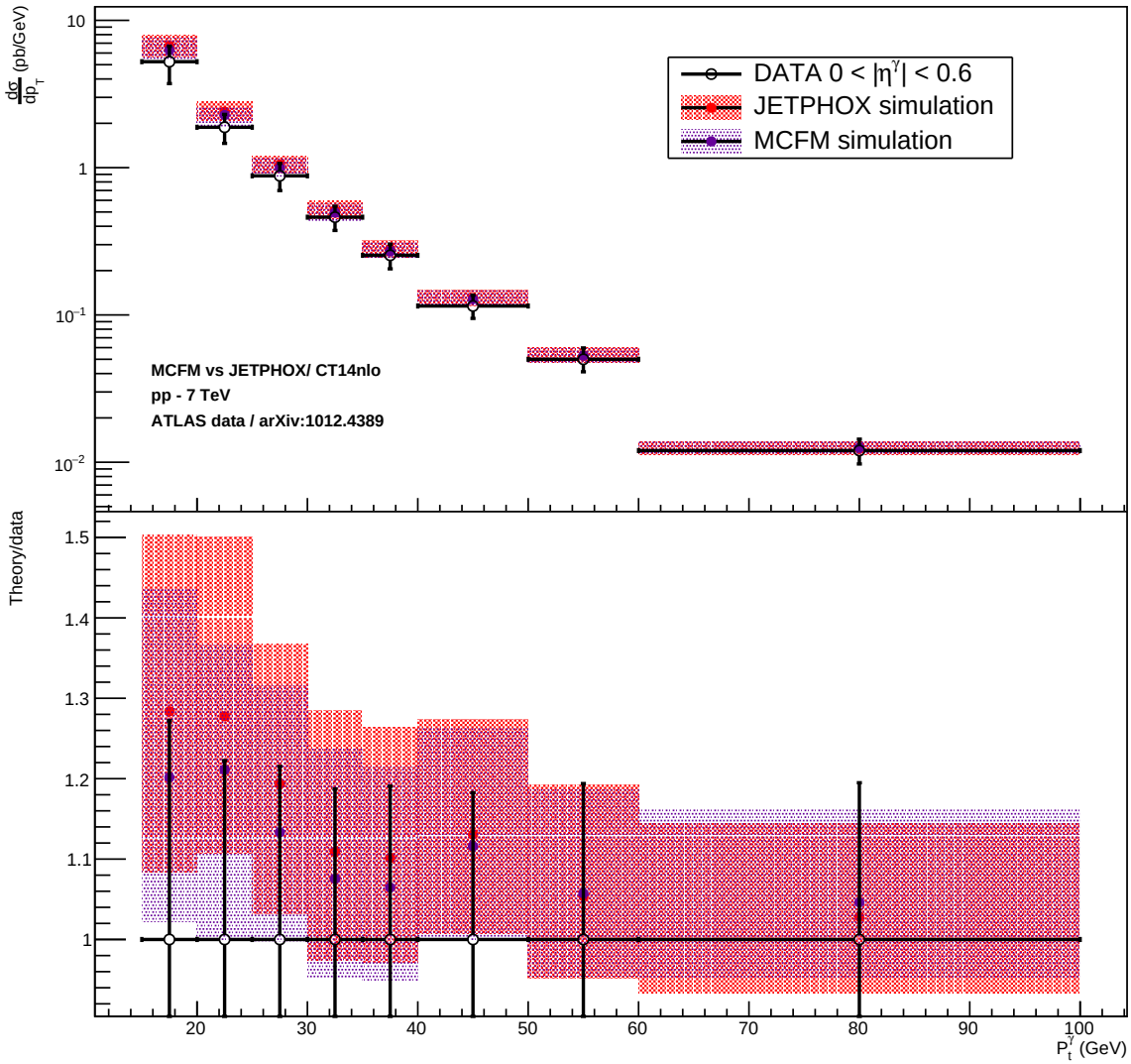


Figure 3.13: Here, JETPHOX and MCFM are compared. The ratio of Theory over Data is shown in the lower panels of the figure and the scale uncertainties of theoretical computations are included. Same data as in Fig. 3.11 and Fig. 3.12.

8 TeV

We have also conducted a detailed study of proton-proton collisions at 8 TeV using the ATLAS data [243]. In Fig. 3.14, we present the ratio of data over theory for $\frac{d\sigma}{dp_T}$ as a function of transverse momentum (p_T). The theoretical predictions were generated using the CT14nlo PDFs, with an updated isolation criterion based on the ATLAS specifications; $E_T^{iso} < 4.8\text{GeV} + 4.2\times10^{-3}E_T^{\gamma}$. In this case, scale uncertainties are shown as the green slope, while PDF uncertainties are presented by vertical green error bars. The PDF uncertainties are computed following the prescription in [245], using the Hessian method adopted by the CT14 collaboration,

$$\Delta O = \frac{1}{2}\sqrt{\sum_k \left(O_k^+ - O_k^-\right)^2}, \tag{3.48}$$

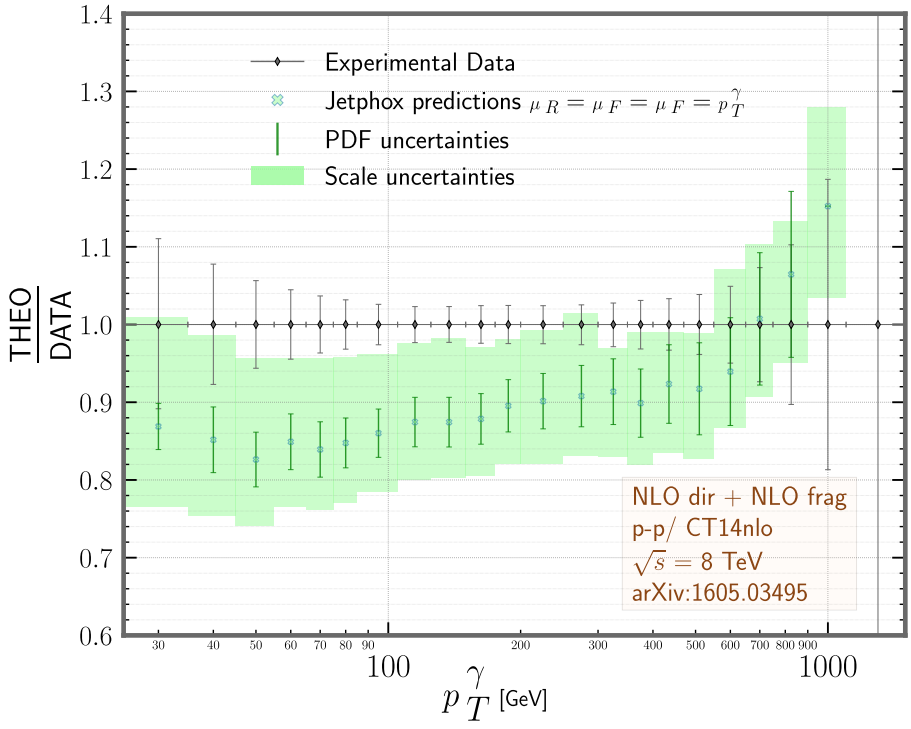


Figure 3.14: The ratio of ATLAS data [243] compared to JETPHOX predictions, including scale and PDF errors.

where O denotes the observable of interest (in this case, the differential cross-section), and $O_k^{\pm}(k = 1, \ldots, 28)$ are the results obtained using the 56 error PDFs of the CT14nlo distribution. As observed, the PDF uncertainties are significantly smaller than the scale variation uncertainties at small p_T ($\simeq 500$ GeV), but they become comparable in magnitude as p_T increases.

Note that the comparison with MCFM is not shown here. However, an in-depth study is conducted in the following chapter, and Sec. 4.2.1 includes calculations made with MCFM for this kinematics.

13 TeV

A comparison between JETPHOX and MCFM is shown in Fig. 3.15 for the higher energy ATLAS data at $\sqrt{s} = 13$ TeV, with the p_T range for the central rapidity region being $125 \leq p_T \leq 2500$ GeV. Excluding the last bin, where the statistics are poor, we observe that JETPHOX predictions tends to lie below the data for many bins, while the MCFM predictions start below the data but end up above it at higher p_T.

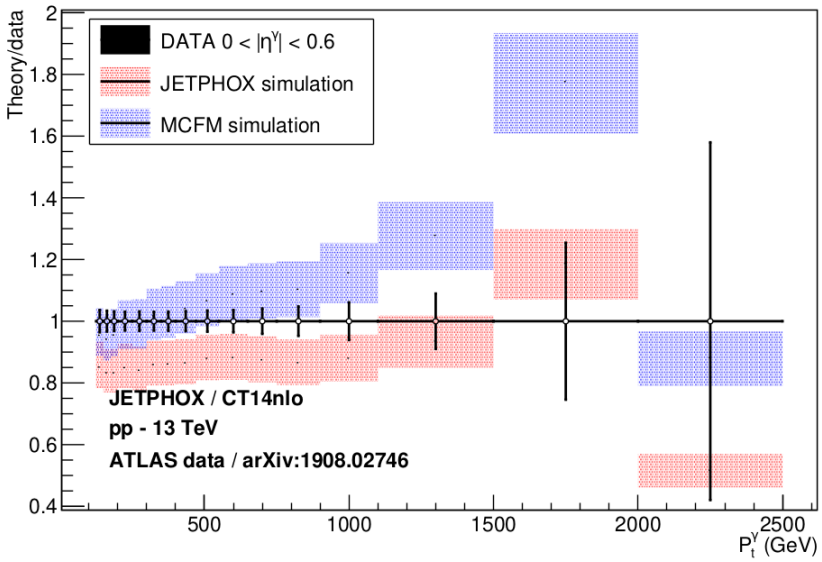


Figure 3.15: A comparison between JETPHOX and MCFM predictions for p-p collisions at 13 TeV [225], for the central rapidity. As previous plots, CT14nlo PDFs are used.

Remarks

While MCFM and JETPHOX accurately describe the data at 7 TeV, a discrepancy emerges between the two codes for the other two kinematics, and the description of the data at 8 TeV and 13 TeV is less accurate. It is important to note that, when taking into account scale uncertainties and potential modifications related to PDFs and their associated uncertainties, it may still be valid to employ either one, without being able to determine a clear preference between the two.

3.3.2 Proton-lead collisions

In the following, we present numerical results for prompt photon production in *p-Pb* collisions at the LHC, with a cms energy of 8.16 TeV, using data from ATLAS and ALICE experiments. In this section, we have employed a modified version of JETPHOX, in which we have allowed two different PDFs to be used for the two incoming hadrons; proton and lead.

As a remainder; nuclear PDFs are parametrized as

$$f_i^A = Z\, f_i^{p/A} + (A - Z)\, f_i^{n/A}\,, \tag{3.49}$$

where f_i^A is the nuclear PDF, $f_i^{p/A}$ and $f_i^{n/A}$ are the PDFs of protons and neutrons[11] inside the nucleus A, respectively and Z is the number of protons in the nucleus. Usually, nuclear corrections are defined as

$$R_i^A(x, Q^2) = \frac{f_i^{p/A}(x, Q^2)}{f_i^p(x, Q^2)}, \tag{3.50}$$

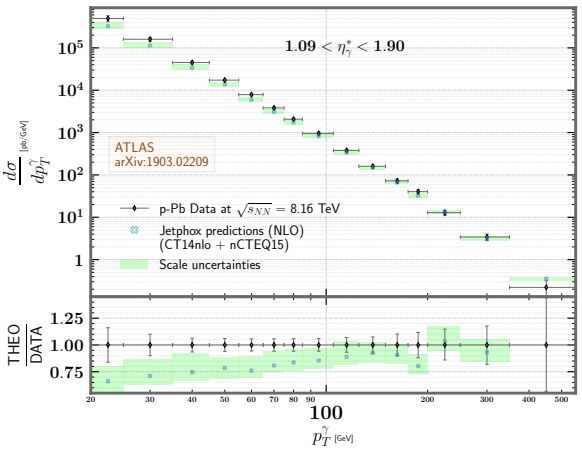
where f_i^p is the PDF of a *free* proton, and R_i^A is the nuclear correction for a nucleus A. JETPHOX includes several nuclear corrections (R_i^A) sets, such as nDS and EKS98, among others[12]. These corrections imply using the same proton PDF for the proton and the nucleus, modified by a nuclear correction factor constrained by the chosen nuclear correction set. However, our preferred approach is to use a dedicated nuclear PDF for the nucleus and a separate proton PDF for the proton. To this end, we have implemented a function in JETPHOX that calls a second PDF set when a nuclear collision occurs.

ATLAS

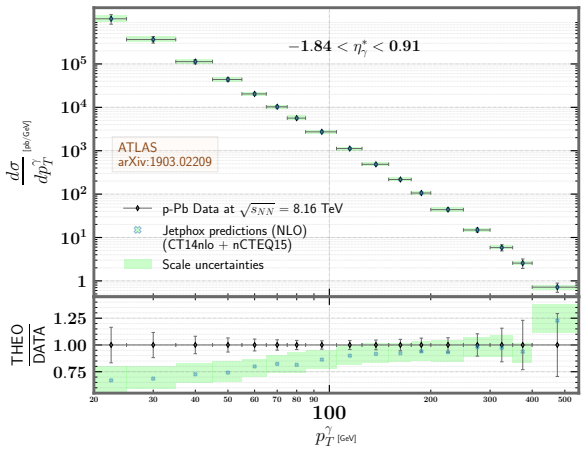
Fig. 3.16 presents data on (isolated) prompt photon production from the ATLAS experiment at 8.16 TeV, for *p-Pb* collisions [187]. The data are divided into three distinct pseudo-rapidity ranges ($1.09 < \eta_\gamma < 1.90$, $-1.84 < \eta_\gamma < 0.91$ and $-2.83 < \eta_\gamma < 2.02$), respectively called forward, central and backward rapidity, covering a transverse momentum (p_T^γ) range of 20 to 550 GeV for the first two, and 20 to 350 GeV for the last one. The theoretical predictions were generated using the CT14nlo PDFs for the proton and the nCTEQ15 PDFs for lead, with the standard ATLAS isolation criterion, as previously described in Sec. 3.3.1. The theoretical uncertainties are displayed in the usual manner. As shown in the upper plots, on a logarithmic scale, the theoretical predictions exhibit

[11] The neutron PDF ($f_i^{n/A}$) is obtained using isospin symmetry from $f_i^{p/A}$.

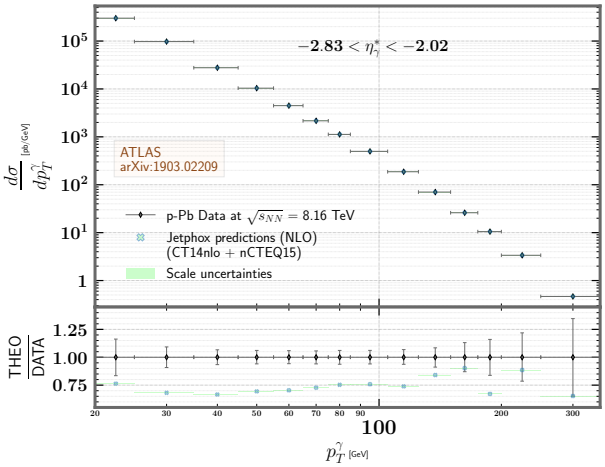
[12] A full list of nuclear corrections is available directly in the input file of JETPHOX.



(a) Forward.



(b) Central.



(c) Backward.

Figure 3.16: ATLAS data [187] for *p-Pb* collisions, compared to JETPHOX predictions, including scale errors for forward and central rapidity. The isolation criterion is the usual ATLAS criterion, as in Sec. 3.3.1. The upper panel shows the absolute distribution and the lower panel shows the ratio of theory prediction over data.

good agreement with the experimental data. The lower panels, showing the ratio of theory to data indicate reasonably good agreement within the scale uncertainties. However, we can note that the predictions are almost always below the data, with the exception of a few bins at large p_T.

ALICE

An analysis at the same cms energy is currently being conducted by the Alice collaboration; focusing on a lower p_T-range. Preliminary results were presented in [246] and we used the data provided therein[13] to compare the *p-Pb* data with JETPHOX predictions, as shown in Fig. 3.17. The JETPHOX predictions show good agreement with data.

[13]The data were kindly provided by the author of [246] through private communication. Update : the complete ALICE analysis and data, have been published by the ALICE collaboration [188] since then.

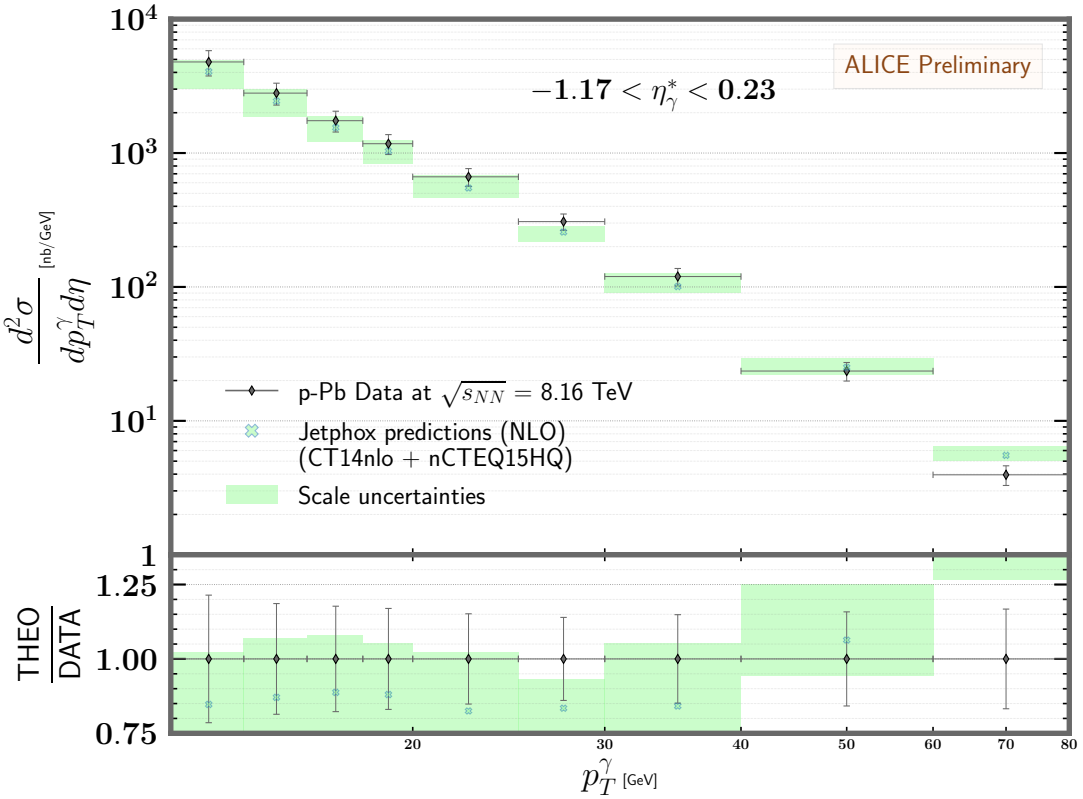


Figure 3.17: ALICE preliminary data from [246] for double differential cross-section of prompt photons, compared to Jetphox. Here, the nCTEQ15HQ nPDFs was used for the nucleus. The isolation criterion is $E_T^{iso} < 1.5$ GeV and $R = 0.4$.

3.4 Conclusions

In conclusion, we can state that direct photon production in p-p collisions is well described by both codes. We did not perform p-A predictions with MCFM to compare with JETPHOX, as we preferred not to use the built-in EKS98 nuclear correction factor [247], but instead to supply two distinct PDFs as input. Implementing such a modification in MCFM appeared to be complex and time-consuming. However, it was not necessary to modify the code for grid production with MCFM, since grids can be generated for p-p collisions and subsequently convoluted with the desired PDFs. In the next chapter, we will proceed to create grids using MCFM and verify that they accurately describe the data.

Chapter 4

Fast calculation of prompt photon production cross-sections with APPLGRID

Introduction

Grids are used in a variety of fields, from simply representing data differently to managing data or performing complex calculations more efficiently. This chapter is dedicated to an in-depth examination of the gridding method, with the objective of enabling these grids to be used in a global analysis of parton distribution functions, particularly within the context of the nCTEQ++ framework.

In Sec. 4.1, we will introduce APPLGRID [248], starting with a general description in Sec. 4.1.1. Next, in Sec. 4.1.2, we will develop a simple case study to establish all the necessary definitions for the following Sec. 4.1.3, where we will apply them to our process, the prompt photon production. Finally, Sec. B.1 gives the computer definition of a grid, as APPLGRID defines it.

Turning to Sec. 4.2, we will use the already available interface of APPLGRID within MCFM in Sec. 4.2.1 and show the results of these MCFM grids in Sec. 4.2.2.

Finally, in Sec. 4.3, we will present our work to obtain grids with JETPHOX. To start with, we will remind the reader in Sec. 4.3.1 of the key ingredients required in the following sections. In Sec. 4.3.2, we will describe how the interface is built and then in Sec. 4.3.3, how to use it. Next, we will explain in Sec. 4.3.4 the modifications in JETPHOX needed to produce grids and finally in Sec. 4.3.5 the results for the first grid produced by JETPHOX at LO.

4.1 General purpose of APPLGRID

4.1.1 *What is a grid?*

In our case, the grid is a look-up table of weights[1] in momentum fraction x and momentum transfer Q, organized by observable bins, order, and subprocesses.

There are different codes to produce grids, like FastNLO [249], or PINEAPPL [250]. We have chosen to work with APPLGRID because of the available interface with MCFM but, in principle, calculations could be done with any of these codes. We will present here the general method of APPLGRID and the underlying calculations.

Creating a grid for a hard scattering process is far from trivial. One can summarize the main steps as:

- Compute the wanted weights with a given Monte-Carlo generator,
- link the code with APPLGRID, using for example an appropriate interface which can write these weights into the grid,
- create a grid that can be provided to users together with a code able to read the grid and to produce fast predictions for the desired observables.

We will follow this thread for the two different codes: MCFM in Sec. 4.2 and JETPHOX in Sec. 4.3. However, we will treat the MCFM part rather as a black box and the JETPHOX part in more detail.

4.1.2 Simple case

First of all, we will present the simple case of DIS, in order to establish all the relevant definitions for the following.

The PDFs:

The PDFs can be represented by storing their values in a two-dimensional grid of points, using an n^{th} order interpolations between those points. A minimum order, $n = 3$ is required to obtain a sufficient accuracy. Due to the x and Q^2-behavior of the PDFs, APPLGRID doesn't use x and Q^2 for storing but a variable transformation $x \to y$ and $Q^2 \to \tau$ which provides a good coverage of x and Q^2 with uniformly spaced grid points :

[1] In our case, these weights are perturbative coefficients, which will be described in more detail below.

$$y(x) = \ln\frac{1}{x} + a \times (1-x) \qquad \text{and} \qquad \tau(Q^2) = \ln(\ln(\frac{Q^2}{\Lambda})). \tag{4.1}$$

The parameters a and Λ can be chosen by the user, as well as the equations themselves. For the x-transformation defined in (4.1), the parameter a is used to intensify the number of points in the large-x region.

The PDFs are then numerically approximated by

$$f(x,Q^2) = \sum_{i=0}^{n}\sum_{j=0}^{n'} f_{k+i,\kappa+j} I_i^{(n)}\left(\frac{y(x)}{\delta y} - k\right) I_j^{(n')}\left(\frac{\tau(Q^2)}{\delta\tau} - \kappa\right), \tag{4.2}$$

where $f_{k+i,\kappa+j}$ is the value of the PDF at the node $(k+i,\kappa+j)$ and n, n' are the interpolation orders which are not necessarily the same and k and κ are the index of the first grid point in y and τ respectively, used for the local interpolation. The interpolation functions $I_i^n(u)$, respectively $I_j^{n'}(u)$ are equal to 1 if $u = i$,[2] otherwise

$$I_i^{(n)}(u) = \frac{(-1)^{n-i}}{i!(n-i)!)}\frac{u(u-1)\cdots(u-n)}{u-i}. \tag{4.3}$$

The Weights:

Let's consider a Monte Carlo generator that produces N events. Each event has an (x,Q^2) value and a weight w , respectively: x_m, Q_m^2 and w_m for $m = 1,\ldots,N$. The final result (for one subprocess) of the Monte Carlo integration can be obtained by:

$$W = \sum_{m=1}^{N} w_m(x_m,Q_m^2)\left(\frac{\alpha_s(Q_m^2)}{2\pi}\right)^{p_m} f(x_m,Q_m^2), \tag{4.4}$$

where p_m is the power of α_s in the event m, and f is the PDF for the flavor considered in the subprocess.

We introduce here the weight grid $W^p_{i_y,i_\tau}$ which is updated by the interpolated weights :

$$W^{p_m}_{k+i,\kappa+j} \longrightarrow W^{p_m}_{k+i,\kappa+j} + w_m I_i^{(n)}\left(\frac{y(x_m)}{\delta y} - k(x_m)\right) I_j^{(n')}\left(\frac{\tau(Q_m^2)}{\delta\tau} - \kappa(Q_m^2)\right). \tag{4.5}$$

It is important to note here that this step increases the time of the calculation. This is the price to pay for modularity. At the end, we don't fill the grid with the originally calculated weights but with their interpolated values, that means at least $n \times n'$ points for one subprocess.

[2]It means we don't need to interpolate, as we're already on a grid node.

We can finally retrieve the initial result of the Monte Carlo program :

$$W = \sum_p \sum_{i_y} \sum_{i_\tau} W^p_{i_y,i_\tau} \left(\frac{\alpha_s(Q^{2(i_\tau)})}{2\pi} \right)^p f(x^{(i_y)}, Q^{2(i_\tau)}) \,, \tag{4.6}$$

where

$$y(x^{(i_y)}) = i_y \delta y \qquad \text{and} \qquad \tau(Q^{2(i_\tau)}) = i_\tau \delta\tau \,. \tag{4.7}$$

At this stage, note that we did not say explicitly what is the weight W because it depends on which observable one wants to compute.

In the case of two incoming hadrons, two different PDFs enter in the calculation and it is then needed to add another dimension for x_2. We obtain :

$$W = \sum_p \sum_{l=0}^{n_{subprocess}} \sum_{i_{y_1}} \sum_{i_{y_2}} \sum_{i_\tau} W^{(p)(l)}_{i_{y_1},i_{y_2},i_\tau} \left(\frac{\alpha_s(Q^{2(i_\tau)})}{2\pi} \right)^p F^{(l)}(x^{(i_{y_1})}, x^{(i_{y_2})}, Q^{2(i_\tau)}) \,, \tag{4.8}$$

where $F^{(l)}$ is just a combination of the two initial PDFs, which takes advantage of symmetries in the matrix elements if they exist.

The Scales:

Since we have calculated the perturbative coefficients order by order in α_s, it is also possible, thanks to the DGLAP evolution equations, to vary the factorization and renormalization scales. In APPLGRID, this variation is possible only by a multiplicative factor of the original scale:

$$\begin{aligned} W(\xi_R, \xi_F) = \sum_{i_y} \sum_{i_\tau} \Bigg\{ & \frac{\alpha_s(\xi_R^2 Q^{2(i_\tau)})}{2\pi}^{p_{LO}} W^{(p_{LO})}_{i_y,i_\tau} f(x^{(i_y)}, \xi_F^2 Q^{2(i_\tau)}) + \\ & \left(\frac{\alpha_s(\xi_R^2 Q^{2(i_\tau)})}{2\pi} \right)^{p_{NLO}} \Big[(W^{(p_{NLO})}_{i_y,i_\tau} + 2\pi\beta_0 p_{LO} \ln \xi_R^2 W^{(p_{LO})}_{i_y,i_\tau}) f(x^{(i_y)}, \xi_F^2 Q^{2(i_\tau)}) \\ & - \ln \xi_F^2 W^{(p_{LO})}_{i_y,i_\tau} (P_O \otimes f)(x^{(i_y)}, \xi_F^2 Q^{2(i_\tau)}) \Big] \Bigg\} \,, \end{aligned} \tag{4.9}$$

where β_0 is the first coefficient of the perturbative expansion of α_s, as defined in Eq. (1.11).

The Center of Mass (CoM) energy:

The weight grid is calculated at a specific energy $\sqrt{s}$. Even so, it is again possible to

calculate the cross-section at a different CoM energy $\sqrt{s'}$, using the transformation

$$x_i' = \frac{\sqrt{s}}{\sqrt{s'}} x_i \qquad \text{and} \qquad W'(\xi_R, \xi_F) = \frac{s}{s'}(\xi_R, \xi_F)\,. \tag{4.10}$$

The following two cases appear and require attention :

- $\sqrt{s'} < \sqrt{s}$: in that case $x_i > 1$ can occur and the PDFs are set to 0,
- $\sqrt{s'} > \sqrt{s}$: in that case, x_i can go down to values that are not enough or at all sampled in the original calculation. And above all, kinematic cuts in the original calculation can make a whole region of interest for the new kinematics disappear.

These two behaviors can be less important if the two energies are not too far apart. For example, in our case, if we want to use the grid produced at $\sqrt{s} = 8$ TeV for $\sqrt{s} = 8.16$ TeV, the ratio tends to 99%, so the coverage of the region of interest (in x and Q) can remain relevant, as long as one avoids the edges and the grid is sufficiently filled across its whole width.

4.1.3 Application to direct photon production

Now that we've outlined each ingredient, we will extend this to the direct photon process at LO and then at NLO. In this case, the final result of the MC algorithm is given by

$$\begin{aligned} W(\xi_R, \xi_F) = \sum_{l=0}^{n_{sub}-1} \sum_{i_{y_1}} \sum_{i_{y_2}} \sum_{i_\tau} \Bigg\{ & \frac{\alpha_s(\xi_R^2 Q^{2(i_\tau)})}{2\pi}^{p_{LO}} W_{i_{y_1}, i_{y_2}, i_\tau}^{(p_{LO})(l)} F^{(l)}(x^{(i_{y_1})}, x^{(i_{y_2})}, \xi_F^2 Q^{2(i_\tau)}) + \\ & \left(\frac{\alpha_s(\xi_R^2 Q^{2(i_\tau)})}{2\pi}\right)^{p_{NLO}} \Big[(W_{i_{y_1}, i_{y_2}, i_\tau}^{(p_{NLO})(l)} + 2\pi\beta_0 p_{LO} \ln \xi_R^2 W_{i_{y_1}, i_{y_2}, i_\tau}^{(p_{LO})(l)}) F^{(l)}(x^{(i_{y_1})}, x^{(i_{y_2})}, \xi_F^2 Q^{2(i_\tau)}) \\ & - \ln \xi_F^2 W_{i_{y_1}, i_{y_2}, i_\tau}^{(p_{LO})(l)} (F_{q_1 \to P_0 \otimes q_1}^{(l)}(x^{(i_{y_1})}, x^{(i_{y_2})}, \xi_F^2 Q^{2(i_\tau)}) + F_{q_2 \to P_0 \otimes q_2}^{(l)}(x^{(i_{y_1})}, x^{(i_{y_2})}, \xi_F^2 Q^{2(i_\tau)})) \Big] \Bigg\}\,. \end{aligned} \tag{4.11}$$

Note that using APPLGRID can be done in different ways. Creating an interface which is finally a third code is not the only solution. The APPLGRID code can also be integrated directly into the original code, then using the APPLGRID functions. However, it is more flexible to use the interface and a good example with MCFM is available. We found it easier to start from this already existing interface and adapt it to JETPHOX[3]. This is why the two interfaces will have the same architecture even if the JETPHOX one is very streamlined since we work with a smaller number of processes.

[3]Another advantage is that this interface could be extended to the entire family of PHOX codes [208].

4.2 Grids with MCFM

The goal of this section is to use APPLGRID in the case of MCFM where an interface is already existing. Once we have control of this we will be ready to turn to the production of APPLGRID grids using JETPHOX in the following section. In order to get started with grid-based calculations with MCFM, we decided to first reproduce the results of Fig. 4.1 from [219]. This figure shows the ratio of theory predictions normalized to ATLAS data for prompt photon production in p-p collisions at 8 TeV as a function of the transverse energy of the photon. Note that a detailed comparison to these data with the exact NLO predictions from JETPHOX was already performed in Sec. 3.3.1 (see Fig. 3.14). Here, our goal is to obtain grid-based MCFM theory predictions which will be shown below in Fig. 4.2.

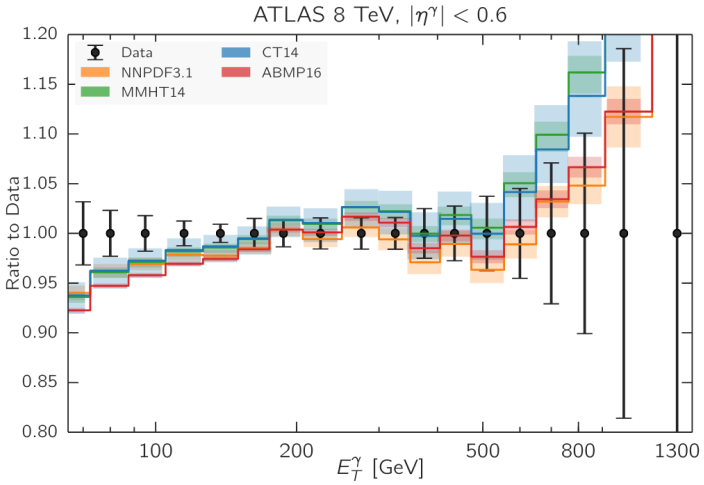


Figure 4.1: Ratio of theory predictions normalized to ATLAS data for prompt photon production in pp-collisions at 8 TeV, as a function of the γ-transverse energy. Figure taken from [219] (part of Fig. 4.1), which serves as a baseline for the comparison in Fig. 4.2.

4.2.1 MCFM interface

An open-source interface of APPLGRID to MCFM is available on the APPLGRID website[4]. The interface (called *mcfm-bridge*) was designed for MCFM v6.8. For this purpose, it is necessary to add the patch *mcfm-patch* provided on the APPLGRID website to MCFM v6.8. This version has been tested to work with version 1.5.35 of APPLGRID. The direct photon process (id: 280) contains an error in the downloadable version 6.8 of MCFM such that it was necessary to add a second patch provided by Juan Rojo[5], one of the

[4] https://applgrid.hepforge.org/
[5] Private communication.

authors of [219] and a member of the NNPDF collaboration. After installing the bridge code and adding the prompt photon process 280 to the interface the latter is ready to use.

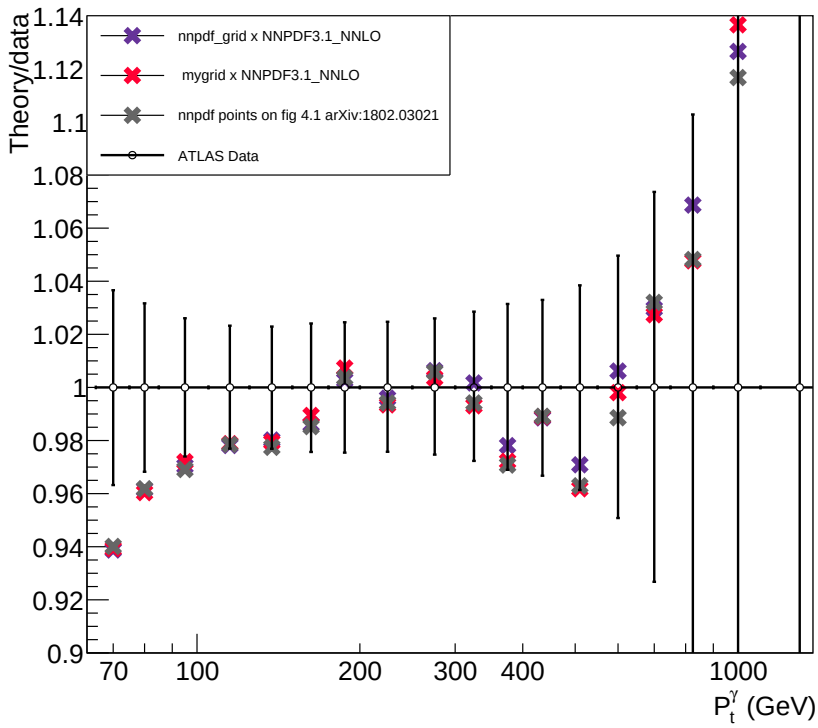


Figure 4.2: The predictions made in Fig. 4.1 of [219], using a data extraction software (PlotDigitizer) are in gray (for details see the main text). The predictions using the grid computed by the NNPDF collaboration, publicly available on GitHub, convoluted with the NLO NNPDF3.1 proton PDF [251] are in blue-violet. Using the same inputs, the predictions using our grid are in red.

The grid-based MCFM predictions at NLO, employing the publicly available grids[6] used in [219], are shown in Fig. 4.2. The results from the convolution between the NLO grids and NNLO NNPDF 3.1 proton PDFs, multiplied by the NNLO K-factor and the electroweak corrections provided in Fig. 3.1 in [219], are shown as violet-blue crosses. Furthermore, the original predictions by Campbell *et al.* in Fig. 4.1 (see orange curve) are shown as gray crosses. Finally, the red crosses show the results of our own grids, generated with the same MCFM input file (also available on GitHub), convoluted with the same NNLO NNPDF 3.1 proton PDFs and multiplied by the same K-factor and electroweak corrections. As can be seen, the difference between the red and violet-blue crosses is less than 1% with exception of the third last bin at $750 < p_T^\gamma < 900$ GeV where a 2% difference can be observed. All in all, the results obtained with the two grids are in very good agreement where the remaining differences are likely due to statistics since

[6] https://github.com/NNPDF/applgrids/tree/377613af0e2f98ec77bc501990d1df862c68f2c4/ATLASPHT12

our grids have been produced with a different number of events[7]. Another reason could be that the grid used by Campbell et al. for predictions in Fig 4.1 of [219] is not exactly the one given publicly on git, but an equivalent grid. This reason is suggested because in principle, the gray and violet points in Fig. 4.2 should match to within the deviation generated by a tool such as PlotDigitizer, which is less than 0.5% in this case. From this point on, we consider that the production of APPLGRID grids with the MCFM interface is under control.

NOTES ON THE ANALYSIS CARRIED OUT BY NNPDF AND THEN NNNPDF

The initial analysis by NNPDF [219] is based on an NNLO analysis [195] including electroweak (EW) and Sudakov corrections [252, 253]. These corrections are required in the NNLO calculation since without them the theoretical predictions no longer agree with the data, as it is shown in Fig. 5 of [195]. However, these EW effects examined in [195] (for the kinematics in Fig. 4.2) are only relevant at large p_T, i.e., $p_T > 100$ GeV [253]. Moreover, it is important to note that these analyses use a smooth-cone[8] isolation, which *completely* suppresses the fragmentation contribution in the cross-section, which is therefore not accounted for. Nevertheless, isolation with the Frixione criterion requires several parameters that have been chosen to mimic the result obtained by JETPHOX in [243], i.e. including photon fragmentation, and the experimental requirement for isolation. In principle, it is desirable for the ease of the calculations to avoid the fragmentation part, especially at NNLO where the NNLO contribution of the fragmentation part is not yet known. That being said, the NNLO calculations performed in the NNPDF analysis [219] are based on NLO calculations of direct photon production by MCFM, resulting in a NLO grid, which contains only direct contributions, to which a K-factor (for NNLO corrections) and electroweak corrections are then applied.

On the other hand, the complete analysis led by D. d'Enterria and J. Rojo in [186], which was carried out entirely with JETPHOX at NLO for direct and fragmentation contributions, demonstrated the potential to constrain the gluon PDF with isolated photon data from p-p runs at the LHC up to 7 TeV[9].

One could then reasonably argue that direct NLO calculations performed using MCFM, which account for fragmentation effects through the specific choice of smooth cone pa-

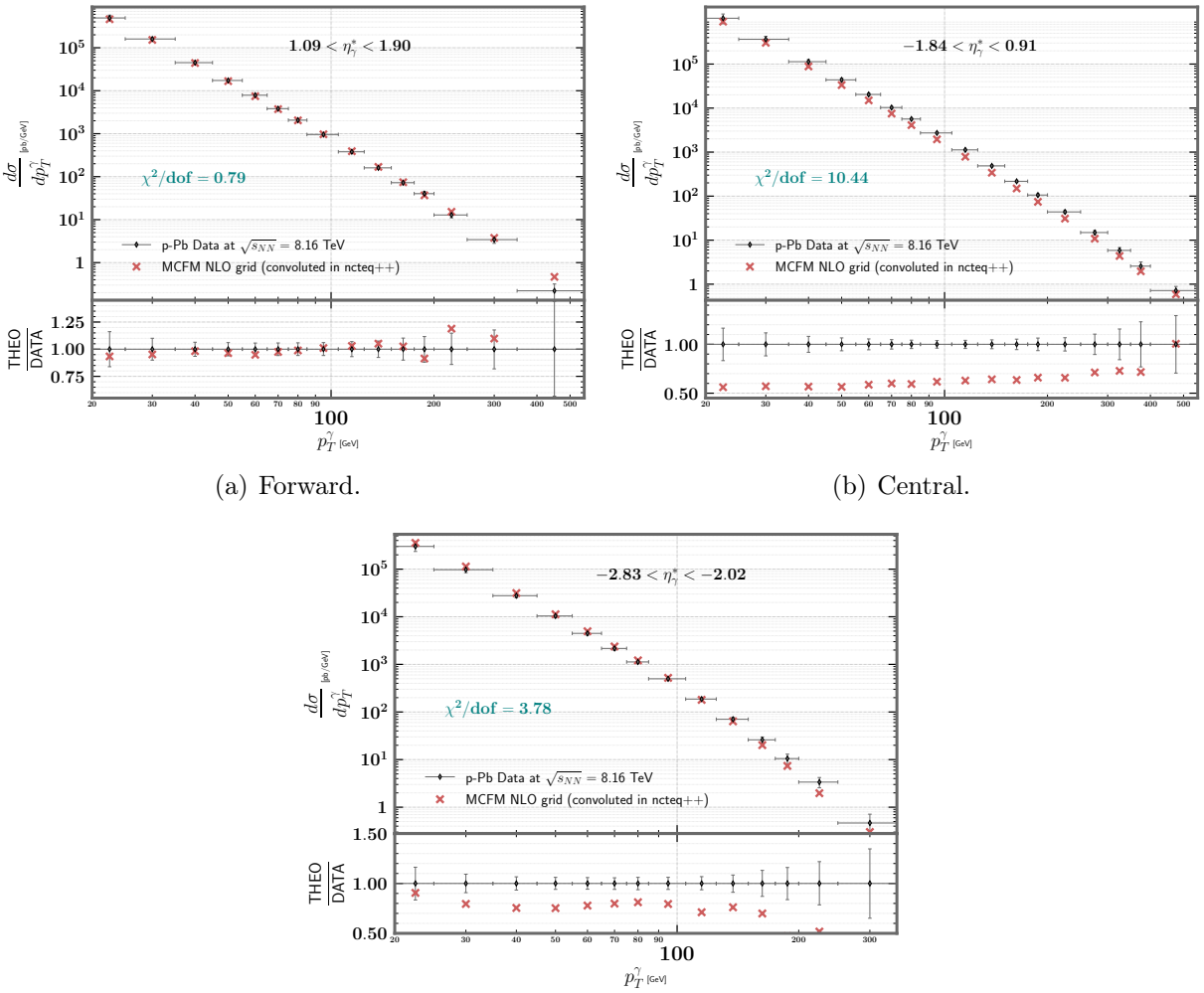
[7] We have used Nevents $= 8 \times 10^7$ for one rapidity bin in a single batch, whereas Campbell *et al.* have used 10 batches with different random seeds and combined the different grids [219] for a resulting grid that contains 3×10^8 events.

[8] The smooth-cone procedure or Frixione isolation [254] is derived from the cone approach (as in JETPHOX) with the aim of minimizing the fragmentation contribution in the isolated cross-section. The smooth-cone procedure involves smooth-cone parameters and implies that photons which come from the fragmentation can be neglected in QCD.

[9] The data at higher energies, as 8 or 13 TeV were not then available to test.

rameters (n, ϵ_γ and R_0) and without EW corrections, may provide a sufficiently precise theoretical prediction for our upcoming NLO analysis. This approach was indeed adopted in the first attempt to include direct photon data from *p-Pb* collisions by the nNNPDF collaboration in [62], whose analysis was also carried out at NLO. In the following section, we will therefore generate grids using MCFM in the same manner as described above, but adapted to the kinematic conditions of *p-Pb* collisions. These predictions will then be compared with the ATLAS data [187].

4.2.2 Grids with MCFM for *p-Pb* collisions



(a) Forward.

(b) Central.

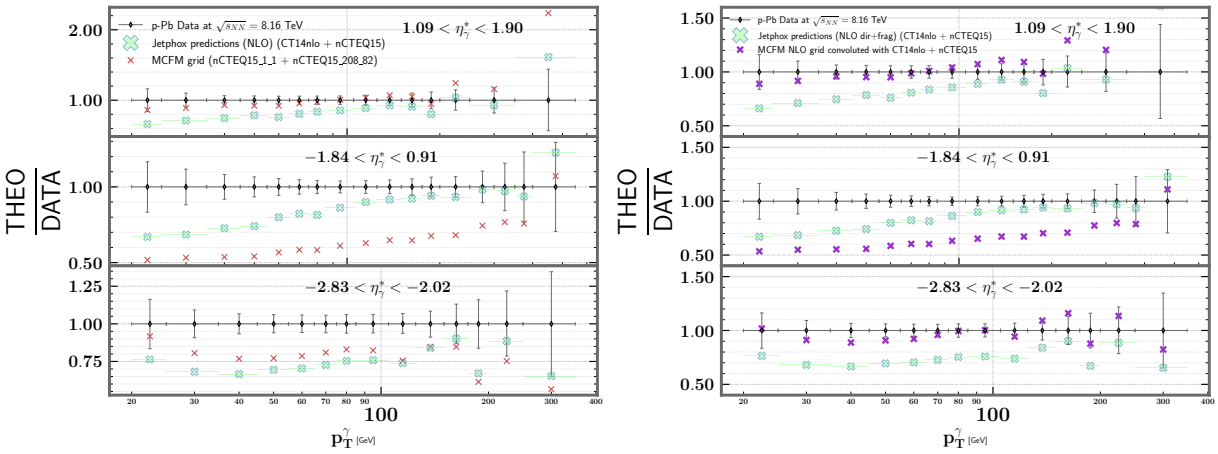
(c) Backward.

Figure 4.3: ATLAS data [187] for *p-Pb* collisions, compared to MCFM grid predictions. The upper panel shows the absolute distribution and the lower panel shows the ratio of the theoretical prediction over data.

The results of our grid-based MCFM predictions at NLO are shown in Fig. 4.3. The differential cross-sections for three rapidity bins; forward, central and backward rapidity,

respectively in Fig. 4.3(a), Fig. 4.3(b) and Fig. 4.3(c) are shown in the upper panel and the ratio of theory predictions over data are shown in the lower panel. As can be seen, the agreement between the experimental data and the theory predictions is quite satisfactory for the 6 orders of magnitude. However, the ratios to the data show a rather variable grid quality.

4.2.3 Comparison of MCFM with JETPHOX



(a) nCTEQ15_1.1 ⊗ nCTEQ15_208.82

(b) CT14nlo ⊗ nCTEQ15_208.82

Figure 4.4: Comparison of the ratio for *p-Pb* data with the MCFM grid, for two different PDFs for the proton; nCTEQ15_1.1 for 4.4(a), CT14nlo for 4.4(b).

Fig. 4.4 is the result of the convolution of the NLO grid produced with MCFM and the nCTEQ15 PDF, in order to compare with Fig. 3.16(a). The results seem similar and it seems impossible to say which of the two theoretical calculations would be more appropriate based on just these two plots. As a comparison, we find a χ^2/dof of **0.79(10.44)[3.78]** for the MCFM grid for the forward(central)[backward] rapidity. As expected, the results are equivalent to the analysis made by NNNPDF in [62], resulting in a rather poor χ^2 ; χ^2/dof = 9.1(10.5),[8.5] - the same recipe, the same result. With JETPHOX predictions, we found χ^2/dof = **7.17(8.50)[13.92]**, which are slightly better values for the two first rapidity ranges.

Conclusions/Recap

Several options are available to us:

1) Continuing with MCFM and APPLGRID using a standard isolation procedure and applying a fragmentation correction factor to see if it improves the results. The issue is that we have never managed to achieve a match of the direct contribution between JETPHOX and MCFM, even when using identical parameters.

Applying a correction derived from JETPHOX to an MCFM calculation appears unjustified unless a near-exact correspondence between the two frameworks can be established. This would also require a complete analysis of the fragmentation contribution, including a parametrization that describes its behavior as a function of the kinematics, FFs used, and the specific isolation criterion.

2) Following the approach taken by nNNPDF in [37], that is, using the R_{pPb} ratios, which can be well described, with a total χ^2/dof =1.03 across the three rapidity bins [10]. We could use a similar technique and hope for comparable results. This approach is promising, as implementing ratios helps to cancel out many theoretical and experimental uncertainties. However, it also introduces the free-proton PDF, which itself comes with its own uncertainties.

3) Generating grids with JETPHOX to test its results and compare them with MCFM calculations in a global fit. This approach also allows us to explore whether we can directly use these predictions in the form of differential cross sections.

The next section will focus on the grid generation with JETPHOX.

[10] Individual rapidities are not provided in [37].

4.3 Grids with Jetphox

4.3.1 Introduction

As we have seen in Chap. 3, JETPHOX provides a good description of the data for *p-Pb* collisions.

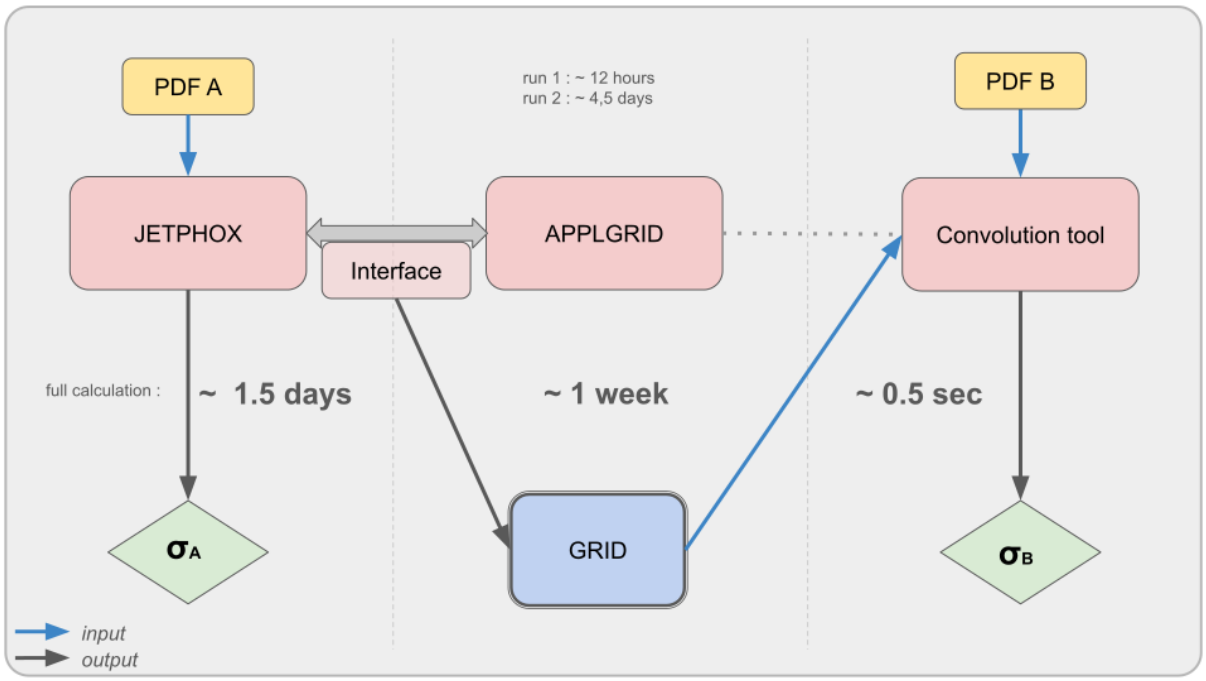


Figure 4.5: Workflow chart of the gridding project with JETPHOX.

That's why we are going to produce grids directly with JETPHOX and APPLGRID by interfacing both codes. We can summarize the workflow in Fig. 4.5. The interface between JETPHOX and APPLGRID is used to collect the *weights*, created by JETPHOX, needed to create the grids. These weights are distributed and ordered in the grid. Finally, the grid can be used to make fast predictions in any other code through a *convolution tool*. As an example, a complete calculation for one rapidity bin, with scale and PDF uncertainties, for about fifteen p_T-bins, takes about 1.5 days. The grid is produced in two runs, the first one (optimization run) consists of defining the boundaries of the grid. The second run consists of filling the grid with the interpolated coefficients and is therefore particularly time-consuming. The calculation time for these two steps is at least several days for the direct contribution at NLO with MCFM and even weeks with JETPHOX. Finally, convoluting the grid with any set of PDFs takes less than a second. To highlight key elements and get a clear picture for the next sections, we rewrite here the Eq. (4.4) of a

Monte Carlo integration, which would apply to the p-$p \to \gamma + X$ case:

$$W = \sum_{l=1}^{n_{sub}} \sum_{m=1}^{N} w_m \left(\frac{\alpha_s(Q_m^2)}{2\pi} \right)^{p_m} F^{(l)}(x_{1,m}, x_{2,m}, Q_m^2) \,, \tag{4.12}$$

as well as Eq. (4.8)

$$W = \sum_{p}^{n_{subprocess}} \sum_{l=0}^{} \sum_{i_{y_1}} \sum_{i_{y_2}} \sum_{i_\tau} W^{(p)(l)}_{i_{y_1}, i_{y_2}, i_\tau} \left(\frac{\alpha_s(Q^{2(i_\tau)})}{2\pi} \right)^{p} F^{(l)}(x^{(i_{y_1})}, x^{(i_{y_2})}, Q^{2(i_\tau)}) \,, \tag{4.13}$$

which allows to recover W with the grid. We also repeat here the corresponding equation for direct photon production at LO (rewritten from Eq. (3.21))

$$\frac{d\sigma^{\mathrm{DIR}}}{d\overrightarrow{p_T} d\eta} = \sum_{i,j=q,\bar{q},g} \int dx_1 dx_2 \; F_{i/H_1}(x_1, M) \; F_{j/H_2}(x_2, M) \; \frac{\alpha_s(\mu_R)}{2\pi} \frac{d\widehat{\sigma}_{ij}}{d\overrightarrow{p_T} d\eta} \,. \tag{4.14}$$

In the following section, we will describe briefly how we construct the interface, and how it works with JETPHOX.

4.3.2 Interface and installation

Most of the JETPHOX code is written in FORTRAN but also contains a small ROOT interface, written in C++. APPLGRID is entirely coded in C++[11]. The interface between the two previous codes is built via the suite of free software available from GNU organization [255]: `libtool`, `autoconf`, `automake`. This step was carried out only once to create the interface, after which it can be installed by the user in the usual way with the

```
$ ./configure
$ make
$ make install
```

commands. The interface then contains three main parts after compilation. Firstly, it contains three main files: `jetphox_grid.h`, `jetphox_grid.cpp` and `jetphox_interface.cpp`. The first one defines the class `appl:jetphox_grid()`, inherited from the main class `appl:grid()` from APPLGRID. It also contains the definitions of the different variables and structures needed for a grid as well as functions for using these variables in the FORTRAN part of the code (through the `extern "C"` command) and the declaration of three functions used in the interface: `fillJetphox()`, `collectWeight()` and `decideSubProcess()`. The `jetphox_grid.cpp` then implements these three functions, appropriately adapted to JETPHOX. This part of the code is designed to JETPHOX and written

[11] The APPLGRID code also contains its own FORTRAN interface, not used here.

once; it is not supposed to be changed by the user when producing different grids. The `jetphox_interface.cpp` file is the one to be modified for each grid creation. It contains the three main functions `book_grid()`, `fill_grid()` and `write_grid()` which we will call in the Fortran part of JETPHOX during the process of event creation[12].

Secondly, a (static) library which contains all the previous functions is built. And thirdly, the interface provides a 'config' tool (`jetphoxbridge-config`), making it easier to implement APPLGRID compilation elements to the Makefile of JETPHOX.

4.3.3 Using the interface

Once the interface is correctly installed, a few steps are required to launch grid creation. First of all, the user needs to enter all the information required to create the skeleton of the grid, in `jetphox_interface.cxx`. It is possible to produce several (`_Ngrids`) grids with several (`_NOBSERVABLES`) observables simultaneously, but here we remain in a simple case with a single observable, i.e. the p_T distribution of the photon, in a single grid named `"pp_collision_TEST.root"`. Then the binning to be applied to the observable(s) is set (here, 6 bins in p_T):

```
static const int _Ngrids = 1;
static const int _NOBSERVABLES = 1;
appl::jetphox_grid* mygrid[_Ngrids];
static const char* gridFiles[_Ngrids] = {"pp_collisions_TEST.root"};
double _pt[7] = {25,35,45,55,65,75,85};
```

Next, in the `book_grid()` function, a few basic parameters are required:

```
// binning information for the grid constructor
  double xLow    = 1.0e-6, xUp = 1.0;
  int    nXbins  = 20;
  int    xorder  = 4;

  double q2Low   = 1.0*1.0;
  double q2Up    = 1.01e8;
  int    nQ2bins = 20;
  int    qorder  = 4;
```

The values `xLow, xUp` are the limits in x wanted by the user and `nXbins, xorder` are the number of bins (before the optimization step) and the order of the interpolation, that

[12]For JETPHOX users, a parallel can be drawn here with JETPHOX's method of creating the NTuple via the functions `init_root_ntuple(), remplie_ntuple(), end_root_ntuple()`.

corresponds to the n in Eq. (4.2) and Eq. (4.3). Similarly, `q2Low, q2Up` are the limits in Q^2, and `nQ2bins, qorder` are the number of Q^2-bins and the order of the interpolation, that corresponds to the n' in Eq. (4.2). Finally, the following variables have to be specified:

```
double apramval=0.;
appl::grid::transformvar(apramval);

int lowest_order = 1;
int nloops = 0;
std::string pdf_function;
pdf_function = "basic"; //"jetphoxLO.config"
```

The variable `apramval`, the a value of Eq. (4.1), is set to 0 here because we're not particularly interested in *extra* filling the region at large-x, but rather in examining the region at small-x. Of course, this can be modified if the statistic is not sufficient at large-x.

The final cross-section grid is internally separated, by orders in α_s, into (`1+nloops`) grids. The `lowest_order` parameter corresponds to the power p in Eq. (4.13). It is set to 1 since the direct photon production is proportional to α_s at LO (unlike DIS where p would be equal to 0 for example). The `pdf_function` contains the name of the config file, used to combine the two initial PDFs and to define the number of subprocesses ($n_{subprocess}$ in Eq. (4.13)). The `"basic"` config corresponds to the general case with all flavor combinations, regardless of the potential symmetries between subprocesses we're working with. For example, for $N_f = 5$, it produces 121 combinations, which can in principle be reduced to 11 combinations at LO with `"jetphoxLO.config"`[13]. At this stage, the minimum necessary elements have been defined, allowing for an initial sampling of the phase space. The aim of this first run is to perform a cross-section calculation with many events, in order to optimize the grid for the second run. This first run doesn't interpolate the grid points, it just remembers the limit values. Once this step has been completed, a second run must be launched, which will read the grid optimizations and this time fill the grid with all the weights $W^{(p)(l)}_{i_{y_1},i_{y_2},i_\tau}$ (Eq. (4.13)).

4.3.4 Modifications in JETPHOX

Initially, JETPHOX can produce NTuples and histograms, so we have added a third output option `"grid"` in the `parameter.indat` file (line 42 of Appendix A.1).

[13] This config file exists but we did not test it yet.

Since we are trying to reconstruct the observable $d\sigma^{\mathbf{DIR}}/d\overrightarrow{p_T}d\eta$, we assume

$$\frac{d\sigma^{\mathbf{DIR}}}{d\overrightarrow{p_T}d\eta} = W, \tag{4.15}$$

and then Eq. (4.14) and Eq. (4.12) are equivalent.

The first thing to do is to identify/find in JETPHOX the corresponding w_m of Eq. (4.12). Comparing these equations it is clear that

$$w_m = \frac{d\widehat{\sigma}_{ij}}{d\overrightarrow{p_T}d\eta}\,. \tag{4.16}$$

In order to identify all the ingredients of these partonic cross-sections in the JETPHOX code, one has to modify the `subroutine sfborn()` where initially, the combinations of PDFs are included at an early stage of the calculation. Once having identified the right coefficients, one has to decouple the combinations of matrix elements (inherited from the symmetry of the process), in order to correspond to the basic decomposition given in the config file of APPLGRID. As a starting point, it is useful to verify the result for a single channel at a time. For this purpose we have constructed a Fortran `module appl_grid` containing a `subroutine weight_for_grid(x1,x2,k0max_o,jnf,sec_part,weightb)`. This module returns a 2-dimensional matrix (`weightb(i,j)` with index i, j corresponding to the incoming parton flavor. These weights (`weightb`) are then multiplied by the required factors given in Eq. (3.32) to build the observable in (4.15).

Secondly, a specific feature of JETPHOX must be taken into account. As mentioned before in Sec. 3.2.1.3, the event generator (SPRING) is initially designed to generate events with a weight of 1 and these weights are put in a histogram to which an overall normalization (σ_{tot}/N, see Eq.(3.45)) will be applied. The problem is that this normalization contains the total integrated cross-section, and this depends explicitly on the PDFs and α_s used. This conflicts with the APPLGRID method. Therefore, we have created a small event generator module (`my_spring()`) to overcome this problem. This module is presented in Appendix B.2.

4.3.5 Results for the Born term

In Fig. 4.6, we compare the LO order cross-section for prompt photon production in p-p collisions at 8 TeV[14], calculated with JETPHOX to the corresponding result obtained with the JETPHOX grid as a function of the transverse momentum p_T of the photon. We have

[14]The kinematics has already been used in Sec. 3.3.1.

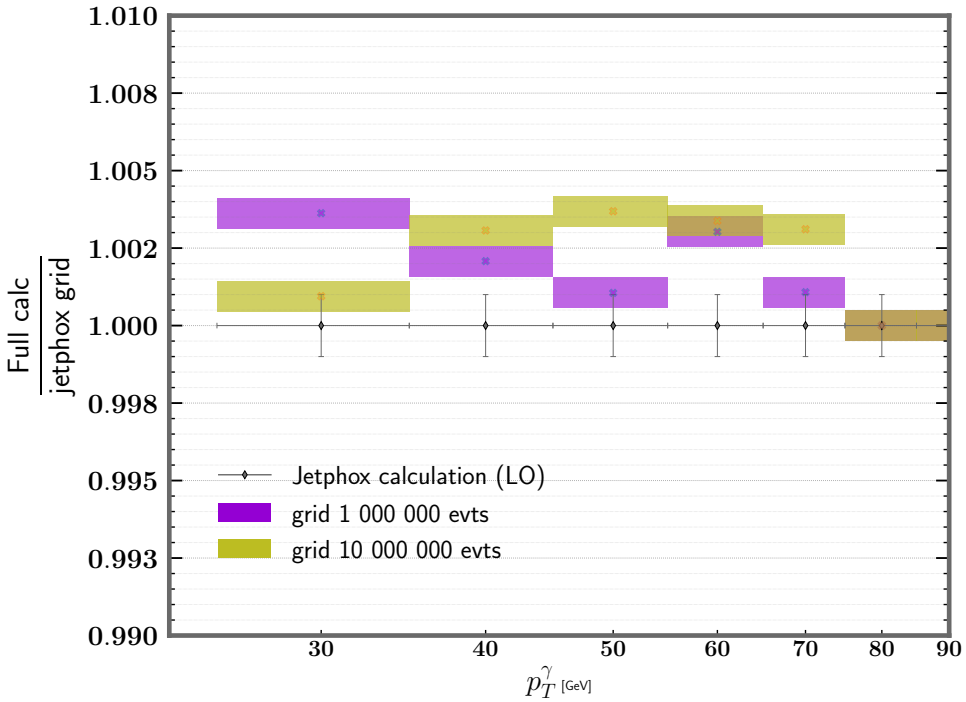


Figure 4.6: Ratio of a JETPHOX full calculation with the a JETPHOX-grid, for two different numbers of events: 10^6 in violet and 10^7 in olive. We present here the six first bins of the kinematics used in Sec. 3.3.1.

used two grids, one created with 10^6 events (in violet) and another one with 10^7 events (in olive). As can be seen in both cases (olive and violet) the fast grid calculations agree with the exact result to within 5 ‰ or better over the entire p_T-range shown. There is no visible improvement of the grid created with 10^7 events.

4.3.6 Outlook for NLO grids and fragmentation

The previous section provides the basis necessary for creating a grid with JETPHOX, and serves as proof of concept. The rest of the work required to achieve NLO calculations will essentially consist of porting the weights defined in Eq. (4.11) for all the JETPHOX functions defined in Table 3.1 for the direct part and their equivalents for fragmentation part. However, for fragmentation, there are two possibilities with different degrees of practicability. Firstly, it would be possible to process only the direct part using a grid and then apply a corrective factor, similar to a K factor. It would then be necessary to verify that the error introduced is very small compared to the complete calculation. Another possible option, which would be even more interesting, would be to add a degree of freedom to equation Eq. (4.11), corresponding to the function $D_{\gamma/k}(z, \mu_{Frag})$ in Eq. (3.22). This would allow calculations to be made for different fragmentation functions.This option would take much longer to implement because it requires modifying and adding new

elements to APPLGRID.

Finally, we can summarize the situation as follows: the first solution is practical and quick, as long as the error introduced is reasonable. The consequence is reduced modularity. The second option is more ambitious and offers greater modularity, but is significantly longer and more complex.

Conclusions and Outlook

The central aim of this thesis has been to explore and strengthen the connection between the theoretical framework of perturbative Quantum Chromodynamics, the phenomenology of hard processes in nuclear collisions, and the practical computational tools required for global analyses of PDFs and nPDFs. By moving from the formal aspects of factorization theorems, through the specific case of target mass corrections in DIS, and finally to the collider phenomenology of prompt photon production and its implementation in fast grid techniques, this work has sought to provide a coherent and useful contribution to the precision study of the partonic structure of nuclei. Chapter 1 provided the theoretical foundations underlying the thesis.

Chapter 2 was devoted to a detailed study of TMCs in lepton–nucleus DIS. Starting from the operator product expansion, we reviewed the derivation of master formulas for TMCs and discussed their physical meaning in the nuclear case. Particular care was taken to clarify the relationship between OPE-based treatments and parton model approaches. We studied the impact of quark masses, nuclear kinematics, and the behavior of nuclear PDFs at large-x and low-Q. Numerical results for both proton and nuclear structure functions demonstrated the quantitative relevance of TMCs and provided a *universal* parametrization for practical applications. In this way, the thesis contributed to a systematic understanding of nuclear TMCs and their implementation in phenomenological analyses.

In Chapter 3, the focus shifted to the production of prompt photons in proton-proton and proton-lead collisions, an important process for probing the gluon distribution in nuclei. We reviewed the theoretical ingredients of photon production, including the direct and fragmentation components, isolation cuts, and higher-order corrections. Using JETPHOX and MCFM, we performed calculations at various c.o.m energies and compared them to available experimental data. The results demonstrated that both JETPHOX and MCFM provide reliable predictions in the relevant kinematic regions, while also highlighting the limitations imposed by fragmentation functions and scale uncertainties. This study confirmed the potential of prompt photons as precision probes of nuclear PDFs.

Finally, Chapter 4 presented the construction and validation of fast grid techniques for prompt photon cross-sections using APPLGRID. We started to use the existing MCFM-APPLGRID interface, but more importantly we interfaced JETPHOX with APPLGRID. We then generated grids, and demonstrated their accuracy by comparison with direct calculations. Importantly, these grids open the door to the systematic inclusion of prompt photon data into global nuclear PDF fits within the nCTEQ framework. This work thus makes a concrete methodological contribution to the future of precision nPDFs analysis.

Taken together, the results of this thesis provide contributions at three levels:

- Theoretical: Clarifying the role of TMCs in the nuclear case, and their consistency with both OPE and parton model pictures.
- Phenomenological: Validating prompt photon production as a tool for constraining nuclear gluon distributions, and benchmarking state-of-the-art codes against available data.
- Practical: Implementing and testing fast grid interfaces for photon production, enabling their use in global fits with nCTEQ++.

Outlook and Future Directions:

While this work has addressed several key questions, it also opens the way to further investigations.

The treatment of TMCs can be further developed by including higher-twist effects. This is particularly relevant in the high-x region, where higher-twist corrections and multi-nucleon correlations become significant. The formalism developed here could be extended and tested against forthcoming data from Jefferson Lab at 12 GeV and, in the longer term, the Electron–Ion Collider (EIC), where precise measurements at moderate and large-x will provide new sensitivity to these effects.

On the phenomenological side, prompt photon production will continue to serve as an essential tool for gluon constraints, particularly in nuclei. With ongoing and future LHC runs, more precise measurements in both p-p and p-Pb collisions will become available, with extended kinematic reach and reduced uncertainties. The Focal experiment will be of major importance in the coming years for this area of study. Contrasting prompt photon production to hadroproduction of heavy quarks, another gluon-sensitive processes used in global analyses of nPDFs, will allow to further test QCD factorization and other mechanisms explaining the relative suppression of heavy quark production in p-Pb collisions at small-x, such as coherent energy loss effects and/or gluon saturation.

About the integration into global analyses, the fast grids developed here represent a first step toward systematically including photon data in nuclear PDF fits.

Furthermore, a new nuclear global analysis, to be published in the coming months, is currently being finalized; **nCTEQ25**. The direct photon process has not yet been incorporated into the data analysis. However, a comparison will be made between the predictions for this process and the earlier nPDFs, as nCTEQ15 and nCTEQ15HQ, to evaluate the impact of the newer one on the theoretical predictions.

Future work will extend these grids to full NLO precision for inclusion in the nCTEQ framework. Together with other processes such as heavy quark production, Drell–Yan, and jet production, these inputs will provide stronger constraints on nuclear gluons and quarks across a wide kinematic range. The ultimate goal is to achieve global fits that combine all data from fixed-target DIS, LHC proton–nucleus collisions, and future EIC measurements in a consistent QCD framework.

Appendix A

Jetphox materials

A.1 Input file

The file *parameter.indat* contains all the parameter that can be easily modified for different analysis.

```
1  ###################################################################################
2  # * This file is treated by a perlscript, so all lines beginning by # are
3  #   not read.
4  # * The ordering of the parameters cannot be changed without changing
5  #   the  perlscript
6  ###################################################################################
7  #
8  #
9  #                  NAMES AND PATHS FOR OUTPUT FILES
10 #
11 #
12 ###################################################################################
13 # Name of the run: string which serves to form :
14 # 1) the names of the input parameter files,
15 # 2) the names of the executable files ,
16 # 3) the path to save the .bs files from the working directory into
17 #    the following directories which will be created if they do not
18 #    exist:
19 #    result.string/dir.string
20 #    result.string/onef.string
21 #    result.string/twof.string
22 # Note that only alpha-numeric characters are allowed, namely
23 # letters, numbers or _
24 #
25 #test
26 arXiv1903_02209_184_091
27 ###################################################################################
28 # Name for the histo: string which serves to form
29 # the names of the histograms or of the ntuple.
30 #test_b
31 arXiv1903_02209_184_091
32 ###################################################################################
33 # String to form the path where the ntuple or the histograms will be put
34 #
35 ../pawres/arXiv1903_02209
36 ###################################################################################
37 # Please select the format of the output: "histo" will give directly
38 # histograms, "ntuple" will produce a ntuple. Here also the Makefile is
```

```
# modified in order to load only the necessary files
#
#histo
ntuple
###################################################################################
#
#
#                          INPUT
#
#
###################################################################################
# Type of initial hadron H1:0 proton, 1 anti-proton, aaazzz nucleus
#
0
#208082
###################################################################################
# Type of initial hadron H2:0 proton, 1 anti-proton, aaazzz nucleus
#
#0
208082
###################################################################################
#                          PDF
#
# The program uses the LHAglue interface for LHAPDF. Please consult
# http://projects.hepforge.org/lhapdf/manual
# to get the LHAGLUE code which labels your pdf's set
#
###################################################################################
# LHAPDF library path
/utils/POOL_SL7/LHAPDF-6.2.1
###################################################################################
# select a PDF set using the explicit name
#
#cteq66=cteq66.LHgrid
#cteq6m=cteq6m.LHgrid
#mstw2008 nlo=MSTW2008nlo68cl.LHgrid
#cteq66.LHgrid
#MRST2004qed_proton.LHgrid
CT14nlo.LHgrid
#CT10.LHgrid
#nCTEQ15_208_82.LHgrid
###################################################################################
# call all PDF error sets (TRUE or FALSE)
#FALSE
TRUE
###################################################################################
# Choice of the nPDF set (for nuclear targets) - must be set to 0 for protons/antiprotons
# 0 no nuclear effects in the PDF
# 1 nDS
# 2 nDSg
# 3 EKS98
# 4 EPS08
# 5 HKN04
# 601-6031 EPS09
# 7 HKN07
# 8 FGS10 (set I)
# 9 FGS10 (set II)
#
101100
#0
#####################################################################
#                      FRAGMENTATION FUNCTION
###################################################################################
# Type of fragmentation functions for hadron/photon 3
#
# conventions: 4 digit number, xxyy
```

```
#
# xx = group label:
#
# photons
#           01: BFG small gluon
#          02: BFG large gluon
# hadrons
#           01: AKK08
#          02: DSS
#          03: HKNS
#          04: AKK05
#          05: BKK
#          06: KKP
#          07: Kretzer
#          08: BFGW best fit
#          09: BFGW large Ng
#          10: BFGW low Ng
#         13: KKKS08 (into heavy-quark hadrons)
#         14: AESSS (eta)
#         15: DSV (lambda)
#
# yy = hadron label:
#
#         00: photon
#         10: (pi^+ + pi^-)/2
#         11: pi^+ + pi^-
#         12: pi^0
#         13: pi^+
#         14: pi^-
#         20: (K^+ + K^-)/2
#         21: K^+ + K^-
#          22: (K^0 + K^0bar)/2
#          23: K^0 + K^0bar
#          24: K^+
#          25: K^-
#          30: (p + pbar)/2
#          31: p + pbar
#          32: p
#          33: pbar
#          40: (h^+ + h^-)/2
#          41: h^+ + h^-
#          42: h^+
#          43: h^-
#        62: eta
#        70: neutral cluster (sum of pi0 and eta)
#          80: (Lambda + Lambdabar)/2
#          81: Lambda + Lambdabar
#        90: D*
#        91: D0
#        92: D+
#        93: sum over all charmed hadrons
#
# For example:
# photons BFG set II : 0200
# h+ + h- DSS : 0241
0200
########################################################################
#                           SCALES
########################################################################
# Choice of the scale:
#       1 (pt3+pt4)*cm,
#       2 sqrt(pt3^2+pt4^2)*cm,
#       3 mgg*cm,
#       4 cm*pt_photon
#       5 cm*pt_photon*sqrt((1+exp(-2*abs(y*)))/2)
# pt3 and pt4 are the transverse momentum of the two photons, mgg is
```

```
# the invariant mass of them, y* = (y_photon-y_jet)/2
#
4
#*********************************************************************
# Value of cm for initial state factorisation scale
#
1.0D0
#*********************************************************************
# Value of cm for renormalisation scale
#
1.0D0
#*********************************************************************
# Value of cm for final state factorisation scale
#
1.0D0
#*********************************************************************
#                          COUPLING, ETC...
#*********************************************************************
# Number of loops in alpha_em:0 constant else running (routine taken from
# jetset)
#
0
#*********************************************************************
# alpha_s is obtained from:
# 0 : LHAPDF
# 1 : Numerical inversion of the solution of the 1-loop R.G equation
# 2 : Numerical inversion of the solution of the 2-loop R.G equation
#
0
#*********************************************************************
# Number of active flavours
#
5
#*********************************************************************
# Value of (hbar*c)^2 in GeV^2 pb to get cross sections in pb
#
.38935D+9
#*********************************************************************
# For the direct part:0 born only,1 box only, 2 born+box
#
2
#*********************************************************************
# Value of PTM in GeV
# Typically PTM must be of the order of 0.5 % of the PT minimum of the
# final photon/hadron
# for LHC, with PT of the photons/hadrons > 25 GeV, PTM = 0.1 GeV works
# for Tevatron, with PT of the photons/hadrons > 12 GeV, PTM = 0.05 GeV
# works
#
0.1D0
#*********************************************************************
# Value of R_th
# R_th = 0.1 or less
#
0.1D0
#*********************************************************************
# Please choose which contribution of the program you want to run:
# dir                direct contribution
# onef                one fragmentation contribution
# the items must be separated by a comma!!!!
# for example: dir,onef will run the two contributions
#
dir,onef
#onef
#dir
#*********************************************************************
```

```
# If TRUE the Higher Order corrections are calculated
# else if FALSE no H.O. corrections are computed
# the number of values here must match the number of contributions
# already selected
# for example if dir,onef has been selected and one wants to compute
# all these three contributions at NLO, you must put here
# TRUE,TRUE if you put only one TRUE the program assumes that you
# want to compute all the contribution at HO
#
TRUE
#FALSE
#*********************************************************************
# If TRUE the Leading Order terms are calculated
# else if FALSE no L.O. terms are computed
# the number of values here not necessarily match the number of contributions
# already selected
#
TRUE
#*********************************************************************
# Set to TRUE if you want to compute the physical integrated cross section
# (for example to  test the PTM or R_th dependence)
# here one assumes that the user wants to use the same option for all
# contributions
#
FALSE
#TRUE
#*********************************************************************
# Set to TRUE if you want to compute differential cross sections through a
# NTUPLE or HISTOGRAMS. It is used to make the grid. In this case, the
# program integrates the absolute value of differential cross sections.
# Note that this flag acts only on the creation of .bs files (creation
# of the grids).
# Here one assumes that the user wants to use the same option for all
# contributions
#
TRUE
#FALSE
#*********************************************************************
# If true only generation. This flag acts only on the generation of
# events just by reading the already created .bs files
# here one assumes that the user wants to use the same option for all
# contributions
#
TRUE
#FALSE
#*********************************************************************
#
#
#                         KINEMATICS
#
#
#*********************************************************************
#0 collider mode, 1 fixed target mode
#
0
#*********************************************************************
# Value of ebeam or sqrt(s) depending on the preceeding flag
#
8160.D0
#*********************************************************************
# Value of the maximal rapidity for the photon
#
0.91D0
#*********************************************************************
# Value of the minimal rapidity for the photon
#
```

```
-1.84D0
#*********************************************************************
#Value of ptmax in GeV for the photon
#
550.0D0
#*********************************************************************
# Value of ptmin in GeV for the photon
#
20.D0
#*********************************************************************
#PARAMETERS FOR ISOLATION
#*********************************************************************
# flag to choose different maximum transverse energy cut deposited in the isolation cone
#         1 value in GeV
#   2 Fraction of photon transverse momentum
#   3 crown with value in GeV (do not put more than 10 values)
#   4 nested cone with value in GeV (do not put more than 10 values) computed from the formula
#     etmax(r) = epsilon*pt*( (1-cos(r))/(1-cos(r_max)) )^(n)
#     for this case enter, instead of etmax, the parameters epsilon and n
#   5 ATLAS criterion : et_deposited < A*pt+B in a cone of size R
# if there is no isolation criterion, choose 1 and put the maximum Et allowed : sqrt(s)/2
5
#*********************************************************************
# Depending on the flag:
#   1 Radius of isolation cone (one value)
#   2 Radius of isolation cone(one value)
#   3 Radius of each crown (not more that 10 values)
#   4 Radius of each nested cone (not more that 10 values)
#   5 Radius of isolation cone (one value)
#
# if several values are required, put them as a list separated with a comma
# ex : 0.4D0,0.35D0,0.3D0,0.25D0,0.2D0,0.15D0,0.1D0
#
0.4D0
#*********************************************************************
# Depending on the flag:
#   1 value of Etmax (in GeV) (one value)
#   2 fraction of the photon transverse momentum (one value)
#   3 value in GeV in each crown (not more that 10 values)
#   4 value in GeV in each nested cone (not more that 10 values)
#   5 value of the parameters A and B (two values)
#
# if several values are required, put them as a list separated with a comma
# ex : 3D0,2.92D0,2.83D0,2.73D0,2.61D0,2.47D0,2.28D0
#
#1.D0,5.D0
#3.D0
4.2D-3,4.8D0
#*********************************************************************
#PARAMETERS FOR JETS
#*********************************************************************
# flag to choose inclusive mode or photon+jet; inclusive: 0, jets: 1
# be careful : if the ntuple option is selected, this option must be 0
# The ntuple is made of photon and partons from which the user has to reconstruct the jet
0
#*********************************************************************
# choice of jet algorithm: "kt": Durham kt, "co": cone, "se": with seeds
# or "d0": midpoint algorithm (to be used with "ho" merging rules)
kt
#*********************************************************************
# value of r_kt
0.4D0
#*********************************************************************
# value of r_cone
0.7D0
#*********************************************************************
```

```
# value of r_sep
1.4D0
#*********************************************************************
# choice of the merging rules: "sn": snow-mass, "ho": houches99
#
sn
#*********************************************************************
# Acceptance of the jet:
# an event is accepted if:
#"gp": the highest pt jet verifies the cut in rapidity and transverse
# momentum
#"up":one of the jets verifies the cut in rapidity and transverse
# momentum
#
gp
#*********************************************************************
# value of ptjet_max
1000.D0
#*********************************************************************
# value of ptjet_min
10.D0
#*********************************************************************
# value of yjet_max
3.37D0
#*********************************************************************
# value of yjet_min
-3.37D0
#*********************************************************************
#
#
#                          BASES/SPRING
#
#
#*********************************************************************
# For all the following items except for the number of tries for SPRING,
# the user can choose different numbers for the different contributions.
# If only one number is specified, the program takes this number for all
# the contributions
#*********************************************************************
# Number of generated events in the ntuple or histograms
#
#4000000 this amount worked in the trial
8000000
#100000
#400
#*********************************************************************
# Seed for the generation only
71345
#*********************************************************************
# Number of iteration for the grid step
#
20
#*********************************************************************
# Number of iteration for the integration step
#
10
#*********************************************************************
# Number of calls per iteration for two to three
#
5000000
#*********************************************************************
# Number of calls per iteration for quasi two to two
#
2500000
#*********************************************************************
# Number of calls per iteration for true two to two
```

```
#
1000000
#############################################################################
# Number of tries for spring
# here one assumes that the user wants to use the same option for all
# contributions
#
1000
#############################################################################
# Accuracy in per cent for Bases
#
0.03D0
#############################################################################
#
#
#                          TEST
#
#
# This section is only for performing tests on specific sub-processes. Do
# not modify the default values below.
#############################################################################
# To select process in direct part:
# In some routines, there is a loop over the number of partonic
# sub-processes. For the direct part, the loop runs from 1 to 2
# the user can select one sub-process by changing the lower (upper) value
# of the loop (for instance putting 1 1 for the two next values the user
# obtains results for qi + qbi only)
#        qi + qk --> jet + ph
#
#        1 : d + u --> jet + ph
#        2 : d + dp --> jet + ph
#        3 : u + up --> jet + ph
#
#        qi + qbk --> jet + ph
#
#        4 : d + ub --> jet + ph
#        5 : d + dpb --> jet + ph
#        6 : u + upb --> jet + ph
#
#        7 : qi + qi --> jet + ph
#
#        qi + qbi --> jet + ph
#        8 : d + db --> jet + ph
#        9 : u + ub --> jet + ph
#
#        10 : qi + g --> jet + ph
#
#        11 : g + g --> jet + ph
#############################################################################
1
11
#############################################################################
# To select process in one fragmentation part:
# Here the loop over the partonic sub-processes runs from 1 to 18
# D means a quark of charge -1/3 (Db anti quark D, Dp another quark of
# charge -1/3 different from D)
# U means a quark of charge 2/3 (Ub anti quark U, Up another quark of
# charge 2/3 different from U)
#        1 : qi + qk --> jet + qk
#        2 : qi + qk --> jet + g
#        3 : qi + qbk --> jet + qbk
#        4 : qi + qbk --> jet + g
#        5 : qi + qi --> jet + qi
#        6 : qi + qi --> jet + g
#        7 : qi + qbi --> jet + qbk
#        8 : qi + qbi --> jet + qbi
```

```
#         9 : qi + qbi --> jet + g
#         10 : qi + g --> jet + qk
#         11 : qi + g --> jet + qbk
#         12 : qi + g --> jet + qbi
#         13 : qi + g --> jet + g
#         14 : qi + g --> jet + qi
#         15 : g + g --> jet + qi
#         16 : g + g --> jet + g
################################################################################
1
16
```

A.2 Output organization

```
TFile *hfile;
TTree *t2;
diphox_evt_t dip;
norma_evt_t norma_cxx;

void Init_ntuple(char* path_rootfile)
{
        // Create a new ROOT binary machine independent file.
        // Note that this file may contain any kind of ROOT objects, histograms,
        // pictures, graphics objects, detector geometries, tracks, events, etc..
        // This file is now becoming the current directory.

        hfile = new TFile(path_rootfile,"RECREATE","");
        // Create the Ntuple
        t2 = new TTree("t2","exemple de Tree avec Diphox");
    t2->SetMaxTreeSize(10000000000);
        //
        t2->Branch("iprov",&dip.iprov,"iprov/I");
        t2->Branch("ntrack",&dip.ntrack,"ntrack/I");
        t2->Branch("nb_member",&dip.nb_member,"nb_member/I");
        t2->Branch("x3",&dip.x3,"x3/D");
        t2->Branch("energy",dip.e,"e[ntrack]/D");
        t2->Branch("px",dip.px,"px[ntrack]/D");
        t2->Branch("py",dip.py,"py[ntrack]/D");
        t2->Branch("pz",dip.pz,"pz[ntrack]/D");
        t2->Branch("pdf_weight",dip.pdf_weight,"pdf_weight[nb_member]/F");
}


//------------later-------------------------------------------------------------

extern TFile *hfile;
extern TTree *t2;
extern norma_evt_t norma_cxx;

void End_ntuple()
{
    // Save all objects in this file
    float my_tab[3];
    my_tab[0] = float(norma_cxx.nb_evt);
    my_tab[1] = norma_cxx.xsec;
    my_tab[2] = norma_cxx.sqrt_s;
    Float_t *tab = my_tab;
    TVectorT<float> *v = new TVectorT<float> (3,tab);
    t2->GetUserInfo()->Add(v);
    hfile = t2->GetCurrentFile();
    hfile->Write();

    hfile->Close();
```

```
}
```

A.3 Analysis

```
#include <TFile.h>
#include <TTree.h>
#include <TKey.h>
#include <iostream>
#include <fstream>
#include "TList.h"
#include "TVectorT.h"
#include "TH1D.h"
#include <algorithm>
#include <vector>

using namespace std;

double find_scales_min_value_for_one_bin(std::vector<double> v)
{
        double minElement = *std::min_element(v.begin(), v.end());
        return minElement;
}

double find_scales_max_value_for_one_bin(std::vector<double> v)
{
        double maxElement = *std::max_element(v.begin(), v.end());
        return maxElement;
}


int find_scales_max_indice_for_one_bin(std::vector<double> v)
{
        int maxElementIndex = std::max_element(v.begin(),v.end()) - v.begin();
        return maxElementIndex;
}

int find_scales_min_indice_for_one_bin(std::vector<double> v)
{
        int minElementIndex = std::min_element(v.begin(),v.end()) - v.begin();
        return minElementIndex;
}


void run_file(string fileName, std::string& treeName, double bin_min, double
 ↪  bin_max,bool pdf_error_flag, double results_tab[]) {
    // Ouvrir le fichier ROOT en mode lecture
    TFile *file = TFile::Open(fileName.c_str());

    // Verifier si le fichier est ouvert avec succes
    if (!file || file->IsZombie()) {
        std::cerr << "Erreur : Impossible d'ouvrir le fichier ROOT." << std::endl;
        ; // Retourne une valeur par defaut pour indiquer une erreur
    }

    // Recuperer la liste des cles dans le fichier ROOT
    TIter next(file->GetListOfKeys());
    TKey *key;
    TTree *tree = nullptr;

    while ((key = (TKey*)next())) {
        // Recuperer l'objet associe a  la cle
        TObject *obj = file->Get(key->GetName());
```

```
58
59          // Verifier si l'objet est un TTree
60          if (obj && obj->InheritsFrom("TTree")) {
61              tree = dynamic_cast<TTree*>(obj);
62              break;
63          }
64      }
65
66      if (!tree) {
67          std::cerr << "Erreur : Aucun TTree trouve dans le fichier ROOT." << std::endl;
68          file->Close();
69           // Retourne une valeur par defaut pour indiquer une erreur
70      }
71
72      // Recuperer le nom du TTree
73      treeName = tree->GetName();
74      // Renvoie le nombre d'entrees dans le TTree (ntuple)
75      double numEntries = static_cast<double>(tree->GetEntries());
76
77      double px[3],py[3],pt[3];
78      float pdf_weight[1000];
79      float weight;
80
81      TList *list = tree->GetUserInfo();
82      TVectorT<float> &v = *((TVectorT<float>*)(list->At(0)));
83      float xsec = v[1];
84      float nb_evt = v[0];
85      float sqrt_s = v[2];
86      float norma = xsec/(nb_evt*(bin_max-bin_min));
87
88      tree->SetBranchAddress("px",px);
89      tree->SetBranchAddress("py",py);
90      tree->SetBranchAddress("pdf_weight",pdf_weight);
91
92      if (pdf_error_flag==0){
93          TH1D *hpt = new TH1D("hpt","histo", 1, bin_min, bin_max);
94          for (int i=0; i<numEntries; i++)
95              {
96                  tree->GetEntry(i);
97                  pt[0] = sqrt(px[0]*px[0]+py[0]*py[0]);
98                  weight = pdf_weight[0];
99                  hpt->Fill(pt[0],weight);
100
101             }
102
103         hpt->Scale(norma);
104         double res = hpt->GetBinContent(1);
105         results_tab[0] = res;
106         results_tab[1] = 0.0;
107         results_tab[2] = 0.0;
108     // Fermer le fichier ROOT
109         file->Close();
110
111     }
112     if (pdf_error_flag==1){
113         int nb_of_member = 57; //for CT14nlo
114         TH1D* hpt[nb_of_member];
115         for (int a=0; a<nb_of_member;a++)
116             {
117                 string tmp = "hist" + to_string(a);
118                 const char *name = tmp.c_str();
119                 hpt[a] = new TH1D(name,"histo", 1, bin_min, bin_max);
120             }
121
122         for (int i=0; i<tree->GetEntries(); i++)
123             {
```

```
124                 tree->GetEntry(i);
125                 pt[0] = sqrt(px[0]*px[0]+py[0]*py[0]);
126                 for (int j=0; j<nb_of_member; j++)
127                     {
128                         weight = pdf_weight[j];
129                         hpt[j]->Fill(pt[0],weight);
130                     }
131             }
132 
133         for (int k=0;k<57; k++)
134             {
135                 hpt[k]->Scale(norma);
136             }
137 
138         double err_plus =0;
139         double err_minus =0;
140         vector<double> values;
141         vector<double> values_minus;
142         double val;
143         double val_minus;
144         for (int k=0;k<28;k++)
145             {
146                 values.clear();
147                 values_minus.clear();
148                 val=0;
149                 val_minus = 0;
150                 values.push_back((hpt[k+1]->GetBinContent(1)) -
    ↪  (hpt[0]->GetBinContent(1)));
151                 values.push_back((hpt[k+1+28]->GetBinContent(1)) -
    ↪  (hpt[0]->GetBinContent(1)));
152 
    ↪  values_minus.push_back((hpt[0]->GetBinContent(1))-(hpt[k+1]->GetBinContent(1)));
153 
    ↪  values_minus.push_back((hpt[0]->GetBinContent(1))-(hpt[k+1+28]->GetBinContent(1))
    ↪  );
154 
155                 if (values[1]<values[0]) val = values[0];
156                 else val = values[1];
157                 if (values_minus[1]<values[0]) val_minus = values_minus[0];
158                 else val_minus = values_minus[1];
159 
160                 err_plus = err_plus + pow(val,2);
161                 err_minus = err_minus + pow(val_minus,2);
162             }
163 
164         results_tab[0]= hpt[0]->GetBinContent(1);
165         results_tab[1]= sqrt(err_plus);
166         results_tab[2]= sqrt(err_minus);
167 
168         file->Close();
169 
170         }
171 
172 }
173 
174 
175 int main() {
176 
177   vector<double> v;
178   vector<double> v_plus;
179   vector<double> v_minus;
180 
181   string temp_filename_ggd;
182   string temp_filename_ggo;
183   string filename_output = "arXiv1012_4389_rap3_positive.csv";
184 
```

```
185   int nb_scales = 15;
186   int nb_bin = 8;
187   double edge[nb_bin+1] ={15.,20.,25.,30.,35.,40.,50.,60.,100.};
188
189   double array[nb_bin];
190   double err_array[nb_bin];
191   double results_analysis[nb_bin][7];
192
193    std::string equivalence_indices_scales[nb_scales] = {"1 1 1","1 1 2","1 2 1","2 1
      ↪  1","1 2 2","2 1 2","2 2 1","1 1 1/2","1 1/2 1","1/2 1 1","1 1/2 1/2","1/2 1
      ↪  1/2","1/2 1/2 1","1/2 1/2 1/2","2 2 2"};
194
195   for (int i=0;i<nb_bin;i++)
196         {
197                 array[i]= edge[i]+(edge[i+1]-edge[i])/2;
198                 err_array[i]=(edge[i+1]-edge[i])/2;
199         }
200
201  ofstream file(filename_output);
202  file<<"bin_number,pt_min,pt_max,dsigma/dpt,err+(scale),err-(scale),scalemax_indice,s
      ↪  calemin_indice,x,errX,err+PDF,err-PDF"<<endl;
203
204   for (int j=0 ;j<nb_bin;j++)
205         {
206                 v.clear();
207                 bool pdf_err = false;
208                 double results_tab_ggd[3];
209                  double results_tab_ggo[3];
210                 for (int i=0; i<nb_scales;i++)
211         {
212                         temp_filename_ggd = "ggdarXiv1012_4389_rap_152_pos_bin" +
      ↪  std::to_string(j) + "scale" + std::to_string(i) + ".root";
213                         temp_filename_ggo = "ggoarXiv1012_4389_rap_152_pos_bin" +
      ↪  std::to_string(j) + "scale" + std::to_string(i) + ".root";
214
215             std::string treeName;
216
217                         if (i==0) pdf_err = true;
218
219             run_file(temp_filename_ggd, treeName, edge[j],edge[j+1],pdf_err,
      ↪  results_tab_ggd);//mettre un argument tableau de 3 entrées
220                          run_file(temp_filename_ggo, treeName,
      ↪  edge[j],edge[j+1],pdf_err, results_tab_ggo);
221
222                         v.push_back(results_tab_ggd[0] + results_tab_ggo[0]);
223
224         }
225
226                 results_analysis[j][0] = v[0];
227                 results_analysis[j][1] = v[0] - find_scales_min_value_for_one_bin(v);
228                 results_analysis[j][2] = find_scales_max_value_for_one_bin(v) - v[0];
229                 results_analysis[j][3] = double
      ↪  (find_scales_min_indice_for_one_bin(v));
230                 results_analysis[j][4] = double
      ↪  (find_scales_max_indice_for_one_bin(v));
231                 results_analysis[j][5]= results_tab_ggd[1];
232                  results_analysis[j][6]= results_tab_ggd[2];
233
234                  file<<j+1<<","<<edge[j]<<","<<edge[j+1]<<","<<results_analysis[j][0]
      ↪  <<","<<results_analysis[j][2]<<","<<results_analysis[j][1]<<","<<equivalence_indi
      ↪  ces_scales[int(results_analysis[j][4])]<<","<<equivalence_indices_scales[int(resu
      ↪  lts_analysis[j][3])]<<","<<array[j]<<","<<err_array[j]<<","<<results_analysis[j][
      ↪  5]<<","<<results_analysis[j][6]<<endl;
235
236
237    }
```

```
238
239   file.close();
240   cout<<"file written : "<<filename_output<<endl;
241   return 0;
242 }
```

Listing 1: Code d'analyse Jetphox en C++

A.4 Table of JETPHOX results

$\lvert\eta^\gamma\rvert < 0.6$ $N_{bins} = 8$	**ATLAS**				**JETPHOX**			
E_T **[GeV]**	$\frac{d\sigma}{dE_T}$	Uncertainties			$\frac{d\sigma}{dE_T}$	Uncertainties		[pb/GeV]
		stat	sys	lum		PDFs	Scales	
15-20	5.24	±0.11	$\pm^{1.3}_{1.4}$	±11%	6.72	$\pm^{0.32}_{0.29}$	$\pm^{1.15}_{1.04}$	–
20-25	1.88	±0.05	±0.36	±11%	2.4	$\pm^{0.11}_{0.09}$	$\pm^{0.42}_{0.32}$	–
25-30	8.8	±0.3	$\pm^{1.6}_{1.5}$	±11%	10.5	$\pm^{0.47}_{0.37}$	$\pm^{1.53}_{1.42}$	$.10^{-1}$
30-35	4.61	±0.16	±0.68	±11%	5.11	$\pm^{0.22}_{0.17}$	$\pm^{0.81}_{0.62}$	$.10^{-1}$
35-40	2.54	±0.11	±0.38	±11%	2.80	$\pm^{0.12}_{0.09}$	$\pm^{0.41}_{0.33}$	$.10^{-1}$
40-50	1.15	±0.05	$\pm^{0.16}_{0.15}$	±11%	1.3	$\pm^{0.05}_{0.04}$	$\pm^{0.16}_{0.14}$	$.10^{-1}$
50-60	5.0	±0.3	$\pm^{0.74}_{0.63}$	±11%	5.27	$\pm^{0.22}_{0.17}$	$\pm^{0.69}_{0.51}$	$.10^{-2}$
60-100	1.2	±0.01	$\pm^{0.07}_{0.18}$	±11%	1.23	$\pm^{0.04}_{0.03}$	$\pm^{0.14}_{0.11}$	$.10^{-2}$

Table A.1: ATLAS 7TeV,$\eta < 0,6$, [241]

$\lvert\eta^{\gamma}\rvert < 0.6$ $N_{bins} = 22$	**ATLAS**				**JETPHOX**			
E_T **[GeV]**	$\frac{d\sigma}{dE_T}$	Uncertainties			$\frac{d\sigma}{dE_T}$	Uncertainties		[pb/GeV]
		stat	sys	lum		PDFs	Scales	
25-35	1.03	±0.00	$\pm^{0.11}_{0.11}$	±0.02	-	-	-	$.10^{3}$
35-45	3.01	±0.01	$\pm^{0.23}_{0.22}$	±0.06	2.56	$\pm^{0.13}_{0.10}$	$\pm^{0.35}_{0.29}$	$.10^{2}$
45-55	1.15	±0.01	$\pm^{0.06}_{0.06}$	±0.02	0.95	$\pm^{0.04}_{0.03}$	$\pm^{0.12}_{0.10}$	$.10^{2}$
55-65	5.03	±0.02	$\pm^{0.20}_{0.20}$	±0.10	4.27	$\pm^{0.18}_{0.13}$	$\pm^{0.48}_{0.42}$	$.10^{1}$
65-75	2.54	±0.01	$\pm^{0.08}_{0.08}$	±0.05	2.13	$\pm^{0.09}_{0.07}$	$\pm^{0.25}_{0.20}$	$.10^{1}$
75-85	1.37	±0.01	$\pm^{0.03}_{0.03}$	±0.03	1.16	$\pm^{0.04}_{0.03}$	$\pm^{0.13}_{0.11}$	$.10^{1}$
85-105	6.37	±0.03	$\pm^{0.11}_{0.11}$	±0.12	8.48	$\pm^{0.20}_{0.15}$	$\pm^{0.57}_{0.48}$	$.10^{0}$
105-125	2.54	±0.01	$\pm^{0.03}_{0.03}$	±0.05	2.22	$\pm^{0.08}_{0.06}$	$\pm^{0.22}_{0.19}$	$.10^{0}$
125-150	1.09	±0.00	$\pm^{0.01}_{0.01}$	±0.02	0.95	$\pm^{0.03}_{0.03}$	$\pm^{0.10}_{0.08}$	$.10^{0}$
150-175	4.84	±0.02	$\pm^{0.07}_{0.07}$	±0.09	4.25	$\pm^{0.16}_{0.12}$	$\pm^{0.40}_{0.36}$	$.10^{-1}$
175-200	2.34	±0.01	$\pm^{0.03}_{0.03}$	±0.04	2.10	$\pm^{0.08}_{0.06}$	$\pm^{0.19}_{0.17}$	$.10^{-1}$
200-250	9.84	±0.05	$\pm^{0.14}_{0.15}$	±0.19	8.87	$\pm^{0.35}_{0.30}$	$\pm^{0.84}_{0.79}$	$.10^{-2}$
250-300	3.42	±0.02	$\pm^{0.05}_{0.06}$	±0.07	3.11	$\pm^{0.14}_{0.13}$	$\pm^{0.31}_{0.26}$	$.10^{-2}$
300-350	1.41	±0.01	$\pm^{0.02}_{0.03}$	±0.03	1.29	$\pm^{0.06}_{0.06}$	$\pm^{0.10}_{0.12}$	$.10^{-2}$
350-400	6.56	±0.10	$\pm^{0.13}_{0.13}$	±0.13	5.90	$\pm^{0.29}_{0.31}$	$\pm^{0.56}_{0.53}$	$.10^{-3}$
400-470	2.84	±0.05	$\pm^{0.06}_{0.06}$	±0.05	2.62	$\pm^{0.14}_{0.16}$	$\pm^{0.22}_{0.25}$	$.10^{-3}$
470-550	1.13	±0.03	$\pm^{0.03}_{0.02}$	±0.02	1.04	$\pm^{0.07}_{0.07}$	$\pm^{0.09}_{0.10}$	$.10^{-3}$
550-650	4.05	±0.15	$\pm^{0.10}_{0.11}$	±0.08	3.80	$\pm^{0.28}_{0.30}$	$\pm^{0.41}_{0.29}$	$.10^{-4}$
650-750	1.39	±0.09	$\pm^{0.04}_{0.04}$	±0.03	1.40	$\pm^{0.12}_{0.12}$	$\pm^{0.14}_{0.14}$	$.10^{-4}$
750-900	4.36	±0.42	$\pm^{0.13}_{0.13}$	±0.08	4.64	$\pm^{0.47}_{0.49}$	$\pm^{0.40}_{0.50}$	$.10^{-5}$
900-1100	9.35	±1.71	$\pm^{0.31}_{0.31}$	±0.18	10.77	-	$\pm^{1.15}_{1.10}$	$.10^{-6}$
1100-1500	6.12	±3.80	$\pm^{0.25}_{0.26}$	±0.12	13.00	-	$\pm^{1.59}_{1.53}$	$.10^{-7}$

Table A.2: ATLAS 8 TeV, $\eta < 0.6$.

Appendix B

Applgrid

B.1 Definition of grid in APPLGRID

The grids are defined as a `class` in C++.

```
1
2  grid(int NQ2=50,  double Q2min=10000.0, double Q2max=25000000.0,  int Q2order=5,
3       int Nx=50,   double xmin=1e-5,     double xmax=0.9,          int xorder=5,
4       int Nobs=20, double obsmin=100.0,  double obsmax=7000.0,
5       std::string genpdf="mcfm_pdf",
6       int leading_order=0, int _nloops=1,
7       std::string transform="f2",
8       std::string qtransform="h0" );
9
10 grid( int Nobs, const double* obsbins,
11       int NQ2=50,  double Q2min=10000.0, double Q2max=25000000.0, int Q2order=5,
12       int Nx=50,   double xmin=1e-5,     double xmax=0.9,         int xorder=5,
13       std::string genpdf="mcfm_pdf",
14       int leading_order=0, int _nloops=1,
15       std::string transform="f2",
16       std::string qtransform="h0",
17       bool isdis=false );
18
19 grid( const std::vector<double>& obs,
20       int NQ2=50,  double Q2min=10000.0, double Q2max=25000000.0,   int Q2order=5,
21       int Nx=50,   double xmin=1e-5,     double xmax=0.9,           int xorder=5,
22       std::string genpdf="mcfm_pdf",
23       int leading_order=0, int _nloops=1,
24       std::string transform="f2",
25       std::string qtransform="h0" );
```

B.2 Modification of a JETPHOX ingredient for gridding : SPRING

Introdution of a FORTRAN90 module that generates weights ≠1,

```
subroutine my_spring(f,idim,res)
  interface
    function f(xx)
      import ki
      real(ki), dimension(50), intent(in) :: xx
      real(ki) :: f
    end function f
  end interface
  integer, intent(in) :: idim
  real(ki), intent(out) :: res
  real(ki), dimension(idim) :: rand
  real(ki), dimension(20) :: xx
  integer :: i
  !
  call random_number(rand)
  xx = 0._ki
  do i=1,idim
    xx(i) = rand(i)
  end do
  res = f(xx)
  !
end subroutine my_spring

```

where `f` is the integrand, `idim` is the (minimal) dimension of the xx vector and `res` call the function `f1dimd` which fill the grid itself with the weight needed and an appropriate xx.

Bibliography

[1] Richard P. Feynman. Very high-energy collisions of hadrons. Phys. Rev. Lett., 23:1415–1417, 1969.

[2] J. D. Bjorken and Emmanuel A. Paschos. Inelastic Electron Proton and gamma Proton Scattering, and the Structure of the Nucleon. Phys. Rev., 185:1975–1982, 1969.

[3] H. Fritzsch, Murray Gell-Mann, and H. Leutwyler. Advantages of the Color Octet Gluon Picture. Phys. Lett. B, 47:365–368, 1973.

[4] D.J. Gross and Frank Wilczek. Asymptotically Free Gauge Theories. 1. Phys.Rev., D8:3633–3652, 1973.

[5] Steven Weinberg. Nonabelian Gauge Theories of the Strong Interactions. Phys.Rev.Lett., 31:494–497, 1973.

[6] H. David Politzer. Reliable Perturbative Results for Strong Interactions? Phys. Rev. Lett., 30:1346–1349, 1973.

[7] John C. Collins and Davison E. Soper. The Theorems of Perturbative QCD. Ann. Rev. Nucl. Part. Sci., 37:383–409, 1987.

[8] John C. Collins, Davison E. Soper, and George F. Sterman. Factorization of Hard Processes in QCD. Adv. Ser. Direct. High Energy Phys., 5:1–91, 1989.

[9] John C. Collins. Hard scattering factorization with heavy quarks: A General treatment. Phys. Rev. D, 58:094002, 1998.

[10] Raymond Brock et al. Handbook of perturbative QCD: Version 1.0. Rev. Mod. Phys., 67:157–248, 1995.

[11] P. Aurenche, J.-Ph Guillet, and E. Pilon. QED, QCD en pratique. Lecture, October 2016.

[12] R. L. Workman and Others. Review of Particle Physics. PTEP, 2022:083C01, 2022.

[13] Thomas Luthe, Andreas Maier, Peter Marquard, and York Schroder. Complete renormalization of QCD at five loops. JHEP, 03:020, 2017.

[14] P. A. Zyla et al. Review of Particle Physics. PTEP, 2020(8):083C01, 2020.

[15] Henry W. Kendall. Deep inelastic scattering: Experiments on the proton and the observation of scaling. Rev. Mod. Phys., 63:597–614, 1991.

[16] Martin Breidenbach, Jerome I. Friedman, Henry W. Kendall, Elliott D. Bloom, D. H. Coward, H. C. DeStaebler, J. Drees, Luke W. Mo, and Richard E. Taylor. Observed behavior of highly inelastic electron-proton scattering. Phys. Rev. Lett., 23:935–939, 1969.

[17] Elliott D. Bloom et al. High-Energy Inelastic e p Scattering at 6-Degrees and 10-Degrees. Phys. Rev. Lett., 23:930–934, 1969.

[18] Curtis G. Callan, Jr. and David J. Gross. High-energy electroproduction and the constitution of the electric current. Phys. Rev. Lett., 22:156–159, 1969.

[19] Murray Gell-Mann. A Schematic Model of Baryons and Mesons. Phys. Lett., 8:214–215, 1964.

[20] Kenneth G. Wilson. Nonlagrangian models of current algebra. Phys. Rev., 179:1499–1512, 1969.

[21] Curtis G. Callan, Jr. and David J. Gross. Bjorken scaling in quantum field theory. Phys. Rev. D, 8:4383–4394, 1973.

[22] John Collins. Foundations of perturbative QCD, volume 32. Cambridge University Press, 11 2013.

[23] Thomas Appelquist and J. Carazzone. Infrared Singularities and Massive Fields. Phys. Rev. D, 11:2856, 1975.

[24] Thomas Appelquist and J. Carazzone. Physical Processes and the Infrared Problem in Gauge Theories. Nucl. Phys. B, 120:77–95, 1977.

[25] T. Kinoshita. Mass singularities of Feynman amplitudes. J. Math. Phys., 3:650–677, 1962.

[26] T. D. Lee and M. Nauenberg. Degenerate Systems and Mass Singularities. Phys. Rev., 133:B1549–B1562, 1964.

[27] Guido Altarelli, R. Keith Ellis, and G. Martinelli. Large Perturbative Corrections to the Drell-Yan Process in QCD. Nucl. Phys. B, 157:461–497, 1979.

[28] William A. Bardeen, A. J. Buras, D. W. Duke, and T. Muta. Deep Inelastic Scattering Beyond the Leading Order in Asymptotically Free Gauge Theories. Phys. Rev. D, 18:3998, 1978.

[29] Yuri L. Dokshitzer. Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics. Sov. Phys. JETP, 46:641–653, 1977.

[30] V. N. Gribov and L. N. Lipatov. Deep inelastic e p scattering in perturbation theory. Sov. J. Nucl. Phys., 15:438–450, 1972.

[31] Guido Altarelli and G. Parisi. Asymptotic Freedom in Parton Language. Nucl. Phys. B, 126:298–318, 1977.

[32] Tie-Jiun Hou et al. New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC. Phys. Rev. D, 103(1):014013, 2021.

[33] A. Accardi, L. T. Brady, W. Melnitchouk, J. F. Owens, and N. Sato. Constraints on large-x parton distributions from new weak boson production and deep-inelastic scattering data. Phys. Rev. D, 93(11):114017, 2016.

[34] Richard D. Ball et al. The path to N^3LO parton distributions. Eur. Phys. J. C, 84(7):659, 2024.

[35] K. Kovarik et al. nCTEQ15 - Global analysis of nuclear parton distributions with uncertainties in the CTEQ framework. Phys. Rev., D93(8):085037, 2016.

[36] Kari J. Eskola, Petja Paakkinen, Hannu Paukkunen, and Carlos A. Salgado. EPPS21: a global QCD analysis of nuclear PDFs. Eur. Phys. J. C, 82(5):413, 2022.

[37] Rabah Abdul Khalek, Rhorry Gauld, Tommaso Giani, Emanuele R. Nocera, Tanjona R. Rabemananjara, and Juan Rojo. nNNPDF3.0: evidence for a modified partonic structure in heavy nuclei. Eur. Phys. J. C, 82(6):507, 2022.

[38] I. Schienbein, J. Y. Yu, C. Keppel, J. G. Morfin, F. Olness, and J. F. Owens. Nuclear parton distribution functions from neutrino deep inelastic scattering. Phys. Rev. D, 77:054013, 2008.

[39] P. Duwentäster, T. Ježo, M. Klasen, K. Kovařík, A. Kusina, K. F. Muzakka, F. I. Olness, R. Ruiz, I. Schienbein, and J. Y. Yu. Impact of heavy quark and quarkonium data on nuclear gluon PDFs. Phys. Rev. D, 105(11):114043, 2022.

[40] I. Schienbein, J. Y. Yu, K. Kovarik, C. Keppel, J. G. Morfin, F. Olness, and J. F. Owens. PDF Nuclear Corrections for Charged and Neutral Current Processes. Phys. Rev. D, 80:094004, 2009.

[41] E. P. Segarra et al. Extending nuclear PDF analyses into the high-x , low-Q^2 region. Phys. Rev. D, 103(11):114015, 2021.

[42] A. Kusina et al. Impact of LHC vector boson production in heavy ion collisions on strange PDFs. Eur. Phys. J. C, 80(10):968, 2020.

[43] P. Duwentäster, L. A. Husová, T. Ježo, M. Klasen, K. Kovařík, A. Kusina, K. F. Muzakka, F. I. Olness, I. Schienbein, and J. Y. Yu. Impact of inclusive hadron production data on nuclear gluon PDFs. Phys. Rev. D, 104:094005, 2021.

[44] K. F. Muzakka et al. Compatibility of neutrino DIS data and its impact on nuclear parton distribution functions. Phys. Rev. D, 106(7):074004, 2022.

[45] A. W. Denniston et al. Modification of Quark-Gluon Distributions in Nuclei by Correlated Nucleon Pairs. Phys. Rev. Lett., 133(15):152502, 2024.

[46] J. Pumplin, D. R. Stump, and W. K. Tung. Multivariate fitting and the error matrix in global analysis of data. Phys. Rev. D, 65:014011, 2001.

[47] J. Pumplin, D. Stump, R. Brock, D. Casey, J. Huston, J. Kalk, H. L. Lai, and W. K. Tung. Uncertainties of predictions from parton distribution functions. 2. The Hessian method. Phys. Rev. D, 65:014013, 2001.

[48] R. Ruiz et al. Target mass corrections in lepton–nucleus DIS: Theory and applications to nuclear PDFs. Prog. Part. Nucl. Phys., 136:104096, 2024.

[49] Richard E. Taylor. Deep inelastic scattering: The Early years. Rev. Mod. Phys., 63:573–595, 1991.

[50] Jerome I. Friedman. Deep inelastic scattering: Comparisons with the quark model. Rev. Mod. Phys., 63:615–629, 1991.

[51] R. Devenish and A. Cooper-Sarkar. Deep inelastic scattering. Oxford University Press, 2004.

[52] S. Bailey, T. Cridge, L. A. Harland-Lang, A. D. Martin, and R. S. Thorne. Parton distributions from LHC, HERA, Tevatron and fixed target data: MSHT20 PDFs. Eur. Phys. J. C, 81(4):341, 2021.

[53] H. Abramowicz et al. Combination of measurements of inclusive deep inelastic $e^{\pm}p$ scattering cross sections and QCD analysis of HERA data. Eur. Phys. J. C, 75(12):580, 2015.

[54] Richard D. Ball et al. Parton distributions from high-precision collider data. Eur. Phys. J. C, 77(10):663, 2017.

[55] S. Alekhin, J. Blümlein, S. Moch, and R. Placakyte. Parton distribution functions, α_s, and heavy-quark masses for LHC Run II. Phys. Rev. D, 96(1):014011, 2017.

[56] M. Hirai, S. Kumano, and T. H. Nagai. Determination of nuclear parton distribution functions and their uncertainties in next-to-leading order. Phys. Rev. C, 76:065207, 2007.

[57] Daniel de Florian, Rodolfo Sassot, Pia Zurita, and Marco Stratmann. Global Analysis of Nuclear Parton Distributions. Phys. Rev. D, 85:074028, 2012.

[58] K. J. Eskola, H. Paukkunen, and C. A. Salgado. EPS09: A New Generation of NLO and LO Nuclear Parton Distribution Functions. JHEP, 04:065, 2009.

[59] Kari J. Eskola, Petja Paakkinen, Hannu Paukkunen, and Carlos A. Salgado. EPPS16: Nuclear parton distributions with LHC data. Eur. Phys. J. C, 77(3):163, 2017.

[60] A. Kusina, F. Lyonnet, D. B. Clark, E. Godat, T. Jezo, K. Kovarik, F. I. Olness, I. Schienbein, and J. Y. Yu. Vector boson production in pPb and PbPb collisions at the LHC and its impact on nCTEQ15 PDFs. Eur. Phys. J. C, 77(7):488, 2017.

[61] Rabah Abdul Khalek, Jacob J. Ethier, and Juan Rojo. Nuclear parton distributions from lepton-nucleus scattering and the impact of an electron-ion collider. Eur. Phys. J. C, 79(6):471, 2019.

[62] Rabah Abdul Khalek, Jacob J. Ethier, Juan Rojo, and Gijs van Weelden. nNNPDF2.0: quark flavor separation in nuclei from LHC data. JHEP, 09:183, 2020.

[63] Rabah Abdul Khalek, Rhorry Gauld, Tommaso Giani, Emanuele R. Nocera, Tanjona R. Rabemananjara, and Juan Rojo. nNNPDF3.0: evidence for a modified partonic structure in heavy nuclei. Eur. Phys. J. C, 82(6):507, 2022.

[64] Marina Walt, Ilkka Helenius, and Werner Vogelsang. Open-source QCD analysis of nuclear parton distribution functions at NLO and NNLO. Phys. Rev. D, 100(9):096015, 2019.

[65] Hamzeh Khanpour, Maryam Soleymaninia, S. Atashbar Tehrani, Hubert Spiesberger, and Vadim Guzey. Nuclear parton distribution functions with uncertainties in a general mass variable flavor number scheme. Phys. Rev. D, 104(3):034010, 2021.

[66] A. Accardi et al. Electron Ion Collider: The Next QCD Frontier: Understanding the glue that binds us all. Eur. Phys. J. A, 52(9):268, 2016.

[67] R. Abdul Khalek et al. Science Requirements and Detector Concepts for the Electron-Ion Collider: EIC Yellow Report. Nucl. Phys. A, 1026:122447, 2022.

[68] Jonathan L. Feng et al. The Forward Physics Facility at the High-Luminosity LHC. J. Phys. G, 50(3):030501, 2023.

[69] J. Arrington et al. Physics with CEBAF at 12 GeV and future opportunities. Prog. Part. Nucl. Phys., 127:103985, 2022.

[70] R. Acciarri et al. Long-Baseline Neutrino Facility (LBNF) and Deep Underground Neutrino Experiment (DUNE): Conceptual Design Report, Volume 2: The Physics Program for DUNE at LBNF. 12 2015.

[71] Pedro AN Machado, Ornella Palamara, and David W Schmitz. The Short-Baseline Neutrino Program at Fermilab. Ann. Rev. Nucl. Part. Sci., 69:363–387, 2019.

[72] L. Alvarez Ruso et al. Theoretical tools for neutrino scattering: interplay between lattice QCD, EFTs, nuclear physics, phenomenology, and neutrino event generators. 3 2022.

[73] Elliott D. Bloom and Frederick J. Gilman. Scaling, Duality, and the Behavior of Resonances in Inelastic electron-Proton Scattering. Phys. Rev. Lett., 25:1140, 1970.

[74] W. Melnitchouk, R. Ent, and C. Keppel. Quark-hadron duality in electron scattering. Phys. Rept., 406:127–301, 2005.

[75] C. Andreopoulos et al. The GENIE Neutrino Monte Carlo Generator. Nucl. Instrum. Meth. A, 614:87–104, 2010.

[76] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. JHEP, 07:079, 2014.

[77] Enrico Bothmann et al. Event Generation with Sherpa 2.2. SciPost Phys., 7(3):034, 2019.

[78] Alfonso Garcia and Aart Heijboer. High-energy neutrino event simulation at NLO in Genie for KM3NeT and other observatories. PoS, ICRC2019:895, 2020.

[79] Alfonso Garcia, Rhorry Gauld, Aart Heijboer, and Juan Rojo. Complete predictions for high-energy neutrino propagation in matter. JCAP, 09:025, 2020.

[80] Christian Bierlich et al. A comprehensive guide to the physics and usage of PYTHIA 8.3. 3 2022.

[81] Joshua Isaacson, William I. Jay, Alessandro Lovato, Pedro A. N. Machado, and Noemi Rocco. Introducing a novel event generator for electron-nucleus and neutrino-nucleus scattering. Phys. Rev. D, 107(3):033007, 2023.

[82] Ilkka Helenius, Marina Walt, and Werner Vogelsang. NNLO nuclear parton distribution functions with electroweak-boson production data from the LHC. Phys. Rev. D, 105(9):094031, 2022.

[83] Richard A. Brandt and Giuliano Preparata. Operator product expansions near the light cone. Nucl. Phys. B, 27:541–567, 1971.

[84] Norman H. Christ, B. Hasslacher, and Alfred H. Mueller. Light cone behavior of perturbation theory. Phys. Rev. D, 6:3543, 1972.

[85] Geoffrey T. Bodwin. Factorization of the Drell-Yan Cross-Section in Perturbation Theory. Phys. Rev. D, 31:2616, 1985. [Erratum: Phys.Rev.D 34, 3932 (1986)].

[86] John C. Collins, Davison E. Soper, and George F. Sterman. Factorization for Short Distance Hadron - Hadron Scattering. Nucl. Phys. B, 261:104–142, 1985.

[87] John C. Collins, Davison E. Soper, and George F. Sterman. Soft Gluons and Factorization. Nucl. Phys. B, 308:833–856, 1988.

[88] Edmond L. Berger and F. Coester. Nuclear Effects in Deep Inelastic Lepton Scattering. Ann. Rev. Nucl. Part. Sci., 37:463–491, 1987.

[89] Howard Georgi and H. David Politzer. Freedom at Moderate Energies: Masses in Color Dynamics. Phys. Rev. D, 14:1829, 1976.

[90] Howard Georgi and H. David Politzer. Precocious Scaling, Rescaling and xi Scaling. Phys. Rev. Lett., 36:1281, 1976. [Erratum: Phys.Rev.Lett. 37, 68 (1976)].

[91] T. Muta. Foundations of quantum chromodynamics. Second edition, volume 57. 1998.

[92] Taizo Muta. Foundations of Quantum Chromodynamics: An Introduction to Perturbative Methods in Gauge Theories, (3rd ed.), volume 78 of World scientific Lecture Notes in Physics. World Scientific, Hackensack, N.J., 3rd edition, 2010.

[93] R. Keith Ellis, W. Furmanski, and R. Petronzio. Power Corrections to the Parton Model in QCD. Nucl. Phys. B, 207:1–14, 1982.

[94] R. Keith Ellis, W. Furmanski, and R. Petronzio. Unraveling Higher Twists. Nucl. Phys. B, 212:29, 1983.

[95] Riccardo Barbieri, John R. Ellis, M. K. Gaillard, and Graham G. Ross. Mass Corrections to Scaling in Deep Inelastic Processes. Nucl. Phys. B, 117:50–76, 1976.

[96] Riccardo Barbieri, John R. Ellis, Mary K. Gaillard, and Graham G. Ross. A Quest for a Wholly Scaling Variable. Phys. Lett. B, 64:171–176, 1976.

[97] Alvaro De Rujula, Howard Georgi, and H. David Politzer. Demythification of Electroproduction, Local Duality and Precocious Scaling. Annals Phys., 103:315, 1977.

[98] S. Kretzer and M. H. Reno. Target mass corrections to electroweak structure functions and perturbative neutrino cross-sections. Phys. Rev. D, 69:034002, 2004.

[99] John C. Collins and George F. Sterman. Soft Partons in QCD. Nucl. Phys. B, 185:172–188, 1981.

[100] A. Accardi, T. Hobbs, and W. Melnitchouk. Hadron mass corrections in semi-inclusive deep inelastic scattering. JHEP, 11:084, 2009.

[101] M. A. G. Aivazis, Frederick I. Olness, and Wu-Ki Tung. Leptoproduction of heavy quarks. 1. General formalism and kinematics of charged current and neutral current production processes. Phys. Rev. D, 50:3085–3101, 1994.

[102] M. A. G. Aivazis, John C. Collins, Fredrick I. Olness, and Wu-Ki Tung. Leptoproduction of heavy quarks. 2. A Unified QCD formulation of charged and neutral current processes from fixed target to collider energies. Phys. Rev. D, 50:3102–3118, 1994.

[103] S. Kretzer and M. H. Reno. Tau neutrino deep inelastic charged current interactions. Phys. Rev. D, 66:113007, 2002.

[104] Ingo Schienbein et al. A Review of Target Mass Corrections. J. Phys. G, 35:053101, 2008.

[105] Robert L. Jaffe. Deep Inelastic Scattering with Application to Nuclear Targets: Lectures at the 1985 Los Alamos School on Relativistic Dynamics and Quark Nuclear Physics. In Research Program at CEBAF (I): Report of the 1985 Summer Study Group, 7 1985.

[106] Jian-wei Qiu. QCD factorization and rescattering in proton nucleus collisions. 5 2003.

[107] K. Kovarik, I. Schienbein, F. I. Olness, J. Y. Yu, C. Keppel, J. G. Morfin, J. F. Owens, and T. Stavreva. Nuclear Corrections in Neutrino-Nucleus DIS and Their Compatibility with Global NPDF Analyses. Phys. Rev. Lett., 106:122301, 2011.

[108] Otto Nachtmann. Positivity constraints for anomalous dimensions. Nucl. Phys. B, 63:237–247, 1973.

[109] P. G. Blunden, W. Melnitchouk, and J. A. Tjon. Two photon exchange and elastic electron proton scattering. Phys. Rev. Lett., 91:142304, 2003.

[110] P. G. Blunden, W. Melnitchouk, and J. A. Tjon. Two-photon exchange in elastic electron-nucleon scattering. Phys. Rev. C, 72:034612, 2005.

[111] J. Arrington, W. Melnitchouk, and J. A. Tjon. Global analysis of proton elastic form factor data with two-photon exchange corrections. Phys. Rev. C, 76:035205, 2007.

[112] J. Arrington, P. G. Blunden, and W. Melnitchouk. Review of two-photon exchange in electron scattering. Prog. Part. Nucl. Phys., 66:782–833, 2011.

[113] Sam B. Treiman, Roman Jackiw, and David J. Gross. Lectures on Current Algebra and Its Applications. Princeton University Press, 1972.

[114] John C. Collins. Renormalization: An Introduction to Renormalization, The Renormalization Group, and the Operator Product Expansion, volume 26 of Cambridge Monographs on Mathematical Physics. Cambridge University Press, Cambridge, 1986.

[115] George F. Sterman. Partons, factorization and resummation, TASI 95. In Theoretical Advanced Study Institute in Elementary Particle Physics (TASI 95): QCD and Beyond, pages 327–408, 6 1995.

[116] R. L. Jaffe and M. Soldate. Twist Four in Electroproduction: Canonical Operators and Coefficient Functions. Phys. Rev. D, 26:49–68, 1982.

[117] A. Vogt. Efficient evolution of unpolarized and polarized parton distributions with QCD-PEGASUS. Comput. Phys. Commun., 170:65–92, 2005.

[118] E. Moffat, W. Melnitchouk, T. C. Rogers, and N. Sato. What are the low-Q and large-x boundaries of collinear QCD factorization theorems? Phys. Rev. D, 95(9):096008, 2017.

[119] William Detmold. Target mass effects in deep-inelastic scattering on the deuteron. Phys. Lett. B, 632:261–269, 2006.

[120] Fredrick I. Olness and Wu-Ki Tung. Factorization of Helicity Amplitudes and Angular Correlations for Electroweak Processes. Phys. Rev. D, 35:833, 1987.

[121] Fredrick I. Olness, R. J. Scalise, and Wu-Ki Tung. Heavy quark hadroproduction in perturbative QCD. Phys. Rev. D, 59:014506, 1999.

[122] R. S. Thorne and R. G. Roberts. A Variable number flavor scheme for charged current heavy flavor structure functions. Eur. Phys. J. C, 19:339–349, 2001.

[123] S. Alekhin, J. Blümlein, and S. Moch. Heavy-flavor PDF evolution and variable-flavor number scheme uncertainties in deep-inelastic scattering. Phys. Rev. D, 102(5):054014, 2020.

[124] S. Alekhin, J. Blumlein, S. Klein, and S. Moch. The 3, 4, and 5-flavor NNLO Parton from Deep-Inelastic-Scattering Data and at Hadron Colliders. Phys. Rev. D, 81:014032, 2010.

[125] D. Abrams et al. Measurement of the Nucleon F_2^n/F_2^p Structure Function Ratio by the Jefferson Lab MARATHON Tritium/Helium-3 Deep Inelastic Scattering Experiment. Phys. Rev. Lett., 128(13):132003, 2022.

[126] C. Cocuzza, C. E. Keppel, H. Liu, W. Melnitchouk, A. Metz, N. Sato, and A. W. Thomas. Isovector EMC Effect from Global QCD Analysis with MARATHON Data. Phys. Rev. Lett., 127(24):242001, 2021.

[127] M. Sajjad Athar and Jorge G. Morfín. Neutrino(antineutrino)–nucleus interactions in the shallow- and deep-inelastic scattering regions. J. Phys. G, 48(3):034001, 2021.

[128] I. Niculescu et al. Direct observation of quark-hadron duality in the free neutron F_2 structure function. Phys. Rev. C, 91(5):055206, 2015.

[129] Sheren A. Alsalmi. Measurement of the Nuclear Dependence of F_2 and R=Sigma_L/Sigma_T in The Nucleon Resonance Region. PhD thesis, Kent State University, Kent State U., 2019.

[130] B. Z. Kopeliovich, J. G. Morfin, and Ivan Schmidt. Nuclear Shadowing in Electro-Weak Interactions. Prog. Part. Nucl. Phys., 68:314–372, 2013.

[131] J. J. Aubert et al. The ratio of the nucleon structure functions $F2_n$ for iron and deuterium. Phys. Lett. B, 123:275–278, 1983.

[132] L. L. Frankfurt and M. I. Strikman. Hard Nuclear Processes and Microscopic Nuclear Structure. Phys. Rept., 160:235–427, 1988.

[133] L. L. Frankfurt, M. I. Strikman, D. B. Day, and M. Sargsian. Evidence for short range correlations from high Q**2 (e, e-prime) reactions. Phys. Rev. C, 48:2451–2461, 1993.

[134] L. L. Frankfurt and M. I. Strikman. High-Energy Phenomena, Short Range Nuclear Structure and QCD. Phys. Rept., 76:215–347, 1981.

[135] M. M. Sargsian et al. Hadrons in the nuclear medium. J. Phys. G, 29(3):R1–R45, 2003.

[136] J. Rozynek and M. C. Birse. Nucleon Nucleon Correlation Effects on Deeply Inelastic Lepton Scattering in the Region X > 1. Phys. Rev. C, 38:2201–2204, 1988.

[137] I. Niculescu, J. Arrington, R. Ent, and C. E. Keppel. Moments of nuclear and nucleon structure functions at low Q**2 and the momentum sum rule. Phys. Rev. C, 73:045206, 2006.

[138] L. A. Kondratyuk and M. Shmatikov. EMC EFFECT AND MULTI - QUARK BAGS IN NUCLEI. Yad. Fiz., 41:222–235, 1985.

[139] M. M. Sargsian. Superfast quarks in the nuclear medium. Nucl. Phys. A, 782:199–206, 2007.

[140] Adam J. Freese, Misak M. Sargsian, and Mark I. Strikman. Probing superfast quarks in nuclei through dijet production at the LHC. Eur. Phys. J. C, 75(11):534, 2015.

[141] Adam J. Freese, Wim Cosyn, and Misak M. Sargsian. QCD evolution of superfast quarks. Phys. Rev. D, 99(11):114019, 2019.

[142] A. Bodek and J. L. Ritchie. Fermi Motion Effects in Deep Inelastic Lepton Scattering from Nuclear Targets. Phys. Rev. D, 23:1070, 1981.

[143] A. Bodek and J. L. Ritchie. Further Studies of Fermi Motion Effects in Lepton Scattering from Nuclear Targets. Phys. Rev. D, 24:1400, 1981.

[144] M. Vakili et al. Nuclear structure functions in the large x large Q**2 kinematic region in neutrino deep inelastic scattering. Phys. Rev. D, 61:052003, 2000.

[145] Stephen Rock. Elastic electron - proton scattering at high Q**2. 11 1986.

[146] J. Arrington et al. Inclusive Scattering from Nuclei at x>1 in the quasielastic and deeply inelastic regimes. JLab Proposal PR12-06-105, pages 1–21, 2006.

[147] N. Fomin et al. Scaling of the F_2 structure function in nuclei and quark distributions at $x > 1$. Phys. Rev. Lett., 105:212502, 2010.

[148] J. Arrington et al. Inclusive electron scattering from nuclei at x approximately = 1. Phys. Rev. C, 53:2248–2251, 1996.

[149] B. W. Filippone et al. Nuclear structure functions at x > 1. Phys. Rev. C, 45:1582–1585, 1992.

[150] A. C. Benvenuti et al. Nuclear structure function in carbon near x = 1. Z. Phys. C, 63:29–36, 1994.

[151] Sergey A. Kulagin and R. Petti. Global study of nuclear structure functions. Nucl. Phys. A, 765:126–187, 2006.

[152] L. T. Brady, A. Accardi, T. J. Hobbs, and W. Melnitchouk. Next-to leading order analysis of target mass corrections to structure functions and asymmetries. Phys. Rev. D, 84:074008, 2011. [Erratum: Phys.Rev.D 85, 039902 (2012)].

[153] M. A. G. Aivazis, Fredrick I. Olness, and Wu-Ki Tung. QCD formulation of charm production in deep inelastic scattering and the sea quark - gluon dichotomy. Phys. Rev. Lett., 65:2339–2342, 1990.

[154] B. A. Kniehl, G. Kramer, I. Schienbein, and H. Spiesberger. Collinear subtractions in hadroproduction of heavy quarks. Eur. Phys. J. C, 41:199–212, 2005.

[155] F. Halzen and Alan D. Martin. QUARKS AND LEPTONS: AN INTRODUCTORY COURSE IN MODERN PARTICLE PHYSICS. John Wiley & Sons,, 1984.

[156] Michael E. Peskin and Daniel V. Schroeder. An Introduction to quantum field theory. Addison-Wesley, Reading, USA, 1995.

[157] Diogo Buarque Franzosi, Olivier Mattelaer, Richard Ruiz, and Sujay Shil. Automated predictions from polarized matrix elements. JHEP, 04:082, 2020.

[158] H. Murayama, I. Watanabe, and Kaoru Hagiwara. HELAS: HELicity amplitude subroutines for Feynman diagram evaluations. 1 1992.

[159] Michael Krämer, Fredrick I. Olness, and Davison E. Soper. Treatment of heavy quarks in deeply inelastic scattering. Phys. Rev. D, 62:096007, 2000.

[160] H. Abramowicz et al. Combination of measurements of inclusive deep inelastic $e^{\pm}p$ scattering cross sections and QCD analysis of HERA data. Eur. Phys. J. C, 75(12):580, 2015.

[161] C. O. Escobar. Photoproduction of Large Transverse Momentum Mesons and Production of Large Transverse Momentum Photons and Leptons in Proton Proton Collisions. Nucl. Phys. B, 98:173–188, 1975.

[162] C. O. Escobar. Direct Production of Photons with Large Transverse Momentum in Proton Proton Collisions. Phys. Rev. D, 15:355, 1977.

[163] Glennys R. Farrar and Steven C. Frautschi. Copious Direct Photon Production as a Possible Resolution of the Prompt Lepton Puzzle. Phys. Rev. Lett., 36:1017, 1976.

[164] F. Halzen. Large p(T) Photons and the Gammaization of High-Energy Hadron Collisions. 4 1977.

[165] F. Halzen and D. M. Scott. Testing QCD in the Hadroproduction of Real and Virtual Photons. Phys. Rev. Lett., 40:1117, 1978.

[166] R. Ruckl, Stanley J. Brodsky, and J. F. Gunion. The Production of Real Photons at Large Transverse Momentum in PP Collisions. Phys. Rev. D, 18:2469–2483, 1978.

[167] R. D. Field. DYNAMICS OF HIGH-ENERGY REACTIONS. In 19th International Conference on High-Energy Physics, page 0743, 8 1978.

[168] P. Aurenche and J. Lindfors. Direct Photon Production Beyond Leading Order in QCD. Nucl. Phys. B, 168:296–314, 1980.

[169] F. Halzen and D. M. Scott. On the Recent Observation of the Hadroproduction of Prompt Photons. Phys. Rev. D, 21:1320, 1980.

[170] R. Baier, J. Engels, and B. Petersson. Correlations with Large Transverse Momentum Photons and the Gluon Structure Function. Z. Phys. C, 6:309–316, 1980.

[171] Edmond L. Berger, E. Braaten, and R. D. Field. Large p(T) Production of Single and Double Photons in Proton Proton and Pion-Proton Collisions. Nucl. Phys. B, 239:52–92, 1984.

[172] P. Aurenche, A. Douiri, R. Baier, M. Fontannaz, and D. Schiff. Prompt Photon Production at Large p(T) in QCD Beyond the Leading Order. Phys. Lett. B, 140:87–92, 1984.

[173] J. F. Owens. Large Momentum Transfer Production of Direct Photons, Jets, and Particles. Rev. Mod. Phys., 59:465, 1987.

[174] P. Aurenche, R. Baier, M. Fontannaz, and D. Schiff. Prompt Photon Production at Large p(T) Scheme Invariant QCD Predictions and Comparison with Experiment. Nucl. Phys. B, 297:661–696, 1988.

[175] P. Aurenche, R. Baier, and M. Fontannaz. Prompt Photon Production at Colliders. Phys. Rev. D, 42:1440–1449, 1990.

[176] Edmond L. Berger and Jian-wei Qiu. Calculations of prompt photon production in QCD. Phys. Rev. D, 44:2002–2024, 1991.

[177] L. E. Gordon and W. Vogelsang. Polarized and unpolarized isolated prompt photon production beyond the leading order. Phys. Rev. D, 50:1901–1916, 1994.

[178] P. Aurenche, M. Fontannaz, J. P. Guillet, A. Kotikov, and E. Pilon. Is factorization for isolated photon cross-sections broken? Phys. Rev. D, 55:1124–1126, 1997.

[179] M. Fontannaz. Theoretical understanding of direct photon production. Pramana, 51:157–167, 1998.

[180] P. Aurenche, M. Fontannaz, J. Ph. Guillet, Bernd A. Kniehl, E. Pilon, and M. Werlen. A Critical phenomenological study of inclusive photon production in hadronic collisions. Eur. Phys. J. C, 9:107–119, 1999.

[181] L. Apanasevich et al. k_T effects in direct photon production. Phys. Rev. D, 59:074007, 1999.

[182] Thomas Becher and Matthew D. Schwartz. Direct photon production with effective field theory. JHEP, 02:040, 2010.

[183] F. Arleo et al. Hard probes in heavy-ion collisions at the LHC: Photon physics in heavy ion collisions at the LHC. 10 2004.

[184] Shreyasi Acharya et al. Direct photon production at low transverse momentum in proton-proton collisions at $\sqrt{\mathbf{s}} = \mathbf{2.76}$ and 8 TeV. Phys. Rev. C, 99(2):024912, 2019.

[185] S. Catani, M. Fontannaz, J. P. Guillet, and E. Pilon. Cross-section of isolated prompt photons in hadron hadron collisions. JHEP, 05:028, 2002.

[186] David d'Enterria and Juan Rojo. Quantitative constraints on the gluon distribution function in the proton from collider isolated-photon data. Nucl. Phys. B, 860:311–338, 2012.

[187] Morad Aaboud et al. Measurement of prompt photon production in $\sqrt{s_{\mathrm{NN}}}$ = 8.16 TeV p+Pb collisions with ATLAS. Phys. Lett. B, 796:230–252, 2019.

[188] Shreyasi Acharya et al. Measurement of isolated prompt photon production in pp and p–Pb collisions at the LHC. Eur. Phys. J. C, 85(12):1407, 2025.

[189] ALICE. Letter of Intent: A Forward Calorimeter (FoCal) in the ALICE experiment. Technical report, CERN, Geneva, 2020.

[190] L. Bourhis, M. Fontannaz, and J. P. Guillet. Quarks and gluon fragmentation functions into photons. Eur. Phys. J. C, 2:529–537, 1998.

[191] H. Baer, J. Ohnemus, and J. F. Owens. A Next-to-leading Logarithm Calculation of Direct Photon Production. Phys. Rev. D, 42:61–71, 1990.

[192] F. Aversa, P. Chiappetta, Mario Greco, and J. P. Guillet. Higher Order Corrections to QCD Jets. Phys. Lett. B, 210:225, 1988.

[193] F. Aversa, Mario Greco, P. Chiappetta, and J. P. Guillet. HIGHER ORDER CORRECTIONS TO QCD JETS: GLUON-GLUON PROCESSES. Phys. Lett. B, 211:465, 1988.

[194] F. Aversa, P. Chiappetta, Mario Greco, and J. P. Guillet. QCD Corrections to Parton-Parton Scattering Processes. Nucl. Phys. B, 327:105, 1989.

[195] John M. Campbell, R. Keith Ellis, and Ciaran Williams. Direct Photon Production at Next-to–Next-to-Leading Order. Phys. Rev. Lett., 118(22):222001, 2017. [Erratum: Phys.Rev.Lett. 124, 259901 (2020)].

[196] John M. Campbell, R. Keith Ellis, and Ciaran Williams. Erratum: Direct photon production at next-to–next-to-leading order [phys. rev. lett. 118, 222001 (2017)]. Phys. Rev. Lett., 124:259901, Jun 2020.

[197] Xuan Chen, Thomas Gehrmann, Nigel Glover, Marius Höfer, and Alexander Huss. Isolated photon and photon+jet production at NNLO QCD accuracy. JHEP, 04:166, 2020.

[198] X. Chen, T. Gehrmann, E. W. N. Glover, M. Höfer, A. Huss, and R. Schürmann. Single photon production at hadron colliders at NNLO QCD with realistic photon isolation. JHEP, 08:094, 2022.

[199] Robin Schuermann, Xuan Chen, Thomas Gehrmann, E. W. N. Glover, Marius Höfer, and Alexander Huss. NNLO Photon Production with Realistic Photon Isolation. PoS, LL2022:034, 2022.

[200] S. Catani, M. Fontannaz, J. P. Guillet, and E. Pilon. Cross-section of isolated prompt photons in hadron hadron collisions. JHEP, 05:028, 2002.

[201] Edmond L. Berger, Xiao-feng Guo, and Jian-wei Qiu. Breakdown of conventional factorization for isolated photon cross-sections. Phys. Rev. Lett., 76:2234–2237, 1996.

[202] Patrick Aurenche, Michel Fontannaz, Jean-Philippe Guillet, Eric Pilon, and Monique Werlen. A New critical study of photon production in hadronic collisions. Phys. Rev. D, 73:094007, 2006.

[203] P. Aurenche, P. Chiappetta, M. Fontannaz, J. P. Guillet, and E. Pilon. Next-to-leading order bremsstrahlung contribution to prompt photon production. Nucl. Phys. B, 399:34–62, 1993.

[204] P. Aurenche, R. Baier, M. Fontannaz, J. F. Owens, and M. Werlen. The Gluon Contents of the Nucleon Probed with Real and Virtual Photons. Phys. Rev. D, 39:3275, 1989.

[205] John M. Campbell and R. Keith Ellis. An Update on vector boson pair production at hadron colliders. Phys. Rev. D, 60:113006, 1999.

[206] John M. Campbell, R. Keith Ellis, and Ciaran Williams. Vector boson pair production at the LHC. JHEP, 07:018, 2011.

[207] John Campbell and Tobias Neumann. Precision Phenomenology with MCFM. JHEP, 12:034, 2019.

[208] https://lapth.cnrs.fr/PHOX_FAMILY/diphox.html.

[209] T. Binoth, J. P. Guillet, E. Pilon, and M. Werlen. A Full next-to-leading order study of direct photon pair production in hadronic collisions. Eur. Phys. J. C, 16:311–330, 2000.

[210] P. Chiappetta, R. Fergani, and J. P. Guillet. Production of two large p(T) hadrons from hadronic collisions. Z. Phys. C, 69:443–457, 1996.

[211] B. W. Harris and J. F. Owens. The Two cutoff phase space slicing method. Phys. Rev. D, 65:094032, 2002.

[212] S. Catani and M. H. Seymour. A General algorithm for calculating jet cross-sections in NLO QCD. Nucl. Phys. B, 485:291–419, 1997. [Erratum: Nucl.Phys.B 510, 503–504 (1998)].

[213] S. Frixione. A General approach to jet cross-sections in QCD. Nucl. Phys. B, 507:295–314, 1997.

[214] Setsuya Kawabata. A New version of the multidimensional integration and event generation package BASES/SPRING. Comput. Phys. Commun., 88:309–326, 1995.

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

[216] John M. Campbell, R. Keith Ellis, and Walter T. Giele. A Multi-Threaded Version of MCFM. Eur. Phys. J. C, 75(6):246, 2015.

[217] Eric Godat et al. PDF Flavor Determination and the nCTEQ PDFs: WZ vector boson production in heavy ion collisions. PoS, DIS2018:009, 2018.

[218] https://mcfm.fnal.gov/.

[219] John M. Campbell, Juan Rojo, Emma Slade, and Ciaran Williams. Direct photon production and PDF fits reloaded. Eur. Phys. J. C, 78(6):470, 2018.

[220] Stefano Carrazza, R. Keith Ellis, and Giulia Zanderighi. QCDLoop: a comprehensive framework for one-loop scalar integrals. Comput. Phys. Commun., 209:134–143, 2016.

[221] A. Gehrmann-De Ridder, T. Gehrmann, and E. W. Nigel Glover. Antenna subtraction at NNLO. JHEP, 09:056, 2005.

[222] A. Huss et al. NNLOJET: a parton-level event generator for jet cross sections at NNLO QCD accuracy. 3 2025.

[223] Stefan Hoeche, Steffen Schumann, and Frank Siegert. Hard photon production and matrix-element parton-shower merging. Phys. Rev. D, 81:034026, 2010.

[224] Frank Siegert. A practical guide to event generation for prompt photon production with Sherpa. J. Phys. G, 44(4):044007, 2017.

[225] Georges Aad et al. Measurement of the inclusive isolated-photon cross section in pp collisions at $\sqrt{s} = 13$ TeV using 36 fb^{-1} of ATLAS data. JHEP, 10:203, 2019.

[226] Thomas Becher, Christian Lorentzen, and Matthew D. Schwartz. Precision Direct Photon and W-Boson Spectra at High p_T and Comparison to LHC Data. Phys. Rev. D, 86:054026, 2012.

[227] Thomas Becher, Guido Bell, Christian Lorentzen, and Stefanie Marti. Transverse-momentum spectra of electroweak bosons near threshold at NNLO. JHEP, 02:004, 2014.

[228] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. JHEP, 06:043, 2010.

[229] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with Parton Shower simulations: the POWHEG method. JHEP, 11:070, 2007.

[230] Paolo Nason. A New method for combining NLO QCD with shower Monte Carlo algorithms. JHEP, 11:040, 2004.

[231] Michael Klasen, Christian Klein-Bösing, and H. Poppenborg. Prompt photon production and photon-jet correlations at the LHC. JHEP, 03:081, 2018.

[232] Tomas Jezo, Michael Klasen, and Florian König. Prompt photon production and photon-hadron jet correlations with POWHEG. JHEP, 11:033, 2016.

[233] R. Frederix, S. Frixione, V. Hirschi, D. Pagani, H. S. Shao, and M. Zaro. The automation of next-to-leading order electroweak calculations. JHEP, 07:185, 2018. [Erratum: JHEP 11, 085 (2021)].

[234] Stefano Frixione, Fabian Stoeckli, Paolo Torrielli, and Bryan R. Webber. NLO QCD corrections in Herwig++ with MC@NLO. JHEP, 01:053, 2011.

[235] Paolo Torrielli and Stefano Frixione. Matching NLO QCD computations with PYTHIA using MC@NLO. JHEP, 04:110, 2010.

[236] The ATLAS collaboration. Measurement of the inclusive isolated prompt photon cross section in pp collisions at $\sqrt{s}$ = 8 tev with the atlas detector. Journal of High Energy Physics, 2016(8):5, Aug 2016.

[237] Vardan Khachatryan et al. Measurement of the Isolated Prompt Photon Production Cross Section in pp Collisions at $\sqrt{s} = 7$˜TeV. Phys. Rev. Lett., 106:082001, 2011.

[238] Serguei Chatrchyan et al. Measurement of the Differential Cross Section for Isolated Prompt Photon Production in pp Collisions at 7 TeV. Phys. Rev. D, 84:052011, 2011.

[239] Albert M Sirunyan et al. Measurement of differential cross sections for inclusive isolated-photon and photon+jets production in proton-proton collisions at $\sqrt{s}$ = 13 TeV. Eur. Phys. J. C, 79(1):20, 2019.

[240] Thomas Boettcher. Direct photon production at LHCb. Nucl. Phys. A, 982:251–254, 2019.

[241] G. Aad et al. Measurement of the inclusive isolated prompt photon cross section in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. Phys. Rev. D, 83:052005, 2011.

[242] Georges Aad et al. Measurement of the inclusive isolated prompt photons cross section in pp collisions at $\sqrt{s}$ = 7 TeV with the ATLAS detector using 4.6 fb^{-1}. Phys. Rev. D, 89(5):052004, 2014.

[243] Georges Aad et al. Measurement of the inclusive isolated prompt photon cross section in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. JHEP, 08:005, 2016.

[244] Sayipjamal Dulat, Tie-Jiun Hou, Jun Gao, Marco Guzzi, Joey Huston, Pavel Nadolsky, Jon Pumplin, Carl Schmidt, Daniel Stump, and C. P. Yuan. New parton distribution functions from a global analysis of quantum chromodynamics. Phys. Rev. D, 93(3):033006, 2016.

[245] Andy Buckley, James Ferrando, Stephen Lloyd, Karl Nordström, Ben Page, Martin Rüfenacht, Marek Schönherr, and Graeme Watt. LHAPDF6: parton density access in the LHC precision era. Eur. Phys. J., C75:132, 2015.

[246] Florian Jonas. Probing the initial state of heavy-ion collisions with isolated prompt photons. PhD thesis, LBL, Berkeley, 5 2023.

[247] K. J. Eskola, V. J. Kolhinen, and C. A. Salgado. The Scale dependent nuclear effects in parton distributions for practical applications. Eur. Phys. J. C, 9:61–68, 1999.

[248] Tancredi Carli, Dan Clements, Amanda Cooper-Sarkar, Claire Gwenlan, Gavin P. Salam, Frank Siegert, Pavel Starovoitov, and Mark Sutton. A posteriori inclusion of parton density functions in NLO QCD final-state calculations at hadron colliders: The APPLGRID Project. Eur. Phys. J. C, 66:503–524, 2010.

[249] T. Kluge, K. Rabbertz, and M. Wobisch. FastNLO: Fast pQCD calculations for PDF fits. In 14th International Workshop on Deep Inelastic Scattering, pages 483–486, 9 2006.

[250] S. Carrazza, E. R. Nocera, C. Schwan, and M. Zaro. PineAPPL: combining EW and QCD corrections for fast evaluation of LHC processes. JHEP, 12:108, 2020.

[251] Richard D. Ball et al. Parton distributions from high-precision collider data. Eur. Phys. J. C, 77(10):663, 2017.

[252] Thomas Becher and Xavier Garcia i Tormo. Electroweak Sudakov effects in W, Z and γ production at large transverse momentum. Phys. Rev. D, 88(1):013009, 2013.

[253] Thomas Becher and Xavier Garcia i Tormo. Addendum: Electroweak Sudakov effects in W, Z and gamma production at large transverse momentum. Phys. Rev. D, 9 2015. [Addendum: Phys.Rev.D 92, 073011 (2015)].

[254] Stefano Frixione. Isolated photons in perturbative QCD. Phys. Lett. B, 429:369–374, 1998.

[255] https://www.gnu.org/software/software.html.