Collinear factorization for deep inelastic scattering structure functions at large Bjorken x_B

JHEP07(2008)090

(http://iopscience.iop.org/1126-6708/2008/07/090)

Published by Institute of Physics Publishing for SISSA

Received: *May 28, 2008*
Revised: *June 26, 2008*
Accepted: *July 8, 2008*
Published: *July 21, 2008*

Collinear factorization for deep inelastic scattering structure functions at large Bjorken x_B

Alberto Accardi[abc] **and Jian-Wei Qiu**[c]

[a] *Hampton University, Hampton, VA, 23668, U.S.A.*
[b] *Jefferson Lab, Newport News, VA 23606, U.S.A.*
[c] *Department of Physics and Astronomy, Iowa State University, Ames, IA 50011-3160, U.S.A.*
E-mail: accardi@jlab.org

Abstract: We examine the uncertainty of perturbative QCD factorization for hadron structure functions in deep inelastic scattering at a large value of the Bjorken variable x_B. We analyze the target mass correction to the structure functions by using the collinear factorization approach in the momentum space. We express the long distance physics of structure functions and the leading target mass corrections in terms of parton distribution functions with the standard operator definition. We compare our result with existing work on the target mass correction. We also discuss the impact of a final-state jet function on the extraction of parton distributions at large fractional momentum x.

Keywords: Deep Inelastic Scattering, Parton Model, QCD.

Contents

JHEP07(2008)090

1. Introduction

Much of the predictive power of perturbative Quantum Chromo Dynamics (pQCD) is contained in factorization theorems and in the universality of non-perturbative hadronic matrix elements [1]. Predictions follow when processes with different hard scatterings but the same matrix elements are compared. In the case of leading power contributions, the universal matrix elements are interpreted as parton (quark or gluon) distribution functions (PDFs). With the PDFs extracted from a global QCD analysis [2 – 4], pQCD has been very successful in interpreting and predicting high-energy scattering processes.

However, significant uncertainties still exist in the PDFs due to the accuracy of experimental data and to unknown higher order corrections to perturbative calculations. In particular, the PDFs are least constrained in the region where the parton momentum fraction $x > 0.5$ for valence quark distributions and $x > 0.3$ for gluon and sea quark distributions [2, 3]. On the other hand, precise PDFs are needed for many reasons [5]. For example, the discovery potential of the Large Hadron Collider (LHC) on new physics as excess in particle/jet spectrum at large momentum requires accurate PDFs at large x and large factorization scale μ. Since PDFs at a large μ are obtained by solving DGLAP

evolution equations with input PDFs at a lower factorization scale, and the evolution feeds the large-x partons at a lower scale to those at a higher scale with smaller momentum fraction x, the precision of PDFs at large μ depends on the accuracy of PDFs at large x and low factorization scale. Furthermore, reliable information on the ratio of $d(x)/u(x)$ as $x \to 1$ could provide very important insights into the non perturbative structure of the nucleon [6, 8, 9, 7] and references therein. However, because of the PDFs steeply falling shape as a function of x as $x \to 1$, and because of the convolution of two PDFs, most observables in hadronic collisions do not provide tight enough constraints to the PDFs at large x. On the other hand, inclusive lepton-hadron deep inelastic scattering (DIS) at large Bjorken x_B is a more direct and clean probe of large-x parton distributions. Recently, experiments at the Jefferson Laboratory have produced DIS data at large x_B with high precision, but at relatively low virtuality Q^2 of the exchanged virtual photon in lepton-hadron collisions [10 – 12]. Experiments measure DIS cross sections, or, equivalently, the DIS structure functions, not PDFs. In order to extract PDFs at large x from these and other data at low Q^2, it is necessary to have theoretical control over power corrections, such as the dynamical power corrections (or high twist effects), $\propto \Lambda^2_{\rm QCD}/Q^2$ with the non-perturbative scale $\Lambda_{\rm QCD} \sim 1/\text{fm}$ [13, 14], the target mass corrections (TMC), $\propto x_B^2 m_N^2/Q^2$ with nucleon mass m_N [15], and possibly, final-state jet mass corrections (JMC), $\propto m_j^2/Q^2$ [16]. These corrections become larger and larger as data approach the kinematic limit $x_B = 1$. In this paper, we examine the uncertainties in extracting PDFs at low Q^2 and large x_B caused by the target mass and jet mass corrections.

At the leading power, the perturbative QCD factorization treatment of DIS cross sections neglects all $1/Q^2$-type power corrections. However, TMC play a somewhat special role. Since the mass of the target is a non-perturbative quantity, the partonic dynamics of short-distance factors in the QCD factorization formalism should not depend on it. Therefore, for any hadronic cross section that can be factorized in perturbative QCD, the effect of TMC should be implicitly included in the definition of the non-perturbative hadron matrix elements, and explicitly accounted for in the kinematic variables of the observables. In this sense, TMC are mostly of kinematic origin. At large x_B and low Q^2, the $x_B^2 m_N^2/Q^2$-type TMC can be an important part of the measured cross sections, and should be identified and removed before we extract the leading power PDFs at large x.

Following the pioneering work by Georgi and Politzer (GP) in as early as 1976 [17], many papers have been written on TMC, in particular, for lepton-hadron DIS. A recent review by Schienbein *et al.* provides a nice summary of this effort [15]. Most existing calculations use the technique of operator product expansion (OPE) to resum m_N^2/Q^2 corrections to the structure function moments. A strong debate has been centered on the inversion of the moment formula [18 – 22]. If we keep the target mass in the DIS kinematics, the Bjorken scaling variable x_B for the DIS cross sections or structure functions needs to be replaced by the Nachtmann variable [23], $\xi = 2x_B/(1 + \sqrt{1 + 4x_B^2 m_N^2/Q^2}) \to x_B$ as $m_N^2/Q^2 \to 0$. If the target mass cannot be neglected at low Q^2, the Nachtmann variable ξ is less than 1 even at $x_B = 1$. Only if one ignores the $x_B = 1$ kinematic threshold, and allows ξ to run up to 1, does the inverse Mellin transformation of the structure function

moments give back the structure functions in x_B space [20, 21]. As a consequence, the inverted structure functions are finite in the unphysical $x_B > 1$ region. The unphysical region has been argued to disappear with the inclusion of power-suppressed higher-twist terms in the computation [18]. Alternatively, many prescriptions have been suggested to fix the moments inversion problem, or to phenomenologically eliminate the unphysical region, see [20 – 22]. None of these prescriptions is entirely satisfactory or unique.

To completely avoid the ambiguities in connection with the structure functions moments and their inversion, it is natural to investigate the TMC in the momentum space without using the OPE and taking the moments. This is most easily done in the context of the field theoretic pQCD parton model, as pioneered by Ellis, Furmanski, and Petronzio in ref. [13]. Recently, Kretzer and Reno applied and compared both approaches in the case of neutrino initiated DIS experiments [24, 25]. In this paper, we revisit the TMC in DIS in terms of the perturbative QCD collinear factorization approach in momentum space and express the long distance physics of structure functions and the leading target mass correction in terms of PDFs that share the same partonic operators with the PDFs of zero hadron mass. In our approach, the momentum space structure functions have no unphysical region. Moreover, our approach can be generalized to semi-inclusive DIS and hadronic collisions, where the OPE is not applicable.

In the collinear factorization approach at the leading power in $1/Q^2$, the short-distance factors are perturbatively calculated with massless final-state light partons. As recently pointed by Collins, Rogers and Stasto in ref. [16], the outgoing parton lines should acquire jet subgraphs/functions to have correct kinematics. The invariant mass in the jet subgraph leads to the before mentioned m_j^2/Q^2-type JMC, which are particularly sensitive to the large-x_B kinematics and the extraction of large-x PDFs. In this paper, we discuss the role of the jet functions in modifying the DIS kinematics in the collinear factorization approach. We neglect the soft interactions between the beam jet and the final-state jet functions, and present a collinear factorization formalism for calculating DIS structure functions with a non trivial jet function. Based on a toy-model estimate, we argue that the JMC has a significant effect on the extraction of PDFs when $x \gtrsim 0.6$. The connection of the jet function with lattice QCD computations of the non-perturbative quark propagator is also discussed.

The rest of our paper is organized as follows. In section 2, we drive the TMC in terms of QCD collinear factorization in momentum space. We explicitly demonstrate that our result has no unphysical region for the DIS structure functions. We compare our result with TMC predicted by other approaches. In section 3, we discuss the JMC. Finally, we present our summary and thoughts on future extensions in section 4. In the main text we limit the discussion to light partons and the transverse and longitudinal structure functions. In the appendices, we generalize our formulae.

JHEP07(2008)090

2. Target mass corrections

The DIS cross section is determined by the hadronic tensor

$$W^{\mu\nu}(p,q) = \frac{1}{8\pi}\int d^4z\, e^{-iq\cdot z}\langle p|J^{\dagger\mu}(z)J^{\nu}(0)|p\rangle\,, \tag{2.1}$$

where p is the nucleon 4-momentum, q is the virtual boson 4-momentum, J^μ is the electromagnetic or electroweak current, and $|p\rangle$ is the hadron wave function. In the impulse approximation the lepton-nucleon interaction proceeds through the scattering of the virtual boson with a parton (quark or gluon) belonging to the nucleon, and having 4-momentum k, see figure 1. With these 4-momenta we can build the following useful invariants:

$$x_B = \frac{-q^2}{2p\cdot q}, \quad Q^2 = -q^2, \quad m_N^2 = p^2, \quad x_f = \frac{-q^2}{2k\cdot q}\,. \tag{2.2}$$

The first 3 invariants, namely, the Bjorken variable x_B, the rest mass m_N of the nucleon and the vector boson virtuality Q^2, are experimentally measurable. We call them "external invariants". The fourth invariant, x_f, is the Bjorken variable for a partonic target and is not experimentally measurable, so we call it "internal".

We work in a class of frames, called collinear frames, defined such that p and q do not have transverse momentum. Then we can decompose p, q and k as follows.

$$\begin{aligned} p^\mu &= p^+\overline{n}^\mu + \frac{m_N^2}{2p_A^+}n^\mu \\ q^\mu &= -\xi p^+\overline{n}^\mu + \frac{Q^2}{2\xi p^+}n^\mu \\ k^\mu &= xp^+\overline{n}^\mu + \frac{k^2+k_T^2}{2xp^+}n^\mu + \vec{k}_\perp^{\,\mu}\,. \end{aligned} \tag{2.3}$$

The light-cone vectors n^μ and $\overline{n}^\mu$ satisfy

$$n^2 = \overline{n}^2 = 0 \qquad n\cdot\overline{n} = 1\,, \tag{2.4}$$

and define the light-cone plus and minus directions, respectively. The plus- and minus-components of a 4-vector a are defined by

$$a^+ = a\cdot n \qquad a^- = a\cdot\overline{n}. \tag{2.5}$$

If we choose $\overline{n} = (1/\sqrt{2}, \vec{0}_\perp, 1/\sqrt{2})$ and $n = (1/\sqrt{2}, \vec{0}_\perp, -1/\sqrt{2})$, we obtain $a^\pm = (a_0 \pm a_3)/\sqrt{2}$. The transverse parton momentum k_T satisfies $k_T\cdot n = k_T\cdot\overline{n} = 0$. The nucleon plus-momentum, p^+, can be interpreted as a parameter for boosts along the z-axis, connecting the target rest frame to the hadron infinite-momentum frame. The parton fractional light-cone momentum with respect to the nucleon is defined as

$$x = k^+/p^+\,, \tag{2.6}$$

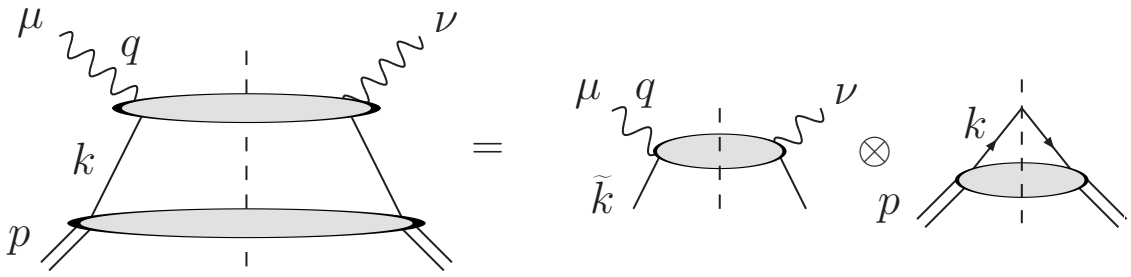

Figure 1: Collinear factorization of the hadronic tensor in the impulse approximation. The top blob represents the interaction of a virtual boson with a parton computed in pQCD at any order in α_s.

and is an internal variable. The virtual boson fractional momentum

$$\xi = -\frac{q^+}{p^+} = \frac{2x_B}{1+\sqrt{1+4x_B^2 m_N^2/Q^2}} \tag{2.7}$$

is an external variable, and coincides with the Nachtmann variable [23]. Note that in the Bjorken limit ($Q^2 \to \infty$ at fixed x_B) $\xi \to x_B$ and we recover the standard kinematics in the massless target approximation. In this paper, we will consider light quarks u, d, s only and set $m^2_{u,d,s} = 0$. In appendix B we will extend our results to heavy quarks.

Collinear factorization for the hadronic tensor can be obtained by expanding the parton momentum k in figure 1 around its positive light-cone component,

$$\widetilde{k}^\mu = x p^+ \overline{n}^\mu \; . \tag{2.8}$$

Correspondingly, we can define the collinear invariant

$$\widetilde{x}_f = \frac{-q^2}{2\widetilde{k}\cdot q} = \frac{\xi}{x} \; . \tag{2.9}$$

According to the QCD factorization theorem [1], the nucleon hadronic tensor can then be factorized as follows:

$$W^{\mu\nu}(p,q) = \sum_f \int \frac{dx}{x}\, \mathcal{H}_f^{\mu\nu}(\widetilde{k},q)\, \varphi_{f/N}(x,Q^2,m_N^2) + O(\Lambda^2/Q^2) \tag{2.10}$$

where $\mathcal{H}_f^{\mu\nu}$ is the short-distance partonic tensor for scattering on a parton of flavor f, and $\varphi_{f/N}$ is the leading twist parton distribution function for a parton of flavor f inside a nucleon N, see figure 1. For example, the quark distribution at leading order in α_s is defined as

$$\varphi_q(x,Q^2,m_N^2) = \int \frac{dz^-}{2\pi} e^{-ixp^+z^-} \langle p|\overline{\psi}(z^- n)\, \frac{\gamma^+}{2}\, \psi(0)|p\rangle \, . \tag{2.11}$$

A proper gauge link between the two fermion field operators is required to have a gauge-invariant parton distribution, but drops out if one chooses the light-cone gauge $n \cdot A = 0$,

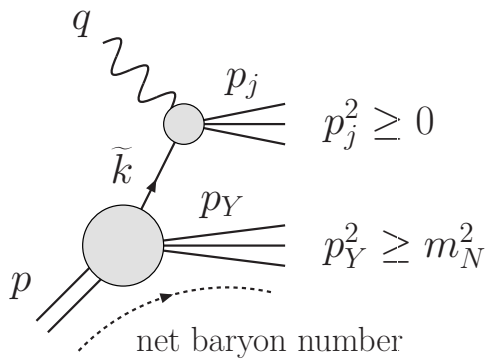

Figure 2: DIS in the impulse approximation, for the special case of an internal on-shell light parton, $k^2 = 0$, relevant to collinear factorization. The current jet has momentum p_j and the target jet has momentum p_Y. The net baryon number is only shown to flow in the target jet (lower part of the graph).

where A is the gluon field. Higher orders in the Taylor expansion are suppressed by powers of Λ^2/Q^2, with Λ a hadronic scale, and contribute to restore gauge invariance in higher twist terms [14]. We will discuss in detail how to obtain such a factorized form in section 3. In eq. (2.10), the partonic tensor $\mathcal{H}^{\mu\nu}$ can be computed perturbatively to any order in α_s, and can depend on the nucleon mass only kinematically through the invariant $\widetilde{x}_f$. Dynamical target mass corrections can enter only through the proton wave function $|p\rangle$, whence the explicit dependence of φ on m_N^2 in eqs. (2.10)-(2.11). From now on we will suppress such dependence for ease of notation. For higher twist terms, the situation is more complicated, because the equations of motion may induce dynamical correlations between lower- and higher-twist terms [13], but we will not discuss this issue here.

Structure functions are obtained by suitable projections of the tensors in eq. (2.10), see appendix B. In this paper, we choose the helicity basis to perform the projection of the $W^{\mu\nu}$ and $\mathcal{H}^{\mu\nu}$ tensors. The transverse and longitudinal structure functions read

$$F_{T,L}(x_B, Q^2, m_N^2) = \sum_f \int \frac{dx}{x} h_{f|T,L}(\widetilde{x}_f, Q^2) \varphi_f(x, Q^2) \ . \tag{2.12}$$

The advantage of the helicity basis is that in the right hand side there are no kinematic prefactors, which would appear when considering the $F_{1,2}$ structure functions, as discussed in ref. [26] and reviewed in appendix B.

Applying the factorized eq. (2.10) without paying attention to the kinematic limits on x, which have been understood in eq. (2.10), and using eq. (2.9), one would obtain what we call the "naïve" TMC in collinear factorization:

$$F_{T,L}^{\rm nv}(x_B, Q^2, m_N^2) = F_{T,L}^{(0)}(\xi, Q^2) \,, \tag{2.13}$$

where $F_{T,L}^{(0)}$ are the structure functions as they would be defined and computed in the massless nucleon limit by setting $m_N^2 = 0$ from the beginning. Indeed, the partonic structure

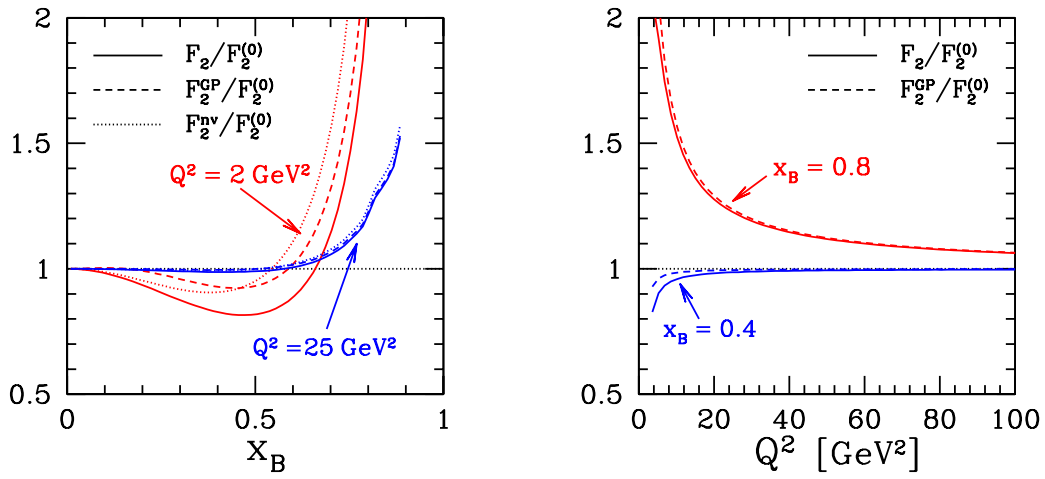

Figure 3: Comparison of prescriptions for NLO target mass corrections to the F_2 structure function. The ratio $F_2/F_2^{(0)}$ is plotted as a function of x_B and Q^2. The structure functions have been computed using MRST2002 parton distributions [3].

functions $h_{f|T,L}$ are independent of the hadron target, and are defined in the same way for the massive and massless nucleon cases. As a consequence of the fact that $F_{T,L}^{(0)}(y,Q^2)$ has support over $0 < y \leq 1$, the target mass corrected $F_{T,L}^{\rm nv}$ can be different from zero in the kinematically forbidden region $1 < x_B \leq 1/(1-m_N^2/Q^2)$. The appearance of such an unphysical region is also a feature of the OPE approach [17, 18], as discussed in the introduction. eq. (2.13) has been introduced in ref. [26] and compared to the OPE approach in refs. [24, 25].

In fact, a closer examination of the handbag diagram kinematics reveals that there is no unphysical region. Let us consider the handbag diagram in the right hand side of figure 1, and limit the discussion to on-shell light quarks or gluons, $k^2 = 0$, in both the initial and final states. The general case of off-shell partons, including heavy quark production is discussed in appendix A. Because of baryon number conservation, the net baryon number must flow either into the target jet or into the current jet. We shall separately examine these two cases. If the net baryon number flows into the target jet (bottom part of figure 2), the jet invariant masses satisfy $m_j^2 = p_j^2 \geq m_f^2$ and $p_Y^2 \geq m_N^2$. Let us consider the invariant momentum square of the process, $s = (p+q)^2 = (p_j + p_Y)^2$. Since the 2 jets are made of on-shell particles, $p_j \cdot p_Y \geq 0$. Hence, $s \geq m_j^2 + m_N^2$. In summary, the current jet mass must satisfy

$$0 \leq m_j^2 \leq s - m_N^2 \ . \tag{2.14}$$

Since $s - m_N^2 = (1/x_B - 1)Q^2$, eq. (2.14) guarantees that the handbag diagram is non-zero only when $x_B \leq 1$, as it must be on general grounds because of baryon number conservation, irrespective of the model used to compute the process. On the other hand, if the net baryon number flows into the current jet (top part of figure 2). The invariant jet

JHEP07(2008)090

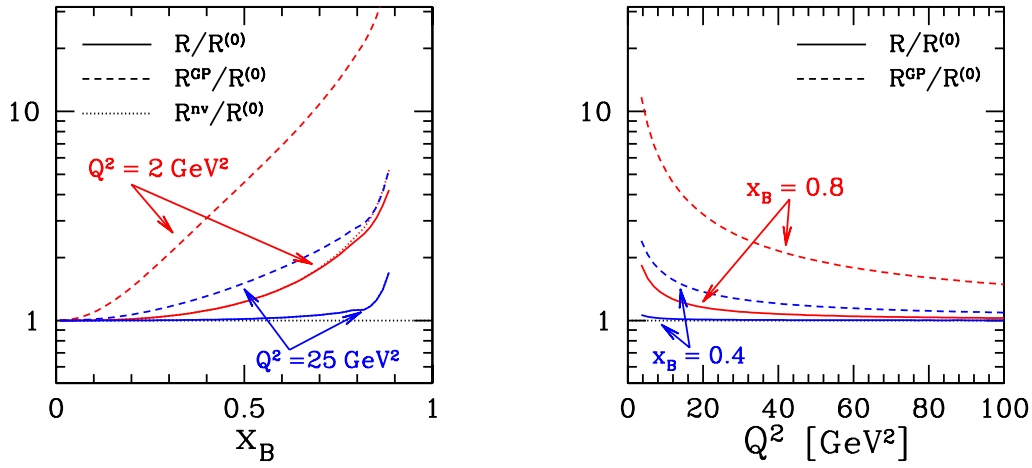

Figure 4: Comparison of prescriptions for NLO target mass corrections to the ratio of the longitudinal and transverse cross sections, $R = \sigma_L/\sigma_T = F_L/F_1$. The ratio $R/R^{(0)}$ is plotted as a function of x_B and Q^2. The structure functions have been computed using MRST2002 parton distributions [3].

masses satisfy $m_j^2 \geq m_N^2$ and $p_Y^2 \geq 0$, so that

$$m_N^2 \leq m_j^2 \leq s\,, \tag{2.15}$$

which again guarantees that the handbag diagram respects the $x_B \leq 1$ limit. Within the collinear factorization approach, the momentum of the active quark entering the short-distance hard part that generates the current jet is approximated to be on mass shell, $\tilde{k}^2 = 0 \ll m_N^2$. That is, the baryon number is very likely to flow into the target jet for the factorized contribution to the DIS cross section, and eq. (2.14) gives the relevant limits on m_j^2. Using $m_j^2 = (\tilde{k}+q)^2 = (1/\widetilde{x}_f - 1)Q^2$ and $\widetilde{x}_f = \xi/x$ in eq. (2.14), we obtain

$$x_B \leq \widetilde{x}_f \leq 1\,, \tag{2.16}$$

which implies the following limits on the dx integration in eq. (2.12):

$$\xi \leq x \leq \frac{\xi}{x_B} \tag{2.17}$$

eqs. (2.16)-(2.17) explicitly guarantee $F_{T,L} = 0$ if $x_B > 1$, so that there is no unphysical region for target mass corrected structure functions:

$$F_{T,L}(x_B, Q^2, m_N^2) = \int_\xi^{\xi/x_B} \frac{dx}{x} h_{f|T,L}(\widetilde{x}_f, Q^2)\varphi_f(x, Q^2)\ . \tag{2.18}$$

Eq. (2.18) is our formula for calculating DIS structure functions with the TMC. As expected, it has the hadron mass dependence explicitly in the integration limits caused by the

JHEP07(2008)090

DIS kinematics and implicitly from the hadron states in the definition of the PDFs. The naïve structure functions (2.13) are obtained when considering $x \leq 1$ as upper integration limit in eq. (2.18). This limit is a general and process-independent consequence of the definition of a parton distribution in the field theoretic parton model [27], but in DIS it is weaker than $x \leq \xi/x_B$, which is induced by 4-momentum and baryon number conservation. In the massless target limit, $m_N^2/Q^2 \to 0$, the constraint (2.17) reduces to $x_B \leq x \leq 1$, and we recover the massless structure functions as we should expect:

$$F_{T,L}(x_B, Q^2, m_N^2) \xrightarrow[m_N^2/Q^2 \to 0]{} F_{T,L}^{(0)}(x_B, Q^2) \; . \tag{2.19}$$

In figure 3, we plot the ratio of the TMC corrected F_2 to the massless $F_2^{(0)}$, with TMC computed using the analog of eq. (2.18), see appendix B.3, the naive prescription (2.13), and the Georgi-Politzer prescription. The corrections are in general quite large at $Q^2 = 2$, but still non negligible at the generally considered "safe" scale $Q^2 = 25\,\mathrm{GeV}^2$. From the right panel of the figure, one can estimate how large Q^2 should be to safely neglect TMC. At $x_B \lesssim 0.5$ the TMC are smaller than 5% if $Q^2 \gtrsim 10\,\mathrm{GeV}^2$. However, at larger x_B, one may need to go to $Q^2 \gtrsim 100\,\mathrm{GeV}^2$ for TMC to become small. Note also the difference between F_2 and F_2^{nv}, which is smaller than 30-40% at $Q^2 = 2$: it gives the size of the contribution of the unphysical region $\xi/x_B < x \leq 1$, which has to be subtracted from the naïve structure function.

The difference between TMC of F_2 (and similarly of $F_{1,T}$) in collinear factorization and in the Georgi-Politzer formalism is smaller than 15-20% at the lowest scale, and rapidly disappears at larger scales. So one is tempted to brush aside the question of what formalism is correct, if willing to accept this level of uncertainty. However, the situation completely changes when considering F_L, or the ratio R of the longitudinal to transverse cross section,

$$R = \frac{\sigma_L}{\sigma_T} = \frac{F_L}{F_1} \, , \qquad R^{(0)} = \frac{F_L^{(0)}}{F_1^{(0)}} \, , \tag{2.20}$$

whose TMC/massless ratios are plotted in figure 4. (Note a factor of $2x_B$ with respect to other common conventions, see appendix B.2.) The TMC of R are much larger than for F_2. Most importantly, the difference between the collinear factorization and Georgi-Politzer TMC is huge, up to a factor 10 (5) at Q^2=2 (25) GeV^2! Therefore, one has to decide which formalism to use. This is especially important for a fit of the gluon PDF, to which F_L is sensitive.

It is also important to note that our formula for TMC in eq. (2.18) explicitly eliminates the kinematically forbidden region $1 < x_B \leq 1/(1 - m_N^2/Q^2)$ because of the integration limits on the parton momentum fraction x. As $x_B \to 1$, structure functions calculated by using eq. (2.18) approach to zero, the kinematic limit, smoothly, except for the lowest order contribution, whose partonic structure functions are derived from the tree level handbag diagram in figure 5. Indeed, by explicit computation at tree level in the approximation of massless quarks, we obtain

$$h_{q|T}(\widetilde{x}_f, Q^2) = \frac{1}{2} \, e_f^2 \, \delta(\widetilde{x}_f - 1) = \frac{1}{2} \, e_f^2 \, x \, \delta(x - \xi) \, , \tag{2.21}$$

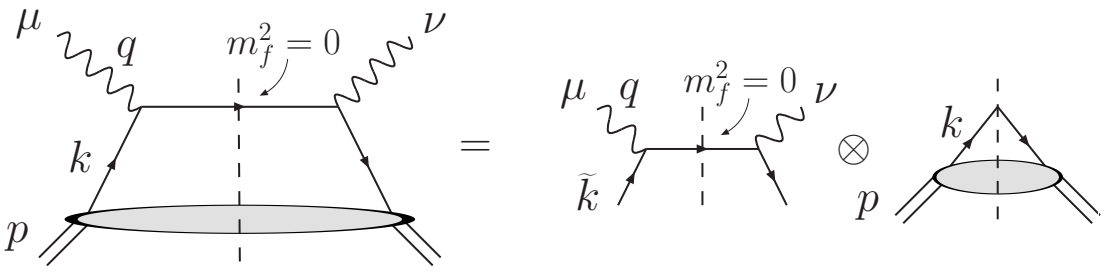

Figure 5: DIS handbag diagram at leading order in α_s.

where e_f is the electric charge of the parton f. Substituting the tree-level partonic structure function into eq. (2.18), the lowest order contribution to the transverse structure function,

$$F_T(x_B, Q^2, m_N^2) = \begin{cases} F_T^{(0)}(\xi, Q^2) & x_B \leq 1 \\ 0 & x_B > 1\,, \end{cases} \tag{2.22}$$

remains positively finite when $x_B \to 1$ and does not vanish as $x_B \to 1$, as shown by the dashed line in figure 6,

This problem exists only at the lowest order and arises because of the δ-function behavior of the partonic structure function (2.21) and the assumption that the final-state is made of a massless quark, $m_f^2 = 0$, as shown in figure 5. The δ-function bypasses the kinematic constraint from the integration limits in eq. (2.18) and forces $x = \xi(x_B)$, which exhibits the mismatch between the phase space for x at the parton level and that for $\xi(x_B)$ at the hadron level. Under the collinear approximation the momentum fraction x for a massless parton can be as large as 1, while the plus momentum fraction of the virtual photon $\xi(x_B)$ smaller than 1 for a finite target mass m_N. As a result, the perturbatively calculated structure functions do not vanish at $x_B = 1$ because the PDFs are finite at $x = \xi(x_B = 1) = 2/(1 + \sqrt{1 + 4m_N^2/Q^2}) < 1$. As we will discuss in the next section, this explicit phase space mismatch at the lowest order could be improved if the single massless quark final-state in figure 5, which is not physical, is replaced by a jet function as shown in figure 7.

We conclude this section by stating that if one is performing global QCD fits of the PDFs in the context of pQCD collinear factorization, our formalism in eq. (2.18) might be the most consistent way to treat TMC, because it expresses the long distance physics of structure functions and the leading target mass correction in terms of PDFs that share the same partonic operators with the PDFs of zero hadron mass. Moreover the structure functions calculated using our formulae do not have the $x_B > 1$ unphysical region and vanish at the $x_B = 1$ kinematic limit except for the lowest order contribution that will be discussed further in next section. The same collinear factorization formalism can be easily and consistently extended to semi-inclusive DIS measurements and hadronic collisions, for which the OPE formalism is not applicable, but which are included in global QCD fits of

JHEP07(2008)090

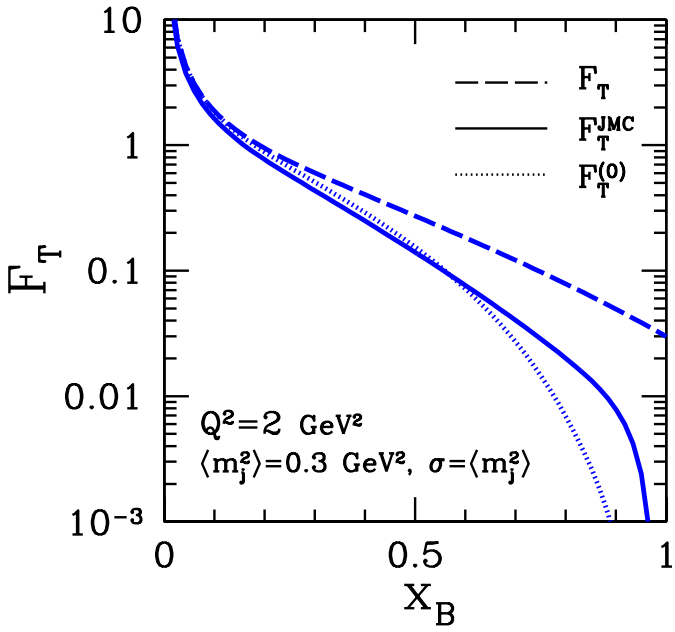

Figure 6: Transverse structure function plotted as a function of x_B, with and without target and jet mass corrections, computed with only light quarks at lowest order in α_s using MRST2001LO parton distributions [36]. The dotted line is the massless structure function. The dashed line includes only TMC, and corresponds to $Z = 1$ in eq. (3.42). The solid line corresponds to JMC coming only from the continuum part ρ of the jet spectral function, $Z = 0$ in eq. (3.42). JMC are computed using a log-normal spectral function with $\langle m_j^2 \rangle = 0.3\,\mathrm{GeV}^2$ and standard deviation $\sigma_{m_j^2} = \langle m_j^2 \rangle$.

parton distributions. Careful analysis of kinematics and conservation laws will guarantee that no unphysical region appears in these observables, as well. The obtained formulae will not merely be an approximation to the TMC for those processes, as argued in [25, 24], but will give the correct answer in the context of pQCD collinear factorization.

3. Jet mass corrections

In this section we discuss the possibility to include a jet function into the lowest order contribution to have a more realistic kinematic constraint on the "single quark" final-state [16]. Hopefully, we can reduce the unphysical positive value of structure functions at $x_B = 1$. As discussed in the last section, this is caused by the δ-function behavior of the partonic structure functions, the assumption that the final-state is made of a massless quark, $m_f^2 = 0$, and the mismatch between the phase space for $\xi(x_B)$ and x.

The assumption that the leading order final-state is made of a massless quark, $m_f^2 = 0$, is clearly unphysical because the quark has to hadronize due to color confinement, so that the current jet will have an invariant mass m_j^2. Then, we may heuristically set $m_f^2 = m_j^2$

JHEP07(2008)090

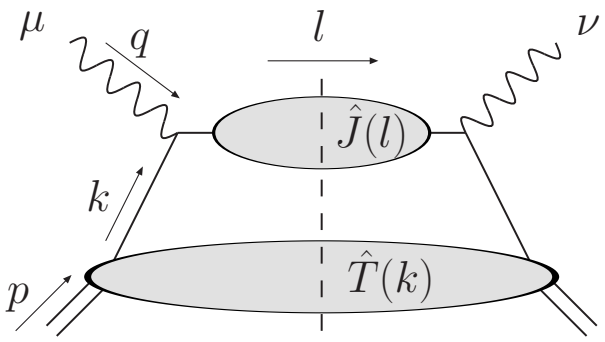

Figure 7: DIS handbag diagram at leading order in α_s generalized to include a jet function $\hat{J}(l)$ beside the target function $\hat{T}(k)$.

for the cut quark line, and substitute $\delta(\widetilde{x}_f - 1)$ with $\delta(\widetilde{x}_f - 1/(1 + m_j^2/Q^2))$ in eq. (2.21):

$$h_{q|T}(\widetilde{x}_f, Q^2) \longrightarrow \frac{1}{2}\, e_f^2\, x\, \delta\left(x - \xi\left(1 + \frac{m_j^2}{Q^2}\right)\right) . \tag{3.1}$$

Furthermore, we may assume that the current jet has an invariant mass probability distribution $J_m(m_j^2)$ normalized to 1, and accordingly smear the structure functions in (2.18):

$$\begin{aligned} F_T^{\mathrm{JMC}}(x_B, Q^2, m_N^2) &= \int_0^\infty dm_j^2 J_m(m_j^2) \int_\xi^{\xi/x_B} \frac{dx}{x} h_{f|T}(\widetilde{x}_f, Q^2) \varphi_f(x, Q^2) \\ &= \int_0^{\frac{1-x_B}{x_B}Q^2} dm_j^2 J_m(m_j^2) F_T^{(0)}\big(\xi(1 + m_j^2/Q^2), Q^2\big) . \end{aligned} \tag{3.2}$$

If $J_m(m_j^2)$ is a sufficiently smooth function of m_j^2, we obtain

$$F_T^{\mathrm{JMC}}(x_B, Q^2, m_N^2) \xrightarrow[x_B \to 1]{} 0 . \tag{3.3}$$

The jet mass corrections (JMC) so introduced are of order $O(m_j^2/Q^2)$. It is easy to see that in the limit $Q^2 \gg \langle m_j^2 \rangle$, the massless $F_T^{(0)}$ decouples from the integration over the jet mass, and we recover the structure function with TMC:

$$F_T^{\mathrm{JMC}}(x_B, Q^2, m_N^2) \xrightarrow[Q^2 \gg \langle m_j^2 \rangle]{} F_T(x_B, Q^2, m_N^2) . \tag{3.4}$$

In the following, we will discuss how to put this Ansatz on a more firm theoretical basis.

3.1 Collinear factorization with a jet function

We aim at including in the DIS handbag at leading order in α_s a suitable jet function to take into account the invariant mass of the jet produced by the hadronization of the struck quark, see figure 7. Note that in computing the DIS cross section with the handbag diagram of figure 7, we are making several assumptions. First, we are assuming that it

makes sense to separate the final state into a current jet and a target jet, respectively the top and the bottom blob. Because of color confinement, this separation can only make sense as an approximation, and is justified for inclusive and semi-inclusive cross sections if the rapidity separation between the 2 jets is large enough. This is in general the case at asymptotically large Q^2. However, at finite Q^2, the rapidity difference between the 2 jets tends to 0 as $x_B \to 1$, and the struck quark may participate in the hadronization process together with the unstruck target partons. Thus, we need to take care in estimating the range in x_B in which the handbag diagram is a meaningful approximation to the DIS process. The second assumption we make, intimately related with the first one, is that color neutralization of the current jet happens via the exchange of soft momenta, which we can neglect when discussing 4-momentum conservation.

In order to obtain a collinear factorization formula, we will closely follow the procedure of Ellis, Furmanski and Petronzio [13]. The hadronic tensor is

$$W^{\mu\nu}(p,q) = \frac{e_f^2}{8\pi} \int \frac{d^4k}{(2\pi)^4} \mathrm{Tr}\big[\hat{T}(k)\gamma^\nu \hat{J}(l)\gamma^\mu\big]\, \mathbb{K}(k,p,q) \ . \tag{3.5}$$

where we considered only 1 flavor for simplicity. The sum over quark flavors will be restored at the end of the computation. We use a hat to denote a matrix in Dirac space. The trace over color indexes can be easily factorized and included in the target function [13]. The remaining trace is over Dirac indexes. The target function $\hat{T}$ is defined as

$$\begin{aligned} \big[\hat{T}(k)\big]_{ij} &= \sum_Y \delta^{(4)}\left(p - k - \sum_{i\in Y} p_i\right) \big|\langle p|k, Y\rangle\big|^2 \\ &= \int d^4z e^{iz\cdot k} \langle p|\overline{\psi}_j(z)\psi_i(0)|p\rangle \,, \end{aligned} \tag{3.6}$$

where $\langle k, Y| = \langle k|\langle Y|$, $\langle Y|$ are all possible final states originating from the target fragmentation, and $\langle k|$ is a parton state of momentum k. Analogously, the jet function $\hat{J}$ is the non-perturbative quark propagator:

$$\begin{aligned} \big[\hat{J}(l)\big]_{ij} &= \sum_Y \delta^{(4)}\left(l - \sum_{i\in Y} p_i\right) \big|\langle l|Y\rangle\big|^2 \\ &= \int d^4z e^{iz\cdot l} \langle 0|\overline{\psi}_j(z)\psi_i(0)|0\rangle \,, \end{aligned} \tag{3.7}$$

and $\langle l|$ is a quark state of momentum l. The jet momentum is constrained by momentum conservation to $l = k+q$, but it is useful to keep it explicit in our formulae. The function $\mathbb{K}$ is included to impose the kinematic constraints, the non-trivial one being $x_f^{\min} \le x_f \le x_f\mathrm{min}$, see appendix A:

$$\mathbb{K}(k,p,q) = \theta(k^+ + q^+)\theta(k^- + q^-)\theta(p^+ - k^+)\theta(p^- - k^-)\theta(x_f - x_B)\theta(1 - x_f) \,, \tag{3.8}$$

where, for light quarks,

$$\begin{aligned} x_f^{\min} &= \frac{x_B}{1 - x_B k^2/Q^2} \\ x_f^{\max} &= \frac{1}{1 - k^2/Q^2} \ . \end{aligned} \tag{3.9}$$

To obtain the leading power contribution, we expand $\hat{T}(k)$ in terms of Dirac matrices and neglect terms that depend on the vector defining the direction of the gauge link in the PDFs, which are suppressed by powers of $1/Q^2$ [14],

$$\hat{T}(k) = \tau_1(k)\hat{\mathbb{I}} + \tau_2(k)\not{k} + \tau_3(k)\gamma_5 + \tau_4(k)\not{k}\gamma_5 \,, \tag{3.10}$$

and, analogously, we expand $\hat{J}$ as :

$$\hat{J}(l) = j_1(l)\hat{\mathbb{I}} + j_2(l)\not{l} + j_3(l)\gamma_5 + j_4(l)\not{l}\gamma_5 \ . \tag{3.11}$$

For massless quarks, $\tau_1 = 0$. The terms $\tau_{3,4}$, which are proportional to γ_5 cancel when computing unpolarized cross sections. In pure QCD, $j_{3,4} = 0$ because of parity invariance, and j_1 only enters in traces with an odd number of γ matrices, hence does not contribute. We are left with the terms proportional to τ_2 and j_2. The dominance of the k^+ and l^- components of k and l in the Breit frame suggests to define

$$\tau_2(k) = \frac{1}{4k^+}\mathrm{Tr}\big[\not{n}\hat{T}(k)\big] = \frac{1}{4k^+}\int d^4z e^{iz\cdot k}\langle p|\overline{\psi}(z)\gamma^+\psi(0)|p\rangle \tag{3.12}$$

$$j_2(l) = \frac{1}{4l^-}\mathrm{Tr}\big[\not{\overline{n}}\hat{J}(l)\big] = \frac{1}{4l^-}\int d^4z e^{iz\cdot l}\langle 0|\overline{\psi}(z)\gamma^-\psi(0)|0\rangle \ . \tag{3.13}$$

After these manipulations, the hadronic tensor reads

$$W^{\mu\nu}(p,q) = \int \frac{dk^+dk^-d^2k_T}{(2\pi)^4}\frac{e_f^2}{8\pi}\mathrm{Tr}\big[\not{k}\gamma^\nu\not{l}\gamma^\mu\big]\, j_2(l)\,\tau_2(k)\,\mathbb{K}(k,p,q)\,, \tag{3.14}$$

where

$$k^\mu = xp^+\overline{n}^\mu + \frac{k^2+k_T^2}{2xp^+}n^\mu + k_T^\mu \tag{3.15}$$

$$l^\mu = (x-\xi)p^+\overline{n}^\mu + \left(\frac{k^2+k_T^2}{2xp^+} + \frac{Q^2}{2\xi p^+}\right)n^\mu + k_T^\mu \ . \tag{3.16}$$

For later use, let us also define

$$\frac{1}{\pi}H_*^{\mu\nu}(k,l) = \frac{e_f^2}{8\pi}\mathrm{Tr}\big[\not{k}\gamma^\nu\not{l}\gamma^\mu\big] \ . \tag{3.17}$$

Our goal is to obtain a factorized expression for the hadronic tensor in terms of collinear parton distribution functions, see for example eq. (2.11). For this purpose we need to let $\int dk^-d^2k_T$ act only on $\tau_2(k)$, which defines the collinear PDF modulo factors of 2. In doing this we will be forced to make approximations on the momenta entering and exiting the hard scattering vertex, viz., k and l. In principle, one would like to avoid it and allow approximations in the computation of the hard scattering tensor only [16]. In this way, one can ensure that the final state obeys 4-momentum conservation, and avoid potentially large errors in region of phase-space close to the kinematic boundaries. While in most cases this is not a problem for inclusive cross sections, it might become very important for exclusive observables. In our case, we want to compute the inclusive DIS cross section at large $x_B \to 1$

JHEP07(2008)090

in collinear factorization: in order to extend the validity of our computation as close as possible to this kinematic boundary, we need to pay attention to the approximations we will make, keep them at a minimum, and estimate the range of validity in x_B and Q^2 of the approximations we will have to make.

The first step in the collinear factorization of the hadronic tensor (3.14), is to expand $H_*^{\mu\nu}$ around the momentum of a collinear and massless quark:

$$H_*^{\mu\nu}(k,l) = H_*^{\mu\nu}(\widetilde{k},\widetilde{l}) + \frac{\partial H_*^{\mu\nu}}{\partial k^\alpha}(k^\alpha - \widetilde{k}^\alpha) + \dots \tag{3.18}$$

where

$$\begin{aligned} \widetilde{k}^\mu &= x p^+ \overline{n}^\mu \\ \widetilde{l}^\mu &= \widetilde{k}^\mu + q^\mu \ . \end{aligned} \tag{3.19}$$

The higher order terms in the expansion are suppressed as powers of Λ^2/Q^2, where Λ^2 is a hadronic scale, and contribute to restore gauge invariance in higher-twist diagrams [14]. In this paper, we will retain only the leading twist term of the expansion. Note that we did not yet make any kinematic approximation: in principle, one may sum over as many higher-twist terms as desired.

The second step involves using the spectral representation of $\hat{J}$ [28] to explicitly introduce the invariant jet mass in the formalism:

$$\hat{J}(l) = \int_0^\infty dm_j^2 \left[J_1(m_j^2)\hat{\mathbb{I}} + J_2(m_j^2)\not{l} \right] 2\pi\delta(l^2 - m_j^2)\theta(l^0)\,, \tag{3.20}$$

where the spectral functions $J_i(m_j^2)$ are positive definite and normalized to 1:

$$\int_0^\infty dm_j^2\, J_i(m_j^2) = 1 \ . \tag{3.21}$$

In particular, by substituting eq. (3.20) into (3.13), we obtain

$$j_2(l) = \int_0^\infty dm_j^2\, J_2(m_j^2)\, 2\pi\delta(l^2 - m_j^2)\,\theta(l^0)\,, \tag{3.22}$$

so that we can interpret m_j as the jet invariant mass, and $J_2(m_j^2)$ as its probability distribution.

In the light-cone gauge $n \cdot A = 0$, the jet spectral function is related to the non perturbative quark propagator:

$$\int_0^\infty dm_j^2\, J_2(m_j^2)\, 2\pi\delta(l^2 - m_j^2)\,\theta(l^0) = \frac{1}{4l^-}\int d^4z e^{iz\cdot l} \mathrm{Tr}\left[\gamma^- \langle 0|\overline{\psi}(z)\psi(0)|0\rangle\right] \ . \tag{3.23}$$

Computations of the non-perturbative quark propagator have been performed in lattice QCD [29] and using Schwinger-Dyson equations, see [30] for a review. However, there are several difficulties in extracting information relevant to the jet spectral function from these computations: (i) the quark-antiquark correlator appearing in (3.23) is typically computed in the Landau gauge instead of the light-cone gauge, (ii) computations are performed

in Euclidean space instead of Minkowski space, (iii) one needs to extract the spectral representation from the computed correlator. The biggest problem is that the analytic structure of the quark propagator is not sufficiently well known to either perform the analytic continuation back to Minkowski space or to extract its spectral representation [30, 31]. As a way to avoid this problem, it would be interesting to see if it is possible to rotate the whole handbag diagram, including its external momenta, to Euclidean space as done in [32] for the computation of the hadronic contribution to the muon anomalous magnetic moment. In this way one would be able to directly use the lattice propagator in the computation of the forward Compton amplitude. Alternatively, one may try to use the light-cone QCD formulation on the lattice discussed in [33], which exploits the Hamiltonian formulation of QCD in order to remain in Minkowski space. A more phenomenological approach to the spectral function will be discussed in the next subsection.

The third step involves our first kinematic approximation. In order to factorize j_2 from the $dk^- d^2k_T$ integrations, we need to approximate $l \to \widetilde{l}$, so that

$$j_2(l) \longrightarrow j_2(\widetilde{l}) = \int_0^\infty dm_j^2 \, J_2(m_j^2) \, 2\pi\delta(\widetilde{l}^2 - m_j^2) \, \theta(l^0) \; . \tag{3.24}$$

Then, the hadronic tensor reads

$$W^{\mu\nu}(p,q) = \int_0^\infty dm_j^2 \, J_2(m_j^2) \int dk^+ \, H_*^{\mu\nu}(\widetilde{k},\widetilde{l}) \, \delta(\widetilde{l}^2 - m_j^2) \int \frac{dk^- d^2k_T}{(2\pi)^4} 2\tau_2(k) \, \mathbb{K}(k,p,q) \tag{3.25}$$

where $\theta(l^0) = \theta(k^0 + q^0)$ is already included in the kinematic constraint function $\mathbb{K}(k,p,q)$. We can expect the approximation (3.24) to be reasonable in a region where $j_2(l)$ does not vary strongly with l. In terms of the spectral representation, this requirement is satisfied if the integral in eq. (3.25) is dominated by values of m_j^2 close to where the jet spectral function has a maximum. We will discuss below the conditions on x_B and Q^2 for which this condition is satisfied. Note that this kinematic approximation only acts on the δ-function in eq. (3.25) so that $J_2(m_j^2)$ has been left unapproximated: in this sense the approximation is the mildest possible compatible with collinear factorization.

The fourth step involves decoupling $\mathbb{K}$ and τ_2 in eq. (3.25). It can be achieved by replacing

$$\mathbb{K}(k,q,p) \longrightarrow \mathbb{K}(\widetilde{k},q,p) = \theta(\widetilde{x}_f - x_B)\theta(1 - \widetilde{x}_f) \; . \tag{3.26}$$

Note that $\theta(\widetilde{k}^0 + q^0) = \theta(p^+ - \widetilde{k}^+) = \theta(p^- - \widetilde{k}^-) = 1$ because of the constraints on $\widetilde{x}_f$. In terms of x,

$$\xi \leq x \leq \xi/x_B \; . \tag{3.27}$$

This is a delicate step because it involves approximating the kinematic constraints, such that the integration over k_T and k^- are unbounded. This clearly is not a good approximation as $x_B \to 1$ [16], in which case the struck parton carries most of the nucleon plus-momentum, so that the minus and transverse components cannot be large. To appreciate

this, consider

$$s = (p+q)^2 = (p_Y + l)^2 \,, \tag{3.28}$$

where p_Y is the total four momentum of the target jet, and we define $m_Y^2 = p_Y^2 \geq 0$, see figure 2. Using the full kinematics of eq. (2.3), we obtain

$$s = \frac{1-\xi}{\xi} Q^2 + (1-\xi) m_N^2 \,. \tag{3.29}$$

On the other hand, in the center-of-mass frame, $\vec{p}_Y = -\vec{l}$ and $\vec{p}_{Y,T} = -\vec{l}_T = -\vec{k}_T$, so that

$$\begin{aligned} s &= (p_Y^0 + l^0)^2 \\ &= \left(\sqrt{m_Y^2 + k_T^2 + (p_Y^3)^2} + \sqrt{m_j^2 + k_T^2 + (l^3)^2} \right)^2 > 4k_T^2 \end{aligned} \tag{3.30}$$

Combining these 2 results, we obtain

$$k_T^2 < \frac{1-\xi}{4\xi} Q^2 \left(1 + \xi \frac{m_N^2}{Q^2} \right) \,. \tag{3.31}$$

As $x_B \to 1$, $\xi \to \xi_{th} \lesssim 1$ so that the $(1-\xi)$ factor tends to close the available k_T phase space. In section 3.3, we will discuss in which region of x_B and Q^2 we may in fact neglect this bound. Using the definition (3.12) of τ_2, the hadronic tensor reads

$$W^{\mu\nu}(p,q) = \int_0^\infty dm_j^2 \, J_2(m_j^2) \int_\xi^{\xi/x_B} \frac{dx}{x} H_*^{\mu\nu}(\widetilde{k}, \widetilde{l}) \, \delta(\widetilde{l}^2 - m_j^2) \varphi_q(x, Q^2) \,, \tag{3.32}$$

where the quark PDF φ_q is defined as in eq. (2.11).

As a last step, we define an on-shell and massless jet momentum for the partonic tensor,

$$\widehat{l}^\mu = \widetilde{l}^- n^\mu = \frac{Q^2}{2\xi p^+} n^\mu \tag{3.33}$$

and replace

$$H_*^{\mu\nu}(\widetilde{k}, \widetilde{l}) \longrightarrow H_*^{\mu\nu}(\widetilde{k}, \widehat{l}) = \frac{e_f^2}{8} \mathrm{Tr}\left[\widetilde{\not{k}} \gamma^\nu \widehat{\not{l}} \gamma^\mu \right] \tag{3.34}$$

This is needed: (i) to ensure that $q_\mu H_*^{\mu\nu} = 0$, hence the gauge invariance of the hadronic tensor, and (ii) to allow use of the Ward identities in proofs of factorization [16]. This approximation, made on the hard scattering coefficient, is less critical then the kinematic approximations previously discussed because it does not change in itself the kinematics of the process. It is analogous to the approximation taken in considering the usual handbag diagram of figure 5 with a massless quark line joining the 2 virtual photon, except that it approximates only the computation of the Dirac traces.

Finally, we define the LO hard scattering tensor

$$\begin{aligned} \mathcal{H}_f^{\mu\nu}(\widetilde{k}, q, m_j^2) &= H_*^{\mu\nu}(\widetilde{k}, \widehat{l}) \delta(\widetilde{l}^2 - m_j^2) & (3.35) \\ &= \mathrm{Tr}\left[\widetilde{\not{k}} \gamma^\nu \widehat{\not{l}} \gamma^\mu \right] \delta(\widetilde{l}^2 - m_j^2) \,, & (3.36) \end{aligned}$$

which for $m_j^2 = 0$ coincides with the LO hard scattering tensor computed for a diagram without jet function, as in eq. (2.10). The hadronic tensor can then be written in factorized form as

$$W^{\mu\nu}(p,q) = \int_0^\infty dm_j^2\, J_2(m_j^2) \int_\xi^{\xi/x_B} \frac{dx}{x} \mathcal{H}_f^{\mu\nu}(\widetilde{k}, q, m_j^2)\, \varphi_q(x, Q^2)\,,$$

which is the central result of this section. The transverse structure function reads

$$F_T(x_B, Q^2, m_N^2) = \int_0^\infty dm_j^2\, J_2(m_j^2) \sum_f \int_\xi^{\xi/x_B} \frac{dx}{x} h_{f|T}(\widetilde{x}_f, Q^2, m_j^2) \varphi_q(x, Q^2)\,, \tag{3.37}$$

with φ_q defined in eq. (2.11). The longitudinal structure function $F_L = 0$ because $h_L = 0$. An explicit computation gives $h_T(\widetilde{x}_f, Q^2) = \frac{1}{2} e_f^2 x \delta\big(x - \xi(1 + m_j^2/Q^2)\big)$ so that at LO

$$F_T(x_B, Q^2, m_N^2) = \int_0^{\frac{1-x_B}{x_B} Q^2} dm_j^2 J_2(m_j^2) F_T^{(0)}\left(\xi\left(1 + \frac{m_j^2}{Q^2}\right), Q^2\right), \tag{3.38}$$

Note that when $Q^2 \gg \langle m_j^2 \rangle$, where $\langle m_j^2 \rangle = \int dm_j^2\, m_j^2 J_2(m_j^2)$, the massless $F_T^{(0)}$ decouples from the integration over the jet mass, and we recover the TMC to the LO structure functions.

3.2 The jet spectral function

Let us discuss more in detail the properties of the nonperturbative quark spectral function J_2, defined in eq. (3.22). Let us start from the definition of the j_2 component of the jet function,

$$\begin{aligned} j_2(l) &= \frac{1}{4l^-} \mathrm{Tr}\big[\gamma^- \hat{J}(l)\big] \\ &= \sum_Y \delta^{(4)}\left(l - \sum_{i \in Y} p_i\right) \langle 0 | \bar{\psi}_f(0) | Y \rangle \gamma^- \langle Y | \psi_f(0) | 0 \rangle\,, \end{aligned} \tag{3.39}$$

with f the quark flavor. For simplicity, we consider only light quark flavors with $m_f \ll m_\pi$. The color c of the quark operator ψ is not neutralized, so that it must appear in the final state $|Y\rangle$. In the physical process, we are assuming that the struck quark's color is neutralized by a soft gluon exchange with the target's remnant. We also assume that we can neglect the soft exchange for the purpose of evaluating the change of kinematics induced by the inclusion of a quark jet function on the lowest order contribution to the inclusive DIS. This assumption is likely valid if the jet and target rapidities are sufficiently separated. However, this might not be the case close to the kinematic limit $x_B = 1$, and the approximation will break down as we shall soon see. Because of color confinement, we may assume that no more than 1 particle in the final state is colored, all the other ones binding into colorless hadrons. The colored particle must be a quark, to match the quark operator's color, and we denote it by $|q_{f'}^c\rangle$. Hence, the final state is made of 1 quark plus an arbitrary number of hadrons, the lightest of which is a pion:

$$|Y\rangle = |q_{f'}^c\rangle |h_1\rangle \cdots |h_N\rangle\,, \tag{3.40}$$

JHEP07(2008)090

with $N \geq 0$ and $f' = f$ when $N = 0$. The spectral function J_2, defined in eq. (3.22), can be written as

$$J_2(m_j^2) = Z\delta(m_j^2 - m_q^2) + (1 - Z)\rho(m_j^2)\,, \tag{3.41}$$

where the δ-function is due to the contribution of the single particle $|q^c_{f'}\rangle$, and $0 < Z < 1$ [28]. The continuous and positive definite function ρ is the contribution of multiparticle states with $N > 0$ in eq. (3.40) and is normalized to 1 because of eq. (3.21). Due to the assumption (3.40), ρ has a bell shape: it is equal to 0 up to $m_j^2 = (m_\pi + m_q)^2 \approx m_\pi^2$, increases up to a maximum and then tends to 0 as $m_j^2 \to \infty$ to satisfy the normalization condition. Using eq. (3.41) in (3.38), we obtain

$$F_T(x_B, Q^2, m_N^2) = Z\, F_T^{(0)}\big(\xi, Q^2\big) + (1 - Z)\int_{m_\pi}^{\frac{1-x_B}{x_B}Q^2} dm_j^2 \rho(m_j^2) F_T^{(0)}\left(\xi\left(1 + \frac{m_j^2}{Q^2}\right), Q^2\right)\,. \tag{3.42}$$

Setting $Z = 1$ is equivalent to calculating the standard handbag diagram without the jet function, and one recovers the TMC formula.

The first term in eq. (3.42) shows that the introduction of the jet function in the handbag diagram goes some way toward softening the problem with the unphysically positive F_T at $x_B = 1$, but does not solve it. The reason is that we cannot kinematically neglect the effect of the color neutralizing soft interactions when we compute the hadbag diagram close to $x_B = 1$, where the rapidity difference between the current and target jets is becoming smaller and smaller. A full solution to this problem is the inclusion of a "soft function", in addition to the target and jet functions, which describes the soft exchanges in the context of fully unintegrated correlation functions [16]. The soft function has essentially the effect of smearing the jet function, avoiding the singular behavior displayed by the δ-function. For a phenomenological inclusion of the soft function in collinear factorization, we can substitute J_2 with a continuous function J_m such that

$$J_m(m_j^2) \xrightarrow[m_j^2 \to 0]{} 0\,, \tag{3.43}$$

because of phase space, and

$$J_m(m_j^2) \xrightarrow[m_j^2 \gg m_\pi^2]{} J_2(m_j^2)\,. \tag{3.44}$$

It can be physically interpreted as the (smeared) jet mass distribution, analogously to J_2, and we will call it smeared jet spectral function. The structure function is then computed as in the Ansatz discussed at the beginning of the section:

$$F_T(x_B, Q^2, m_N^2) = \int_{m_\pi}^{\frac{1-x_B}{x_B}Q^2} dm_j^2 J_m(m_j^2) F_T^{(0)}\left(\xi\left(1 + \frac{m_j^2}{Q^2}\right), Q^2\right)\,. \tag{3.45}$$

We note that the jet spectral function J_2 is defined as a quark correlation function in vacuum, therefore it is process-independent. On the other hand, the smeared jet function

J_m is process-dependent because it effectively includes the soft momentum exchange with the target. As a result, it's shape at $m_j^2 \lesssim m_\pi^2$ might depend on x_B and Q^2. However, the average jet mass squared, $\langle m_j^2 \rangle_m = \int_0^\infty dm_j^2\, m_j^2\, J_m(m_j^2)$ should exhibit a small sensitivity on x_B and Q^2 because we may expect $\langle m_j^2 \rangle \gg m_\pi^2$, see section 3.3. Eq. (3.45) is a reasonable approximation to the full handbag diagram computation in the region of (x_B, Q^2) phase space where the integration over dm_j^2 in eq. (3.38) extends well beyond the peak of the continuum $\rho(m_j^2)$, namely if

$$\frac{1-x_B}{x_B} Q^2 \gtrsim \langle m_j^2 \rangle_\rho \,, \tag{3.46}$$

where

$$\langle m_j^2 \rangle_\rho = \int_{m_\pi^2}^\infty dm_j^2\, m_j^2 J_2(m_j^2) \,. \tag{3.47}$$

In these conditions, the structure function (3.45) is not much sensitive to the behavior of the jet function at small m_j^2, where J_m may substantially differ from J_2.

For practical applications of JMC to global QCD fits of the PDFs, it is necessary to develop a flexible enough and realistic parametrization of the smeared jet spectral function J_m. For this purpose, one may try to use a Monte Carlo simulation of DIS events in order to generate enough data and test possible parametrizations. One may also study the invariant jet mass distribution in $e^+ + e^- \to$ jets events, where the same jet function $\hat{J}$ discussed in this section appears in the LO cross-section. However, these studies lie outside the scope of this paper, and we leave them for the future.

3.3 Numerical estimates

In order to obtain an estimate for the magnitude of JMC and of the present theoretical uncertainty, we employ a toy model for the jet spectral function. Let's consider a bell-shaped smooth function such as the log-normal distribution

$$f(x;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{(\log x - \mu)^2}{2\sigma^2}\right] \tag{3.48}$$

where

$$\begin{aligned} \mu &= \frac{1}{2} \log\left(\frac{\bar{x}^4}{\bar{x}^2 + \sigma_x^2}\right) \\ \sigma &= \left[\log\left(\frac{\sigma_x^2}{\bar{x}^2} + 1\right)\right]^{\frac{1}{2}} \,, \end{aligned} \tag{3.49}$$

and $\bar{x}$ and σ_x are the average value of x and its standard deviation. Then, we can parametrize the continuum part ρ of the toy jet mass distribution in eq. (3.38) in terms of the average jet mass $\langle m_j^2 \rangle_\rho$ and its standard deviation $\sigma_{m_j^2}$:

$$\rho(m_j^2) = f(m_j^2 - m_\pi^2; \mu, \sigma) \,, \tag{3.50}$$

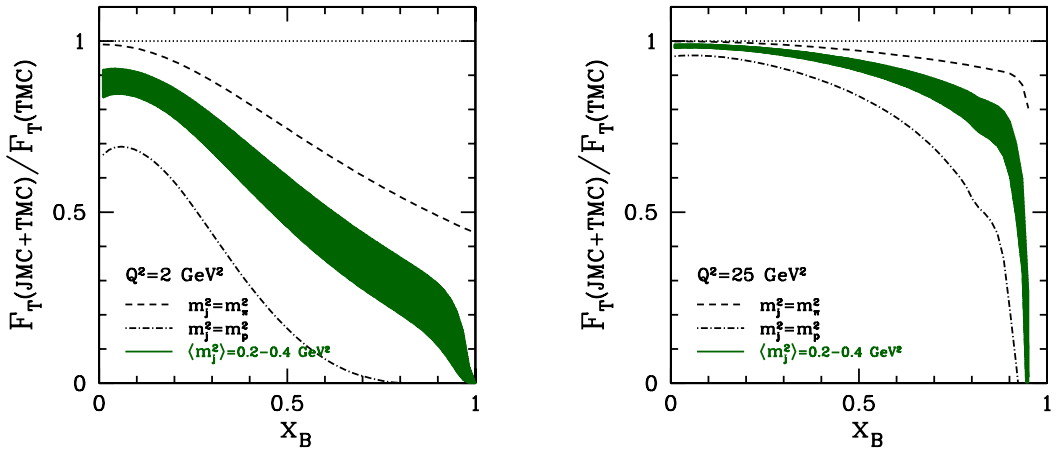

Figure 8: Effect of jet mass corrections on F_T, computed with a toy jet spectral function as described in the text. Plotted is the ratio of F_T with both TMC and JMC to F_T with only TMC included, as a function of x_B for $Q^2 = 2$ and 25 GeV^2. The shaded band corresponds to a log-normal jet mass distribution with $\langle m_j^2\rangle = 0.2 - 0.4\,\mathrm{GeV}^2$ and $\sigma_{m_j^2} = \langle m_j^2\rangle - 2\langle m_j^2\rangle$. The dashed and dot-dashed lines corresponds to delta functions at $m_j^2 = m_\pi^2$ and $m_j^2 = m_N^2$, respectively.

with

$$\begin{aligned} \mu &= \langle m_j^2\rangle_\rho - m_\pi^2 \\ \sigma &= \sigma_{m_j^2} \end{aligned} \tag{3.51}$$

in units of GeV^2. From the typical particle multiplicity of the current jet at the JLab energy, we estimate $\langle m_j^2\rangle_\rho = 0.2 - 0.4\,\mathrm{GeV}^2$, and assume $\sigma_{m_j^2} = C\langle m_j^2\rangle_\rho$ with $C = 1 - 2$.

In figure 6, we plot the JMC to the transverse structure function as obtained in eq. (3.38) by neglecting soft momentum exchanges. The dashed line corresponds to $Z = 1$, and is equivalent to computing only TMC. The solid line corresponds to JMC coming only from the continuum part ρ of the jet spectral function, i.e., $Z = 0$. For comparison, the massless structure function is plotted as a dotted line. The true jet mass corrected F_T should lie somewhere in between because $0 < Z < 1$, in general. With the smearing due to soft interactions, see eq. (3.45), the true F_T will tend to 0 as $x_B \to 1$.

The sensitivity of JMC to the $\langle m_j^2\rangle_\rho$ value can be gauged from figure 8, where we plotted the ratio of the $Z = 1$ TMC-only structure function to the $Z = 0$ jet mass corrected structure function. For comparison, we also use $\rho(m_j^2) = \delta(m_j^2 - m_\pi^2)$ and $\rho(m_j^2) = \delta(m_j^2 - m_N^2)$, considered as extreme cases of JMC. In the absence of a better knowledge of the value of Z and $\langle m_j^2\rangle$, the overall theoretical uncertainty on JMC can be quite large, especially at low $Q^2 = 2\,\mathrm{GeV}^2$, and is still non-negligible at $Q^2 = 25\,\mathrm{GeV}^2$. At moderate $x_B \lesssim 0.6$ it is of the same order of magnitude of the TMC corrections to the massless F_T.

Finally, we want to estimate in which region of x_B and Q^2 the kinematic approximations involved in step 3 and 4 of the factorization procedure are expected to be valid.

JHEP07(2008)090

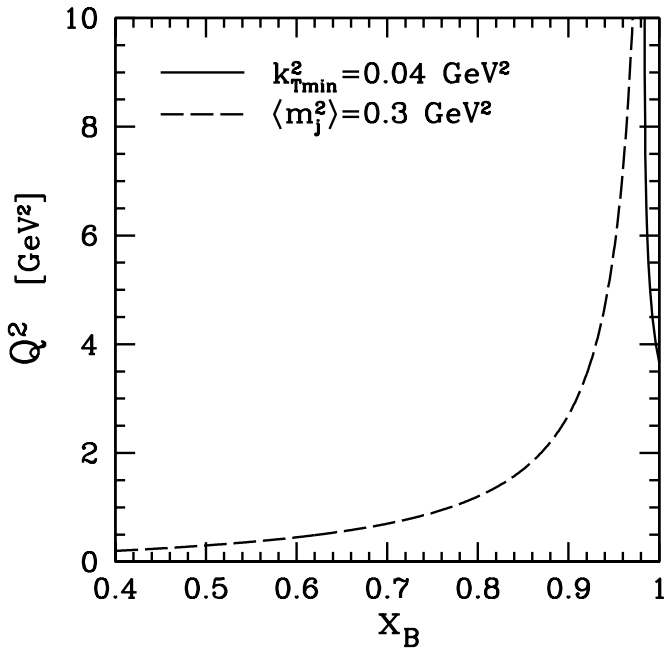

Figure 9: Range of validity of kinematic approximations used in deriving TMC and JMC. The solid line corresponds to eq. (3.52) with $\langle m_j^2 \rangle = 0.3\,\mathrm{GeV}^2$; the dashed line corresponds to eq. (3.53) and (3.54), with $k^2_{T\min} = 0.04\,\mathrm{GeV}^2$.

As we discussed after eq. (3.25), replacing $j_2(l)$ with $j_2(\widetilde{l})$ makes sense only if the integral over dm_j^2 is dominated by $m_j^2 \approx m^2_{j|\max}$, where the spectral function J_2 has a maximum, hence minimal slope. Looking at the integration limits in eq. (3.45), and noticing that for a probability distribution with the properties of the jet spectral function it is typically true that $m^2_{j|\max} \lesssim \langle m_j^2 \rangle \lesssim \langle m_j^2 \rangle_\rho$, where $\langle m_j^2 \rangle = \int_0^\infty dm_j^2\, m_j^2\, J_m(m_j^2)$, we obtain the following condition:

$$\frac{1-x_B}{x_B} Q^2 \gtrsim \langle m_j^2 \rangle_\rho \ . \tag{3.52}$$

Note that it coincides with the condition (3.46) that insures we can indeed approximate $J_2 \approx J_m$ for the computation of inclusive DIS cross section. It also guarantees some rapidity separation between the current and target jets, which is needed to justify the handbag diagram in the first place. The approximation of step 4 consisted in neglecting the integration limits on dk^- and d^2k_T. For the transverse momentum, we need at least to make sure that the average $\langle k_T^2 \rangle$ is well below the upper limit derived in eq. (3.31):

$$\langle k_T^2 \rangle \ll \frac{1-\xi}{4\xi} Q^2 \left(1 + \xi \frac{m_N^2}{Q^2} \right) \ . \tag{3.53}$$

The difficulty is that we cannot estimate $\langle k_T^2 \rangle$ within collinear factorization. To do this, we would need to resort to unintegrated PDF [34, 35], which are still integrated over dk^-, or

to the more recently proposed fully unintegrated PDF [16]. For a rough estimate, we may use the uncertainty principle and set the minimum transverse momentum $k^2_{T\min} = 1/r^2_N \approx 0.04\,\mathrm{GeV}^2$, where r^2_N is the nucleon radius. pQCD evolution will then broaden it roughly according to

$$\langle k^2_T \rangle = k^2_{T\min}[1 + C \log(Q^2/k^2_{T\min})]\,, \tag{3.54}$$

with C a constant of order 1. The borders of the confidence region for the discussed TMC and JMC are plotted in figure 9 using the above estimate for $\langle k^2_T \rangle$, and $\langle m^2_j \rangle = 0.3\,\mathrm{GeV}^2$.

4. Summary and conclusions

In the first part of this paper, we computed the target mass corrections to unpolarized DIS structure functions in the context of collinear factorization. Because of the non-perturbative nature of target mass, we emphasized that for any factorizable hadronic observable, the TMC can only appear explicitly in kinematic variables and implicitly in definitions of non-perturbative hadron matrix elements. The momentum space approach allowed us to avoid the ambiguities related to the moments inversion which affect the OPE treatment of Georgi and Politzer. In particular, we could respect 4-momentum and baryon number conservation, and obtain TMC corrected structure functions without unphysical contributions at $x_B > 1$. When performing global QCD fits of the PDFs in the context of pQCD collinear factorization, the procedure presented in this paper might be the most consistent way to treat TMC, because it expresses the long distance physics of structure functions, and the leading target mass correction, in terms of PDFs that share the same partonic operators with the PDFs of zero hadron mass. Hence it allows to unambiguously separate the kinematic effects of the target's mass from its dynamical contribution to parton matrix elements and the PDFs.

Our formalism for TMC in eq. (2.18) is valid at leading twist and any order in α_s. Calculating TMC for the power-suppressed higher-twist contributions to the structure functions is a non-trivial [13] but important issue for measuring the size of parton correlations in the nucleon wave-function, which we leave to a future effort. The leading-twist formalism can be easily extended to polarized DIS structure functions [37], for which a correct evaluation of TMC is even more important than in the unpolarized case because the bulk of available data is in fact in the large-x_B domain. The extension to semi-inclusive DIS and to hadronic collisions is also very important, in order to fully include TMC in global PDF fits. An example is the Drell-Yan cross-section at large Feynman x_F, which has the potential to further constrain large-x PDFs [38]. It is also straightforward to extend the TMC analysis to DIS on nuclear targets, in order to include the effects of nucleon binding and Fermi motion [39]. This is especially important for studying the large-x neutron PDFs and the d/u ratio, which are extracted from data taken with a Deuterium target.

In the second part of the paper, we examined the impact of a final-state jet function on the extraction of PDFs at large x_B. We proposed to write the leading order hadronic tensor, hence the lowest order contribution to DIS cross section, in terms of the spectral representation J_2 of the jet function, which has the physical meaning of invariant jet mass

JHEP07(2008)090

distribution. We evaluated the impact of JMC on the leading order DIS structure functions, and found it to be potentially large even at not so small values of photon virtuality such as $Q^2 = 25\,\mathrm{GeV}^2$. In the NLO cross-section, the impact of JMC is likely to be reduced, because a non-zero jet invariant mass can be produced in the hard scattering beyond tree level, but is still potentially large. We also evaluated the range of validity in x_B and Q^2 of the approximations we made.

For practical applications to global fits of PDFs, it is important to investigate the shape and properties of the smeared jet spectral function J_m, which effectively includes the neglected soft momentum exchanges in the final state. This can be phenomenologically done using a Monte-Carlo simulation and then trying several parametrizations of J_m. In a more fundamental approach, we noticed that the jet spectral function J_2 is related to the non-perturbative quark propagator, which can be computed in lattice QCD or using Schwinger-Dyson equations. To avoid the difficulties connected to the analytic continuation to Minkowski space, one may try and rotate the whole handbag diagram to Euclidean space, or use a Hamiltonian-based formulation of lattice QCD.

In conclusion, the obtained results on TMC and JMC will be very important when using large-x_B and low-Q^2 data on DIS structure function (like those obtained at Jefferson Lab) to extract reliable PDFs at large-x, and to disentangle kinematic effects from the dynamically interesting higher-twist parton correlations. The discussed extensions of our formalism to other processes will allow a full inclusion of TMC and JMC in global QCD fits of parton distribution functions.

Acknowledgments

We thank C. Aubin, T. Blum, M. E. Christy, J. C. Collins, V. Guzey, C. E. Keppel, P. Maris, W. Melnitchouk, D. Richards, C. Weiss for useful discussions and suggestions. This work has been supported in part by the U.S. Department of Energy, Office of Nuclear Physics, under contract DE-AC02-06CH11357 and contract DE-AC05-06OR23177 under which Jefferson Science Associates, LLC operates the Thomas Jefferson National Accelerator Facility, and under grant DE-FG02-87ER40371 and contract DE-AC02-98CH10886.

A. Kinematic constraints at finite Q^2

Let us consider the handbag diagram for a DIS process on a nucleon target, as depicted in the right hand side of figure 1. We repeat the kinematic analysis of the handbag diagram performed in section 2, but for the general case of an off-shell bound parton of momentum k, and $k^2 \lesssim m_f^2$. The limit of on-shell quarks of mass m_f^2, relevant to collinear factorization, can be obtained setting $k^2 = m_f^2$ and $x_f = \widetilde{x}_f$ in the formulae below.

We consider the scattering of a generic vector boson $(\gamma, W^{\pm}, Z)$ on a parton of flavor f of mass m_f. The lowest order couplings are displayed in figure 10. The masses of the quarks (other than f) coupled to the vector boson are m_1 and m_2. The current jet mass must satisfy

$$m_j^2 \geq s_{\mathrm{th}} \tag{A.1}$$

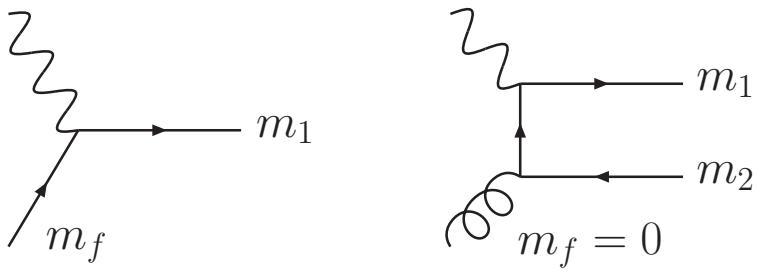

Figure 10: Lowest order couplings of a generic vector boson $(\gamma, W^{\pm}, Z)$ to a parton of flavor f and mass m_f. The masses of the produced quarks are m_1 and m_2. Left: boson-quark scattering $(m_2 = 0)$. Right: boson-gluon fusion.

where

$$s_{\rm th} = (m_1 + m_2)^2 \ . \tag{A.2}$$

As discussed in section 2, the net baryon number is likely to flow through the bottom of the handbag diagram for the leading DIS contribution that is given by the collinear factorization formalism. Therefore,

$$s_{\rm th} \leq m_j^2 \leq s - m_N^2 \ . \tag{A.3}$$

Using $m_j^2 = (q+k)^2 = k^2 + (1/x_f - 1)Q^2$ we obtain

$$\frac{x_B}{1 - x_B k^2/Q^2} \leq x_f \leq \frac{1}{1 + (s_{\rm th} - k^2)/Q^2} \ . \tag{A.4}$$

Using $m_j^2 = (Q^2 + \frac{\xi}{x}k^2)(\frac{x}{\xi} - 1)$, eq. (A.4) can alternatively be expressed as limits over the fractional momentum $x = k^+/p^+$:

$$x^{\rm min} \leq x \leq x^{\rm max} \tag{A.5}$$

where

$$\begin{aligned} x^{\rm min} &= \xi \frac{Q^2 + s_{\rm th} - k^2 + \Delta[k^2, -Q^2, s_{\rm th}]}{Q^2} \\ x^{\rm max} &= \xi \frac{Q^2 + s - m_N^2 - k^2 + \Delta[k^2, -Q^2, s - m_N^2]}{Q^2} \\ \Delta[a,b,c] &= \sqrt{a^2 + b^2 + c^2 - 2(ab + bc + ca)} \ . \end{aligned} \tag{A.6}$$

We finally note that

$$x_f = \frac{\xi}{x} \frac{1}{1 - \frac{\xi^2}{x^2}\frac{k^2}{Q^2}} \ . \tag{A.7}$$

JHEP07(2008)090

B. Invariant and helicity structure functions

B.1 Helicity structure functions

We work in a collinear frame and for generality we keep the quark mass different from zero. For collinear on-shell partons we have from eq. (2.3)

$$\begin{aligned} p^\mu &= p^+ \overline{n}^\mu + \frac{m_N^2}{2p_A^+} n^\mu \\ q^\mu &= -\xi p^+ \overline{n}^\mu + \frac{Q^2}{2\xi p^+} n^\mu \\ \widetilde{k}^\mu &= x p^+ \overline{n}^\mu + \frac{m_f^2}{2x p^+} n^\mu \,, \end{aligned} \tag{B.1}$$

where m_f is the mass of the parton of flavor f. For later use, we define the shorthands

$$\rho_B^2 = 1 + 4x_B^2 \frac{m_N^2}{Q^2} \qquad \rho_f^2 = 1 + 4x_f^2 \frac{m_f^2}{Q^2} \,, \tag{B.2}$$

where, as in eq. (2.2),

$$x_B = \frac{-q^2}{2p\cdot q} \qquad x_f = \frac{-q^2}{2\widetilde{k}\cdot q} \,. \tag{B.3}$$

Following [26], we define the longitudinal, transverse and scalar polarization vectors with respect to the virtual photon momentum q and a reference vector p,

$$\begin{aligned} \varepsilon_0^\mu(p,q) &= \frac{-q^2 p^\mu + (p\cdot q) q^\mu}{\sqrt{-q^2}[(p\cdot q)^2 - q^2 p^2]} = \frac{-q^2 p^\mu + (p\cdot q) q^\mu}{\sqrt{-q^2}(p\cdot q)\rho^2(p,q)} \\ \varepsilon_\pm^\mu(p,q) &= \frac{1}{\sqrt{2}}(0, \pm 1, -i, 0) \\ \varepsilon_q^\mu(p,q) &= \frac{q^\mu}{\sqrt{-q^2}} \,, \end{aligned} \tag{B.4}$$

where

$$\rho^2(p,q) = 1 - p^2 q^2/(p\cdot q)^2 \,. \tag{B.5}$$

It is immediate to verify that $\rho^2(p,q) = \rho_B^2$ and $\rho^2(\widetilde{k},q) = \rho_f^2$. The polarization vectors satisfy the following conditions

$$\begin{aligned} \varepsilon_\lambda \cdot \varepsilon_{\lambda'} &= 0 \qquad && \text{for } \lambda \neq \lambda' \\ \varepsilon_\lambda \cdot \varepsilon_\lambda &= 1 \qquad && \text{for } \lambda = 0, +, - \\ \varepsilon_q \cdot \varepsilon_q &= -1 \end{aligned} \tag{B.6}$$

and, in particular, $q\cdot\varepsilon_0 = q\cdot\varepsilon_\pm = 0$. The helicity structure functions F_λ are defined as projections of the hadronic tensor:

$$F_\lambda(x_B, Q^2) = P_\lambda^{\mu\nu}(p,q) W_{\mu\nu}(p,q) \tag{B.7}$$

with $\lambda = L, T, A, S, \{0q\}, [0q]$. The longitudinal, transverse, axial, scalar, and mixed projectors $P_\lambda^{\mu\nu}$ are

$$\begin{aligned}
P_L^{\mu\nu}(p,q) &= \varepsilon_0^\mu(p,q)\varepsilon_0^{\nu*}(p,q) \\
P_T^{\mu\nu}(p,q) &= \varepsilon_+^\mu(p,q)\varepsilon_+^{\nu*}(p,q) + \varepsilon_-^\mu(p,q)\varepsilon_-^{\nu*}(p,q) \\
P_A^{\mu\nu}(p,q) &= \varepsilon_+^\mu(p,q)\varepsilon_+^{\nu*}(p,q) - \varepsilon_-^\mu(p,q)\varepsilon_-^{\nu*}(p,q) \\
P_q^{\mu\nu}(p,q) &= \varepsilon_q^\mu(p,q)\varepsilon_q^{\nu*}(p,q) \\
P_{\{0q\}}^{\mu\nu}(p,q) &= \varepsilon_0^\mu(p,q)\varepsilon_q^{\nu*}(p,q) + \varepsilon_q^\mu(p,q)\varepsilon_0^{\nu*}(p,q) \\
P_{[0q]}^{\mu\nu}(p,q) &= \varepsilon_0^\mu(p,q)\varepsilon_q^{\nu*}(p,q) - \varepsilon_q^\mu(p,q)\varepsilon_0^{\nu*}(p,q)\ .
\end{aligned} \tag{B.8}$$

Using

$$\varepsilon_+^\mu(p,q)\varepsilon_+^{\nu*}(p,q) - \varepsilon_-^\mu(p,q)\varepsilon_-^{\nu*}(p,q) = \frac{-i\varepsilon^{\mu\nu\alpha\beta}p_\alpha q_\beta}{(p\cdot q)\rho_B} \tag{B.9}$$

$$\varepsilon_+^\mu(p,q)\varepsilon_+^{\nu*}(p,q) + \varepsilon_-^\mu(p,q)\varepsilon_-^{\nu*}(p,q) = -g^{\mu\nu} + \varepsilon_0^\mu(p,q)\varepsilon_0^{\nu*}(p,q) - \varepsilon_q^\mu(p,q)\varepsilon_q^{\nu*}(p,q)\,, \tag{B.10}$$

one easily sees that

$$\begin{aligned}
F_T(x_B,Q^2) &= -W_\mu^\mu(p,q) + F_L(x_B,Q^2) - F_q(x_B,Q^2) \\
F_A(x_B,Q^2) &= \frac{-i\varepsilon^{\mu\nu\alpha\beta}p_\alpha q_\beta}{(p\cdot q)\rho_B}W_{\mu\nu}(p,q)\ .
\end{aligned} \tag{B.11}$$

Even if not apparent from eq. (B.4), a consequence of the normalization conditions is that the reference vector has the only function to define the $t-z$ and transverse planes in conjunction with q^μ: as long as it lays in the $t-z$ plane, a different reference vector defines the same polarization vectors [26]. For example, $\varepsilon_\lambda^\mu(p,q) = \varepsilon_\lambda^\mu(\widetilde{k},q)$. As we will see, choosing $\widetilde{k}$ instead of p is convenient when defining the parton level helicity structure functions, which read

$$h_\lambda(x_f,Q^2) = P_\lambda^{\mu\nu}(\widetilde{k},q)\mathcal{H}_{\mu\nu}(\widetilde{k},q) \tag{B.12}$$

and satisfy identities analogous to eq. (B.11), with $p\to\widetilde{k}$.

B.2 Invariant structure functions

For a generic lepton-hadron scattering, we define the hadronic F_i and partonic h_i invariant structure functions with $i = 1,\ldots,6$ by the following tensor decomposition of the hadronic tensor:

$$\begin{aligned}
W^{\mu\nu}(p,q) = &\left(-g^{\mu\nu} + \frac{q^\mu q^\nu}{q^2}\right)F_1(x_B,Q^2) \\
&+ \left(p^\mu - q^\mu\frac{p\cdot q}{q^2}\right)\left(p^\nu - q^\nu\frac{p\cdot q}{q^2}\right)\frac{F_2(x_B,Q^2)}{p\cdot q} \\
&- i\varepsilon^{\mu\nu\alpha\beta}p_\alpha q_\beta\frac{F_3(x_B,Q^2)}{p\cdot q} - \frac{q^\mu q^\nu}{q^2}F_4(x_B,Q^2) \\
&- \frac{p^\mu q^\nu + q^\mu p^\nu}{2p\cdot q}F_5(x_B,Q^2) + \frac{p^\mu q^\nu - q^\mu p^\nu}{2p\cdot q}F_6(x_B,Q^2)
\end{aligned} \tag{B.13}$$

and

$$\mathcal{H}^{\mu\nu}(\widetilde{k},q) = \Big(-g^{\mu\nu} + \frac{q^\mu q^\nu}{q^2}\Big) h_1(\widetilde{x}_f,Q^2) \tag{B.14}$$
$$+ \left(\widetilde{k}^\mu - q^\mu \frac{\widetilde{k}\cdot q}{q^2}\right)\left(\widetilde{k}^\nu - q^\nu \frac{\widetilde{k}\cdot q}{q^2}\right)\frac{h_2(\widetilde{x}_f,Q^2)}{\widetilde{k}\cdot q}$$
$$- i\varepsilon^{\mu\nu\alpha\beta}\widetilde{k}_\alpha q_\beta \frac{h_3(\widetilde{x}_f,Q^2)}{\widetilde{k}\cdot q} - \frac{q^\mu q^\nu}{q^2} h_4(\widetilde{x}_f,Q^2)$$
$$- \frac{\widetilde{k}^\mu q^\nu + q^\mu \widetilde{k}^\nu}{2\widetilde{k}\cdot q} h_5(\widetilde{x}_f,Q^2) + \frac{\widetilde{k}^\mu q^\nu - q^\mu \widetilde{k}^\nu}{2\widetilde{k}\cdot q} h_6(\widetilde{x}_f,Q^2)\ .$$

These 2 definitions differ from the notation of ref. [26] in the chosen denominators. Our definitions have the advantage of displaying a duality between the hadron and parton level, which can be obtained from each other by exchanging $p \leftrightarrow \widetilde{k}$, and lead to a lesser degree of mixing between the hadron and parton structure functions under collinear factorization, see eq. (B.25). By applying the projectors (B.8) to eqs. (B.13)-(B.14), it is straightforward to show that

$$\begin{aligned}
F_L &= -F_1 + \frac{\rho_B^2}{2x_B} F_2 & h_L &= -h_1 + \frac{\rho_f^2}{2\widetilde{x}_f} h_2 \\
F_T &= 2F_1 & h_T &= 2h_1 \\
F_A &= \rho_B F_3 & h_A &= \rho_f h_3 \\
F_S &= F_4 - F_5 & h_S &= h_4 - h_5 \\
F_{\{0q\}} &= -\rho_B F_5 & h_{\{0q\}} &= -\rho_f h_5 \\
F_{[0q]} &= -\rho_B F_6 & h_{[0q]} &= -\rho_f h_6\,,
\end{aligned} \tag{B.15}$$

where we understood the dependence of $F_{i\lambda}$ on (x_B,Q^2) and of $h_{i\lambda}$ on $(\widetilde{x}_f,Q^2)$ for ease of notation. Note that F_L differs by a factor of $2x_B$ with respect to other common conventions. In our notation, the ratio R of transverse and longitudinal electron-nucleon cross sections reads

$$R = \frac{\sigma_T}{\sigma_L} = \frac{F_L}{F_1}\ . \tag{B.16}$$

B.3 Collinear factorization for structure functions

As discussed in section 2 and appendix A, the collinear factorization theorem states that

$$W^{\mu\nu}(p,q) = \sum_f \int \frac{dx}{x}\,\theta(\widetilde{x}_f^{\max} - \widetilde{x}_f)\theta(\widetilde{x}_f - \widetilde{x}_f^{\min})\mathcal{H}_f^{\mu\nu}(\widetilde{k},q)\,\varphi_{f/N}(x,Q^2) \tag{B.17}$$

where

$$\widetilde{x}_f = \frac{\xi}{x}\frac{1}{1-\frac{\xi^2}{x^2}\frac{m_f^2}{Q^2}} \tag{B.18}$$

$$\widetilde{x}_f^{\min} = \frac{x_B}{1 - x_B m_f^2/Q^2} \tag{B.19}$$

$$\widetilde{x}_f^{\max} = \frac{1}{1 + (s_{\rm th} - m_f^2)/Q^2}\ . \tag{B.20}$$

The corresponding limits of integration on dx, namely $x^{\min}$ and $x^{\max}$, can be read off eq. (A.6) setting $k^2 = m_f^2$. As discussed in appendix B.1, $P_\lambda^{\mu\nu}(p,q) = P_\lambda^{\mu\nu}(\widetilde{k},q)$, hence the factorization theorem for helicity structure functions reads

$$F_\lambda(x_B, Q^2, m_N^2) = \sum_f \int_{x^{\min}}^{x^{\max}} \frac{dx}{x}\, h_\lambda^f(\widetilde{x}_f, Q^2)\, \varphi_{f/N}(x, Q^2) \tag{B.21}$$

$$= \sum_f \int_{\widetilde{x}_f^{\min}}^{\widetilde{x}_f^{\max}} \frac{d\widetilde{x}_f}{\widetilde{x}_f}\, h_\lambda^f(\widetilde{x}_f, Q^2)\, \varphi_{f/N}\left(\frac{\xi}{\xi_f}, Q^2\right), \tag{B.22}$$

where

$$\xi_f = \frac{2\widetilde{x}_f}{1 + \sqrt{1 + 4\widetilde{x}_f^2 m_f^2/Q^2}}\ . \tag{B.23}$$

The last line of eq. (B.22) is particularly interesting, because the Nachtmann variable ξ only appears in the argument of φ, without touching the integration limits. In shorthand notation, where we highlight the dependence on x_B and ξ and suppress that on m_N^2 and Q^2, and understand the sum over f, the helicity structure functions read

$$F_\lambda(x_B) \equiv h_\lambda^f \otimes \varphi_{f/N}(\xi)\ . \tag{B.24}$$

For the invariant structure functions, kinematic prefactors often appear:

$$\begin{aligned}
F_1(x_B) &= h_1^f \otimes \varphi_{f/N}(\xi) \\
F_2(x_B) &= \frac{x_B}{\widetilde{x}_f}\frac{\rho_f^2}{\rho_B^2} h_2^f \otimes \varphi_{f/N}(\xi) \\
F_{3,5,6}(x_B) &= \frac{\rho_f}{\rho_B} h_{3,5,6}^f \otimes \varphi_{f/N}(\xi) \\
F_4(x_B) &= h_4^f \otimes \varphi_{f/N}(\xi) + \left(\frac{\rho_f}{\rho_B} - 1\right) h_5^f \otimes \varphi_{f/N}(\xi)\ .
\end{aligned} \tag{B.25}$$

The "massless structure functions" can be obtained by setting $m_N^2 = 0$, hence, $\xi = x_B$ in eqs. (B.24)-(B.25):

$$F_{\lambda,i}^{(0)}(x_B) = F_{\lambda,i}(x_B)|_{m_N^2=0}\ . \tag{B.26}$$

In this definition we left the quark mass m_f arbitrary.

The "naïve" target mass corrected structure functions $F^{\rm nv}$ are obtained by using $x \leq 1$ as an upper limit of integration over dx in eq. (B.22). This limit is a general and process-independent consequence of the definition of a parton distribution in the field theoretic parton model [27], but in DIS it is weaker than $x \leq x^{\max}$, which is induced by 4-momentum and baryon number conservation as discussed in section 2. In detail, the naïve helicity structure functions read

$$F_\lambda^{\rm nv}(x_B) = \sum_f \int_{x^{\min}}^{1} \frac{dx}{x}\, h_\lambda(\widetilde{x}_f, Q^2)\, \varphi_{f/N}(x, Q^2) \tag{B.27}$$

Using the definition of massless structure functions, one finds

$$\begin{aligned} F_{1,\lambda}^{\rm nv}(x_B) &= F_{1,\lambda}^{(0)}(\xi) \\ F_2^{\rm nv}(x_B) &= \frac{1}{\rho_B^2}\frac{x_B}{\xi}F_2^{(0)}(\xi) \\ F_{3,4,5}^{\rm nv}(x_B) &= \frac{1}{\rho_B}F_{3,4,5}^{(0)}(\xi) \\ F_4^{\rm nv}(x_B) &= F_4^{(0)}(\xi) + \frac{1-\rho_B}{\rho_B}F_5^{(0)}(\xi)\ . \end{aligned} \tag{B.28}$$

These formulae have already appeared in [26, 25, 24], modulo the change of notation discussed in appendix B.2. As already noted in the main text, they are non-zero in the unphysical region $x_B > 1$.

B.4 Structure functions with Jet Mass Corrections

At LO, the helicity structure functions with jet mass corrections read

$$F_\lambda(x_B, Q^2, m_N^2) = \int_0^{\frac{1-x_B}{x_B}Q^2} dm_j^2 J_2(m_j^2) F_\lambda^{(0)}\big(\xi(1+m_j^2/Q^2), Q^2\big)\ . \tag{B.29}$$

The JMC to invariant structure functions can be obtained from eqs. (B.29) and (B.15). Suppressing the Q^2 and m_N^2 dependence of the structure functions for ease of notation, we obtain:

$$\begin{aligned} F_1^{\rm JMC}(x_B) &= \int_0^{\frac{1-x_B}{x_B}Q^2} dm_j^2 J_2(m_j^2) F_1^{(0)}\big(\xi(1+m_j^2/Q^2)\big) \\ F_2^{\rm JMC}(x_B) &= \int_0^{\frac{1-x_B}{x_B}Q^2} dm_j^2 J_2(m_j^2) \frac{1}{\rho_B^2}\frac{x_B}{\xi(1+m_j^2/Q^2)} F_2^{(0)}\big(\xi(1+m_j^2/Q^2)\big) \\ F_{3,5,6}^{\rm JMC}(x_B) &= \frac{1}{\rho_B}\int_0^{\frac{1-x_B}{x_B}Q^2} dm_j^2 J_2(m_j^2) F_{3,5,6}^{(0)}\big(\xi(1+m_j^2/Q^2)\big) \\ F_4^{\rm JMC}(x_B) &= \int_0^{\frac{1-x_B}{x_B}Q^2} dm_j^2 J_2(m_j^2)\left\{ F_4^{(0)}\big(\xi(1+m_j^2/Q^2)\big) + \frac{1-\rho_B}{\rho_B}F_5^{(0)}\big(\xi(1+m_j^2/Q^2)\big)\right\}\ . \end{aligned} \tag{B.30}$$

C. Target mass corrections in the OPE formalism

We collect here for completeness the target mass corrections to the electromagnetic structure functions obtained in the operator product expansion formalism of De Rujula, Georgi and Politzer [17, 18], see also [15] for a thorough review and discussion:

$$\begin{aligned} F_1^{\rm GP}(x_B, Q^2) &= \frac{x_B}{\rho_B}\left[\frac{F_1^{(0)}(\xi, Q^2)}{\xi} + \frac{m_N^2 x_B}{Q^2\rho_B}\Delta_2(x_B, Q^2)\right] \\ F_2^{\rm GP}(x_B, Q^2) &= \frac{x_B^2}{\rho_B^3}\left[\frac{F_2^{(0)}(\xi, Q^2)}{\xi^2} + 6\frac{m_N^2 x_B}{Q^2\rho_B}\Delta_2(x_B, Q^2)\right] \\ F_L^{\rm GP}(x_B, Q^2) &= \frac{x_B}{\rho_B}\left[\frac{F_L^{(0)}(\xi, Q^2)}{\xi} + 2\frac{m_N^2 x_B}{Q^2\rho_B}\Delta_2(x_B, Q^2)\right] \end{aligned} \tag{C.1}$$

where

$$\Delta_2(x_B, Q^2) = \int_\xi^1 dv \left[1 + 2\frac{m_N^2 x_B}{Q^2 \rho_B}(v - \xi)\right] \frac{F_2^{(0)}(v, Q^2)}{v^2}\,, \tag{C.2}$$

and $F_i^{(0)}$ are the perturbative structure functions computed in the massless target approximation $m_N^2/Q^2 \to 0$. Formulae for the $F_{3-6}^{\rm GP}$ structure functions can be found in ref. [40].

Note that in the notation of appendix B, differently from ref. [15], the longitudinal structure function is defined such that

$$F_L(x_B) = \frac{\rho^2}{2x_B} F_2(x_B) - F_1(x_B)\,, \tag{C.3}$$

and

$$R(x_B) \equiv \frac{\sigma_L(x_B)}{\sigma_T(x_B)} = \frac{F_L(x_B)}{F_1(x_B)}. \tag{C.4}$$

This notation is explained in detail in the appendices of ref. [26]. Combining eqs. (C.3) and (C.4) we obtain

$$F_1(x_B) = \frac{\rho_B^2}{2x_B} \frac{F_2(x_B)}{1 + R(x_B)} \tag{C.5}$$

in agreement with ref. [15].

Equations (C.1) have been used to compute the OPE target mass corrections in figures 3 and 4. Note that both $F_L^{\rm GP}$ and $F_1^{\rm GP}$ receive a correction from an integral of $F_2^{(0)} \gg F_{1,L}^{(0)}$. This explains the large size of the target mass corrections for the OPE curves of figure 4.

References

[1] For a review, see J.C. Collins, D.E. Soper and G. Sterman, *Factorization of hard processes in QCD*, *Adv. Ser. Direct. High Energy Phys.* **5** (1988) 1 [hep-ph/0409313].

[2] J. Pumplin et al., *New generation of parton distributions with uncertainties from global QCD analysis*, *JHEP* **07** (2002) 012 [hep-ph/0201195].

[3] A.D. Martin, R.G. Roberts, W.J. Stirling and R.S. Thorne, *Uncertainties of predictions from parton distributions. I: experimental errors*, *Eur. Phys. J.* **C 28** (2003) 455 [hep-ph/0211080].

[4] M. Gluck, E. Reya and A. Vogt, *Dynamical parton distributions revisited*, *Eur. Phys. J.* **C 5** (1998) 461 [hep-ph/9806404].

[5] S. Kuhlmann et al., *Large-x parton distributions*, *Phys. Lett.* **B 476** (2000) 291 [hep-ph/9912283].

[6] W. Melnitchouk and A.W. Thomas, *Neutron/proton structure function ratio at large x*, *Phys. Lett.* **B 377** (1996) 11 [nucl-th/9602038].

[7] G.R. Farrar and D.R. Jackson, *Pion and nucleon structure functions near $x = 1$*, *Phys. Rev. Lett.* **35** (1975) 1416.

JHEP07(2008)090

[8] S.J. Brodsky, M. Burkardt and I. Schmidt, *Perturbative QCD constraints on the shape of polarized quark and gluon distributions*, *Nucl. Phys.* **B 441** (1995) 197 [hep-ph/9401328].

[9] N. Isgur, *Valence quark spin distribution functions*, *Phys. Rev.* **D 59** (1999) 034013 [hep-ph/9809255].

[10] V. Tvaskis et al., *Longitudinal-transverse separations of structure functions at low Q^2 for hydrogen and deuterium*, *Phys. Rev. Lett.* **98** (2007) 142301 [nucl-ex/0611023]; Jefferson Lab Hall C E94-110 collaboration, Y. Liang et al., *Measurement of $R = \sigma(L)/\sigma(T)$ and the separated longitudinal and transverse structure functions in the nucleon resonance region*, nucl-ex/0410027.

[11] W. Melnitchouk, *Jefferson lab phenomenology: an overview*, *Nucl. Phys.* **141** *(Proc. Suppl.)* (2005) 151.

[12] L.S. Cardman, *Physics at the Thomas Jefferson national accelerator facility*, *Eur. Phys. J.* **A28S1** (2006) 7.

[13] R.K. Ellis, W. Furmanski and R. Petronzio, *Unraveling higher twists*, *Nucl. Phys.* **B 212** (1983) 29.

[14] J.-W. Qiu, *Twist four contributions to the parton structure functions*, *Phys. Rev.* **D 42** (1990) 30.

[15] For a review, see I. Schienbein et al., *A review of target mass corrections*, *J. Phys.* **G 35** (2008) 053101 [arXiv:0709.1775].

[16] J.C. Collins, T.C. Rogers and A.M. Stasto, *Fully unintegrated parton correlation functions and factorization in lowest order hard scattering*, *Phys. Rev.* **D 77** (2008) 085009 [arXiv:0708.2833].

[17] H. Georgi and H.D. Politzer, *Freedom at moderate energies: masses in color dynamics*, *Phys. Rev.* **D 14** (1976) 1829.

[18] A. De Rujula, H. Georgi and H.D. Politzer, *Demythification of electroproduction, local duality and precocious scaling*, *Ann. Phys. (NY)* **103** (1977) 315.

[19] D.J. Gross, S.B. Treiman and F.A. Wilczek, *Mass corrections in deep inelastic scattering*, *Phys. Rev.* **D 15** (1977) 2486.

[20] P.W. Johnson and W.K. Tung, *Structure functions and their moments as testing ground for QCD: the Pandora's box of mass effects*, in the Proceedings of *NEUTRINO* 79, Bergen, Norway, June 18–22, 1979.

[21] K. Bitar, P.W. Johnson and W.-k. Tung, *QCD asymptotics and kinematic thresholds in deep inelastic scattering*, *Phys. Lett.* **B 83** (1979) 114.

[22] F.M. Steffens and W. Melnitchouk, *Target mass corrections revisited*, *Phys. Rev.* **C 73** (2006) 055202 [nucl-th/0603014].

[23] O. Nachtmann, *Positivity constraints for anomalous dimensions*, *Nucl. Phys.* **B 63** (1973) 237.

[24] S. Kretzer and M.H. Reno, *Tau neutrino deep inelastic charged current interactions*, *Phys. Rev.* **D 66** (2002) 113007 [hep-ph/0208187].

[25] S. Kretzer and M.H. Reno, *Target mass corrections to electro-weak structure functions and perturbative neutrino cross sections*, *Phys. Rev.* **D 69** (2004) 034002 [hep-ph/0307023].

JHEP07(2008)090

[26] M.A.G. Aivazis, F.I. Olness and W.-K. Tung, *Leptoproduction of heavy quarks. 1. General formalism and kinematics of charged current and neutral current production processes*, *Phys. Rev.* **D 50** (1994) 3085 [hep-ph/9312318].

[27] R.L. Jaffe, *Parton distribution functions for twist four*, *Nucl. Phys.* **B 229** (1983) 205.

[28] S. Weinberg, *The quantum theory of fields. Volume* 1*: foundations*, Cambridge University Press, Cambridge U.K. (1995).

[29] P.O. Bowman et al., *Unquenched quark propagator in Landau gauge*, *Phys. Rev.* **D 71** (2005) 054507 [hep-lat/0501019].

[30] C.S. Fischer, *Infrared properties of QCD from Dyson-Schwinger equations*, *J. Phys.* **G 32** (2006) R253 [hep-ph/0605173].

[31] R. Alkofer, W. Detmold, C.S. Fischer and P. Maris, *Analytic properties of the Landau gauge gluon and quark propagators*, *Phys. Rev.* **D 70** (2004) 014014 [hep-ph/0309077].

[32] T. Blum, *Lattice calculation of the lowest order hadronic contribution to the muon anomalous magnetic moment*, *Phys. Rev. Lett.* **91** (2003) 052001 [hep-lat/0212018].

[33] D. Grunewald, E.M. Ilgenfritz, E.V. Prokhvatilov and H.J. Pirner, *Formulating light cone QCD on the lattice*, *Phys. Rev.* **D 77** (2008) 014512 [arXiv:0711.0620].

[34] J.C. Collins and D.E. Soper, *Parton distribution and decay functions*, *Nucl. Phys.* **B 194** (1982) 445.

[35] M.A. Kimber, A.D. Martin and M.G. Ryskin, *Unintegrated parton distributions*, *Phys. Rev.* **D 63** (2001) 114027 [hep-ph/0101348].

[36] A.D. Martin, R.G. Roberts, W.J. Stirling and R.S. Thorne, *NNLO global parton analysis*, *Phys. Lett.* **B 531** (2002) 216 [hep-ph/0201127].

[37] A. Accardi and W. Melnitchouk, in preparation.

[38] J.F. Owens et al., *The Impact of new neutrino DIS and Drell-Yan data on large-x parton distributions*, *Phys. Rev.* **D 75** (2007) 054030 [hep-ph/0702159].

[39] A. Accardi, J.W. Qiu and J. P. Vary, in preparation.

[40] J. Blumlein and A. Tkabladze, *Target mass corrections for polarized structure functions and new sum rules*, *Nucl. Phys.* **B 553** (1999) 427 [hep-ph/9812478].

JHEP07(2008)090