PHYSICAL REVIEW D **91,** 074020 (2015)

Understanding the large-distance behavior of transverse-momentum-dependent parton densities and the Collins-Soper evolution kernel

John Collins[*]

104 Davey Lab., Penn State University, University Park, Pennsylvania 16802, USA

Ted Rogers[†]

C.N. Yang Institute for Theoretical Physics, Stony Brook University, Stony Brook, New York 11794, USA; Department of Physics, Southern Methodist University, Dallas, Texas 75275, USA; Department of Physics, Old Dominion University, Norfolk, Virginia 23529, USA and Theory Center, Jefferson Lab, 12000 Jefferson Avenue, Newport News, Virginia 23606, USA

(Received 23 January 2015; published 15 April 2015)

There is considerable controversy about the size and importance of nonperturbative contributions to the evolution of transverse-momentum-dependent (TMD) parton distribution functions. Standard fits to relatively high-energy Drell–Yan data give evolution that when taken to lower Q is too rapid to be consistent with recent data in semi-inclusive deeply inelastic scattering. Some authors provide very different forms for TMD evolution, even arguing that nonperturbative contributions at large transverse distance b_T are not needed or are irrelevant. Here, we systematically analyze the issues, both perturbative and nonperturbative. We make a motivated proposal for the parametrization of the nonperturbative part of the TMD evolution kernel that could give consistency: with the variety of apparently conflicting data, with theoretical perturbative calculations where they are applicable, and with general theoretical nonperturbative constraints on correlation functions at large distances. We propose and use a scheme- and scale-independent function $A(b_T)$ that gives a tool to compare and diagnose different proposals for TMD evolution. We also advocate for phenomenological studies of $A(b_T)$ as a probe of TMD evolution. The results are important generally for applications of TMD factorization. In particular, they are important to making predictions for proposed polarized Drell–Yan experiments to measure the Sivers function.

DOI: 10.1103/PhysRevD.91.074020 PACS numbers: 12.38.-t, 12.38.Bx, 12.38.Cy, 12.38.Aw

I. INTRODUCTION

Factorization involving transverse-momentum-dependent (TMD) parton densities and fragmentation functions is important to understanding a variety of hard-scattering reactions in QCD. The domain of utility of such "TMD factorization" is where there is a relevant measured transverse momentum, q_T, that is much less than the large scale Q of the hard scattering.

Although there has been substantial success (e.g., Refs. [1–5]) in fitting data with TMD factorization in the standard framework of Collins, Soper, and Sterman (CSS) [6–9], a number of recent papers—e.g., Refs. [10–13]—disagree with the way this has been done, as we will explain in more detail in Sec. V. There appear to be substantially different predictions for lower-energy experiments, and there even appear to be inconsistencies in how the phenomenology of TMD factorization is to be implemented.

These problems particularly impact proposals for experiments [14–16] for the Drell–Yan process with a transversely polarized hadron. The proposed experiments are designed (among other things) to test the predicted reversal of sign [17] of the Sivers function [18,19] between semi-inclusive deeply inelastic scattering (SIDIS) and the Drell–Yan process. Predictions for polarized Drell–Yan use the Sivers functions measured in SIDIS at relatively low Q (e.g., $Q \sim \sqrt{2.4}$ GeV at HERMES [20,21]). However, to make the predictions, one needs to use correct evolution of the TMD functions to the higher Q of the polarized Drell–Yan experiments. The same applies when one wants to analyze data on the Collins function both from the HERMES and COMPASS experiments and from e^+e^- annihilation from the Belle [22,23] and *BABAR* [24] experiments.

The main issue concerns the evolution of the TMD parton densities (from which arises a dilution of the fractional Sivers asymmetry and other similar quantities). In the CSS framework, the evolution is governed by the function called[1] $\tilde{K}(b_T)$ [6–9], with $\boldsymbol{b}_T$ being a transverse position variable. The results of standard fits [1,2] of the nonperturbative part of $\tilde{K}$ to unpolarized Drell–Yan data at $Q \gtrsim 9$ GeV give what Sun and Yuan [12,13] argue to be a too rapid evolution of TMD functions at the lower values of Q relevant for currently available measurements of the Sivers function.

[*]jcc8@psu.edu
[†]tedconantrogers@gmail.com

[1]Essentially the same function is notated $-2D$ by Echevarría *et al.* [25] and $-F_{q\bar{q}}$ by Becher and Neubert [10].

However, Aybat *et al.* [26] find that the rather rapid evolution given by the fit of Brock-Landry-Nadolsky-Yuan (BLNY) [1] does agree with the change in the Sivers asymmetry between the values of Q for the HERMES [20] and preliminary COMPASS [27] data. More recently, Aidala *et al.* [28] make a very detailed examination of the HERMES and COMPASS data. They agree that there is indeed a discrepancy between these data and the predictions based on the earlier Drell–Yan data but argue that the low- and high-energy data are mostly probing different regions of transverse position; thus, the discrepancy concerns the extrapolation of a parametrization of nonperturbative physics outside the region where it was fitted.

Another issue is that there is more than one formulation of TMD factorization and that these do not always appear to be compatible. There is the original CSS formulation in Refs. [6–8] and its new version in Ref. [9]. There is the related but different formulation by Ji, Ma, and Yuan [29]. There are at least three not obviously identical formulations in the framework of Soft-Collinear Effective Theory (SCET): by Becher and Neubert [10]; by Echevarría, Idilbi, and Scimemi [25]; and by Mantry and Petriello [30,31]. Furthermore, although Sun and Yuan [12,13] explicitly base their work on the CSS method, the actual formula they used to fit data differs substantially from those used in the original fits to Drell–Yan data. Sun and Yuan's formula represents a certain approximation that is essentially identical to one given by Boer in Ref. [32], Eq. (144). The different formulas give rather different results, as seen in a recent paper by Boer [33].

To understand the origin of the proliferation in methods, it is helpful to note that there is often a clash of motivations for applying TMD factorization in specific phenomenological situations.

One increasingly prominent motivation is to study the intrinsic transverse motion related to nonperturbative binding and hadronic physics. For this it is desirable to have a cross section where the effects of intrinsic transverse momentum do not become washed out by perturbative gluon radiation. For these applications Q is taken to be large enough for an overall TMD factorization theorem to be valid, while also being small enough that perturbative radiation does not greatly obscure interesting nonperturbative dynamics associated with the hadron wave function.[2] The measurement of TMD parton densities and fragmentation functions then gives a useful probe of nonperturbative transverse momentum dynamics inside a hadron. Examples of recent work where this is the driving motivation include Refs. [34–40].

By contrast, at high-energy hadron colliders, one is often interested in studying purely perturbative phenomena in cross sections with large transverse momentum or in utilizing perturbative QCD calculations as part of searches for new physics beyond the Standard Model. TMD factorization contains a useful way of resumming large logarithms in perturbative calculations. Examples of recent work where this second motivation is the primary focus include Refs. [10,41–44]. In these applications, sensitivity to nonperturbative transverse momentum phenomena could understandably be seen as a nuisance and a source of uncertainty, while these same phenomena are the main objects of study in the situation described in the previous paragraph.

Our perspective is that the two motivations described above are best seen as complementary rather than conflicting. Both the perturbative calculability of the small b_T dependence and the universality of the large-b_T dependence are important features of the TMD factorization theorem, and predictive power is optimized when both are fully exploited. Verifying the universality of the large-b_T behavior can provide valuable insight into aspects of the nonperturbative dynamics relevant to studies of hadron structure. Moreover, even for measurements at relatively large Q, where sensitivity to the nonperturbative region of b_T is expected to be small, nonperturbative components are potentially necessary if very high-precision calculations are desired.[3] Thus, the TMD factorization theorem, with its accommodation of both types of behavior, simultaneously addresses both of the motivations described above within a single formalism.

One of the issues that now becomes prominent is the nature of the evolution of the TMD functions between different energies. Notably, there is disagreement on whether there are nonperturbative contributions to TMD evolution and how significant they are.

A nonperturbative contribution to the evolution might appear troublesome at first glance. Nonperturbative inputs in perturbative calculations are sources of uncertainty. However, the Collins-Soper kernel, including its nonperturbative components, has what we call "strong" universality. Namely, it is process independent, but it is also insensitive to the types of external hadrons involved, any polarization dependence, the flavors of the quarks, and the scale Q. [The only dependence is on the color representation of the colliding parton—there is, in principle, a different nonperturbative evolution for quark and gluon TMD parton distribution functions (pdfs).] In this sense, it has a much greater degree of universality than, say, collinear pdfs, which, while independent of the process, are dependent upon the nature and identity of the parent hadron, including the polarization, and on the flavor and polarization of the quark. Once the nonperturbative TMD *evolution* is known, it can be used in all situations where a TMD factorization theorem is valid. Verifying this strongly universal nature of the

[2]A standard example is the Sivers function, where associated with intrinsic nonperturbative transverse momentum is a transverse single-spin asymmetry with a process-dependent sign.

[3]One example is the W mass where the large-b_T behavior is an important source of uncertainty in precision experimental constraints. See, for example, p. 57 of Ref. [45]. See also a discussion of related issues in Ref. [46].

nonperturbative evolution is an important test of the TMD-factorization theorem itself. Moreover, its relationship to matrix elements of the fundamental field operators is known —it is essentially a nonperturbative contribution to the vacuum expectation value of a certain kind of Wilson loop. [This follows immediately from the definitions of the TMD pdfs; see, for example, Ref. [9], Eqs. (13.47) and (13.48).] Therefore, determining its influence on cross sections provides a probe of fundamental nonperturbative quark-gluon dynamics, which can then be compared with nonperturbative approaches to understanding QCD matrix elements. Such nonperturbative methods could include, for example, lattice-based calculations [47–49].

Now, the TMD functions are normally used in their Fourier transformed versions, in terms of a transverse position variable $\mathbf{b}_T$. CSS took it as obvious that a large enough value of b_T is in the region of nonperturbative physics in QCD. They therefore argued that one must provide some kind of cutoff on perturbative calculations and insert a parametrization, to be fit to data, to handle nonperturbative contributions at large b_T. Such fits were done in Refs. [1,2]. But other papers have either avoided [10,12,13] the use of a nonperturbative contribution to the function $\tilde{K}(b_T)$ that controls evolution or have argued [11] that the nonperturbative contribution is not needed until well beyond the region of b_T that was regarded as nonperturbative in the earlier fits. Furthermore, Kulesza, Sterman, and Vogelsang [50] have presented a method in which the Landau pole manifested in perturbative calculations in b_T-space is avoided by a contour deformation, and the nonperturbative behavior associated with large b_T is inferred from power law behavior obtained in perturbation theory and extrapolated to large b_T. In Refs. [51,52], the sizes of power corrections were estimated on the basis of renormalon studies.

In this paper, we give a detailed analysis of the issues and propose methods and solutions for resolving the disagreements.

First, we survey a sample of the different ways solutions to the evolution equations can be written. This will illustrate how different but essentially equivalent styles of presentation may emphasize particular specific features while preserving the broad, underlying predictive power of the TMD factorization formalism. For instance, we will note the connection between a TMD parton model language and the standard presentation of the results of the CSS formalism.

Then, as a diagnostic tool, we define a master function[4] $A(b_T)$ that can be used to quantify the evolution of the shape of TMD functions, separately from the normalization. This function is scheme- and renormalization-group-independent, and the function is predicted by QCD to be independent of all other kinematic variables (e.g., Q) and of parton and hadron flavor. Because A is scheme independent, a comparison of A between different formalisms and approximations can be used as a diagnostic to determine where in b_T the methods disagree. This then indicates which experiments would give sensitivity to the differences. Since A is a scheme-independent property of QCD and is a function of only one variable, we propose that calculations and fits for TMD evolution can be presented as including a determination of A in particular regions of b_T and that it would be useful to obtain a global fit for A. The result would be a universal function that controls the evolution with energy of the shape of all the many TMD functions (for color triplet quarks). As a first step in such an analysis, we use the master function A to compare evolution in various formalisms and from different analyses of data.

Finally, we examine phenomenological and theoretical issues concerning the nonperturbative large-b_T behavior of the TMD pdfs and of $\tilde{K}$. We will argue that the standard Gaussian parametrizations of the TMD functions give the wrong limiting behavior at asymptotically large b_T. Instead, we agree with Schweitzer *et al.* [53] that TMD parton densities and fragmentation functions should decay exponentially at large b_T, with a decay length corresponding to the mass of the lowest relevant state.

As regards the CSS evolution kernel $\tilde{K}$, the same argument suggests that it goes to a constant at large b_T (and hence that our master function A goes to zero). This is also suggested phenomenologically by the apparent slowness of TMD evolution at low Q, where phenomena are dominated by the nonperturbative range of large b_T. We will suggest possible parametrizations that can reconcile our proposed large-b_T asymptote with previous fits that use a quadratic b_T^2 dependence for $\tilde{K}$. These quadratic parametrizations correspond to a Q-dependent width for the standard Gaussian ansatz for TMD functions. The quadratic form for $\tilde{K}$ (and hence A) can only be valid over a limited range of moderately large b_T. One should not continue the b_T^2 form to the larger values of b_T that are important for processes at low Q. The result is then that the evolution of TMD pdfs is much weaker at low Q than would otherwise happen.

A related proposal for the nonperturbative form was given by Collins and Soper [54]. Their form was logarithmic instead of quadratic, and the particular formula was designed to provide a better match to the perturbative part of $\tilde{K}$ in the extrapolation to large b_T.

A number of the ideas in this paper have been discussed and presented at various conferences and workshops, so they are already becoming current (e.g., Ref. [55]). The particular aim of the present paper is to give a detailed, systematic, and unified account of the issues.

[4]The function in fact equals the one that CSS [8] called A. But we now give a more general definition, not tied to a particular version of the CSS formalism in the perturbative region. In the CSS formalism, our definition gives $A = -\partial\tilde{K}/\partial \ln b_T^2$. Contrary to CSS, we now treat A as a function of b_T instead of $\alpha_s(1/b_T)$, as is more appropriate when examining its behavior beyond perturbation theory.

The outline of this paper is as follows. First, in Sec. II, we review the CSS formalism for TMD factorization in the form given in Ref. [9], together with an analysis of various forms of the solution of the evolution equations. In Sec. III, we review the universality properties of the functions in the formalism. Then in Sec. IV, we motivate and define the master function A. In Sec. V, we assess various approximations for $\tilde{K}$ and show the phenomenological difficulties referred to earlier. In Sec. VII, we propose new forms for parametrizing the large-b_T region.

II. CSS FORMALISM

In this section, we review TMD factorization for the Drell–Yan cross section in the version of the CSS formalism recently derived in Ref. [9], in which can be found justifications of assertions in this section that are otherwise unreferenced. (The results presented are for the case that all quarks are light. In the presence of heavy quarks, as in real QCD, the modifications have not been completely worked out, to the best of our knowledge. See, however, Ref. [56] for work in this direction in the specific case of heavy quark production.)

It should be emphasized that the CSS formalism in Ref. [9] is, in fact, equivalent to the earlier formalism of Refs. [6–8], but with a different organization of the factors that is intended to be more convenient and with improved operator definitions of the TMD functions. This implies, in particular, that high-order calculations for quantities like anomalous dimensions performed for the original version of the CSS formalism can be carried over to the new version, after allowing for an effective scheme change. Thus, use of the updated formalisms does not necessarily imply an increased degree of complexity, relative to previous work, for doing calculations.

Results of the same structure apply to other processes of interest, e.g., SIDIS [57–59] and inclusive production of two hadrons in e^+e^- annihilation [6]. It will be sufficient to use Drell–Yan scattering as our main example.

Much of the presentation in this section follows previous treatments. The reader should be warned that some steps may require familiarity with more complete derivations such as can be found in Refs. [6–9]. We will emphasize aspects that are particularly relevant to comparison with data, to the predictive power of the formalism, and to the comparison with other formalisms and approximations. In particular, we will give a careful account of the nature of quantities that receive nonperturbative contributions, notably the function $\tilde{K}$ that controls the evolution of TMD densities.

A. Fundamental equations

The Drell–Yan process is the production of a high-mass lepton-pair in a high-energy collision of two hadrons, $A + B \to l + \bar{l} + X$, with the restriction to production of the lepton pair through a virtual photon and/or Z boson[5] in the lowest order in electroweak interactions.

Kinematic variables are defined as follows: The momenta of the incoming hadrons are p_A and p_B, the momentum of the lepton pair is q, polar angles θ and ϕ in the Collins–Soper frame [60] are used to specify the directions of the lepton momenta, and the element of the solid angle for the lepton direction is $d\Omega$. Certain auxiliary variables are convenient for the factorization formalism. To define these, we use light-front coordinates in the overall center-of-mass frame, with A and B moving in the $+z$ and $-z$ directions. Then $q^\pm = (q^0 \pm q^3)/\sqrt{2}$, and we let the mass, rapidity, and transverse momentum of the lepton pair be $Q = \sqrt{q^2}$, $y = \frac{1}{2}\ln(q^+/q^-)$, and $\mathbf{q}_T$. Longitudinal momentum fractions are defined by $x_A = Qe^y/\sqrt{s}$ and $x_B = Qe^{-y}/\sqrt{s}$, where $\sqrt{s}$ is the overall center-of-mass energy. Factorization applies up to power suppressed corrections when s and Q are made large with fixed x_A and x_B.

1. Factorization

The factorization formula is

$$\begin{aligned}\frac{d\sigma}{d^4q\,d\Omega} = \frac{2}{s}\sum_j \frac{d\hat{\sigma}_{j\bar{j}}(Q,\mu,\alpha_s(\mu))}{d\Omega} \\ \times \int d^2\boldsymbol{b}_T e^{i\boldsymbol{q}_T\cdot\boldsymbol{b}_T} \tilde{f}_{j/A}(x_A,\boldsymbol{b}_T;\zeta_A,\mu) \\ \times \tilde{f}_{\bar{j}/B}(x_B,\boldsymbol{b}_T;\zeta_B,\mu) \\ + \text{polarization terms} + \text{high-}q_T\text{ term }(Y) \\ + \text{power-suppressed.} \qquad (1)\end{aligned}$$

Here, $\tilde{f}_{j/H}(x,\boldsymbol{b}_T;\zeta,\mu)$ is the TMD density of a quark of flavor j in hadron H, but Fourier transformed to transverse coordinate space. A suitable definition is given in Ref. [9], Sec. 13.15, with the change of direction of the Wilson lines appropriate to the Drell–Yan process [9], Sec. 14.15. The sum over j is over all flavors of quark and antiquark. The hard-scattering coefficient is $d\hat{\sigma}_{j\bar{j}}(Q,\mu,\alpha_s(\mu))/d\Omega$, normalized like a cross section, for production of the lepton pair from a quark-antiquark collision, with renormalization scale μ. When the renormalization scale is taken to be of order Q, the hard scattering is perturbatively calculable because of the asymptotic freedom of QCD. The parton densities have two scale arguments: One is the renormalization scale μ, and the other is a scale ζ that is used to parametrize how the effects of soft-gluon radiation are partitioned between the two TMD parton densities. The two ζ scales obey $\zeta_A\zeta_B = Q^4$.

[5]The modifications for other related processes, e.g., W production are, of course, elementary.

It is convenient to define a contribution to the $\boldsymbol{b}_{\mathrm{T}}$-space integrand by

$$\tilde{W}_j(\boldsymbol{b}_{\mathrm{T}};Q) \equiv Q^2 \frac{\mathrm{d}\hat{\sigma}_{j\bar{j}}(Q,\mu,\alpha_s(\mu))}{\mathrm{d}\Omega} \tilde{f}_{j/A}(x_A,\boldsymbol{b}_{\mathrm{T}};\zeta_A,\mu) \times \tilde{f}_{\bar{j}/B}(x_B,\boldsymbol{b}_{\mathrm{T}};\zeta_B,\mu). \quad (2)$$

Note that this quantity includes the hard part and that the normalization differs from that of a similar quantity in Ref. [8]. The overall factor of Q^2 in Eq. (2) is to make $\tilde{W}_j$ dimensionless. It is also convenient to define a summed $\tilde{W}$ of which the Fourier transform corresponds to the cross section itself:

$$\tilde{W} = \sum_j \tilde{W}_j. \quad (3)$$

Then the first line of Eq. (1) is

$$\frac{2}{sQ^2} \sum_j \int \mathrm{d}^2\boldsymbol{b}_{\mathrm{T}} e^{i\boldsymbol{q}_{\mathrm{T}}\cdot\boldsymbol{b}_{\mathrm{T}}} \tilde{W}_j(\boldsymbol{b}_{\mathrm{T}};Q), \quad (4)$$

i.e.,

$$\frac{2}{sQ^2} \int \mathrm{d}^2\boldsymbol{b}_{\mathrm{T}} e^{i\boldsymbol{q}_{\mathrm{T}}\cdot\boldsymbol{b}_{\mathrm{T}}} \tilde{W}(\boldsymbol{b}_{\mathrm{T}};Q). \quad (5)$$

The hard scattering is computed from graphs for the massless on-shell quark-antiquark-to-lepton-pair reaction with subtractions for collinear and soft regions appropriate to the definition of the TMD parton densities that are used. The leptons' angular distribution is computed by using the valid approximation that in the hard scattering the incoming quark j and antiquark $\bar{\jmath}$ move in exactly the $+z$ and $-z$ directions in the Collins–Soper frame.

The cross section, with its angular distribution, can be expressed in terms of a hadronic tensor $W^{\mu\nu}(q,p_A,p_B)$ and corresponding scalar structure functions (W_1, etc.) in the standard way [61–64]: Given that only annihilation through an electroweak boson is involved, the cross section is written in terms of the product of $W^{\mu\nu}$ and a lowest-order leptonic factor. The hadronic tensor is decomposed in terms of standard basis tensors times scalar structure functions. There is a corresponding decomposition of the cross section in terms of basis functions for the dependence on the lepton polar angles, including spin dependence [65–67]. By matching the basis function expansion for the full cross section and the hard-scattering cross section (or some equivalent method), one obtains from Eq. (1) corresponding TMD factorization formulas for the structure functions.

The first line in Eq. (1) gives the cross section when $q_{\mathrm{T}} \ll Q$, the hadrons are unpolarized, and quark polarization is ignored. In that case, the TMD parton densities are independent of the azimuthal angle of $\boldsymbol{b}_{\mathrm{T}}$. When hadron and quark polarization are taken into account, further similar terms arise; these can be characterized in terms of suitable polarization-dependent TMD densities [68,69]. For example, there is the Sivers function [18,19], which gives the azimuthal dependence of the TMD density of unpolarized quarks in a transversely polarized proton.

Treatment of the polarization-dependent terms involves a mostly straightforward generalization of the CSS method; see, for example, Refs. [33,70,71].

A further addition to the formula is present because the first part of Eq. (1), including the polarization-dependent terms, gives a valid approximation to the cross section only when $q_{\mathrm{T}} \ll Q$. At large q_{T}, ordinary collinear factorization is valid. Therefore, a correction term is added so that the total is correct to the leading power of Q for any value of q_{T}; the correction term is called Y by CSS [6,8].

2. Summary of subsidiary results

Most of the predictive power of TMD factorization comes not from the factorization formula (1) alone but from its combination with further results, as follows.

First, there are evolution equations [Sec. II A 3, Eqs. (6)–(9)] for the ζ and μ dependence of the factors. These equations enable the parton densities to be written in terms of parton densities at fixed scales. The parton densities are universal across processes.[6]

The second source of predictive power is from perturbative calculations of the kernels of the evolution equations. These include anomalous dimensions and the universal function $\tilde{K}(b_{\mathrm{T}},\mu)$ that controls ζ dependence. With the aid of a renormalization-group transformation, $\tilde{K}(b_{\mathrm{T}},\mu)$ can be calculated perturbatively when b_{T} is not too large. [See Sec. II A 3, Eqs. (10)–(12) below.]

The universality of the various nonperturbative functions, as summarized in Sec. III below, gives further predictions. This especially concerns the function $\tilde{K}(b_{\mathrm{T}},\mu)$, which is nonperturbative for large b_{T}.

Another source of predictions is that, after evolution is applied to set ζ and μ^2 to be of order Q^2, the hard scattering, $\mathrm{d}\hat{\sigma}_{j\bar{j}}(Q,\mu,\alpha_s(\mu))/\mathrm{d}\Omega$, is perturbative; that is, it can be expanded in powers of the effective coupling $\alpha_s(Q)$ at a high scale, without the logarithmic enhancements of coefficients that would otherwise occur.

Finally, there is a kind of operator product expansion (OPE) for the TMD parton densities at small b_{T}. (See Sec. II A 4.) In a theory, such as a super-renormalizable nongauge theory, where the elementary parton model is valid, the value of a coordinate-space TMD parton density at zero b_{T} equals the integral over all transverse momenta of the momentum space TMD function, by elementary properties of Fourier transforms. The result is also the corresponding integrated parton density. See Ref. [9], Sec. 6.8 for details.

[6]Apart from the predicted sign reversal of T-odd functions, like the Sivers function, between Drell–Yan and SIDIS.

But in renormalizable theories and especially QCD, there is a strong enough singularity at $b_T \to 0$ that such results must be modified (e.g., [9], Chap. 13) and the appropriate modification is the OPE at small b_T. This enables the TMD functions at small b_T to be expressed in terms of ordinary integrated parton densities and perturbatively calculable coefficient functions. The coefficient functions in this OPE are currently known[7] to order α_s^2 [73,74]. Intuitively, the OPE can be characterized by saying that, when a momentum-space TMD density is integrated over transverse momenta up to order Q, the result is the integrated parton density at scale Q plus perturbative corrections of order $\alpha_s(Q)$.

Naturally, Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) evolution also enters here, so that the OPE plus DGLAP evolution gives the TMD parton densities at small b_T in terms of ordinary parton densities at a fixed scale.

It should be added that, even without any of these subsidiary results, the factorization in Eq. (2) alone provides predictions, since, as regards the dependence on the longitudinal momentum fraction parameters x_A and x_B, the cross section is a function jointly of both variables. But each parton density depends only on one of these variables.

3. Evolution equations

The CSS evolution equation for the ζ dependence of the TMD parton densities is

$$\frac{\partial \ln \tilde{f}_{f/H}(x, b_T; \zeta; \mu)}{\partial \ln \sqrt{\zeta}} = \tilde{K}(b_T; \mu). \tag{6}$$

The kernel $\tilde{K}$ is independent of the flavor and spin of the quark, of the nature of the hadron target, and of the momentum fraction x. It is also the same for fragmentation functions as well as parton densities and is the same between the versions of parton densities for Drell–Yan and the SIDIS processes, and for all the different polarized parton densities. A different kernel does appear in gluon densities, since $\tilde{K}$ depends on the color representation carried by the parton. Note that both the parton densities and the kernel $\tilde{K}$ have contributions from the infrared or long-distance domain,[8] and hence these functions depend on quark masses, as well as on the coupling $\alpha_s(\mu)$, but we have not indicated this dependence explicitly.

The renormalization-group (RG) equation for the kernel is

$$\frac{d\tilde{K}(b_T; \mu)}{d \ln \mu} = -\gamma_K(\alpha_s(\mu)), \tag{7}$$

and for the parton densities, the RG equation is

$$\frac{d \ln \tilde{f}_{j/H}(x, b_T; \zeta; \mu)}{d \ln \mu} = \gamma_j(\alpha_s(\mu); 1) - \frac{1}{2}\gamma_K(\alpha_s(\mu)) \ln \frac{\zeta}{\mu^2}, \tag{8}$$

in the notation of Ref. [9]. The RG coefficient γ_j is specific to quark j. However, the relevant calculations are the same for all flavors of spin-$\frac{1}{2}$ quark. The ζ dependence on the right-hand side is determined from the fact that differentiation of a parton density with respect to μ commutes with differentiation with respect to ζ. [An alternative notation for the whole of the right-hand side of (8) is $\gamma_j(\alpha_s(\mu); \zeta/\mu^2)$.]

An RG equation for the hard scattering follows from the RG invariance of physical cross sections:

$$\frac{d}{d \ln \mu} \ln \left[\frac{d\hat{\sigma}_{j\bar{j}}(Q, \mu, \alpha_s(\mu))}{d\Omega} \right] = -2\gamma_j(\alpha_s(\mu); 1) + \gamma_K(\alpha_s(\mu)) \ln \frac{Q^2}{\mu^2}. \tag{9}$$

In our calculations, we will need the one-loop values for the above quantities and the two-loop value of γ_K[9]:

$$\gamma_K(\alpha_s(\mu)) = 2C_F \frac{\alpha_s(\mu)}{\pi} + \left(\frac{\alpha_s(\mu)}{\pi}\right)^2 C_F \left[C_A \left(\frac{67}{18} - \frac{\pi^2}{6} \right) - \frac{10}{9} T_F n_f \right] + O(\alpha_s(\mu)^3), \tag{10}$$

$$\tilde{K}(b_T; \mu) = -\frac{\alpha_s(\mu)}{\pi} C_F \left[\ln \frac{b_T^2 \mu^2}{4} + 2\gamma_E \right] + O(\alpha_s(\mu)^2), \tag{11}$$

$$\gamma_j(\alpha_s(\mu); 1) = \frac{3C_F}{2} \frac{\alpha_s(\mu)}{\pi} + O(\alpha_s(\mu)^2). \tag{12}$$

4. Small-b_T expansion

At small b_T, the unpolarized TMD parton densities can be expressed in terms of the corresponding integrated parton densities, $f_{k/H}(x; \mu)$, by a kind of OPE:

[7] Catani *et al.* [72] also give the results for a number of high-order calculations for a version of CSS resummation. Since there may be a scheme change compared with the formalism that we and the authors of Refs. [73,74] used, it remains to check consistency of the different calculations.

[8] For $\tilde{K}$, infrared contributions are power suppressed at small b_T but not at large b_T.

[9] See Ref. [9] for one-loop calculations of γ_K from its definition. The value to three-loop order was found by Moch, Vermaseren, and Vogt [75]; they computed a quantity they call A, which is our $\gamma_K/2$—see their Eq. (2.4). Their value was recently confirmed by Grozin *et al.* [76].

$$\tilde{f}_{j/H}(x, b_T; \zeta; \mu) = \sum_k \int_{x-}^{1+} \frac{d\xi}{\xi} \tilde{C}_{j/k}(x/\xi, b_T; \zeta, \mu, \alpha_s(\mu)) f_{k/H}(\xi; \mu) + O[(m b_T)^p]. \quad (13)$$

Here, the sum is over all flavors k of parton: quarks, antiquarks, and gluons. When b_T is small, the coefficient functions, $\tilde{C}_{j/k}$, can be usefully expanded in perturbation theory, provided that $\sqrt{\zeta}$ and μ are comparable to $1/b_T$, so that large logarithms involving these parameters are not present. Corrections to the OPE are suppressed by a power of b_T, as indicated by the last term in (13).

The lowest-order coefficient is effectively unity:

$$\tilde{C}_{j/k}(x/\xi, b_T; \zeta, \mu, \alpha_s(\mu)) = \delta_{jk}\delta(\xi/x - 1) + O(\alpha_s). \quad (14)$$

See Refs. [73,74] for the coefficient functions to order α_s^2.

An OPE of the same form as (13) applies also to the helicity and transversity densities (but generally with coefficient functions that differ beyond lowest order). The proofs work the same way. In each case, one can characterize the OPE in the same way as in last part of Sec. II A 2: It formulates the QCD corrections to the parton-model idea that the integral of a TMD density over all transverse momenta is the corresponding integrated density.

As for the other (polarization-dependent) TMD densities, like the Sivers function, generalizations of Eq. (13) apply—see e.g., Ref. [77], Eq. (9). They relate these other TMD densities to more involved quantities associated with matrix elements of higher-twist operators (e.g., the Qiu–Sterman function [78,79]). As such, they are less useful because, if nothing else, the Qiu–Sterman function is much less well measured than conventional unpolarized parton densities. Thus, it will often be useful not to apply an OPE to the Sivers function, etc.

B. Solutions

The evolution equations can be used to reformulate the factorization formula in such a way that:

(i) Universality properties are exhibited. In particular, functions with nonperturbative content, like parton densities, are at a single fixed scale.
(ii) Perturbatively calculated quantities have no large logarithms in their expansions in powers of α_s.

We present useful solutions in several forms. The different forms can be used to emphasize different aspects of the physics.

As a starting point, we can use the original factorization formula (1) with ζ set to Q^2, and μ proportional to Q:

$$\frac{d\sigma}{d^4q\, d\Omega} = \frac{2}{s}\sum_j \frac{d\hat{\sigma}_{jj}(Q, \mu_Q, \alpha_s(\mu_Q))}{d\Omega} \int d^2\boldsymbol{b}_T e^{i\boldsymbol{q}_T\cdot\boldsymbol{b}_T} \times \tilde{f}_{j/A}(x_A, \boldsymbol{b}_T; Q^2, \mu_Q)\tilde{f}_{\bar{j}/B}(x_B, \boldsymbol{b}_T; Q^2, \mu_Q) + \text{polarization terms} + \text{high-}q_T\text{ term }(Y) + \text{power-suppressed}. \quad (15)$$

Here, $\mu_Q = C_2 Q$, and, as usual, the constant C_2 can be chosen with the aim of optimizing the accuracy of perturbation theory for $d\hat{\sigma}$. This formula exhibits a parton-model form with a perturbatively calculable hard scattering, while characteristic QCD effects are hidden inside the Q dependence of the parton densities.

A corresponding equation to (15) is used in ordinary collinear factorization; there, one has a perturbatively calculable hard scattering convoluted with parton densities at a scale proportional to Q. Then one uses DGLAP evolution to express the Q-dependent collinear parton densities in terms of those at a fixed reference scale. We apply the same strategy to TMD factorization. Differences in the implementation arise from three sources: (1) DGLAP evolution results in a complicated convolution to relate collinear parton densities at different scales; this convolution is hidden inside numerical computer codes for its implementation. For the TMD case, the solutions to the evolution equations are simple enough to be exhibited explicitly. (2) There are two scale arguments ζ and μ in the TMD case; this just reflects two sources of scale dependence as implemented in the technical definition of the TMD parton densities. (3) There is an extra variable $\boldsymbol{b}_T$ in some of the functions, and extra steps in the analysis are used to treat this.

In the rest of this section, we will first evolve the TMD parton densities and $\tilde{K}$ so that factorization is presented in terms of these functions at fixed values of their scale arguments. This displays the universality properties of parton densities and of the CSS kernel $\tilde{K}$ and is especially suitable when these functions are in a nonperturbative region. It also exhibits in a rather direct manner the nature of the evolution with Q and the strong predictions that can be obtained even without the use of perturbation theory to compute TMD parton densities and $\tilde{K}(b_T)$ at small b_T. We will also exhibit a modified form of this solution which more clearly displays the fact that measurements to give an unambiguous determination of $\tilde{K}(b_T)$ require changes of the center-of-mass energy $\sqrt{s}$, not merely changes in Q at fixed $\sqrt{s}$.

Next, we exploit RG transformations to allow perturbative expansions without logarithms, when appropriate, and we arrange for explicitly defined separate functions that contain the nonperturbative parts. The result is a version of a formula given by CSS in Ref. [8]. It utilizes perturbative information both for $\tilde{K}$ and for the OPE for the TMD densities and therefore makes maximum use of the

predictive power of perturbative calculations. But the structure of the formula, (22) below, does not make clear the fact that it uses TMD parton densities in a form close to that of the parton model.

We will also present a modified form of this solution that uses perturbative information only for $\tilde{K}$ and effectively treats the TMD densities as if they are functions to be obtained from data. The reason for this is that it is common that low-energy data have been fitted with TMD parton densities in a pure parton-model formalism [39,80–89]. Normally a Gaussian dependence on b_T is used in such fits; they are intended to take account of nonperturbative contributions to the TMD functions. This is an approximation that ignores the predicted small-b_T dependence in real QCD. It is, of course, important to know how such TMD functions evolve.

The multiplicity of different forms of solution of the evolution equations gives a potentially confusing set of different ways of using TMD factorization phenomenologically. In fact, the different forms of solution correspond to different methods that have appeared in the literature (typically in conjunction with approximations). What should be clear from the order of presentation in this paper is that they are all related to a single unifying factorization framework.

1. Fixed scales

The solution with the parton densities and the CSS kernel at fixed scales is

$$\begin{aligned}\frac{d\sigma}{d^4q\,d\Omega} = {}& \frac{2}{s}\sum_j \frac{d\hat{\sigma}_{j\bar{j}}(Q,\mu_Q,\alpha_s(\mu_Q))}{d\Omega}\int d^2\boldsymbol{b}_T e^{i\boldsymbol{q}_T\cdot\boldsymbol{b}_T}\tilde{f}_{j/A}(x_A,\boldsymbol{b}_T;Q_0^2,\mu_0)\tilde{f}_{\bar{j}/B}(x_B,\boldsymbol{b}_T;Q_0^2,\mu_0)\\ &\times\left(\frac{Q^2}{Q_0^2}\right)^{\tilde{K}(b_T;\mu_0)}\exp\left\{\int_{\mu_0}^{\mu_Q}\frac{d\mu'}{\mu'}\left[2\gamma_j(\alpha_s(\mu');1)-\ln\frac{Q^2}{(\mu')^2}\gamma_K(\alpha_s(\mu'))\right]\right\}\\ &+\text{polarized terms}+\text{large-}q_T\text{ correction},\ Y+\text{p.s.c.}\end{aligned} \tag{16}$$

Here, Q_0 and μ_0 are chosen fixed reference scales. They have exactly the same status as a similar parameter that is used in DGLAP evolution of ordinary parton densities (e.g., Ref. [90]), and that is often denoted by Q_0. In both cases, there is a functional form of parton densities at the fixed reference scale (or scales), and evolution has been used to obtain Q-dependent parton densities, as used in Eq. (15). The solution (16) can be obtained from (15) by first applying the RG equation (8) for TMD densities and then the Collins-Soper equation (6), to express the Q-dependent densities in terms of the densities at the scales used in (16). It will generally be convenient to set $\mu_0 = \mu_{Q_0} = C_2Q_0$. In Eq. (16) (and in similar later equations) "p.s.c." is a shorthand for "power suppressed corrections."

Although the reference scale μ_0 is in principle arbitrary, it should in practice be chosen large enough to be treated as being in the perturbative region. This allows finite-order perturbative calculations of the anomalous dimensions, γ_j and γ_K, to be appropriate for all the values in the integral over μ'. Given the choice $\mu_0 = C_2Q_0$, it will generally be sensible to choose Q_0 to be near the lower end of the range of Q for the data to which one applies factorization. This is exactly the same as with typical implementations of DGLAP evolution.

In the hard scattering, we have preserved $\mu_Q = C_2Q$, so that it has no large logarithms in its perturbative coefficients and can therefore be effectively calculated by low-order perturbation theory in powers of $\alpha_s(\mu_Q)$.

It is convenient to notate the evolution factor on the second line of (16) as $e^{-S(b_T,Q,Q_0,\mu_0)}$, where

$$\begin{aligned}&S(b_T,Q,Q_0,\mu_0)\\ &\quad=-\tilde{K}(b_T;\mu_0)\ln\frac{Q^2}{Q_0^2}+\int_{\mu_0}^{\mu_Q}\frac{d\mu'}{\mu'}\\ &\qquad\times\left[-2\gamma_j(\alpha_s(\mu');1)+\ln\frac{Q^2}{(\mu')^2}\gamma_K(\alpha_s(\mu'))\right].\end{aligned} \tag{17}$$

The two equations (15) and (16) exhibit a very close relationship to a TMD parton model formula. Equation (15) is of a parton model form except that: (a) the hard part has perturbative higher-order corrections, (b) the TMD parton densities are scale dependent, and (c) there is a Y-term. But when the cross section is expressed in terms of TMD parton densities at the reference scales, we find in Eq. (16) a factor e^{-S} that gives the important effects of gluon radiation.

Furthermore, the solution (16) exhibits universality properties of the TMD densities that are the same as in the parton model. That is, under all circumstances where a TMD factorization theorem holds, the same TMD densities, functions of x and $\boldsymbol{b}_T$, are used, up to possible factors of -1 for T-odd functions. Both the perturbatively calculable hard scattering and the factor involving perturbatively calculable anomalous dimensions only affect the

normalization of the cross section, but not its shape as a function of q_T. The remaining factor $(Q^2/Q_0^2)^{\tilde{K}(b_T;\mu_0)} = \exp[\tilde{K}(b_T;\mu_0)\ln(Q^2/Q_0^2)]$ gives a Q-dependent change in the shape of the distribution, a very characteristic effect of gluonic emission in a gauge theory.

As is well known, a minor modification to universality arises because the appropriate TMD parton densities differ between processes of the Drell–Yan type and those of the SIDIS type. The operators in the definition of the TMD densities use oppositely directed Wilson lines in the two cases. Most TMD densities are numerically unchanged, but T-odd densities, like the Sivers function, change sign [17].

It is important to recall that the derivation of Eq. (16) depends on the TMD factorization and evolution equations and that these in turn depend on properties of the particular definitions used for the TMD parton densities; see Ref. [9], Eqs. (13.106) and (13.108) and the discussions leading up to this definition. (These remarks apply equally to the original CSS derivations [6–8].)

2. Measuring CSS evolution

Suppose we temporarily ignored the perturbative information about b_T dependence available at small b_T for the CSS kernel and for the TMD densities. Then one could determine $\tilde{K}$ from a limited set of data with variable Q at fixed x_A and x_B (up to errors associated with the unknown power-suppressed corrections). The evolution of the cross section at every other value of x_A, x_B, and Q would then be determined. The TMD densities can be determined from data from an experiment at one value of s (aside from the issue of flavor dependence that can be best analyzed with data from other processes like SIDIS).

However, if one examines data in a single experiment, i.e., at fixed s, the x dependence of the TMD functions is confounded with the Q dependence of the factor involving $\tilde{K}$, since $Q^2 = x_A x_B s$. To exhibit the fact that $\tilde{K}$ can only be determined by varying s with x_A and x_B fixed, it may therefore be convenient to use CSS evolution to change the choice of the ζ argument of the TMD densities so as to make the corresponding $\tilde{K}$-dependent factor a function of s instead of Q. For this purpose, an appropriate solution is

$$\begin{aligned}\frac{d\sigma}{d^4q\,d\Omega} = \frac{2}{s}\sum_j \frac{d\hat{\sigma}_{j\bar{j}}(Q,\mu_Q,\alpha_s(\mu_Q))}{d\Omega}\int d^2\boldsymbol{b}_T e^{i\boldsymbol{q}_T\cdot\boldsymbol{b}_T}\tilde{f}_{j/A}(x_A,\boldsymbol{b}_T;x_A^2Q_0^2,\mu_0)\tilde{f}_{\bar{j}/B}(x_B,\boldsymbol{b}_T;x_B^2Q_0^2,\mu_0)\\ \times\left(\frac{s}{Q_0^2}\right)^{\tilde{K}(b_T;\mu_0)}\exp\left\{\int_{\mu_0}^{\mu_Q}\frac{d\mu'}{\mu'}\left[2\gamma_j(\alpha_s(\mu');1)-\ln\frac{Q^2}{(\mu')^2}\gamma_K(\alpha_s(\mu'))\right]\right\}\\ +\text{polarized terms}+\text{large-}q_T\text{ correction},\ Y+\text{p.s.c.}\end{aligned} \tag{18}$$

One might choose $Q_0^2 = s_0$, where s_0 is the value of s for some particularly important set of data. Given fixed values of Q_0 and μ_0, the functions $\tilde{f}_{j/H}(x,\boldsymbol{b}_T;x^2Q_0^2,\mu_0)$ are just like differently defined TMD densities. That is, as regards the kinematic variables x and $\boldsymbol{b}_T$, they are functions of the same two variables as the TMD functions in Eq. (16). They have a definite relation to the versions of the densities used in the previous solution.

Which is the most appropriate form to use in practice is not so clear. Furthermore, the simple use of these solutions to fit data ignores two sources of perturbatively accessible information for b_T dependence: the calculation of $\tilde{K}$ and of the coefficient functions in the OPE (13). So in Sec. II B 3, we will give another form of solution, (22) below, that is more suitable for using perturbative calculations in combination with fits in the nonperturbative domain.

However, that solution obscures an important unifying property that is clearly visible in Eqs. (16) and (18). This is, quite simply, the existence of the QCD entities that are the TMD densities and the CSS kernel $\tilde{K}$. When using (22), we can regard the purpose of the associated perturbative calculations and of the fitting as providing useful estimates for the TMD densities and for the function $\tilde{K}$ as they appear in Eqs. (16) and (18). One can imagine the result of a global fit of TMD factorization being presented as tables of the TMD functions and of $\tilde{K}(b_T)$, just as with global fits of integrated parton densities. The fitting process could use the complicated formula (22), but the application of the results to predict cross sections could use a simpler formula, e.g., the original factorization formula (1) or one of the fixed scale formulas (16) and (18). If results of a global fit were presented for evolved Q-dependent TMD densities (as is routine for collinear parton densities), then users would not even have to use (16), (18), or (22); they could just use the simple parton-model-like form (15). (See Ref. [91] for recent efforts in this direction.)

It is important to note that, with the form of solution given in Eqs. (16) and (18), it is dangerous to estimate the kernel $\tilde{K}(b_T;\mu_0)$ by a simple fixed-order perturbative expansion with the renormalization scale set to μ_0. This is because there is an integral over all $\boldsymbol{b}_T$, and inevitably there will be large logarithmic corrections from higher-order terms. Thus, a low-order perturbative expansion of $\tilde{K}(b_T;\mu_0)$ can give inaccurate results, especially if Q/Q_0 differs substantially from unity. Of course, the integral also extends into a clearly nonperturbative region of large b_T.

3. Solution optimized for perturbative calculations

We now present a solution that corresponds to one presented by CSS. The aim is to allow the maximum use of perturbative calculations. First, the small-b_T expansion is applied to the TMD parton densities. Then, to allow the effective use of fixed-order perturbative calculations, the evolution equations are applied so that, in the functions $\tilde{K}$ and $\tilde{C}_{j/f}$, μ and $\sqrt{\zeta}$ are of order $1/b_T$. Thus, when b_T is sufficiently much smaller than $1/\Lambda_{QCD}$, these functions are given by their expansions in powers of a small coupling, $\alpha_s(1/b_T)$, and no large logarithms of μb_T, etc., are present.

Note, however, that perturbatively calculated functions may appear in an exponent—as for $\tilde{K}$ and the anomalous dimensions. Thus, any errors in a perturbative calculation of such a function can be magnified by a large logarithm.

In any case, perturbative calculations are not applicable at large enough values of b_T, which, given our knowledge of QCD, is a nonperturbative region. An indication of where the nonperturbative region is likely to be quantitatively important is given by an analysis by Schweitzer, Strikman, and Weiss [53]. Using a chiral effective theory, they found that there are two relevant nonperturbative distance scales: a chiral scale 0.3 fm $= 1.5$ GeV^{-1} and a confinement scale 1 fm $= 5$ GeV^{-1}. At large b_T, they find that a TMD density behaves like an exponential $e^{-b_T/l}$ times a power of b_T, with l being a characteristic scale. The chiral and confinement scales manifest themselves in the large-b_T dependence of the sea and valence quark densities.

To get maximum predictive information, one should therefore combine the use of perturbative calculations at small b_T with fits to data to measure $\tilde{K}$ and the TMD densities at large b_T. Undoubtedly, fits to data will eventually be supplemented by further constraints from nonperturbative calculations like those of Ref. [53] from chiral models and those of Ref. [48] from lattice gauge theory.

Since the integral in Eq. (1) extends from $b_T = 0$ to $b_T = \infty$, one cannot avoid using parton densities and $\tilde{K}$ in the nonperturbative large-b_T region.[10] Therefore, it is necessary to combine nonperturbative information with perturbative calculations. CSS [8] provided a prescription[11] for doing this; we will call their method the "b_* method," after the name of a variable defined by CSS.

They first defined quantities with a smooth upper cutoff on transverse distance, at a chosen value b_{max}, with the use of the following function of $\boldsymbol{b}_T$:

$$\boldsymbol{b}_* = \frac{\boldsymbol{b}_T}{\sqrt{1 + b_T^2/b_{max}^2}}. \tag{19}$$

What are called the perturbative parts of the TMD densities and of $\tilde{K}$ were defined by replacing $\boldsymbol{b}_T$ by $\boldsymbol{b}_*$. Then, the nonperturbative[12] parts were defined as whatever is left over. This idea is implemented with the aid of functions $g_{j/H}(x, b_T; b_{max})$ and $g_K(b_T; b_{max})$ defined by

$$g_K(b_T; b_{max}) = -\tilde{K}(b_T, \mu) + \tilde{K}(b_*, \mu) \tag{20}$$

and

$$e^{-g_{j/H}(x,\boldsymbol{b}_T;b_{max})} = \frac{\tilde{f}_{j/H}(x, \boldsymbol{b}_T; \zeta, \mu)}{\tilde{f}_{j/H}(x, \boldsymbol{b}_*; \zeta, \mu)} e^{g_K(b_T;b_{max}) \ln(\sqrt{\zeta}/Q_0)}. \tag{21}$$

Here, Q_0 is a chosen reference scale that simply determines how much of the TMD density is in $e^{-g_{j/H}}$ and how much is put into the exponential of g_K times a logarithm that appears in Eq. (21) and in Eq. (22) below. As indicated by our notation, we will choose Q_0 here to have the same value as in Eq. (16). We treat $g_{j/H}$ and g_K as needing to be fit to data.

Both of $g_{j/H}$ and g_K vanish approximately[13] like b_T^2 at small b_T, from their definition, and become significant when b_T approaches b_{max} and beyond.

Both functions are independent of both μ and ζ. This is because there is an exact cancellation in the terms obtained by applying the CSS and RG equations to the quantities on the right of Eqs. (20) and (21). The functions do depend, however, on the choice of the value of b_{max} and on the particular CSS prescription for segregating nonperturbative information. It is the full TMD parton densities and the function $\tilde{K}$ that are independent of b_{max} and of the use of the b_* prescription of CSS.

As regards the possible flavor and x dependence of g_K and $g_{j/H}$, this follows from that of the corresponding parent functions, i.e., $\tilde{K}$ and the TMD parton densities. Since $\tilde{K}$ is independent of quark flavor, hadron flavor, and parton x, so is g_K. But the TMD parton densities can depend on quark and hadron flavor and on x, so the same is true of the $g_{j/H}$ functions.

[10] An important, but separate, practical question is whether or not the integrand in Eq. (1) is large enough in the nonperturbative region for the details of the nonperturbative parametrization to matter in the context of particular calculations.

[11] The CSS prescription is not the only possibility. See Refs. [4,92] for one alternative. In addition Bozzi *et al.* Ref. [41], (17), motivated by Ref. [93], proposed a modification to improve the behavior of the formalism at small b_T. This involves the replacement of b_T^2 by $b_T^2 + 4e^{-2\gamma_E}/Q^2$ in the parton densities and evolution factors in (15), (16), and our other solutions, together with a consequent change in Y, as computed in Ref. [41], App. B. This is probably a generally useful prescription.

[12] "Nonperturbative" is somewhat of a misnomer. If b_{max} were chosen to be excessively small, the values of the nonperturbative parts near $b_T = b_{max}$ could be reliably estimated perturbatively.

[13] As we will see in Sec. VI, the existence of perturbatively controlled logarithmic singular behavior of $\tilde{K}$ and $\tilde{f}$ at small b_T implies that $g_{j/H}$ and g_K are not exactly quadratic at small b_T.

Given these definitions, the evolution equations and the small-$b_{\rm T}$ expansion can be used to write the factorization formula as

$$\begin{aligned}\frac{{\rm d}\sigma}{{\rm d}^4q{\rm d}\Omega} = \frac{2}{s}\sum_{j,j_A,j_B}\frac{{\rm d}\hat{\sigma}_{j\bar{j}}(Q,\mu_Q,\alpha_s(\mu_Q))}{{\rm d}\Omega}\int\frac{{\rm d}^2\boldsymbol{b}_{\rm T}}{(2\pi)^2}e^{i\boldsymbol{q}_{\rm T}\cdot\boldsymbol{b}_{\rm T}}\\ \times e^{-g_{j/A}(x_A,b_{\rm T};b_{\rm max})}\int_{x_A}^1\frac{{\rm d}\hat{x}_A}{\hat{x}_A}f_{j_A/A}(\hat{x}_A;\mu_{b_*})\tilde{C}_{j/j_A}\left(\frac{x_A}{\hat{x}_A},b_*;\mu_{b_*}^2,\mu_{b_*},\alpha_s(\mu_{b_*})\right)\\ \times e^{-g_{\bar{j}/B}(x_B,b_{\rm T};b_{\rm max})}\int_{x_B}^1\frac{{\rm d}\hat{x}_B}{\hat{x}_B}f_{j_B/B}(\hat{x}_B;\mu_{b_*})\tilde{C}_{\bar{j}/j_B}\left(\frac{x_B}{\hat{x}_B},b_*;\mu_{b_*}^2,\mu_{b_*},\alpha_s(\mu_{b_*})\right)\\ \times\left(\frac{Q^2}{Q_0^2}\right)^{-g_K(b_{\rm T};b_{\rm max})}\left(\frac{Q^2}{\mu_{b_*}^2}\right)^{\tilde{K}(b_*;\mu_{b_*})}\exp\left\{\int_{\mu_{b_*}}^{\mu_Q}\frac{{\rm d}\mu'}{\mu'}\left[2\gamma_j(\alpha_s(\mu');1)-\ln\frac{Q^2}{(\mu')^2}\gamma_K(\alpha_s(\mu'))\right]\right\}\\ +\text{polarized terms}+\text{large-}q_{\rm T}\text{ correction},\ Y+\text{p.s.c.}\end{aligned}\tag{22}$$

Here, μ_{b_*} is chosen to allow perturbative calculations of b_*-dependent quantities without large logarithms,

$$\mu_{b_*} = C_1/b_*,\tag{23}$$

where C_1 is a numerical constant typically chosen to be $C_1 = 2e^{-\gamma_{\rm E}}$.

4. Fixed scale densities with perturbative organization of CSS evolution

As mentioned in Sec. II A 4, an OPE of the form of (13) applies not only to the unpolarized TMD densities but also to the helicity and transversity densities, i.e., to those TMD densities that have a corresponding integrated density (of the "twist-2" kind). Thus, the form of solution in the first part of Eq. (22) applies also to the terms with the helicity and transversity densities [94].

But for the remaining terms, e.g., those with a Sivers function, the appropriate OPE is of a different form, involving twist-3 distributions like the Qiu–Sterman function. Therefore, for these terms, it is useful to have a version of Eq. (22) that does not apply the OPE to the TMD densities. Even when the simple OPEs can be applied, it can still be useful to treat the TMD functions at a fixed scale as being measured from data. In this new form of solution, we retain the TMD parton densities as in Eq. (16), but we do RG improvement on $\tilde{K}$ to use its perturbative expansion optimally; we also set $\mu_0 = \mu_{Q_0} = C_2Q_0$ so that we arrange for the evolution factor to be unity at $Q = Q_0$. The result is

$$\begin{aligned}\frac{{\rm d}\sigma}{{\rm d}^4q{\rm d}\Omega} = \frac{2}{s}\sum_{j}\frac{{\rm d}\hat{\sigma}_{j\bar{j}}(Q,\mu_Q,\alpha_s(\mu_Q))}{{\rm d}\Omega}\int{\rm d}^2\boldsymbol{b}_{\rm T}e^{i\boldsymbol{q}_{\rm T}\cdot\boldsymbol{b}_{\rm T}}\tilde{f}_{j/A}(x_A,\boldsymbol{b}_{\rm T};Q_0^2,\mu_{Q_0})\tilde{f}_{\bar{j}/B}(x_B,\boldsymbol{b}_{\rm T};Q_0^2,\mu_{Q_0})\\ \times\exp\left\{\left[-g_K(b_{\rm T};b_{\rm max})+\tilde{K}(b_*;\mu_{b_*})-\int_{\mu_{b_*}}^{\mu_{Q_0}}\frac{{\rm d}\mu'}{\mu'}\gamma_K(\alpha_s(\mu'))\right]\ln\frac{Q^2}{Q_0^2}\right\}\\ \times\exp\left\{\int_{\mu_{Q_0}}^{\mu_Q}\frac{{\rm d}\mu'}{\mu'}\left[2\gamma_j(\alpha_s(\mu');1)-\ln\frac{Q^2}{(\mu')^2}\gamma_K(\alpha_s(\mu'))\right]\right\}\\ +\text{polarized terms}+\text{large-}q_{\rm T}\text{ correction},\ Y+\text{p.s.c.}\end{aligned}\tag{24}$$

The exponentials are unity when $Q = Q_0$. That allows the fitting of parton densities as in the parton model at this scale (with small perturbative corrections from the hard scattering). Then the exponentials show how to do evolution to other energies in terms of perturbative quantities without logarithms and a single nonperturbative function. The above solution is related to a form of solution given in Ref. [28].

In both Eqs. (22) and (24), the strategy in organizing the solution was to arrange that quantities to be calculated perturbatively are used in a region where the coupling is in a perturbative region and that there are no large logarithms in the expansions of these quantities. (The logarithms in ordinary fixed coupling expansions have all moved into the explicit logarithms and into the integrals over μ'.) Thus, the

general size of the errors due to truncation of perturbation expansions can be quantified.

Quantities that are not to be calculated perturbatively in a given form of solution [e.g., $\tilde{f}_{j/A}(x_A, \boldsymbol{b}_T; Q_0^2, \mu_{Q_0})$ in Eq. (24)] are assumed to be fitted or treated using nonperturbative methods.

5. Choice of parameters

There are a number of arbitrary parameters in the solutions (16), (18), (22), and (24), notably $b_{\max}$, μ_0, and Q_0. Their occurrence might appear to reduce the predictive power of the formalism. However, these appearances are misleading.

The trickiest case is that of $b_{\max}$. It is sometimes said that $b_{\max}$ is a parameter to be fitted to data, e.g., Ref. [11], Sec. I. But as can be seen from the definitions and derivation summarized above, this is not the case. Equation (22) is true independently of the choice of $b_{\max}$. When $b_{\max}$ is changed, the so-called nonperturbative functions, g_K and $g_{j/H}$, defined in (20) and (21), change their form. In fits to data, the results should be equivalent, provided that the parametrizations used are flexible enough.

However, if limited fixed parametrizations are used for g_K and $g_{j/H}$, they may work better with one value of $b_{\max}$ than another. Also, if an excessively small value of $b_{\max}$ is used, much of the fitting of the functions will be devoted to recovering their dependence on b_T in a region where perturbative calculations would be adequate. We will see symptoms of this later.

The status of the other parameters μ_0 and Q_0 is easier to explain. These are just like the scale used to specify a measured value of the strong coupling, or the scale at which initial parton densities are parametrized for DGLAP evolution in global fits to ordinary integrated parton densities. For example, it is commonly chosen to report the value of the strong coupling at the scale of the mass of the Z boson. The choice $\mu = m_Z$ is essentially arbitrary (provided that it is in a perturbative region). No matter what scale is chosen, there is one number (a value of the coupling) that needs to be reported as the result of a fit to data. Once a particular scale is chosen, the meaning of the coupling's numerical value is fixed. If someone prefers a different choice of scale, then the original numerical coupling $\alpha_s(m_Z)$ may be transformed unambiguously to a value at the new scale, without any gain or loss of predictive power. The same remarks apply to our scales Q_0 and μ_0. The only exception could be due to the influence of errors caused by truncation of perturbation series, which is not an issue of principle.

III. UNIVERSALITY PROPERTIES

The issues that motivated this paper concerned the fitting of TMD parton densities and their evolution in one collection of data and the use of the results to predict other experimental data. However, in a more comprehensive view of TMD factorization, there are properties of the factors that concern their different kinds of universality. So in this section, we review the universality properties. Proofs of the statements made are to be found in Ref. [9] and elsewhere. The issues are particularly important for the nonperturbative functions, but they apply equally to corresponding perturbatively calculable quantities. Some of the universality properties were not fully explicitly derived prior to Ref. [9].

First, we examine the function $\tilde{K}(b_T; \mu)$ that controls TMD evolution (and its corresponding anomalous dimension γ_K and "nonperturbative function" $g_K(b_T; b_{\max})$). Since $\tilde{K}$ is a property of Wilson line operators, it depends on the color representation of the two partons entering the hard scattering, but not on quark flavor, hadron flavor, Q, nor parton fractional momentum. Thus, for all processes involving quarks, the same $\tilde{K}$ is used. It was also proved not to depend on whether the quarks are initial state or final state, so the same $\tilde{K}$ applies to all versions of the Drell–Yan process, SIDIS, and e^+e^--annihilation. An immediate implication is that in Eq. (16), etc., the part of the evolution factor that involves $\tilde{K}$ (and hence γ_K) is a common (b_T-dependent) factor in the b_T-integrand, independent of which term it applies to in the sum over parton flavors j, j_A, and j_B.

The most important real-world situation in which a different value of $\tilde{K}$ is needed is in the Drell–Yan-like process of Higgs production by gluon-gluon annihilation in hadron-hadron collisions, since gluons are color octet.

The anomalous dimension γ_j, for the TMD quark densities, arises from a mass-independent calculation with spin-1/2 quarks. It is therefore independent of quark flavor (and also of hadron flavor and parton momentum fraction). However, in hypothetical extensions of QCD, there could be scalar color-triplet quarks (like squarks in a supersymmetric theory), and these could have different anomalous dimensions. Gluons (not to mention gluinos) have a different anomalous dimension.

The hard scattering depends on the process, but only on the variables available for the partons initiating the hard scattering. It is of course perturbatively calculable and has well-known dependence on partonic flavor. There is an interesting partial universality of higher-order corrections. For example, in the electromagnetic hard scattering for Drell–Yan, the ratios of the one- and two-loop corrections to the lowest-order term are quark-flavor independent (when quark masses are neglected). Flavor dependence first arises at order α_s^3, where the virtual photon can couple to a quark loop that has a different flavor than the quark and antiquark initiating the hard scattering. However, the loop corrections are generally different between "timelike" processes (e.g., Drell–Yan) and "spacelike" processes (e.g., SIDIS). In addition, there are potential (and calculable) differences between the hard scattering for unpolarized quarks and the parts that depend on quark polarization.

As for the complete TMD functions (parton densities and fragmentation functions), the nonperturbative parts that do not arise from $\tilde{K}$ are in general all different and can depend on the flavors of the parton and hadron. The shape of the nonperturbative b_T dependence can depend on both the values of x and on the flavor. It is therefore in general incorrect to assume that the nonperturbative b_T dependence is a universal factor times the corresponding integrated distribution. The nonperturbative modeling by Schweitzer, Strikman, and Weiss [53] is important in suggesting a large difference between the b_T dependence for sea and valence quarks.

Each particular TMD parton density is the same in all processes where it is used, aside from the effects of evolution, and aside from the predicted reversal of sign [17] of "time-reversal-odd" functions (Sivers function, etc.) between Drell–Yan and SIDIS.

The universality properties of the nonperturbative functions $g_{j/A}(x, b_T; b_{max})$ match those of the corresponding TMD functions. [However, as pointed out above Eq. (24), the use of functions like $g_{j/A}$ applies, in its simplest form, only to the TMD densities that correspond to the standard integrated densities, i.e., the unpolarized, helicity, and transversity densities.]

The expansion coefficients $\tilde{C}$ in the OPE for TMD parton densities depend on the color of the partons involved. To the extent that quark masses are neglected, the dependence of $\tilde{C}$ on flavor is governed by exact flavor symmetry. Beyond lowest order, they do depend on the polarization type of the TMD functions, e.g., unpolarized TMD parton density as compared with the coefficients for the corresponding transversity TMD parton densities. They can be different between the expansions for TMD parton densities and for TMD fragmentation functions.

IV. SINGLE MASTER FUNCTION FOR CS EVOLUTION OF TMD DENSITIES

A. Definition and properties

In this section, we show how to gain a more unified view of TMD evolution. The starting point is the first form of solution (16) of the evolution equations. There, the TMD densities are all independent of Q, and the Q dependence, for each combination of flavors of quark entering the hard scattering in the first two lines arises from three sources:

(i) the $Q^{2\tilde{K}(b_T,\mu_0)}$ factor,
(ii) the exponential of anomalous dimensions,
(iii) the coupling $\alpha_s(\mu_Q)$ in the hard scattering $d\hat{\sigma}_{j\bar{j}}$.

We first observe that only the first item gives dependence on b_T, and therefore only this item gives a Q-dependent change in the shape of b_T distribution, which would then be reflected in the distribution of the cross section in the transverse momentum. This statement is valid for the contribution of a particular quark flavor. If different quark flavors have different intrinsic transverse-momentum distributions, then a change in the relative normalization of the different flavor terms would be a source of Q dependence in the shape of the transverse-momentum dependence of the cross section. However, the first two items in the list are flavor independent. Moreover, as regards the hard scattering, the ratios of one- and two-loop corrections relative to the lowest graph are flavor independent, as we observed in Sec. III. Thus, flavor dependence occurs only in the third item in the above list and only at the rather high order $\alpha_s^3(Q)$.

Hence, to a good approximation, the Q dependence in the cross section is merely an overall factor in the summed integrand, $\tilde{W}$, as in (5). This factor is a Q-dependent normalization times the $Q^{2\tilde{K}(b_T,\mu_0)}$ factor that affects the shape. This implies that a measurement of the cross section alone is, in principle, sufficient to test the evolution in Q and to give a measurement of $\tilde{K}$. Hence, for dealing with evolution, there is essentially no need to do a decomposition by parton flavor, even though the evolution kernel $\tilde{K}$ is defined as a property of the individual TMD parton densities. In this sense, the situation is quite different from the one of testing the evolution of ordinary integrated parton densities.

What is also striking is that the same flavor independence and evolution factor apply to all cases involving triplet quarks, not only to unpolarized Drell–Yan but also to polarized cases, e.g., with the Sivers function, to SIDIS, and to back-to-back hadron production in e^+e^- annihilation.

Now, consider the contribution of a particular flavor. We defined $\tilde{W}_j$ in Eq. (2). Differentiating it with respect to Q^2 (or s) at fixed x_A and x_B gives

$$\begin{aligned}\frac{\partial \ln \tilde{W}_j(\boldsymbol{b}_T, Q, x_A, x_B)}{\partial \ln Q^2} &= \frac{\partial \ln \tilde{W}_j(\boldsymbol{b}_T, Q, x_A, x_B)}{\partial \ln s} = \tilde{K}(b_T;\mu) + G_{j\bar{j}}^{\mathrm{DY}}(\alpha_s(\mu), Q/\mu) \\ &= \tilde{K}(b_T;\mu_0) + G_{j\bar{j}}^{\mathrm{DY}}(\alpha_s(\mu_Q), Q/\mu_Q) - \int_{\mu_0}^{\mu_Q} \frac{d\mu'}{\mu'} \gamma_K(\alpha_s(\mu')) \\ &= -g_K(b_T; b_{max}) + \tilde{K}(b_*;\mu_{b_*}) + G_{j\bar{j}}^{\mathrm{DY}}(\alpha_s(\mu_Q), Q/\mu_Q) - \int_{\mu_{b_*}}^{\mu_Q} \frac{d\mu'}{\mu'} \gamma_K(\alpha_s(\mu')), \end{aligned} \tag{25}$$

where $G_{j\bar{j}}^{\mathrm{DY}}$ is defined by

$$G_{j\bar{j}}^{\mathrm{DY}}(\alpha_s(\mu), Q/\mu) = \frac{\partial}{\partial \ln Q^2}\left[\ln \frac{Q^2 \mathrm{d}\hat{\sigma}_{j\bar{j}}(Q,\mu,\alpha_s(\mu))}{\mathrm{d}\Omega}\right]. \tag{26}$$

If μ is set to $\mu_Q = C_2 Q$, proportional to Q, we can write

$$G_{j\bar{j}}^{\mathrm{DY}}(\alpha_s(\mu_Q), Q/\mu_Q) = \frac{\mathrm{d}}{\mathrm{d}\ln Q^2}\left[\ln \frac{Q^2 \mathrm{d}\hat{\sigma}_{j\bar{j}}(Q,\mu_Q,\alpha_s(\mu_Q))}{\mathrm{d}\Omega}\right] + \gamma_j(\alpha_s(\mu_Q);1) - \gamma_K(\alpha_s(\mu_Q))\ln\frac{Q}{\mu_Q}. \tag{27}$$

Notice that the derivative with respect to $\ln Q^2$ in Eq. (27) is a total derivative, unlike the partial derivative in Eq. (26). That is, it acts on both the Q and the μ_Q arguments of $\hat{\sigma}$, and so the first term in the derivative of the hard cross section in Eq. (27) is of order $\alpha_s(\mu_Q)^2$. The notation on the right-hand side of Eq. (25) corresponds to a notation in the CSS papers. All the derivatives are at fixed x_A and x_B.

Notice also that, from our previous argument about flavor dependence of the hard scattering, the lowest order in which $G_{j\bar{j}}^{\mathrm{DY}}$ is flavor dependent is $\alpha_s(Q)^4$; a flavor dependence of G arises from the Q-derivative of the coupling in a three-loop graph (in the electromagnetic case). Process dependence of G (e.g., between Drell-Yan (DY) and SIDIS) would in contrast arise at $\alpha_s(Q)^2$, from the process dependence of the one-loop hard scattering. Given this (small) process dependence, in contrast to the flavor and process independence of the other terms, we have notated G with "DY" for the process under discussion.

The three forms on the right-hand side of Eq. (25) have the following significance: In the first line, all quantities are at a fixed common renormalization scale μ. In the second line, RG improvement is made on G so that it can be calculated perturbatively with small errors, while $\tilde{K}$ is kept at a fixed scale. In the last line, a RG improvement and the CSS prescription are applied to $\tilde{K}$, to allow perturbative calculations of $\tilde{K}$ supplemented by a parametrization of any nonperturbative part, by $g_K(b_{\mathrm{T}}; b_{\max})$.

Everything on the right-hand side of Eq. (25) is the same as the evolution of the standard exponent $-S$ defined by Eq. (17), except for the addition of a term,

$$\frac{\mathrm{d}}{\mathrm{d}\ln Q^2}\left[\ln \frac{Q^2 \mathrm{d}\hat{\sigma}_{j\bar{j}}(Q,\mu_Q,\alpha_s(\mu_Q))}{\mathrm{d}\Omega}\right], \tag{28}$$

that is associated with the running coupling in higher-order corrections to the hard scattering.

The change with Q of the shape of the b_{T} dependence is governed solely by the $\tilde{K}(b_{\mathrm{T}};\mu_0)$ term. All the remaining terms on the right of Eq. (25) give only a change in the normalization, since they are independent of b_{T}. There appears to be an arbitrariness, because of the choice of μ_0. However, because of the RG equation (7), a change of μ_0 in $\tilde{K}(b_{\mathrm{T}};\mu_0)$ only adds a constant, independent of b_{T}, and therefore only affects the normalization of the cross section but not its shape. Of course, the μ_0 dependence in $\tilde{K}(b_{\mathrm{T}};\mu_0)$ is canceled by μ_0 dependence of the integral over γ_K in the first line of Eq. (25).

Note that, to the good extent that $G_{j\bar{j}}^{\mathrm{DY}}$ can be approximated as flavor independent, Eq. (25) applies equally to the full b_{T}-space integrand $\tilde{W} = \sum_j \tilde{W}_j$ as well as to its individual flavor components.

It can be useful to discuss the evolution of the shape of the q_{T} dependence of the cross section, and hence of the b_{T} dependence of $\tilde{W}_j$, separately from the evolution of the normalization. Therefore, it would be useful to describe it by a function of b_{T} that does not have an arbitrary additive constant defined only by an abstract renormalization scheme, unlike $\tilde{K}$.

One possibility would be to subtract the value of $\tilde{K}$ at one fixed value b_c of b_{T}, by defining

$$\begin{aligned}\hat{K}(b_{\mathrm{T}}) &= \frac{\partial \ln \tilde{W}_j(b_{\mathrm{T}}, Q, x_A, x_B)}{\partial \ln Q^2} - \text{value with } b_{\mathrm{T}} \mapsto b_c \\ &= \tilde{K}(b_{\mathrm{T}};\mu_0) - \tilde{K}(b_c;\mu_0).\end{aligned} \tag{29}$$

Although this subtracted quantity still has an arbitrary parameter, the parameter is, so to speak, a data-related quantity. Note that the dependence on the RG scale μ_0 cancels. The significance of the definition in the first line of (29) is that it can be applied in any formalism for working with a TMD cross section. This is in contrast to $\tilde{K}$, the definition of which is within a specific CSS-like formalism. Sometimes presenting results in terms of $\hat{K}$ will allow a convenient comparison of different calculations.

Instead, we now propose what may be a better definition of a universal measure of the evolution of the shape of TMD functions, with no arbitrary scale at all. The definition is motivated by observing that any Q-dependent change in the b_{T}-shape of $\tilde{W}_j$ arises from $\tilde{K}(b_{\mathrm{T}})$ not being a constant. So the derivative of $\tilde{K}(b_{\mathrm{T}})$ gives the relevant information. Therefore, let us define

$$A(b_T) = -\frac{\partial}{\partial \ln b_T^2}\frac{\partial}{\partial \ln Q^2} \ln \tilde{W}_j(b_T, Q, x_A, x_B), \quad (30)$$

with the derivatives again being at fixed x_A and x_B. Logarithmic derivatives are used so that $A(b_T)$ is dimensionless. We have chosen a convention where A is defined to be the negative of a derivative of $\tilde{W}_j$, so that A will be a generally positive function. If we make the good approximation of neglecting the flavor dependence of the hard scattering, then we can apply definition (30) to the flavor summed integrand $\tilde{W}$ instead of its flavor components $\tilde{W}_j$.

We have constructed the definition of A so that it applies not only in the formulation of TMD factorization that we have presented but in any other similar formalism where the cross section is given as a Fourier transform of a quantity like $\tilde{W}_j$ (summed over j). These are typically some kind of TMD factorization or of resummation formalism, together with various approximations. Therefore, in principle, it could depend on all the kinematic parameters and on flavor as well as having expected dependence on b_T. By use of CSS-style TMD factorization, we will show that the true A in QCD does not depend on any of these other variables.

From the solution (16) of the evolution equations, from the definitions (19) and (20) for the b_* prescription, and from the RG equation (7) for $\tilde{K}$, we obtain

$$\begin{aligned} A(b_T) &= -\frac{\partial \tilde{K}(b_T;\mu_0)}{\partial \ln b_T^2} \\ &= -\frac{\partial \tilde{K}(b_T;\mu)}{\partial \ln b_T^2}, \\ &= \frac{\partial}{\partial \ln b_T^2}\Big[g_K(b_T;b_{\max}) - \tilde{K}(b_*;\mu)\Big] \\ &= \frac{\partial g_K(b_T;b_{\max})}{\partial \ln b_T^2} + \frac{b_*^2}{b_T^2}A(b_*;\mu) \\ &= \frac{\partial g_K(b_T;b_{\max})}{\partial \ln b_T^2} + \frac{b_*^2}{b_T^2}A(b_*;\mu_{b_*}) \\ &= \frac{\partial}{\partial \ln b_T^2}\Big[g_K(b_T;b_{\max}) - \tilde{K}(b_*;\mu)\Big]\Big|_{\mu\mapsto\mu_{b_*}}. \end{aligned} \quad (31)$$

In the second line, we have changed the value of the renormalization scale. Since a RG transformation of $\tilde{K}$ adds to it a constant, independent of b_T, we have RG invariance of its derivative with respect to b_T, and hence of A:

$$\frac{dA(b_T;\mu)}{d\ln\mu} = 0. \quad (32)$$

In the last two lines of (31), we have chosen to set the value of μ to μ_{b_*} in order give a suitable form for the use of finite-order perturbation theory, without large logarithms of μb_*. As usual in these situations, although the exact value of A does not depend on μ, a finite-order truncation of its perturbative expansion does, by an amount of order the first omitted term. The notation on the last line of (31) is to emphasize that the derivative of $\tilde{K}$ is to be taken at the fixed renormalization scale. Only after that is the renormalization scale set to its final value of order $1/b_*$. The factors of b_*^2/b_T^2 arise from the following calculation:

$$\frac{\partial \ln b_*^2}{\partial \ln b_T^2} = \frac{1}{1 + b_T^2/b_{\max}^2} = \frac{b_*^2}{b_T^2}. \quad (33)$$

We will review the two-loop calculation of A below, where we also show the equality of our master function A with the function of the same name defined by CSS.

In the later parts of (31), we have notated A with an additional scale argument. This indicates that when perturbative calculations are performed, a RG scale must be chosen and used with a corresponding value of the running coupling (and in principle running masses). More correctly, we should remember that the full set of arguments of A include all the parameters of QCD. To indicate, when necessary, the relevant arguments and parameters, we write $A(b_T) = A(b_T;\mu) = A(b_T;\mu,\alpha_s(\mu),m(\mu))$, where the number of arguments that we actually choose to write depends on the context. In the situations where we actually use perturbation theory, the (light) quark masses will normally be approximated by zero. But in a bigger context, including nonperturbative physics, the arguments must also include the μ-dependent masses of the quarks. Since A is defined from a RG-invariant quantity $\tilde{W}_j$, RG invariance applies also to A, as in Eq. (32).

Given its definition (30), we can regard the function $A(b_T)$ as a fundamental property of parton physics in QCD, independently of any particular factorization scheme and of particular techniques for its calculation. Therefore, we propose that it is a good master function for analyzing the evolution of the shape of transverse-momentum distributions. The function is directly related to experimental data, from Eq. (30). But on the theoretical side, it is a property of certain Wilson loops; this arises via TMD factorization from Eq. (31) and the definition of $\tilde{K}$ in the recent formulation of TMD factorization in Ref. [9].

Both the perturbative calculation of A and the fits of the large-b_T behavior of $\tilde{K}$ agree that $A(b_T)$ is generally positive. This implies that $\tilde{K}(b_T)$ decreases when b_T increases. Hence, when Q is increased, the $\boldsymbol{b}_T$-space integrand $\tilde{W}$ undergoes a larger fractional decrease at larger b_T than at smaller b_T. Thus, the shape in b_T of $\tilde{W}$ undergoes a shift to where it is dominated by ever smaller b_T as Q increases, as is well known, from, e.g., Ref. [2]. Correspondingly, the transverse-momentum distribution broadens. The function A codes these properties.

An important prediction of QCD is that A is independent of the process, of Q, and of the kinematic variables x_A and x_B; it is also independent of quark and hadron flavor. These

are highly nontrivial predictions of QCD dynamics, since the lack of dependence on these variables is not guaranteed merely by the definition of A in the first line of Eq. (30).

Without the CSS b_* prescription, a purely perturbative calculation from Eq. (11) gives the one-loop term in A

$$A(b_{\rm T}) = \frac{\alpha_s(C_1/b_{\rm T})C_F}{\pi} + O(\alpha_s(C_1/b_{\rm T})^2), \tag{34}$$

where we have made the standard choice of scale $\mu = C_1/b_{\rm T}$.

B. Equality with CSS's A; presentation of two-loop value

Now, in Ref. [8], CSS transformed their TMD factorization formula to a form where

$$\tilde{W}_j = (Q\text{-independent factor}) \times \exp\left\{-\int_{C_1^2/b_{\rm T}^2}^{C_2^2Q^2} \frac{{\rm d}\mu'^2}{\mu'^2}\left[A_{\rm CSS}(\alpha_s(\mu');C_1)\ln\left(\frac{C_2^2Q^2}{\mu'^2}\right) + B_{\rm CSS}(\alpha_s(\mu');C_1,C_2)\right]\right\}, \tag{35}$$

with the quantities $A_{\rm CSS}$ and $B_{\rm CSS}$ being those defined by CSS. See, for example, Eq. (5.1) of Ref. [8]. Applying our definition of A, Eq. (30) gives

$$A(b_{\rm T};\mu) = A_{\rm CSS}(\alpha_s(C_1/b_{\rm T});C_1). \tag{36}$$

Thus, we must numerically identify A, as we defined it, with CSS's quantity of the same name.[14] This is not an accident, of course. CSS's construction involved a transformation of their original TMD factorization formula so that the quantities involved could be related to properties of cross sections.

However, despite the numerical equality of our A with $A_{\rm CSS}$, differences exist in how the arguments of the function are presented. Our A is treated primarily as a function of $b_{\rm T}$; but it has as extra parameters the renormalization scale μ and the parameters of QCD, $A = A(b_{\rm T};\mu,\alpha_s(\mu),m(\mu))$, subject to RG invariance. It is defined for all $b_{\rm T}$. By its definition, $A(b_{\rm T})$ is scale independent for all $b_{\rm T}$, as expressed by Eq. (32). Therefore, despite the extra arguments, A should be treated as a function of the single variable $b_{\rm T}$, when we make plots as in later sections.

In contrast, CSS [8] presented A as a function of α_s alone. This is because they focused on the perturbative region for $b_{\rm T}$, and $A_{\rm CSS}$ arose in the construction of a particular transformation of their solution of the evolution equations—see their Eqs. (3.11)–(3.15). Moreover, because of the transformation they made of their formula, they did not actually use $A_{\rm CSS}(\alpha_s)$ with the coupling evaluated at the scale $C_1/b_{\rm T}$ as it appears in Eq. (36). Instead, after the transformation to obtain (35), they have $A_{\rm CSS}(\alpha_s(\mu))$ evaluated at a general scale inside an integral over μ in a purely perturbative context. Thus, it is completely unobvious that $A_{\rm CSS}(\alpha_s)$ is to be regarded as a function of $b_{\rm T}$ as we do. Indeed, $A_{\rm CSS}$ gives the appearance in (35) of being a modified version of γ_K.

Our definition is in a sense more general, and the use of $b_{\rm T}$ as the primary argument implies that there is a definite measurable functional form for A in the nonperturbative region, without the need to transform to a function of coupling, which would appear to be entailed by the CSS definition if taken as it stands. A consequence is that, for any methods and approximations that are proposed for some kind of TMD factorization, the function $A(b_{\rm T})$ can be obtained and compared with the same function obtained in another method. It is not restricted to perturbative uses.

This equality of our new definition A with that of CSS allows us to use the results of its calculation by Kodaira and Trentadue [95] and by Davies and Stirling [96]. We find

$$A(b_{\rm T}) = \frac{\alpha_s(\mu)C_F}{\pi} + \left(\frac{\alpha_s(\mu)}{\pi}\right)^2 C_F\left[\left(\frac{67}{36}-\frac{\pi^2}{12}\right)C_A - \frac{5}{9}T_R n_f + \left(\frac{11}{12}C_A - \frac{1}{3}T_R n_f\right)\ln\left(\frac{\mu^2 b_{\rm T}^2}{4e^{-2\gamma_E}}\right)\right] + O(\alpha_s(\mu)^3). \tag{37}$$

This differs from the Davies–Stirling result by the addition of the logarithmic term to give invariance under a change of μ. For the Davies–Stirling calculation, the CSS definition of A was used, which, via Eq. (36), can be interpreted as our A when μ is set to $2e^{-\gamma_E}/b_{\rm T}$.

From this, we can verify the known two-loop value of γ_K, given in Eq. (10), by the manipulations

$$\begin{aligned}\gamma_K(\alpha_s) &= -\frac{{\rm d}\tilde{K}}{{\rm d}\ln\mu}\\ &= -\frac{\partial\tilde{K}(b_{\rm T};\mu,\alpha_s(\mu))}{\partial\ln\mu} - \frac{{\rm d}\alpha_s(\mu)}{{\rm d}\ln\mu}\frac{\partial\tilde{K}(b_{\rm T};\mu,\alpha_s(\mu))}{\partial\alpha_s(\mu)}\\ &= -\frac{\partial\tilde{K}(b_{\rm T};\mu,\alpha_s(\mu))}{\partial\ln b_{\rm T}} - \frac{{\rm d}\alpha_s(\mu)}{{\rm d}\ln\mu}\frac{\partial\tilde{K}(b_{\rm T};\mu,\alpha_s(\mu))}{\partial\alpha_s(\mu)}\\ &= 2A(b_{\rm T};\mu,\alpha_s(\mu)) - \frac{{\rm d}\alpha_s(\mu)}{{\rm d}\ln\mu}\frac{\partial\tilde{K}(b_{\rm T};\mu,\alpha_s(\mu))}{\partial\alpha_s(\mu)},\end{aligned} \tag{38}$$

[14] *Warning:* Moch, Vermaseren, and Vogt [75] gave a calculation at $O(\alpha_s^3)$ of a quantity they called A. What they calculate is in fact $\frac{1}{2}\gamma_K$ rather than A as defined by CSS (and us). This can be checked from the definition (2.4) that Ref. [75] uses for A. Essentially the same observation was made by Becher and Neubert in Ref. [10].

given that quark masses are neglected. Then, γ_K to two loops, Eq. (10), is obtained from the value of A to two loops together with the one-loop values of $\tilde{K}$ and of the evolution of α_s. This also agrees with the result given in Moch, Vermaseren, and Vogt [75] in their Eq. (3.8), which is for what they call A but which is $\frac{1}{2}\gamma_K$ in our notation. They give the three-loop value in their Eq. (3.9).

The only dependence on the specific method for implementing perturbation theory is in the choice of μ and, in particular, in the way of optimizing the relationship to $b_{\rm T}$. If μ is set to $C_1/b_{\rm T}$, then $\alpha_s(C_1/b_{\rm T})$ is small at small $b_{\rm T}$, with finite perturbative coefficients in the expansion in Eq. (37). If C_1 is excessively large or small, the coefficients beyond order α_s become large, and the validity of a truncated perturbative expansion becomes suspect. The common choice is $C_1 = 2e^{-\gamma_E}$.

V. ASSESSMENT OF APPROXIMATIONS AND FITS

As we already mentioned, there appears to be incompatibility between different articles that have compared TMD factorization with data. In this section, we will first review a selection of these works. Then, we will present the results in terms of the master function $A(b_{\rm T})$, defined in Sec. IV, to assess their actual compatibility and validity, or lack thereof.

A. Review of fits, etc.

Some authors (e.g., Refs. [1–3,92]) have used the CSS formalism with the form of solution given in Eq. (22) (or some variant). These fits use data to make fits for the functions $g_K(b_{\rm T};b_{\rm max})$ and $g_{j/H}(x,\boldsymbol{b}_{\rm T})$ that were intended to parametrize nonperturbative $b_{\rm T}$ dependence and that were defined in Eqs. (20) and (21). However, different analyses obtain quite different numerical values. Others argue (e.g., Refs. [10–13,32]) that they can successfully completely avoid the use of the b_* method (or any similar matching method) and the function $g_K(b_{\rm T};b_{\rm max})$.

1. Fits to Drell–Yan data with CSS formalism

Landry *et al.* [1] made fits to Drell–Yan data using the CSS formalism, including the CSS treatment of nonperturbative regions by the b_* method. They chose $b_{\rm max}$ to be[15] $0.5\ {\rm GeV}^{-1} = 0.1$ fm. Their best fit used quadratic functions for $g_{j/A}$ and g_K, without flavor dependence or hadron species dependence for $g_{j/A}$,

$$\begin{aligned}&\exp\left[-g_{j/A}(x_A,b_{\rm T};b_{\rm max})-g_{j/B}(x_B,b_{\rm T};b_{\rm max})\right.\\&\quad\left.-g_K(b_{\rm T};b_{\rm max})\ln(Q^2/Q_0^2)\right]\\&=\exp\left\{-\left[g_1+g_2\ln\frac{Q}{2Q_0}+g_1g_3\ln(100x_Ax_B)\right]b_{\rm T}^2\right\},\end{aligned}\tag{39}$$

with $Q_0 = 1.6$ GeV. This factor alone, when Fourier transformed to transverse-momentum space, would give a Gaussian shape to the transverse-momentum distribution, and the $q_{\rm T}$ distribution would broaden as Q increases. The measured values of the parameters were

$$g_1 = 0.21^{+0.01}_{-0.01}\ {\rm GeV}^2,\tag{40a}$$

$$g_2 = 0.68^{+0.01}_{-0.02}\ {\rm GeV}^2,\tag{40b}$$

$$g_3 = -0.6^{+0.05}_{-0.04}.\tag{40c}$$

$$({\rm BLNY}, b_{\rm max} = 0.5\ {\rm GeV}^{-1} = 0.1\ {\rm fm}).$$

It is the g_2 coefficient that corresponds to the function $g_K(b_{\rm T};b_{\rm max})$ that embodies the nonperturbative behavior of the CSS evolution kernel $\tilde{K}$. The fitted form of the function is then

$$g_K(b_{\rm T};b_{\rm max}) = \frac{g_2}{2}b_{\rm T}^2 = \frac{0.68^{+0.01}_{-0.02}}{2}b_{\rm T}^2.$$
$$({\rm BLNY}, b_{\rm max} = 0.5\ {\rm GeV}^{-1} = 0.1\ {\rm fm}).\tag{41}$$

Updated fits were made by Konychev and Nadolsky (KN) [2]. They quantified the effect of changing $b_{\rm max}$ on the fits. At $0.5\ {\rm GeV}^{-1} = 0.1$ fm, their value for g_K was compatible with (41). But when $b_{\rm max}$ was increased, not only did the fit become somewhat better (with an optimum near $b_{\rm max} = 1.5\ {\rm GeV}^{-1} = 0.3$ fm), but the coefficient g_2 was substantially smaller, giving

$$g_K(b_{\rm T};b_{\rm max}) = \frac{0.184\pm0.018}{2}b_{\rm T}^2$$
$$({\rm KN}, b_{\rm max} = 1.5\ {\rm GeV}^{-1} = 0.3\ {\rm fm}, C_1 = 2e^{-\gamma_E}).\tag{42}$$

If one treated these fits as giving the true large-$b_{\rm T}$ behavior of g_K, and hence of $\tilde{K}$, the two fits would be strongly incompatible. However, this is not a legitimate deduction, as can be seen from Fig. 1 (which is Fig. 4 of Ref. [2]). There, the integrand in TMD factorization is plotted for several different fits, including the ones mentioned above. The integrand is the factor $b_{\rm T}\tilde{W}(b_{\rm T})$ in an integral of the form

$$\int_0^\infty b_{\rm T}\tilde{W}(b_{\rm T})J_0(q_{\rm T}b_{\rm T})db_{\rm T},\tag{43}$$

[15]To allow an easy comparison with the size of a proton, we give numerical distances in units of fm as well as $\rm GeV^{-1}$.

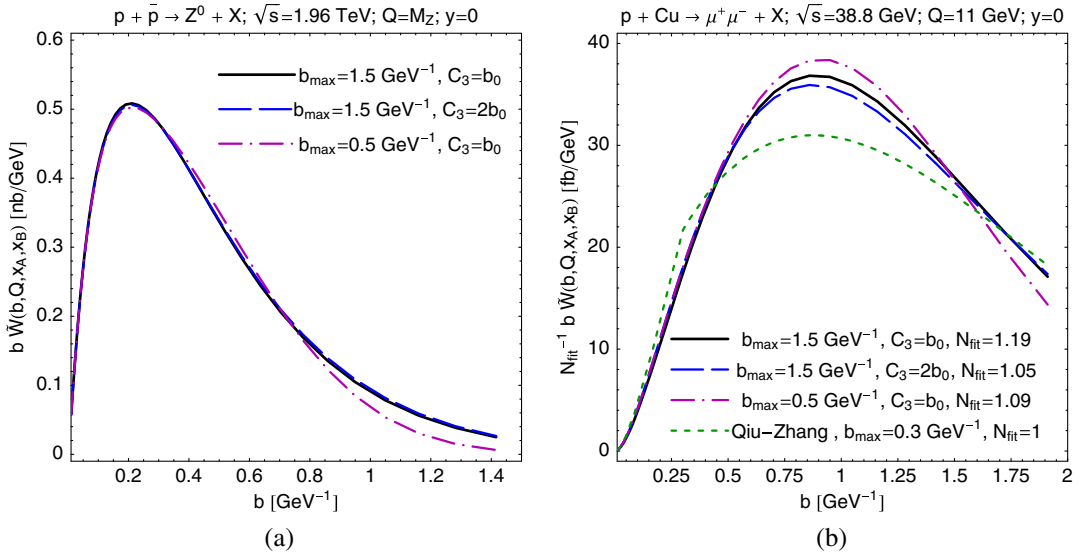


FIG. 1 (color online). b_T-space integrand for the Drell–Yan process. These plots are from Fig. 4 of Ref. [2] and show the results of using different values of b_{max}: (a) for Z production $\sqrt{s} = 1.96$ TeV; (b) for $\sqrt{s} = 38.8$ GeV and $Q = 11$ GeV. The two fits with $b_{max} = 1.5$ GeV^{-1} correspond to two different choices for the ratio $C_3 = \mu_{b_*} b_*$, which gives a measure of sensitivity to truncation of perturbative expansions. The curve labeled "Qiu–Zhang," with $b_{max} = 0.3$ GeV^{-1}, uses the Qiu–Zhang [3,92] parametrization. The normalization of $\tilde{W}$ differs from that defined in Eq. (2).

which is obtained from the TMD factorization formula by performing the integral over the azimuth of $\boldsymbol{b}_T$ to obtain a Bessel function.

In Fig. 1(a) is shown the situation at large Q, specifically the integrand for production of the Z boson. At the right-hand edge, at $b_T = 1.4$ GeV^{-1}, the difference between the fits for g_K, such as those in Eqs. (41) and (42), manifests itself as a difference between the curves of more than a factor of 2. However, although this is a large *relative* difference, it is in a region where the integrand is small so that the absolute difference is quite small. Thus, the factor-of-2 change in the tail of the integrand only has a small effect on the cross section and does not greatly affect fitting within uncertainties at these large values of Q. The dominant values of b_T that contribute to the cross section are in a perturbative region, around 0.1 to 0.5 GeV^{-1}, i.e., 0.02 to 0.1 fm. Here, the curves are very close to each other. This makes quantitatively manifest a phenomenon originally found by Parisi and Petronzio [97]: When Q is sufficiently large, the whole of the transverse-momentum distribution, even down to $q_T = 0$, is determined by a range of b_T appropriate to perturbative phenomena. That is, the only important nonperturbative information is in the use of ordinary *integrated* parton densities, and no significant extra nonperturbative information is needed for the q_T distribution. Thus, the sensitivity to nonperturbative parameters for the large-b_T behavior becomes very weak, provided only that the large-b_T behavior is parametrized in a physically reasonable way. One can therefore assert that at large enough Q the nonperturbative information is only used qualitatively, to ensure that the integrand is so small in the nonperturbative region that its detailed behavior is unimportant.

However, if one examines lower-energy data, the suppression of the large-b_T region is weaker, and correspondingly the relevant values of b_T are larger. For example, Fig. 1(b) shows that when Q is around 10 GeV the *peak* of the integrand is near $b_T \simeq 1$ GeV$^{-1} = 0.2$ fm, closer to nonperturbative phenomena. There are noticeable differences between the curves for different values of b_{max}, but the relative differences are much smaller than the factor-of-2 differences seen at the edge of graph (a).

Current discussions involve experiments and data at even lower Q, while still having Q large enough that it seems reasonable to use TMD factorization. The effects of evolution imply that these experiments probe substantially larger values of b_T that are in a more clearly nonperturbative region. In these contexts, one is often specifically interested in accessing the large-b_T region in efforts to study nonperturbative properties of hadron structure. However, there is a clear danger that the fits resulting in Eqs. (41) and (42) were primarily sensitive to smaller values of b_T than are needed for the low-Q data. Hence, the extrapolation to larger b_T of the quadratic fit beyond the region where it was determined by the fitted data could well misrepresent the true consequences of QCD.

In fact, a difficulty immediately manifests itself, as particularly emphasized by Sun and Yuan [13]: There are regions of Q, x_A, and x_B where one might reasonably expect factorization to continue to work, but where the coefficient of b_T^2 in the exponent in Eq. (39), with the BLNY values, reverses sign to become positive, given

the values in Eq. (40). An example is with $Q = 3$ GeV and $x_A = x_B = 0.3$. With this reversed sign, the integral over b_T for the cross section is badly divergent at $b_T = \infty$. Even the Konychev–Nadolsky fit with its lower value of g_2 does not [13] adequately account for the data at lower Q than where the fit was made. (See also Ref. [28] for a confirming analysis.)

This suggests the following hypothesis: The fits in Refs. [1,2] determine g_K and $g_{j/H}$ in a range of relatively low b_T, say between about 0.5 and 2 GeV^{-1}, but their extrapolation to larger b_T is wrong. The function g_K that concerns the Q dependence should not continue to rise like b_T^2; instead, it should flatten. Then, the evolution at low Q is substantially slower than the extrapolation of the old fits would give. This is without necessarily invalidating the earlier fits.

2. Work without nonperturbative evolution function

Several references, e.g., Refs. [10–13,32], use a CSS-like formalism, but without a fitted function to parametrize nonperturbative large-b_T evolution. (However, all agree in using fitted nonperturbative function(s) for the Q-independent part of the TMD parton densities.)

Echevarría *et al.* [11] argued that one can avoid the use of b_{max} and the corresponding parametrized function g_K. At the most fundamental level, they have TMD factorization and evolution results that are equivalent [98,99] to a version of the CSS formalism. So the differences really only concern how the formalism is exploited. They apply a resummation procedure to the function $\tilde{K}(b_T)$ (which is $-2D$ in their notation). They argue that the resummed formula applies in a range of b_T up to about a limiting value they term b_C. The value of this limit is given as $b_C = 6$ GeV$^{-1} = 1.2$ fm, when the scale μ is 5 GeV.

Beyond this distance scale, they agree that nonperturbative information in $\tilde{K}$ is needed. However, they argue that, since the coordinate-space parton density is already small at the limiting value $b_T = b_C$, the nonperturbative information in $\tilde{K}$ is irrelevant. If this argument is really valid, it increases the predictive power of TMD factorization.

Unfortunately, this argument is not consistent with known nonperturbative scales. First, the distance scale quoted, 1.2 fm, is clearly in a region where nonperturbative physics is important. For example, it is larger than the confinement scale found in Ref. [53] and a factor of 4 larger than the chiral scale in the same reference. Moreover, to obtain a small parton density at this value of b_T, Echevarría *et al.* use a Gaussian ansatz for the large-b_T behavior,

$$e^{-b_T^2\langle p_T^2\rangle/4}, \tag{44}$$

where the value of $\langle p_T^2\rangle = 0.38$ GeV^2 is taken from a fit in Ref. [87]. The Gaussian factor is evidently describing nonperturbative effects, which are not in any of the Feynman graphs used, even with resummation. The distance scale associated with this factor is

$$\frac{2}{\sqrt{\langle p_T^2\rangle}} = 3.2 \text{ GeV}^{-1} = 0.65 \text{ fm}. \tag{45}$$

Quite reasonably this is roughly midway between the chiral and confinement scales determined from very different theoretical considerations in Ref. [53].

It follows then that Green functions already have substantial nonperturbative contributions when transverse distances reach the value in (45). Hence, a perturbative calculation, even a resummed calculation, cannot be expected to be accurate at or beyond this scale. Wherever in b_T nonperturbative contributions are important in a TMD parton density $\tilde{f}_{j/H}(x, b_T)$, one should also expect them to be important for the evolution kernel $\tilde{K}$, and one has not at all evaded the need to use a nonperturbative contribution to it, either extracted by fitting to data or by nonperturbative theoretical methods in QCD theory (or, better, both). In Sec. V C, we will give a further analysis of the argument given in Ref. [11].

Given the above quantitative estimates of the onset of nonperturbative physics, the previously used value of $b_{max} = 1.5$ GeV$^{-1} = 0.3$ fm is reasonable. A substantially smaller value is excessively conservative, while increasing it by more than about a factor of 2 goes too far into the nonperturbative region.

We conclude that Echevarría *et al.* [11] use perturbatively based calculations for $\tilde{K}$ (admittedly with resummation) in a region where nonperturbative effects are important and that in other parts of their calculation, nonperturbative effects are important in the same region.

Becher and Neubert [10] also use a related formalism without including nonperturbative effects at large b_T. But they only claim that their formalism is valid when transverse momentum is much larger than the QCD scale, i.e., $q_T \gg \Lambda_{QCD}$. In that situation, the Fourier transform probes variations of the integrand on a scale $1/q_T$, and it is dominated by relatively small distances of order $1/q_T \ll 1/\Lambda_{QCD}$, and thus the nonperturbative contributions can be numerically unimportant.

In contrast, the full TMD factorization method described above is valid for transverse momenta down to zero, and the nonperturbative large-b_T region is important, if Q is not too large. The fact that the nonperturbative region is important is established by the essentially universal use of a Gaussian form for TMD parton densities at large b_T in fitting data.

Further work without a nonperturbative function for evolution is by Sun and Yuan [12,13]. They use a certain perturbative approximation to the exponent S of the full evolution factor, with S being defined as in Eq. (17). The approximation is essentially the same as one used by Boer in Refs. [32], Eq. (144) and [33], Eq. (34). We will refer to this as the Boer–Sun–Yuan (BSY) approximation.

The approximation was devised with the aim of expressing the TMD factorization formula in terms of TMD densities, $\tilde{f}_{j/H}(x_A, \boldsymbol{b}_T; x_A^2 Q_0^2, \mu_0)$, with a fixed scale. (More recent work by Boer in Refs. [100,101] and by Sun–Yuan in Ref. [55] no longer uses the BSY method.)

First, given the legitimate choices that $\mu_0 = Q_0$ and $\mu_Q = Q$, the full exponent (17) can be written as

$$S(b_T, Q, Q_0, Q_0) = \int_{Q_0}^{Q} \frac{d\mu'}{\mu'} \left[-2\tilde{K}(b_T; \alpha_s(Q_0), Q_0) - 2\gamma_j(\alpha_s(\mu'); 1) + \ln \frac{Q^2}{(\mu')^2} \gamma_K(\alpha_s(\mu')) \right]. \quad (46)$$

Here, we have made explicit the coupling argument $\alpha_s(Q_0)$ of the function $\tilde{K}$, since its scale does not match the scale in the coupling elsewhere in the integrand. This equation is equivalent to the exponent in Ref. [29], Eq. (86). The approximations made in Refs. [12,13,32] are to replace all quantities in Eq. (46) by their one-loop approximations and to replace the coupling $\alpha_s(Q_0)$ in $\tilde{K}$ by $\alpha_s(\mu')$, to give

$$S_{\rm BSY}(b_T, Q, Q_0, Q_0) \equiv 2C_F \int_{Q_0}^{Q} \frac{d\mu'}{\mu'} \frac{\alpha_s(\mu')}{\pi} \left[\ln \frac{Q^2}{(\mu')^2} + \ln \frac{b_T^2 Q_0^2}{4} + 2\gamma_E - \frac{3}{2} \right]. \quad (47)$$

This approximation is questionable for three reasons. First, the actual value of the coupling $\alpha_s(Q_0)$ in $\tilde{K}$ has been replaced by $\alpha_s(\mu')$. Second, the perturbation series for $\tilde{K}$ has been truncated at one loop, despite the fact that higher orders will have higher powers of $\ln(b_T Q_0)$, so it is not obvious that the higher-order corrections are small compared with lowest order. The exponent is used in an integral over all b_T, so the integral includes regions of b_T where the logarithms are arbitrarily large. Finally, we recall that in analyzing HERMES and COMPASS data important regions of b_T are in the domain of nonperturbative phenomena, and therefore where a perturbative approximation is generally invalid.

Observe that the b_T dependence in Eq. (47) results in a power-law b_T dependence in e^{-S} proportional to

$$\left(\frac{1}{b_T^2} \right)^{a(Q, Q_0)}, \quad (48)$$

where

$$a(Q, Q_0) = 2C_F \int_{Q_0}^{Q} \frac{d\mu'}{\mu'} \frac{\alpha_s(\mu')}{\pi}. \quad (49)$$

We get a divergence at $b_T = 0$ that is worse than what is present in a correct formula, and we get a suppression at large b_T.

Of course, the in-principle inadequacy of the approximation does not in itself show that the approximation is actually numerically inadequate in its application to data over a limited range of Q. To see its effects in practice, we need to evaluate the size of the error, which we will do later.

As already mentioned, the reasoning leading to the BSY approximation was aimed at expressing the TMD factorization formula in terms of TMD densities at a fixed scale. This is better achieved by our Eq. (18), where a RG transformation has been applied to $\tilde{K}$ to remove large logarithms. However, that formula also shows that the need to parameterize the nonperturbative contribution to evolution has not at all been evaded.

3. Comparisons with single-spin asymmetry data at fairly low Q

In contrast to the established good fits for Drell–Yan at relatively large Q, the situation at the lower values of Q for the HERMES and COMPASS data is rather confusing, especially for the spin-dependent TMD functions.

Aybat, Prokudin, and Rogers [26] (APR) used the results of the Landry *et al.* [1] fit given in Eq. (41) for evolution and used a fit by Anselmino *et al.* [102] for the Sivers function at HERMES at $\langle Q^2 \rangle_{\rm HERMES} \simeq 2.4\ {\rm GeV}^2$. (Average values of kinematic variables were used in the calculations.) The result seemed to be a successful prediction/postdiction of COMPASS data at $\langle Q^2 \rangle_{\rm COMPASS} \simeq 3.8\ {\rm GeV}^2$. However, the Sivers function extracted from COMPASS data was sensitive to significantly smaller x values than the HERMES data. Hence, while the apparent consistency with the BLNY $g_K(b_T; b_{\rm max})$ seen by APR is compelling, the relatively large variations between HERMES and COMPASS data were most likely due to a combination of variations in both x dependence and evolution in Q, rather than a test of Q evolution alone. See, for example, the discussion of this on p. 6 of Ref. [27] where it is explained how restricting to smaller y likely enhances the influence of very small Q and larger x, leading to a larger Sivers effect and greater direct agreement with HERMES data without explicit evolution.

Anselmino *et al.* [103] made a comprehensive analysis of both the HERMES and COMPASS data. They found improved fits with TMD evolution than without, but the effect does not seem as dramatic as in the results of APR, even though they used same BLNY form (41) for TMD evolution.

A related group of authors [88] fit the transversity and Collins functions to data from HERMES, COMPASS, and Belle, covering a substantial range of Q, but did not need to use TMD evolution to get a good fit.

Finally, Sun and Yuan [12,13] were able to get agreement with much data with their approximation (47) for the

evolution exponent. This exponent gives much less rapid evolution than is implied from the Landry *et al.* [1] fit. It should be noted that the *lowest* Q used by Landry *et al.* is about the *highest* Q considered by Sun and Yuan. This is the clue that we will exploit to motivate the idea that compatibility between the results might be obtained once one allows for the fact that they probe different ranges of b_{T}.

B. Comparison using master function A

We now use $A(b_{\mathrm{T}})$ to analyze some of the phenomenological approaches.

1. BLNY and KN

First, in Fig. 2 we show comparisons between the KN and BLNY fits, described in Sec. V A 1, and a purely perturbative calculation. The two sets of plots in the figure differ only in their horizontal scales for b_{T}. Graph (a) has a linear scale out to $b_{\mathrm{T}} = 5\ \mathrm{GeV}^{-1} = 1$ fm and thus emphasizes the nonperturbative region. Graph (b) has a logarithmic scale and goes only up to $b_{\mathrm{T}} = 1\ \mathrm{GeV}^{-1} = 0.2$ fm; it thus shows a primarily perturbative region for b_{T}. In graph (a), we have also indicated some significant scales for nonperturbative physics, from Ref. [53].

A baseline for all the comparisons is the perturbative value in Eq. (37). With the two-loop approximation and the indicated choice of scale for the strong coupling, we will call this the "purely perturbative, but RG-improved calculation" of A. It is plotted in red in the figures. We calculate the strong coupling with three active flavors, the standard two-loop formula in terms of $\Lambda_3 = 339$ MeV, with the value of Λ_3 being from a recent summary of world data by Bethke [104].

At low b_{T}, the coupling is small, and the QCD prediction for A is accurate up to higher-order corrections. When b_{T} is decreased to zero, $A(b_{\mathrm{T}})$ slowly decreases to zero, like $1/\ln(1/\Lambda b_{\mathrm{T}})$. But when b_{T} is increased sufficiently, a fixed-order approximation diverges at the Landau pole, as illustrated by the red line in Fig. 2(a); a perturbative approximation is evidently not trustworthy there.

The BLNY and KN fits use the CSS b_* prescription and a quadratic form for $g_K(b_{\mathrm{T}}; b_{\max})$. From Eq. (31), we get

$$A(b_{\mathrm{T}})_{\mathrm{BLNY,KN}} = \frac{g_2}{2} b_{\mathrm{T}}^2 + \frac{1}{1 + b_{\mathrm{T}}^2/b_{\max}^2} (\text{Two-loop } A(b_*) \text{ from (37)}) + O(\alpha_s(\mu_{b_*})^3). \quad (50)$$

For the BLNY fit, the blue dashed curves in each of the plots in Fig. 2 give the perturbative part of A, i.e., the second term on the second line of Eq. (50). When b_{T} is well below the cutoff $b_{\max,\mathrm{BLNY}} = 0.5\ \mathrm{GeV}^{-1}$, this perturbative part matches the purely perturbative value in red. But at larger b_{T}, it decreases to zero; the corresponding contribution to $\tilde{K}$ goes to a constant.

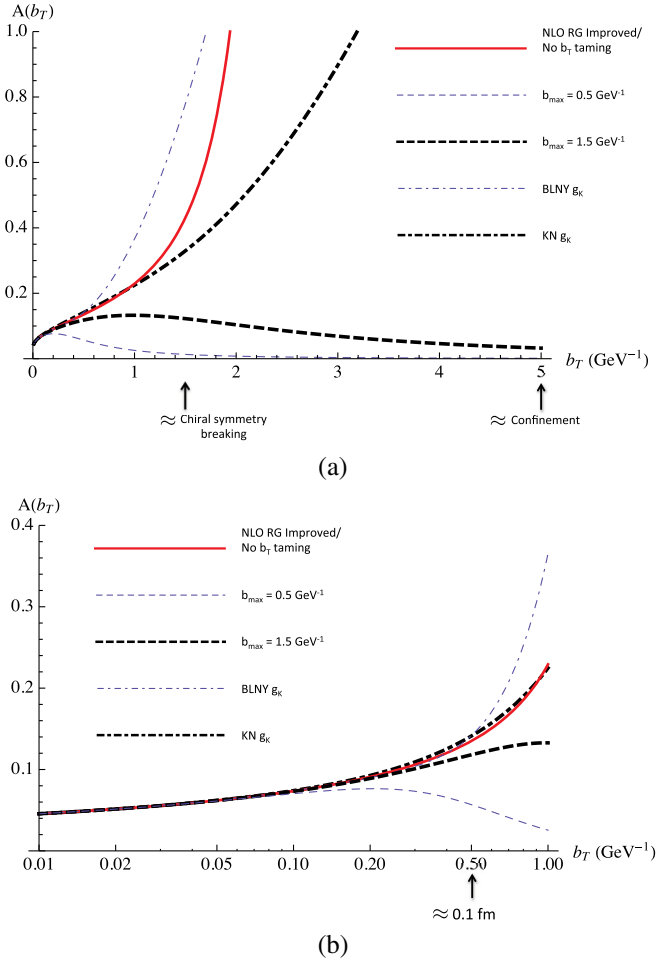


FIG. 2 (color online). The $A(b_{\mathrm{T}})$ function defined in Eq. (30) for the BLNY and KN fits, together with some of its components. The thick, solid, red curve shows a purely perturbative, but RG-improved, calculation, using the leading-order and next-to-leading-order terms in Eq. (37). The thin, blue, dashed curve shows the same quantity with the CS taming procedure, i.e., the second term in (50), with the value $b_{\max} = 0.5\ \mathrm{GeV}^{-1}$ for the BLNY fit. The thin, blue, dot-dashed curve shows the result after also adding in the $g_K(b_{\mathrm{T}}; b_{\max} = 0.5\ \mathrm{GeV}^{-1})$ function from BLNY [1]; i.e., it is the actual $A(b_{\mathrm{T}})$ for the BLNY fit. The thick, black, dashed, and dot-dashed curves similarly show the cutoff perturbative component and the full A for the KN fit [2] with $b_{\max} = 1.5\ \mathrm{GeV}^{-1}$. In (a), we plot the results with a linear scale out to $5\ \mathrm{GeV}^{-1} = 1$ fm. In (b), we show the same functions, but with a logarithmic horizontal scale extending to $1\ \mathrm{GeV}^{-1} = 0.2$ fm, to show better the region of small b_{T}, to which large-Q interactions are most sensitive.

When the fitted quadratic term for g_K is added—see Eq. (41) for the coefficient—we get the blue dash-dotted curve. This starts by compensating the effect of the $b_{\max}$ cutoff on the perturbative term and then rapidly gets large.

For the KN fit, with $b_{\max} = 1.5\ \mathrm{GeV}^{-1}$, the corresponding results are shown by the black curves. The black dashed

curve is the cutoff perturbative part, and the black dot-dashed curve is the result of adding in the fitted function for g_K, Eq. (42).

Observe first that the full BLNY curve matches the purely perturbative term very closely up to $b_T = 0.5\ \mathrm{GeV}^{-1}$, even though the perturbative term in Eq. (50) has already been substantially reduced by the b_* cutoff. The difference between the full BLNY value for A and the cutoff perturbative term is the g_K function, which BLNY fitted to data. Thus, one can reasonably say that BLNY verifies a perturbative prediction.

The KN fit matches the purely perturbative calculation slightly less well at $0.5\ \mathrm{GeV}^{-1}$. But then it matches its trend substantially farther out than does the BLNY fit, to a bit beyond $1.5\ \mathrm{GeV}^{-1}$, admittedly in a region where the accurate applicability of perturbation theory might be debatable. Beyond that, it falls below the perturbative curve. KN's fit has a notably better fit to the data, as measured by χ^2, so we should regard the KN fit as more correctly corresponding to reality.

Recall from Fig. 1 that the b_T-space integrand decreases rapidly beyond around $1.5\ \mathrm{GeV}^{-1}$, for the lower-energy data, and that the decrease starts even earlier for collider data. Thus, although there are large differences at large b_T between the A functions in the two fits, the data used are not sensitive to the higher values of b_T, say beyond $2\ \mathrm{GeV}^{-1}$; this insensitivity is even more marked at higher energy, as is seen in Fig. 1(a).

For a more incisive measurement at large b_T, one must use data at lower Q, such as is provided by HERMES and COMPASS.

2. Boer–Sun–Yuan

We now examine the consequences of the BSY approximation [12,13,32], which gave Eq. (47) for the exponent in the b_T integrand. We apply the definition of A, the first line of Eq. (30), to the value of $\tilde{W}$ that corresponds to Eq. (47). This gives

$$A(b_T; Q)_{\mathrm{BSY}} \approx \frac{\alpha_s(Q) C_F}{\pi}. \tag{51}$$

Now, a highly nontrivial prediction of TMD factorization in QCD is that $A(b_T)$ is independent of other kinematic variables, notably Q, when TMD factorization is applied to leading power in q_T/Q. So the Q dependence in (51) is already in contradiction with the prediction.

To show some of its implications, we have plotted in Fig. 3 values for the BSY A at $Q = 1$ GeV (roughly appropriate for HERMES) and at $Q = 10$ GeV (roughly appropriate for the lower-energy Drell–Yan data used in the KN and BLNY fits). These are, respectively, the brown solid and blue dashed curves. In addition, we show the purely perturbative term (red solid line) and the full KN and BLNY fitted values (black and blue dot-dashed lines).

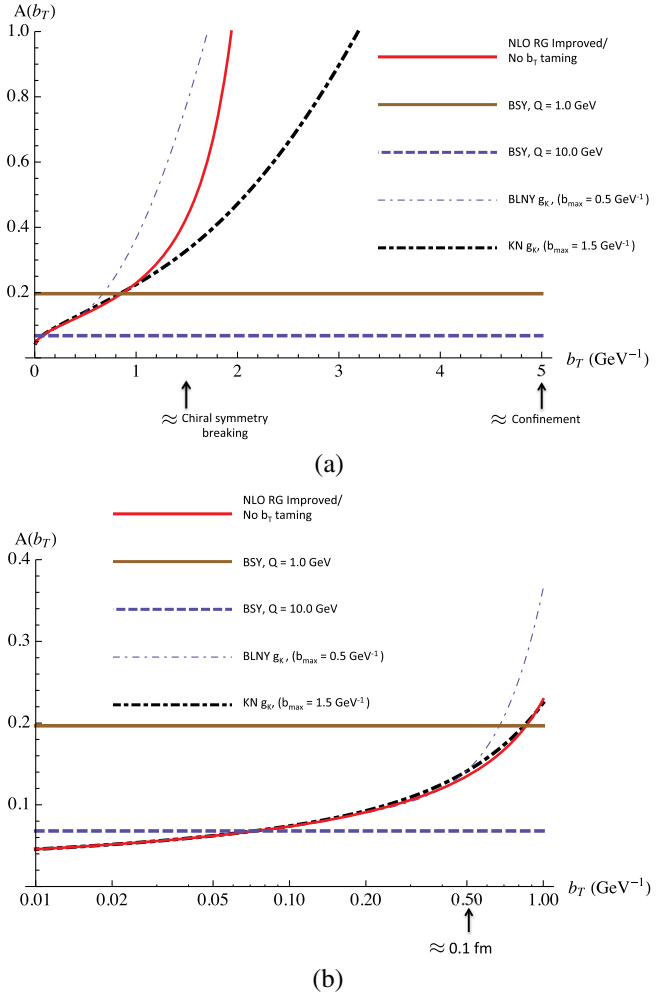


FIG. 3 (color online). Like Fig. 2, but showing a comparison between the BSY approximation, the KN fit, and the purely perturbative, RG-improved, calculation. The thick, red, solid curve is the next-to-leading-order perturbative calculation. The thick, black, dot-dashed, and thin, blue dot-dashed curves are the same KN and BLNY curves as in Fig. 2. The solid brown curve is for A at $Q = 1.0$ GeV in the BSY [13] method, from Eq. (51). The dashed blue curve is the BSY calculation for $Q = 10.0$ GeV. In (a), we plot the results with a linear scale out to $5\ \mathrm{GeV}^{-1} = 1$ fm. In (b), we show the same functions, but with a logarithmic horizontal scale extending to $1\ \mathrm{GeV}^{-1} = 0.2$ fm, to show better the region of small b_T, to which large Q interactions are most sensitive.

To assess the effect of A on the cross sections under different kinematic conditions, we need to know the typical values of b_T that are important. We measure these by the point at which the integrand $b_T \tilde{W}(b_T)$ falls to half its peak value. (We note that at asymptotically large Q the peak occurs at around $b_T \sim 1/Q$, but the width falls as a slower power of Q [97]; this is quite different than for a Gaussian.) At the Z, the left-hand graph in Fig. 1 gives a width of about $0.5\ \mathrm{GeV}^{-1}$. For Q around 11 GeV, the right-hand plot gives a width of about $1.8\ \mathrm{GeV}^{-1}$.

In contrast, transverse-momentum distributions at HERMES and COMPASS are fit reasonably well by Gaussians,

$$e^{-p_T^2/\langle p_T^2\rangle}, \tag{52}$$

with the mean-square transverse momentum $\langle p_T^2\rangle$ being around 0.2 to 0.3 GeV^2. Fourier transformed, this gives

$$e^{-b_T^2\langle p_T^2\rangle/4}, \tag{53}$$

which gives a width in position space of very roughly $2/\sqrt{\langle p_T^2\rangle}$, perhaps 4 GeV^{-1}.

In Fig. 3, the brown curve, for $Q = 1$ GeV, which is appropriate for low-energy data, intersects the KN and perturbative curves at $b_T \simeq 1$ GeV^{-1} and falls below them at higher b_T. Thus, the BSY approximation gives slower evolution than KN at the lowest values of Q.

When the energy is increased, the BSY form of A becomes the blue dashed curve. This matches the QCD-predicted form at around $b_T \simeq 0.1$ GeV^{-1} and is well below the QCD perturbative prediction in a region of b_T that is relevant for the successful KN and BLNY fits, where the fits themselves confirmed the accuracy of the perturbative predictions for $b_T \lesssim 1$ GeV^{-1}. As Sun and Yuan themselves acknowledge, their approximation is not adequate for the higher-energy data. The best that their approximation can manage is that the fall of the resulting function A with Q roughly matches the values of the true A at the values of b_T probed at that Q; these distances decrease as Q increases.

From the success of the BLNY and KN fits, together with the knowledge that perturbative QCD predictions are valid in the perturbative region of b_T, we should regard the values of $A(b_T)$ given by the KN fit as reliable in the region $b_T \lesssim 2$ GeV^{-1}, roughly. Larger values of b_T are unimportant for the Drell–Yan data fit by KN. But it is larger values of b_T that dominate for the conditions of the HERMES and COMPASS experiments. Therefore, we should expect that the true A turns down above about 2 GeV^{-1}. This would also match the conclusions of Aidala *et al.* [28]. Most importantly, we should be able to achieve this while preserving the goodness of fit for the high-energy Drell–Yan data, not to mention predictions for cross sections at high-energy hadron-hadron colliders.

In the future, it may become practical to analyze data directly in coordinate space, as proposed in Ref. [105]. This would circumvent the complications associated with Fourier transforms or convolution integrals.

3. Qiu–Zhang

In Fig. 4, we show comparisons for the result for $A(b_T)$ in the Qiu–Zhang fit in Refs. [3,92]. The Qiu–Zhang result is the brown line with a step at $b_{max} = 0.5$ GeV^{-1}. They use a sharp cutoff at $b_T = b_{max}$ instead of CSS's smooth cutoff. Below b_{max}, they use exactly the perturbatively calculated formula. Above b_{max}, they freeze b_T at b_{max} and

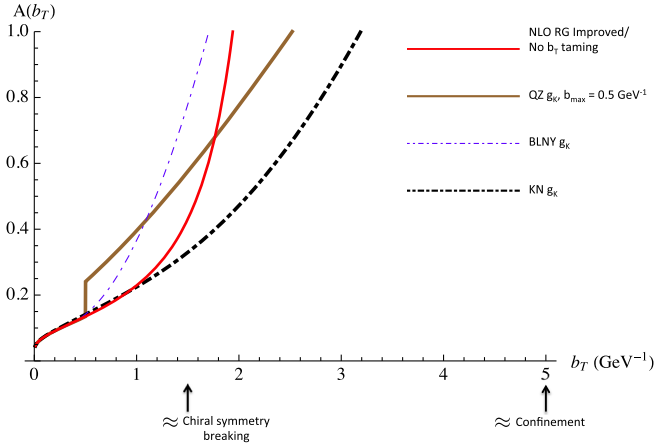


FIG. 4 (color online). Like Fig. 2(a), but now we show the comparison between $A(b_T)$ for the Qiu–Zhang model of Refs. [3,92] (brown line with step), the purely perturbative RG-improved calculation (red), and the full A functions from the BLNY fit (thin blue dot-dashed) and the KN fit (thick black dot-dashed). Since by construction the Qiu–Zhang method agrees with the RG-improved perturbative calculation below the b_{max} cutoff, we do not bother with the logarithmic plot for low b_T.

multiply the b_T-space integrand by a parametrized form corresponding to TMD factorization and evolution,

$$\tilde{W}(b_T,\ldots) = \begin{cases} \tilde{W}^{(\mathrm{pert})}(b_T,\ldots) & \text{if } b_T < b_{max}, \\ \tilde{W}^{(\mathrm{pert})}(b_{max},\ldots)F_{QZ}^{NP}(b,\ldots) & \text{if } b_T > b_{max}, \end{cases} \tag{54}$$

where

$$F_{QZ}^{NP}(b_T,\ldots) = \exp\left\{-\ln\left(\frac{Q^2 b_{max}^2}{c^2}\right)[g_1(b_T^{2\alpha} - b_{max}^{2\alpha}) + g_2(b_T^2 - b_{max}^2)] - \bar{g}_2(b_T^2 - b_{max}^2)\right\}, \tag{55}$$

and α, g_1, g_2, and $\bar{g}_2$ are parameters to be fit to data.

The perturbative formula is used for small b_T. By construction, $\tilde{W}$ is continuous at $b_T = b_{max}$. The result seen in Fig. 4 is that $A(b_T)$ contains a discontinuity at $b_T = b_{max}$.

C. Echevarría–Idilbi–Schäfer–Scimemi

Echevarría, Idilbi, Schäfer, and Scimemi (EISS) [11] argue that the practical effect of the nonperturbative part of $\tilde{K}$ is small. This conclusion is obtained from an argument that the range of b_T in which perturbative calculations are valid is a factor of 2 or more larger than has normally been considered appropriate. Although underlying their derivations are TMD factorization and evolution formulas that are

exactly equivalent to those we presented in Sec. II A, their method of solution is very different.

In our solutions (following CSS), we used the RG equation (7) to evolve $\tilde{K}$ to a scale μ of order $1/b_T$, in order to avoid large logarithms of μb_T in the perturbative expansion of $\tilde{K}$ in powers of $\alpha_s(\mu)$. Expressed in terms of the quantity $D^R(b_T;\mu) = -\frac{1}{2}\tilde{K}(b_T;\mu)$ used by EISS, this gives

$$D^R(b_T;\mu) = D^R(b_T; C_1/b_T) + \frac{1}{2}\int_{C_1/b_T}^{\mu} \frac{d\mu'}{\mu'}\gamma_K(\alpha_s(\mu')). \tag{56}$$

For a numerical estimate of $D^R(b_T;\mu)$, our method is to use the right-hand-side of Eq. (56) with truncated perturbative calculations for $D^R(b_T; C_1/b_T)$ and for $\gamma_K(\alpha_s(\mu'))$ and with a truncated perturbation expansion of the β function that controls the running of the coupling by

$$\frac{d\alpha_s(\mu)/4\pi}{d\ln\mu} = \beta(\alpha_s(\mu)) = -2\sum_{n=0}^{\infty}\beta_n\left(\frac{\alpha_s(\mu)}{4\pi}\right)^{n+2}. \tag{57}$$

Errors in the resulting approximation can be estimated from the truncation errors in the three perturbative expansions used. When μ is fixed in a perturbative region and b_T is increased beyond $1/\mu$, the largest of the errors is controlled by the size of $\alpha_s(C_1/b_T)$. When b_T is large enough, the inapplicability of perturbation theory is signaled by a Landau singularity in the value used for $\alpha_s(C_1/b_T)$ and a corresponding singularity in the *approximated* $D^R(b_T;\mu)$. Strictly speaking, a full argument for the inapplicability of perturbation theory at large distances arises not only from the large size of the effective coupling, by itself, but also from the knowledge that perturbation series in quantum field theory are often asymptotic series rather than actually convergent series. This results in an ultimate limit to the accuracy of perturbative approximations, no matter how many terms in the series are used.

Instead, EISS use an explicit resummation of logarithms for the perturbative expansion of $D^R(b_T;\mu)$, i.e., for the left-hand side of (56). The logarithms are powers of $L_\perp(\mu b_T) = \ln(\mu^2 b_T^2/4e^{-2\gamma_E})$. In massless perturbation theory for D^R, the coefficient of $(\alpha_s(\mu))^n$ is a polynomial of degree n in $L_\perp$,

$$D^R(b_T;\mu) = \sum_{n=1}^{\infty}\sum_{p=0}^{n} D_{np}a^n L_\perp^p, \tag{58}$$

where $a(\mu) = \alpha_s(\mu)/4\pi$. The leading-logarithm (LL) resummation is the sum of the terms with the maximal number of logarithms, the next-to-leading-logarithm (NLL) term is the sum of the terms with one fewer logarithm, etc. To implement resummation, EISS define $X(\mu b_T, b_T) = a(\mu)\beta_0 L_\perp(\mu b_T)$ and then reorganize the perturbation series as

$$D^R(b_T;\mu) = \sum_{n=0}^{\infty} a^n D_n(X), \tag{59}$$

where

$$D_n(X) = \sum_{j=j_{\min}(n)}^{\infty}\frac{D_{j+n,j}}{\beta_0^j}X^j. \tag{60}$$

Here, $j_{\min}(0) = 1$, and $j_{\min}(n) = 0$ for $n \geq 1$. The term with D_n sums the nth-to-leading logarithms. The LL approximation is $D_0(X)$, the NLL approximation is $D_0(X) + aD_1(X)$, etc. EISS provide formulas for resummations at the LL, NLL, and next-to-next-to-leading-logarithm (NNLL) approximations. They give a derivation in their Appendix B from the RG equation for D^R and its solution (56).

If one-loop approximations are used for γ_K, for the β function that controls the running of the α_s, and for $D^R(b_T; C_1/b_T)$ with $C_1 = 2e^{-\gamma_E}$, then the right-hand side of (56) gives the same results as the LL resummation.

But beyond that level, there are generally differences between using the resummation formula (59) truncated to some order in a and directly using the right-hand side of the solution (56) with corresponding truncations in the quantities that appear in it.

Ultimately what matters for making predictions for a cross sections is the size (or expected size) of the errors in whatever approximation is used to estimate $D^R(b_T;\mu)$.

A symptom of one source of error is visible in EISS's Fig. 1 in Ref. [11]. Here are shown plots of the results of their resummations, for $\mu = \sqrt{2.4}$ GeV and for $\mu = 5$ GeV. (In the plot, the symbol Q_i is used instead of our μ.) If the exact $D^R(b_T;\mu)$ were plotted as a function of b_T for two different values of μ, then the two curves would differ only by a simple vertical shift. This follows directly from the right-hand side of (56), which implies that

$$D^R(b_T;\mu_2) - D^R(b_T;\mu_1) = \frac{1}{2}\int_{\mu_1}^{\mu_2}\frac{d\mu'}{\mu'}\gamma_K(\alpha_s(\mu')). \tag{61}$$

These results follow from the RG equation for D^R:

$$\frac{dD^R(b_T;\mu)}{d\ln\mu} = \frac{1}{2}\gamma_K(\alpha_s(\mu)). \tag{62}$$

In going from $\mu = \sqrt{2.4}$ to $\mu = 5$ GeV, the shift is upward, since γ_K is dominated by its positive leading-order term. We now examine Fig. 1 in Ref. [11]. At small b_T, below about 2 or even 3 GeV^{-1}, the upward shift is easily seen. But at larger b_T, the plots disagree with the prediction: The calculated D^R decreases substantially instead of increasing as μ increases.

Of course, an approximate calculation of D^R need not exactly obey a property known to hold for the exact

quantity. But the deviations from the predicted b_T-independent shift give a lower bound on the error in the approximation. From the plots, we deduce that the calculation is no longer trustworthy as b_T approaches about 3 GeV^{-1}. This therefore falsifies EISS's assertion that their resummation is valid up to around $b_c = 6$ GeV^{-1} when $\mu = 5$ GeV.

That the NLL and higher approximations do not give a simple vertical shift can be seen by differentiating $D_0 + aD_1$ with respect to μ (or Q_i). From the formulas given by EISS, it is found that the value of the derivative depends on $L_\perp$ and hence on b_T. The result is of order a^2 (or higher) times a function of X. Thus, it is of a form that would be compensated by including higher-order terms in the expansion in nonleading logarithms.

We next obtain the expected size of the truncation errors for (59) both from the explicit formulas given by EISS for D_0, D_1, and D_2 and from their derivation in Appendix B of Ref. [11]. The properties we will list below can be seen explicitly in EISS's formulas up to $n = 2$. They can be proved to hold for all n, which is not too hard to do starting from the derivation in their Appendix B.

Each higher term has one factor more of $a(\mu)$, i.e., of $\alpha_s(\mu)$. But it is not this factor alone that is the true expansion parameter relevant for the size of the higher terms. Each term can be written as a polynomial in $1/(1 - X(b_T\mu, \mu))$ and $\ln(1 - X(b_T\mu, \mu))$. When $b_T\mu$ gets large enough, $X(b_T\mu, \mu)$ approaches 1 so that the terms exhibit a Landau singularity. The leading singularity for the $a(\mu)^n$ term is proportional to $[\frac{\ln(1-X(b_T\mu,\mu))}{1-X(b_T\mu,\mu)}]^n$, when $n \geq 1$. That is, each extra power of $a(\mu)$ is accompanied by one extra power of $\frac{\ln(1-X(b_T\mu,\mu))}{1-X(b_T\mu,\mu)}$. Hence, the effective expansion parameter, when $X(b_T\mu, \mu)$ gets near to 1, is not $a(\mu)$, but $\frac{a(\mu)\ln(1-X(b_T\mu,\mu))}{1-X(b_T\mu,\mu)}$. This is roughly the effective coupling (divided by 4π) at a scale proportional to $1/b_T$, i.e., $a(C_1/b_T)$. [Note that if the one-loop approximation is used for β, then the corresponding approximation to the effective coupling is $\alpha_s(C_1/b_T) = \alpha_s(\mu)/(1 - X(b_T\mu, \mu))$.]

We therefore deduce that the series in leading and nonleading logarithms becomes inapplicable when $a/(1 - X(b_T\mu, \mu))$ is of order unity, i.e., when $1 - X(b_T\mu, \mu)$ is of order $a(\mu)$. This implies that the other terms in the polynomial in $1/(1 - X(b_T\mu, \mu))$ and $\ln(1 - X(b_T\mu, \mu))$ are a smaller relative correction since they are smaller in size than the term with the leading singularity.

The result is that the error in a truncation of the expansion in leading and nonleading logarithms is governed by the same parameter $a(C_1/b_T)$ that determines the sizes of truncation errors in the direct use of the solution of the evolution equation, i.e., the truncation errors for the right-hand side of Eq. (56).

So nothing in terms of accuracy has been gained by a use of literal resummation instead of the use of the right-hand side of (56) with truncations in perturbative expansions of quantities that have no logarithms. What is lost is relative simplicity in the formulas and a transparency in the nature of the errors.

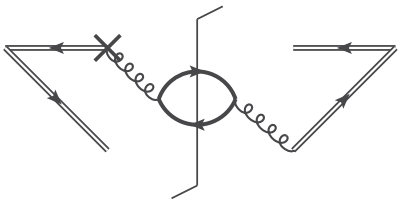


FIG. 5. Typical two-loop graph with a heavy quark loop that contributes to $\tilde{K}$. The notation is as in Ref. [9]: The cross denotes a derivative of a spacelike Wilson line with respect to its rapidity.

Another important issue is EISS's treatment of heavy quarks. Their resummation is obtained from calculations in Ref. [75] of massless Feynman graphs. Now, we know that propagators of heavy particles decay exponentially at large distances, like e^{-mb_T} (aside from a power law that is irrelevant here). The distances we are concerned with are for potential nonperturbative contributions to $\tilde{K}$ or D^R, i.e., for b_T above about 1 GeV^{-1}. In this region of b_T, the exponential e^{-mb_T} is very important for both the charm and bottom quarks. For an approximation using massless quarks to be suitable, this region of b_T is such that one must decouple the heavy quarks first. That is, the appropriate number of active quark flavors is 3.

Furthermore, when $m_q b_T$ is not small for some quark of mass m_q, the dependence on b_T is nontrivial; one cannot expect a simple logarithmic dependence. An example of a two-loop graph with a heavy quark that contributes to $\tilde{K}$ is shown in Fig. 5. This and similar graphs have a heavy quark loop correcting a gluon propagator in any of the Wilson-line matrix elements that occur in the definition of $\tilde{K}$.

EISS first estimate the location of the Landau singularity by setting the number of active flavors to 5 and using the value[16] $\Lambda_5 = 157$ MeV for the scale parameterizing the running coupling. They calculate a corresponding distance scale $b_{\Lambda_{\rm QCD}} \simeq 7$ GeV^{-1}; this calculation corresponds to $b_{\Lambda_{\rm QCD}} = 2e^{-\gamma_E}/\Lambda_5$.

But, as we observed, the appropriate number of active flavors is 3. Then, one should use [104] $\Lambda_3 = 339$ GeV, more than a factor of 2 larger. This brings the corresponding distance scale down to $b_{\Lambda_{\rm QCD}} \simeq 3$ GeV$^{-1} = 0.6$ fm, which is much more in line with standard ideas of distance scales where nonperturbative physics is important.

Having obtained a form for D^R (or $\tilde{K}$) at some low scale μ_i, one naturally wishes to evolve it to a much larger value of μ, whenever processes with large Q are to be treated.

[16]Note that the value of Λ_5 used by EISS is somewhat smaller than the measured value [104].

As is usual with similar matrix elements (e.g., densities of light quarks), one should then evolve it by Eq. (7), used in regions with different numbers of active quark flavors, together with the relevant matching conditions. This evolution preserves the property that the difference (61) between D^R at different scales is independent of b_T. (As mentioned earlier, a full treatment of the effects of heavy quarks on TMD factorization remains to be worked out; we hope to deal with this in a future article.)

In any case, it is entirely incorrect to use a massless approximation in perturbative calculations of D^R when $m_q b_T$ is not small. So performing a resummation with four or five active flavors, as EISS do, is incorrect when b_T is in the region of 1 GeV^{-1} upward.

We have one final concern with the EISS approach. This is that it implements the transition from the perturbative part of D^R to the nonperturbative part through a sharp cutoff at a certain position $b_T = b_c$. For b_T below b_c, EISS calculate D^R from resummed perturbation theory alone and use a parametrized form only for larger b_T. However, as explained in the next section, such a sharp boundary between perturbative and nonperturbative regions does not correspond to what happens in actual quantum field theory. CSS's motivation for a smooth cutoff remains valid. (Of course, there is the possibility that RG-improved or resummed perturbation theory remains accurate at larger b_T than has been assumed earlier.)

Hence, we conclude that, in contrast with the implication of the EISS results, one does indeed need to allow for a nonperturbative parametrization in the phenomenologically important range of b_T beginning above roughly 1 or 2 GeV^{-1}. A separate question, to be discussed below, is whether changes are needed compared with previously accepted parametrizations.

We note that the more recent phenomenological work of Ref. [106], which utilizes the EISS method, does in fact consider a nonperturbative part for D^R. However, the problems discussed above remain in their calculations.

VI. SMALL-b_T BEHAVIOR OF g_K

The function $g_K(b_T; b_{\max})$ was defined in Eq. (20) as (the negative of) the difference between the exact $\tilde{K}$ and a cutoff version restricted to a roughly perturbative region of b_T. It is power suppressed when b_T is much less than $b_{\max}$. CSS deliberately chose a smooth cutoff because a normal calculation of $\tilde{K}$ uses RG-improved perturbation theory with truncations of the RG functions $\beta(\alpha_s)$ and $\gamma_K(\alpha_s)$ (or some resummation with similar accuracy). The accuracy of such a calculation gradually degrades as b_T increases. Nonperturbative phenomena are not restricted to a sharp range of b_T but contribute at all b_T; but they are expected to be power suppressed at small b_T.

Hence, when Eq. (22) is actually used to fit data, there is a subtle shift in the meaning of $g_K(b_T; b_{\max})$. Instead of g_K defined in Eq. (20), the fits should be regarded as actually measuring the following quantity[17]:

$$\hat{g}_K(b_T; b_{\max}) = -\tilde{K}(b_T, \mu_0)_{\text{exact}} + \tilde{K}(b_*, \mu_0)_{\text{approx}}. \quad (63)$$

Here, the perturbatively approximated $\tilde{K}$ is defined as follows. From the RG equation for $\tilde{K}$, we find

$$\tilde{K}(b_T; \mu_0; \alpha_s(\mu_0)) = \tilde{K}(b_T; \mu_{b_*}; \alpha_s(\mu_{b_*})) - \int_{\mu_{b_*}}^{\mu_0} \frac{d\mu'}{\mu'} \gamma_K(\alpha_s(\mu')), \quad (64)$$

where the renormalization scale in $\tilde{K}$ on the right-hand side is chosen to avoid large logarithms in its perturbation series. Then, we define $\tilde{K}(b_*, \mu_0)_{\text{approx}}$ by applying to the right-hand side of (64) truncated perturbation theory for $\tilde{K}$, γ_K, and for the function β controlling evolution of the coupling. [Note that this form of approximation preserves the property of the exact $\tilde{K}(b_T; \mu_0)$ that its dependence on b_T and μ_0 is the sum of a function of b_T and a function of μ_0.] Thus, a fit of g_K can allow both for nonperturbative phenomena and for uncalculated higher-order terms in the perturbative part of $\tilde{K}$.

To better understand the changed g_K, we write it in two forms:

$$\hat{g}_K(b_T; b_{\max}) = [-\tilde{K}(b_T, \mu_0)_{\text{exact}} + \tilde{K}(b_*, \mu_0)_{\text{exact}}] + [-\tilde{K}(b_*, \mu_0)_{\text{exact}} + \tilde{K}(b_*, \mu_0)_{\text{approx}}] \quad (65a)$$

$$= [-\tilde{K}(b_T, \mu_0)_{\text{exact}} + \tilde{K}(b_T, \mu_0)_{\text{approx}}] + [-\tilde{K}(b_T, \mu_0)_{\text{approx}} + \tilde{K}(b_*, \mu_0)_{\text{approx}}]. \quad (65b)$$

In (65a) the quantity in the first brackets is g_K as originally defined, and the second brackets contain the error in using truncated perturbative methods to compute $\tilde{K}(b_*)$. In (65b) the first brackets show the difference between the exact $\tilde{K}$ and a truncated perturbatively based estimate, while the second brackets give a perturbative approximation to g_K.

If $b_{\max}$ is chosen conservatively (as in the BLNY fits), then perturbatively based calculations of $\tilde{K}$ are applicable for the whole region of b_T less than $b_{\max}$, and even at somewhat larger b_T. Actual fits for g_K, or rather $\hat{g}_K$, particularly with a simple quadratic approximation, are a compromise between reproducing g_K in a region where it is predicted and fitting g_K at larger b_T where it is less

[17] It is not entirely clear which value μ_0 of the renormalization scale should be used here but that only affects an additive constant, with no dependence on b_T.

perturbative. Even so, we expect to estimate, roughly, the small-b_T behavior of g_K from perturbative calculations. We can regard such an estimate as giving a property of the first term on the right-hand side of (65a), i.e., of g_K itself. Alternatively, it gives a property of the second term on the right-hand side of (65b). The validity of perturbation theory when b_T is small is coded in a small value for the other term on each line, which is a difference between the exact $\tilde{K}$ and its perturbative estimate.

Real nonperturbative physics is at larger b_T, and as we will see in more detail in Sec. VII, a simple extrapolation of g_K from small b_T is likely to be wrong.

Once a less conservative value of b_{max} is chosen, more of the fitting is concerned with effects beyond those predicted by low-order perturbation theory. This is exhibited on the right-hand side of (65a), where the first term is the exact g_K and the second term gives the error in replacing the exact value of $\tilde{K}(b_*, \mu_0)$ by a perturbatively based approximation.

We now show how to predict approximately the quadratic behavior of g_K when $b_T \lesssim b_{max}$. This amounts to an examination of the second term on the right-hand side of (65b). We will find that the results roughly reproduce the values of the coefficient g_2 in Eq. (39) that were fitted by BLNY and KN.

If $\tilde{K}$ were an analytic function of b_T around $b_T = 0$, then g_K as defined by Eq. (20) would be correctly given by a quadratic in b_T at small b_T. But in fact, $\tilde{K}$ has a mild singularity at $b_T = 0$, as is verified by doing a renormalization-group improvement, as in (64). Because the effective coupling $\alpha_s(\mu_{b_*})$ is not analytic at $b_T = 0$, neither is $\tilde{K}$. This mildly modifies the quadratic small-b_T behavior of g_K. But normally we are not concerned with accurately approximating g_K at very small b_T, precisely because g_K is small there and has little effect on the cross section. What we need to obtain is an approximation that is useful when b_T gets closer to b_{max}.

To get a simple approximation, we first set $\mu = \mu_{b_*}$ in the definition of g_K, to remove large logarithms:

$$g_K(b_T; b_{max}) = -\tilde{K}(b_T; \mu_{b_*}; \alpha_s(\mu_{b_*})) + \tilde{K}(b_*; \mu_{b_*}; \alpha_s(\mu_{b_*})). \tag{66}$$

We assume that b_T is less than b_{max}. Then, there are no large logarithms involving b_T or b_*. Using the lowest-order formula for $\tilde{K}$ gives

$$g_K(b_T; b_{max}) \simeq \frac{\alpha_s(C_1/b_*)C_F}{\pi} \ln(1 + b_T^2/b_{max}^2). \tag{67}$$

This has b_T^2 behavior at small b_T but a slower rise above b_{max}. It is the form used in Ref. [54] to optimize matching between the perturbative calculation and $g_K(b_T; b_{max})$ at moderate b_T.

To compare with fitted values of g_K with $g_K = \frac{1}{2} g_2 b_T^2$, we propose two methods. One is to expand at small b_T,

$$g_K(b_T; b_{max}) \simeq \frac{\alpha_s(C_1/b_*)C_F}{\pi} \frac{b_T^2}{b_{max}^2}, \tag{68}$$

and then to replace C_1/b_* by C_1/b_{max}, since fits for g_K concern b_T not far from b_{max}. Then, we equate the coefficients of b_T^2 in this formula and in the fitted g_K to obtain

$$g_2 \simeq \frac{2\alpha_s(C_1/b_{max})C_F}{\pi b_{max}^2} \text{ (by small-}b_T\text{ expansion)}. \tag{69}$$

The other method is to equate the derivatives with respect to b_{max}^2 at $b_T = b_{max}$; this may be more representative of how g_K affects the evolution of the cross section because this is where g_K gives a substantial correction to the cutoff $\tilde{K}$. The result gives an estimate that is a factor of 2 smaller:

$$g_2 \simeq \frac{\alpha_s(C_1/b_{max})C_F}{\pi b_{max}^2} \text{ (by derivative at } b_{max}\text{)}. \tag{70}$$

Neither method can exactly reproduce the fitted g_K, since the perturbative estimate for g_K has a different functional form than the fitted g_K; the best we can do is an approximate match.

To obtain numerical values, we use the two-loop parametrization of $\alpha_s(\mu)$ from Ref. [107] with three active flavors of quark. We make the standard choice $C_1 = 2e^{-\gamma_E}$. For the two standard values $b_{max} = 0.5\ \text{GeV}^{-1}$ and $b_{max} = 1.5\ \text{GeV}^{-1}$, we find

$$\left.\frac{C_F}{\pi}\frac{1}{b_{max}^2}\alpha_s(C_1/b_{max})\right|_{b_{max}=0.5\ \text{GeV}^{-1}} \approx 0.45\ \text{GeV}^2, \tag{71}$$

$$\left.\frac{C_F}{\pi}\frac{1}{b_{max}^2}\alpha_s(C_1/b_{max})\right|_{b_{max}=1.5\ \text{GeV}^{-1}} \approx 0.13\ \text{GeV}^2. \tag{72}$$

We compare with the measured values in Table I. We see a rough agreement, with the two methods of matching a value of g_2 to (67) giving results that bracket the measured value. We deduce that some of the work in the fits simply reproduces perturbative predictions in a region where

TABLE I.

g_2 values in quadratic parametrizations:

b_{max}	Fitted	Expansion method	Derivative method
$0.5\ \text{GeV}^{-1}$	$0.68^{+0.01}_{-0.02}\ \text{GeV}^2$	$0.9\ \text{GeV}^2$	$0.45\ \text{GeV}^2$
$1.5\ \text{GeV}^{-1}$	$0.18 \pm 0.02\ \text{GeV}^2$	$0.26\ \text{GeV}^2$	$0.13\ \text{GeV}^2$

the predictions have a useful, if approximate, validity. We also deduce that the values of $b_{\max}$ are conservative. If one wants to genuinely measure the nonperturbative part of g_K, one needs a more general parametrization, and one needs to ensure that data are used that are sensitive to higher values of b_T. We will address this issue in the next section.

Of course, the above estimates are crude and meant only to check for general consistency. At large b_T, Eq. (67) is not expected to be an accurate parametrization of $g_K(b_T; b_{\max})$. First, it is based on an extrapolation of a low-order perturbative calculation. Second, at large b_T, it depends strongly on $b_{\max}$. The complete TMD factorization formalism is $b_{\max}$ independent, and fully optimized fits should approximately reflect this if they are to account for large-b_T behavior.

Notice that the arguments for approximately quadratic behavior for $g_K(b_T)$ at small b_T equally apply to the functions $g_{j/H}$ defined in Eq. (21). This small-b_T behavior corresponds, after exponentiation, to a Gaussian for a TMD parton density.

We should emphasize that our result that perturbation theory approximately reproduces the fitted values of g_2 does not imply that it should get them exactly correct; the fitted values have also to allow for both uncalculated higher-order perturbative terms and for genuinely nonperturbative effects.

VII. LARGE-b_T BEHAVIOR OF CORRELATION FUNCTION

A. General properties

Appropriate parametrizations for the nonperturbative large-b_T behavior of TMD parton densities and of the CSS kernel $\tilde{K}$ need to be informed by the expectations from the general principles of quantum field theory. All of these quantities are certain kinds of Euclidean correlation function. Therefore, we generally expect them to decay exponentially (supplemented by a power law),

$$\frac{1}{{b_T}^{\alpha}} e^{-m b_T}, \tag{73}$$

for large distance b_T. Here, m is the mass of the lowest mass state that can be exchanged in the relevant channel. The exponent α depends on the dimensionality of the problem.

Contributions to the correlation functions arise from quantities like $\langle P|\mathrm{op}(\boldsymbol{b}_T)|X\rangle\langle X|\mathrm{op}(0)|P\rangle$. Exceptions to the property of exponential decay only arise when the theory has massless particles, or when we have vacuum matrix elements. Nonperturbative QCD has only massive states, so the exception of masslessness does not apply. The issue of vacuum matrix elements does not arise for quark and gluon densities in a hadron, but it does happen for $\tilde{K}$, at least in perturbation theory, as we will discuss later.

The decaying exponential behavior of coordinate-space parton densities at large b_T is illustrated by the calculations by Schweitzer *et al.* [53] in a chiral model. It can also be illustrated by simple perturbative calculations with massive fields.[18]

From many one-loop calculations of the TMD quantities of interest, we know that a typical integral giving b_T dependence is of the form

$$\int d^2\boldsymbol{k}_T \frac{e^{i\boldsymbol{k}_T\cdot\boldsymbol{b}_T}}{k_T^2 + m^2}. \tag{74}$$

Note that m is generally an x-dependent function of actual particle masses. We can systematically obtain the large-b_T behavior by deforming the integral over $\boldsymbol{k}_T$ into the space of complex momenta, so as to make the real part of the exponent negative. The dominant behavior is from the neighborhood of the pole at $k_T^2 = -m^2$. Simple power counting gives the dominant region of $\boldsymbol{k}_T$ as

$$\boldsymbol{k}_T \sim \left[im + O(1/b_T), O(\sqrt{m/b_T})\right], \tag{75}$$

where Euclidean coordinates $\boldsymbol{k}_T = (k_x, k_y)$ are used in a situation where $\boldsymbol{b}_T = (b_T, 0)$ is in the x direction.

Since the lowest-mass state is a property of the theory, this suggests (but does not strictly prove) that the exponential (and probably the accompanying power of b_T) is valid for the TMD parton density independently of ζ (or Q). This implies that the nonperturbative part of $\tilde{K}$ goes to a constant at large b_T, to preserve the exponential and the power-law behavior of b_T in Eq. (73). Naturally, the numerical coefficient of the exponential decreases when ζ (and hence Q) increases; this corresponds to the known qualitative behavior of TMD parton densities.

For $\tilde{K}$, this matches what is obtained from perturbative calculations with a massive gluon (with the mass mimicking the effect of massive states in nonperturbative QCD). Relevant formulas can be found in Ref. [9]. Graphs with only emission of virtual particles, i.e., with the vacuum for the final state, are independent of b_T. Graphs with particle emission into the final state decay exponentially. The commonly assumed quadratic behavior of $\tilde{K}$ and Gaussian behavior of TMD parton densities can only be an approximation, valid at best only for moderate b_T.

We therefore propose that the following constraints be applied to nonperturbative parametrizations:

(1) The nonperturbative TMD parton density at a starting value of ζ has the above exponential behavior, as

[18]Parton densities are matrix elements of gauge-invariant quark and gluon operators in hadron states. What the appropriate nonperturbative final states $|X\rangle$ should be is not entirely obvious in a theory with color confinement. This issue also gives a potential loophole in the argument for exponential decay at large b_T.

coded in a linear large-b_T behavior of the functions $g_{j/H}$ in Eq. (21).

(2) For the functions $g_{j/H}$, there should therefore be a transition from approximately quadratic low-b_T behavior to linear high-b_T behavior.

(3) The nonperturbative part of $\tilde{K}$ goes to a constant at large b_T.

(4) This constant is negative, so that the large-b_T tail is reduced as ζ increases.[19]

(5) Correspondingly, the master function A goes to zero: $A(\infty) = 0$.

Thus, a key property of the dynamics of nonperturbative QCD is the number $\tilde{K}(\infty;\mu_0)$ (at a chosen reference scale μ_0). This should be extracted from fits to data. The value at other scales is given by the RG applied in a perturbative region. Thus,

$$\tilde{K}(\infty;\mu_Q) = \tilde{K}(\infty;\mu_0) - \int_{\mu_0}^{\mu_Q} \frac{d\mu'}{\mu'} \gamma_K(\alpha_s(\mu')). \tag{76}$$

As for the function $g_K(b_T;b_{max})$, it follows from its definition that at a *minimum* it should obey:

(i) At large b_T, $g_K(b_T;b_{max})$ goes to a constant, such that $\tilde{K}(b_T;Q)$ goes to a negative constant.

(ii) At small b_T, $g_K(b_T;b_{max})$ is (approximately) quadratic, with roughly the coefficient found in Sec. VI.

Note that Tafat [52] argues that $\tilde{K}$ is proportional to b_T at large b_T. We do not adequately understand the justification of Tafat's argument. In any case, all such properties are subject to experimental testing.

B. Dependence of g_K on b_{max}

Since the exact $\tilde{K}$ is independent of b_{max}, it is useful to devise methods of parametrizing g_K to achieve this to a useful accuracy, with the aid of perturbative calculations like those in Sec. VI. This will add (approximate) constraints on g_K beyond the two just listed. In the next section, we will show a parametrization obeying the constraints.

The extra constraints are to arrange automatic b_{max} independence in certain important regions and that g_K agrees with its perturbative calculation when the perturbative calculation is valid:

(1) The form of $g_K(b_T;b_{max})$ is such that

$$\underset{b_T \ll b_{max}}{\text{asy}} \frac{d}{db_{max}} g_K(b_T;b_{max})\bigg|_{\text{parametrized}} = \underset{b_T \ll b_{max}}{\text{asy}} \frac{d}{db_{max}} g_K(b_T;b_{max})\bigg|_{\text{truncated PT}}. \tag{77}$$

Here, the notation asy (for "asymptote") denotes the extraction of the approximately quadratic behavior at small b_T, with neglect of higher power corrections. The notation "truncated PT" on the right-hand side indicates an application of the calculational methods of Sec. VI, but possibly taken to higher order.

(2) When $b_T \to \infty$, the value of $g_K(\infty;b_{max})$ is to be arranged so that $\tilde{K}(\infty)$ is independent of b_{max}.

To implement this, consider the RG-improved form (64) for $\tilde{K}$. We differentiate with respect to b_{max} and take the limit $b_T \to \infty$. We also replace the perturbative part of the right-hand side by its truncated approximation. Since the full $\tilde{K}$ is independent of b_{max}, we find

$$\frac{d}{d\ln b_{max}} g_K(b_T = \infty; b_{max}) = \left[\frac{d\tilde{K}(b_{max}; C_1/b_{max})}{d\ln b_{max}} - \gamma_K(\alpha_s(C_1/b_{max}))\right]_{\text{truncated PT}}, \tag{78}$$

which is our second requirement on the parametrization of g_K. It determines the b_{max} dependence of $g_K(b_T = \infty; b_{max})$.

C. Simple parametrization of $g_K(b_T;b_{max})$

Here, we propose a very simple example of a single-parameter form for $g_K(b_T;b_{max})$ that follows the strategy established by the above conditions. While not unique, it is useful for illustrating the basic properties that are desired for a parametrization of $g_K(b_T;b_{max})$. The general principle is that we enforce b_{max} independence in the asymptotic small- and large-b_T limits and that the perturbatively predicted small-b_T behavior is reproduced. The functional form of g_K interpolates smoothly between the small- and large-b_T regions, and one parameter determines the numerical value of $\tilde{K}$ at $b_T = \infty$. Modifications and additions to this function can be proposed later if required to get a better fit to data.

The proposed form is

$$g_K(b_T;b_{max}) = g_0(b_{max})\left(1 - \exp\left[-\frac{C_F \alpha_s(\mu_{b_*}) b_T^2}{\pi g_0(b_{max}) b_{max}^2}\right]\right), \tag{79}$$

[19]The value of $\tilde{K}(b_T,\mu)$ does depend on the RG scale μ. At large scales, the RG evolution and the positivity of the leading-order term in the anomalous dimension γ_K ensures that $\tilde{K}$ is strongly negative at large b_T. But the negativity should apply even at fairly low μ.

where $g_0(b_{\max})$ is a function of $b_{\max}$ that we will determine. Expanding in small $b_T^2 \ll b_{\max}^2$ gives

$$g_K(b_T; b_{\max}) = \frac{C_F}{\pi}\frac{b_T^2}{b_{\max}^2}\alpha_s(\mu_{b_*}) + O\left(\frac{b_T^4 C_F^2 \alpha_s(\mu_{b_*})^2}{b_{\max}^4 \pi^2 g_0(b_{\max})}\right), \tag{80}$$

which matches the $O(\alpha_s b_T^2/b_{\max}^2)$ term in Eq. (68), independently of what the function $g_0(b_{\max})$ is. Thus, g_K both agrees with the predicted small b_T behavior and gives $b_{\max}$ independence of the full $\tilde{K}$ in this region. See also Ref. [108] for more discussion of the motivation for Eq. (79).

Note that a (RG improved and truncated) perturbation expansion for $\tilde{K}(b_*)$ is only valid if $b_{\max}$ is not so large that one encounters the Landau pole before the b_* taming sets in. On the other hand, if $b_{\max}$ is too small, then an expansion of a perturbative treatment of $(-\tilde{K}(b_T, \mu) + \tilde{K}(b_*, \mu))$ around small $b_T^2/b_{\max}^2$ is sensitive to quartic b_T^4 and higher-power terms, even in regions of small b_T where a perturbative description is valid. Thus, Eq. (79) is optimized for choices of $b_{\max}$ roughly in the transition region between perturbative and nonperturbative b_T.

To enforce that $\tilde{K}(b_T = \infty)$ is independent of $b_{\max}$, we solve Eq. (78) with the leading-order expressions for γ_K and $\tilde{K}$ from Eqs. (10), (11). This gives

$$g_0(b_{\max}) = g_0(b_{\max,0}) + \frac{2C_F}{\pi}\int_{C_1/b_{\max,0}}^{C_1/b_{\max}} \frac{d\mu'}{\mu'}\alpha_s(\mu'). \tag{81}$$

Here, $b_{\max,0}$ is a reference value for b_T used to fix g_0; it is effectively an integration constant. If one chooses a particular value of $b_{\max}$ when fitting the TMD factorization formula to data, then the single numerical value $g_0(b_{\max})$ is a parameter of the fit. If a second fit were made with a different value of $b_{\max}$, then we would expect the fitted value of $g_0(b_{\max})$ at the new $b_{\max}$ to be given (approximately) in terms of the old value by Eq. (81).

At large b_T, the resulting $\tilde{K}$ goes to a constant, as planned, and at small b_T, g_K is an approximate power series in b_T^2. In principle, we may also include a treatment of $b_{\max}$ independence in higher powers of b_T in the region of intermediate b_T. Since the resulting corrections are small, we regard the current approximation as adequate for the moment. If in Eq. (79) we take the limit $g_0 \to \infty$, this is equivalent to setting $g_K(b_T; b_{\max})$ to zero and describing the entire range of $\tilde{K}(b_T; Q)$ using the cutoff form $\tilde{K}(b_*; Q)$. If, in addition, we take the limit $b_{\max} \to \infty$, and use the perturbative formula for $\tilde{K}$, then the result for $\tilde{K}(b_T; \mu_Q)$ in Eq. (64) is to use the perturbative RG-improved version for all b_T.

Improvements can of course be made by using higher-order terms in the perturbative expansion of $\tilde{K}$ and γ_K.

We now display an example of results obtained using the parametrization in Eq. (79). In Fig. 6, we have plotted the $\tilde{K}(b_T; Q)$ obtained using Eqs. (79) and (81). For the reference value of $b_{\max}$, we choose $b_{\max,0} = 1.5\ \text{GeV}^{-1}$. The goal of this paper is only to illustrate the general method, so we need only to estimate a reasonable value for $g_0(b_{\max} = 1.5\ \text{GeV}^{-1})$. To do this, we note that, if one assumes that perturbation theory is at least roughly applicable up to $b_T \sim 1.5\ \text{GeV}^{-1}$, then the leading quadratic

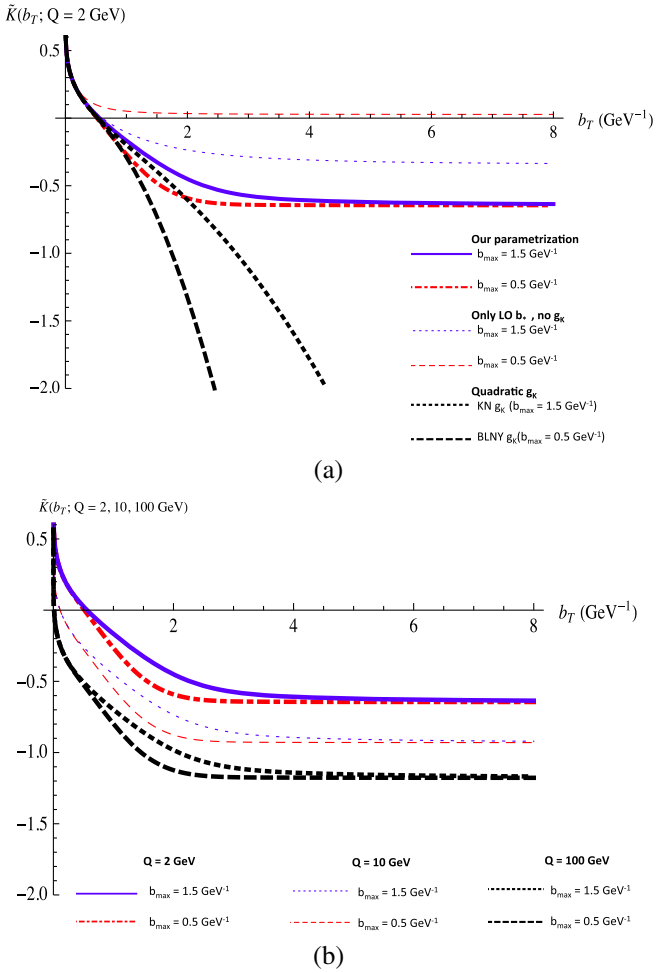


FIG. 6 (color online). (a) Comparison of different treatments of $g_K(b_T; b_{\max})$ in calculation of $\tilde{K}(b_T; Q)$ for $Q = 2.0$ GeV. The solid blue and dot-dashed red curves are the calculation using the $g_K(b_T; b_{\max})$ parametrization from Eq. (79) with, respectively, $b_{\max} = 1.5$ and $0.5\ \text{GeV}^{-1}$. The value at $b_T = \infty$ is set by $g(b_{\max} = 1.5\ \text{GeV}^{-1}) = 0.3$. The thin dotted blue and dashed red curves are the leading-order RG improved calculations of $\tilde{K}(b_*; Q)$ using $b_{\max} = 1.5$ and $0.5\ \text{GeV}^{-1}$ and *no* $g_K(b_T; b_{\max})$. The black dotted and dashed curves are using the KN and BLNY fits for $g_K(b_T; b_{\max})$ with $b_{\max} = 1.5$ and $0.5\ \text{GeV}^{-1}$ respectively. (b) Calculation of $\tilde{K}(b_T; Q)$ using Eq. (79) with $b_{\max} = 1.5$ and $0.5\ \text{GeV}^{-1}$ and several values of Q: $Q = 2$, 10, and 100 GeV. *Note the change in the meaning of the line types between graphs (a) and (b).*

terms should dominate in the expansion in Eq. (80) for the region of $b_T \lesssim 1.5\ \mathrm{GeV}^{-1}$. Thus, for the quartic terms in Eq. (80) to be small at $b_T \sim b_{\max}$, we must have

$$\frac{C_F \alpha_s(\mu_{b_*})}{\pi g_0(b_{\max})} \lesssim 1 \tag{82}$$

or

$$g_0(b_{\max} = 1.5\ \mathrm{GeV}^{-1}) \gtrsim \frac{C_F \alpha_s(C_1/1.5)}{\pi} \approx 0.3. \tag{83}$$

Therefore, for the sake of illustration, we will use $g_0(b_{\max} = 1.5\ \mathrm{GeV}^{-1}) = 0.3$. In the future, this parameter should be extracted from actual fits.

We have chosen to plot $\tilde{K}(b_T; Q)$ in Fig. 6(a) at the relatively low scale of $Q = 2.0$ GeV. The result of using $b_{\max} = 1.5\ \mathrm{GeV}^{-1}$ is shown as the solid blue curve, while the result of switching to $b_{\max} = 0.5\ \mathrm{GeV}^{-1}$ using Eq. (81) is shown as the red dot-dashed curve. For comparison, Fig. 6(a) also shows the calculation of $\tilde{K}(b_*; Q = 2\ \mathrm{GeV})$ in the leading-order RG improved approximation, but with no $g_K(b_T; b_{\max})$ term at all (thin blue dotted and red dashed curves for $b_{\max} = 1.5\ \mathrm{GeV}^{-1}$ and $b_{\max} = 0.5\ \mathrm{GeV}^{-1}$, respectively). Finally, we have shown the calculation with the KN($b_{\max} = 1.5\ \mathrm{GeV}^{-1}$) and BLNY ($b_{\max} = 0.5\ \mathrm{GeV}^{-1}$) parametrizations (thick black dotted and dashed curves), which use the quadratic form for $g_K(b_T; b_{\max})$ from Eqs. (39) and (42). Figure 6(a) illustrates how the parametrization from Eq. (79) undergoes a comparatively small variation with $b_{\max}$, relative to methods that use a quadratic $g_K(b_T; b_{\max})$ or no $g_K(b_T; b_{\max})$ at all.

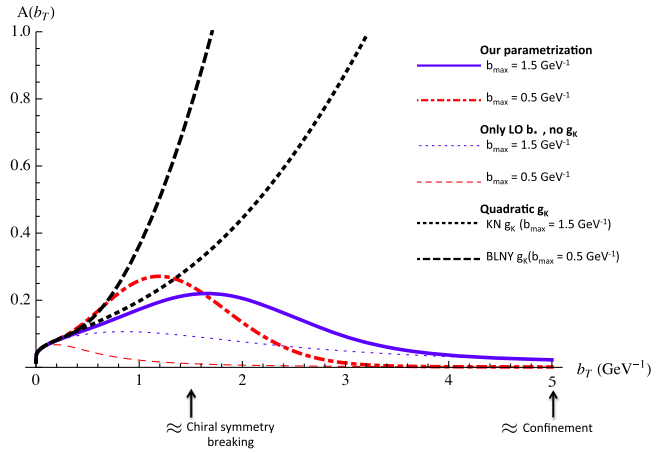
Most of the residual variation in the parametrization from Eq. (79) is confined to the intermediate region between about $b_T \sim 1.0$ and $b_T \sim 3.0\ \mathrm{GeV}^{-1}$ where one expects sensitivity to the precise details of the physics involved in the transition between a perturbative and a purely nonperturbative description of the b_T dependence. At both very large and very small b_T, there is no sensitivity to $b_{\max}$, as imposed by our construction.

The true $\tilde{K}(b_*; \mu)$ is independent of $b_{\max}$ everywhere, so the gap between, for example, the blue and red thick curves in Fig. 6(a) at around $b_T \sim 2.0\ \mathrm{GeV}^{-1}$ is symptomatic of the limitations of the simplistic parametrization of Eq. (79). Our present aim is simply to design a simple functional form that greatly reduces the $b_{\max}$ dependence. It is suggestive of improvements that can be made, by including higher-order calculations and greater knowledge of the nonperturbative effects that reduce the dependence even further.

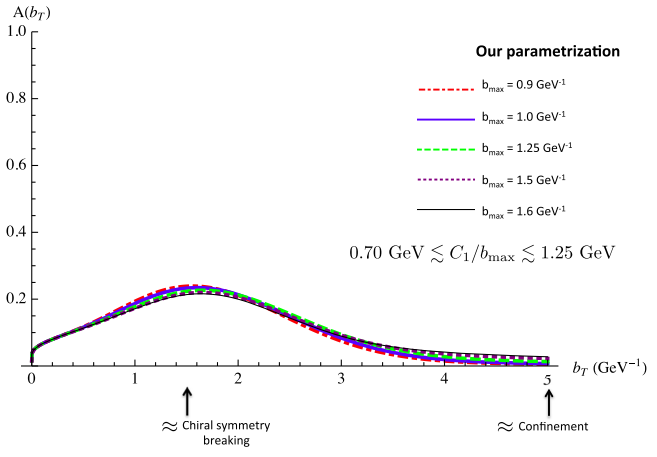
In Fig. 6(b), we have again plotted the $\tilde{K}(b_T; Q)$ obtained using Eqs. (79) and (81), for $b_{\max} = 1.5\ \mathrm{GeV}^{-1}$ and $b_{\max} = 0.5\ \mathrm{GeV}^{-1}$, but now we show the result of using $Q = 2, 10$, and 100 GeV. The variation in Q results only in a vertical shift of $\tilde{K}(b_T; Q)$. This is an exact property of the CS kernel, and the vertical shift needs to be accounted for when addressing the evolution of the normalization of the cross section.

As explained in Sect. IV, the physical consequences of $\tilde{K}(b_T; Q)$ for the evolution of the shape of transverse distributions are conveniently illustrated through the master function $A(b_T)$ defined in Eq. (30). Therefore, in Fig. 7(a) we have plotted the versions of the function $A(b_T)$ that correspond to the curves in Fig. 6(a).

In principle, like $\tilde{K}(b_T; Q)$, $A(b_T)$ should also be exactly independent of $b_{\max}$. The problem of the instability of the quadratic parametrization for $g_K(b_T; b_{\max})$ is made clear in Fig. 7(a) in the large-b_T region. There, we see dramatic differences between the results of the two standard parametrizations, with different $b_{\max}$. These differences result in large changes in the evolution with Q of the shape of the



(a)



(b)

FIG. 7 (color online). (a) The values of the function $A(b_T)$ that corresponds to $\tilde{K}(b_T)$ for the curves in Fig. 6(a). (b) The function $A(b_T)$ for a selection of $b_{\max}$ values in the range $0.7\ \mathrm{GeV} \lesssim C_1/b_{\max} \lesssim 1.25$ GeV. They use (79) and (80), with our parameter values. For $b \lesssim 2.5\ \mathrm{GeV}^{-1}$, we have checked that $A(b_T)$ varies by less than 20% with the changes in the cutoff scale corresponding to the range plot (b), so long as $0.2 \lesssim g_0(1.5\ \mathrm{GeV}^{-1}) \lesssim 0.6$.

integrand function $\tilde{W}(b_T)$. The variation in $A(b_T)$ with b_{max} obtained with the parametrization in Eq. (79) is noticeable but much smaller, and it approaches zero at very large b_T.

Note that the variation of b_{max} in Fig. 7(a) corresponds to a rather large variation in the lower cutoff on the hard scale: $0.75\ \text{GeV} \lesssim C_1/b_{max} \lesssim 2.25\ \text{GeV}$. For the smallest values of b_{max}, the expansion of the perturbative expression for $(-\tilde{K}(b_T,\mu) + \tilde{K}(b_*,\mu))$ is sensitive to quartic powers of b_T/b_{max} (and higher powers) even for quite small values of b_T.

If we instead restrict to a slightly narrower window of $0.70\ \text{GeV} \lesssim C_1/b_{max} \lesssim 1.25\ \text{GeV}$, sensitivity to b_{max} becomes essentially negligible, as illustrated by Fig. 7(b).

D. Evolution of the normalization

Many discussions of TMD evolution focus on the variation with Q of the shape of transverse-momentum distributions. Among the reasons are that this is a particularly recognizable effect of evolution, while accurate measurements of normalizations are more difficult. In addition, the quadratic form commonly used for $g_K(b_T)$ predicts important shape changes even at low Q. However, it is also useful to ask how the normalization evolves.

Our proposal that $\tilde{K}(b_T)$ goes to a constant as $b_T \to \infty$ changes this situation in an interesting way. The effect of the nonperturbative part of $\tilde{K}$ is now to change the normalization of the Fourier transformed cross section $\tilde{W}$ at large b_T instead of changing its shape dramatically. The variation of normalization with Q is approximately a power law as we now show. Measurement of this power law would be particularly useful for determining the value of $\tilde{K}$ at $b_T = \infty$.

It follows from Eq. (22) that the shape of the b_T-space integrand at large b_T is determined by the functions $g_{j/H}(b_T)$, which can be regarded as parametrizing the intrinsic transverse-momentum distribution of the quarks.

By replacing $\tilde{K}(b_T)$ by its asymptotic value $\tilde{K}(\infty)$, we now show that the effect of evolution is to give the normalization an approximate power dependence on Q. This can be seen from Eq. (25), where for large b_T

$$\frac{\partial \ln \tilde{W}(\boldsymbol{b}_T, Q, x_A, x_B)}{\partial \ln Q^2} \simeq \tilde{K}(\infty;\mu_0) + G(\alpha_s(\mu_Q), Q/\mu_Q) - \int_{\mu_0}^{\mu_Q} \frac{d\mu'}{\mu'} \gamma_K(\alpha_s(\mu')). \tag{84}$$

The variation of the right-hand side with $\ln Q$ is an order-$\alpha_s(Q)$ effect. If to a first approximation we neglect this variation, we find a power law,

$$\tilde{W}(b_T, Q, x_A, x_B) \simeq \tilde{W}(b_T, Q_0, x_A, x_B)\left(\frac{Q_0^2}{Q^2}\right)^a, \tag{85}$$

with

$$\begin{aligned} a &= -K(\infty;Q_0) - G(\alpha_s(Q_0), Q/Q_0) \\ &= g_K(\infty;b_{max}) - \tilde{K}(b_{max};C_1/b_{max}) - G(\alpha_s(Q_0), Q/Q_0) \\ &\quad + \int_{C_1/b_{max}}^{Q_0} \frac{d\mu'}{\mu'} \gamma_K(\alpha_s(\mu')). \end{aligned} \tag{86}$$

As Q increases, the right-hand side of (84) becomes more negative, so the decrease of the normalization becomes even stronger. At small b_T, the situation is totally different, of course, as can be read off the plots in Fig. 6, for example. The power law only applies at large b_T.

At low values of Q, around a GeV or two at the lower limit of TMD factorization's applicability, the power law just derived gives a corresponding decrease in the cross section at small transverse momentum. The approximate Bjorken scaling of the total cross section is restored by a compensating change at larger transverse momentum, which comes from relatively small b_T.

VIII. SUMMARY AND CONCLUSION

There is a wide variety of sometimes apparently contradictory methods and results in the theoretical treatment and analysis of transverse momentum dependent cross sections, especially as regards the nature and importance of nonperturbative contributions from large transverse distances. We presented a systematic analysis of the issues. The basis of the logic is a TMD factorization theorem, together with the associated evolution equations, etc., for the TMD parton densities (and fragmentation functions). Several different forms of solution were presented, each emphasizing particular aspects of the physics. These have mostly been seen in the literature before, but here they are unified by the link to scale-dependent TMD densities as the foundation of the reasoning.

The evolution of the shape of TMD functions is primarily governed by a function $\tilde{K}(b_T,\mu)$, which appears under different names in some SCET-based formalisms. Although this function is strongly universal, it is nontrivial to gain a unified view of it.

From an experimental point of view, the range of b_T dominantly probed depends on the kinematic region of the data so that no single experiment can measure or test $\tilde{K}$ for all b_T. Moreover, $\tilde{K}$ is scale dependent. Although the scale dependence is just a perturbatively calculable upward shift as μ increases, independent of b_T, it does add an important complication in measuring $\tilde{K}$ and in testing measured and predicted values.

We proposed a master function $A(b_T)$ as a measure of the evolution of the shape of TMD functions. Unlike $\tilde{K}$, it is independent of both scheme and scale (but is related to $\tilde{K}$ by a derivative). We showed how the function $A(b_T)$ can be used to diagnose disagreements between different

methods and approximations for TMD cross sections. We suggest that the results of calculations and fits should include a presentation of the resulting values of $A(b_T)$, and one aim should be to find the values of $A(b_T)$, in addition to $\tilde{K}(b_T; Q)$, for all b_T, as an important property of QCD.

A further complication concerns the predictability of $\tilde{K}$. At low b_T, perturbative calculations, supplemented by RG improvement or by explicit resummation, are accurate. As b_T is increased, these predictions gradually become less accurate, and at large enough b_T (beyond about 2 or 3 GeV^{-1}), $\tilde{K}$ becomes essentially nonperturbative. We pointed out that the use of resummation instead of RG improvement does not at all change this situation; the accuracy of either kind of perturbatively based calculation is governed by the value of $\alpha_s(1/b_T)$.

Sensitivity to large b_T in $A(b_T)$, for its part, opens new avenues of opportunity for probing nonperturbative properties of QCD. The relevant objects are the vacuum expectation values of Wilson loops, which are basic nonperturbative subjects of interest in, for example, lattice QCD methods and have already attracted interest in TMD studies [47–49].

When Q is increased, the dominant range of b_T needed to calculate the cross section shifts to ever smaller values. Thus, even dramatic differences in the form of $\tilde{K}$ at large b_T typically have little effect on TMD cross sections at large enough Q because, at large b_T, the b_T integrand is exponentiated to a small value. There is a stability to the evolution toward larger Q. In contrast, backward evolution in Q is unstable. It is found that evolution with Q is in fact considerably less rapid at low Q than is given by a backward evolution from standard fits to Drell–Yan data.

(Even at relatively large Q, however, some knowledge of the large-b_T behavior may be desirable if very high degrees of precision are needed.)

From general principles about correlation functions at large distances, we argued that $\tilde{K}(b_T, \mu)$ should go to a constant as $b_T \to \infty$. This contrasts with the widely used quadratic parametrizations. We proposed a new form for a one-parameter approximation for interpolating between the perturbative result for $\tilde{K}$ at small b_T and a constant at large b_T. An important task is to fit the constant from data that are sensitive to evolution with b_T in the range of 3 to 4 GeV^{-1} upward. Our parametrization is intended to approximately agree with standard Drell-Yan fits in the region of b_T to which they are sensitive, while giving slower evolution of the shape of the cross section at lower Q. It is also arranged to give automatically weak sensitivity to CSS's parameter $b_{\max}$. Of course, our parametrization can be supplemented by higher-order perturbative calculations where available and by extra parametrized functions for the nonperturbative part. But the new parametrization is designed so that these corrections should be relatively weak.

It is worth emphasizing that important and interesting physics is encoded in both the large and small b_T regions of cross sections like Eq. (1). Depending on the specific underlying motivation for applying TMD factorization, different ranges of b_T may be of greater or lesser interest, and the relative importance of small and large-b_T contributions depends on the size of Q. However, a good TMD factorization formalism incorporates both types of behavior and smoothly relates a diverse range of different observables with different degrees of relative sensitivity to large and small b_T.

As a conclusion, we propose that one important aim of theoretical and phenomenological work in QCD should be to obtain accurate values of $A(b_T)$ and also of $\tilde{K}(b_T, \mu)$ over a wide range of b_T. The results will have a similar significance to the well-known global fits of collinear parton densities and fragmentation functions, which provide definitive values, and uncertainties, for these functions.[20] We propose that $\tilde{K}$ goes to a constant at large b_T instead of being quadratic in b_T.

One important benefit of presenting results directly for $A(b_T)$ and $\tilde{K}(b_T, \mu)$ should be a much better understanding of how the transition occurs from perturbative behavior at small b_T to nonperturbative behavior at large b_T. This transition should be gradual rather than very sudden. The strongly universal nature of any nonperturbative contribution to $\tilde{K}$, mentioned earlier, gives broad implications for these results.

ACKNOWLEDGMENTS

T. R. is supported in part by the National Science Foundation under Grants No. PHY-0969739 and No. PHY-1316617. T. R. also acknowledges support from the University of Michigan and the Lightner-Sams Foundation. J. C. and T. R. are supported by the U.S. Department of Energy under Grant No. DE-SC0008745. This work was also supported by the DOE Contract No. DE-AC05-06OR23177, under which Jefferson Science Associates, LLC, operates Jefferson Lab. We acknowledge useful conversations with D. Boer, L. Gamberg, A. Idilbi, P. Nadolsky, and G. Sterman.

[20] In the range of variables where they are actually determined, of course.

[1] F. Landry, R. Brock, P. M. Nadolsky, and C.-P. Yuan, Phys. Rev. D **67**, 073016 (2003).
[2] A. V. Konychev and P. M. Nadolsky, Phys. Lett. B **633**, 710 (2006).
[3] J.-W. Qiu and X.-F. Zhang, Phys. Rev. D **63**, 114011 (2001).
[4] J.-W. Qiu and X.-F. Zhang, arXiv:hep-ph/0205115.
[5] G. I. Fai, J.-W. Qiu, and X.-F. Zhang, Phys. Lett. B **567**, 243 (2003).
[6] J. C. Collins and D. E. Soper, Nucl. Phys. **B193**, 381 (1981); **B213**, 545(E) (1983).
[7] J. C. Collins and D. E. Soper, Nucl. Phys. **B194**, 445 (1982).
[8] J. C. Collins, D. E. Soper, and G. Sterman, Nucl. Phys. **B250**, 199 (1985).
[9] J. C. Collins, *Foundations of Perturbative QCD* (Cambridge University Press, Cambridge, England, 2011).
[10] T. Becher and M. Neubert, Eur. Phys. J. C **71**, 1665 (2011).
[11] M. G. Echevarría, A. Idilbi, A. Schäfer, and I. Scimemi, Eur. Phys. J. C **73**, 2636 (2013).
[12] P. Sun and F. Yuan, Phys. Rev. D **88**, 034016 (2013).
[13] P. Sun and F. Yuan, Phys. Rev. D **88**, 114012 (2013).
[14] V. Barone *et al.* (PAX Collaboration), arXiv:hep-ex/0505054.
[15] M. Liu, X. Jiang, D. Crabb, J. Chen, and M. Bai, J. Phys. Conf. Ser. **295**, 012164 (2011).
[16] W. Erni *et al.* (PANDA Collaboration), arXiv:0903.3905.
[17] J. C. Collins, Phys. Lett. B **536**, 43 (2002).
[18] D. W. Sivers, Phys. Rev. D **41**, 83 (1990).
[19] D. W. Sivers, Phys. Rev. D **43**, 261 (1991).
[20] A. Airapetian *et al.* (HERMES Collaboration), Phys. Rev. Lett. **103**, 152002 (2009).
[21] A. Airapetian *et al.* (HERMES Collaboration), Phys. Lett. B **693**, 11 (2010).
[22] R. Seidl *et al.* (Belle), Phys. Rev. Lett. **96**, 232002 (2006).
[23] R. Seidl *et al.* (Belle), Phys. Rev. D **78**, 032011 (2008).
[24] J. Lees *et al.* (*BABAR* Collaboration), Phys. Rev. D **90**, 052003 (2014).
[25] M. G. Echevarría, A. Idilbi, and I. Scimemi, J. High Energy Phys. 07 (2012) 002.
[26] S. M. Aybat, A. Prokudin, and T. C. Rogers, Phys. Rev. Lett. **108**, 242003 (2012).
[27] F. Bradamante (COMPASS Collaboration), Nuovo Cimento C **035N2**, 107 (2012).
[28] C. Aidala, B. Field, L. Gamberg, and T. Rogers, Phys. Rev. D **89**, 094002 (2014).
[29] X.-D. Ji, J.-P. Ma, and F. Yuan, Phys. Rev. D **71**, 034005 (2005).
[30] S. Mantry and F. Petriello, Phys. Rev. D **81**, 093007 (2010).
[31] S. Mantry and F. Petriello, Phys. Rev. D **84**, 014030 (2011).
[32] D. Boer, Nucl. Phys. **B806**, 23 (2009).
[33] D. Boer, Nucl. Phys. **B874**, 217 (2013).
[34] H. Avakian, A. V. Efremov, P. Schweitzer, O. V. Teryaev, F. Yuan, and P. Zavada, Mod. Phys. Lett. A **24**, 2995 (2009).
[35] A. Courtoy, S. Scopetta, and V. Vento, Phys. Rev. D **79**, 074001 (2009).
[36] C. Lorcé and B. Pasquini, Phys. Rev. D **84**, 034039 (2011).
[37] M. Burkardt, Phys. Rev. D **88**, 014014 (2013).
[38] M. Anselmino, M. Boglione, and S. Melis, arXiv:1209.1541.
[39] A. Signori, A. Bacchetta, M. Radici, and G. Schnell, J. High Energy Phys. 11 (2013) 194.
[40] M. G. Echevarria, A. Idilbi, Z. -B. Kang, and I. Vitev, Phys. Rev. D **89**, 074013 (2014).
[41] G. Bozzi, S. Catani, G. Ferrera, D. de Florian, and M. Grazzini, Phys. Lett. B **696**, 207 (2011).
[42] A. Banfi, M. Dasgupta, and S. Marzani, Phys. Lett. B **701**, 75 (2011).
[43] M. Guzzi, P. M. Nadolsky, and B. Wang, Phys. Rev. D **90**, 014030 (2014).
[44] P. Meade, H. Ramani, and M. Zeng, Phys. Rev. D **90**, 114006 (2014).
[45] R. Lopes de Sa, Ph.D. Thesis, Stony Brook University, 2013).
[46] P. M. Nadolsky, AIP Conf. Proc. **753**, 158 (2005).
[47] B. U. Musch, P. Hägler, J. W. Negele, and A. Schäfer, Phys. Rev. D **83**, 094507 (2011).
[48] B. Musch, P. Hägler, M. Engelhardt, J. Negele, and A. Schäfer, Phys. Rev. D **85**, 094510 (2012).
[49] X. Ji, P. Sun, X. Xiong, and F. Yuan, arXiv:1405.7640.
[50] A. Kulesza, G. F. Sterman, and W. Vogelsang, Phys. Rev. D **66**, 014011 (2002).
[51] G. P. Korchemsky and G. F. Sterman, Nucl. Phys. **B437**, 415 (1995).
[52] S. Tafat, J. High Energy Phys. 05 (2001) 004.
[53] P. Schweitzer, M. Strikman, and C. Weiss, J. High Energy Phys. 01 (2013) 163.
[54] J. C. Collins and D. E. Soper, Nucl. Phys. **B284**, 253 (1987).
[55] P. Sun, J. Isaacson, C. P. Yuan, and F. Yuan, arXiv:1406.3073.
[56] P. M. Nadolsky, N. Kidonakis, F. Olness, and C. Yuan, Phys. Rev. D **67**, 074015 (2003).
[57] R. Meng, F. I. Olness, and D. E. Soper, Nucl. Phys. **B371**, 79 (1992).
[58] R. Meng, F. I. Olness, and D. E. Soper, Phys. Rev. D **54**, 1919 (1996).
[59] P. Nadolsky, D. R. Stump, and C. P. Yuan, Phys. Rev. D **61**, 014003 (1999).
[60] J. C. Collins and D. E. Soper, Phys. Rev. D **16**, 2219 (1977).
[61] C. S. Lam and W.-K. Tung, Phys. Rev. D **18**, 2447 (1978).
[62] E. Mirkes, Nucl. Phys. **B387**, 3 (1992).
[63] J. P. Ralston and D. E. Soper, Nucl. Phys. **B152**, 109 (1979).
[64] J. T. Donohue and S. A. Gottlieb, Phys. Rev. D **23**, 2577 (1981).
[65] C. Balázs, J.-W. Qiu, and C. Yuan, Phys. Lett. B **355**, 548 (1995).
[66] C. Balázs and C. P. Yuan, Phys. Rev. D **56**, 5558 (1997).
[67] S. Arnold, A. Metz, and M. Schlegel, Phys. Rev. D **79**, 034005 (2009).
[68] A. Bacchetta, M. Diehl, K. Goeke, A. Metz, P. J. Mulders, and M. Schlegel, J. High Energy Phys. 02 (2007) 093.
[69] P. J. Mulders and R. D. Tangerman, Nucl. Phys. **B461**, 197 (1996).
[70] D. Boer, Nucl. Phys. **B603**, 195 (2001).

[71] S. M. Aybat, J. C. Collins, J.-W. Qiu, and T. C. Rogers, Phys. Rev. D **85**, 034043 (2012).
[72] S. Catani, L. Cieri, D. de Florian, G. Ferrera, and M. Grazzini, Eur. Phys. J. C **72**, 2195 (2012).
[73] T. Gehrmann, T. Lubbert, and L. L. Yang, Phys. Rev. Lett. **109**, 242003 (2012).
[74] T. Gehrmann, T. Luebbert, and L. L. Yang, J. High Energy Phys. 06 (2014) 155.
[75] S. Moch, J. Vermaseren, and A. Vogt, J. High Energy Phys. 08 (2005) 049.
[76] A. Grozin, J. M. Henn, G. P. Korchemsky, and P. Marquard, Phys. Rev. Lett. **114**, 062006 (2015).
[77] Z.-B. Kang, B.-W. Xiao, and F. Yuan, Phys. Rev. Lett. **107**, 152002 (2011).
[78] J.-W. Qiu and G. Sterman, Nucl. Phys. **B353**, 105 (1991).
[79] J.-W. Qiu and G. Sterman, Nucl. Phys. **B353**, 137 (1991).
[80] M. Anselmino, M. Boglione, and F. Murgia, Phys. Lett. B **362**, 164 (1995).
[81] M. Anselmino, M. Boglione, U. D'Alesio, E. Leader, and F. Murgia, Phys. Rev. D **70**, 074025 (2004).
[82] M. Anselmino, M. Boglione, U. D'Alesio, A. Kotzinian, F. Murgia, and A. Prokudin, Phys. Rev. D **72**, 094007 (2005).
[83] J. Collins, A. Efremov, K. Goeke, S. Menzel, A. Metz, and P. Schweitzer, Phys. Rev. D **73**, 014021 (2006).
[84] J. Collins, A. Efremov, K. Goeke, M. Perdekamp, S. Menzel, B. Meredith, A. Metz, and P. Schweitzer, Phys. Rev. D **73**, 094023 (2006).
[85] M. Anselmino *et al.*, arXiv:hep-ph/0511017.
[86] M. Anselmino, M. Boglione, U. D'Alesio, S. Melis, F. Murgia, and A. Prokudin, Phys. Rev. D **79**, 054010 (2009).
[87] P. Schweitzer, T. Teckentrup, and A. Metz, Phys. Rev. D **81**, 094019 (2010).
[88] M. Anselmino, M. Boglione, U. D'Alesio, S. Melis, F. Murgia, and A. Prokudin, Phys. Rev. D **87**, 094019 (2013).
[89] M. Anselmino, M. Boglione, U. D'Alesio, S. Melis, F. Murgia, and A. Prokudin, Phys. Rev. D **88**, 054023 (2013).
[90] J. Owens, A. Accardi, and W. Melnitchouk, Phys. Rev. D **87**, 094012 (2013).
[91] F. Hautmann, H. Jung, M. Krämer, P. J. Mulders, E. R. Nocera, T. C. Rogers, and A. Signori, Eur. Phys. J. C **74**, 3220 (2014).
[92] J.-W. Qiu and X.-F. Zhang, Phys. Rev. Lett. **86**, 2724 (2001).
[93] S. Catani, L. Trentadue, G. Turnock, and B. Webber, Nucl. Phys. **B407**, 3 (1993).
[94] A. Bacchetta and A. Prokudin, Nucl. Phys. **B875**, 536 (2013).
[95] J. Kodaira and L. Trentadue, Phys. Lett. B **112**, 66 (1982).
[96] C. Davies and W. J. Stirling, Nucl. Phys. **B244**, 337 (1984).
[97] G. Parisi and R. Petronzio, Nucl. Phys. **B154**, 427 (1979).
[98] J. C. Collins and T. C. Rogers, Phys. Rev. D **87**, 034018 (2013).
[99] J. Collins, arXiv:1212.5974.
[100] D. Boer, Int. J. Mod. Phys. Conf. Ser. **25**, 1460004 (2014).
[101] D. Boer and W. J. den Dunnen, Nucl. Phys. **B886**, 421 (2014).
[102] M. Anselmino *et al.*, arXiv:1107.4446.
[103] M. Anselmino, M. Boglione, and S. Melis, Phys. Rev. D **86**, 014028 (2012).
[104] S. Bethke, Nucl. Phys. B, Proc. Suppl. **234**, 229 (2013).
[105] D. Boer, L. Gamberg, B. Musch, and A. Prokudin, J. High Energy Phys. 10 (2011) 021.
[106] U. D'Alesio, M. G. Echevarria, S. Melis, and I. Scimemi, J. High Energy Phys. 11 (2014) 098.
[107] S. Bethke, Eur. Phys. J. C **64**, 689 (2009).
[108] J. Collins, EPJ Web Conf. **85**, 01002 (2015).