Annual Review of Nuclear and Particle Science

Nuclear Parton Distribution Functions After the First Decade of LHC Data

Michael Klasen[1] and Hannu Paukkunen[2,3]

[1]Institute for Theoretical Physics, University of Münster, Münster, Germany; email: michael.klasen@uni-muenster.de

[2]Department of Physics, University of Jyväskylä, Jyväskylä, Finland; email: hannu.paukkunen@jyu.fi

[3]Helsinki Institute of Physics, Helsinki, Finland

Keywords

quantum chromodynamics, nuclear structure, parton distribution functions, collider physics, future developments

Abstract

We present a review of the conceptual basis, current knowledge, and recent progress regarding global analysis of nuclear parton distribution functions (PDFs). After introducing the theoretical foundations and methodological approaches for the extraction of nuclear PDFs from experimental data, we discuss how different measurements in fixed-target and collider experiments provide increasingly precise constraints on various aspects of nuclear PDFs, including shadowing, antishadowing, the EMC effect, Fermi motion, flavor separation, deuteron binding, and target-mass and other higher-twist effects. Particular emphasis is given to measurements carried out in proton–lead collisions at the Large Hadron Collider, which have revolutionized the global analysis during the past decade. These measurements include electroweak boson, jet, light hadron, and heavy flavor observables. Finally, we outline the expected impact of the future Electron Ion Collider and discuss the role and interplay of nuclear PDFs with other branches of nuclear, particle, and astroparticle physics.

Contents

1. INTRODUCTION

Nuclear structure at high energies is an important current research topic; it is relevant not only to our understanding of the fundamental quark and gluon dynamics in protons and neutrons bound in nuclei but also to elucidating the formation, properties, and evolution of a deconfined state of hadronic matter that existed in the early Universe—the so-called quark–gluon plasma (QGP). Nuclear parton distribution functions (PDFs) encode cold binding effects in nuclei (1, 2), determine the preequilibrium phase and initial-state phase transition to the QGP (3), and are correlated with the precise extraction of important QGP properties, such as its temperature (4) and the final-state phase transition during chemical freeze-out (5). While the evolution of PDFs with the scale Q^2, at which they are probed, can be computed in perturbative quantum chromodynamics (QCD) (6–9), their dependence on the longitudinal parton momentum fraction inside the hadron is nonperturbative and must be fitted to experimental data. Traditionally, deep inelastic scattering (DIS) of charged leptons or neutrinos and Drell–Yan (DY) dilepton production with fixed targets have

provided the bulk of the data. However, over the last decade, collider data from the Large Hadron Collider (LHC) at CERN and the Relativistic Heavy Ion Collider (RHIC) at BNL have led to significant improvements in our knowledge of collinear, unpolarized nuclear PDFs, which are reviewed here.

In the naive parton model, the double-differential charged lepton DIS cross section per nucleon,

$$\frac{\mathrm{d}^2\sigma^{\ell A}}{\mathrm{d}x\mathrm{d}y} = \frac{2\pi\alpha^2}{Q^4} s\left[1+(1-y)^2\right] F_2^A(x), \qquad 1.$$

is directly related to the nuclear structure function $F_2^A(x) = x\sum_q e_q^2 f_q^A(x)$ and thus the quark PDFs $f_q^A(x)$ in the nucleus A, while the gluon PDF and Q^2 dependence enter only in the QCD-improved parton model (see Section 2). Here, α is the electromagnetic fine-structure constant, e_q is the fractional charge of quark q, Q^2 is the virtuality of the exchanged photon, $s = Q^2/(xy)$ is the hadronic center-of-mass energy, y is the lepton inelasticity, and x is the Bjorken variable. The nuclear structure functions differ from free-nucleon structure functions not only due to an admixture of protons and neutrons but also in various other ways depending on the region in x. At small x, a depletion (called shadowing) is observed in $F_2^A(x)$ at $x \lesssim 0.05$, followed with increasing x by an enhancement (called antishadowing) at $0.05 \lesssim x \lesssim 0.3$, then another depletion—the famous EMC effect—at $0.3 \lesssim x \lesssim 0.7$, and ultimately an enhancement due to Fermi motion at $0.7 \lesssim x$. The ratios of large-A isoscalar structure functions to the structure function of the deuteron (D)—a loosely bound isoscalar state of a proton (p) and a neutron (n)—can be parameterized and fitted, albeit with considerable uncertainties, to SLAC (10, 11) and NMC (12) data in the following form (13):

$$R(x) = 1.10 - 0.36x - 0.28e^{-21.9x} + 2.77x^{14.4}, \qquad 2.$$

that is, without a dependence on the nuclear mass number A or the hard scale Q^2. However, a logarithmic decrease in A has been observed in the EMC region in the 1984 SLAC data, and a logarithmic increase in Q^2 has been observed in the shadowing region in the more precise NMC data (14). To first approximation, this is also how quark PDFs in a bound nucleon N differ from the free-nucleon PDFs.

This review focuses on model-independent global fits of nuclear PDFs and the progress due to LHC data over the last decade. To obtain a first impression of the nuclear dynamics, it is nonetheless interesting to discuss the main historical experimental measurements and theoretical interpretations in the different x regions.

Shadowing had been known to be present in real and virtual photon scattering on nuclei since the 1970s (15, 16). At the hadron level, it can be interpreted by assuming that the photon fluctuates from its pointlike state into a superposition of vector mesons (ρ, ω, ϕ), which then interact strongly with the nucleons on the surface of the target nucleus (vector meson dominance). Such nucleons absorb most of the incoming "hadron" flux and therefore cast a shadow onto the inner ones (17). At the parton level and in the nuclear rest frame, the photon can be seen to split into a quark–antiquark dipole with lifetime τ, which scatters coherently from multiple partons in the nucleus if $\tau \geq R_A \sim A^{1/3}$ fm or $x \leq 1/(2M_N R_A) \sim 0.1A^{-1/3}$, resulting again in a reduced nuclear cross section (18–20).

The relative motion of nucleons inside the nucleus was considered first in the 1970s for the deuteron (21) and also in the 1980s for heavier nuclei near the Fermi surface (22–24). The structure function of a nucleon in a nucleus can then be expressed as a convolution,

$$F_2^A(x_N) = \int_{x_N}^{A} \mathrm{d}y f_A(y) F_2^N\left(\frac{x_N}{y}\right), \text{where } f_A(y) \sim \frac{1}{\sqrt{2\pi}\,\Delta_A} \exp\left\{\frac{-[y-(1-\delta_A)]^2}{2\Delta_A^2}\right\}, \qquad 3.$$

of the nucleon structure function $F_2^N(x_N)$ with the nucleon momentum distribution $f_A(y)$. Its peak is shifted away from unity due to soft nuclear interactions by an amount $\delta_A \sim 0.04$, which corresponds roughly to the ratio of nucleon separation energy over its mass. The width Δ_A is determined by a fraction of the Fermi momentum $k_F \sim 250$ MeV divided by the nucleon mass $M_N \sim 1$ GeV and is thus small. Equation 3 can therefore be approximated by a simple rescaling, $F_2^A(x_N) = F_2^N(x_N/(1-\delta_A))$. The net result of this rescaling is to deplete the partons in the intermediate-x_N region, implying $F_2^A/F_2^B < 1$ for $A > B$, and to enrich the large-$x_N \sim 1$ region with $F_2^A/F_2^B > 1$. The region $x_N > 1$ can be modeled by modifying the Gaussian ansatz for $f_A(y)$ in Equation 3 (e.g., with a power law tail; 25) and also in deconfinement or cluster models (2), but this region is usually neglected (26).

The discovery of a suppression of F_2^{Fe}/F_2^D at $x_N = 0.65$ of ~0.89 in 1983 by the EMC Collaboration in muon DIS (27) and its confirmation in reanalyzed iron and aluminum SLAC data from the early 1970s (28, 29) came as a big surprise, since Fermi motion models predicted an enhancement of ~1.25 at this value of x_N and a suppression only for $x_N < 0.5$. It triggered many theoretical explanations at both the nuclear and partonic levels, and a consensus has yet to emerge (30). Models with nucleons as the only degrees of freedom in the nucleus must be incomplete, since the convolution in Equation 3 violates baryon number and momentum sum rules. The missing momentum could be carried by pions, whose exchanges lead to an intermediate-range nuclear attraction of 300 to 500 MeV, which is canceled by short-distance vector exchanges of almost equal size. The net effect is an average binding energy of 8 MeV per nucleon as observed (31). However, one would then expect an enhancement of antiquarks and therefore of the DY process, which has not been seen (32). The failure of nucleon-only and nucleon–pion models indicates that the nucleon structure itself is modified by the medium. The parton model interpretation of the EMC effect is that the medium reduces the number of high-momentum quarks. This momentum reduction leads, via the uncertainty principle, to the notion that quarks in nuclei are confined in a larger volume than that of a free nucleon. There are two proposals to realize this simple idea: Either scalar and vector mean-field effects cause bound nucleons to be larger than free ones, or short-range correlations (SRCs) cause the nucleon structure to be modified by including either NN^* configurations or deconfined six-quark configurations that are orthogonal to the two-nucleon wave functions. Interestingly, two-nucleon SRCs might explain the observed linear correlation between the magnitude of the EMC effect at $0.3 \leq x_N \leq 0.7$ and the size of the plateau observed in quasi-elastic scattering at $1.5 \leq x_N \leq 2$, which would solve the single-nucleon sum rule problem (33, 34).

A similar compensation mechanism could be at work in the antishadowing region, which is imposed by shadowing through the momentum sum rule. In the Breit frame, small-momentum quarks and gluons, because of the uncertainty principle, spread over a distance comparable to the nucleon–nucleon separation. Quarks and gluons from different nucleons can then overlap spatially and fuse, thus increasing the density of high-momentum partons (antishadowing) at the expense of that of lower-momentum ones (shadowing) (35). In perturbative QCD, this process is flavor dependent, and $q\bar{q} \rightarrow g$ fusion results, for instance, in shadowing for antiquarks and antishadowing for gluons. The fact that there is no clear evidence of antishadowing in the DY process can be interpreted either with an important role of valence (v) quarks or as a consequence of the evolution in Q^2 (1, 2).

The nonperturbative nature of nuclear interactions, the need for phenomenological models, and the incongruous nuclear and partonic interpretations of the effects described above are strong motivations to parameterize and fit nuclear PDFs to the available data in a model-independent way. Improving on Equation 2, Eskola parameterized the ratio of heavier nuclear structure functions over deuterons separately in each region at the starting scale Q_0^2, matched it at the transition

points (whose definition depended on A), and evolved it in Q^2 (36). As observed experimentally (16) and predicted theoretically (37), shadowing then vanished only very slowly toward larger Q^2, in particular for quarks and antiquarks. When the parameterized ratio was fitted to DIS and pA DY data, while imposing baryon number and momentum conservation, the nuclear data could be described fairly independently of the underlying proton PDFs (38). A rigorous statistical analysis of DIS data initially led to rather large values of χ^2 per degree of freedom (df) of 1.82 to 1.93 (39), which could, however, be reduced to 1.35 in leading-order (LO) and 1.21 in next-to-leading-order (NLO) QCD using more precise DIS and pA DY data (40, 41). Collider (RHIC) data on π^0 production from 2006 introduced sensitivity to the gluon density beyond scaling violations with a resulting $\chi^2/\mathrm{df} = 0.79$ at LO and NLO (42). An equally good value of $\chi^2/\mathrm{df} = 0.83$ was obtained at NLO with a similar dataset but with a direct A-dependent parameterization of nuclear PDFs at the starting scale Q_0^2 (43). The inclusion of neutrino DIS proved to be more difficult (44), triggering a discussion about the universality of nuclear effects in charged lepton and neutrino scattering (45–49). In the last decade, a wealth of LHC data on electroweak boson, photon, light and heavy hadron, and jet production has become available for proton–lead (p+Pb) collisions, which has already had a significant impact on the determination of nuclear PDFs (50–52). These modern developments are reviewed thoroughly in this article, whereas other recent reviews have mostly focused on proton PDFs (53–55) or sketched a larger multidimensional picture of the nucleus (56).

The remainder of this article is organized as follows. In Section 2, we briefly review the theoretical foundations of nuclear DIS and its factorization. In Section 3, we describe the different methodological approaches to global fits of nuclear PDFs. In Sections 4 and 5, we discuss the impact of the different experimental data in roughly chronological order; our main focus is on LHC data. The impact of the future Electron Ion Collider (EIC) and connections to other fields in nuclear, particle, and astroparticle physics (lattice QCD, the search for gluon saturation, the QGP, and astrophysical phenomena) are briefly addressed in Section 6 before we conclude the review in Section 7.

2. THEORETICAL FOUNDATIONS

We start our discussion of nuclear PDFs by reviewing their theoretical foundations both within the operator product expansion (OPE) and the QCD-improved parton model, including target-mass and other higher-twist effects.

2.1. Factorization in the Operator Product Expansion and the Quantum Chromodynamics–Improved Parton Model

High-energy lepton DIS is the key process for studying the hadronic structure of nucleons N or nuclei A with mass M_N or $M_A = A M_N$ and four-momentum P_N or P_A in terms of their partonic (quark and gluon) degrees of freedom (see **Figure 1**) (57, 58). The charged lepton ℓ and (anti)neutrino ν have incoming four-momentum k and outgoing four-momentum k', the squared center-of-mass energy is $s_{N,A} = (k + P_{N,A})^2$, and X represents all final-state hadrons with total four-momentum P_X and squared mass $W^2_{N,A} = (q + P_{N,A})^2$. In the laboratory frame, the lepton energy loss is $\nu = q \cdot P_{N,A}/M_{N,A} = E - E'$, and the exchanged vector boson V has squared momentum transfer $Q^2 = -q^2 > 0$.

The inclusive differential cross section $\mathrm{d}\tilde{\sigma}^A \sim L^{\mu\nu}\tilde{W}^A_{\mu\nu}$ can be written as a combination of a pointlike leptonic tensor $L^{\mu\nu}$ and the hadronic tensor

$$\tilde{W}^A_{\mu\nu}(P_A, q) = \frac{1}{4\pi}\int \mathrm{d}^4z\, e^{iq\cdot z}\langle A|J^\dagger_\mu(z)\, J_\nu(0)|A\rangle = \frac{1}{4\pi}\ \mathrm{disc.}\ \tilde{T}_{\mu\nu}(P_A, q), \qquad 4.$$

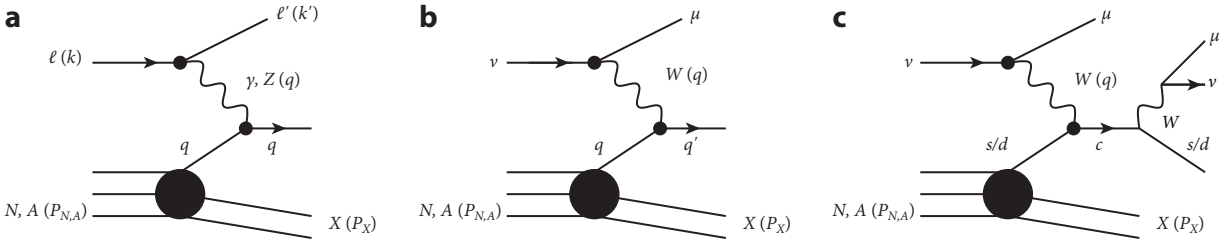

Figure 1

Leading-order diagrams for (*a*) neutral-current deep inelastic scattering with charged leptons, (*b*) charged-current deep inelastic scattering with neutrinos or antineutrinos, and (*c*) charm dimuon production. The charm quark is understood to hadronize before the semileptonic decay.

where the latter is given in terms of a product of hadronic currents and can be related to the discontinuity of the virtual forward Compton scattering amplitude:

$$\tilde{T}^A_{\mu\nu}(P_A, q) = \int \mathrm{d}^4 z \; e^{iq\cdot z} \; \langle A|\mathcal{T} J^\dagger_\mu(z)\, J_\nu(0)|A\rangle. \qquad 5.$$

The OPE then allows one to expand the hadronic matrix element of the forward scattering amplitude in a complete set of local operators (59):

$$\tilde{T}^A_{\mu\nu}(P_A, q) = -2i \sum_{j,\tau,n} c^{j,\mu_1\cdots\mu_n}_{\tau,\mu\nu} \langle A|O^{j,\tau}_{\mu_1\cdots\mu_n}|A\rangle \; = -2i \sum_{j,k} \frac{2^{2k}}{Q^{4k}} C_j^{2k} A_{2k} \tilde{\Pi}^{j,k}_{\mu\nu} \; + \; \mathcal{O}(\tau > 2), \qquad 6.$$

where $c^{j,\mu_1\cdots\mu_n}_{\tau,\mu\nu}$ denotes the hard scattering, τ denotes the twist of the operator O (defined as its mass dimension minus its spin), and j denotes different operators with the same twist. Up to power corrections, one can identify the product of the perturbative Wilson coefficients C_i^{2k} and reduced hadronic matrix elements A_{2k} as integer Mellin moments of structure functions

$$\int_0^1 \mathrm{d}y \, y^{2k-1} \; \tilde{F}^A_i(y, Q^2) = C_i^{2k} A_{2k} + \mathcal{O}(\tau > 2), \qquad 7.$$

with $y^{2k-1} \to y^{2k-2}$ for $i = 2$. The Lorentz structure in terms of metric tensors and momenta is encoded in $\tilde{\Pi}^{j,k}_{\mu\nu}$.

In the QCD-improved parton model, the nuclear structure functions

$$\tilde{F}^A_i(x_A, Q^2) = \sum_{j=q,g} \int_{x_A}^1 \frac{\mathrm{d}y_A}{y_A} \; C_{i,j} \tilde{f}^A_j(y_A, Q^2) + \mathcal{O}(\tau > 2) \qquad 8.$$

depend on the Bjorken scaling variable $x_{N,A} = Q^2/(2q \cdot P_{N,A}) = Q^2/(2\nu M_{N,A})$ with $x_N \in [0, A]$ ($x_A = x_N/A \in [0, 1]$) with logarithmic scaling violation in Q^2 (see below). They are given as convolutions of target-independent short-distance Wilson coefficients C_i and C_j with universal nuclear PDFs $\tilde{f}^A_j$. Inspection of Equations 7 and 8 shows that nuclear PDFs can be understood as moments of matrix elements of local twist-two operators composed of quark and gluon fields (60). The nuclear PDFs $\tilde{f}^A_i(x_A, Q^2)$ above are related to the more familiar average-nucleon nuclear PDFs $f^A_i(x_N, Q^2)$ through $f^A_i(x_N, Q^2) = \tilde{f}^A_i(x_A, Q^2)/A$. This rescaling is a key step that allows us to compare structure functions across different nuclei, including the free nucleon. The evolution of the PDFs with the scale Q^2 is perturbatively calculable and given by the Dokshitzer–Gribov–Lipatov–Altarelli–Parisi (DGLAP) evolution equations (6–9):

$$\frac{\mathrm{d}f^A_i(x_N, Q^2)}{\mathrm{d}\ln Q^2} = \frac{\alpha_s(Q^2)}{2\pi} \int_{x_N}^A \frac{\mathrm{d}y_N}{y_N} \; P_{ij}\left(\frac{x_N}{y_N}\right) \; f^A_j(y_N, Q^2), \qquad 9.$$

where α_s is the QCD coupling and P_{ij} denotes the partonic splitting functions. Furthermore, the PDFs satisfy sum rules due to charge, baryon number, and momentum conservation,

$$\int_0^A \mathrm{d}x_N \{f_{u_v}^A, f_{d_v}^A\}(x_N, Q^2) = \{2Z + N, Z + 2N\}, \quad \int_0^A \mathrm{d}x_N x_N \sum_i f_i^A(x_N, Q^2) = A, \qquad 10.$$

where Z is the electric charge of the nucleus with baryon number $A = Z + N$. It is therefore common (but not necessary) to decompose the nuclear PDFs as

$$f_i^A(x_N, Q^2) = \frac{Z}{A} f_i^{p/A}(x_N, Q^2) + \frac{A - Z}{A} f_i^{n/A}(x_N, Q^2), \qquad 11.$$

where the bound-neutron PDFs $f_i^{n/A}(x_N, Q^2)$ are commonly obtained from those of the bound-proton $f_i^{p/A}(x_N, Q^2)$ by assuming isospin symmetry:

$$f_{u,\bar{u}}^{n/A}(x_N, Q^2) = f_{d,\bar{d}}^{p/A}(x_N, Q^2), \quad f_{d,\bar{d}}^{n/A}(x_N, Q^2) = f_{u,\bar{u}}^{p/A}(x_N, Q^2). \qquad 12.$$

In principle, the above integrations extend to A, although the dominant support of the PDFs is expected to be in the region $x_A \leq 1/A$, or $x_N \leq 1$. One therefore usually assumes $f_i^A(x_N, Q^2) = 0$ for $x_N > 1$, which has the advantage that the same evolution equations can be used for all nuclei in the interval $x \in [0, 1]$.

Accounting for the heavy quark masses is essential in an accurate description of the free-proton data (61). Standard methods to handle the quark masses include general-mass (GM) variable-flavor-number schemes (VFNSs) (62) such as the simplified Aivazis–Collins–Olness–Tung (SACOT) schemes (63–65) or the fixed-order plus next-to-leading logarithms (FONLL) schemes (66, 67), which provide systematic ways to interpolate between the fixed-flavor-number scheme (FFNS), in which heavy quarks are not considered as partons, and the zero-mass (ZM) VFNS, in which heavy quarks are treated as massless partons.

2.2. Target-Mass Corrections

Target-mass corrections (TMCs) can be discussed in terms of collinear factorization (68) or can be obtained from the OPE by inverting moments of structure functions (59) (see Equation 7). They can be written in the following general form (69):

$$F_i^{A,\mathrm{TMC}}(x_N, Q^2) = \sum_j A_i^j F_j^A(\xi_N, Q^2) + B_i^j h_j^A(\xi_N, Q^2) + C_i g_2^A(\xi_N, Q^2), \qquad 13.$$

where $\xi_N = 2x_N/(1 + r_N)$ is the Nachtmann variable (70) with $r_N = \sqrt{1 + 4x_N^2 M_N^2/Q^2}$, and

$$h_j^A(\xi_N, Q^2) \sim \int_{\xi_N}^A \mathrm{d}\xi_N' \frac{F_j^A(\xi_N', Q^2)}{\xi_N'} \quad \text{and} \quad g_2^A(\xi_N, Q^2) = \int_{\xi_N}^A \mathrm{d}\xi_N' h_2^A(\xi_N', Q^2) \qquad 14.$$

are auxiliary functions with $1/\xi_N' \to 1/{\xi_N'}^2$ for $j = 2$. Specifically, the proportionality factor in h_2^A is unity, and we have the following:

$$F_2^{A,\mathrm{TMC}}(x_N, Q^2) = \left(\frac{x_N^2}{\xi_N^2 r_N^3}\right) F_2^A(\xi_N, Q^2) + \left(\frac{6M_N^2 x_N^3}{Q^2 r_N^4}\right) h_2^A(\xi_N, Q^2) + \left(\frac{12M_N^4 x_N^4}{Q^4 r_N^5}\right) g_2^A(\xi_N, Q^2). \qquad 15.$$

Quark masses modify ξ_N by $\xi_N \to R_{ij}\xi_N$, where the factor R_{ij} depends on the incoming and outgoing quark masses m_i and m_j. In the case of $m_i = 0$, one obtains the slow-rescaling limit $R_{ij} = 1 + (nm_j)^2/Q^2$ with $n = 1$ for charged-current (CC) DIS (71) and $n = 2$ for neutral-current (NC) DIS (61).

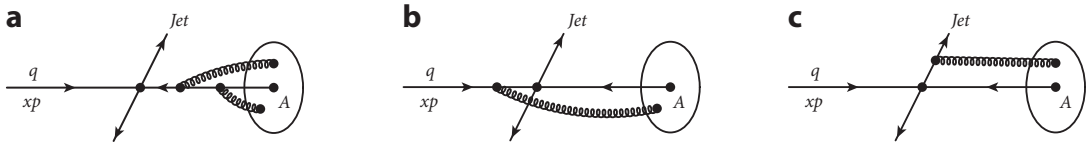

Figure 2

Classification of multiple parton scattering in a nuclear medium: (*a*) interactions internal to the nucleus, (*b*) initial-state interactions, and (*c*) final-state interactions. Figure adapted from Reference 72.

2.3. Higher-Twist Corrections

Interactions internal to the nucleus as in **Figure 2*a*** change the nuclear PDFs with respect to those of the free nucleon. However, since only a single parton participates in the hard scattering, the structure functions can still be factorized as in Equation 8, and the leading-twist nuclear PDFs can be parameterized at an initial scale Q_0, evolved with (in principle A-dependent) evolution equations and fitted to experimental data or modeled theoretically.

Next-to-leading power corrections to Equation 8 of $\mathcal{O}(r_{\mathrm{T}}^2 \sim 1/p_{\mathrm{T}}^2)$, $\mathcal{O}(m_{\mathrm{J}}^2/p_{\mathrm{T}}^2)$, and $\mathcal{O}(\alpha_s(Q^2)\Lambda^2/Q^2)$ arise from the transverse size r_{T} of the initial nucleus, the nonvanishing invariant mass m_{J} of the final jet, and initial-state and final-state interactions involving more than one parton as shown in **Figure 2*b*,*c***. In hadron–nucleus collisions, both are enhanced by $A^{1/3}$ due to the large density of soft partons in the nucleus. For initial-state interactions, $\Lambda^2 \sim 0.01$ GeV^2 is the squared scale of the twist-four correlation function of single parton pairs and proportional to the transverse field strength. For final-state interactions, long-range soft parton interactions must also be considered. It can be shown that the $A^{1/3}$ enhancement can be factorized to all powers in hadron–nucleus but not nucleus–nucleus collisions and then involves correlation functions of multiple parton pairs. In general, however, even the hadron–nucleus DY cross section cannot be factorized beyond next-to-leading power (72).

3. GLOBAL ANALYSES OF NUCLEAR PARTON DISTRIBUTION FUNCTIONS

The inverse problem of extracting nuclear PDFs from experimental data is approached in a similar way as the determination of (free) proton PDFs (55); that is, all global analyses are based on optimizing the correspondence between theoretical calculations and experimental measurements by minimizing a figure-of-merit function that is typically of the following form:

$$\chi^2 = \sum_{i,j} (D_i - T_i) C_{ij}^{-1} (D_j - T_j). \qquad 16.$$

Here, D_i denotes the experimental values for observables in the fit, and T_i denotes the corresponding theoretical values, which depend on the PDFs. The covariance matrix is defined as $C_{ij} = \sigma_i^2 \delta_{ij} + \sum_\alpha \bar{\sigma}_{i\alpha} \bar{\sigma}_{j\alpha}$, where σ_i is the total uncorrelated uncertainty added in quadrature and $\bar{\sigma}_{i\alpha}$ is the correlated systematic uncertainty from source α. How to exactly assign values for σ_i and $\bar{\sigma}_{i\alpha}$ varies from one analysis to another and depends on whether uncertainties are multiplicative or additive (73, 74). In the case of fitting nuclear PDFs, only a few datasets provide complete information on the correlated systematic uncertainties, and in most cases only the overall normalization uncertainty is given.

An essential part of global PDF analyses is the propagation of experimental uncertainties into the PDFs. From the practical point of view, the two principal methods are the Hessian (75, 76) and Monte Carlo methods (77, 78). The Hessian uncertainty analysis (75, 76) is based on expanding

the χ^2 function in the vicinity of its minimum value χ_0^2:

$$\chi^2 \approx \chi_0^2 + \sum_{i,j} H_{ij}\delta a_i \delta a_j = \chi_0^2 + \sum_i z_i^2, \qquad 17.$$

where δa_i indicates deviations from the best-fit parameters and H_{ij} is the second-derivative matrix, the Hessian matrix. In the second step one diagonalizes the Hessian matrix by finding its eigendirections. The PDF error sets $S_k^{\pm}$ are then defined as deviations along these eigendirections to positive or negative directions such that χ^2 increases by a fixed amount $\Delta\chi^2$. The value for $\Delta\chi^2$ can be defined in various ways. A common feature in the current global analyses is that $\Delta\chi^2$ is of the order of the number of fit parameters for 68% confidence level (CL) uncertainties and somewhat higher for 90% CL uncertainties. In the plots of this review, the uncertainties $(\delta X)^{\pm}$ for a given PDF-dependent quantity X are calculated by the asymmetric prescription (79),

$$(\delta X)^{\pm} = \sqrt{\sum_k \left[\begin{matrix}\max\\ \min\end{matrix}\left[X(S_k^+) - X(S_0), X(S_k^-) - X(S_0), 0\right]\right]^2}, \qquad 18.$$

where S_0 denotes the best fit. Of the fits discussed in this review, nCTEQ15HQ (50), EPPS21 (51), TUJU21 (80), and KSASG20 (81) make use of the Hessian method. The Monte Carlo method is based on preparing several fits in which the central values of the experimental data have been randomly shifted within the uncertainties. In the case of uncorrelated experimental uncertainties, the fitted data points are obtained as follows:

$$D_i \to D_i\,(1 + \sigma_i R_i), \qquad 19.$$

where R_i is taken from a Gaussian distribution centered around 0 and with a unit standard deviation. The above formula can also be generalized to the case of correlated uncertainties (82, 83). In the nNNPDF3.0 fit (52), the 100% uncertainty band for a given quantity X is then defined as the minimum and maximum values obtained by calculating X with all PDF replicas. The 90% uncertainty is defined by disregarding the highest and lowest 10%.

3.1. The nCTEQ Framework

In the nCTEQ15 NLO analysis (43) and its sequels, the ansatz for the nuclear PDFs at $Q_0 = 1.3$ GeV follows the CTEQ6M parameterization (84):

$$x f_i^{p/A}(x, Q_0^2) = c_{0i} x^{c_{1i}} (1-x)^{c_{2i}} e^{c_{3i}x} (1 + e^{c_{4i}} x)^{c_{5i}}, \qquad 20.$$

where $i = u_v, d_v, g, \bar{u} + \bar{d}$, and $s + \bar{s}$, and we have dropped the index N of x, while

$$\frac{f_{\bar{d}}^{p/A}(x, Q_0^2)}{f_{\bar{u}}^{p/A}(x, Q_0^2)} = c_0' x^{c_1'} (1-x)^{c_2'} + (1 + c_3' x)(1-x)^{c_4'}. \qquad 21.$$

The normalization coefficients c_{0i} are constrained by the momentum and valence quark sum rules. The proton baseline is similar to the fit CTEQ6.1M (85) but has minimal influence from nuclear data (86). There are currently no uncertainties associated with this proton baseline PDF. The A dependence of the parameterization is directly included in the coefficients:

$$c_{ki} \longrightarrow c_{ki}(A) \equiv p_{ki} + a_{ki}(1 - A^{-b_{ki}}), \quad k = \{1, \ldots, 5\}. \qquad 22.$$

The 16 free parameters in nCTEQ15 describe the x dependence in u_v, d_v, g, and $\bar{u} + \bar{d}$, while the parameters in $\bar{d}/\bar{u}$ were fixed, as was $s = \bar{s} = \kappa(\bar{u} + \bar{d})/2$. Starting with the fits nCTEQ15WZ (87) and nCTEQ15WZ+SIH (88), including weak (W and Z) boson and single inclusive hadron

(SIH) production at the LHC, three free parameters were added for $s + \bar{s}$. The 38 error PDFs are obtained with the Hessian method and a tolerance of $\Delta\chi^2 = 35$.

In the latest nCTEQ fit, nCTEQ15HQ (50), the weakly fragmentation-function (FF)-dependent SIH data were complemented by open heavy quark and quarkonium (HQ) production data from the LHC that had shown great potential to constrain the gluon in preceding reweighting studies (89, 90). These data were fitted with a data-driven method (91), in which the cross sections of hadrons A_1 and A_2 are taken to be dominated by the gluon–gluon subprocesses,

$$\sigma(A_1A_2 \to \mathcal{Q} + X) = \int \mathrm{d}x_1 \mathrm{d}x_2 f_g^{A_1}(x_1, \mu^2) f_g^{A_2}(x_2, \mu^2) \frac{1}{2\hat{s}} \overline{|\mathcal{A}_{gg\to\mathcal{Q}+X}|^2} \mathrm{dPS}, \tag{23.}$$

where $\mathcal{Q} = D^0, J/\psi, B \to J/\psi, \Upsilon, \psi'$, and $B \to \psi'$; the squared factorization scale μ^2 is related to the geometric mean of $M_\mathcal{Q}^2$ and p_T^2; and dPS denotes the two-particle phase space. The effective matrix elements $\overline{|\mathcal{A}_{gg\to\mathcal{Q}+X}|^2}$ are parameterized by a generalized Crystal Ball function:

$$\overline{|\mathcal{A}_{gg\to\mathcal{Q}+X}|^2} = \frac{\lambda^2\kappa\hat{s}}{M_\mathcal{Q}^2} e^{a|y|} \times \begin{cases} e^{-\kappa\frac{p_\mathrm{T}^2}{M_\mathcal{Q}^2}} & \text{if } p_\mathrm{T} \le \langle p_\mathrm{T}\rangle \\ e^{-\kappa\frac{\langle p_\mathrm{T}\rangle^2}{M_\mathcal{Q}^2}} \left(1 + \frac{\kappa}{n}\frac{p_\mathrm{T}^2 - \langle p_\mathrm{T}\rangle^2}{M_\mathcal{Q}^2}\right)^{-n} & \text{if } p_\mathrm{T} > \langle p_\mathrm{T}\rangle \end{cases}, \tag{24.}$$

where $M_\mathcal{Q}$ denotes the mass of particle $\mathcal{Q}$, $\hat{s} = x_1x_2s$, and p_T and y correspond to the transverse momentum and rapidity of $\mathcal{Q}$. The free parameters $\lambda, \kappa, \langle p_\mathrm{T}\rangle, n$, and a are fitted for each final state to pp data. The fits agree with NLO GM-VFNS (92) and NRQCD (93) calculations within their scale uncertainties. In nCTEQ15HQ, the LHC p+Pb data are consistently included in the fit as absolute cross sections in the case of all observables.

3.2. The EPPS Framework

The latest EPPS analysis, EPPS21 (51), is rooted in a series of global fits (42, 94–97) that parameterize the bound-proton PDFs at the starting scale $Q_0 = 1.3$ GeV as follows:

$$f_i^{p/A}(x, Q_0^2) = R_i^{p/A}(x, Q_0^2) f_i^p(x, Q_0^2). \tag{25.}$$

Here, $i = u_v, d_v, g, \bar{u}, \bar{d}, s$ (with $s = \bar{s}$), and the free-proton PDFs $f_i^p(x, Q^2)$ are taken from the fit CT18A (98), which includes more LHC pp data sensitive to strange quarks than the default CT18 fit. This reduces the dependence of the proton baseline on the ν+Fe DIS data, which is also part of the CT18A analysis. The nuclear modifications $R_i^{p/A}(x, Q_0^2)$ are parameterized through 24 free parameters. The parameterization is piecewise smooth in x, so that parameters controlling different x regions mix as little as possible; that is,

$$R_i^{p/A}(x, Q_0^2) = \begin{cases} a_{0i} + a_{1i}(x - x_{ai})\left[e^{-xa_{2i}/x_{ai}} - e^{-a_{2i}}\right], & x \le x_{ai} \\ b_{0i}x^{b_{1i}}(1-x)^{b_{2i}}e^{xb_{3i}}, & x_{ai} \le x \le x_{ei} \\ c_{0i} + c_{1i}(c_{2i} - x)(1-x)^{-\beta_i}, & x_{ei} \le x \le 1, \end{cases} \tag{26.}$$

where x_{ai} and x_{ei} are the locations of the anticipated antishadowing maximum and minimum of the EMC effect, respectively. The A dependence is encoded in such a way that larger nuclei tend to have larger nuclear effects at $x = 0, x_{ai}, x_{ei}$ through

$$R_i^{p/A}(x, Q_0^2) = 1 + \left[R_i^{p/A_\mathrm{ref}}(x, Q_0^2) - 1\right]\left(\frac{A}{A_\mathrm{ref}}\right)^{\gamma_i}, \ \gamma_i > 0, \ A_\mathrm{ref} = 12. \tag{27.}$$

However, for very small nuclei, a monotonic A scaling is not necessarily a justified assumption (e.g., certain small nuclei are more tightly bound), and such deviations are also allowed in the

EPPS parameterization through $R_i^{p/A}(x, Q_0^2) \longrightarrow 1 + f_A\left[R_i^{p/A}(x, Q_0^2) - 1\right]$, where $f_A = 1$ by default. The parameterization is applied for $A \geq 3$, while for smaller nuclei the nuclear modification is set to unity. This is in line with the baseline CT18A proton PDFs (98), which include DIS data on deuteron targets with no nuclear corrections (99). The parameterization of $R_i^{p/A}(x, Q_0^2)$ (and thereby the nuclear PDFs) is not restricted to be positive at small x. Even if the positivity were imposed at Q_0, the backward evolution to smaller Q^2 would result in negative values, especially for gluons. Requiring the positivity at Q_0 thus appears too restrictive and would induce an increased dependence on the parameterization scale Q_0.

The EPPS analyses use ratios of cross sections or structure functions whenever possible to remove the dependence on free-proton PDFs as much as possible. Ratios of cross sections have the additional advantage that they are perturbatively much more stable than absolute cross sections, reducing the risk of fitting missing higher-order effects into the nuclear modifications. Furthermore, experimental uncertainties—known and unknown—can be expected to cancel, for instance, the one from the luminosity. The uncertainties in the EPPS21 analysis are evaluated through the Hessian method with a global tolerance of $\Delta\chi^2 = 33$. In addition, the dependence of nuclear modifications on the free-proton PDFs is mapped by repeating the fit with each of the 58 CT18A error sets. As a result, the EPPS21 fit comes with 106 error sets, which are correlated with the CT18A error sets.

3.3. The nNNPDF Framework

In the nNNPDF3.0 NLO analysis (52), six independent combinations of nuclear PDFs are parameterized at $Q_0 = 1$ GeV on an evolution basis:

$$x f_i^{p/A}(x, Q_0^2) = B_i x^{\alpha_i}(1-x)^{\beta_i} \mathrm{NN}_i(x, A),\ i = \Sigma, T_3, T_8, V, V_3, g, \qquad 28.$$

where Σ, T_3, T_8, V, and V_3 label certain combinations of quark PDFs (53), g is the gluon, and $\mathrm{NN}_f(x, A)$ represents the value of the neuron in the output layer of the neural network associated with each independent nuclear PDF. The normalization coefficients $B_\Sigma = B_{T_3} = B_{T_8} = 1$, while B_V, B_{V_3}, and B_g enforce the momentum and valence sum rules and are determined at Q_0 for each value of A. The preprocessing exponents α_i and β_i are required to control the small- and large-x behavior of the nuclear PDFs. They are fitted simultaneously with the network parameters. The exponents α_V and α_{V_3} are restricted to lie in the range [0, 5] during the fit to ensure integrability of the valence distributions. The other exponents α_i are restricted to the range $[-1, 5]$, consistent with momentum sum rule requirements, while the exponents β_i lie in the range [1, 10]. The figure of merit is defined as follows:

$$\chi^2_{\mathrm{fit}} = \chi^2_{t_0} + \kappa^2_{\mathrm{pos}} + \kappa^2_{\mathrm{BC}}, \qquad 29.$$

where the first term is the contribution from experimental data with a covariance matrix that takes into account the normalization uncertainties of the different datasets with a self-consistent iterative t_0 procedure (74). The second term, κ^2_{pos}, imposes the positivity of physical cross sections, and the third term, κ^2_{BC}, ensures that in the limit $A \to 1$, the nNNPDF predictions reduce to those of the free-proton boundary condition using a grid of 100 points, half of which are distributed logarithmically between $x = 10^{-6}$ and 0.1; the remaining half are linearly distributed between 0.1 and 0.7. Ratios of open heavy quark production (i.e., D^0 meson) data are included with the reweighting method.

The free-proton baseline PDFs are fitted separately with the NNPDF3.1 methodology (100) to all datasets included in the NNPDF4.0 NLO analysis (101) except those involving nuclei with $A \geq 2$. Each replica of nNNPDF3.0 has been fitted using a randomly chosen replica of proton

PDFs coming from these fits. As a result, the uncertainties of nNNPDF3.0 reduce to those of the proton baseline in the $A \to 1$ limit, and the correlations between free-proton and nuclear PDFs are charted and available for the users. The $A \geq 2$ data that were removed from the proton fit are included in the fit of nuclear PDFs. Consequently, the nuclear effects in the deuteron are fitted in a model-independent way.

A common challenge in training neural network models is the choice of the hyperparameters, such as their architecture and activation functions, the optimization algorithm, and learning rates. Here, kernel density estimators (102, 103) have outperformed random or grid searches in selecting the most promising sets. One observes, for instance, that a network with one hidden layer and 25 nodes beats a network with two hidden layers.

3.4. Next-to-Next-to-Leading-Order and Model-Dependent Approaches

A few analyses have been performed in next-to-next-to-leading-order (NNLO) QCD, albeit with restricted datasets. For example, the nNNPDF1.0 analysis (104) fitted only the light quark singlet, octet, and gluon PDFs to the charged lepton DIS data. The TUJU19 analysis (105) took a further step by complementing the charged lepton DIS data with neutrino DIS data in the framework of the open-source tool xFitter (106). Using an ansatz,

$$x f_i^{p/A}\left(x, Q_0^2\right) = c_{0i}\, x^{c_{1i}} (1-x)^{c_{2i}} \left(1 + c_{3i}\, x + c_{4i}\, x^2\right), \tag{30.}$$

and a definition of χ^2 that followed the HERAPDF2.0 analysis (107), the free-proton baseline was first fitted to the HERA data. Due to the limited dataset, $s = \bar{s} = \bar{u} = \bar{d}$ had to be assumed, but with 13 free parameters and a tolerance of $\Delta\chi^2 = 20$, a proton baseline was obtained that was very similar to that from the HERAPDF2.0 analysis. These proton PDFs were then used in the fit to the heavy nuclei data, in which the A dependence was directly encoded in the fit parameters c_k as in the nCTEQ fits (43), including the case of the deuteron. In the updated TUJU21 analysis (80), LHC data for weak boson production were also included in both the proton and nuclear fits. However, it was still not possible to consider the flavor decomposition in the sea quark sector. The TUJU19 and TUJU21 analyses were the first to consider both the proton and heavier nuclei in the same framework, even though they did not provide a common Hessian matrix.

The KSASG20 analysis (81) is based on CT18 (98) free-proton PDFs and uses a nuclear modification factor of cubic HKN form (39, 41)

$$\mathcal{W}_i(x, A) = 1 + \left(1 - \frac{1}{A^\alpha}\right) \frac{a_i(A) + b_i(A)\, x + c_i(A)\, x^2 + d_i(A)\, x^3}{(1-x)^{\beta_i}}, \tag{31.}$$

which is flexible enough to accommodate both shadowing and antishadowing effects. From nuclear volume and surface contributions, one obtains $\alpha = 1/3$ (108), the a_i control shadowing, and β_i are related to Fermi motion. Only charged lepton and neutrino DIS data, supplemented with fixed-target DY data, were fitted with a very restricted flavor decomposition.

Model-dependent nuclear PDFs have been proposed based on the leading-twist approximation for nuclear shadowing with antishadowing constrained by the momentum sum rule (19), SRCs of nucleon pairs motivated by the EMC effect (109), and the four-component Kulagin–Petti model (110, 111), which can reproduce a variety of DIS, fixed-target DY (112), and LHC weak boson data (113). As the viewpoint of the present review is a data-based global analysis of nuclear PDFs, we will not discuss the modeling of nuclear effects further.

3.5. Overview of Global Nuclear Parton Distribution Function Analyses

The key features of the nuclear PDF analyses described above are summarized in **Table 1**. In addition to the main methodological assumptions and input parameters, we list the fitted experimental data types, grouped into fixed-target and collider data. Since TUJU21 and KSASG20

Table 1 Key features of recent global analyses of nuclear PDFs

Analysis	nCTEQ15HQ (50)	EPPS21 (51)	nNNPDF3.0 (52)	TUJU21 (80)	KSASG20 (81)
Theoretical input					
Perturbative order	NLO	NLO	NLO	NNLO	NNLO
Heavy quark scheme	SACOT-χ	SACOT-χ	FONLL	FONLL	FONLL
Value of $\alpha_s(M_Z)$	0.118	0.118	0.118	0.118	0.118
Charm mass	1.3 GeV	1.3 GeV	1.51 GeV	1.43 GeV	1.3 GeV
Bottom mass	4.5 GeV	4.75 GeV	4.92 GeV	4.5 GeV	4.75 GeV
Input scale Q_0	1.3 GeV	1.3 GeV	1.0 GeV	1.3 GeV	1.3 GeV
Data points	1,484	2,077	2,188	2,410	4,353
Independent flavors	5	6	6	4	3
Parameterization	Analytic	Analytic	Neural network	Analytic	Analytic
Free parameters	19	24	256	16	18
Error analysis	Hessian	Hessian	Monte Carlo	Hessian	Hessian
Tolerance	$\Delta\chi^2 = 35$	$\Delta\chi^2 = 33$	N/A	$\Delta\chi^2 = 50$	$\Delta\chi^2 = 20$
Proton PDF	~CTEQ6.1	CT18A	~NNPDF4.0	~HERAPDF2.0	CT18
Proton PDF correlations	—	✓	✓	—	—
Deuteron corrections	(✓)[a,b]	✓[c]	✓	✓	✓
Fixed-target data					
SLAC/EMC/NMC NC DIS	✓	✓	✓	✓	✓
Cut on Q^2	4 GeV^2	1.69 GeV^2	3.5 GeV^2	3.5 GeV^2	1.2 GeV^2
Cut on W^2	12.25 GeV^2	3.24 GeV^2	12.5 GeV^2	12.0 GeV^2	—
JLab NC DIS	(✓)[a]	✓	—	—	✓
CHORUS/CDHSW CC DIS	(✓/—)[b]	✓/—	✓/—	✓/✓	✓/✓
NuTeV/CCFR 2μ CC DIS	(✓/✓)[b]	—	✓/—	—	—
pA DY	✓	✓	✓	—	✓
πA DY	—	✓	—	—	—
Collider data					
Z bosons	✓	✓	✓	✓	—
$W^{\pm}$ bosons	✓	✓	✓	✓	—
Light hadrons	✓	✓[d]	—	—	—
Cut on p_T	3 GeV	3 GeV	—	—	—
Jets	—	✓	✓	—	—
Prompt photons	—	—	✓	—	—
Prompt D^0	✓	✓	✓[e]	—	—
Cut on p_T	3 GeV	3 GeV	0 GeV	—	—
Quarkonia (J/ψ, ψ', Υ)	✓	—	—	—	—

Abbreviations: CC DIS, charged-current deep inelastic scattering; DY, Drell–Yan process; FONLL, fixed-order plus next-to-leading logarithms; N/A, not applicable; NC DIS, neutral-current deep inelastic scattering; NLO, next-to-leading order; NNLO, next-to-next-to-leading order; PDF, parton distribution function; SACOT, simplified Aivazis–Collins–Olness–Tung.

[a]Included in the nCTEQ15HIX analysis (26).

[b]Included in the nCTEQ15ν analysis (158).

[c]Through CT18A.

[d]Only π^0 in D+Au.

[e]Only forward ($y > 0$).

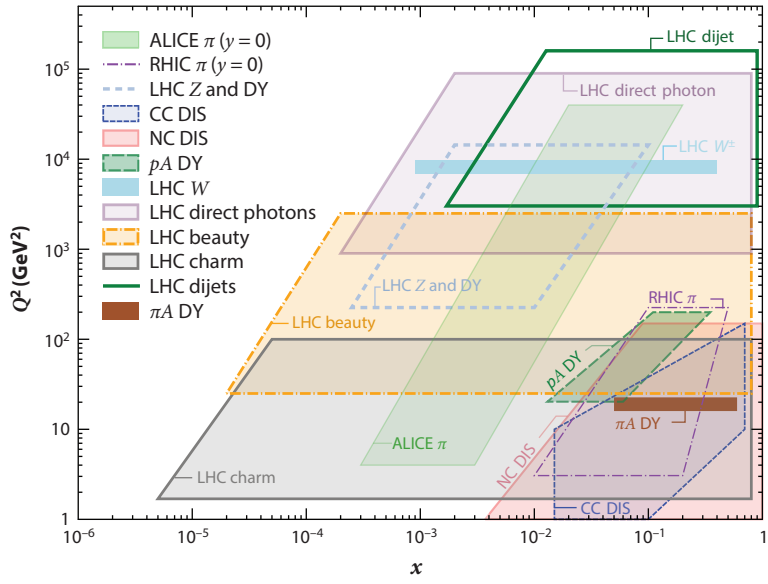

Figure 3

An illustration of the x and Q^2 regions probed by the current ℓA, πA, and pA data included in the global analyses of nuclear parton distribution functions. Abbreviations: CC DIS, charged-current deep inelastic scattering; DY, Drell–Yan process; LHC, Large Hadron Collider; NC DIS, neutral-current deep inelastic scattering; RHIC, Relativistic Heavy Ion Collider.

include (almost) no collider data, we focus on nCTEQ15HQ, EPPS21, and nNNPDF3.0 in the following. **Figure 3** shows the regions of x and Q^2 covered by the world data and included in these analyses. As can be appreciated from the plot, the LHC data taken during the first decade of pA runs have radically expanded the available range in both x and Q^2 and have diversified the global analysis. With this large coverage, the question of process independence of nuclear PDFs can now be addressed much more convincingly.

Figure 4 compares the nuclear modifications of the lead nucleus PDFs at $Q^2 = 10$ GeV^2 from EPPS21 (51), nCTEQ15HQ (50), and nNNPDF3.0 (52). Qualitatively, there is good overall agreement between all three within the 90% CL uncertainty bands. Closer inspection nevertheless reveals significant differences both among the central values and the widths of the uncertainty bands in several distributions and x regions. Due to the precise fixed-target NC DIS data, the nuclear modifications of u and d quarks (and, to a lesser extent, antiquarks) are well constrained at $x \sim 0.1$, though the widths of the error bands differ by at least a factor of two. Larger uncertainties remain in the Fermi motion region, in particular in nNNPDF3.0, where the x dependence at the starting scale is not parametric. In the case of $\bar{u}$ and $\bar{d}$, the fit nCTEQ15HQ comes with the smallest uncertainties, which could, however, be due to not fitting $\bar{u}$ and $\bar{d}$ separately. Thanks to the LHC data, the gluon uncertainties are now much better constrained than in the previous rounds of global fits down to $x \sim 10^{-5}$, and they also affect the sea (anti)quarks and their uncertainties at higher Q^2. Below $x \sim 10^{-3}$, the nCTEQ15HQ and nNNPDF3.0 gluon uncertainty bands do not overlap (as reflected in some of the plots comparing theoretical predictions and LHC data in Section 5). The largest uncertainties are seen for the strange quark distributions, which are constrained only by—to some extent problematic—neutrino data and by LHC weak boson data, where the strange quark originates, however, mostly from gluon splittings. In **Supplemental Figure 1**, we provide a comparison of the absolute nuclear PDFs.

Supplemental Material >

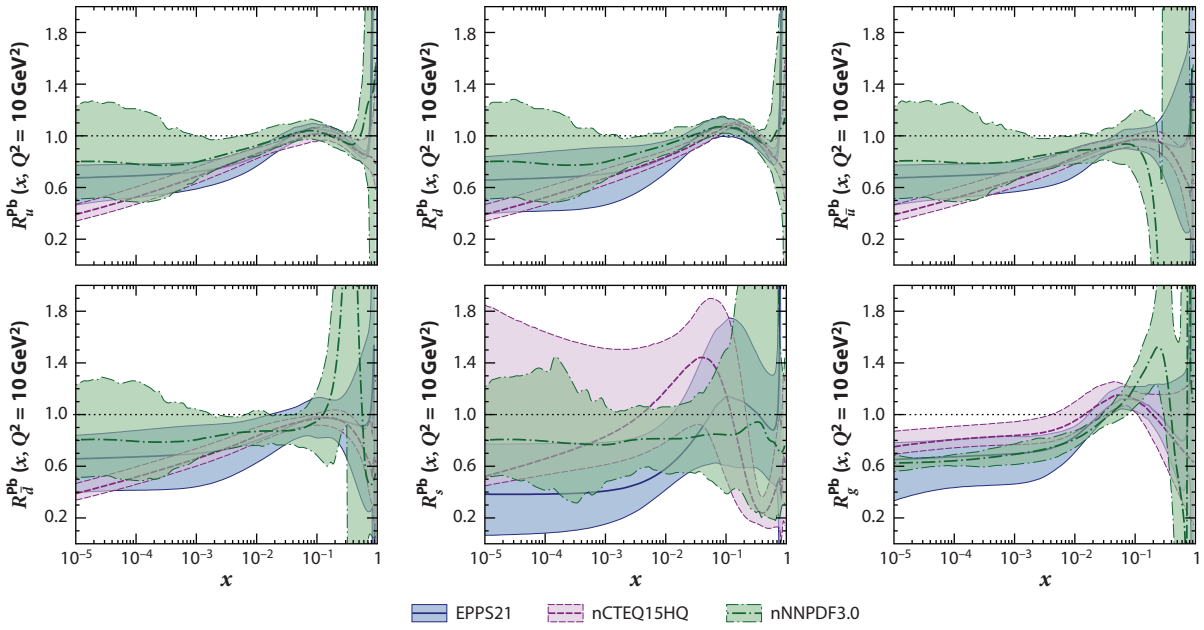

Figure 4

Comparison of the ^{208}Pb nuclear modifications resulting from the EPPS21 (*solid blue lines*) (51), nCTEQ15HQ (*dashed purple lines*) (50), and nNNPDF3.0 (*dotted-dashed green lines*) (52) global analyses of nuclear PDFs—that is, the PDFs of lead divided by the summed PDFs of 82 free protons and 126 free neutrons. Uncertainty bands (*shaded areas*) correspond to 90% confidence levels. Abbreviation: PDF, parton distribution function.

4. FIXED-TARGET DATA

We now turn to a detailed discussion of the available experimental data and their impact on global nuclear PDF analyses in roughly chronological order. We first focus on the early fixed-target NC and CC DIS and DY data, but we also review the more recent JLab DIS data.

4.1. Early Deep Inelastic Scattering Data and Constraints on Quarks

Measurements of fixed-target electron and muon NC DIS on various nuclei from SLAC, EMC, and NMC, as well as from the BCDMS, FNAL, and HERMES experiments, form the backbone of global nuclear PDF determinations (2). Most of these datasets can be fitted with an excellent χ^2/df, but a few of them, such as the 1988 EMC measurement of $F_2^{\mathrm{Sn}}/F_2^{\mathrm{D}}$ (114), are difficult to describe (41, 50, 52, 80, 81, 115). In contrast, the NMC $F_2^{\mathrm{Sn}}/F_2^{\mathrm{C}}$ data (14, 116) can be fitted well (50–52, 80, 81). Other outliers include the E665 data (117) on $F_2^{\mathrm{C}}/F_2^{\mathrm{D}}$, $F_2^{\mathrm{Ca}}/F_2^{\mathrm{D}}$, and $F_2^{\mathrm{Pb}}/F_2^{\mathrm{D}}$, though, for instance, the ratio $(F_2^{\mathrm{Pb}}/F_2^{\mathrm{D}})/(F_2^{\mathrm{C}}/F_2^{\mathrm{D}}) = F_2^{\mathrm{Pb}}/F_2^{\mathrm{C}}$ is consistent with the NMC data (116).

The kinematic reach of these fixed-target data, $x \gtrsim 5 \times 10^{-3}$ and $Q^2 \lesssim 140\ \mathrm{GeV}^2$, is naturally more limited than that of the HERA experiments H1 and ZEUS, which provide the bulk of the data in free-proton analyses (54, 55). In addition, cuts are often applied to limit the effects of TMCs and other higher-twist corrections, which could be larger in nuclear reactions (72). At low Q^2, NC DIS is governed by virtual photon exchange. In the kinematic region of fixed-target experiments, the cross section

$$\frac{\mathrm{d}^2\sigma^{\ell A}}{\mathrm{d}x\mathrm{d}Q^2} = \frac{4\pi\alpha^2}{Q^4}\left[F_2^A(x,Q^2)\left(\frac{y^2}{2}+1-y-\frac{xyM^2}{s-M^2}\right)-xy^2F_{\mathrm{L}}^A(x,Q^2)\right] \qquad 32.$$

is dominated by the structure function $F_2^A(x, Q^2)$, which at LO is sensitive only to the squared charge-weighted sum of quarks and antiquarks (see Section 1). As the ℓA DIS data that enter the global fits are given in terms of ratios,

$$\frac{\mathrm{d}\sigma^{\ell A_1}}{\mathrm{d}\sigma^{\ell A_2}} \approx \frac{F_2^{A_1}}{F_2^{A_2}}, \tag{33.}$$

where A_2 is typically D or carbon (C), the DIS data can mainly directly constrain the overall nuclear modification of valence and sea quarks. The contributions of gluons enter the cross section only at order α_s, and the direct constraints for the gluon densities are therefore weak. However, the gluons drive the Q^2 dependence of $F_2^A(x, Q^2)$ at small values of x (118):

$$\frac{\mathrm{d}F_2^A(x, Q^2)}{\mathrm{d}\log Q^2} \approx \frac{10\alpha_s(Q^2)}{27\pi} x f_g^A(2x, Q^2),\ x \to 0. \tag{34.}$$

Through this relation, it was understood early on that the Q^2 dependence of the ratios $F_2^{\mathrm{Sn}}/F_2^{\mathrm{C}}$ measured by the NMC Collaboration (14) around $x \approx 0.01\ldots0.02$ and $Q^2 \approx 1\ldots10$ GeV^2 can constrain the A dependence of the gluon nuclear modifications (119). In particular, a very strong A dependence of gluons would contradict the measured positive Q^2 slopes of $F_2^{\mathrm{Sn}}/F_2^{\mathrm{C}}$. There are also similar HERMES data for $F_2^{\mathrm{Kr}}/F_2^{\mathrm{D}}$ (120), but the Q^2 lever arm is not as long in the perturbative regime. In principle, the longitudinal structure function F_{L}^A carries a direct sensitivity to the gluon (121), but data are scarce (122). Likewise, the cross sections for charm production would provide more direct information on the gluons—modulo a possible intrinsic charm PDF (123)—but not many data are available (124, 125).

Since F_2^A probes predominantly the squared charge-weighted sum of quark PDFs, the flavor decomposition is also difficult to pin down. To understand this, let us write, for instance, the valence quark distributions as follows:

$$f_{u_v}^A = R_v^A \left(\frac{Z}{A} f_{u_v}^p + \frac{A-Z}{A} f_{d_v}^p\right) + \delta R_v^A \left(\frac{2Z}{A} - 1\right) \frac{f_{u_v}^p f_{d_v}^p}{f_{u_v}^p + f_{d_v}^p}, \tag{35.}$$

$$f_{d_v}^A = R_v^A \left(\frac{Z}{A} f_{d_v}^p + \frac{A-Z}{A} f_{u_v}^p\right) - \delta R_v^A \left(\frac{2Z}{A} - 1\right) \frac{f_{u_v}^p f_{d_v}^p}{f_{u_v}^p + f_{d_v}^p}, \tag{36.}$$

where $R_v^A \equiv (R_{u_v}^{p/A} f_{u_v}^p + R_{d_v}^{p/A} d_{d_v}^p)/(f_{u_v}^p + f_{d_v}^p)$ is an average nuclear modification of the valence quarks, and the difference is $\delta R_v^A \equiv R_{u_v}^{p/A} - R_{d_v}^{p/A}$. The first terms dominate, and thus the large-x data constrain very tightly the average modification R_v^A. Having data for several different combinations of Z and A, combined with the fact that the nuclear effects are expected to scale with A, will give also constraints on δR_v^A and thereby to the mutual differences of nuclear effects in up and down valence quarks. The same reasoning naturally applies for the up and down sea quarks. Equations 35 and 36 also clearly demonstrate the anticorrelation of the nuclear effects between up and down quarks. For an isoscalar nucleus $Z = A/2$, the up and down quark distributions are always equal.

To facilitate the interpretation of nuclear effects, many early NC DIS experiments corrected their data for isospin effects—that is, for the unequal numbers of protons and neutrons in heavy nuclei compared with the deuteron. These data were long taken at face value (and are still done so, e.g., in nNNPDF3.0) and fitted by setting $Z = N = A/2$. However, this is not necessary in global fits and has in the past even caused some confusion about the nuclear valence quark modification (97). Comparing $F_2^A = [ZF_2^{p/A} + (A-Z)F_2^{n/A}]/A$ (see Equation 11) with the isoscalar expression

$\hat{F}_2^A = [F_2^{p/A} + F_2^{n/A}]/2$ leads to $\hat{F}_2^A = \beta F_2^A$ with

$$\beta = \frac{A}{2}\left(1 + \frac{F_2^{n/A}}{F_2^{p/A}}\right) / \left(Z + (A-Z)\frac{F_2^{n/A}}{F_2^{p/A}}\right). \quad 37.$$

The experiments then assumed $F_2^{n/A}/F_2^{p/A} = F_2^n/F_2^p$ and parameterized this ratio from DIS data on protons and deuterons with, for instance, $1 - 0.8x$ (11) or $0.92 - 0.86x$ (126). These functions then allow one to calculate β and either apply it to the theoretical calculations (105) or remove the isoscalar correction from the data and fit the true nuclei (26, 97).

4.2. Deep Inelastic Scattering at High x and Nuclear Effects in the Deuteron

Recent precise JLab measurements taken with 6- to 10-GeV electron beams on various nuclear targets (127–130) at low to intermediate Q^2 and W^2 can considerably reduce the nuclear PDF uncertainties in the high-x region (26, 131). However, this region requires good control over potential TMCs (see Section 2.2) and other higher-twist corrections (see Section 2.3), hadronic resonances, and (as with the other fixed-target DIS data) nuclear effects in the deuteron. TMCs scale with powers of M_N^2/Q^2 and are thus suppressed even for heavy nuclei. Their leading effect is a shift in the probed momentum fraction x_N to the Nachtmann variable ξ_N (69). When the kinematic cuts are relaxed from $Q^2 > 4$ GeV^2 and $W^2 > 12.25$ GeV^2 to $Q^2 > 1.69$ GeV^2 and $W^2 > 2.89$ GeV^2, the subleading target-mass effects provide a uniform shift of less than 1% for all nuclei, leaving the ratios $F_2^A/F_2^{\rm D}$ unaffected (26). Some global analyses therefore include the JLab data with lower kinematic cuts (26, 51, 81, 131). The compatibility of the JLab data with global fits can be taken as evidence that other higher-twist effects can be neglected (51). They can, however, also be parameterized as

$$F_2^A(x,Q) \to F_2^A(x,Q)\left[1 + \frac{A^{1/3} h_0 x^{h_1}(1 + h_2 x)}{Q^2}\right], \quad 38.$$

where the values $\{h_0, h_1, h_2\} = \{-3.3\ \text{GeV}^2, 1.9, -2.1\}$ come from the CJ15 proton PDF analysis (132) and $A^{1/3}$ scaling is assumed (see Section 2.3). This form of higher-twist corrections leads to a slight reduction of F_2^A at intermediate $x \sim 0.3$ and a substantial enhancement at high x and $Q^2 < 16$ GeV^2, and it improves the global χ^2 by around 3% within the nCTEQ15HIX global analysis (26).

In the resonance region at very low $W^2 \in [1.21; 2.89]$ GeV^2, the nuclear effects at large x are surprisingly similar to those in DIS, which may signal the applicability of quark hadron duality due to the averaging over nuclear resonances (133). As described in Section 1, Fermi motion can be accounted for by a convolution of the nucleon structure function with the nucleon momentum distribution or effectively by a rescaling of the variable x. The rise of $F_2^A/F_2^{\rm D}$ at large x can be well described by the following parameterization (26):

$$x' = x - \varepsilon\, x^\kappa \log_{10} A. \quad 39.$$

As the PDFs decrease with x, the negative shift ensures that the transformed function is larger than the unmodified one and nonvanishing as $x \to 1$. The overall size of the rescaling effect is controlled by $\varepsilon, \kappa > 0$ ensures that only the large-x region is modified, and the $\log_{10} A$ term implies an increasing modification across the full range of nuclear A values from the proton ($A = 1$) to lead ($A = 208$). A good description of the JLab data is obtained with $\kappa = 10$ and $\epsilon \sim 0.03$. While this suggests that it may be possible to expand the kinematic reach to $W^2 < 2.89$ GeV^2, the resonance region is currently avoided in all global fits (see **Table 1**).

Given that the nuclear DIS data are usually presented as ratios of $F_2^A/F_2^{\rm D}$ and that most global fits of proton PDFs use deuteron data as well, good control over the nuclear effects in the deuteron

is required. The deuteron is much more loosely bound than heavier nuclei and is therefore often approximated as an isoscalar combination of a free proton and neutron. However, its structure at large x is still modified by Fermi motion, nuclear binding, and off-shell effects, while at small x rescattering still induces some shadowing. These nuclear effects are of the order of a few percent in the available DIS data and below 1% in the available DY data. They can be accounted for in different ways; one option is to use a free-proton baseline fitted to deuteron DIS data without nuclear corrections (51). In this case the deuteron nuclear effects are, to some extent, fitted into the u versus d quark flavor separation. Then no additional nuclear correction should be applied, although this also has been done, and in such a case the effect of double counting should be quantified (81). Alternatively, one can rescale the fitted F_2^A/F_2^{D} data by a ratio of F_2^{D}/F_2^p (26) as modeled, for instance, within the CJ15 global proton analysis (132; see also 134). If this is done with a free-proton PDF that already includes deuteron data (26), the same deuteron correction should be applied in both cases (86). A third possibility is to fit the free proton without deuteron data and then the deuteron in the same way as the other nuclei (80). The theoretical treatment of the deuteron affects the description of all NC DIS data and thus also the question of the compatibility of CC DIS data with NC DIS and electroweak boson production at the LHC (see Sections 4.4 and 5.1).

4.3. The Drell–Yan Process and Constraints on Antiquarks

The DY process—the inclusive production of electroweak gauge bosons in hadron collisions, followed by a leptonic decay of the gauge boson—has been of enormous historical importance for the quark flavor separation in protons (53). For heavier nuclei, fixed-target measurements have been made in the FNAL E605 (135), E772 (32), and E866 (136) experiments in pA collisions covering several nuclei from carbon to tungsten in the kinematic range $x > 10^{-2}$ and dilepton mass $M_{\ell\ell} < 15$ GeV. At such low values of $M_{\ell\ell}$, well below the Z boson peak, the DY process is dominated by an off-shell intermediate photon with the cross section differential in the lepton pair rapidity $y_{\ell\ell}$ and $M_{\ell\ell}$ given by

$$\frac{\mathrm{d}^2\sigma_{\mathrm{DY}}^{pA}}{\mathrm{d}M_{\ell\ell}\mathrm{d}y_{\ell\ell}} \sim \sum_q e_q^2 \left[f_q^p(x_1) f_{\bar{q}}^A(x_2) + f_{\bar{q}}^p(x_1) f_q^A(x_2) \right] \text{ with } x_{1,2} = \frac{M_{\ell\ell} e^{\pm y_{\ell\ell}}}{\sqrt{s}}, \tag{40.}$$

which tests a squared charge-weighted combination of quarks and antiquarks. Fixed-target experiments generally have larger acceptance in the $x_1 \gg x_2$ region, where x_1 is defined with respect to the proton beam, and in this case the first term in Equation 40 is the dominant one, where $f_q^p(x_1)$ is mostly determined by the valence quark content of the proton. For isoscalar nuclei, $f_{\bar{u}}^A = f_{\bar{d}}^A$, and the cross section ratios between pA and pD collisions, measured in the E772 and E866 experiments, become

$$\left.\frac{\mathrm{d}\sigma_{\mathrm{DY}}^{pA}}{\mathrm{d}\sigma_{\mathrm{DY}}^{p\mathrm{D}}}\right|_{\text{isoscalar } A} \approx \frac{f_{\bar{u}}^A(x_2)}{f_{\bar{u}}^{\mathrm{D}}(x_2)} = \frac{f_{\bar{d}}^A(x_2)}{f_{\bar{d}}^{\mathrm{D}}(x_2)}. \tag{41.}$$

As a result, the measured DY ratios are sensitive to the nuclear modifications of sea quark distributions at $x_2 \sim 0.03\ldots0.3$ (137, 138). In principle, the dependence on $M_{\ell\ell}$ should also retain sensitivity to the gluon through the DGLAP evolution. Combining the photon-mediated DIS and DY measurements historically provided the first chance to disentangle the nuclear effects in valence and sea quarks, leading to the conclusion that there was not such a clear antishadowing effect in sea quarks as there was for valence quarks. The E605 data (135) for p+Cu collisions are given in

terms of absolute cross sections and are often used in fits of proton PDFs. In the future, fixed-target pA data from the FNAL E906/SeaQuest experiment (139) are expected to improve the precision of the available data. The renewed facilities at RHIC should also enable new measurements of the DY process (140), and similar measurements are planned at the LHCb experiment at the LHC (141).

In principle, the pion–nucleus DY process has the potential to constrain the flavor decomposition of the valence quarks (142). It depends on the pion PDFs, but this dependence cancels largely in ratios of nuclear cross sections. Unfortunately, the precision of the πA DY data from the CERN NA3 (143), NA10 (144), and FNAL E615 (145) experiments is not high enough to provide significant discrimination power on top of the DIS data. However, the new CERN-based facility AMBER (146) may be able to improve upon the current precision.

4.4. Neutrino Deep Inelastic Scattering Data and Flavor Separation

Because of the weak nature of neutrino interactions, heavy nuclear targets such as iron or lead have traditionally been used to obtain CC DIS data with sufficient statistics. Despite the nuclear targets, these data have routinely been included in global analyses of proton PDFs, with or without nuclear corrections (147), and have formed the principal constraint for a possible s versus $\bar{s}$ asymmetry (148). In addition, determinations of the weak mixing angle in neutrino DIS (149) have relied on a sufficient understanding of the nuclear structure (150–152). In the case of CC neutrino DIS, the differential cross section is

$$\frac{\mathrm{d}^2\sigma^{\nu,\bar{\nu}A}}{\mathrm{d}x\mathrm{d}y} = \frac{G_\mathrm{F}^2 M_W^4}{2\pi xyQ^2}\left(\frac{Q^2}{Q^2+M_W^2}\right)^2\left[\left(1-y-\frac{x^2y^2M_N^2}{Q^2}\right)F_2^{\nu,\bar{\nu}A}+y^2xF_1^{\nu,\bar{\nu}A}\pm\left(y-\frac{y^2}{2}\right)xF_3^{\nu,\bar{\nu}A}\right], \quad 42.$$

where G_F is the Fermi constant, M_W is the W boson mass, and the plus–minus sign is taken for incoming neutrinos (+) and antineutrinos (−). At LO and high Q^2,

$$\mathrm{d}^2\sigma^{\nu A} \propto \left(f_d^A + f_s^A + f_b^A\right) + (1-y)^2\left(f_{\bar{u}}^A + f_{\bar{c}}^A\right), \quad 43.$$

$$\mathrm{d}^2\sigma^{\bar{\nu}A} \propto \left(f_{\bar{d}}^A + f_{\bar{s}}^A + f_{\bar{b}}^A\right) + (1-y)^2\left(f_u^A + f_c^A\right). \quad 44.$$

Due to the suppressing factor $(1-y)^2$, there is an increased sensitivity to the strange quark distribution in comparison to NC charged lepton DIS, particularly for antineutrinos. In addition, the up and down quarks enter the cross sections with different weights than in the case of NC charged lepton DIS. Adding the neutrino data thus helps in constraining the differences between nuclear effects in up and down quarks (see Equations 35 and 36). However, in a global fit these data are also sensitive to the assumed proton PDFs.

Data on inclusive neutrino DIS have been taken by, for instance, the CDHSW (153), CCFR (154), and NuTeV (13) experiments on iron and by the CHORUS experiment on lead (155). Charm production has been measured as well through muonic decays of produced charmed hadrons (see **Figure 1**) (156, 157). In principle, all of these data should be relevant for nuclear PDFs, but they are only partially included in global analyses due to concerns about possible mutual tensions between the neutrino datasets, tensions with the charged lepton DIS data, and the fact that some of these neutrino data are in some cases already used in the proton PDF fits that are used as baselines in the fits of nuclear PDFs. Another difficulty is that there are no references from νp or νD scattering, so the data are reported as absolute cross sections.

Figure 5 shows neutrino (**Figure 5*a***) and antineutrino (**Figure 5*b***) cross sections divided by the theoretical NLO predictions with CT18A proton PDFs in the SACOT-χ scheme, including

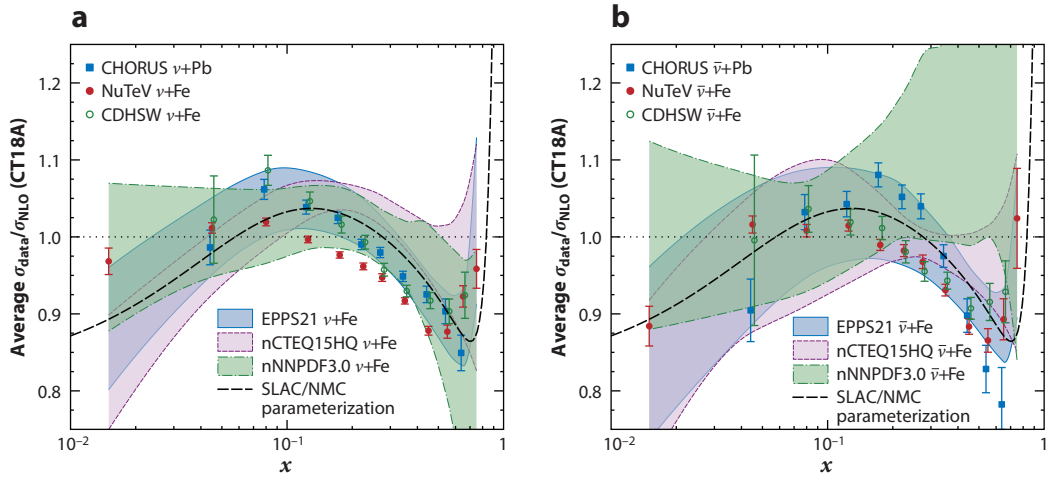

Figure 5

Average ratios of (*a*) neutrino cross sections and (*b*) antineutrino cross sections, as measured in the CHORUS (155), NuTeV (13), and CDHSW (153) experiments, to a theoretical next-to-leading-order prediction with CT18A parton distribution functions within the kinematic range of $Q^2 > 4$ GeV^2 and $W^2 > 12.25$ GeV^2. The data are compared with EPPS21 (51), nCTEQ15HQ (158), and nNNPDF3.0 (52) predictions. The SLAC/NMC neutral-current deep inelastic scattering parameterization (see Equation 2) is also shown as a reference.

approximate TMCs and electroweak corrections as used in Reference 45. The ratios are evaluated as weighted averages over $Q^2 > 4$ GeV^2 and $W^2 > 12.25$ GeV^2 as in Reference 158. The obtained ratios are compared with predictions for NuTeV data using EPPS21 (51), nCTEQ15HQ (158), and nNNPDF3.0 (52). Also, the SLAC/NMC NC DIS parameterization (Equation 2) is shown for comparison. Tensions between different datasets, nuclear PDFs, and data and theory are clearly visible. The largest differences between nuclear PDFs occur in the case of antineutrino DIS at $x \gtrsim 0.2$, where the nNNPDF3.0 values are significantly above those from EPPS21 and nCTEQ15HQ. This can be explained by the large enhancement of $\bar{d}$ and s densities in nNNPDF3.0 compared with those in EPPS21 and nCTEQ15HQ (see **Supplemental Figure 1**). From the datasets, the NuTeV neutrino data in particular stand out from the others, but larger deviations between the CHORUS and CDHSW data can also be observed if less restrictive kinematic cuts are imposed and electroweak corrections are neglected (158). To some degree, the observed tensions can be alleviated by normalizing the data by the cross sections integrated over x and y (45, 47), by neglecting the NuTeV systematic error correlations, and by introducing additional cuts in x (158). The current consensus seems to be that at least the CHORUS data can be included in global analyses without significant tensions. In addition, the charm dimuon data are used in nCTEQ15ν (158) and nNNPDF3.0 (52), and the CDHSW data are used in TUJU21 (80) and KSASG20 (81). There have been speculations about differences in nuclear shadowing in CC and NC processes (20), at least at low Q^2 (19), but $W^{\pm}$ and Z production at the LHC probing nuclear PDFs at significantly higher Q^2 can be fitted well in global analyses.

Supplemental Material >

In the future, novel neutrino–nucleus DIS data may become available through dedicated experiments measuring neutrinos produced in high-luminosity pp collisions at the LHC. Indeed, the first observations of such collider neutrinos have already been made by the FASER (159) and SND@LHC (160) Collaborations. The impact of such future measurements on nuclear PDFs has been considered recently in Reference 161.

5. COLLIDER DATA

We now turn to a discussion of the LHC (p+Pb) as well as the RHIC (D+Au) measurements used in global fits of nuclear PDFs. Within collinear factorization, the pA (or DA) cross sections

$$\mathrm{d}\sigma(pA \to \mathcal{O} + X) = \sum_{i,j[,k]} f_i^p \otimes f_j^A \otimes \mathrm{d}\hat{\sigma}(ij \to \mathcal{O}, [k] + X) \, [\otimes D_k^{\mathcal{O}}] \qquad 45.$$

for the observable $\mathcal{O}$ involve convolutions of (nuclear) PDFs $f_{i,j}^{p_rA}$ with perturbative partonic cross sections $\mathrm{d}\hat{\sigma}$ and, in the case of inclusive light or heavy flavor hadron (h) production, nonperturbative FFs D_k^h. These are numerically costly and in many cases must be evaluated with precomputed grids (162). Since one of the colliding objects is a proton and the nuclear PDFs also depend on the proton, absolute LHC cross sections depend directly and indirectly on the proton PDFs. To reduce this dependence and also cancel other theoretical and experimental uncertainties, the nuclear modification ratio,

$$R_{pA} \equiv \mathrm{d}\sigma(pA \to \mathcal{O} + X) \big/ \mathrm{d}\sigma(pp \to \mathcal{O} + X), \qquad 46.$$

and the forward-to-backward ratio,

$$R_{\mathrm{FB}} \equiv \mathrm{d}\sigma(pA \to \mathcal{O} + X)\big|_{y>0} \big/ \mathrm{d}\sigma(pA \to \mathcal{O} + X)\big|_{y<0}, \qquad 47.$$

are often introduced. Here y refers to the rapidity of the observable $\mathcal{O}$.

In fits of collider data, normalization uncertainties originating from the luminosity measurements play a special role. For example, the measured and calculated $R_{p+\mathrm{Pb}}$ values for hadron production in the $y \gg 0$ region (small x_N) at the LHC are often rather flat, and changes in the nuclear PDFs can be compensated by treating the normalization uncertainty as a correlated systematic uncertainty (see, e.g., 158). If the luminosity uncertainty is common for $y > 0$ and $y < 0$, the R_{FB} is free from this additional freedom. Today, all global fits of nuclear PDFs account for the systematic normalization uncertainties.

5.1. Electroweak Bosons

The first global analysis of nuclear PDFs to include W and Z boson data from p+Pb collisions was EPPS16 (97). However, the impact of the Run 1 ATLAS (163) and CMS (164, 165) data was still rather limited due to low statistics. Since then, measurements for these electroweak processes have been published from all four LHC experiments at $\sqrt{s} = 5.02$ TeV (Run 1) and from the ALICE, CMS, and LHCb experiments at $\sqrt{s} = 8.16$ TeV (Run 2). They are now, in different combinations, used in all recent global analyses (50–52) and also in the NNLO and model-dependent fits TUJU21 (80) and KP16 (113). The available data are summarized in **Table 2**.

At the moment, the most stringent constraints come from the Run 2 CMS W boson data (170). **Figure 6** compares the nuclear modification ratio $R_{p+\mathrm{Pb}}$ constructed from the CMS Run 2 (170, 171) measurements with NLO calculations using the EPPS21 (51), nCTEQ15HQ (50), and nNNPDF3.0 (52) nuclear PDFs. While the EPPS21 analysis included the $R_{p+\mathrm{Pb}}$ data shown in the figure, the nCTEQ15HQ and nNNPDF3.0 analyses fitted absolute p+Pb cross sections. As one can see, the spread between the different predictions is still rather significant. In comparison to a calculation with no nuclear effects—that is, 82 free protons and 126 free neutrons—the data indicate a clear sign of shadowing at forward rapidities or $x \ll 1$. The relative ordering of EPPS21, nCTEQ15HQ, and nNNPDF3.0 values follows that of the corresponding gluon shadowing in **Figure 4**. Also, as can be seen from **Figure 4**, even after inclusion of these electroweak data, the overall variation in the strange quark PDF is still quite significant, which indicates that the constraints for the nuclear strange quark PDFs are still not very strong.

Table 2 Summary of Z, $W^{\pm}$, and low-invariant-mass Z/γ^* rapidity distributions available from p+Pb collisions at the Large Hadron Collider

Dataset	nCTEQ15HQ (50)	EPPS21 (51)	nNNPDF3.0 (52)	TUJU21 (80)	KP16 (113)
Run 1					
ATLAS Z (163)	✓	✓	✓	✓	✓
CMS Z (164)	✓	✓	✓	✓	✓
ALICE Z (166)	—	—	✓[a]	—	—
LHCb Z (167)	✓	—	✓[a]	—	—
ATLAS $W^{\pm}$ (168)[b]	✓	—	—	—	✓
CMS $W^{\pm}$ (165)	✓	✓	✓	—	—
ALICE $W^{\pm}$ (166)	✓	—	✓[a]	—	—
Run 2					
CMS Z (172)	—	—	✓[a]	—	—
CMS Z/γ^* (172)	—	—	✓[a]	—	—
ALICE Z (346)	—	—	✓[a]	—	—
LHCb Z (169)	—	—	—	—	—
CMS $W^{\pm}$ (170, 171)	✓	✓[c]	✓	✓	—
ALICE $W^{\pm}$ (347)	—	—	—	—	—

[a]Added in nNNPDF3.0 (52).
[b]Preliminary data.
[c]Added in EPPS21 (51).

The CMS Run 2 measurement for Z boson production (172) reports similarly small uncertainties. However, it is not possible to obtain a good quantitative description of these data with any nuclear PDFs because of the large fluctuations of the data around midrapidity ($y_{\ell\ell} = 0$), which lead, for instance, to an R_{FB} that does not tend to unity toward $y_{\ell\ell} \to 0$ as one would expect. Along with the on-shell Z production, the CMS experiment also measured low-mass cross sections in

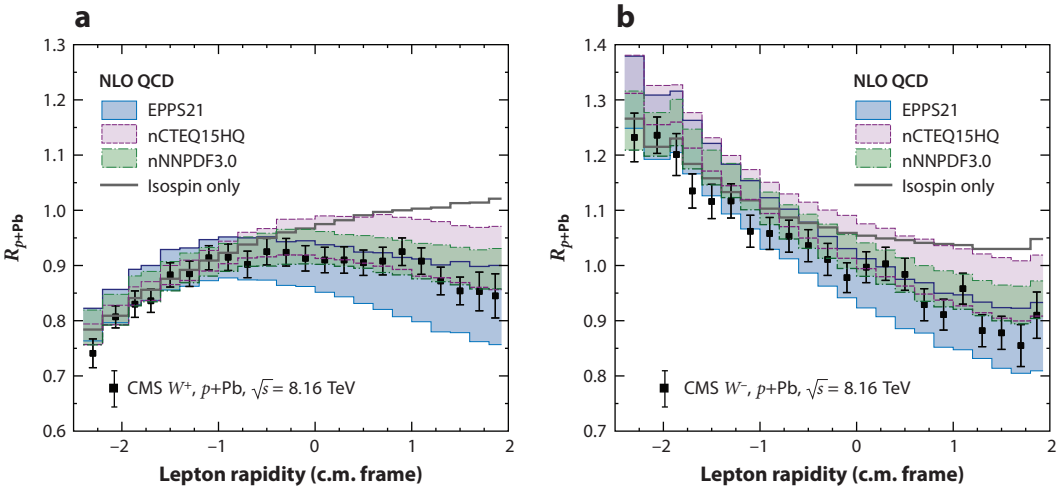

Figure 6

Nuclear modification ratios for (*a*) W^+ bosons and (*b*) W^- bosons at CMS Run 2 (170, 171) compared with EPPS21 (51), nCTEQ15HQ (50), nNNPDF3.0 (52), and a calculation with 82 free protons and 126 free neutrons. Abbreviations: c.m., center of mass; NLO, next-to-leading order; QCD, quantum chromodynamics.

the window 15 GeV $< M_{\ell\ell} <$ 60 GeV. Within the TUJU21 analysis (80), it was noticed that to simultaneously reproduce the normalization of the CMS low-mass and Z cross sections, the NNLO QCD corrections appear to be necessary. This is the first time the necessity of NNLO corrections has been seen in the case of p+Pb collisions.

Dielectron pairs have also been measured in the ALICE experiment at Run 1 in the low-mass region $M_{\ell\ell} < 3$ GeV and with $0 < p_{\mathrm{T},\ell\ell} < 8$ GeV (173), which is in principle very sensitive to the gluon density and avoids the fragmentation contribution present for real photons (174–176). Currently, the data are unfortunately still dominated by the heavy flavor (c, b) decay background, but the statistics should be improved in Run 2 and the background should be reducible with heavy flavor tagging, in particular in the LHCb experiment (141).

5.2. Photons

Another electroweak probe of nuclear PDFs is the prompt production of real photons with finite p_{T} (177–180). It proceeds directly through quark–antiquark annihilation ($q\bar{q} \to g\gamma$) and the QCD Compton process ($qg \to q\gamma$), which dominates at large p_{T} (181). The radiation of massless photons from quarks in pure QCD processes gives rise to a photon fragmentation contribution (182–184), which is important at small p_{T}. Isolating the photon and thus reducing the surrounding hadronic energy (e.g., in a cone) suppresses the fragmentation component and also the nonprompt photon background from hadronic (in particular pion) decays (185).

Direct photon production in proton (and pion) nucleus collisions was first measured in fixed-target mode at the Fermilab E706 experiment (186). PHENIX (187) and STAR (188) measurements in D+Au collisions at RHIC are also available. At the LHC, high-p_{T} isolated photons in p+Pb (189) collisions have been measured by the ATLAS Collaboration. Ratios of these ATLAS cross section measurements to the corresponding pp data (190) have a reduced sensitivity to missing higher-order effects, FFs, and proton PDFs. They can be reasonably described at NLO QCD and are therefore included in the nNNPDF3.0 (52) analysis. The absolute pp and p+Pb cross sections are, however, underestimated by NLO QCD by up to 30% at the lowest values of $p_{\mathrm{T}} \sim 20$ GeV. This could indicate the need to include NNLO corrections (191). The impact of the ATLAS prompt photon data in the global fit is small compared with dijet and heavy flavor production due to larger uncertainties. Looking toward the future, the ALICE Collaboration has proposed building a new forward calorimeter (FoCal) for LHC Run 4 (192, 193), which would be optimized for direct photons in the rapidity region $3.2 < \eta < 5.8$.

5.3. Light Hadrons

As discussed in Section 4, before the beginning of the LHC era, not much was known about the nuclear gluon PDFs. The first direct evidence for the presence of shadowing, antishadowing, and the EMC effect in gluons came from inclusive hadron production in DA collisions at RHIC. This process involves a gluon contribution already at LO and is therefore a candidate to constrain the nuclear gluons in the perturbative region—that is, when the transverse momentum of the hadron is sufficiently large. This possibility was first discussed in References 40, 95, and 194 in the light of early RHIC data (195–198). The first global analysis to fit this type of data was EPS08 (96), which included negatively charged hadron data from the BRAHMS experiment (196) as well as PHENIX (197, 199) and STAR (200) pion data. It was, however, noticed that the rapidity dependence of the BRAHMS data at low values of $p_{\mathrm{T}} \gtrsim 2$ GeV was too strong to be optimally reproduced within a global fit, inducing tensions with the NMC data for $F_2^{\mathrm{Sn}}/F_2^{\mathrm{C}}$ (14). These negatively charged hadron BRAHMS data were eventually dropped from the EPS09 analysis (42) when it was noticed that it was difficult to reproduce even the pp reference data. The later EP(P)S analyses (51, 97) have

retained only PHENIX π^0 data (199), while nCTEQ15 (43) included these and updated STAR π^0 data (188).

Inclusive hadron production is sensitive not only to PDFs but also to the final-state hadronization encoded in the parton-to-hadron FFs. In the EP(P)S and nCTEQ fits, the FFs are taken from global fits of hadron production in e^+e^-, eN, and pp collisions (201–205). The sensitivity to FFs was recently studied within the nCTEQ15WZ+SIH (88) analysis, which also propagated the FF uncertainties, when available, into the fit. Combining the latest RHIC D+Au data on pions, kaons, and η mesons (188, 199, 200, 206) with the corresponding ALICE p+Pb measurements (207–209) led to a consistent description of the data and to a considerable reduction in the nuclear gluon uncertainty. The fact that consistent global fits down to $p_T > 3$ GeV are possible can also be taken as an indication that higher-twist final-state rescattering (see Section 2.3) is indeed a subleading effect. The inclusive hadron data have therefore been retained as well in the latest nCTEQ15HQ analysis (50). Alternatively, the RHIC data in D+Au collisions have also been interpreted in terms of nuclear-modified FFs (210, 211), which were used in the DSSZ (115) global analysis of nuclear PDFs and resulted in reduced nuclear effects for the gluon PDF.

The latest LHC measurements of $R_{p+\mathrm{Pb}}$ for high-p_T neutral pion production come from the LHCb experiment (212), complementing the ALICE midrapidity data with forward/backward measurements. While the forward-rapidity data agree with the NLO predictions with nuclear PDFs, there appears to be a slight normalization difference between the predictions and the LHCb data at negative rapidities. The preliminary LHCb data for η mesons look consistent with nuclear PDFs (213). These data are not yet included in global fits. The LHCb forward pion data also agree well with the LHCb forward charged hadron ($h^\pm$) data (214). The corresponding backward data are, however, in disagreement with the nuclear PDF predictions, which hints that the baryon production in the lead-going direction cannot be described solely within the factorization. The same issue is visible at midrapidity as well (215–218) and is more pronounced at RHIC (197, 198). Moreover, it has been noticed that even in simpler pp collisions at LHC energies, the collinear factorization around midrapidity appears to be applicable only at $p_T \gtrsim 10$ GeV for $h^\pm$ production (219). As a result, only the production of mesons is considered in global fits of nuclear PDFs.

5.4. Jets

Jet measurements in p+Pb collisions probe the intermediate- to large-x regime of nuclear PDFs at large interaction scales ($Q^2 \gtrsim 10^3$ GeV^2). A complication in p+Pb compared with pp collisions is the significantly larger background from the underlying event. Indeed, in Glauber-type models, an average p+Pb collision contains around 7 ± 5 pN interactions (220). To reduce the model dependence regarding how the multiparton interactions (MPIs) are dealt with, the jet p_T must therefore be large enough, or the jet cone must be small enough, to reduce the probability of particles from MPIs to occupy the same phase space. However, at cone sizes that are too small, the jet cross sections become unstable due to an incomplete cancellation of infrared divergences. Also, the hadronization corrections, which tend to widen the partonic jets, grow. While in pp collisions all nonperturbative corrections are applied to the theoretical predictions, the p+Pb data have already been subtracted for the backgrounds from MPIs. This works rather well; the obtained ratios $R_{p+\mathrm{Pb}}$ are broadly consistent with the expectations from nuclear PDFs.

Currently, the most constraining data are the Run 1 CMS dijet data differential in the average p_T (p_T^{ave}) and rapidity (η_{dijet}) of the two jets (221). They supersede the earlier dijet data (222), which were included in the EPPS16 analysis. The cross sections are normalized to the rapidity-integrated cross section, so that most of the systematic uncertainties cancel. The resulting spectra in pp collisions are then so precise that they challenge the theoretical description, with the NLO

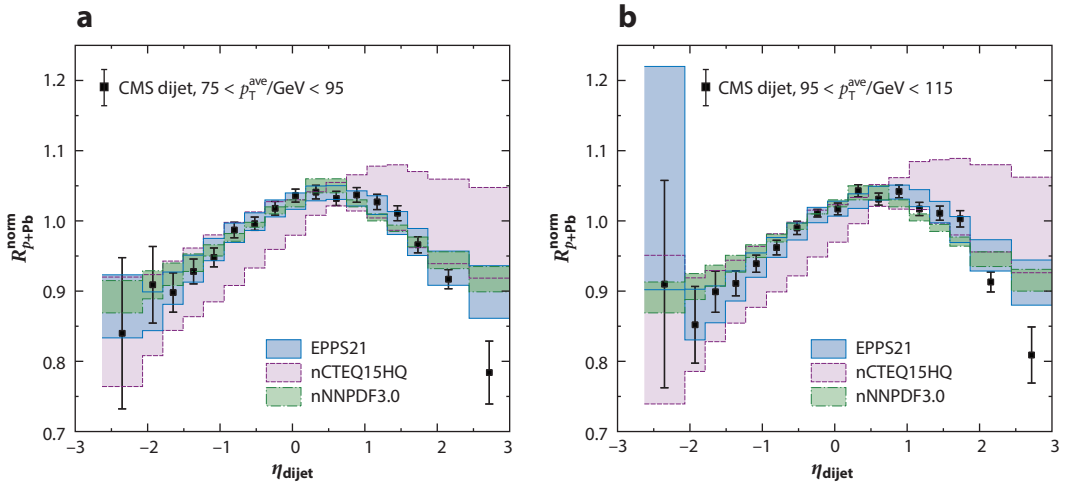

Figure 7

Comparison of the normalized CMS dijet nuclear modification ratios (221) with next-to-leading-order calculations using EPPS21, nCTEQ15HQ, and nNNPDF3.0 nuclear parton distribution functions.

perturbative QCD calculations in tension with the data (223). The impact of the NNLO corrections (224) is still unclear. Despite this tension, ratios of normalized cross sections between p+Pb and pp collisions, R^{norm}_{p+Pb}, are broadly consistent with nuclear PDFs and included in the EPPS21 and nNNPDF3.0 global fits, where they have a large impact on the gluon. However, the most forward data points at the edge of the detector acceptance indicate a suppression, which cannot be fitted. This fact calls into question the reliability of extracting nuclear PDFs from these data.

Some of the CMS dijet data are compared in **Figure 7** with NLO calculations using the EPPS21, nCTEQ15HQ, and nNNPDF3.0 fits. The depletion in the backward direction, $\eta_{dijet} \ll 0$, indicates a presence of an EMC effect for gluons (though it is partially obscured due to the contribution of valence quarks), whereas the depletion in the forward direction, $\eta_{dijet} \gg 0$, is consistent with the expected gluon shadowing. At $\eta_{dijet} \gg 0$, the less-shadowed gluons of nCTEQ15HQ seen in **Figure 4** lead to the observed higher prediction for R^{norm}_{p+Pb}. The nCTEQ15HQ error bands are the widest, as these data were not included in the fit. The EPPS21 and nNNPDF3.0 uncertainties are smaller, and near $\eta_{dijet} \sim 1$ there is even a discrepancy among them. Data for single inclusive jets are also available (225–227), but the uncertainties are clearly larger than in the dijet measurements.

5.5. Heavy Quarks and Quarkonia

The possibilities of constraining the gluon PDF with inclusive heavy flavor production at the LHC have been actively investigated in pp (228–231) as well as p+Pb (50, 89, 90, 232, 233) collisions. However, the theoretical approaches vary from one analysis to another, and this is also the case in the fits of nuclear PDFs that include heavy quark data. The EPPS group uses GM-VFNS calculations (92, 234, 235) in which heavy quarks are active partons above the mass thresholds, resumming collinear logarithms from the initial- and final-state radiation and thereby matching with the variable flavor structure of the nuclear PDFs. The nNNPDF group employs an approach that supplements fixed-order calculations with a similar, though less complete, resummation of collinear logarithms through parton showers (236–238). The nCTEQ15HQ group relies on the effective matrix element fitting (239) introduced in Section 3.1. It uses the fact that in fixed-order

Table 3 Heavy quark production data available from p+Pb collisions at the Large Hadron Collider

Observable $\mathcal{O}$	D^0	J/ψ	$\Upsilon(1S)$	$\psi(2S)$	B^0, $B^\pm$	c jet	b jet
Run 1							
ATLAS	—	✓ (251, 252)[a]	✓ (252)[a]	✓ (252)[a]	—	—	—
CMS	—	✓ (253)[a]	✓ (254)	✓ (255)[a]	—	✓ (242)	✓ (243)
ALICE	✓ (256, 257, 272)[a]	✓ (258,[a] 259,[a] 260)	✓ (261)	✓ (262)[a]	—	—	✓ (244)
LHCb	✓ (271)[a,b,c]	✓ (263)[a]	✓ (264)	—	—	—	—
Run 2							
ALICE	—	✓ (265,[a] 266)	✓ (267)[a]	✓ (268)[a]	—	—	—
LHCb	✓ (273)	✓ (269)[a]	✓ (270)[a]	—	✓ (245)	—	—
Fixed target							
LHCb	✓ (248, 249)	✓ (248, 250)	—	✓ (250)	—	—	—

[a]Included in nCTEQ15HQ (50).
[b]Included in EPPS21 (51).
[c]Included in nNNPDF3.0 (52).

calculations, the gg initial state dominates at low p_T and allows fitting both open heavy quark and quarkonium production including the hadronization process, which for quarkonia remains to be fully understood (240).

The four LHC collaborations have collected a vast dataset on D^0, B^0, $B^\pm$, J/ψ, Υ, and ψ' mesons (see **Table 3**), which allows one to extend the range in x_N to below 10^{-5}, that is, more than an order of magnitude lower than the electroweak boson production (and more than two orders of magnitude lower than the jet production) at the LHC at scales from m_c^2 to 10^3 GeV^2 (see **Figure 3**). Including these data even partially, the gluon uncertainties of nCTEQ15HQ, EPPS21, and nNNPDF3.0 have shrunk considerably below $x_N = 10^{-2}$ in comparison to those of their respective predecessors nCTEQ15WZ+SIH (88), EPPS16 (97), and nNNPDF2.0 (241). While not included in the current global fits, the CMS Collaboration has also measured c jets (242) and b jets (243), the ALICE Collaboration has measured b jets (244), the LHCb Collaboration has measured inclusive B meson (245) production (discussed, e.g., in Reference 235), and the ALICE Collaboration has measured heavy flavor decay electrons (246, 247). First heavy flavor measurements have also been carried out by the LHCb Collaboration in fixed-target mode with different nuclei (He, Ar, Ne) (248–250). This may eventually allow for the study of the A dependence of nuclear PDFs.

An important dataset in the current global fits is the LHCb Run 1 D^0 measurement (271), which is included in all three fits. In the forward direction ($y \gg 0$, small x), the nuclear modification ratio $R_{p+\mathrm{Pb}}$ shows a clear suppression that is consistent with shadowing. In the backward direction ($y \ll 0$, larger x) at the intersection between shadowing and antishadowing, $R_{p+\mathrm{Pb}}$ is closer to unity. This behavior is consistent with the CMS dijet and $W^\pm$ data. The ALICE D meson data (272) lie at midrapidity between the LHCb acceptance and have a somewhat different normalization. The recent LHCb Run 2 D^0 data (273) are consistent with nuclear PDF predictions in the forward direction (shadowing) but indicate a stronger suppression than expected in the backward direction. Given that these $R_{p+\mathrm{Pb}}$ data use a pp reference interpolated between 5 and 13 TeV, R_{FB} could arguably be more accurate. **Figure 8*a*** compares the new LHCb Run 2 measurement with the predictions obtained using EPPS21, nCTEQ15HQ, and nNNPDF3.0 PDFs in an NLO GM-VFNS calculation (234). Despite the fact that all three use the 5-TeV p+Pb D^0 data as an input, there are still significant differences among the predictions. Recently, preliminary LHCb Run 1 data on the $R_{p+\mathrm{Pb}}$ of D^+ and D_s^+ have also appeared (274). They are consistent with the D^0 results at $y \gg 0$, but the D^+ data deviate from the D^0 results at $y \ll 0$.

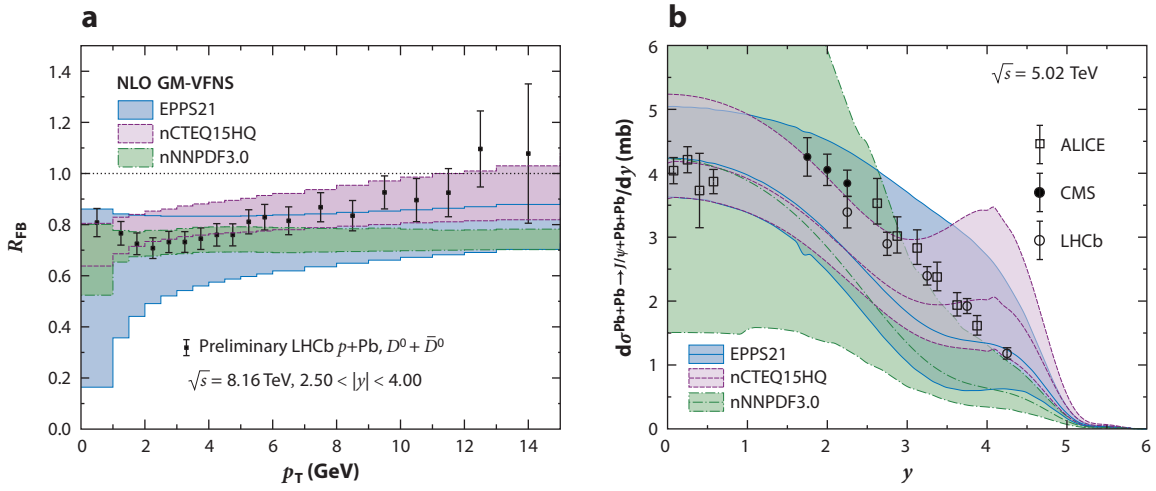

Figure 8

(*a*) The LHCb Run 2 forward-to-backward ratio (273) compared with NLO GM-VFNS calculations (234) using the EPPS21, nCTEQ15HQ, and nNNPDF3.0 nuclear PDFs. (*b*) The LHC Run 2 data (288–291) for exclusive J/ψ production in Pb+Pb collisions compared with NLO calculations using the EPPS21, nCTEQ15HQ, and nNNPDF3.0 nuclear PDFs. The factorization scales $\mu \sim 2.2$ GeV have been chosen to match the ALICE data at $y = 0$. Abbreviations: GM-VFNS, general-mass variable-flavor-number scheme; LHC, Large Hadron Collider; NLO, next-to-leading order; PDF, parton distribution function.

The prospects of using top quark production in p+Pb and Pb+Pb collisions to understand nuclear PDFs were first quantitatively discussed in Reference 275. While the large mass of the top quark renders the production cross sections small compared with those of charm and beauty production, the process was predicted to be visible at the LHC. Total top quark cross sections have thereafter been measured in the CMS (276) and ATLAS (277) experiments as well as in Pb+Pb collisions in the CMS experiment (278). The ATLAS measurement in p+Pb collisions is consistent with the nCTEQ15HQ, EPPS21, and TUJU21 but not the nNNPDF3.0 nuclear PDFs.

5.6. Exclusive and Inclusive Observables in Ultraperipheral Collisions

Ultraperipheral collisions (UPCs) of ions are interactions in which the approaching nuclei do not touch. Instead, they interact at a distance due to their strong electromagnetic fields (279, 280). In comparison to typical minimum-bias p+Pb (let alone Pb+Pb) collisions, far fewer background processes take place, and the signal processes are thus easier to isolate.

The exclusive production of J/ψ mesons in UPCs has triggered particular interest. The process is dominated by the exchange of an almost real photon. In photon–nucleus collisions, the PDFs appear already at the level of the matrix element:

$$\mathcal{M}(\gamma + A \to J/\psi + A) \sim T_g \otimes f_g^A + \sum_q T_q \otimes f_q^A. \qquad 48.$$

When squared to obtain a cross section, the latter becomes extremely sensitive to PDFs. Several LO studies have been performed in the past (281–284), but the first NLO calculations for Pb+Pb collisions have appeared only very recently (285, 286) despite the fact that the NLO coefficient functions T_g and T_q have been known for some time (287). **Figure 8*b*** compares NLO calculations from several recent nuclear PDFs with the combined experimental data from the LHC (288–291). The factorization scales have been chosen to match the ALICE data at midrapidity.

While the central theory values do not reproduce the behavior of the data particularly well, the nuclear PDF error bands are much wider than the data uncertainties. This indicates that these data should further constrain the nuclear PDFs. Unfortunately, the process is perturbatively unstable: At LO, only gluons contribute, while at NLO there is also a contribution from the quark singlet, which can even dominate at NLO. The reason is that the LO and NLO gluon contributions enter T_g with opposite signs, and there is a significant cancellation between the two. It has been argued that the theoretical uncertainties could be brought under better control by summing logarithmically enhanced contributions at small x (292) and by considering power corrections in the coefficient functions (293). There are also other theoretical uncertainties, such as those associated with modeling of the photon flux and how nuclear generalized PDFs and collinear PDFs are related (294).

In addition to exclusive observables, inclusive processes such as dijet photoproduction are sensitive to nuclear PDFs. Here, a photon emitted from one nucleus breaks up the other to produce a hadronic final state that contains two hard jets. This process has been calculated at LO with parton showers (295) and at NLO (296). Preliminary measurements from the ATLAS experiment also exist (297, 298), and the 2017 measurement qualitatively agrees with the NLO calculation. However, the ATLAS data make use of forward neutrons and rapidity gaps to resolve the photon-going direction on an event-by-event basis, and this excludes the diffractive component of UPCs, which is included in NLO calculations with standard PDFs (299). The imposed experimental conditions also require further modeling associated with preventing a Coulomb breakup of the photon-emitting nucleus and with finite-size effects.

6. OTHER DEVELOPMENTS

6.1. Electron Ion Collider

The EIC currently under construction at BNL will extend the kinematic region of lepton–nucleus scattering compared with that of fixed-target experiments by one order of magnitude in x (to a few times 10^{-4}) and Q^2 (to 10^3 GeV^2). This region is of course still considerably smaller than the one accessible at the LHC (see **Figure 3**), but the environment will be much cleaner than in pA collisions, making it easier to disentangle cold nuclear matter (i.e., leading-twist, factorizable) effects from other possible (higher-twist, nonfactorizable) contributions. In addition, the EIC will allow for measurements on a large variety of nuclei in NC and CC DIS as well as in photoproduction and with inclusive, semi-inclusive (e.g., identified hadron), and exclusive (e.g., diffractive) final states (300, 301).

While F_2^A is sensitive to the momentum distributions of gluons mainly through scaling violations (see Equation 34), the EIC will also allow for measurements of the longitudinal structure function (302, 303):

$$F_{\mathrm{L}}^A(x,Q^2) = \frac{\alpha_s(Q^2)}{2\pi}x^2\int_x^1 \frac{\mathrm{d}z}{z^3}\left[\frac{8}{3}F_2(x,Q^2) + 4\sum_q e_q^2\left(1-\frac{x}{z}\right)z f_g^A(x,Q^2)\right], \qquad 49.$$

which has a direct contribution from gluons. Measuring F_{L}^A will, however, require operation at different center-of-mass energies. The impact of inclusive NC DIS measurements at the EIC on global fits of nuclear PDFs has been investigated in References 301, 304, and 305 in the frameworks of nCTEQ15WZ, EPPS16, and nNNPDF2.0 with several nuclei and beam energy configurations (see also 104). At low Q^2, the predicted impact is significant for all partonic flavors other than the strange quark. At higher Q^2, the better-constrained gluon also leads to a better-constrained strange quark PDF. Charm tagging allows one to access the reduced charm cross

section

$$\sigma_{\mathrm{red}}^{c\bar{c}} = \frac{\mathrm{d}^2\sigma^{c\bar{c}}}{\mathrm{d}x\mathrm{d}Q^2}\frac{xQ^4}{2\pi\alpha^2[1+(1-y)^2]} = F_2^{c\bar{c}} - \frac{y^2}{1+(1-y)^2}F_{\mathrm{L}}^{c\bar{c}} \qquad 50.$$

related to the charm structure functions $F_2^{c\bar{c}}$ and $F_{\mathrm{L}}^{c\bar{c}}$, which are sensitive to the gluon and a potential intrinsic charm content in the nucleon, and thus to further reduce the uncertainties (304, 306). Dijets in DIS (307, 308) and photoproduction (309–311) as well as charm jets (312) provide further information, the latter in particular on the strange quark in CC DIS. As observed at the LHC, exclusive vector meson production is highly sensitive to the gluon (313, 314). Diffractive final states with a large rapidity gap or identified hadrons in the forward direction will allow access to the completely unknown territory of diffractive nuclear PDFs (315, 316).

Taken together, the EIC measurements will allow for a greater parametric flexibility in the x dependence at the starting scale Q_0^2 similar to that of proton PDFs, lead to a more reliable A dependence, make parameterizations possible not just in A but also in Z, and allow researchers to move away from the nuclear stability line and to study mirror nuclei. EIC measurements should also help answer the question of potential different nuclear effects in CC and NC DIS, shedding light on shadowing, gluon saturation (317–319), transverse momentum distributions (320, 321), the transition to the color glass condensate (322, 323), and the EMC effect across a wide range of A and energy scales. With polarized beams of light nuclei (^{2}H, ^{3}H, ^{3}He), even the polarized EMC effect could be investigated (300, 301).

6.2. Lattice Quantum Chromodynamics

In the nonperturbative lattice QCD approach, the four-dimensional space time is discretized and QCD regularized on a finite Euclidean lattice. Correlation functions are then computed numerically in the path integral formalism using methods adapted from statistical mechanics, and the results are extrapolated to the continuum and infinite-volume limits. To make contact with experimental data, lattice QCD calculations must demonstrate control over all sources of systematic uncertainty, including discretization effects, extrapolation from unphysical pion masses, finite-volume effects, and renormalization of composite operators.

Light cone quantities like PDFs cannot be calculated directly on a Euclidean lattice. Instead, the traditional approach has been to determine the matrix elements of local twist-two operators that can be related to the Mellin moments of PDFs. In principle, given a sufficient number of Mellin moments, PDFs can be reconstructed from the inverse Mellin transform. In practice, however, the calculation is limited to the lowest three moments because power-divergent mixing occurs between twist-two operators. Three moments are insufficient to fully reconstruct the momentum dependence of the PDFs without significant model dependence. The lowest three moments do, however, provide useful information both as benchmarks of lattice calculations and as constraints in global extractions of PDFs (324).

Direct extractions of the x dependence of PDFs have been attempted based on quasi-PDFs in large-momentum effective theory (325), pseudo-PDFs (326), and other methods. Quasi-PDFs are defined as Fourier transforms of the matrix elements, whereas pseudo-PDFs are transforms in Ioffe time. For positive (u–d) and negative ($\bar{u}$–$\bar{d}$) isovector quark combinations, the quasi-PDFs at the physical pion mass were found to agree with global fits at large $x > 0.1$ and $x > 0.4$, respectively (327). Gluon quantities are much noisier than quark disconnected loops and require calculations with very high statistics. Up to perturbative matching and power corrections, the Fourier transform of the gluon quasi-PDF was found to be compatible with that of global fits within the statistical uncertainty (328). Calculating the small-x behavior requires larger boost momenta, as this results in a faster decay of the matrix elements, so that truncations in the Fourier

transform matter less (329). Fourier transforms of strange and charm quark PDFs have also been obtained; the former are about five times larger, and both are smaller than those in global fits, possibly because of missing contributions from other flavor distributions. A full analysis of lattice QCD systematics must still be performed. Nevertheless, the strangeness asymmetry (s–$\bar{s}$) in the region of $0.3 < x < 0.8$ was found to be very small with high precision compared with the uncertainty in global fits (330). Including these lattice data in a global fit therefore greatly reduces the size of the s–$\bar{s}$ error band in the large-x region (331).

Lattice QCD studies of nuclear structure are currently restricted to low A and unphysical pion masses. In particular, the ratio of the longitudinal momentum fraction carried by the positive isovector quark combination in ^{3}He to the one in the free nucleon was found to be consistent with unity at the few-percent level. This is in agreement with, but more precise than, current determinations from global fits. Including this lattice result in the nNNPDF2.0 global fit framework reduces the uncertainty on the isovector momentum fraction ratio by a factor of 2.5 and leads to a more precise extraction of the u and d quark distributions in ^{3}He (332). Previously, the nuclear modification of the gluon momentum fraction for $A \leq 3$ was found to be less than ∼10%. This is consistent with expectations from phenomenological quark distributions and the momentum sum rule (333).

6.3. Relations to Other Phenomena

The global analysis of nuclear PDFs outlined in this review indicates that, at sufficiently high interaction scales, collinear factorization is a consistent way to describe lepton–nucleus and proton–nucleus collisions. However, toward low interaction scales there are theoretical reasons to believe that the role of higher-twist effects, suppressed by inverse powers of the interaction scale, becomes increasingly important compared with the role of such effects in simpler lepton–proton and proton–proton collisions. In the parton model, these effects arise from processes where two initial-state partons recombine (334–337). In general, these processes modify the linear DGLAP evolution by slowing down the evolution at small x and increasing it at intermediate x. Since the spatial density of partons is higher in large nuclei, the effect should be more pronounced there and lead to dynamically generated shadowing and antishadowing (338). At very low Q^2, the recombination of an even larger number of partons becomes eventually important and, when resummed, can be interpreted as saturation (3, 339). Finding a conclusive signature of saturation has proved rather difficult; calculations based on nuclear PDFs and the saturation picture give very similar results. Even the study of forward D meson production in p+Pb collisions, which probes the nucleus down to $x \sim 10^{-5}$ at nearly nonperturbative interaction scales, has not revealed clear deviations from the linear DGLAP dynamics. In addition, the resummation of small-x BFKL logarithms in the language of PDFs has a tendency to slow down the linear DGLAP evolution (340, 341), further complicating the search for true saturation effects.

There is a heated discussion regarding whether in very central, high-multiplicity p+Pb collisions, and even in smaller systems like pp and γA collisions, one creates a state of matter that has collective liquid-type properties—a droplet of QGP (342). Some characteristic features have been experimentally observed (343), which typically consist of correlations between particles. However, the measurements can also be interpreted in terms of the initial-state geometry (344). Furthermore, the way in which final-state QCD particles hadronize into a color-neutral state has been shown to display features that could accidentally be attributed to a liquid-like behavior (345). It should also be kept in mind that jet quenching has not been observed in p+Pb collisions, a fact that challenges the QGP picture of small systems and reflects a concrete difference between observations in p+Pb and Pb+Pb collisions. Multiplicity-integrated p+Pb cross sections appear to

be consistent with collinear factorization and process-independent nuclear PDFs. The relations between the observations discussed above remain open questions at this moment.

In the case of heavy ion collisions, the formation of a QGP is now a generally accepted phenomenon. Nevertheless, even in heavy ion collisions, the LHC data for electroweak boson (346–349) and high-p_T direct photon production (350–352) are consistent with collinear factorization and process-independent nuclear PDFs (353). Thus, there is no reason to believe that the initial state of heavy ion collisions would not be dictated by nuclear PDFs. This idea has been pursued in the Eskola–Kolhinen–Ruuskanen–Tuominen model of heavy ion collisions, the most recent versions of which (354) apply NLO perturbative QCD calculations and impact-parameter-dependent nuclear PDFs (355) to compute the initial conditions for the subsequent fluid-dynamical evolution of the system.

Nuclear PDFs also find use in the field of neutrino astronomy and cosmic-ray physics (356–358). Interactions of neutrinos coming from outer space can be measured using large neutrino telescopes such as IceCube, KM3NeT, and Baikal, where the neutrinos interact with water or ice. Precise theoretical calculations of the cross sections require nuclear PDFs as an input. In addition, protons from astrophysical sources can collide with the air molecules in the atmosphere, and such interactions can produce neutrinos. Precise calculations of the cross sections for these secondary neutrinos require nuclear PDFs as well.

7. CONCLUSION

During the last 25 years of research in nuclear PDFs, the field has undergone enormous development. Methodologically, simple fits, performed by eye at LO, have matured into rigorous statistical analyses, including machine learning techniques, at NLO and NNLO with full error estimates. Nonetheless, global nuclear PDF analyses are still driven by experimental measurements, and in this respect the p+Pb collisions carried out during the past decade at the LHC have opened up a wide, previously unexplored regime in terms of both kinematics and processes.

SUMMARY POINTS

1. Despite the theoretical and experimental advances, there are still significant differences among the independent global analyses of nuclear parton distribution functions (PDFs) in terms of both the extracted nuclear modifications of PDFs and the absolute nuclear PDFs. In several places, the central values of a given analysis can be outside the error bands of the others, and in some cases even the error bands of two given analyses do not overlap. The widths of the error bands also can be very different. The most significant factors behind the observed differences can be attributed to (*a*) the assumed form of the nonperturbative parameterization of nuclear PDFs at low Q^2, (*b*) the data selection, (*c*) the fitting of absolute cross sections versus ratios of cross sections, (*d*) the theoretical treatment of heavy flavors, and (*e*) the use of different baseline free-proton PDFs. These differences also lead to visible effects in observables. This underscores the need to carry out the global analysis independently in several groups. To faithfully chart the theoretical uncertainties in quantities that depend on nuclear PDFs, it is thus recommended to use more than one set of nuclear PDFs.
2. At the moment, next-to-leading-order (NLO) accuracy is the standard in the field of nuclear PDFs; the full next-to-next-to-leading-order (NNLO) accuracy is limited by the

existence or public availability of cross section codes. Most *pA* data can be well described within the NLO calculations, but in the case of some observables there is a confirmed (Drell–Yan process below the *Z* peak) or conjectured (prompt photons, jets) need for NNLO quantum chromodynamics. The NNLO accuracy also reduces the theoretical uncertainties in observables sensitive to the effects of partonic saturation, offers a standard candle to Glauber modeling of heavy ion collisions through precise predictions for electroweak observables, and should lead to more precise predictions for astrophysical applications.

3. In the long run, the global analysis of nuclear PDFs should be extended to include the proton and the deuteron. To date, most of the free-proton fits still use heavy-target—particularly neutrino deep inelastic scattering—data to constrain the full flavor decomposition. However, these same data can be (and are) also taken as constraints on the nuclear modifications of PDFs. Fully charting the interplay between the two requires simultaneous extraction of the free-proton and nuclear PDFs.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The authors thank their nCTEQ and EPPS colleagues for their collaboration and useful discussions, E. Nocera for providing the nNNPDF3.0 values for **Figure 7**, and V. Guzey for providing the nCTEQ15HQ values for **Figure 8*b***. M.K. thanks his ALICE colleagues for their collaboration and acknowledges funding from the German Federal Ministry of Education and Research (BMBF; project 05P21PMCAA) and the German Research Foundation (DFG; GRK 2149 and SFB 1225 "Isoquant," project 273811115). H.P. acknowledges funding from the Academy of Finland through the Center of Excellence in Quark Matter (project 346326). The results shown in **Figure 6** and in **Figure 8*a*** have been computed for this review using computing resources of the Finnish IT Center for Science (CSC; project jyy2580).

LITERATURE CITED

1. Frankfurt LL, Strikman MI. *Phys. Rep.* 160:235 (1988)
2. Arneodo M. *Phys. Rep.* 240:301 (1994)
3. Gelis F, Iancu E, Jalilian-Marian J, Venugopalan R. *Annu. Rev. Nucl. Part. Sci.* 60:463 (2010)
4. Klasen M, Klein-Bösing C, König F, Wessels JP. *J. High Energy Phys.* 1310:119 (2013)
5. Andronic A, Braun-Munzinger P, Redlich K, Stachel J. *Nature* 561:321 (2018)
6. Gribov VN, Lipatov LN. *Sov. J. Nucl. Phys.* 15:438 (1972)
7. Gribov VN, Lipatov LN. *Sov. J. Nucl. Phys.* 15:675 (1972)
8. Dokshitzer YL. *Sov. Phys. JETP* 46:641 (1977)
9. Altarelli G, Parisi G. *Nucl. Phys. B* 126:298 (1977)
10. Arnold RG, et al. *Phys. Rev. Lett.* 52:727 (1984)
11. Gomez J, et al. *Phys. Rev. D* 49:4348 (1994)
12. Amaudruz P, et al. (New Muon Collab.) *Nucl. Phys. B* 441:3 (1995)
13. Tzanov M, et al. (NuTeV Collab.) *Phys. Rev. D* 74:012008 (2006)
14. Arneodo M, et al. (New Muon Collab.) *Nucl. Phys. B* 481:23 (1996)
15. Caldwell DO, et al. *Phys. Rev. Lett.* 42:553 (1979)

16. Goodman MS, et al. *Phys. Rev. Lett.* 47:293 (1981)
17. Bauer TH, Spital RD, Yennie DR, Pipkin FM. *Rev. Mod. Phys.* 50:261 (1978)
18. Armesto N. *J. Phys. G* 32:R367 (2006)
19. Frankfurt L, Guzey V, Strikman M. *Phys. Rep.* 512:255 (2012)
20. Kopeliovich BZ, Morfin JG, Schmidt I. *Prog. Part. Nucl. Phys.* 68:314 (2013)
21. Atwood WB, West GB. *Phys. Rev. D* 7:773 (1973)
22. Bodek A, Ritchie JL. *Phys. Rev. D* 23:1070 (1981)
23. Frankfurt LL, Strikman MI. *Nucl. Phys. B* 181:22 (1981)
24. Saito K, Uchiyama T. *Z. Phys. A* 322:299 (1985)
25. Ciofi degli Atti C, Simula Y. *Phys. Rev. C* 53:1689 (1996)
26. Segarra EP, et al. *Phys. Rev. D* 103:114015 (2021)
27. Aubert JJ, et al. (Eur. Muon Collab.) *Phys. Lett. B* 123:275 (1983)
28. Bodek A, et al. *Phys. Rev. Lett.* 50:1431 (1983)
29. Bodek A, et al. *Phys. Rev. Lett.* 51:534 (1983)
30. Geesaman DF, Saito K, Thomas AW. *Annu. Rev. Nucl. Part. Sci.* 45:337 (1995)
31. Thomas AW. *Int. J. Mod. Phys. E* 27:1840001 (2019)
32. Alde DM, et al. (E772 Collab.) *Phys. Rev. Lett.* 64:2479 (1990)
33. Malace S, Gaskell D, Higinbotham DW, Cloet I. *Int. J. Mod. Phys. E* 23:1430013 (2014)
34. Hen O, Miller GA, Piasetzky E, Weinstein LB. *Rev. Mod. Phys.* 89:045002 (2017)
35. Nikolaev NN, Zakharov VI. *Phys. Lett. B* 55:397 (1975)
36. Eskola KJ. *Nucl. Phys. B* 400:240 (1993)
37. Qiu JW. *Nucl. Phys. B* 291:746 (1987)
38. Eskola KJ, Kolhinen VJ, Salgado CA. *Eur. Phys. J. C* 9:61 (1999)
39. Hirai M, Kumano S, Miyama M. *Phys. Rev. D* 64:034003 (2001)
40. de Florian D, Sassot R. *Phys. Rev. D* 69:074028 (2004)
41. Hirai M, Kumano S, Nagai TH. *Phys. Rev. C* 76:065207 (2007)
42. Eskola KJ, Paukkunen H, Salgado CA. *J. High Energy Phys.* 0904:065 (2009)
43. Kovarik K, et al. *Phys. Rev. D* 93:085037 (2016)
44. Schienbein I, et al. *Phys. Rev. D* 77:054013 (2008)
45. Paukkunen H, Salgado CA. *J. High Energy Phys.* 1007:32 (2010)
46. Kovarik K, et al. *Phys. Rev. Lett.* 106:122301 (2011)
47. Paukkunen H, Salgado CA. *Phys. Rev. Lett.* 110:212301 (2013)
48. Nakamura SX, et al. *Rep. Prog. Phys.* 80:056301 (2017)
49. Kalantarians N, Keppel C, Christy ME. *Phys. Rev. C* 96:032201 (2017)
50. Duwentäster P, et al. *Phys. Rev. D* 105:114043 (2022)
51. Eskola KJ, Paakkinen P, Paukkunen H, Salgado CA. *Eur. Phys. J. C* 82:413 (2022)
52. Abdul Khalek R, et al. *Eur. Phys. J. C* 82:507 (2022)
53. Gao J, Harland-Lang L, Rojo J. *Phys. Rep.* 742:1 (2018)
54. Kovařík K, Nadolsky PM, Soper DE. *Rev. Mod. Phys.* 92:045003 (2020)
55. Ethier JJ, Nocera ER. *Annu. Rev. Nucl. Part. Sci.* 70:43 (2020)
56. Klein SR, Mäntysaari H. *Nat. Rev. Phys.* 1:662 (2019)
57. Blümlein J. *Prog. Part. Nucl. Phys.* 69:28 (2013)
58. Blümlein J. arXiv:2306.01362 [hep-ph] (2023)
59. Georgi H, Politzer HD. *Phys. Rev. D* 14:1829 (1976)
60. Collins JC, Soper DE, Sterman GF. *Adv. Ser. Dir. High Energy Phys.* 5:1 (1989)
61. Nadolsky PM, Tung WK. *Phys. Rev. D* 79:113014 (2009)
62. Thorne RS, Tung WK. arXiv:0809.0714 [hep-ph] (2008)
63. Aivazis MAG, Collins JC, Olness FI, Tung WK. *Phys. Rev. D* 50:3102 (1994)
64. Krämer M, Olness FI, Soper DE. *Phys. Rev. D* 62:096007 (2000)
65. Tung WK, Kretzer S, Schmidt C. *J. Phys. G* 28:983 (2002)
66. Cacciari M, Greco M, Nason P. *J. High Energy Phys.* 9805:007 (1998)
67. Forte S, Laenen E, Nason P, Rojo J. *Nucl. Phys. B* 834:116 (2010)

68. Accardi A, Qiu JW. *J. High Energy Phys.* 0807:090 (2008)
69. Ruiz R, et al. arXiv:2301.07715 [hep-ph] (2023)
70. Nachtmann O. *Nucl. Phys. B* 63:237 (1973)
71. Barnett RM. *Phys. Rev. Lett.* 36:1163 (1976)
72. Accardi A, et al. arXiv:hep-ph/0308248 (2003)
73. d'Agostini G. *Nucl. Instrum. Methods A* 346:306 (1994)
74. Ball RD, et al. (NNPDF Collab.) *J. High Energy Phys.* 1005:75 (2010)
75. Pumplin J, Stump DR, Tung WK. *Phys. Rev. D* 65:014011 (2001)
76. Pumplin J, et al. *Phys. Rev. D* 65:014013 (2001)
77. Giele WT, Keller S. *Phys. Rev. D* 58:094023 (1998)
78. Giele WT, Keller SA, Kosower DA. arXiv:hep-ph/0104052 (2001)
79. Nadolsky PM, Sullivan Z. *eConf* C010630:P510 (2001)
80. Helenius I, Walt M, Vogelsang W. *Phys. Rev. D* 105:9 (2022)
81. Khanpour H, et al. *Phys. Rev. D* 104:034010 (2021)
82. Forte S, Garrido L, Latorre JI, Piccione A. *J. High Energy Phys.* 0205:062 (2002)
83. Ball RD, et al. (NNPDF Collab.) *Nucl. Phys. B* 809:1 (2009). Erratum. *Nucl. Phys. B* 816:293 (2009)
84. Pumplin J, et al. *J. High Energy Phys.* 0207:012 (2002)
85. Stump D, et al. *J. High Energy Phys.* 0310:046 (2003)
86. Owens JF, et al. *Phys. Rev. D* 75:054030 (2007)
87. Kusina A, et al. *Eur. Phys. J. C* 80:968 (2020)
88. Duwentäster P, et al. *Phys. Rev. D* 104:094005 (2021)
89. Eskola KJ, Helenius I, Paakkinen P, Paukkunen H. *J. High Energy Phys.* 2005:37 (2020)
90. Kusina A, Lansberg JP, Schienbein I, Shao HS. *Phys. Rev. D* 104:014010 (2021)
91. Kom CH, Kulesza A, Stirling WJ. *Phys. Rev. Lett.* 107:082002 (2011)
92. Kniehl BA, Kramer G, Schienbein I, Spiesberger H. *Phys. Rev. D* 71:014018 (2005)
93. Butenschoen M, Kniehl BA. *Phys. Rev. Lett.* 106:022003 (2011)
94. Eskola KJ, Kolhinen VJ, Ruuskanen PV. *Nucl. Phys. B* 535:351 (1998)
95. Eskola KJ, Kolhinen VJ, Paukkunen H, Salgado CA. *J. High Energy Phys.* 0705:002 (2007)
96. Eskola KJ, Paukkunen H, Salgado CA. *J. High Energy Phys.* 0807:102 (2008)
97. Eskola KJ, Paakkinen P, Paukkunen H, Salgado CA. *Eur. Phys. J. C* 77:163 (2017)
98. Hou TJ, et al. *Phys. Rev. D* 103:014013 (2021)
99. Accardi A, Hobbs TJ, Jing X, Nadolsky PM. *Eur. Phys. J. C* 81:603 (2021)
100. Ball RD, et al. (NNPDF Collab.) *Eur. Phys. J. C* 77:663 (2017)
101. Ball RD, et al. (NNPDF Collab.) *Eur. Phys. J. C* 82:428 (2022)
102. Rosenblatt M. *Ann. Math. Stat.* 27:832 (1956)
103. Parzen E. *Ann. Math. Stat.* 33:1065 (1962)
104. Abdul Khalek R, et al. (NNPDF Collab.) *Eur. Phys. J. C* 79:471 (2019)
105. Walt M, Helenius I, Vogelsang W. *Phys. Rev. D* 100:096015 (2019)
106. Bertone V, et al. (xFitter Dev. Team) *Proc. Sci.* DIS2017:203 (2018)
107. Abramowicz H, et al. (H1/ZEUS Collab.) *Eur. Phys. J. C* 75:580 (2015)
108. Sick I, Day D. *Phys. Lett. B* 274:16 (1992)
109. Denniston AW, et al. arXiv:2312.16293 [hep-ph] (2023)
110. Kulagin SA, Petti R. *Nucl. Phys. A* 765:126 (2006)
111. Kulagin SA, Petti R. *Phys. Rev. D* 76:094023 (2007)
112. Kulagin SA, Petti R. *Phys. Rev. C* 90:045204 (2014)
113. Ru P, Kulagin SA, Petti R, Zhang BW. *Phys. Rev. D* 94:113013 (2016)
114. Ashman J, et al. (Eur. Muon Collab.) *Phys. Lett. B* 202:603 (1988)
115. de Florian D, Sassot R, Zurita P, Stratmann M. *Phys. Rev. D* 85:074028 (2012)
116. Arneodo M, et al. (New Muon Collab.) *Nucl. Phys. B* 481:3 (1996)
117. Adams MR, et al. (E665 Collab.) *Z. Phys. C* 67:403 (1995)
118. Prytz K. *Phys. Lett. B* 311:286 (1993)
119. Eskola KJ, Honkanen H, Kolhinen VJ, Salgado CA. *Phys. Lett. B* 532:222 (2002)

120. Airapetian A, et al. (HERMES Collab.) arXiv:hep-ex/0210068 (2002)
121. Armesto N, Paukkunen H, Salgado CA, Tywoniuk K. *Phys. Lett. B* 694:38 (2011)
122. Ackerstaff K, et al. (HERMES Collab.) *Phys. Lett. B* 475:386 (2000). Erratum. *Phys. Lett. B* 567:339 (2003)
123. Ball RD, et al. (NNPDF Collab.) *Nature* 608:483 (2022)
124. Aubert JJ, et al. (Eur. Muon Collab.) *Phys. Lett. B* 94:96 (1980)
125. Aubert JJ, et al. (Eur. Muon Collab.) *Nucl. Phys. B* 213:31 (1983)
126. Ashman J, et al. (Eur. Muon Collab.) *Z. Phys. C* 57:211 (1993)
127. Seely J, et al. (Hall C Collab.) *Phys. Rev. Lett.* 103:202301 (2009)
128. Arrington J, et al. (Hall C Collab.) *Phys. Rev. C* 104:065203 (2021)
129. Karki A, et al. (Hall C Collab.) arXiv:2207.03850 [nucl-ex] (2022)
130. Schmookler B, et al. (CLAS Collab.) *Nature* 566:354 (2019)
131. Paukkunen H, Zurita P. *Eur. Phys. J. C* 80:381 (2020)
132. Accardi A, et al. *Phys. Rev. D* 93:114017 (2016)
133. Arrington J, et al. *Phys. Rev. C* 73:035205 (2006)
134. Martin AD, et al. *Eur. Phys. J. C* 73:2318 (2013)
135. Moreno G, et al. (E605 Collab.) *Phys. Rev. D* 43:2815 (1991)
136. Vasilev MA, et al. (E605 Collab.) *Phys. Rev. Lett.* 83:2304 (1999)
137. Ellis SD, Stirling WJ. *Phys. Lett. B* 256:258 (1991)
138. Martin AD, Roberts RG, Stirling WJ, Thorne RS. *Eur. Phys. J. C* 4:463 (1998)
139. Lin PJ. Report FERMILAB-THESIS-2017-18, Fermi Natl. Accel. Lab., Batavia, IL (2017)
140. Helenius I, et al. *Phys. Rev. D* 100:014004 (2019)
141. Aaij R, et al. (LHCb Collab.) Report LHCb-CONF-2018-005, CERN, Geneva (2018)
142. Paakkinen P, Eskola KJ, Paukkunen H. *Phys. Lett. B* 768:7 (2017)
143. Badier J, et al. (NA3 Collab.) *Phys. Lett. B* 104:335 (1981)
144. Bordalo P, et al. (NA10 Collab.) *Phys. Lett. B* 193:368 (1987)
145. Heinrich JG, et al. (E615 Collab.) *Phys. Rev. Lett.* 63:356 (1989)
146. Adams B, et al. (COMPASS++/AMBER Collab.) arXiv:1808.00848 [hep-ex] (2018)
147. Ball RD, et al. (NNPDF Collab.) *Eur. Phys. J. C* 79:282 (2019)
148. Mason D, et al. (NuTeV Collab.) *Phys. Rev. Lett.* 99:192001 (2007)
149. Zeller GP, et al. (NuTeV Collab.) *Phys. Rev. Lett.* 88:091802 (2002). Erratum. *Phys. Rev. Lett.* 90:239902 (2003)
150. Hirai M, Kumano S, Nagai TH. *Phys. Rev. D* 71:113007 (2005)
151. Eskola KJ, Paukkunen H. *J. High Energy Phys.* 0606:008 (2006)
152. Workman RL, et al. (Part. Data Group) *PTEP* 2022:083C01 (2022)
153. Berge JP, et al. *Z. Phys. C* 49:187 (1991)
154. Yang UK, et al. (CCFR Collab.) *Phys. Rev. Lett.* 86:2742 (2001)
155. Onengut G, et al. (CHORUS Collab.) *Phys. Lett. B* 632:65 (2006)
156. Goncharov M, et al. (CCFR/NuTeV Collab.) *Phys. Rev. D* 64:112006 (2001)
157. Samoylov O, et al. (NOMAD Collab.) *Nucl. Phys. B* 876:339 (2013)
158. Muzakka KF, et al. *Phys. Rev. D* 106:074004 (2022)
159. Abreu H, et al. (FASER Collab.) *Phys. Rev. Lett.* 131:3 (2023)
160. Albanese R, et al. (SND@LHC Collab.) *Phys. Rev. Lett.* 131:031802 (2023)
161. Cruz-Martinez JM, et al. arXiv:2309.09581 [hep-ph] (2023)
162. Carli T, et al. *Eur. Phys. J. C* 66:503 (2010)
163. Aad G, et al. (ATLAS Collab.) *Phys. Rev. C* 92:044915 (2015)
164. Khachatryan V, et al. (CMS Collab.) *Phys. Lett. B* 759:36 (2016)
165. Khachatryan V, et al. (CMS Collab.) *Phys. Lett. B* 750:565 (2015)
166. Adam J, et al. (ALICE Collab.) *J. High Energy Phys.* 1702:77 (2017)
167. Aaij R, et al. (LHCb Collab.) *J. High Energy Phys.* 1409:30 (2014)
168. Aad G, et al. (ATLAS Collab.) Report ATLAS-CONF-2015-056, CERN, Geneva (2015)
169. Aaij R, et al. (LHCb Collab.) *J. High Energy Phys.* 2306:22 (2023)
170. Sirunyan AM, et al. (CMS Collab.) *Phys. Lett. B* 800:135048 (2020)

171. Khachatryan V, et al. (CMS Collab.) *Eur. Phys. J. C* 76:469 (2016)
172. Sirunyan AM, et al. (CMS Collab.) *J. High Energy Phys.* 2105:182 (2021)
173. Acharya S, et al. (ALICE Collab.) *Phys. Rev. C* 102:055204 (2020)
174. Berger EL, Gordon LE, Klasen M. *Phys. Rev. D* 58:074012 (1998)
175. Klasen M, Brandt M. *Phys. Rev. D* 88:054002 (2013)
176. Brandt M, Klasen M, König F. *Nucl. Phys. A* 927:78 (2014)
177. Arleo F, Gousset T. *Phys. Lett. B* 660:181 (2008)
178. Brenner Mariotto C, Goncalves VP. *Phys. Rev. C* 78:037901 (2008)
179. Arleo F, Eskola KJ, Paukkunen H, Salgado CA. *J. High Energy Phys.* 1104:55 (2011)
180. Helenius I, Eskola KJ, Paukkunen H. *J. High Energy Phys.* 1409:138 (2014)
181. Klasen M. *Rev. Mod. Phys.* 74:1221 (2002)
182. Glück M, Reya E, Vogt A. *Phys. Rev. D* 48:116 (1993). Erratum. *Phys. Rev. D* 51:1427 (1995)
183. Bourhis L, Fontannaz M, Guillet JP. *Eur. Phys. J. C* 2:529 (1998)
184. Klasen M, König F. *Eur. Phys. J. C* 74:3009 (2014)
185. Frixione S. *Phys. Lett. B* 429:369 (1998)
186. Apanasevich L, et al. (Fermilab E706 Collab.) *Phys. Rev. D* 70:092009 (2004)
187. Adare A, et al. (PHENIX Collab.) *Phys. Rev. C* 87:054907 (2013)
188. Abelev BI, et al. (STAR Collab.) *Phys. Rev. C* 81:064904 (2010)
189. Aaboud M, et al. (ATLAS Collab.) *Phys. Lett. B* 796:230 (2019)
190. Aad G, et al. (ATLAS Collab.) *J. High Energy Phys.* 1608:5 (2016)
191. Campbell JM, Ellis RK, Williams C. *Phys. Rev. Lett.* 118(22):222001 (2017). Erratum. *Phys. Rev. Lett.* 124:259901 (2020)
192. Acharya S, et al. (ALICE Collab.) Report CERN-LHCC-2020-009, CERN, Geneva (2020)
193. Acharya S, et al. (ALICE Collab.) Report ALICE-PUBLIC-2023-001, CERN, Geneva (2023)
194. Vogt R. *Phys. Rev. C* 70:064902 (2004)
195. Arsene I, et al. (BRAHMS Collab.) *Phys. Rev. Lett.* 91:072305 (2003)
196. Arsene I, et al. (BRAHMS Collab.) *Phys. Rev. Lett.* 93:242303 (2004)
197. Adler SS, et al. (PHENIX Collab.) *Phys. Rev. Lett.* 91:072303 (2003)
198. Adams J, et al. (STAR Collab.) *Phys. Rev. Lett.* 91:072304 (2003)
199. Adler SS, et al. (PHENIX Collab.) *Phys. Rev. Lett.* 98:172302 (2007)
200. Adams J, et al. (STAR Collab.) *Phys. Lett. B* 637:161 (2006)
201. Kniehl BA, Kramer G, Pötter B. *Nucl. Phys. B* 582:514 (2000)
202. Aidala CA, et al. *Phys. Rev. D* 83:034002 (2011)
203. de Florian D, et al. *Phys. Rev. D* 91:014035 (2015)
204. de Florian D, et al. *Phys. Rev. D* 95:094019 (2017)
205. Moffat E, et al. (JLab Angular Momentum Collab.) *Phys. Rev. D* 104:016015 (2021)
206. Adare A, et al. (PHENIX Collab.) *Phys. Rev. C* 88:024906 (2013)
207. Acharya S, et al. (ALICE Collab.) *Eur. Phys. J. C* 78:624 (2018)
208. Adam J, et al. (ALICE Collab.) *Phys. Lett. B* 760:720 (2016)
209. Acharya S, et al. (ALICE Collab.) *Phys. Lett. B* 827:136943 (2022)
210. Sassot R, Stratmann M, Zurita P. *Phys. Rev. D* 81:054001 (2010)
211. Zurita P. arXiv:2101.01088 [hep-ph] (2021)
212. Aaij R, et al. (LHCb Collab.) *Phys. Rev. Lett.* 131:042302 (2023)
213. Aaij R, et al. (LHCb Collab.) arXiv:2310.17326 [nucl-ex] (2023)
214. Aaij R, et al. (LHCb Collab.) *Phys. Rev. Lett.* 128:142004 (2022)
215. Aad G, et al. (ATLAS Collab.) *J. High Energy Phys.* 2307:74 (2023)
216. Khachatryan V, et al. (CMS Collab.) *Eur. Phys. J. C* 75:237 (2015)
217. Khachatryan V, et al. (CMS Collab.) *J. High Energy Phys.* 1704:39 (2017)
218. Acharya S, et al. (ALICE Collab.) *J. High Energy Phys.* 1811:13 (2018)
219. d'Enterria D, Eskola KJ, Helenius I, Paukkunen H. *Nucl. Phys. B* 883:615 (2014)
220. Loizides C, Kamin J, d'Enterria D. *Phys. Rev. C* 97:054910 (2018). Erratum. *Phys. Rev. C* 99:019901 (2019)

221. Sirunyan AM, et al. (CMS Collab.) *Phys. Rev. Lett.* 121:062002 (2018)
222. Chatrchyan S, et al. (CMS Collab.) *Eur. Phys. J. C* 74:2951 (2014)
223. Eskola KJ, Paakkinen P, Paukkunen H. *Eur. Phys. J. C* 79:511 (2019)
224. Currie J, et al. *Phys. Rev. Lett.* 119:152001 (2017)
225. Aad G, et al. (ATLAS Collab.) *Phys. Lett. B* 748:392 (2015)
226. Khachatryan V, et al. (CMS Collab.) *Eur. Phys. J. C* 76:372 (2016)
227. Acharya S, et al. (ALICE Collab.) arXiv:2307.10860 [nucl-ex] (2023)
228. Cacciari M, Mangano ML, Nason P. *Eur. Phys. J. C* 75:610 (2015)
229. Gauld R, Rojo J, Rottoli L, Talbert J. *J. High Energy Phys.* 1511:9 (2015)
230. Gauld R, Rojo J. *Phys. Rev. Lett.* 118:072001 (2017)
231. Zenaiev O, et al. (PROSA Collab.) *Eur. Phys. J. C* 75:396 (2015)
232. Gauld R. *Phys. Rev. D* 93:014001 (2016)
233. Kusina A, Lansberg JP, Schienbein I, Shao HS. *Phys. Rev. Lett.* 121:052004 (2018)
234. Helenius I, Paukkunen H. *J. High Energy Phys.* 1805:196 (2018)
235. Helenius I, Paukkunen H. *J. High Energy Phys.* 2307:54 (2023)
236. Nason P. *J. High Energy Phys.* 0411:040 (2004)
237. Frixione S, Nason P, Oleari C. *J. High Energy Phys.* 0711:070 (2007)
238. Alioli S, Nason P, Oleari C, Re E. *J. High Energy Phys.* 1006:43 (2010)
239. Lansberg JP, Shao HS. *Eur. Phys. J. C* 77:1 (2017)
240. Andronic A, et al. *Eur. Phys. J. C* 76:107 (2016)
241. Abdul Khalek R, Ethier JJ, Rojo J, van Weelden G. *J. High Energy Phys.* 2009:183 (2020)
242. Sirunyan AM, et al. (CMS Collab.) *Phys. Lett. B* 772:306 (2017)
243. Khachatryan V, et al. (CMS Collab.) *Phys. Lett. B* 754:59 (2016)
244. Acharya S, et al. (ALICE Collab.) *J. High Energy Phys.* 2201:178 (2022)
245. Aaij R, et al. (LHCb Collab.) *Phys. Rev. D* 99:052011 (2019)
246. Acharya S, et al. (ALICE Collab.) *J. High Energy Phys.* 2002:77 (2020)
247. Acharya S, et al. (ALICE Collab.) *J. High Energy Phys.* 2308:6 (2023)
248. Aaij R, et al. (LHCb Collab.) *Phys. Rev. Lett.* 122:132002 (2019)
249. Aaij R, et al. (LHCb Collab.) *Eur. Phys. J. C* 83:541 (2023)
250. Aaij R, et al. (LHCb Collab.) *Eur. Phys. J. C* 83:625 (2023)
251. Aad G, et al. (ATLAS Collab.) *Phys. Rev. C* 92:034904 (2015)
252. Aaboud M, et al. (ATLAS Collab.) *Eur. Phys. J. C* 78:171 (2018)
253. Sirunyan AM, et al. (CMS Collab.) *Eur. Phys. J. C* 77:269 (2017)
254. Tumasyan A, et al. (CMS Collab.) *Phys. Lett. B* 835:137397 (2022)
255. Sirunyan AM, et al. (CMS Collab.) *Phys. Lett. B* 790:509 (2019)
256. Abelev BB, et al. (ALICE Collab.) *Phys. Rev. Lett.* 113:232301 (2014)
257. Adam J, et al. (ALICE Collab.) *Phys. Rev. C* 94:054908 (2016)
258. Abelev BB, et al. (ALICE Collab.) *J. High Energy Phys.* 1402:73 (2014)
259. Adam J, et al. (ALICE Collab.) *J. High Energy Phys.* 1506:55 (2015)
260. Acharya S, et al. (ALICE Collab.) *J. High Energy Phys.* 2206:11 (2022)
261. Abelev BB, et al. (ALICE Collab.) *Phys. Lett. B* 740:105 (2015)
262. Abelev BB, et al. (ALICE Collab.) *J. High Energy Phys.* 1412:73 (2014)
263. Aaij R, et al. (LHCb Collab.) *J. High Energy Phys.* 1402:72 (2014)
264. Aaij R, et al. (LHCb Collab.) *J. High Energy Phys.* 1407:94 (2014)
265. Acharya S, et al. (ALICE Collab.) *J. High Energy Phys.* 1807:160 (2018)
266. Acharya S, et al. (ALICE Collab.) *J. High Energy Phys.* 2307:137 (2023)
267. Acharya S, et al. (ALICE Collab.) *Phys. Lett. B* 806:135486 (2020)
268. Acharya S, et al. (ALICE Collab.) *J. High Energy Phys.* 2007:237 (2020)
269. Aaij R, et al. (LHCb Collab.) *Phys. Lett. B* 774:159 (2017)
270. Aaij R, et al. (LHCb Collab.) *J. High Energy Phys.* 1811:194 (2018). Erratum. *J. High Energy Phys.* 2002:93 (2020)
271. Aaij R, et al. (LHCb Collab.) *J. High Energy Phys.* 1710:90 (2017)

272. Acharya S, et al. (ALICE Collab.) *J. High Energy Phys.* 1912:92 (2019)
273. Bezshyiko I, et al. (LHCb Collab.) *Phys. Rev. Lett.* 131:102301 (2023)
274. Aaij R, et al. (LHCb Collab.) *J. High Energy Phys.* 2401:70 (2024)
275. d'Enterria D, Krajczár K, Paukkunen H. *Phys. Lett. B* 746:64 (2015)
276. Sirunyan AM, et al. (CMS Collab.) *Phys. Rev. Lett.* 119:242001 (2017)
277. Aad G, et al. (ATLAS Collab.) Report ATLAS-CONF-2023-063, CERN, Geneva (2023)
278. Sirunyan AM, et al. (CMS Collab.) *Phys. Rev. Lett.* 125:222001 (2020)
279. Bertulani CA, Klein SR, Nystrand J. *Annu. Rev. Nucl. Part. Sci.* 55:271 (2005)
280. Baltz AJ, et al. *Phys. Rep.* 458:1 (2008)
281. Adeluyi A, Bertulani C. *Phys. Rev. C* 84:024916 (2011)
282. Adeluyi A, Bertulani CA. *Phys. Rev. C* 85:044904 (2012)
283. Guzey V, Zhalov M. *J. High Energy Phys.* 1310:207 (2013)
284. Guzey V, Kryshen E, Strikman M, Zhalov M. *Phys. Lett. B* 726:290 (2013)
285. Eskola KJ, et al. *Phys. Rev. C* 106:035202 (2022)
286. Eskola KJ, et al. *Phys. Rev. C* 107:044912 (2023)
287. Ivanov DY, Schafer A, Szymanowski L, Krasnikov G. *Eur. Phys. J. C* 34:297 (2004). Erratum. *Eur. Phys. J. C* 75:75 (2015)
288. Acharya S, et al. (ALICE Collab.) *Phys. Lett. B* 798:134926 (2019)
289. Acharya S, et al. (ALICE Collab.) *Eur. Phys. J. C* 81:712 (2021)
290. Aaij R, et al. (LHCb Collab.) *J. High Energy Phys.* 2306:146 (2023)
291. Tumasyan A, et al. (CMS Collab.) *Phys. Rev. Lett.* 131:262301 (2023)
292. Jones SP, Martin AD, Ryskin MG, Teubner T. *J. Phys. G* 43:035002 (2016)
293. Jones SP, Martin AD, Ryskin MG, Teubner T. *Eur. Phys. J. C* 76:633 (2016)
294. Dutrieux H, Winn M, Bertone V. *Phys. Rev. D* 107:114019 (2023)
295. Helenius I. *Proc. Sci.* HardProbes2018:118 (2018)
296. Guzey V, Klasen M. *Phys. Rev. C* 99:065202 (2019)
297. Aad G, et al. (ATLAS Collab.) Report ATLAS-CONF-2017-011, CERN, Geneva (2017)
298. Aad G, et al. (ATLAS Collab.) Report ATLAS-CONF-2022-021, CERN, Geneva (2022)
299. Guzey V, Klasen M. *Phys. Rev. D* 104:114013 (2021)
300. Accardi A, et al. *Eur. Phys. J. A* 52:268 (2016)
301. Abdul Khalek R, et al. *Nucl. Phys. A* 1026:122447 (2022)
302. Altarelli G, Martinelli G. *Phys. Lett. B* 76:89 (1978)
303. Bandeira YB, Goncalves VP. *Eur. Phys. J. A* 59:19 (2023)
304. Aschenauer EC, et al. *Phys. Rev. D* 96:114005 (2017)
305. Khalek RA, Ethier JJ, Nocera ER, Rojo J. *Phys. Rev. D* 103:096005 (2021)
306. Kelsey M, et al. *Phys. Rev. D* 104:054002 (2021)
307. Klasen M, Kovarik K, Potthoff J. *Phys. Rev. D* 95:094013 (2017)
308. Arratia M, Song Y, Ringer F, Jacak BV. *Phys. Rev. C* 101:065204 (2020)
309. Klasen M, Kovařík K. *Phys. Rev. D* 97:114013 (2018)
310. Aschenauer EC, Lee K, BS Page, Ringer F. *Phys. Rev. D* 101:054028 (2020)
311. Guzey V, Klasen M. *Phys. Rev. C* 102:065201 (2020)
312. Arratia M, et al. *Phys. Rev. D* 103:074023 (2021)
313. Chen G, Li Y, Tuchin K, Vary JP. *Phys. Rev. C* 100:025208 (2019)
314. Goncalves VP, Martins DE, Sena CR. *Nucl. Phys. A* 1004:122055 (2020)
315. Guzey V, Klasen M. *J. High Energy Phys.* 2005:74 (2020)
316. Deák M, Stašto AM, Strikman M. *Phys. Rev. D* 103:014022 (2021)
317. Marquet C, Moldes MR, Zurita P. *Phys. Lett. B* 772:607 (2017)
318. Mäntysaari H, Mueller N, Salazar F, Schenke B. *Phys. Rev. Lett.* 124:112301 (2020)
319. Tong XB, BW Xiao, Zhang YY. *Phys. Rev. Lett.* 130:151902 (2023)
320. Dumitru A, Skokov V, Ullrich T. *Phys. Rev. C* 99:015204 (2019)
321. Caucal P, et al. *J. High Energy Phys.* 2308:62 (2023)
322. Caucal P, Salazar F, Venugopalan R. *J. High Energy Phys.* 2111:222 (2021)

323. Liu HY, et al. *Phys. Rev. Lett.* 130:18 (2023)
324. Lin HW, et al. *Prog. Part. Nucl. Phys.* 100:107 (2018)
325. Ji X, et al. *Rev. Mod. Phys.* 93:035005 (2021)
326. Radyushkin A. *Phys. Lett. B* 767:314 (2017)
327. Lin HW, et al. (LP3 Collab.) *Phys. Rev. D* 98:054504 (2018)
328. Fan ZY, et al. *Phys. Rev. Lett.* 121:242001 (2018)
329. Constantinou M, et al. *Prog. Part. Nucl. Phys.* 121:103908 (2021)
330. Zhang R, Lin HW, Yoon B. *Phys. Rev. D* 104:094511 (2021)
331. Hou TJ, Lin HW, Yan M, Yuan CP. *Phys. Rev. D* 107:076018 (2023)
332. Detmold W, et al. (NPLQCD Collab.) *Phys. Rev. Lett.* 126:202001 (2021)
333. Winter F, et al. *Phys. Rev. D* 96:094512 (2017)
334. Gribov LV, Levin EM, Ryskin MG. *Phys. Rep.* 100:1 (1983)
335. Mueller AH, Qiu JW. *Nucl. Phys. B* 268:427 (1986)
336. Zhu W. *Nucl. Phys. B* 551:245 (1999)
337. Zhu W, Ruan JH. *Nucl. Phys. B* 559:378 (1999)
338. Rausch J, Guzey V, Klasen M. *Phys. Rev. D* 107:054003 (2023)
339. Morreale A, Salazar F. *Universe* 7:312 (2021)
340. Bonvini M, Marzani S, Peraro T. *Eur. Phys. J. C* 76:597 (2016)
341. Ball RD, et al. *Eur. Phys. J. C* 78:321 (2018)
342. Pasechnik R, Šumbera M. *Universe* 3:7 (2017)
343. Nagle JL, Zajc WA. *Annu. Rev. Nucl. Part. Sci.* 68:211 (2018)
344. Schlichting S, Tribedy P. *Adv. High Energy Phys.* 2016:8460349 (2016)
345. Ortiz Velasquez A, et al. *Phys. Rev. Lett.* 111:042001 (2013)
346. Acharya S, et al. (ALICE Collab.) *J. High Energy Phys.* 2009:76 (2020)
347. Acharya S, et al. (ALICE Collab.) *J. High Energy Phys.* 2305:36 (2023)
348. Aad G, et al. (ATLAS Collab.) *Eur. Phys. J. C* 79:935 (2019)
349. Aad G, et al. (ATLAS Collab.) *Phys. Lett. B* 802:135262 (2020)
350. Aad G, et al. (ATLAS Collab.) *Phys. Rev. C* 93:034914 (2016)
351. Chatrchyan S, et al. (CMS Collab.) *Phys. Lett. B* 710:256 (2012)
352. Afanasiev S, et al. (PHENIX Collab.) *Phys. Rev. Lett.* 109:152302 (2012)
353. Eskola KJ, Helenius I, Kuha M, Paukkunen H. *Phys. Rev. Lett.* 125:212301 (2020)
354. Niemi H, Eskola KJ, Paatelainen R. *Phys. Rev. C* 93:024907 (2016)
355. Helenius I, Eskola KJ, Honkanen H, Salgado CA. *J. High Energy Phys.* 1207:73 (2012)
356. Bhattacharya A, et al. *J. High Energy Phys.* 1611:167 (2016)
357. Bertone V, Gauld R, Rojo J. *J. High Energy Phys.* 1901:217 (2019)
358. Reno MH. *Annu. Rev. Nucl. Part. Sci.* 73:181 (2023)