Heavy quark expansion and preasymptotic corrections to decay widths in the 't Hooft model

Ikaros Bigi

Department of Physics, University of Notre Dame du Lac, Notre Dame, Indiana 46556

Nikolai Uraltsev

Department of Physics, University of Notre Dame du Lac, Notre Dame, Indiana 46556
and Petersburg Nuclear Physics Institute, Gatchina, St. Petersburg, 188350, Russia
(Received 12 February 1999; published 12 November 1999)

We address nonperturbative power corrections to inclusive decay widths of heavy flavor hadrons in the context of the 't Hooft model (two-dimensional QCD at $N_c \to \infty$), with the emphasis on the ''spectator-dependent'' effects, i.e., those sensitive to the flavor of the spectator. The summation of exclusive widths is performed analytically using the 't Hooft equation. We show that the $1/m_Q$ expansion of both the weak annihilation and Pauli interference widths coincides with the OPE predictions to the computed orders. Violation of local duality in the inclusive widths is quantified, and the new example is identified where the OPE prediction and the actual effect are completely saturated by a single final state. The qualitative aspects of quark hadronization emerging from the analysis in the 't Hooft model are discussed. Certain aspects of the summation of spectator-independent hadronic weak decay widths are given in more detail, which were not spelled out previously. We also give some useful details of the $1/m_Q$ expansion in the 't Hooft model.
[S0556-2821(99)02121-9]

PACS number(s): 12.38.Aw, 12.39.Hg, 13.35.Dx, 23.70.+j

I. INTRODUCTION

The decays of heavy flavor hadrons H_Q are shaped by nonperturbative strong interaction dynamics which, at first sight, completely obscures most of the properties of the underlying weak interactions self-manifest at the quark level. Suffice it to say that the actual hadrons, rather than quarks are observed in the final state. The actual dynamics of confinement in QCD to a large extent remains mysterious. Nevertheless, significant progress has been achieved in describing heavy flavor decays applying the formalism based on Wilson's operator product expansion (OPE) [1,2]. In particular, it became possible to quantify the effects of the confining domain on the inclusive decay rates. This theory is in the mature stage now (see Refs. [3,4], and references therein).

Among the general statements derived for the heavy quark decays, we mention the following.

The absence of $\Lambda_{\rm QCD}/m_Q$ corrections to all types of fully inclusive decay widths [5,6].[1]

The leading nonperturbative corrections arise in order $1/m_Q^2$ and are given by the expectation values μ_π^2, μ_G^2 of the two heavy quark operators, kinetic and chromomagnetic. While the first effect is universal amounting to the correction $-\mu_\pi^2/2m_Q^2$, the Wilson coefficient for the second one depends on the considered process. Both, however, are insensitive to the flavor of the spectator(s) (''flavor-independent'' corrections) [5,6].

The widths are determined by the short-distance running quark masses $m_Q(\mu)$ [8]. These are shielded against uncontrollable corrections from the infrared domain which would otherwise bring in uncertainty $\delta m_Q/m_Q \sim \Lambda_{\rm QCD}/m_Q$.

The effects sensing the spectator flavor *per se*, emerge at the level $1/m_Q^3$ [9,10,5]. They are conventionally called weak annihilation (WA) in mesons, weak scattering (WS) in baryons, and Pauli interference (PI) in both systems. Their magnitudes are given by the expectation values of local four-quark operators.[2]

For practical applications we should keep the following in mind (for a recent dedicated discussion, see Refs. [11,12, 4]): Good control over the perturbative expansion must be established to address power-suppressed effects; the consistent Wilsonian OPE requires introducing the separation of ''hard'' and ''soft'' scales, with the borderline μ serving as the normalization point in the effective theory; one has to allow, in principle, for short-distance (small-coupling regime) effects that are not directly expandable in the powers of the strong coupling; account must be taken of the fact that the OPE power series are only asymptotic [13], and reconstructing from them the actual Minkowskian observable, generally speaking, potentially leaves out the oscillating (sign-alternating) contributions suppressed, in a certain interval of energies, by only a power of the high momentum scale. This is compounded by the fact that in practice one can typically determine only the first few terms in the power expansion.

The last item in the list is behind the phenomenon of

---

[1]The OPE for the inclusive widths, actually, is *a priori* governed by the energy release rather than literally m_Q [7]. For simplicity, we do not distinguish between them parametrically unless it becomes essential.

[2]In the context of the heavy quark expansion, local operators have a more narrow meaning denoting the generic operator of the form $\bar{Q}OQ$, with O being a local operator involving only light degrees of freedom.
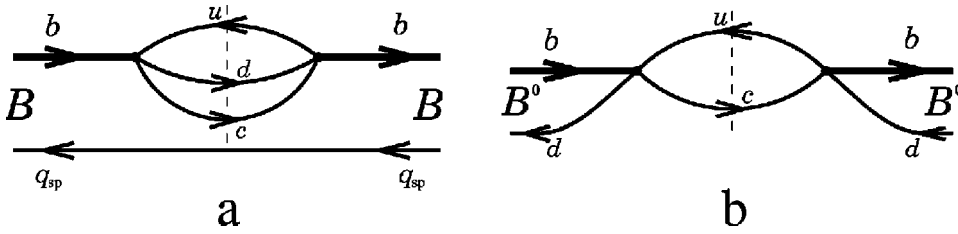
FIG. 1. (a) Quark diagram describing the leading quasifree term in the decay width. (b) Bare quark diagram for WA.

violation of local parton-hadron duality; in many cases it is among the primary factors potentially limiting the accuracy of the theoretical expansion. In the actual QCD these technical complications are often interrelated. Therefore, it is instructive to investigate the OPE in a simplified setting where these elements can be disentangled. As explained in Ref. [11], this is achieved in QCD formulated in $1+1$ dimensions. Additionally, employing the limit $N_c \to \infty$ one arrives at the exactly solvable 't Hooft model where all the features can be traced explicitly. It is important that the 't Hooft model maintains the crucial feature of QCD—quark confinement—which is often believed to be tightly related to the violation of local duality. Yet in $1+1$ dimensions confinement appears already in the perturbative expansion.

The 't Hooft model has often been used as a theoretical laboratory for exploring various field-theoretic approaches [14]. Most recently the OPE for the inclusive widths and the related sum rules in the heavy flavor transitions [15] were analytically studied in Ref. [11], where a perfect match between the OPE power expansion and the actual asymptotics of the widths was found. The known high-energy asymptotics of the spectrum in the model allowed us to determine the violation of local duality in the inclusive widths at large m_Q. As expected, it obeyed the general constraints imposed by the OPE. Moreover, at least in the framework of this simplified model, the main features of duality violation could be inferred from the parton-level analysis itself, the working tools of the OPE. The suppression of the duality-violating component in Γ_{H_Q} was found to be rather strong, with the power of $1/m_Q$, however, depending essentially on the particulars of the considered model and the process.

Reference [11] focused on flavor-independent effects. To this end it was assumed that the spectator quark q_{sp} has a flavor different from all quarks in the final state, thus ruling out both WA and PI. The OPE analysis of these effects is also straightforward. Nevertheless, they may be of independent interest for several reasons.

First, WA and PI represent power-suppressed and thus purely preasymptotic effect. In such a situation one may expect a later onset of duality and more significant violations of local duality. Since the above effects are numerically enhanced for actual charm and beauty hadrons, studying this question has practical importance.

Another reason to look more closely at the spectator-dependent effects is related to the color-flow considerations usually employed in the context of the large-N_c perspective on QCD, and interpreting the OPE predictions in terms of hadronic states. In the case of the quasifree quark decay width or the WA processes one finds a rather straightforward correspondence between the OPE expressions and the had-

ronic contributions already in the simplest quark picture where quark allocation over the final state hadrons is unambiguous (such a description is expected to hold at $N_c \to \infty$). Let us consider, for example, the free parton decay diagram Fig. 1(a). The $\bar{u}d$ pair is in a colorless state and typically has a large momentum $q^2 \sim m_b^2$ flowing through it. It is then naturally dual to the contributions from the hadronic resonances in the $V-A$ channel (in particular when integrated over q^2), much in the same way as in e^+e^- annihilation or hadronic τ decays. The c quark together with the spectator antiquark produces another string of hadronic excitation. Furthermore, the interaction between these two hadronic clusters can naturally be small at large m_b. WA, Fig. 1(b), looks even simpler in this respect; we will discuss it in detail later on.

The hadronic picture of the processes underlying PI *a priori* is less obvious, Fig. 2. The $\bar{u}$ quark produced in the decay must be slow to interfere with the valence $\bar{u}$. The large momentum here flows through the diquark loop (cd) which therefore represents the ''hard core'' of the process. The practical OPE, effectively, prescribes to replace the propagation of this diquark by a nearly free di-fermion loop, which amounts to evaluating its absorptive part as if the production of the free quarks was considered. Basically, no distinction emerges compared to the color-singlet $\bar{q}q'$ pairs in Fig. 1. This may leave one with the feeling of discomfort, for no colored states (in particular, with the diquark content) is present in the physical spectrum. In other words, the diquark configuration *per se* cannot be dual to the mesonic states at any arbitrary large momentum transfer.

Alternatively, one can combine a ''hard'' quark from the loop in Fig. 2 with the slow spectator antiquark to have a color-singlet meson-like configuration. However, such a pair naively is not ''hard'': at least in the perturbative partonic picture with $p_{sp} \sim m_{sp} \to 0$ its invariant mass vanishes irrespective of m_b. While such reasoning is clearly of the hand-waving variety, it illustrates nevertheless that interference effects are more subtle.

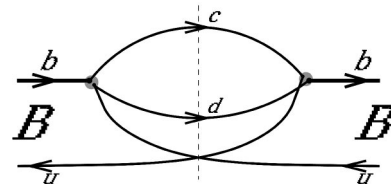A more troublesome feature of the interference is also

FIG. 2. Quark diagram for PI in B meson decays. The weak vertices are broken to show the color flow yielding the leading-N_c contribution.

illustrated by the observation made in the early 1990s by Shifman [16]. He considered a more general scenario with both charged- and neutral-current type interactions, as described by the effective weak Lagrangian

$$\mathcal{L} = -\frac{G}{\sqrt{2}}\{a_1[\bar{c}\gamma_\mu(1-\gamma_5)b][\bar{d}\gamma^\mu(1-\gamma_5)u]$$

$$+ a_2[\bar{d}\gamma_\mu(1-\gamma_5)b][\bar{c}\gamma^\mu(1-\gamma_5)u]\}+\text{H.c.} \quad (1)$$

The leading (rather than the power-suppressed spectator-dependent) width was addressed. The parton result depends on the color factors a_1, a_2 in the following way:

$$\Gamma_Q \sim N_c\left(a_1^2 + a_2^2 + \frac{2}{N_c}a_1 a_2\right). \quad (2)$$

On the other hand, the usual counting rules yield the decay amplitudes into the two-meson final states in the form

$$\mathcal{M} \propto \sqrt{N_c}\left(a_1 + \frac{1}{N_c}a_2\right) \quad \text{for ``}D_s\pi^-\text{'' states},$$

$$\mathcal{M} \propto \sqrt{N_c}\left(a_2 + \frac{1}{N_c}a_1\right) \quad \text{for ``}DK\text{'' states}, \quad (3)$$

where, for illustrative purposes, we call q_{sp} the strange quark to simplify distinguishing between the two different ways to pair the quarks into mesons. (Since we discuss the leading free-parton amplitude, the flavor of the spectator is chosen to be different from all other quarks in the process.) Adopting the rules (3) one gets

$$\Gamma_Q \sim N_c\left(a_1^2 + a_2^2 + \frac{4}{N_c}a_1 a_2\right) \quad (4)$$

more or less independently of the dynamics. While the dependence for the terms $\sim a_1^2$ and $\sim a_2^2$ is reproduced, there is a clear mismatch between Eqs. (2) and (4) in the term describing the interference of the two different color amplitudes [16].

There is little doubt that the formal OPE asymptotics must work at arbitrary N_c. The arguments above might suggest, however, that the onset of duality is delayed for suppressed effects, for example, grow with N_c.

In reality, we do not think that there is convincing evidence supporting such reservations about applying the OPE to flavor-dependent corrections. To provide an additional justification, we have explicitly analyzed both PI and WA in the 't Hooft model. We have found complete consistency with the OPE, with the onset of duality largely independent of the details. As a matter of fact, the parton-deduced OPE expression for PI appears to be *exact* in the chiral limit when all involved quarks (but Q) are massless. The resolution of the above paradoxes emerges in a rather straightforward manner as well; we will comment on them in subsequent sections.

We note that we disagree with the claims of the recent paper [17] which found a mismatch between the actual WA

width and the OPE-based prediction, relying on numerical computations. We have determined the leading effect analytically and showed it to coincide with the OPE result. We comment on the apparent drawbacks in the analysis of Ref. [17] in Sec. VI.

The paper is organized as follows. After this introduction, in Sec. II we sketch the aspects of the 't Hooft model important for addressing weak decays. In Sec. III we analyze the effects of WA at $N_c \to \infty$ and analytically compute the large-m_Q asymptotics of the corresponding heavy meson weak decay width, with technical details given in Appendix A. Section IV addresses PI; we analytically compute this width up to terms Γ_{H_Q}/m_Q^3 and find full agreement with the expressions obtained in the OPE. The effects of local duality violation at large m_Q are quantified. The special case—with massless final-state quarks—is identified where duality violation is totally absent from the spectator-dependent part of the width. In Sec. V we present a more detailed derivation of the total decay width up to corrections $1/m_Q^3$ explicitly accounting for nonzero light-quark masses, to demonstrate consistency with the OPE (a detailed description of this analysis had been omitted from Ref. [11]). Section VI comments on the analyses which have claimed observing inapplicability of the OPE predictions based on numerical computations. Section VII comprises conclusions and overlook and outlines our perspective on the problem of OPE and duality violation in the decays of heavy flavor hadrons.

Most technicalities are relegated to Appendixes. Appendix B collects a number of relations useful in constructing analytic $1/m_Q$ expansion in the 't Hooft model and summing the exclusive widths. In particular, we give simple expressions for the leading terms in the transition amplitudes in Appendix B 2, perform the differential fixed-q^2 semileptonic decay width summation up to $1/m_Q^2$ corrections in Appendix B 3, prove the OPE prescription for the domain of large q^2 and demonstrate the proper functional form of the transition amplitudes in Appendix B 4. The expression for the IW functions in terms of the 't Hooft eigenfunctions is quoted in Appendix C. Appendix D reports a direct covariant computation of the perturbative radiative corrections performed while working on Ref. [11]; it shows that the result coincides with what is obtained by summing exclusive decay channels.

II. THE 't HOOFT MODEL AND HEAVY QUARK DECAYS

The 't Hooft model, the $1+1$ QCD with $N_c \to \infty$ has been described in many papers [18–21]. The first dedicated studies of heavy quarks in the 't Hooft model date back to the early 1990s [22,23]. Recent paper [11] specifically addressed heavy quark decays and the OPE in this model. Here we only recapitulate some basic features.

The Lagrangian has the form

$$\mathcal{L}_{1+1} = -\frac{1}{4g_s^2}G_{\mu\nu}^a G_{\mu\nu}^a + \sum \bar{\psi}_i(i\slashed{D} - m_i)\psi_i,$$

$$iD_\mu = i\partial_\mu + A_\mu^a T^a. \quad (5)$$

The coupling g_s has dimension of mass. With the above normalization of the gauge field, A_μ still has dimension of mass as in $D=4$. The fermion fields $\psi(x)$, however, carry a dimension of $m^{1/2}$.

The OPE analysis is carried out universally for arbitrary number of colors, and so far N_c is kept finite. Anticipating the large N_c limit for the final analysis, we define a parameter β

$$\beta^2 = \frac{g_s^2}{2\pi}\left(N_c - \frac{1}{N_c}\right), \quad (6)$$

that remains finite at $N_c \to \infty$. It plays the role of the nonperturbative scale Λ_{QCD}.

Following the actual standard model, we choose the weak decay interaction of the current-current form. Since in $D=2$ the axial current is related to the vector one, $J_\mu^A = \epsilon_{\mu\nu}J_\nu^V$, we simply consider the $V\times V$ interaction

$$\mathcal{L}_{\text{weak}} = -\frac{G}{\sqrt{2}}(\bar{q}\gamma_\mu Q)(\bar{\psi}_a\gamma^\mu\psi_b) + \text{H.c.}, \quad (7)$$

where the dimensionless G is an analogue of the Fermi constant. For semileptonic decays $\psi_{a,b}$ are colorless (leptonic) fields. In what follows our main interest lies in nonleptonic decays with $\psi_{a,b}$ being the quark fields. To make the notations more transparent, we adhere to the cases of interest in actual QCD and denote the ψ fields as u and d quarks, while Q will be a synonym of the b quark, and q called c quark (whether we chose $m_q \gg \Lambda_{\text{QCD}}$ or consider $m_q \lesssim \Lambda_{\text{QCD}}$). The spectator quark q_{sp} can be either u or d (for studying WA or PI), or different in flavor from both.

To address inclusive widths of a heavy flavor hadron H_Q one considers the forward transition amplitude appearing in the second order in the decay interaction [9]:

$$\Gamma_{H_Q} = 2\,\text{Im}\int d^D x\,\frac{1}{2M_{H_Q}}\langle H_Q|iT\{\mathcal{L}_{\text{weak}}(x)\mathcal{L}_{\text{weak}}^\dagger(0)\}|H_Q\rangle. \quad (8)$$

In the limit $N_c \to \infty$, with H_Q being the mesonic $(Q\bar{q}_{\text{sp}})$ states, factorization of the amplitudes holds, which takes the following form for the transition operator:

$$\int d^D x\langle H_Q|iT\{\mathcal{L}_{\text{weak}}(x)\mathcal{L}_{\text{weak}}^\dagger(0)\}|H_Q\rangle$$

$$= \frac{G^2}{2}\int d^D x\,T^{\mu\nu}(x)\Pi_{\mu\nu}(x), \quad (9)$$

where we have introduced the "semileptonic" $T_{\mu\nu}$ and "hadronic" $\Pi_{\mu\nu}$ tensors:

$$\Pi_{\mu\nu}(x) = \langle 0|iT\{\bar{d}(x)\gamma_\mu u(x)\bar{u}(0)\gamma_\nu d(0)\}|0\rangle, \quad (10)$$

$$T^{\mu\nu}(x) = \langle H_Q|iT\{\bar{q}(x)\gamma^\mu Q(x)\bar{Q}(0)\gamma^\nu q(0)\}|H_Q\rangle. \quad (11)$$

The Cutkosky rules then yield

$$\Gamma_{H_Q} = G^2\frac{1}{M_{H_Q}}\int d^D x\,\text{Im}\,T^{\mu\nu}(x)\text{Im}\,\Pi_{\mu\nu}(x). \quad (12)$$

The factorized representation of the decay width holds only at $N_c \to \infty$ where the momenta of the $\psi_a\bar{\psi}_b$-pair and $(q\bar{q}_{\text{sp}})$ become observables separately. In other words, in this limit there is a rigid quark allocation over the particular hadronic final state and factorization of the corresponding amplitudes, and there is no "cross talk" between them. Yet, Eq. (12) represents a certain observable at arbitrary N_c and, as such, enjoys the full rights of being studied regardless of the details of the model. In particular, at large energy release it is a short-distance observable and can be subjected to an OPE anatomy. In what follows we will discuss this quantity and refer to it as the inclusive decay width as motivated by the large-N_c limit.

It is worth noting at this point that the qualitative difference between nonleptonic and semileptonic inclusive widths disappears for $N_c \to \infty$. The nonleptonic width is given directly in terms of the differential semileptonic distributions $d\Gamma^{\text{sl}}/dq^2$ (though, in $D=2$ one may have to consider the decays with massive leptons as well). Indeed, with $m_u = m_d$ as an example, one has (in the momentum representation)

$$\Pi_{\mu\nu}(q^2) = \frac{1}{\pi}\Pi(q^2)(q^2\delta_{\mu\nu} - q_\mu q_\nu),$$

$$\rho(q^2) \equiv -\frac{1}{\pi}\text{Im}\,\Pi(q^2) \quad (13)$$

and

$$\Gamma_{H_Q}^{\text{nl}} = \int dq^2\rho(q^2)\Gamma_{\text{sl}}(q^2) \quad \text{with}$$

$$\Gamma_{\text{sl}}(q^2) = \frac{1}{\rho_{\text{lept}}(q^2)}\frac{d\Gamma_{H_Q}^{\text{sl}}}{dq^2}. \quad (14)$$

In $D=2$ the correlator of vector currents for *massless* quarks is known exactly and is very simple:

$$\Pi(q^2) = \frac{N_c}{q^2}, \quad \rho(q^2) = N_c\delta(q^2). \quad (15)$$

With nonzero quark masses the spectral density shifts upward, to the mass scale $\sim \beta m$ or m^2. A high-energy tail in ρ also appears $\sim N_c(m_u^2 + m_d^2)/q^4$. This will be quantified in Sec. III.

More specific for heavy quark decays is the "semileptonic" part $T_{\mu\nu}(x)$, Eq. (11). The general color counting rules determine its N_c behavior:

$$T_{\mu\nu}(x) \sim N_c^1. \quad (16)$$

Such a leading contribution, however, can arise only with the vacuum as intermediate state; all other contributions scale as N_c^0, or even are further suppressed. The vacuum intermedi-

ate state is possible only when the decay quark q has the same flavor as q_{sp}. This is the effect belonging to WA. Therefore, one has

$$\Gamma_{nl}^{WA} \sim N_c^2, \quad \Gamma_{sl}^{WA} \sim N_c^1, \tag{17}$$

$$\Gamma_{nl} \sim N_c^1, \quad \Gamma_{sl} \sim N_c^0, \quad \text{at} \quad q \neq q_{sp}. \tag{18}$$

Since WA is a leading-N_c effect, vacuum factorization saturates $T_{\mu\nu}^{WA}$ at $N_c \to \infty$, and the effect takes the simplest form. This is the subject of the next section.

On the other hand, the ''usual'' nonspectator widths are formally subleading in N_c (even though they may yield the dominant contribution to the decay width for a particular type of the heavy meson). For such amplitudes the naive factorization does not hold, and the explicit expressions take a far less trivial form. In the context of the OPE, this emerges as ''color-disfavored'' structure of the resulting local operators, so that *a priori* the factorization cannot be applied to evaluate their expectation values [11].

In the limit $N_c \to \infty$ the spectrum of $1+1$ QCD consists of mesonic quark-antiquark bound states which are stable under strong interactions. The meson masses are given by eigenvalues of the 't Hooft equation

$$M_n^2 \varphi_n(x) = \left[\frac{m_1^2 - \beta^2}{x} + \frac{m_2^2 - \beta^2}{1-x} \right] \varphi_n(x) - \beta^2 \int_0^1 dy \frac{\varphi_n(y)}{(y-x)^2}, \tag{19}$$

where $m_{1,2}$ are the bare quark masses of the constituents, and the integral is understood in the principal value prescription. The solutions to the equation are the light-cone wave functions $\varphi(x)$, with $x \in [0,1]$ having the meaning of the portion of momentum carried by the (first) quark. They are singular at $x=0$ and $x=1$ where their behavior is given by x^{γ_0} and $(1-x)^{\gamma_1}$, respectively, with $\gamma_{0,1}$ defined by the following conditions:

$$\frac{\pi \gamma_0}{\tan \pi \gamma_0} = -\frac{m_1^2 - \beta^2}{\beta^2}, \quad \frac{\pi \gamma_1}{\tan \pi \gamma_1} = -\frac{m_2^2 - \beta^2}{\beta^2}. \tag{20}$$

In full analogy with nonrelativistic quantum mechanics, the eigenfunctions φ_n form a basis (complete in the physical space):

$$\int_0^1 dx \, \varphi_n(x) \varphi_k(x) = \delta_{nk}, \quad \sum_n \varphi_n(x) \varphi_n(y) = \delta(x-y). \tag{21}$$

The weak decay constant of a particular meson is given by

$$f_n = \sqrt{\frac{N_c}{\pi}} \int_0^1 dx \, \varphi_n(x) \tag{22}$$

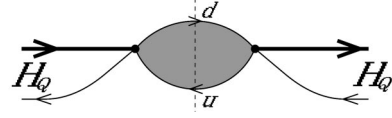and the polarization tensor of vector currents (at $m_u = m_d$) takes the form

FIG. 3. WA correction to the inclusive decay width in the large-N_c limit. The shaded loop depicts the exact polarization operator.

$$\Pi(q^2) = \pi \sum_n \frac{f_n^2}{q^2 - M_n^2}, \quad \rho(q^2) = \pi \sum_n f_n^2 \delta(q^2 - M_n^2). \tag{23}$$

As mentioned above, at $m_u = m_d = 0$ one has $M_0 = 0$ and $f_0 = \sqrt{N_c/\pi}$, but for all excitations $f_n = 0$.

The transition form factors between two mesonic states that define the nonannihilation widths for large N_c, are of order N_c^0. Since the weak quark currents $\bar{Q}q$ are formally of order N_c^1, these form factors are ''subleading'' in the same sense as was discussed previously and, in general have a more complicated form corresponding to the first order correction in the $1/N_c$ expansion [20,21].

III. WEAK ANNIHILATION AT $N_c \to \infty$

WA in the decays of heavy mesons becomes possible when one of the quarks produced in the weak vertex has the same flavor as the spectator antiquark. We assume $q = q_{sp}$, in our notations. As detailed in the preceding section, in this case there is a single contribution to the transition tensor $T_{\mu\nu}$ proportional to N_c and leading to $\Gamma_{H_Q} \sim N_c^2$. This is associated with the vacuum intermediate state, and is given by[3]

$$T_{\mu\nu}(x) = i e^{-iPx} \vartheta(x_0) \langle H_Q | \bar{Q} \gamma_\nu q | 0 \rangle \langle 0 | \bar{q} \gamma_\mu Q | H_Q \rangle \tag{24}$$

(with P_α denoting the momentum of the decaying heavy flavor hadron H_Q), so that, in the momentum representation

$$\text{Im} \, T_{\mu\nu}(q) = \frac{1}{2} (2\pi)^D \delta^D(P - q) \langle H_Q | \bar{Q} \gamma_\nu q | 0 \rangle$$
$$\times \langle 0 | \bar{q} \gamma_\mu Q | H_Q \rangle. \tag{25}$$

This expression is valid in arbitrary dimension for any choice of the weak current—in general, one only must replace $\bar{Q} \gamma_\mu q$ by the appropriate quark bilinear. Therefore, at $N_c \to \infty$ one has

$$\Gamma_{H_Q}^{WA} = G^2 \frac{1}{2M_{H_Q}} \langle H_Q | \bar{Q} \gamma_\mu q | 0 \rangle$$
$$\times \langle 0 | \bar{q} \gamma_\nu Q | H_Q \rangle \text{Im} \, \Pi_{\mu\nu}(M_{H_Q}^2), \tag{26}$$

which is illustrated in Fig. 3. It can be traced that the OPE

----

[3]We neglect the contribution of another, two-particle state $|H_Q(P)H_Q(P)\rangle$, also corresponding to vacuum factorization, but yielding the u-channel discontinuity.

corresponds to the same expression if the expectation values of all the higher-dimension four-quark operators reduce to their vacuum factorized values (for earlier discussion of WA in a similar context, see Ref. [24]). The latter formally holds, in turn, at $N_c \to \infty$.

In $D=2$ for pseudoscalar H_Q one has $\langle 0 | \bar{q} \gamma_\mu Q | H_Q \rangle = i f_{H_Q} \epsilon_{\mu\nu} P_\nu$. For simplicity, we will further limit ourselves by the case $m_u = m_d$. Then

$$\Gamma_{H_Q}^{\text{WA}} = \frac{G^2}{2} f_{H_Q}^2 M_{H_Q}^3 \rho(M_{H_Q}^2). \tag{27}$$

Strictly speaking, in practical applications of the OPE, $\Pi_{\mu\nu}$ itself is usually likewise expanded in $1/m_Q^2$. Also, the deviation of $M_{H_Q}^2$ from m_Q^2 and/or the values of f_{H_Q} are expanded around their asymptotic values at $m_Q \to \infty$. Therefore, the sensible check of duality for practical OPE in WA in the framework of the large-N_c approximation is only comparison of the actual behavior of Im $\Pi(q^2)$ at large q^2 with its OPE expansion obtained from the deep Euclidean domain.

For massless u and d quarks, the exact polarization operator of the vector currents is given by Eq. (15); the WA width, therefore, vanishes. A nonzero result is obtained if one considers a scalar (pseudoscalar) polarization operator, or if m_u or m_d do not vanish. The absorptive part $\rho(q^2)$ is saturated by the comb of narrow resonances with heights $\sim N_c$ and widths $\sim 1/N_c$. Therefore, the formal limit $N_c \to \infty$ requires an alternative to point-to-point comparison of the actual hadronic probabilities with the parton-calculated, or OPE-improved short-distance expansion, even at arbitrary large energies. This implies a certain smearing procedure for the actual hadronic probabilities.

Note that, according to Eq. (26) the width—however singular it is—always remains integrable around the resonances [see also the discussion below, Eqs. (28), (29)–(33)]. By virtue of the dispersion relations the integral of the decay width is expressed via the transition amplitude in the complex plane. This amplitude is regular even in the formal limit $N_c \to \infty$ when the resonances become infinitely narrow.

Smearing enters naturally when one considers the ''imaginary'' part $1/2i[\Pi(s)-\Pi(s^*)]$ at complex s, somewhat away from the physical cut at $s>0$. According to a dispersion relation it amounts to averaging the physical cross section $R(s)$ with a specific weight,

$$\frac{1}{2i}[\Pi(s)-\Pi(s^*)] = \frac{1}{\pi} \int d\tilde{s} \frac{\Delta}{(\tilde{s}-s_0)^2 + \Delta^2} R(\tilde{s}),$$

$$s = s_0 + i\Delta. \tag{28}$$

One can also use different choices of the smearing function having singularities away from the physical cut.

A similar procedure, in principle, is required for the inclusive decays of heavy flavors. Strictly speaking, one must introduce the complex variable ω to study the analytic properties of the transition amplitude in question [24,12,11]:

$$\mathcal{A}(\omega) = \int d^D x \, e^{-i\omega(vx)} \langle H_Q | iT\{\mathcal{L}_{\text{weak}}(x)\mathcal{L}_{\text{weak}}^\dagger(0)\} | H_Q \rangle. \tag{29}$$

It can be visualized as the transition amplitude governing the total (weak) cross section of the scattering of a fictitious spurion particle S on the heavy quark

$$S(q) + H_Q(p) \to \text{light hadrons}, \tag{30}$$

or the weak decay width in the process

$$Q \to \text{quarks (leptons)} + S. \tag{31}$$

Such processes would appear if the weak decay Lagrangian is modified from, say the conventional four-fermion form to the ''four-fermion+spurion'' interaction

$$\mathcal{L}_{\text{weak}}(x) \to S(x)\mathcal{L}_{\text{weak}}(x). \tag{32}$$

For simplicity, it is convenient to assume, as in Eq. (29) that the spurion field does not carry spacelike momentum.

The amplitude $\mathcal{A}(\omega)$ has the usual analytic properties, and the discontinuity across the physical cut at which the point $\omega=0$ is located, describes the total decay width we are interested in. The OPE for the inclusive widths relies on the fact that the short-distance expansion of $\mathcal{A}(\omega)$ runs in $1/(\omega - E_r)$ and can be applied near the physical point $\omega=0$ exactly as in e^+e^- annihilation near a positive value of $s \gg \Lambda_{\text{QCD}}^2$. ($E_r$ denotes energy release.) To the same extent, in principle, a certain smearing can be required if the hadronic probabilities still exhibit the resonance structure.

Thus, there is no theoretical peculiarity in the asymptotic applications of the OPE for nonleptonic widths. It does not create a conceptual difference to perform a short-distance expansion of a single quark Green function (semileptonic widths or deep inelastic scattering), the product of two Green functions (e^+e^- annihilation) or the product of three quark Green functions (the nonleptonic widths).

Alternatively, smearing in ω can be phrased as smearing over the interval of m_Q. Indeed, in the heavy quark limit the amplitudes depend on just the combination $m_b - \omega$,

$$\mathcal{A}(\omega, m_Q) \simeq \mathcal{A}(0, m_Q - \omega) \tag{33}$$

(there are power corrections to this relation associated with explicit mass effects in the initial state). Therefore, in practical terms one can phrase the smearing as an averaging over the interval of the heavy quark mass, which may look more transparent.

After this general digression, we now return to specifically WA in two-dimensional QCD. It is commonly accepted that, for the two-point current correlators, both at $N_c \to \infty$ or finite N_c, the properly averaged absorptive hadronic parts asymptotically coincide with the leading OPE expression given by the free quark diagram. As was mentioned above, for massless quarks this property holds *identically* for vector and axial currents. For the scalar current the asymptotic correspondence in the 't Hooft model has been illustrated already in Ref. [19] (for a recent discussion and earlier references, see Ref. [25]). For the WA width, however, we need

the $m_{u,d}$-suppressed effects. The OPE in $D=2$ yields, at $m_u=m_d=m$ (for arbitrary N_c),

$$\Pi(q^2)=N_c\left[\frac{1}{q^2}+\frac{2m^2}{q^4\sqrt{1-4m^2/q^2}}\ln\frac{\sqrt{1-4m^2/q^2}+1}{\sqrt{1-4m^2/q^2}-1}\right]$$
$$-\frac{2\pi\langle 0|m_u\bar{u}u+m_d\bar{d}d|0\rangle}{q^4}+\mathcal{O}\left(\frac{m^2\ln q^2}{q^6}\right),$$

$$\rho(q^2)=N_c\frac{2m^2}{q^4}\frac{1}{\sqrt{1-4m^2/q^2}}+2\pi\langle 0|m_u\bar{u}u+m_d\bar{d}d|0\rangle$$
$$\times\delta'(q^2)+\mathcal{O}\left(\frac{m^2}{q^6}\right), \qquad (34)$$

where the first term in both equations is just the free quark loop. There is little reason to doubt the OPE for the subleading terms either. Nevertheless, it is instructive to give here the direct derivation of the next-to-leading term $\sim(m_u^2+m_d^2)/q^4$ in $\rho(q^2)$ directly from the 't Hooft equation.

We follow here the approach of Ref. [11] based on sum rules. In the context of the Euclidean polarization operator similar considerations ascend to the earliest papers on the model, Refs. [19,20]. To simplify the expressions, we will suppress the explicit powers of N_c which enter in a trivial way, and usually will also omit the mass scale factor β, assuming that all energies are measured in units of β. Then Eqs. (22), (23) take the form

$$\Pi(q^2)=\sum_n\frac{c_n^2}{q^2-M_n^2},\quad \rho(q^2)=\sum_n c_n^2\delta(q^2-M_n^2) \qquad (35)$$

with

$$c_n=\int_0^1 dx\,\varphi_n(x). \qquad (36)$$

The completeness of eigenstates yields

$$\sum_n c_n^2=\sum_n\int_0^1 dx\,dy\,\varphi_n(x)\varphi_n(y)=1. \qquad (37)$$

On the other hand, integrating the 't Hooft equation from 0 to 1 we get

$$c_n=\int_0^1 dx\,\varphi_n(x)=\frac{1}{M_n^2}\int_0^1 dx\left(\frac{m_d^2}{x}+\frac{m_u^2}{1-x}\right)\varphi_n(x). \qquad (38)$$

Therefore, we get the second sum rule

$$\sum_n M_n^2 c_n^2=\sum_n\int_0^1 dx\,\varphi_n(x)\int_0^1 dy\left(\frac{m_d^2}{y}+\frac{m_u^2}{1-y}\right)\varphi_n(y)$$
$$=\int_0^1 dx\left(\frac{m_d^2}{x}+\frac{m_u^2}{1-x}\right). \qquad (39)$$

The integral logarithmically diverges at $x\to 0$ and $x\to 1$, which corresponds to the behavior

$$\Pi(q^2)\overset{q^2\to\infty}{\simeq}\frac{1}{q^2}-\frac{m_u^2+m_d^2}{q^4}\ln\frac{q^2}{m^2}+\mathcal{O}\left(\frac{1}{q^4}\right) \qquad (40)$$

given by the free quark diagram, Eq. (34). The divergence of the sum in Eq. (39) is associated with the high excitations n. Therefore, quantifying the divergence allows one to determine the asymptotic behavior of c_n^2.

To render the sum in Eq. (39) finite we must introduce an ultraviolet regularization. For the logarithmic divergence the exact way is not essential—one is to add a hard cutoff factor $\vartheta(\Lambda^2-M_n^2)$. For analytic computations the Borel-type regularization by the factor $e^{-M_n^2/\Lambda^2}$ is usually convenient.

For the regularized sums (we mark them with the superscript Λ) the completeness condition is modified,

$$\sum_n^\Lambda\varphi_n(x)\varphi_n(y)=G(x,y;\Lambda), \qquad (41)$$

and the Green function G becomes a "finite-width" δ-like distribution with the width

$$\Delta\sim\frac{1}{\Lambda^2}. \qquad (42)$$

This regularizes the sum in Eq. (39):

$$\sum_{M_n^2<\Lambda^2}M_n^2 c_n^2=\int_0^{\Lambda^2}q^2 dq^2\,\rho(q^2)$$
$$=\int_0^1 dx\,dy\left(\frac{m_d^2}{x}+\frac{m_u^2}{1-x}\right)G(x,y;\Lambda)$$
$$=(m_u^2+m_d^2)(\ln\Lambda^2+\text{const}). \qquad (43)$$

One has, for instance, for the sum over an interval of highly excited states

$$\int_{\Lambda_1^2}^{\Lambda_2^2}q^2 dq^2\,\rho(q^2)=\sum_{\Lambda_1^2<M_n^2<\Lambda_2^2}M_n^2 c_n^2$$
$$=(m_u^2+m_d^2)\ln\frac{\Lambda_2^2}{\Lambda_1^2}+\mathcal{O}\left(\frac{1}{\Lambda^2}\right). \qquad (44)$$

The sum rule (43) proves that the asymptotics of the smeared $\text{Im}\,\Pi(q^2)$ coincides with the free quark loop result through terms m^2/q^4. It is easy to see that the nontrivial corrections in the OPE also emerge only with higher-order terms in $1/q^2$. Note that Eqs. (43), (44) hold both for light ($m\ll\beta$) and heavy ($m\gg\beta$) quarks. However, for the asymptotics to start, the condition $\Lambda\gg m_{u,d}$ must be observed.

Since $J_\mu^5=\epsilon_{\mu\nu}J_\nu$ and using equation of motion $\partial_\mu J_\mu^5=(m_u+m_d)\bar{u}i\gamma_5 d$, by the same token we showed the leading-order duality between the hadronic saturation and the

partonic expression for the absorptive part of the pseudo-scalar current. A direct derivation in the same approach is described in Appendix A.

As expected, for large M_n one finds the residues $c_n \sim m_{u,d}$. Let us note that for light quarks c_n are only linear in m_q: since for light quarks $\varphi_n(x) \sim x^{m_u}$ at $x \to 0$ (and likewise at $x \to 1$), the end points of integration in Eq. (38) bring in the $1/m_q$ enhancement. Combining the sum rule Eq. (43) with the asymptotics of the 't Hooft eigenvalues

$$M_n^2 \simeq \beta^2 \pi^2 n,$$

we obtain

$$c_n^2 \simeq \pi^2 \beta^2 \frac{m_u^2 + m_d^2}{M_n^4} \simeq \frac{m_u^2 + m_d^2}{\pi^2 \beta^2 n^2}. \tag{45}$$

Again, these asymptotics are valid if ''averaged'' over an interval of n.

It must be noted that the explicit constant in Eq. (42) is not important. A more detailed derivation of the large-Λ asymptotics uses the semiclassical expansion of the 't Hooft wave functions. We show in Appendix A that the domain of integration where $x < 1/\Lambda^2$ or $y < 1/\Lambda^2$ yields only a finite contribution to the integral in Eq. (43) (and likewise in the vicinity of $x = 1$ or $y = 1$). At the same time, in the domain $x, y \gg 1/\Lambda^2$ the approximation $G(x, y; \Lambda) = \delta(x - y)$ is applicable.

With the relation for the WA width Eq. (26), the comparison of Eq. (44) and the OPE asymptotics Eq. (34) demonstrates that the smeared width in the 't Hooft model coincides with the OPE width at least through terms $m_{u,d}^2/m_Q^2$.

IV. PAULI INTERFERENCE

In this section we address the effect of interference in the weak decay width of the heavy mesons. As explained in the Introduction, it has an independent interest. Similar to the partonic free-quark decay width, PI is a ''subleading'' $1/N_c$ effect, with $\Gamma^{PI} \sim N_c$ rather than N_c^2. Therefore, the expressions for the amplitudes are not as trivial as for WA. Nevertheless, it is not difficult to demonstrate that, again, the quark-based OPE predictions coincide with the actual hadronic widths.

To incorporate PI we must have the flavor of the anti-quark produced in the decay of virtual W coinciding with the flavor of the spectator; we shall call it u. Moreover, the weak decay Lagrangian must contain two different color structures to have PI at the same order in N_c as the free partonic width. So, we adopt, for simplicity,

$$\mathcal{L}_{\text{weak}} = -\frac{G}{\sqrt{2}} [a_1 (\bar{c} \gamma_\mu b)(\bar{d} \gamma^\mu u) + a_2 (\bar{d} \gamma_\mu b)(\bar{c} \gamma^\mu u)] + \text{H.c.}, \tag{46}$$

where, again for notational transparency, we identified Q with b and called q by c.

In this case the decay width has three terms

$$\Gamma_{H_Q} = \frac{G^2}{2} N_c (a_1^2 \Gamma_1 + a_2^2 \Gamma_2 + 2 a_1 a_2 \Gamma_{12}), \tag{47}$$

where Γ_1 and Γ_2 are $\mathcal{O}(N_c^0)$. Clearly, $\Gamma_1(m_b, m_c, m_u, m_d) = \Gamma_2(m_b, m_d, m_u, m_c)$ holds.

The asymptotics of the noninterference width Γ_1 (Γ_2) for the 't Hooft model was calculated in Ref. [11] and shown to be given by the OPE one. Now we address the analogous question for Γ_{12}.

The leading (in m_Q) contribution to the decay width described by the free parton diagram in Fig. 1(a) suggests that $\Gamma_{12} \sim 1/N_c$. For example, for usual $V - A$ interaction in $D = 4$ one would have

$$\Gamma_{12}^{\text{parton}} = \frac{1}{N_c} \Gamma_1^{\text{parton}} = \frac{1}{N_c} \Gamma_2^{\text{parton}} \tag{48}$$

(the explicit factor depends on the Lorentz structure of $\mathcal{L}_{\text{weak}}$). Such N_c-subleading effects are rather complicated. This suppression, however, is not always present [26]. As discussed earlier, invoking the spectator quark through the spectator-dependent effects such as WA or PI can bring in an N_c enhancement by effectively eliminating the generic $1/N_c$ suppression of the free quark width. As a result, at the price of a power suppression in m_Q one can have the N_c-unsuppressed manifestation of the interference of the two color amplitudes in $\mathcal{L}_{\text{weak}}$,

$$\Gamma_{12}^{(\text{PI})} \sim \mathcal{O}(N_c^0). \tag{49}$$

Thus, on the one hand, studying PI allows one to address the interference of the color amplitudes in a straightforward way relying on the $1/N_c$ expansion. On the other hand, considering the term $\sim a_1 a_2$ in the decay width in the limit $N_c \to \infty$ automatically singles out the power-suppressed effect of PI. This goes in contrast with the usual situation where isolating PI formally requires subtracting the decay width of the similar heavy flavor hadron with the spectator(s) having the same mass but with the flavor which is sterile in weak interactions.

The simple quark diagram describing PI is shown in Fig. 2. To leading order it generates the operator

$$\hat{\Gamma}^{\text{PI}} = -2 a_1 a_2 \frac{G^2}{2K} \left\{ \left(1 - \frac{m_c^2 + m_d^2}{m_Q^2} \right) (\bar{b} \gamma_\mu \gamma_5 u)(\bar{u} \gamma_\mu \gamma_5 b) \right.$$
$$\left. - \frac{2 m_c m_d}{m_Q^2} [(\bar{b} u)(\bar{u} b) + (\bar{b} i \gamma_5 u)(\bar{u} i \gamma_5 b)] \right\}, \tag{50}$$

with $2 m_Q K$ having the meaning of the quark spacelike momentum in the final state:

$$K = \left[\left(1 - \frac{(m_c + m_d)^2}{m_Q^2} \right) \left(1 - \frac{(m_c - m_d)^2}{m_Q^2} \right) \right]^{1/2}.$$

It is worth noting that this contribution is not chirally suppressed. Therefore, it is meaningful and convenient to consider it in the limit $m_c = m_d = 0$.

For B^- mesons having $\bar{u}$ spectator, the operators in Eq. (50) have the N_c-favorable color structure and, therefore, their expectation values are given by vacuum factorization:

$$\frac{1}{2M_B}\langle B^-|(\bar{b}\gamma_\mu\gamma_5 u)(\bar{u}\gamma_\mu\gamma_5 b)|B^-\rangle = \frac{1}{2}f_B^2 M_B,$$

$$\frac{1}{2M_B}\langle B^-|(\bar{b}u)(\bar{u}b)+(\bar{b}i\gamma_5 u)(\bar{u}i\gamma_5 b)|B^-\rangle$$

$$= \frac{f_B^2 M_B^3}{2(m_b+m_u)^2}. \qquad (51)$$

In particular, at $m_c=m_d=0$ one gets

$$\Gamma^{\text{PI}}=-2a_1 a_2 \frac{G^2}{4}f_B^2 M_B. \qquad (52)$$

We note that Γ^{PI} asymptotically approaches a constant when $m_Q\to\infty$.

It is interesting that there are no $1/m_Q$ corrections (at small $m_{c,d}$) to the above result. This is a peculiarity of two dimensions where the absorptive part of the (di)quark loop in Fig. 2 scales as the momentum to the zeroth power and, thus, does not depend on whether one uses p_b or P_B as the momentum flowing into it. The corrections to the Wilson coefficient as well as other higher-order operators can induce only terms suppressed by at least two powers of inverse mass.

Let us now consider the decays in terms of hadrons. In the absence of WA, the leading-N_c final states are pairs of mesons. The partial decay width $B\to D_k^0\pi_n^-$ takes the general form

$$\Gamma_{kn}=\frac{G^2}{8M_B^2|\vec{p}|}[a_1^2|\mathcal{A}_k\mathcal{B}_n|^2+a_2^2|\mathcal{A}_n\mathcal{B}_k|^2$$
$$+2a_1 a_2\,\text{Re}\,\mathcal{A}_k\mathcal{A}_n^*\mathcal{B}_k^*\mathcal{B}_n], \qquad (53)$$

where $\mathcal{A}$ and $\mathcal{B}$ schematically denote the "multiperipheral" $B\to k$ transition amplitudes and the "pointlike" meson creation amplitudes, respectively:

$$\mathcal{A}_k\sim\langle k|J_\mu|B\rangle, \quad \mathcal{B}_n\sim\langle n|J_\mu|0\rangle. \qquad (54)$$

We denote by $\vec{p}$ the rest-frame momentum of the final state mesons. The PI term is then given by the sum

$$\Gamma^{\text{PI}}=2a_1 a_2\frac{G^2}{8M_B^2}\sum_{k,n}\frac{1}{|\vec{p}|}\mathcal{A}_k\mathcal{A}_n^*\mathcal{B}_k^*\mathcal{B}_n. \qquad (55)$$

Both Γ_1 and Γ_2 are saturated by the final states of the type $D_k^0\pi_n^-$ with various excitation indices k and n. However, the production mechanism differs: while the "charge-current" interaction $\sim a_1$ produces π_n^- by the weak current "pointlike" and D_k^0 in a "multiperipheral" way [see Fig. 4(a)], the situation reverses for the "neutral-current" amplitudes proportional to a_2, Fig. 4(b). These two sources of the final state mesons have distinct features for heavy enough Q:
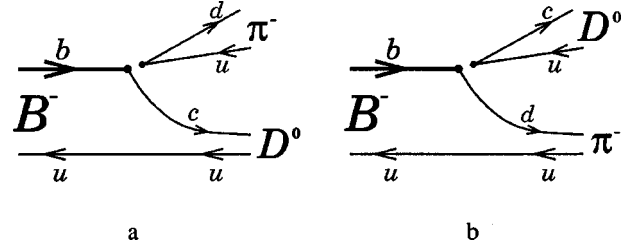

FIG. 4. Large-N_c decay amplitudes induced by charge-current (a) and neutral-current (b) terms in the weak decay Lagrangian.

the multiperipherally produced mesons have the mass squared distributed in the interval from 0 to $\sim\Lambda_{\text{QCD}}m_Q$. The bulk of the pointlike produced mesons have the mass squared[4] $\lesssim m_q\Lambda_{\text{QCD}}$ or m_q^2. Reference [11] demonstrated these OPE-suggested facts explicitly in the 't Hooft model.

As a result, interference becomes possible only at a small, $\sim 1/m_Q$ slice of the principal decay channels. This qualitatively explains the $1/m_Q$ power suppression of PI which is automatic in the OPE.

We will now demonstrate the quantitative matching between the OPE-based calculation and the hadronic saturation of the interference width. To make the proof most transparent, we start with the simplest possible case when all final state quarks u,d,c are massless. While not affecting the OPE analysis, this limit significantly simplifies the expressions for the individual hadronic amplitudes, as explained in Ref. [11]. In the case at hand, for example, only $n=0$ survives for the decay amplitude $\sim a_1$ [Fig. 4(a)] and $k=0$ for the amplitude $\sim a_2$ [Fig. 4(b)]. The interference then resides in the single final state containing the lowest lying massless D^0 and π^-. Moreover, the corresponding transition amplitudes $\mathcal{A}$ between two mesons take particularly simple form at $q^2=0$ in terms of their 't Hooft wave functions [11]:

$$q_\mu\frac{1}{2M_B}\langle k|\epsilon_{\mu\nu}J_\nu|B\rangle=-q_z\int_0^1 dx\,\varphi_k(x)\varphi_B(x), \qquad (56)$$

where we have recalled that $\mathcal{B}_0\sim i\epsilon_{\mu\nu}f_0 P_\nu^{(0)}$, and therefore considered only the relevant light-cone component of the amplitude. Then we have

$$\Gamma^{\text{PI}}=-2a_1 a_2\frac{G^2 M_B N_c}{4\pi}\left|\int_0^1 dx\,\varphi_B(x)\right|^2=-2a_1 a_2\frac{G^2}{4}f_B^2 M_B \qquad (57)$$

[we have used the fact that $\varphi_0(x)=1$, $f_0=\sqrt{N_c/\pi}$ for massless quarks]. The minus sign emerges since the direction of the vector playing the role of $\vec{q}$ is opposite for the two interfering amplitudes.

Thus, the OPE asymptotics Eq. (52) is exactly reproduced. Apparently, there is no violation of local duality *at all* for PI in the case $m_u=m_d=m_c=0$. This is not surprising—in this limit the only threshold in Γ^{PI} occurs at zero mass, and

_____

[4]For the vectorlike current; it would be evenly spread from about m^2 to m_Q^2 if the weak vertex were scalar.

the OPE series can have the same convergent properties in Minkowskian as in Euclidean space.

With $m_{u,d,c} \neq 0$ the interference effects are saturated by several final state pairs of mesons, even if the masses are small compared to m_Q. It is still not difficult, though, to check that the leading OPE term Eq. (52) is reproduced. We keep in mind that at nonzero masses the width exhibits the threshold singularities due to the singular two-body phase space $1/|\vec{p}|$ in $D=2$. Since it is integrable, the threshold spikes do not affect the width smeared over the interval of mass $\Delta m_Q \sim 1/m_Q$.

The idea of the proof is suggested by the detailed kinematic duality between the partonic and hadronic probabilities. The bulk of pointlike-produced mesons have masses squared M_n^2 not exceeding β^2 or m_q^2, while for multiperipherally created mesons k this scale is $\sim \beta m_Q$ or $m_{sp} m_Q$. More precisely [11], for the decay rates $\sim a_1^2$ [Fig. 4(a)],

$$\frac{1}{\Gamma_{tot}} \sum_k \sum_{M_n > \text{const } m_Q} \Gamma_{kn} \propto \frac{m_{u,d}^2}{m_Q^2}, \tag{58}$$

$$\frac{1}{\Gamma_{tot}} \sum_{M_n \ll m_Q} \sum_{M_k > \text{const } m_Q} \Gamma_{kn} \propto \frac{1}{m_Q^5}. \tag{59}$$

Then, calculating the width we can expand around the free quark kinematics $M_n = M_k = 0$. In particular, we set

$$\frac{1}{|\vec{p}|} = \frac{2}{M_B} \left(1 + \frac{M_k^2 + M_n^2}{M_B^2} + \cdots \right). \tag{60}$$

Additionally, we can expand the transition form factors in amplitudes $\mathcal{A}_k$ in q^2:

$$\mathcal{A}_k(M_n^2) \simeq \mathcal{A}_k(0) + \frac{M_n^2}{m_Q^2} m_Q^2 \frac{d\mathcal{A}_k}{dq^2} \bigg|_{q^2=0}, \tag{61}$$

and likewise for $\mathcal{A}_n(M_k^2)$. In factoring out $1/m_Q^2$ in the slope of the amplitude we accounted for the fact that it scales as $1/m_Q^2$ in this kinematics. Indeed, the t-channel resonances have masses exceeding m_Q, and the kinematics (the fractions of the light-cone momenta entering computation of the transition amplitudes, see the next section and Appendix B) likewise depend on q^2 only as q^2/m_Q^2.

To obtain Γ^{PI} with an accuracy $1/m_Q^2$ of the free quark width, we actually expand the particular two-body decay amplitude only in q^2, that is, do not neglect M_k^2 dependence for the amplitude $\sim a_1$ or M_n^2 dependence for the amplitude which is proportional to a_2. The expressions for the decay amplitudes at $q^2 = 0$ are very simple [11]:

$$\mathcal{M}_{kn} = G \sqrt{\frac{N_c}{2\pi}} (a_1 \mathcal{M}_{kn}^{(1)} + a_2 \mathcal{M}_{kn}^{(2)}), \tag{62}$$

$$\mathcal{M}_{kn}^{(1)} = (M_B^2 - M_k^2) \int_0^1 dx\, \varphi_B(x) \varphi_k(x) \int_0^1 dy\, \varphi_n(y),$$

$$\mathcal{M}_{kn}^{(2)} = -(M_B^2 - M_n^2) \int_0^1 dx\, \varphi_B(x) \varphi_n(x) \int_0^1 dy\, \varphi_k(y),$$

$$2|\vec{p}| M_B \simeq M_B^2 - M_k^2 - M_n^2.$$

Then we have

$$\sum_{k,n} \Gamma_{kn}^{PI} \simeq -2a_1 a_2 \frac{G^2 M_B N_c}{4\pi} \sum_{k,n} \int_0^1 dx\, \varphi_B(x) \varphi_k(x)$$

$$\times \int_0^1 dy\, \varphi_B(y) \varphi_n(y) \int_0^1 dz\, \varphi_n(z) \int_0^1 dt\, \varphi_k(t)$$

$$= -2a_1 a_2 \frac{G^2 M_B N_c}{4\pi M_B^2} \left| \int_0^1 dx\, \varphi_B(x) \right|^2, \tag{63}$$

or

$$\Gamma^{PI} = -2a_1 a_2 \frac{G^2}{4} f_B^2 M_B. \tag{64}$$

We extended summation over k and l in Eq. (63) to include all states, since the contribution of additional, kinematically forbidden meson pairs is suppressed by high powers of $1/m_Q$.

This expression is valid up to the relative $1/m_Q^2$ corrections. Indeed, the leftover effect of the slope of the transition form factor F_{Bk} is quadratic in m_q/m_Q. For example, using representation Eq. (38) we obtain a sum rule which allows one to cast it in the form

$$\sum_n \delta \mathcal{M}_{kn}^{(1)} \mathcal{M}_{kn}^{(2)} = \sum_n \frac{\partial}{\partial q^2} F_{Bk} M_n^2 c_n \int_0^1 dy\, \varphi_B(y) \varphi_n(y) c_k$$

$$= c_k \frac{\partial}{\partial q^2} F_{Bk} \int_0^1 dy\, \varphi_B(y) \left(\frac{m_d^2}{y} + \frac{m_u^2}{1-y} \right) \tag{65}$$

(and likewise for $\delta \mathcal{M}_{kn}^{(2)}$). The convergence of the integral over y shows that this effect is saturated at small n and is of order m_q^2/m_Q^2 or $m_q \beta/m_Q^2$, whichever is larger.

A more accurate consideration reveals that the two amplitudes in Eqs. (62) have the factors $(-1)^n$ and $(-1)^k$, respectively, and their product, additionally, the factor $(-1)^{(P_{H_Q}+n+k)}$. [The latter is related to the opposite direction of $\vec{q}$ in the two amplitudes and is readily understood since this is a parity-conserving decay $H_Q \rightarrow k+n$ with the meson parities P_{H_Q}, $(-1)^n$ and $(-1)^k$.] Therefore, the sign of Γ^{PI} is given by the parity of H_Q, which is manifest for the OPE result in $D=2$, cf. Eq. (51). Thus, we see that Γ^{PI} agrees with the expression given by the free quark loop just to the accuracy suggested by the OPE.

It is not difficult to estimate the effects of violation of local duality in PI related to the thresholds, for small but nonvanishing m_q. Since the two-body phase space is singular, different ways to gauge its strength will yield different power of its asymptotic suppression. Full information is just given by the nature of the threshold singularity, the scaling of the corresponding residues, and the asymptotic distance between the principal thresholds. This would show the con-

tribution to PI of any new decay channel, close to the mass where it opens, where the corresponding width is not literally given by the OPE.

It turns out that the magnitude of local duality violation in PI essentially depends on the relation between the final state masses. The strongest effect comes from the kinematics where one of the mesons belongs to low excitations while another has the large mass close to m_b.

The case when $m_u = m_d = 0$ (but $m_c > 0$) is somewhat special. Here one of the interfering decay amplitudes vanishes at the thresholds, and Γ^{PI} simply experiences a finite jump:

$$|\delta\Gamma^{\mathrm{PI}}_{k0}| \simeq \frac{G^2}{2} 2a_1 a_2 \times \mathrm{const}\, f_B m_c \frac{\beta^{9/2}}{M_{H_Q}^{9/2}} \vartheta(M_{H_Q} - M_k),$$

$$M_{k+1} - M_k \simeq \frac{\pi^2 \beta^2}{2M_{H_Q}}. \tag{66}$$

Here we used the semiclassical calculations of the transitions to highly excites states given in Sec. III, Eq. (45) and in Ref. [11], Eq. (79). The latter estimate for the $H_Q \to D^k$ transition amplitude is valid up to a factor of order one; accepting it at face value would yield $\sqrt{3\pi}$ for the constant in Eq. (66).

The referred asymptotics determined the absolute magnitude of the decay amplitudes relevant for usual decay probabilities, but not their sign which plays a role in interference which can be both constructive and destructive. A more careful analysis suggests that the relative sign of the two amplitudes alternates for successive thresholds. Therefore, at $m_u \simeq m_d \ll \beta$ and $m_c \ll m_Q$ we have the following ansatz:

$$\delta\Gamma^{\mathrm{PI}}_{\mathrm{osc}} \simeq \frac{G^2}{2} 2a_1 a_2 \,\mathrm{const}\, \sqrt{3\pi} f_B m_c \frac{\beta^{9/2}}{M_{H_Q}^{9/2}}$$

$$\times \sum_k (-1)^k \vartheta(M_{H_Q} - \pi\beta\sqrt{k}). \tag{67}$$

The amplitude of oscillations in PI scales down at least as $1/m_Q^5$. At large m_Q the threshold widths are much smaller than even the individual principal widths saturating Γ^{PI} at $m_{c,u,d} \neq 0$.

When $m_{u,d}$ are nonzero, the picture changes essentially in two respects. First, neither decay amplitude vanishes at the threshold, since the two-momentum of the lighter meson does not vanish: $q_0 \simeq M_1$ instead of $q_0 = |\vec{q}| \simeq M_{H_Q} - M_{\mathrm{thr}}$ if $q^2 = 0$. Second, the phase space factor $1/|\vec{p}|$ becomes now $[2M_1(M_{H_Q} - M_{\mathrm{thr}})]^{-1/2}$ vs $1/(M_{H_Q} - M_{\mathrm{thr}})$ for $M_1 = 0$. (M_1 is the mass of the lighter meson and its momentum is called q here. We assume that M_1 is much larger than the resonance spacing $\sim \beta^2/m_Q$.) Otherwise, the scaling of the transition amplitudes remains the same. Therefore, in this case we have

$$\delta\Gamma^{\mathrm{PI}}_{\mathrm{osc}} \propto \frac{G^2}{2} 2a_1 a_2 m_c m_\pi^{1/2} \frac{\beta^5}{M_{H_Q}^5} \sum_k (-1)^k \frac{\vartheta(M_{H_Q} - M_{\mathrm{thr}}^{(k)})}{\sqrt{M_{H_Q} - M_{\mathrm{thr}}^{(k)}}},$$

$$M_{\mathrm{thr}}^{(k)} \simeq m_\pi + \pi\beta\sqrt{k}. \tag{68}$$

Strictly speaking, at $m_{u,d} \gtrsim \beta$ additional light meson states contribute, and the pattern of the threshold spikes in Γ^{PI} becomes less even reflecting the superposition of a number of similar structures. Additionally, at $m_u \sim m_c$ the individual sign-alternating behavior becomes more complicated.

In principle, with all final-state masses not vanishing, there are thresholds corresponding to decays where both final-state mesons have masses constituting a finite fraction of m_Q. The threshold amplitudes for such decays, however, are too strongly suppressed, since both interfering amplitudes have the chiral and the form factor suppression:

$$|\mathcal{M}_{kn}|^2_{\mathrm{PI}} \propto \frac{m_q^2 \beta^9}{m_Q^7} \quad \text{at} \quad k,n \sim \frac{m_Q^2}{\beta^2}.$$

The phase space for such decays is also smaller since $|\vec{p}| \sim \sqrt{m_Q}(M_{H_Q} - M_{\mathrm{thr}})^{1/2}$. Additionally, these thresholds are spaced very closely, at distances scaling as β^4/m_Q^3. Therefore, they are subdominant for duality violation. They are related to the subseries of the OPE terms which appear only beyond the tree-level perturbative computations.

To conclude this section, let us describe the physical picture which emerges from the analysis. In particular, we can see how the interpretation problem mentioned in the Introduction is resolved. As expected, the explicit analysis yielded nothing about colored diquark correlator, directly. Instead, we observe the duality of differently combined quark-antiquark pairs to the hadronic states: one energetic quark (c or d) is to be combined with the "wee" spectator antiquark or slow $\bar{u}$ produced in the weak vertex. It is the pair of quarks picked up from the different final state mesons that corresponds to the large invariant mass in the quark diagram. The completeness of the hadronic states—or, in other words, the duality between the parton-level and mesonic states—is achieved already for a single fast moving decay quark when it picks up a slow spectator. In particular, the "hardness" of these processes determining the applicability of the quasifree approximation, is governed by the energy of the fast quark rather than by the invariant mass of the pair.

There is nothing wrong with considering the colored diquark loop as nearly free. Since the overall color is conserved in the perturbative diagrams, in the full graph for the meson decay which would include explicitly propagation of the spectator, there is always a color mate for any quark in the "partonic" part of the diagram. Moreover, if the leading-N_c contribution is considered, such a color pairing (i.e., which pair must be embodied into a meson) is unambiguous.

Of course, the invariant *mass* of a single, even fast on-shell quark vanishes. However, this does not make the inclusive probability for it to hadronize by picking up spectator and forming a meson, a "soft" quantity. For it is not the invariant mass but the (rest frame) momentum that determines the hardness. Indeed, the color of the initial static heavy quark Q is compensated by the slow spectator. This initial distribution of the color field marks the rest frame and makes the hardness parameter for the total probability to

look noninvariant if the final state is considered perturbatively as a pair of free partons.

V. TOTAL (SPECTATOR-FREE) WIDTH THROUGH $1/m_Q^2$

The inclusive decay widths of heavy hadrons in the 't Hooft model in the absence of the flavor-dependent spectator effects were considered in detail in Ref. [11]. It was demonstrated that the analytic summation of the widths for the accessible two-body modes reproduces the $1/m_Q$ expansion of the widths in the OPE, at least through the terms high enough in $1/m_Q$. In particular, the hadronic width does not have any $1/m_Q$ correction which would not be present in the OPE.

The analysis was performed for arbitrary m_c and m_{sp}, but simplified significantly when $m_u = m_d = 0$ was set (in the notations of the present paper). Since there is little doubt that the dependence of the hadronic width on m_u and m_d is suppressed by at least two powers of $1/m_Q$, this simplification cannot affect the conclusion regarding possible non-OPE $1/m_Q$ terms in the width. Nevertheless, we find it instructive to describe the direct computation of the terms $\sim m_{u,d}^2/m_Q^2$ in the width based on the 't Hooft eigenstate problem, following the approach of Ref. [11] and the analysis of the previous sections.[5] In particular, it illustrates that the case of nonzero masses is not any different from $m_u = m_d = 0$.

As before, we assume for simplicity that $m_u = m_d$, so that the $\bar{u} \gamma_\mu d$ current is strictly conserved, and represent the large-N_c nonleptonic decay width as an integral of the differential semileptonic width $\Gamma_{sl}(q^2)$ over q^2 weighted with the spectral density $\rho(q^2)$, Eq. (14). The upper limit of integration q_{max}^2 comes from vanishing of $\Gamma_{sl}(q^2)$ at $q^2 > (M_{H_Q} - M_0^D)^2$.

$\Gamma_{sl}(0)$ was analytically calculated in Ref. [11]:

$$\Gamma_{sl}(0) = \frac{G^2}{4\pi} \frac{m_Q^2 - m_q^2}{m_Q} \left[\frac{m_Q}{M_{H_Q}} \int_0^1 \frac{dx}{x} \varphi_{H_Q}^2(x) + \mathcal{O}\left(\frac{1}{m_Q^5}\right) \right]$$

$$= \frac{G^2}{4\pi} \frac{m_Q^2 - m_q^2}{m_Q} \left[\frac{\langle H_Q | \chi_Q^\dagger (m_Q/i\partial_-) \chi_Q | H_Q \rangle}{2 M_{H_Q}} \right.$$

$$\left. + \mathcal{O}\left(\frac{1}{m_Q^5}\right) \right], \qquad (69)$$

which coincides with its OPE expansion.

On the other hand, $\rho(q^2)$ does not vanish at $q^2 > 0$ only due to nonzero $m_{u,d}$. Therefore, following the approach of the previous section, we expand $\rho(q^2)$ in m_q^2/q^2 at large q^2 and, simultaneously, $\Gamma_{sl}(q^2)$ in q^2/m_Q^2 at $q^2 = 0$. To this end we write the width as

---

[5]The consideration below was elaborated while working on Ref. [11], but was not included in the final version for the sake of brevity.

$$\Gamma_{H_Q} = \int_0^{q_{max}^2} dq^2 \, \rho(q^2) \Gamma_{sl}(q^2) = \Gamma_{sl}(0) \int_0^\infty dq^2 \, \rho(q^2)$$

$$+ \int_0^\infty dq^2 \, \rho(q^2) [\Gamma_{sl}(q^2) - \Gamma_{sl}(0)]. \qquad (70)$$

Since at $q^2 \sim m_Q^2$ the spectral density $\rho(q^2)$ is explicitly proportional to m_q^2/m_Q^2, in the second integral we can use for $\Gamma_{sl}(q^2) - \Gamma_{sl}(0)$ its leading-order approximation. With $\int dq^2 \rho(q^2) = N_c$, the first term *exactly* reproduces the corresponding term in the OPE, Eq. (69).

All transition form factors are expandable in q^2/m_Q^2 (except, possibly, the point $q^2 \to m_Q^2$). Therefore, the smeared width $\Gamma_{sl}(q^2)$ is likewise expandable in q^2/m_Q^2, and so is Γ_{H_Q}—at least up to small corrections $\sim 1/m_Q^3$ associated with the domain of q^2 close to m_Q^2. Thus, the second term scales only as $m_{u,d}^2/m_Q^2$.

In order to calculate this term, we can use the following facts regarding the smeared width.

The exact ''semileptonic'' width $\Gamma_{sl}(q^2)$ coincides with the free width $\Gamma_{sl}^{tree}(q^2)$ to the leading order in $1/m_Q$ when $m_Q - \sqrt{q^2} \gg \beta$.

The smeared $\Gamma_{sl}(q^2)$ is an analytic function of q^2 and is expandable in $1/(m_Q - \sqrt{q^2})$.

The smeared $\Gamma_{sl}(q^2)$ does not blow up at $q^2 \to m_Q^2$. We shall comment on them below. Accepting these three facts for now, and neglecting m_q^2 compared to m_Q^2 we find

$$\Gamma_{sl}^{tree}(q^2) = \frac{G^2}{4\pi} m_Q \vartheta(m_Q^2 - q^2), \qquad (71)$$

and

$$\int_0^\infty dq^2 \, \rho(q^2) [\Gamma_{sl}(q^2) - \Gamma_{sl}(0)]$$

$$= \int_0^\infty dq^2 [\Gamma_{sl}(q^2) - \Gamma_{sl}(0)] \frac{1}{q^4} \frac{d}{d \ln q^2} \int_0^{q^2} dt \, t \rho(t)$$

$$= - \frac{G^2 m_Q N_c}{4\pi} \int_{m_Q^2}^\infty \frac{dq^2}{q^4} (m_u^2 + m_d^2)$$

$$= - N_c \frac{m_u^2 + m_d^2}{m_Q^2} \Gamma_{sl}(0). \qquad (72)$$

Here we have used the sum rules Eqs. (43), (44).

Thus, we get through order $1/m_Q^2$

$$\Gamma_{H_Q} = - \frac{G^2}{4\pi} N_c \frac{m_Q^2 - m_c^2 - m_u^2 - m_d^2}{m_Q} \left(1 + \frac{\beta^2}{m_Q^2} - \frac{\mu_\pi^2}{2m_Q^2} \right)$$

$$+ \mathcal{O}\left(\frac{\beta^3, m_q^3}{m_Q^2}\right), \qquad (73)$$
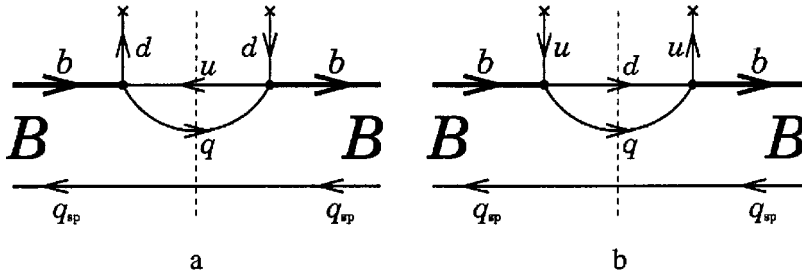
where the kinetic expectation value

FIG. 5. Diagrams responsible for the linear in $m_{u,d}$ corrections to Γ_{H_Q}.

$$\mu_\pi^2 = m_Q^2 \left[\int_0^1 dx\, x^2 \varphi_{H_Q}^2(x) - \left(\int_0^1 dx\, x \varphi_{H_Q}^2(x) \right)^2 \right] \quad (74)$$

represents the low-momentum part of the $1/m_Q$ expansion of the integral in Eq. (69), whereas the term β^2/m_Q^2 accounts for its "hard" part [11]. The OPE result is thus reproduced explicitly.

Neglecting m_c compared to m_Q was essential in the above computation. For, at $m_c \neq 0$ the unique *linear* in $m_{u,d}$ effect appears which was calculated in Ref. [11], Sec. VI A:

$$\frac{\Delta\Gamma_{H_Q}}{\Gamma_{H_Q}} = -\frac{4\pi}{N_c} \langle 0 | m_u \bar u u + m_d \bar d d | 0 \rangle \frac{m_Q m_c}{(m_Q^2 - m_c^2)^2}. \quad (75)$$

Of course, it has the direct OPE counterpart, see Fig. 5. At $m_c \sim \beta$ it scales as $\beta^2(|m_u| + |m_d|)/m_Q^3$.

It is straightforward to obtain the $1/m_Q^3$ term in Γ_{H_Q} as well. The least trivial corrections not related to m_u, m_d are all incorporated in Eq. (69). The only remaining part is the $1/m_Q$ term in the explicit $(m_u^2 + m_d^2)/m_Q^2$ correction. It is associated with the domain of maximal q^2, i.e., $m_Q - \sqrt{q^2} \sim m_c, \beta$. It has a logarithmic enhancement, $[(m_u^2 + m_d^2)m_c/m_Q^3]\ln[m_Q/m_c]$ coming from the domain $m_c \ll m_Q - \sqrt{q^2} \ll m_Q$ (and, therefore, this log conforms with the one in the free quark phase space). On the other hand, at nonequal m_u and m_d and light c and q_{sp} the smeared width contains the chiral log of the form $[(m_u - m_d)^2\beta/m_Q^3]\ln[\beta/(m_c + m_{sp})]$, from the domain $\sqrt{\beta(m_c + m_{sp})} < m_Q - \sqrt{q^2} < \beta$. These contributions are given by the corresponding four-quark expectation values $\langle H_Q | \bar Q \Gamma_\mu c \bar c \Gamma_\nu Q | H_Q \rangle$, in agreement with the general proof of Ref. [24].[6] However, these operators being

---

[6] At $m_u = m_d$ these expectation values are multiplied by $\delta_{\mu\nu} - v_\mu v_\nu$ with $v = P_{H_Q}/M_{H_Q}$, containing only the spacelike component, and then the contribution of the lowest pseudoscalar, "pion" intermediate state comes proportional to its momentum. From the point of view of hadronic decay modes, the pion threshold decay amplitude likewise vanishes at $m_u = m_d$ since, for the conserved vector current, only the current-meson couplings $\sim \epsilon_{\mu\alpha}$ survive, and the decay amplitudes into the mesons with the same parity as H_Q vanish at the threshold. These suppressions are eliminated when $m_u \neq m_d$, and the chiral log appears both in the four-quark expectation value due to the pion contribution in the timelike component, and in the smeared width due to the pion phase space $\sim 1/(M_{H_Q} - M_{thr})$.

N_c subleading (the expectation values scaling as N_c rather than N_c^2, unless WA is possible), their values are anyway expressed in a rather ugly way in terms of various 't Hooft eigenfunctions both in the s and t channel. Calculating this maximal-q^2 effect requires straightforward $1/m_Q$ expansion of the form factors and wave functions, which is not too instructive. This is briefly outlined in Appendix B.

Now it is time to comment on the assumptions used in deriving the correction $-(m_u^2 + m_d^2)/m_Q^2$. The first one was related to the semileptonic width at nonzero q^2. Of course, even the stronger statement holds: $\Gamma_{sl}(q^2)$ coincides with $\Gamma_{sl}^{tree}(q^2)$ up to the terms $\sim \beta^2/(m_Q - \sqrt{q^2})^2$. Proving this goes along the same explicit $1/m_Q$ expansion of the wavefunctions and the form factors [11]. Namely, to order $1/m_Q$ one still has the transition form factors determined only by the overlap of the initial φ_{H_Q} and the final φ_k wave functions. The only difference is that $q_- \neq 0$ at $q^2 \neq 0$ and, as a result, the arguments of the wave functions change. This purely kinematic modification accounts for all changes to order $1/m_Q$. Some technicalities are given in Appendix B.

The situation changes at order $1/m_Q^2$. The peculiarity of the point $q^2 = 0$ is that the vertices do not renormalize at $q^2 = 0$. More precisely, in the light-cone gauge the vertex corrections in Fig. 6(a) are proportional to q_- and thus vanish, to all orders, at $q^2 = 0$ [11]. The only surviving contribution is the effect of renormalization of the external quark legs, Fig. 6(b) which is readily summed up to all orders. In
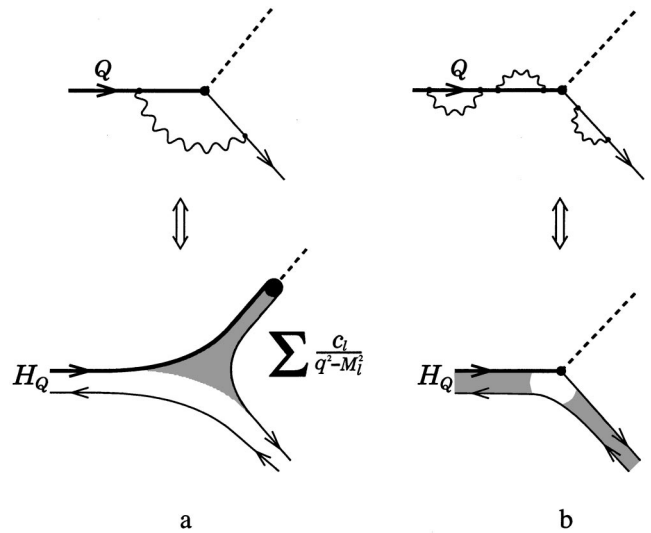

FIG. 6. Vertex (a) and external quark leg (b) renormalizations of weak decay amplitudes, and their hadronic counterparts.

the light-cone formalism it consists in overall (IR divergent) shift in the reference point for the light-cone energy p_+, and the dispersion term $-\beta^2/(2p_-)$ formally coinciding with the replacement $m^2 \to m^2 - \beta^2$ for all quark flavors.

This perturbative computation has an exact analogy for the actual meson form factors. The absence of the vertex corrections means the absence of the t-channel resonance contributions at $q_- = 0$. This fact is directly seen in the explicit expressions for the form factors in the kinematics where the fraction of the total light-cone momentum ω corresponding to the momentum transfer q, goes to zero. All the strong interaction effects occur in the initial and the final states and are described by using the exact eigenfunctions instead of the plane waves.

At $q_- \neq 0$ the vertex corrections appear to order g_s^2, and already the first g_s^2 correction in the coefficient function of the operator $\bar{Q}Q$ becomes different from $(1 - \beta^2/m_Q^2)^{-1/2}$. The vertex corrections are dual to the t-channel resonance contribution in the exact amplitude, and they also appear explicitly with the factors q_- and β^2. Additionally, the original tree-level overlap is modified both kinematically and due to the additional terms in the expansion of the $\bar{Q}\gamma_\mu q$ current in terms of the light-cone spinors. All this is directly observed in the explicit expressions for the meson transition form factors.

Regarding the two other assumptions, the fact that the transition form factors are regular functions of q^2 can be seen, of course, from their most general analytic expressions. Another way to visualize this is to use the approximate scaling behavior

$$\frac{1}{\sqrt{2m_Q}} \langle k | J_\mu(q) | H_Q(m_Q) \rangle \simeq \frac{1}{\sqrt{2m_Q'}} \langle k | J_\mu(q') | H_Q(m_Q') \rangle,$$

(76)

where q' is adjusted in respect to m_Q' to have the same rest-frame momentum of the final-state meson k:

$$q'^2 = \frac{M_{H_Q}(m_Q)}{M_{H_Q}(m_Q')} q^2 + M_{H_Q}(m_Q')[M_{H_Q}(m_Q') - M_{H_Q}(m_Q)]$$

$$+ M_k^2 \left(1 - \frac{M_{H_Q}(m_Q)}{M_{H_Q}(m_Q')} \right).$$

(77)

This freedom can be used, say, to make momentum transfer q lightlike,

$$m_Q' \simeq m_Q - \left[\frac{m_Q^2 - M_k^2 + q^2}{2m_Q} - \sqrt{\left(\frac{m_Q^2 - M_k^2 - q^2}{2m_Q} \right)^2 + \frac{q^2 M_k^2}{m_Q^2}} \right]$$

(78)

and then use the exact representation of the amplitudes *via* the simple wave functions overlaps. This trick allows one, for example, to see the strong suppression of the transition amplitudes to highly excited states $M_k^2 \sim m_Q^2 \gg \beta m_Q$ at non-zero q^2 as well, without going into details described in Appendix B. The relation Eq. (76) is violated by the β^{2n} radia-

tive corrections and by the subleading $1/m_Q$ terms in the expansion of the current J_μ, which are both power suppressed here. It can be derived directly from the explicit solution of the model, see Secs. 2 and 4 of Appendix B.

VI. COMMENTS ON THE LITERATURE

Two recent papers [27,17] claimed to have established inapplicability of the OPE for the heavy flavor widths considering the solvable 't Hooft model. Reference [27] addressed the conventional spectator-independent decay channels. Numerically evaluating the possible two-body decay rates for the values of m_Q up to $m_Q = 14$, it found a small systematic excess of the decay width over the free quark diagram which was fitted as

$$\frac{\Gamma_{H_Q} - \Gamma_Q}{\Gamma_Q} \sim 0.15 \frac{\beta}{m_Q}.$$

(79)

This was regarded as the demonstration of (*a priori* proclaimed) nonexistence of the OPE for the nonleptonic widths. Since in the large-N_c limit the difference between the nonleptonic and semileptonic widths disappears, this, if true, would mandate the same absence of the OPE for the semileptonic widths as well, an obvious fact ignored by the authors.

In contrast, Ref. [11] accomplished the analytic summation of the large-m_Q width in the 't Hooft model, and no deviation from the OPE was found to the high enough orders in $1/m_Q$ (the exact power addressed depended on the values of the final state masses). In particular, the absence of the $1/m_Q$ corrections to the parton result for $m_{u,d} \sim \beta \ll m_Q$ was very transparent. Additionally, the correspondence was established between the step-by-step quark-gluon based OPE computations and the matching contributions in the integrals determining the meson form factors in terms of the 't Hooft wave functions. This allows one to compare the OPE computations to the hadronic saturation at the intermediate stages rather than only for the final result.

What could go wrong with the numerical analysis of Ref. [27]? We note that the apparent ''effect'' was really small numerically and, in the fiducial range of $m_Q/\beta \sim 8$ to 10 constituted only about 1%. Moreover, a closer look at the plots of Ref. [27] shows that the reported discrepancy somewhat increased toward the upper values of m_Q. As a result, a better fit of the numerical points of Ref. [27] would be achieved assuming small $\mathcal{O}(1/m_Q^0)$ (not power suppressed) corrections to the width, with the numerical coefficient $\sim \mathcal{O}(10^{-2})$. Unfortunately, *a priori* targeting the $1/m_Q$ terms, the authors did not explore the possibility of alternative interpretations.

Incidentally, the scale of the claimed discrepancy lies just in the magnitude range of the $1/m_Q^2$ effects from the OPE which could be *a priori* expected at $m_Q \approx 10$. It turns out that the numerical points for $\Gamma(m_Q)$ in Ref. [27] can be well fitted by the leading-m_Q expression and adjustable β^2/m_Q^2 terms with the coefficient $\mathcal{O}(1)$. Adding arbitrary corrections $\sim \beta^3/m_Q^3$ with the larger coefficient up to 6–8 allows really perfect fits. Since the actual $1/m_Q^3$ corrections include the

expectation values of the four-quark operators which are not given by factorization, they are uncertain. Their estimates showed that they are typically significantly enhanced compared to the naive dimensional estimate $\sim \beta$.

Thus, the question would remain open before the actual OPE corrections are computed. This was accomplished in Ref. [11]. While the radiative correction enhances the width by the factor $1+\beta^2/(2m_Q^2)$ to order $1/m_Q^2$, the kinetic term $-\mu_\pi^2/(2m_Q^2)$ tends to suppress it.[7] Therefore, if, as stated in Ref. [27], the plotted values refer to the bare mass m_Q, it seems that the numerical calculation reported by Grinstein and Lebed yielded the result exceeding the asymptotic width by an amount ranging from a fraction to a per cent.

The most apparent resolution, in our opinion, is that the analysis of Ref. [27] simply does not control the accuracy of the numerical computations of the amplitudes at the required level of a few per mill. This is not surprising and roots to the well-known problems of numerical computations. The width at large m_Q is saturated by highly excited states whose wave functions oscillate fast. This standard problem of the numerical computations of the semiclassical transition overlaps is additionally plagued by the general complexity of the numerical solutions of the 't Hooft equation and the encountered singular integrals. These problems were alluded to in Ref. [27].

The general lesson one draws from this comparison is not new: it is not easy to correctly evaluate the actual asymptotic width in too straightforward numerical summations of many exclusive widths, were it a computer-simulated approximate solution for a theoretical model with—theoretically speaking—potentially unlimited accuracy, or the actual experimental data. The approximations made in practical applications usually go across preserving the subtle interplay of different effects which underlies delicate cancellations shaping the size and scaling of the power suppressed nonperturbative effects.

The recent paper Ref. [17] announced even more drastic numerical mismatch of the actual hadronic width vs the OPE considering WA, even in the leading order at which the effect appears. In Secs. III and IV, on the contrary, we analytically computed the asymptotics of the spectator-dependent widths and found exact correspondence with the OPE. As a matter of fact, Ref. [17] contains self-contradicting assertions: It was stated that the (smeared) current-current vacuum correlator [of the type determining, say, $\sigma(e^+e^- \rightarrow \text{hadrons})$] is known to obey the OPE. Simultaneously, the difference was claimed established for the effect of WA in the context relying on the factorization expression Eq. (27). The starting expression—whether correct or not for a particular model—thus equates the validity of the OPE for the WA nonleptonic width to its applicability for the vacuum current correlator. Figure 7 shows the ratio of the (smeared) values of the absorptive parts of the correlator
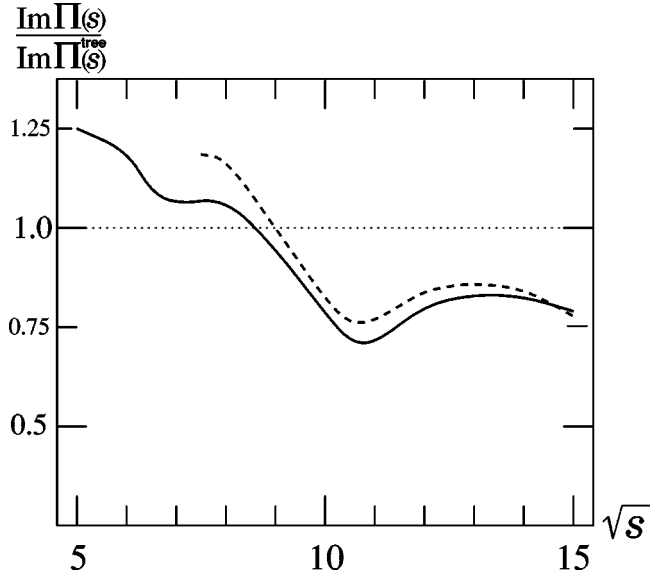

FIG. 7. The actual-to-partonic ratio of the absorptive part of the vector current correlator at $m_u = m_d = 0.56\beta$, according to Ref. [17], for $N_c = 10$ (solid line) and $N_c = 20$ (dashed line). Energy scale is in units of β. Smearing procedure has been applied for the hadronic cross section.

$$\frac{\text{Im}\,\Pi_{\mu\mu}(q^2)}{\text{Im}\,\Pi_{\mu\mu}^{\text{tree}}(q^2)}$$

according to Ref. [17]. The variance from unity, in general, does seem to be present.

We do not think that this warrants one to become cautious in applying the OPE to the polarization operator. There are certain drawbacks in the computations of Ref. [17]. For unknown reasons the authors presumed that the width Eq. (27) becomes nonintegrable for the large-N_c theory and, therefore, cannot be sensibly smeared.[8] As a result, instead of considering the current correlator in the 't Hooft model itself, a rather *ad hoc* ansatz was adopted which was intended to mimic nonvanishing widths of the resonances at finite N_c. The simple resonant representation for $\Pi(q^2)$,

$$\Pi(q^2) = N_c \sum_n \frac{c_n^2}{q^2 - M_n^2 + i\epsilon},$$

$$-\text{Im}\,\Pi(q^2) = \pi N_c \sum_n c_n^2 \delta(q^2 - M_n^2) \qquad (80)$$

was replaced by a complicated model where $\text{Im}\,\Pi(q^2)$ was due to the two-meson states, with their production amplitude containing the resonance terms

----

[7]The value of μ_π^2 for quark masses used in Ref. [27] has been recently evaluated by Lebed to be about $0.8\beta^2$ (private communication).

[8]It is obvious that for any regularization consistent with unitarity, the width remains integrable, with the integral around the spike independent of the regularization. This is ensured by the dispersion relation; see Sec. III.

$$\frac{\sqrt{N_c}c_n}{q^2 - M_n^2 + iM_n\Gamma_n}.$$

At large N_c the residues $\sqrt{N_c}c_n \sim \sqrt{N_c}$ while the resonance decay amplitudes $\mathcal{M}_{nkl} \sim 1/\sqrt{N_c}$; the widths $\Gamma_n \sim 1/N_c$.

As soon as Γ_n are small enough, such a prescription does not differ from the proper spectral density Eq. (80) being a concrete functional choice of the δ distributions,

$$\int_{|q^2 - M_n^2| \gg M_n\Gamma_n} dq^2 \frac{c_n^2}{|q^2 - M_n^2 + iM_n\Gamma_n|^2} = \pi c_n^2 \frac{1}{M_n\Gamma_n},$$
(81)

provided each Γ_n is saturated by the included decay modes:

$$\sum_{k,l} |\mathcal{M}(n \to kl)|^2 \Phi_{kl} = 2M_n\Gamma_n.$$
(82)

Here Φ_{kl} denote the two-body decay phase space. Under these constraints such a model—if not true—is at least legitimate. However, two conditions must be observed: First, the resonances do not overlap, $\Gamma_n \ll |M_{n \pm 1} - M_n|$. Additionally, the partial decay amplitudes and the phase space factors must be practically constant within the (total) width of the individual meson.

Both constraints are satisfied in the formal limit $N_c \to \infty$. However, in practice the width grows with the mass of the resonance, while the spacing between the successive ones decreases. Therefore, for finite N_c adopted in Ref. [17] (let alone the considered case $N_c = 1$) the first condition was not well respected. Regarding the second constraint, the problem gets additionally aggravated by the singular two-body phase space in $D = 2$. Therefore, it seems quite probable that the reported disagreement is rooted to the inconsistencies of the model for $\text{Im}\,\Pi(q^2)$ adopted in Ref. [17].

A detailed look at the plots displayed there incidentally provides support for this conjecture. The plots for (the smeared) $\text{Im}\,\Pi(q^2)$ show a rather unphysical shoulder at $m_Q \approx 10$ which coincides with the mass of the resonance having abnormally large width (Fig. 9 of Ref. [17]). On the other hand, the agreement between the hadronic and quark widths is surprisingly good just at *lower* masses where the resonances are more narrow. The point where the curves start to diverge, apparently shifts upward with increasing N_c.

Finally, a simpler question remains open about the accuracy of numerical computations employed in Ref. [17]. Taken at face value, even adopted sampling of m_Q when only 2 to 3 points fall inside a separate sharp resonance peak, seems insufficient to evaluate the smeared width reliably. Whether these effects can explain the observed discrepancy $\sim 20\%$ at $m_Q \sim 10$ to 15, remains to be clarified.[9]

---

[9]Since the effect in question by itself is $1/q^2$ in the polarization operator, it could be *a priori* conceivable to have subleading $1/q^3$ corrections as large as 20%. However, the OPE ensures that the corrections are suppressed by at least two powers of q, and thus must be small; they are readily calculated.

It is worth reiterating that the computational difficulties— quite significant in the analysis of Ref. [17]—to a large extent were a handmade problem. Both c_n and M_n are readily computed without cumbersome triple overlaps involving singular integrals, and in any case had been determined to construct the authors' model. Computing the smeared width directly from Eq. (80) would be then quite straightforward.

VII. CONCLUSIONS

We have examined the inclusive decay widths of heavy flavor mesons in the 't Hooft model in the context of the heavy quark expansion, paying attention to the spectator-dependent effects sensitive to the flavor of the spectator quark. To the order the high-energy asymptotics are calculated, there is no deviation from the OPE predictions, either for semileptonic, nonleptonic decays or for the $e^+e^- \to$ hadrons-type processes—as anticipated.

We confirm that there is no difference for the OPE whether a semileptonic or nonleptonic width is considered. What matters is only whether the particular observable can be represented as the *complete* discontinuity of the properly constructed correlation function over the cut in a suitable "hard" variable. This variable ω is the same for both semileptonic and nonleptonic decays, Eq. (29). After that the only difference is the number of the quark Green functions to be multiplied and to perform the expansion of the product in the complex plane. The validity of the OPE, of course, cannot depend on these technicalities.

This does not mean, however, that for all types of decays the predictions of practical OPE truncated after the first few terms, work with equal accuracy at a fixed mass m_Q. On the contrary, considerations of Ref. [28] suggest that the effective "hardness" scale can be smaller than literally the energy release. Accordingly, higher onset of duality and larger deviations for nonleptonic widths were obtained in Ref. [12] in the instanton-based model.

Therefore, in our opinion, attempts to check (let alone to disprove) the OPE itself in the concrete model are hardly meaningful beyond illustrative purposes. What has a potential of providing useful insights, is studying the behavior of the contributions violating local duality, so far the least understood theoretically subject.

The "practical" OPE yields the width in the power expansion

$$\frac{\Gamma_{H_Q}}{\Gamma_0} = A_0 + \frac{A_1}{m_Q} + \frac{A_2}{m_Q^2} + \cdots.$$
(83)

If the series in $1/m_Q$ were convergent (to the actual ratio), Γ_{H_Q} would have been an analytic function of m_Q above a certain mass m_0 pointing to the onset of the *exact* local parton-hadron duality. The actual Γ_{H_Q} is definitely nonanalytic at any threshold (whether or not the amplitude vanishes at the threshold). Thus, the "radius of convergence" cannot correspond to the mass smaller than the threshold mass. Since in the actual QCD the thresholds exist at arbitrary high

energy, the power expansion in Eq. (83) can be only asymptotic, with formally zero radius of convergence in $1/m_Q$.

In practice, the true threshold singularities are expected to be strongly suppressed at large energies, and the corresponding uncertainties in the OPE series quite small. In actual QCD they are expected to be exponentially suppressed eventually, though, possibly, starting at larger energies. In the intermediate domain they can decrease as a certain power and must *oscillate*.

As was illustrated in Ref. [11], the power expansions like Eq. (83) are meaningful even beyond the power suppression where the duality-violating oscillations show up. In the case of the heavy quark widths where mass m_Q cannot be varied in experiment, the size of the duality-violating component may set the practical bound for calculating the widths. Thus, it is important to have an idea about its size. We emphasize that one should always include the leading QCD effects to the partonic expressions, rather than compare the actual observable with the bare quark result. In the model considered in Ref. [11], incorporating the power corrections from the practical OPE suppressed the apparent deviations by more than an order of magnitude.

It is worth noting that we identified the case where the exact quark-hadron duality in the interference width is saturated on a *single* final state. It is realized in the chiral limit for the final state quarks. It seems to be just an opposite case to the classical small velocity limit for semileptonic decays noted by Shifman and Voloshin in 1986 [7]. While applicable only for large N_c and in $D=2$, the present case is peculiar in that the duality is not affected by subleading power corrections. Both cases serve as a counter-example to the lore often purportingly equating the accuracy of local duality to the proliferation of the final state channels in the process in question.

The performed analysis elucidates how the general duality between the partonic and hadronic widths works out its way at large energies. It can be traced that, to the leading order in $1/m_Q$, the duality is simply the completeness of eigenstates of the hadronic Hamiltonian. It is not even required to explicitly solve the 't Hooft equation to establish the leading free quark result for the width, but just to know that the solutions form a complete basis.[10] The absence of the non-OPE terms already at the next to leading order $1/m_Q$ is a dynamical fact requiring, for example, a proper solution of the bound-state equations. This enters *via* the average of the moments of the invariant hadronic mass in the final state.

The 't Hooft model gives an explicit example of the duality-violating effects, at least of the "minimal," resonance-related nature. As a general feature, we observe that the duality between the appropriately averaged (to eliminate the threshold singularities) hadronic widths and the truncated OPE predictions sets in numerically well rather

early, after only a couple of the principal thresholds, and the local duality works much better in the heavy quark widths, nonleptonic and semileptonic than in $\sigma(e^+e^- \rightarrow \text{hadrons})$. This applies to both the qualitative behavior (generally, a larger density of resonances implying a more narrow minimal interval of smearing) and to the strength of the threshold singularities as well as the threshold residues. The qualitative discussion of the underlying reason can be found in Ref. [29], Sec. 3.5.3.

Can these lessons be transferred to actual QCD? Unfortunately, there are some essential differences between it and the explored 't Hooft model, which must be important for the local duality violation.

The singular $1/|\vec{p}|$ two-body phase space in $D=2$ strongly enhances the threshold singularities, compared to $|\vec{p}|$ in $D=4$. In the actual QCD enhancement of the nonsmooth behavior can rather be expected only from single resonances with masses close to M_B. While infinitely narrow at $N_c \rightarrow \infty$, they acquire significant width for $n_f/N_c=1$ which leads to drastic flattening of the resonance-related combs when the width becomes comparable to the distance between the successive resonances. Additionally, one expects a denser resonance structure, at least asymptotically, in the actual QCD in four dimensions (even at $N_c \rightarrow \infty$) than the equal in M^2 spacing in $D=2$. All this would lead to suppression of the duality violation.

Finally, similar types of the condensate corrections lead to a weaker suppression in $D=2$ due to the smaller dimension of the corresponding operators. This provides a larger room for various possible nonperturbative effects.

These obvious differences would optimistically suggest that the 't Hooft model represents, in a sense, an upper bound for violations of local duality in QCD. While this is not excluded, such implications must be regarded with caution.

In fact, the largest duality violating effects in the nonleptonic heavy flavor decays can be expected from the resonance structures in the combined $(\bar{q}q\bar{q}q)$ channel in the final state, embedding the quarks belonging to both the "semileptonic" and "hadronic" subprocesses which do not decouple completely. Such states are lost in the 't Hooft model. Moreover, the limit $N_c \rightarrow \infty$ simply erases the difference between the nonleptonic and semileptonic decays in this respect.

Another peculiarity of QCD in two dimensions is absence of the real gluonic degrees of freedom and of the perturbative logarithmic short-distance corrections typical for $D=4$ (in contrast to the powerlike ones in the super-renormalizable $D=2$ QCD). So far we have no clues if the onset of duality for excitation of the gluonic degrees of freedom is the same or noticeably larger than for the processes describing the evolution of the "valence" quarks. Correspondingly, while observing a good qualitative duality between the bare quark computations and the actual hadronic probabilities already at energies ~ 1 GeV, one may have to ascend to higher energies in the Minkowski domain to reach quantitative agreement at the level of the perturbative corrections. Let us note in this respect that the "real gluons" in jet physics are reliably observed in experiment only at rather high energies. Unfortunately, two-dimensional theories do not allow one to

---

[10]The Multhopp technique employed in Ref. [27] does not automatically respect orthogonality of solutions with truncation and, therefore, in principle can lead to overestimating even the leading-m_Q coefficient.

study this interesting aspect of local gluon-hadron duality even in the accessible model settings.

As emphasized above, we found that the OPE predictions universally hold for the inclusive widths. In this respect, the origin of the paradox in mismatch between the size of the interference term on the quark and meson languages, Eqs. (2) and (4), is quite transparent. The mismatch emerges in the N_c-suppressed term. Addressing such effects requires a more accurate control of decay amplitudes beyond the leading-N_c counting rules. In particular, it is necessary to account for $1/N_c$ corrections in the color-allowed amplitudes rather than use oversimplified prescription such as Eq. (3) (a possible resolution was conjectured already in Ref. [16]).

A closer look reveals that the problem originates just on the quark configurations where all four quarks in the intermediate state have the same color. This configuration is $1/N_c$ suppressed and can be simply discarded for the leading-N_c probabilities. Within this color combination, there is an ambiguity of allocating the quarks over the colorless mesons which is absent otherwise. It is not difficult to see that the naive computations based on the amplitude prescription Eq. (3) amount to counting the two possibilities independently, which is not justified *a priori*. Moreover, it is a clear double counting at least in the simple quark picture. To phrase it differently, the bases of the final states used in the naive color rules such as Eq. (3), are nonorthogonal beyond the leading order in N_c.

With the complicated confinement dynamics, we cannot know beforehand what are the actual decay amplitudes to the particular hadronic states in this case—and, additionally, the simple two-meson picture of the final state must be, in general, extended when going beyond the leading order in N_c. However, it seems obvious that the naive prescription have little chance to be true in a complete theory if it violates general requirements even in the simplest case of almost free constituents.

This, incidentally, is an illustration of the fact that the naive factorization of decay amplitudes *must* be violated at some point at the $1/N_c$ level—either in the corrections to the color-allowed amplitudes, or at the leading level for color-suppressed amplitudes, or in both. A dedicated study of the above inconsistency was undertaken in the framework of the nonrelativistic quark model in Ref. [30]. It was found that the problem is indeed resolved there in this way, with the concrete dynamical mechanism dominating the modification of the $1/N_c$-suppressed amplitudes depending on details of the model.

In Sec. IV, considering the spectator-dependent preasymptotic correction we formulated the problem of calculating the interference width in the way that allowed us to avoid the complications associated with the analysis of the N_c-subleading transition amplitudes, and to study it in the theoretically clean environment. Incorporating the spectator quark *via* PI made interference an effect appearing in the same order in N_c as the noninterference widths. Then the hadron-based computation could be performed consistently, and the OPE prediction was readily reproduced.

APPENDIX A: 't HOOFT WAVE FUNCTIONS IN THE SEMICLASSICAL REGIME AND THE UV DIVERGENCES

Here we outline a more accurate determination of the scaling with Λ of the smearing parameter Δ in Eqs. (41)–(44). At $M_n \gg \beta$, $m_{u,d}$ the solutions of the 't Hooft equations are nearly free and approach the massless solutions [18,31]

$$\varphi_n(x) \simeq \sqrt{2} \cos \pi n x, \quad M_n^2 \approx \pi^2 \beta^2 n. \quad (A1)$$

The first expression holds outside the end-point domains $x \to 0$ and $x \to 1$ bounded by the "classical turning points"

$$x, 1-x \simeq \frac{m^2, \beta^2}{M_n^2}. \quad (A2)$$

We again imply the simplified case of equal quark masses $m_u = m_d = m$. The semiclassical wave functions (A1) determine the ultraviolet-singular part of the Green functions.

We shall adopt the exponential regularization of the sums. Then the Green function $G(x,y;\Lambda)$ in Eq. (41) has a meaning of the Euclidean Green function for the 't Hooft light-cone Hamiltonian, with $1/\Lambda^2$ corresponding to the Euclidean (imaginary) light-cone "time" ix_+, and x,y being "coordinates" in the space of the light-cone momentum fractions. Formally, the problem is equivalent to usual one-dimensional quantum mechanics on the interval [0,1] with Hamiltonian given by the right-hand side (RHS) of Eq. (19). The first (local) terms play a role of the potential whereas the integral term includes the counterpart of the kinetic energy which is now approximately $\pi|k|$ for the "momentum" k much larger than β.

The regularized summation is straightforward and yields literally for $G(x,y;\Lambda)$

$$\frac{1}{4} \frac{\sinh \epsilon}{\sinh^2(\epsilon/2) + \sin^2(\pi/2)(x-y)}$$

$$+ \frac{1}{4} \frac{\sinh \epsilon}{\sinh^2(\epsilon/2) + \sin^2(\pi/2)(x+y)}, \quad \epsilon \equiv \frac{\pi^2 \beta^2}{\Lambda^2}.$$

$$(A3)$$

Since the small-n eigenfunctions can be quite different from Eq. (A1), there can be, *a priori*, additional regular (ϵ-independent) terms in Eq. (A3). However, since at $\epsilon \to 0$ the limit of G must yield δ function, they are absent.

The first term in Eq. (A3) clearly yields $\delta(x-y)$ at $\epsilon \to 0$:

$$\frac{1}{4}\frac{\sinh\epsilon}{\sinh^2(\epsilon/2)+\sin^2(\pi/2)(x-y)}\approx\frac{\epsilon}{\epsilon^2+\pi^2(x-y)^2},$$
(A4)

with the width Δ at finite ϵ amounting to

$$\Delta=\frac{\epsilon}{\pi}=\frac{\pi\beta^2}{\Lambda^2}.$$
(A5)

The second term in Eq. (A3) seems to appear when both x and y are small, $x,y\lesssim\epsilon$ or $1-x,1-y\lesssim\epsilon$. In this domain, however, wave functions $\varphi_n(x)$ with $n\lesssim\Lambda^2$ are different (suppressed), and in reality this term must be discarded. In practice the domain of such small x must be treated separately, which is illustrated below. Therefore, we adopt

$$G(x,y;\Lambda)\simeq\frac{1}{4}\frac{\sinh\epsilon}{\sinh^2(\epsilon/2)+\sin^2(\pi/2)(x-y)}\quad\text{at}$$

$$\frac{1}{\Lambda^2}\lesssim x,y\lesssim1-\frac{1}{\Lambda^2}.$$
(A6)

To address the end-point domains x or $(1-x)\sim1/\Lambda^2$ we need to account for the corresponding behavior of $\varphi_n(x)$ in the "classically forbidden" domain, using the language of ordinary quantum mechanics. In these domains the wave functions are suppressed:

$$\varphi_n(x)\sim\text{const}\begin{cases}\left(\dfrac{x}{x_t}\right)^{\gamma_1}&x\lesssim x_t\ll1,\\[2ex]\left(\dfrac{1-x}{1-x_t}\right)^{\gamma_2}&1-x\lesssim1-x_t\ll1,\end{cases}\quad x_t\approx\frac{m^2,\beta^2}{\Lambda^2}.$$
(A7)

Now we consider the regularized sum rule Eq. (43):

$$I_1(\Lambda^2)=\int_0^\infty q^2dq^2e^{-q^2/\Lambda^2}\rho(q^2)$$

$$=\int_0^1dx\,dy\left(\frac{m^2}{x}+\frac{m^2}{1-x}\right)G(x,y;\Lambda).$$
(A8)

Using $G(x,y;\Lambda)$ in the form Eq. (A6) we get

$$I_1(\Lambda^2)=2m^2\left[\ln\frac{\Lambda^2}{\pi\beta^2}+\text{const}\right],$$
(A9)

where the factor 2 reflects the contributions of both $x\to0$ and $x\to1$. The logarithm is saturated at

$$|x-y|\lesssim\frac{\beta^2}{\Lambda^2},\quad\frac{\beta^2}{\Lambda^2}\ll x\ll1$$
(A10)

(and likewise with $x\to1-x$, $y\to1-y$). In this domain the expression for $G(x,y;\Lambda)$ is legitimate.

Now we show directly that the domain $x\lesssim\beta^2/\Lambda^2$ yields only a constant, that is, a nonsingular contribution at $\Lambda\to\infty$. First, we note that Eq. (38) ensures an upper bound on c_n, const/M_n^2. More precisely,

$$|c_n|\lesssim\text{const}\frac{m^2}{\gamma}\frac{1}{M_n^2}\propto\frac{1}{n}.$$
(A11)

This, of course, follows also from Eqs. (44), (45).[11] Then we can bound from above the small-x contribution as

$$\left|\int_{x\lesssim1/\Lambda^2}dx\,dy\left(\frac{m^2}{x}+\frac{m^2}{1-x}\right)G(x,y;\Lambda)\right|$$

$$\lesssim\text{const}\frac{m^2}{\gamma}\sum_{n=1}^\infty\frac{e^{-\pi^2\beta^2n/\Lambda^2}}{n}\int_0^{\Lambda^{-2}}dx\frac{m^2}{x}(\pi^2nx)^\gamma$$

$$\lesssim\text{const}\frac{m^4}{\gamma^2}\sum_{n=1}^\infty\frac{e^{-\pi^2\beta^2n/\Lambda^2}}{n}\left(\frac{n}{\Lambda^2}\right)^\gamma\simeq\text{const}\frac{m^4}{\gamma^3\beta^2}.$$
(A12)

Here we assumed the cutoff in the integral over x at $\sim1/\Lambda^2$, up to an arbitrary constant. The dimensionful factor can be either β^2 or m^2, whichever is larger and determines the position of the "classical turning" point. The key property is the convergence of both the integral over small x and the sum over n due to the suppression of $\varphi_n(x)$ in the "classically forbidden" domain. Thus, the translation rule for the cutoff energy into the regularization Eq. (42) in the integral over x,y is obtained directly from the 't Hooft equation.

Using the similar technique, it is easy to establish the asymptotic duality directly for the pseudoscalar current correlator

$$\Pi_P(q^2)=N_c\sum_n\frac{d_n^2}{M_n^2-q^2},$$

$$\frac{1}{\pi}\text{Im}\,\Pi_P(q^2)=N_c\sum_n d_n^2\delta(M_n^2-q^2),$$
(A13)

where

$$d_n=\sqrt{\frac{\pi}{N_c}}\langle n|\bar{d}i\gamma_5u|0\rangle=\frac{1}{2}\int_0^1dx\left(\frac{m_d}{x}+\frac{m_u}{1-x}\right)\varphi_n(x)$$

$$=\begin{cases}\dfrac{1}{m_u+m_d}\displaystyle\int_0^1dx\left(\dfrac{m_d^2}{x}+\dfrac{m_u^2}{1-x}\right)\varphi_n(x)&(\text{even }n),\\[2ex]0&(\text{odd }n).\end{cases}$$
(A14)

(The first form is a direct representation, whereas the second expresses the pseudoscalar density as the divergence of the axial current. A certain parity relation [19] ascending to 't Hooft ensures they both hold.) To find the asymptotics of d_n we consider the logarithmically divergent sum

---

[11]While the exact coefficient in Eqs. (43)–(45) is a subtle thing addressed here, the power of n itself is simple enough.

$$\int dq^2 \frac{\rho_P(q^2)}{q^2} \propto \sum_n \frac{d_n^2}{M_n^2}$$

$$= \sum_n \frac{1}{2(m_u+m_d)M_n^2} \int_0^1 dx \left(\frac{m_d}{x} + \frac{m_u}{1-x}\right)$$

$$\times \varphi_n(x) \int_0^1 dy \left(\frac{m_d^2}{y} + \frac{m_u^2}{1-y}\right) \varphi_n(y).$$

(A15)

Equation (38) allows one to rewrite it as

$$\sum_n \frac{d_n^2}{M_n^2} = \frac{1}{2(m_u+m_d)} \sum_n \int_0^1 dx \left(\frac{m_d}{x} + \frac{m_u}{1-x}\right)$$

$$\times \varphi_n(x) \int_0^1 dy \, \varphi_n(y),$$

(A16)

which returns us to Eq. (39).

APPENDIX B: $1/m_Q$ EXPANSION OF THE HEAVY QUARK WEAK DECAY AMPLITUDES

As discussed in Sec. V, the perturbative corrections to the weak decay vertex appear to order β^2. In the light-front formalism they vanish at the kinematic point $q^2=0$ [11], since it can be realized as the configuration with $q_-=0$, for which no physical states in the t channel is possible. The loop corrections proportional to the dimensionful coupling β^2 can be, therefore, inversely proportional to m_Q^2 or $E_{\rm rel}^2 \sim (m_Q - m_q - \sqrt{q^2})^2$. Since at $m_Q - m_q - \sqrt{q^2} \sim \beta$ the process is "soft," we do not consider here this domain, and do not distinguish between the scales of m_Q and $E_{\rm rel}$. It is important that no dynamic gluon degrees of freedom exist in $D=2$. Therefore, in the 't Hooft model vertex corrections separately are perturbatively infrared finite in physical gauges. This was explicitly illustrated in Ref. [11].

As a result, the actual transition amplitudes $\langle k|J|H_Q\rangle$ up to terms β^2/m_Q^2 must be given by only some overlaps of the initial and final state wave functions, at arbitrary [though not too close to $(m_Q-m_q)^2$] values of q^2. We assume that the initial state H_Q is either a ground state or has a finite (not

scaling with m_Q) excitation number. The final state k can be arbitrary. We will show that the explicit expressions for the transition amplitudes in the 't Hooft model exhibit this parton-deduced property.

Let us consider, for example, the representation used in Ref. [27], although we put it in a slightly modified form similar to that of Refs. [32,22]. One introduces the kinematic variable ω which depends on q^2 and the final state meson mass:

$$\omega = \frac{1}{2}\left[1 + \frac{q^2 - M_k^2}{M_{H_Q}^2} - \sqrt{1 - 2\frac{q^2 + M_k^2}{M_{H_Q}^2} + \left(\frac{q^2 - M_k^2}{M_{H_Q}^2}\right)^2}\right],$$

(B1)

which, in the light-cone formalism has a meaning of the fraction of the momentum seen in the infinite-momentum frame carried by the particle with mass $\sqrt{q^2}$ in the two body decay of B meson, if another particle has mass M_k. This fraction has two possible values corresponding to the two possible directions of meson k in the rest frame. We chose the above branch to have $q_- \to 0$ as $q^2 \to 0$, as in Ref. [11]. The light-cone fraction ω has a very simple meaning in the rest frame as well:

$$(1-\omega)M_{H_Q} = (|\vec{p}_k| + E_k)_{\rm c.m.}, \quad \omega M_{H_Q} = (-|\vec{q}| + q_0)_{\rm c.m.}.$$

(B2)

Two ω-dependent $H_Q D^{(k)}$ overlaps are then considered

$$\mathcal{C}_k(\omega) = -\frac{1-\omega}{\omega} \int_0^1 dy \, \varphi_k(y)\varphi_{H_Q}[1-(1-\omega)(1-y)],$$

$$\mathcal{D}_k(\omega) = -\omega \int_0^1 dy \, \frac{\varphi_k(y)}{y} \frac{\varphi_{H_Q}[1-(1-\omega)(1-y)]}{1-(1-\omega)(1-y)}$$

(B3)

which correspond to the effects unrelated to vertex corrections.

Additionally, in the t channel of the decay process one can have various bound states of B_c mesons; we reserve index l for them, and their masses will be denoted μ_l. For each t-channel resonance there is an ω-dependent triple-meson ("$H_Q B_c^{(l)} D^{(k)}$") overlap

$$F_{lk}(\omega) = \omega(1-\omega) \int_0^1 dx \int_0^1 dy \, \frac{\varphi_l^{B_c}(x)\varphi_k(y)}{[\omega(1-x) + (1-\omega)y]^2} \{\varphi_{H_Q}(\omega x) - \varphi_{H_Q}[1-(1-\omega)(1-y)]\}.$$

(B4)

With these notations one has for the transition amplitudes

$$\epsilon_{\mu\nu}q_\mu\langle k|\bar{q}\gamma_\nu Q|H_Q(P)\rangle = \frac{\beta^2\sqrt{\pi}}{\sqrt{N_c}} \sum_l \frac{q^2 + (-1)^l\mu_l^2}{q^2 - \mu_l^2} f_l F_{lk}(\omega) - q^2 \mathcal{C}_k(\omega) + m_Q m_q \mathcal{D}_k(\omega),$$

(B5)

$$q_\mu\langle k|\bar{q}\gamma_\mu Q|H_Q(P)\rangle = \frac{\beta^2\sqrt{\pi}}{\sqrt{N_c}} \sum_l \frac{-q^2 + (-1)^l\mu_l^2}{q^2 - \mu_l^2} f_l F_{lk}(\omega) + q^2 \mathcal{C}_k(\omega) + m_Q m_q \mathcal{D}_k(\omega).$$

(B6)

Here, for convenience, we wrote the invariant combinations of the amplitudes instead of the two Lorentz components of the current separately. For example, the decay amplitude $B \to D^{(k)} + \pi^{(n)}$ in the considered case of vectorlike interactions takes the form

$$\mathcal{M}_{kn} = \frac{G}{\sqrt{2\pi}} c_n \left[\sum_l \frac{-P_n q^2 + (-1)^l \mu_l^2}{q^2 - \mu_l^2} c_l F_{lk}(\omega) \right. $$

$$\left. + P_n q^2 \mathcal{C}_k(\omega) + m_Q m_q \mathcal{D}_k(\omega) \right], \quad (B7)$$

where P_n is the parity $(-1)^{n+1}$ of the $\pi^{(n)}$ state. Here and below throughout this appendix we suppressed the factors N_c and put dimensionful $\beta = 1$. It is easy to see that at $q^2 = 0$ ($\omega \to 0$) only the term $\sim \mathcal{C}_k$ survives and this expression indeed reduces to Eqs. (56), (62).

We shall show that the first term in this relation associated with the dynamics in the t channel, is $1/m_Q^2$-suppressed compared to the terms given by overlaps $\mathcal{C}$ and $\mathcal{D}$. This implements the approximate "on-shell" condition for the heavy quark Q; in this case it can be accomplished employing the $1/m_Q$ expansion of the initial-state wave function φ_{H_Q}.

1. Nonrelativistic expansion in the 't Hooft equation

To analyze the heavy quark system, it is advantageous to introduce the nonrelativistic variables:[12]

$$M_n = m_Q + \epsilon_n, \quad t = (1-x) m_Q, \quad \text{and}$$

$$\Psi_n(t) = \frac{1}{\sqrt{m_Q}} \varphi_n \left(1 - \frac{t}{m_Q} \right), \quad (B8)$$

in terms of which the equation takes the form

$$\left(\epsilon_n + \frac{\epsilon_n^2 + 1}{2 m_Q} \right) \Psi_n(t) = \left(\frac{m_{sp}^2 - 1}{2t} + \frac{t}{2} \frac{1 - 1/m_Q^2}{1 - t/m_Q} \right) \Psi_n(t)$$

$$- \frac{1}{2} \int_0^{m_Q} ds \frac{\Psi_n(s)}{(t-s)^2}. \quad (B9)$$

The asymptotics of $\Psi_n(t)$ at $1 \ll t \ll m_Q$ is given by

$$\Psi_n(t) \simeq \frac{F_n}{t^3} \quad \text{with} \quad F_n = \sqrt{\pi m_Q} f_n = \int_0^{m_Q} dt \, \Psi_n(t). \quad (B10)$$

One can extend the bound-state problem Eq. (B9) from $[0, m_Q]$ to the whole interval $[0, \infty)$ if the exact linear potential term $t/2[(1 - 1/m_Q^2)/(1 - t/m_Q)]$ is replaced by its formal expansion in t/m_Q. This is justified literally only up to $1/m_Q^4$ terms ($1/m_Q^3$ in the wave function), when the ultraviolet divergences of the static heavy quark expansion first show

up. Further terms can be obtained using the explicit large-t asymptotics in Eq. (B10) and the end-point behavior of $\Psi_n(t)$ at $t \to m_Q$. For example, the improved nonrelativistic equation takes the form

$$\left(\epsilon_n + \frac{\epsilon_n^2 + 1}{2 m_Q} \right) \Psi_n(t) = \frac{m_{sp}^2 - 1}{2t} \Psi_n(t) + \frac{1}{2} \left[\left(1 - \frac{1}{m_Q^2} \right) t \right.$$

$$\left. + \left(1 - \frac{1}{m_Q^2} \right) \frac{t^2}{m_Q} + \frac{t^3}{m_Q^2} \right] \Psi_n(t)$$

$$- \frac{1}{2} \int_0^\infty ds \, \Psi_n(s) \frac{1}{(t-s)^2}. \quad (B11)$$

Equation (B11) can be viewed as a usual variational problem for the nonlocal Hamiltonian defined as

$$\langle \Psi | \mathcal{H} | \Psi' \rangle = \int_0^\infty dt \, \Psi(t) \left[\frac{m_{sp}^2 - 1}{2t} + \frac{1}{2} \left(1 - \frac{1}{m_Q^2} \right) t \right.$$

$$\left. + \left(1 - \frac{1}{m_Q^2} \right) \frac{t^2}{2 m_Q} + \frac{t^3}{2 m_Q^2} \right] \Psi'(t)$$

$$- \frac{1}{2} \int_0^\infty dt \int_0^\infty ds \frac{\Psi(t) \Psi'(s)}{(s-t)^2}. \quad (B12)$$

For example, the analogue of the nonrelativistic expansion for the heavy hadron mass takes the form

$$M_{H_Q} - m_Q = \langle t \rangle - \frac{1}{2 m_Q} + \frac{3 \langle t^2 \rangle - \langle t \rangle^2}{2 m_Q}$$

$$+ \frac{4 \langle t^3 \rangle - 3 \langle t \rangle \langle t^2 \rangle + \langle t \rangle^3}{2 m_Q^2} - \frac{\langle t \rangle}{2 m_Q^2} + \mathcal{O} \left(\frac{1}{m_Q^3} \right), \quad (B13)$$

with the average $\langle \cdots \rangle$ defined in the standard way as the integral over t with the weight $|\Psi(t)|^2$. Here all averages are calculated with the finite-m_Q $\Psi(t)$. To exclude the Coulomb interaction term we used certain relations which are derived in the way analogous to the virial theorem in quantum mechanics. Namely, for any eigenfunction $\Psi_n(t)$ we can consider the average of $\mathcal{H}$ over the trial function $\sqrt{\lambda} \Psi_n(\lambda t)$, and require a minimum at $\lambda = 1$ (a similar trick was used in Ref. [23] for the case of infinite m_Q). In this way one obtains

$$(m_{sp}^2 - 1) \left\langle \frac{1}{t} \right\rangle - \langle \mathcal{V} \rangle - \left(1 - \frac{1}{m_Q^2} \right) \langle t \rangle - 2 \frac{\langle t^2 \rangle}{m_Q} - 3 \frac{\langle t^3 \rangle}{m_Q^2}$$

$$= \mathcal{O} \left(\frac{1}{m_Q^3} \right) \quad (B14)$$

while

------

[12]This standard for the infinite momentum frame procedure in the context of the 't Hooft model was first considered in Refs. [23, 22].

$$\epsilon_n + \frac{\epsilon_n^2 + 1}{2m_Q} = \frac{1}{2}\left((m_{\mathrm{sp}}^2 - 1)\left\langle\frac{1}{t}\right\rangle - \langle\mathcal{V}\rangle\right) + \frac{1}{2}\left(1 - \frac{1}{m_Q^2}\right)\langle t\rangle$$

$$+ \frac{1}{2}\frac{\langle t^2\rangle}{m_Q} + \frac{1}{2}\frac{\langle t^3\rangle}{m_Q^2} + \mathcal{O}\left(\frac{1}{m_Q^3}\right), \tag{B15}$$

where $\langle\mathcal{V}\rangle$ denotes the expectation value of the integral term in the equation.

The bound-state–independent term $-1/2m_Q$ in Eq. (B13) deserves a special note: it represents the short-distance renormalization of the bare mass we used, originating from momenta $\sim m_Q$. There is no infrared part of the mass: it enters at the scale much larger than Λ_{QCD}. The effect of smaller momenta is described by the nonperturbative 't Hooft wave function rather than infrared-divergent perturbative diagram.

Let us note that the operator $t/2$ is associated with the breaking of scale invariance of the static 't Hooft equation [i.e., $\Psi(\lambda t)$ is the solution of the equation with the linear term $\lambda^2 t/2$]. In other words, the commutation relation holds $[t(d/dt),\mathcal{H}_\infty] = t - \mathcal{H}_\infty$, where $\mathcal{H}_\infty$ is the static 't Hooft

Hamiltonian, Eq. (B12) with $m_Q = \infty$. This fact can be used to obtain the variation of $\langle t\rangle$ when including perturbations, in the analogy to the case of QCD (Ref. [15], Sec. II):

$$\delta\langle t\rangle = -l\langle\delta\mathcal{H}_l\rangle$$

for the perturbation $\delta\mathcal{H}_l$ which is a homogeneous rank-l functional of t. This follows from the usual operator relations such as $\langle iT\{[\mathcal{H},A],B\}\rangle = \langle[A,B]\rangle$. For instance, to order $1/m_Q$ one has $\langle t\rangle = \langle t\rangle_\infty - \langle t^2\rangle/m_Q$, so that in Eq. (B13) $M_{H_Q} - m_Q \simeq \langle t\rangle_\infty + \langle t^2 - \bar{t}^2\rangle/(2m_Q)$, as it should be.

2. Nonrelativistic expansion of the decay amplitude

In the $1/m_Q$ expansion of the $H_Q \to D^{(k)}$ amplitude we pass from $\varphi_{H_Q}(x)$ to the nonrelativistic wave functions $\Psi_{H_Q}(t)$ *and*, in the case of the triple vertices F_{lk}, rewrite $\varphi_l^{B_c}(x)$ via the corresponding $\Psi_l^{B_c}(u)$. The nonrelativistic approximation for c quark is not employed here, however. Then

$$\mathcal{C}_k(\omega) = -\frac{1}{\omega\sqrt{m_Q}}\int_0^{(1-\omega)m_Q} dt\,\Psi_{H_Q}(t)\varphi_k\left(1 - \frac{t}{(1-\omega)m_Q}\right), \tag{B16}$$

$$\mathcal{D}_m(\omega) = -\frac{\omega}{(1-\omega)\sqrt{m_Q}}\int_0^{(1-\omega)m_Q}\frac{dt}{1-t/m_Q}\Psi_{H_Q}(t)\frac{\varphi_k(1-t/(1-\omega)m_Q)}{1-t/(1-\omega)m_Q}, \tag{B17}$$

and, for the triple vertex,

$$F_{lk}(\omega) = \omega(1-\omega)\int_0^{m_Q} du\int_0^1 dy\,\frac{\Psi_l^{B_c}(u)\varphi_k(y)}{\left[(1-\omega)y + \omega u/m_Q\right]^2}\{\Psi_{H_Q}[\omega u + (1-\omega)m_Q] - \Psi_{H_Q}[(1-\omega)(1-y)m_Q]\}. \tag{B18}$$

At arbitrary q^2 we have $\omega \sim 1$; nevertheless, at large enough energy release $(1-\omega)m_Q \gg 1$ still holds. This parameter defines the ''hardness'' and is used in the $1/m_Q$ expansion.

The nonrelativistic expansion amounts to assuming that the support of the Ψ functions is limited to a finite interval of the argument of order 1, and extending the integration over t and u to infinity. With the fiducial domain of integration $t,u \sim 1$, it is readily seen that $\mathcal{C}_k$ and $\mathcal{D}_k$ scale as $m_Q^{-1/2}$ and lead to the properly m_Q-behaved transition amplitudes. On the other hand, $F_{lk} \sim 1/m_Q$ and are accompanied by the factors $c_l \sim m_Q^{-1/2}$. The sum over l is effectively cut off above $l \sim m_Q$ where μ_l exceeds m_Q. Altogether, the terms with F_{lk} yield corrections to the decay amplitudes suppressed by at least $1/m_Q^2$.

The leading heavy-quark transition amplitudes governed by the overlap factors $\mathcal{C}$ and $\mathcal{D}$, Eqs. (B16), (B17), exhibit explicitly the proper functional dependence on the combination of m_Q and q^2: the inner product of wave functions depends only on $(1-\omega)m_Q$ whose value just fixes the energy (or momentum) of the final state hadron in the rest frame,

Eq. (B2). In reality, this property of the leading-m_Q transition amplitudes is more general and is not related to the smallness of the perturbative corrections. It holds at small energy release as well, where the vertex corrections are not power suppressed; this is addressed below at the end of Appendix B.

3. ''Semileptonic'' width $\Gamma_{\mathrm{sl}}(q^2)$ at arbitrary q^2 with $1/m_Q$ accuracy

Here we illustrate how the parton expression for $\Gamma_{\mathrm{sl}}(q^2)$ is reproduced at nonzero q^2. We assume that the width averaged over an interval of m_Q or q^2 is considered, so that the threshold factors become nonsingular. The analysis differs from the case of $q^2 = 0$ anatomized in Ref. [11] only in technical details. In particular, the width is still saturated by the states with $M_k^2 \lesssim (1 - q^2/m_Q^2)m_Q$, and the summation over the final states can be extended to infinity. The decay amplitudes are given by the wave function overlaps to this accuracy. Let us assume for simplicity that $m_c \ll m_Q$. Then only the term $\sim \mathcal{C}_k$ survives:

$$\Gamma(q^2) = \frac{G^2}{4\pi} \sum_k \frac{1}{2M_{H_Q}^2 |\vec{p}_k|} \frac{q^4}{\omega_k^2 m_Q} \int_0^{(1-\omega_k)m_Q} dt \int_0^{(1-\omega_k)m_Q} ds \, \Psi_{H_Q}(t) \Psi_{H_Q}(s) \varphi_k \left(1 - \frac{t}{(1-\omega_k)m_Q}\right) \varphi_k \left(1 - \frac{s}{(1-\omega_k)m_Q}\right).$$
(B19)

The leading term in the expansion of the width emerges if we neglect M_k compared to $(1-\omega)m_Q$. Then $\omega_k = q^2/M_{H_Q}^2$ and $2|\vec{p}_k| = (1-\omega_k)M_{H_Q}$ are k independent, and we get up to the power corrections

$$\Gamma(q^2) = \frac{G^2}{4\pi} \frac{1}{1-\omega} \int_0^\infty dt \, ds \, \Psi_{H_Q}(t) \Psi_{H_Q}(s) \delta\left(\frac{t-s}{(1-\omega)m_Q}\right) = \frac{G^2 m_Q}{4\pi} \left[1 + \mathcal{O}\left(\frac{1}{m_Q}\right)\right].$$
(B20)

Account for the $1/m_Q$ effects requires expanding $|\vec{p}_k|$ and ω_k in $M_k^2/(m_Q^2 - q^2)$, and leads to the ''second'' sum rule for the average of M_k^2 [11]. Here is how it works.

The partial width $\Gamma_k(q^2)$ to $1/m_Q$ accuracy takes the form

$$\Gamma_k(q^2) = \frac{G^2}{4\pi} \frac{M_{H_Q}^2}{m_Q(1-\omega_0)} \left\{ (1-\omega_0)m_Q + \frac{M_k^2}{m_Q^2}\left[-\frac{1}{(1-\omega_0)} \int dt \, ds \, \Psi_{H_Q}(t) \Psi_{H_Q}(s) \varphi_k \left(1 - \frac{t}{(1-\omega_0)m_Q}\right) \varphi_k \left(1 - \frac{s}{(1-\omega_0)m_Q}\right) \right.\right.$$
$$\left.\left. + \frac{2\omega_0}{(1-\omega_0)^2} \int dt \, ds \, t \Psi_{H_Q}'(t) \Psi_{H_Q}(s) \varphi_k \left(1 - \frac{t}{(1-\omega_0)m_Q}\right) \varphi_k \left(1 - \frac{s}{(1-\omega_0)m_Q}\right) \right]\right\},$$
(B21)

where $\omega_0 = q^2/M_{H_Q}^2$.

Using the relation

$$\sum_k M_k^2 \varphi_k(x) \varphi_k(y) = \left(\frac{m_c^2 - 1}{x} + \frac{m_{\rm sp}^2 - 1}{1-x}\right)\delta(x-y) - \frac{1}{(x-y)^2}$$
(B22)

and Eq. (B11) to the leading order in $1/m_Q$, we obtain

$$\sum_k M_k^2 \int ds \, \Psi_{H_Q}(s) \varphi_k \left(1 - \frac{t}{(1-\omega_0)m_Q}\right) \varphi_k \left(1 - \frac{s}{(1-\omega_0)m_Q}\right) = (1-\omega_0)^2 m_Q^2 (2\epsilon - t) \Psi_{H_Q}(t)$$
(B23)

$(\epsilon = M_{H_Q} - m_Q)$. Hence, we arrive at

$$\Gamma(q^2) = \frac{G^2}{4\pi} M_{H_Q} \left\{ 1 + \frac{1}{m_Q}\left[\int dt (t - 2\epsilon) \Psi_{H_Q}^2(t) + \frac{2\omega_0}{1-\omega_0} \int dt \, t(2\epsilon - t) \Psi_{H_Q}'(t) \Psi_{H_Q}(t) \right] + \mathcal{O}\left(\frac{1}{m_Q^2}\right) \right\}.$$
(B24)

To evaluate the last integral we note that for any $f(t)$

$$\int dt \, \Psi_{H_Q}'(t) \Psi_{H_Q}(t) f(t) = \frac{1}{2} \int dt \left[\frac{d}{dt}[\Psi_{H_Q}^2(t) f(t)] - \Psi_{H_Q}^2(t) f'(t)\right] = -\frac{1}{2}\langle f'(t)\rangle.$$
(B25)

Therefore, Eq. (B24) takes the form

$$\Gamma(q^2) = \frac{G^2 M_{H_Q}}{4\pi} \left\{ 1 - \frac{2\epsilon - \langle t\rangle}{m_Q} + \frac{4\omega_0}{1-\omega_0}\frac{\langle t\rangle - \epsilon}{m_Q} + \mathcal{O}\left(\frac{1}{m_Q^2}\right) \right\} = \frac{G^2 m_Q}{4\pi} \left\{ 1 + \mathcal{O}\left(\frac{1}{m_Q^2}\right) \right\},$$
(B26)

where we recalled that $\epsilon - \langle t\rangle = \mathcal{O}(1/m_Q)$.

The next, $1/m_Q^2$ corrections, on the other hand, are not only limited to the term $-\mu_\pi^2/2m_Q^2$, but include also those coming from the perturbative vertex renormalization and from the sensitivity of the decay width $Q \to q + \phi$ with $m_\phi^2 \neq 0$ to the short-distance quark mass renormalization $m^2 \to m^2 - \beta^2$.

4. Decays at maximal q^2

At q^2 close to the energy release, $(m_Q - m_q)^2 - q^2 \lesssim \beta m_Q$ [or at $(m_Q - m_q) \lesssim \beta$] the decay processes are not hard but proceed over the time intervals $\gtrsim 1/\beta$. As such, they in general are sensitive to the t-channel evolution as well. OPE does not allow to compute these widths in the short-distance expansion directly even to the leading order in

$1/m_Q$. However, it relates the overall width associated with this domain of q^2 (i.e., integrated over q^2) to the expectation value of the local four-fermion operator:

$$\Gamma^{\text{end point}} \simeq G^2 \bar{\rho}(m_Q^2)(P_\mu P_\nu - \delta_{\mu\nu} M_{H_Q}^2)$$

$$\times \frac{1}{2M_{H_Q}} \langle H_Q(P)|\bar{Q}\gamma_\mu q\bar{q}\gamma_\nu Q|H_Q(P)\rangle \quad \text{(B27)}$$

(this fact was proven in Ref. [24] although was used to calculate certain preasymptotic corrections to the heavy quark widths since the original papers [9]). Complicated strong interaction dynamics shows up here as a nontrivial expectation value of this operator (for light q; it is perturbatively calculable when q is heavy). In the 't Hooft model it essentially depends on the details of the lowest t-channel states. As has been already discussed a few times, a smeared decay width is assumed here, which is reflected in Eq. (B27): it incorporates the spectral density $\bar{\rho}(m_Q^2)$ averaged over an interval. With a continuous $\rho(q^2)$ there would be no resonance structure in $\Gamma(m_Q)$.

Referring to paper [24] for the formal derivation of Eq. (B27), here we illustrate it schematically in a transparent way. Let us consider a particular meson k in the final state. The corresponding partial decay probability (in the rest frame) is given by

$$\Gamma_k(q^2) = \frac{G^2}{2} \frac{2}{8M_{H_Q}^2|\vec{p}_k(q^2)|} \frac{1}{\pi}\rho(q^2)(q_\mu q_\nu - \delta_{\mu\nu}q^2)$$

$$\times \langle H_Q(P)|\bar{Q}\gamma_\mu q(0)|k(\vec{p}_k)\rangle$$

$$\times \langle k(\vec{p}_k)|\bar{q}\gamma_\nu Q(0)|H_Q(P)\rangle. \quad \text{(B28)}$$

Integrating over some interval of q^2 we can pass to the variable $|\vec{p}_k|$ according to

$$q_0 = \frac{M_{H_Q}^2 + q^2 - M_k^2}{2M_{H_Q}}, \quad dq_0 = -dE_k \quad E_k dE_k = |\vec{p}_k|d|\vec{p}_k|$$

and have

$$\int dq^2 \Gamma_k(q^2) = G^2 \int \frac{2d|\vec{p}_k|}{2\pi \, 2E_k} \rho[q^2(|\vec{p}_k|)][(P-p_k)_\mu$$

$$\times (P-p_k)_\nu - \delta_{\mu\nu}(P-p_k)^2]$$

$$\times \frac{1}{2M_{H_Q}}\langle H_Q(P)|\bar{Q}\gamma_\mu q(0)|k(\vec{p}_k)\rangle$$

$$\times \langle k(\vec{p}_k)|\bar{q}\gamma_\nu Q(0)|H_Q(P)\rangle. \quad \text{(B29)}$$

The factor of 2 above corresponds to two possible directions of $\vec{p}_k$ and is the unit "sphere" surface area in one space dimension. If the smeared width is considered, the corresponding smeared spectral density $\bar{\rho}(q^2)$ replaces $\rho(q^2)$ in the above equation, since both the decay amplitudes and phase space depend on the combination $m_Q - \sqrt{q^2}$ (for simplicity we imply that $q^2 \gg m_q^2$).

With the smooth $\bar{\rho}(q^2)$, we can neglect p_k compared to M_{H_Q} in Eq. (B29) and have, up to $1/m_Q$ corrections,

$$\int dq^2 \Gamma_k(q^2) = \frac{G^2}{2M_{H_Q}} m_Q^2 \bar{\rho}(m_Q^2)(v_\mu v_\nu - \delta_{\mu\nu})$$

$$\times \int \frac{d\vec{p}_k}{2\pi \, 2E_k} \langle H_Q(P)|\bar{Q}\gamma_\mu q|k_{\vec{p}}\rangle$$

$$\times \langle k_{\vec{p}}|\bar{q}\gamma_\nu Q|H_Q(P)\rangle. \quad \text{(B30)}$$

The corrections to this expression appear due to the dependence of $\bar{\rho}(q^2)$ on q^2 near $q^2 \simeq m_Q^2$ and are power-suppressed.

If we formally summed over the final states k, the RHS of the above equation would become

$$\frac{G^2}{2M_{H_Q}} m_Q^2 \bar{\rho}(m_Q^2)(v_\mu v_\nu - \delta_{\mu\nu}) \sum_k \int \frac{d\vec{p}_k}{2\pi \, 2E_k}$$

$$\times \langle H_Q(P)|\bar{Q}\gamma_\mu q(0)|k_{\vec{p}}\rangle \langle k_{\vec{p}}|\bar{q}\gamma_\nu Q(0)|H_Q(P)\rangle$$

$$= \frac{G^2}{2} m_Q^2 \bar{\rho}(m_Q^2)(v_\mu v_\nu - \delta_{\mu\nu}) \frac{1}{2M_{H_Q}}$$

$$\times \langle H_Q(P)|(\bar{Q}\gamma_\mu q)(\bar{q}\gamma_\nu Q)(0)|H_Q(P)\rangle, \quad \text{(B31)}$$

which shows the representation of the matrix element as a sum over intermediate states. The integration over q^2 replaces $dq^2/(2\pi M_{H_Q}|\vec{p}_k|)$ in the partial decay width by $d\vec{p}_k/(2\pi 2E_k)$ which is just the quantum mechanical summation over the intermediate states.

Of course, this correspondence is not accidental for $D = 2$. Indeed, the general two-body phase space is given by

$$\Phi_2(M_{H_Q}^2, q^2, M_k^2)$$

$$= \int \frac{d^D l}{(2\pi)^D} \int \frac{d\Omega_{\vec{p}} d|\vec{p}|^{D-1}dp_0}{(2\pi)^D}(2\pi)^2$$

$$\times \delta_+(l^2 - q^2)\delta_+(p^2 - M_k^2)(2\pi)^D \delta^D(P - p - l). \quad \text{(B32)}$$

To obtain the decay width we multiply it by the transition amplitudes $\langle H_Q|\bar{Q}\gamma_\mu q|k\rangle$, $\langle k|\bar{q}\gamma_\nu Q|H_Q\rangle$, by $(1/\pi)\text{Im}\,\Pi_{\mu\nu}(q)$ and by $(2\pi)^D/(2M_{H_Q})$. Integrating over q^2, we get

$$\int dq^2 \Gamma_k(q^2) = \frac{G^2}{2} \frac{1}{2M_{H_Q}} \int dq^2 \Phi_2(M_{H_Q}^2, q^2, M_k^2)$$

$$\times \langle H_Q|\bar{Q}\gamma_\mu q|k_{\vec{p}}\rangle \langle k_{\vec{p}}|\bar{q}\gamma_\nu Q|H_Q\rangle$$

$$= G^2 \int \frac{d^{D-1}\vec{p}_k}{(2\pi)^{D-1}2E_k} \frac{1}{\pi}\text{Im}\,\Pi_{\mu\nu}(P-p_k)$$

$$\times \frac{1}{2M_{H_Q}}\langle H_Q|\bar{Q}\gamma_\mu q|k_{\vec{p}}\rangle \langle k_{\vec{p}}|\bar{q}\gamma_\nu Q|H_Q\rangle. \quad \text{(B33)}$$
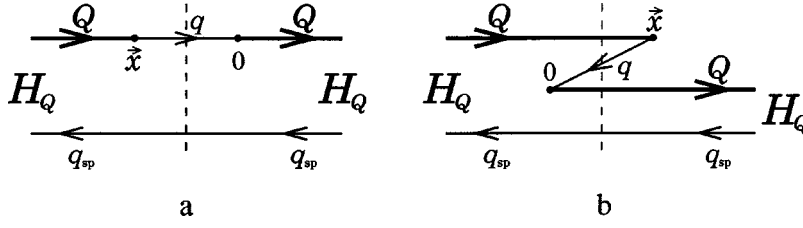
FIG. 8. ''s channel'' with $n_Q=0$ (a) and ''u channel'' with $n_Q=2$ (b) intermediate states saturating the four-quark expectation values in full QCD. Both types are described by ''soft'' transition amplitudes and contribute simultaneously to the forward $\mathcal{O}(G_F^2)$ transition amplitude near maximal q^2. Only the targeted ''s-channel'' states appear in the nonrelativistic effective theory. The two points with weak currents are shown separated in space ($\vec{x}\neq 0$) for clarity.

With $\text{Im}\,\Pi_{\mu\nu}(P-p_k)$ approximated by its (smeared) value at $q=m_Q v$, this is the contribution of the given state k with mass M_k (whether meson, baryon, or a multiparticle state) to the expectation value of the local four-quark operator $\langle H_Q(P)|(\bar{Q}\gamma_\mu q)(\bar{q}\gamma_\nu Q)|H_Q(P)\rangle$ [24].

Of course, in reality we do not want to extend the summation to *all* states with arbitrary large energies, which would correspond to integrating over q^2 down to $q^2=-\infty$. The large-q^2 domain is cut at some $q^2\lesssim(m_Q-\Delta)^2\simeq m_Q^2 -2m_Q\Delta$, with $\Delta\gtrsim\Lambda_{\text{QCD}}$. This means that the heavy quark expectation value in Eq. (B27) corresponds to the normalization point $\mu=\Delta$. We note that the normalization point enters as the cutoff over the *energy* of the intermediate states, the common case for the heavy flavor systems (see Refs. [4,29], and references therein).

It is worth reiterating an important point regarding the end-point domain: the matrix elements in Eqs. (B27), (B31), (B33) are those in the effective theory rather than in full QCD—even if there were no high-momentum gluons in the latter. The representation of the four-fermion expectation values

$$\langle H_Q(P)|(\bar{Q}\gamma_\mu q)(\bar{q}\gamma_\nu Q)|H_Q(P)\rangle$$

$$=\sum_k\int\frac{d^{D-1}\vec{p}_k}{(2\pi)^{D-1}2E_k}\langle H_Q|\bar{Q}\gamma_\mu q|k_{\vec{p}}\rangle\langle k_{\vec{p}}|\bar{q}\gamma_\nu Q|H_Q\rangle$$

$$(\text{B34})$$

in the full theory includes not only the intermediate states without heavy quark Q, Fig. 8(a), but also those containing the pair of Q and $\bar{Q}$, Fig. 8(b). The latter are not related to the (perturbative) heavy quark loops and appear even in the free theory—for example, the state $H_Q(P)+(\bar{Q}+q)$ with $\bar{Q}$ and q having small spacelike momenta $\lesssim\Delta$ totaling to $\vec{p}_k$. It can be simply $H_Q(P)+\tilde{H}_Q(\vec{p})$. These states appear since the Lorentz-covariant currents $\bar{Q}\Gamma q$, together with operators creating Q contain also operators annihilating antiquark $\bar{Q}$, which is not the case in the nonrelativistic field theory. Such amplitudes are shaped by the time scales $\sim 1/\Delta$ and are as nonperturbative as transitions into the usual light hadrons. Imposing the cut on the virtuality itself does not help to eliminate such spurious states with ''valence'' Q. In particular, integrating out the high-momentum modes of quark and gluon fields, or imposing a cut on the gluon momentum (vir-

tuality) in the full theory—while generating an effective low-energy theory—would not yield relevant operators, at least in this context.

The extra contributions of Fig. 8(b) are counterpart of the intermediate states in the transition operator which are nearly on shell for the reverse ordered product of currents $(\bar{q}\gamma_\nu Q)(\bar{Q}\gamma_\mu q)$ in their time-ordered product

$$\hat{T}_{\mu\nu}(q)=\int d^D x\, e^{-iqx} iT\{\bar{q}\gamma_\nu Q(0)\bar{Q}\gamma_\mu q(x)\}\quad(\text{B35})$$

at $x_0<0$. They are responsible for the u-channel cut of the Lorentz-covariant forward transition amplitude $T_{\mu\nu}$ in Eq. (11). In the kinematics $q^2\to m_Q^2$, $q_0\to m_Q$ the transitions to both s channel states with $n_Q=n_{\bar{Q}}=0$ and u channel states with $n_Q=n_{\bar{Q}}=1$ are equally ''soft'' and nonperturbative, and cannot be disentangled in the single amplitude $T_{\mu\nu}(q)$.

On the contrary, only the necessary light intermediate states are present when the proper nonrelativistic effective theory of heavy quarks is considered, where $Q(x)$ do not include antiquark operators, either creation or annihilation. The expectation values determining the corrections to the widths must be understood only in this sense.

We conclude this appendix by noting that the leading-m_Q transition amplitudes $H_Q\to k$ obey the stated scaling in respect to m_Q and q^2, i.e., for the particular final state k depend on one combination $(m_Q^2+M_k^2-q^2)/m_Q$ having the meaning of the rest-frame energy E_k:

$$\frac{1}{\sqrt{2m_Q}}\langle k|\bar{q}\Gamma Q|H_Q\rangle_{q^2,m_Q}\simeq\frac{1}{\sqrt{2m_{Q'}}}\langle k|\bar{q}\Gamma Q'|H_{Q'}\rangle_{q'^2,m_Q'}$$

$$(\text{B36})$$

if

$$\frac{m_Q^2+M_k^2-q^2}{m_Q}=\frac{m_{Q'}^2+M_k^2-q'^2}{m_{Q'}}.$$

This functional dependence ensures, for example, that the four-quark expectation values discussed above have a finite m_Q-independent limit as $m_Q\to\infty$.

In terms of the light-cone parameters Eq. (B36) says that the form factors, up to certain powers of $\sqrt{m_Q}$ must be functions of $(1-\omega)m_Q$, see relation (B2). The explicit expressions (B3)–(B7) do exhibit this property, although in slightly different ways in different kinematic domains. To locate it,

we once again can be guided by the duality between the various pieces in the hadronic expressions based on the 't Hooft equation, and the contributions coming from the corresponding Feynman diagrams. This duality exists in the physical gauge and is transparent when the light-cone formalism is used. Additionally, the scaling behavior of various perturbative effects can be anticipated beforehand if general OPE facts are considered. For example, the high infrared stability of the inclusive widths, together with the fact that no gluon degrees of freedom (and, therefore, no bremsstrahlung) exist in $D=2$, leads to the finiteness of the vertex corrections and allows us to estimate their magnitude.

It has been mentioned already that Eq. (B36) holds at large energy $E_k \gg \beta$: then the amplitudes are given only by the simple overlaps $\mathcal{C}_k$ and $\mathcal{D}_k$. The integrals in Eqs. (B16) and (B17) manifestly depend only on $(1-\omega)m_Q$ if $\Psi_{H_Q}(t)$ are nonzero only for $t \lesssim \beta$.

A different situation takes place at maximal q^2, for example, for light final state quarks. The triple-overlap terms proportional to F_{lk} are then of the same order. However, in this kinematics F_{lk} show the same dependence on $(1-\omega)m_Q$. Indeed, in this case $\omega \to 1$, and they take the form

$$F_{lk}(\omega) \simeq -m_Q^2(1-\omega) \int_0^\infty du \int_0^1 dy \frac{\Psi_l^{B_c}(u)\varphi_k(y)}{[m_Q(1-\omega)y+u]^2}$$

$$\times \Psi_{H_Q}[m_Q(1-\omega)(1-y)]. \quad \text{(B37)}$$

The position of the excited states in the t channel is also driven by the nonrelativistic expression $\mu_l \simeq m_Q + \tilde{\epsilon}_l$ with m_Q-independent $\tilde{\epsilon}_l$. Therefore, the pole factors in front of F_{lk} are simply proportional to $1/m_Q$:

$$\frac{q^2, \mu_l^2}{q^2-\mu_l^2} \simeq \frac{m_Q}{-2(\Delta+\tilde{\epsilon}_l)} \quad \text{for } q^2=(m_Q-\Delta)^2 \quad \text{with}$$

$$\Delta \ll m_Q. \quad \text{(B38)}$$

On the other hand, Δ differs from $(1-\omega)m_Q$ by just a (M_k-dependent) constant,

$$E_k \simeq M_{H_Q} - m_Q + \Delta = \epsilon_0 + \Delta,$$

$$E_k + \sqrt{E_k^2 - M_k^2} = (1-\omega)m_Q.$$

Thus, together with the factors $c_l \propto m_Q^{-1/2}$ this term possesses the required dependence on $(1-\omega)m_Q$ and overall power scaling in m_Q.

This reasoning may seem inapplicable when both Q and q are heavy quarks, in particular, at small $m_Q - m_q$. The t-channel $B_c^{(l)}$ wave functions are then nonrelativistic in respect to both quarks, and must be treated accordingly. This, however, is a transition between two heavy quarks, and even the classic case of the small velocity [7]. As mentioned above, in $D=2$ the amplitudes are finite even at $\vec{v} \neq 0$ and the perturbative corrections are inversely proportional to the heavy quark mass (but not to powers of $1/q^2$ which blows up at the safe zero recoil point). And indeed, it is easy to find

that F_{lk} in Eq. (B18) are universally suppressed by the mass scale $E_k \sim (1-\omega)m_Q$ whether m_q is small or large. Therefore, the functional relation Eq. (B36) holds in the 't Hooft model for any decay kinematics.

APPENDIX C: THE IW FUNCTION AND THE HEAVY QUARK DISTRIBUTION FUNCTIONS

The IW function ξ is most simply defined as the flavor-diagonal vector form factor

$$\langle H_Q(p')|\bar{Q}\gamma_\mu Q|H_Q(p)\rangle = \xi\left(\frac{(pp')}{M_{H_Q}^2}\right)(p+p')_\mu$$

in the large-m_Q limit. The expression for it quoted below can be obtained using the $1/m_Q$ expansion of the amplitudes in the previous section, if we employ the nonrelativistic expansion of the final state as well. Alternatively, we can use directly the simple universal expressions Eq. (56) at $q^2=0$ relying on the scaling relations (76)–(78) by adjusting the final-state quark mass:

$$M_{H_{Q'}} = z M_{H_Q} \quad \text{with} \quad w \equiv (vv') = \frac{1+z^2}{2z},$$

$$z = w \pm \sqrt{w^2-1}. \quad \text{(C1)}$$

The perturbative corrections are finite and vanish at $m_Q \to \infty$ in two dimensions. In this way we obtain

$$\xi(w) = \frac{2}{1+w+\sqrt{w^2-1}} \int_0^\infty dt\, \Psi_{H_Q}(t)\Psi_{H_Q}([w-\sqrt{w^2-1}]t)$$

$$= \frac{2\sqrt{z}}{1+z} \int_0^\infty dt\, \Psi_{H_Q}\left(\frac{t}{\sqrt{z}}\right)\Psi_{H_Q}(\sqrt{z}t). \quad \text{(C2)}$$

The last expression was used to make representation more symmetric. Here $\Psi_{H_Q}(t)$ is the solution of the static 't Hooft equation

$$\epsilon\Psi(t) = \left(\frac{m_{\text{sp}}^2-\beta^2}{2t} + \frac{t}{2}\right)\Psi(t) - \frac{\beta^2}{2}\int_0^\infty ds\, \Psi(s)\frac{1}{(t-s)^2}. \quad \text{(C3)}$$

Expression (C2) is easily generalized for the case of inelastic transitions:

$$\xi_{\text{inel}}^{kn}(w) = \frac{2\sqrt{z}}{1+z}\int_0^\infty dt\, \Psi_{H_Q}^{(k)}\left(\frac{t}{\sqrt{z}}\right)\Psi_{H_Q}^{(n)}(\sqrt{z}t). \quad \text{(C4)}$$

This agrees with the expression obtained in Ref. [22]. We found a direct proof (it will be presented elsewhere) that the right-hand side of Eq. (C4) does not change (up to a sign) if $z \to 1/z$. This ensures the same result for both solutions for z in terms of w, the last of Eqs. (C1).

It is interesting to note that the absolute maximum for the (elastic) IW function which would correspond to the structureless pointlike heavy flavor hadron, is given by just the factor[13]

$$\xi_{\max}(w) = \sqrt{\frac{2}{1+w}} = \frac{2\sqrt{z}}{1+z}. \tag{C5}$$

It would be saturated by the light cone wave function $\Psi(t) \sim c/\sqrt{t}$. Indeed, one has

$$\xi(z) = \left(\frac{2\sqrt{z}}{1+z}\right)^{2+2\alpha} \quad \text{for} \quad \Psi(t) = t^{\alpha} e^{-\mu t}. \tag{C6}$$

Of course, such a situation is not realized for any actual hadronic state in the 't Hooft model. Constructing such a hard-core wave function requires a coherent superposition of all the excited eigenstates.

Studying the decay distribution for the weak transitions into light quarks, $m_c^2 \ll m_Q \Lambda_{\mathrm{QCD}}$ allows one to determine the light-cone distribution function. In particular, it emerges directly in the decays at $q^2 = 0$ as the differential decay probability vs q_0 or M_{hadr}^2 [33]:

$$\frac{1}{\Gamma_{\mathrm{sl}}(0)} \frac{d\Gamma_{\mathrm{sl}}(0)^{m_Q \to \infty}}{dM_{\mathrm{hadr}}^2} = m_Q(M_{H_Q} - m_Q)$$

$$\times F\left(1 - \frac{M_{\mathrm{hadr}}^2}{m_Q(M_{H_Q} - m_Q)}\right),$$

$$q_0 = \frac{M_{H_Q}^2 - M_{\mathrm{hadr}}^2}{2M_{H_Q}} \quad \text{at} \quad q^2 = 0, \tag{C7}$$

where $\Gamma_{\mathrm{sl}}(0)$ as a function of M_{hadr}^2 is

$$\sum_{M_k^2 < M_{\mathrm{hadr}}^2} \Gamma_{\mathrm{sl}}^{(k)}(0).$$

Using the leading term in the $1/m_Q$ expansion of the decay amplitudes, one can check that $F(t)$ coincides with $\Psi_{H_Q}^2(t)$. To obtain the distribution over M_k^2, it is again convenient to use the explicit cutoff dependence of the ultraviolet-regularized Green function $G(x,y;\Lambda)$, as it was done in Sec. III. Earlier discussion for the 't Hooft model can be found in Refs. [20,23].

---

[13]This function is not analytic and rather has a branch point near $w = -1$ (or $q^2 = 4m_Q^2$). This is just a reflection of the Fermi statistics of heavy quarks confined in the bosonic state. This means that it is impossible to construct a heavy pointlike meson from fermionic constituents, with the radius much smaller than the mass. For light constituents this is possible as exemplified by the chiral pion.

The small-x behavior of the heavy-quark distribution function for decays into light quarks has, in general, a noninteger power depending on the spectator mass

$$F(x) \propto x^{2\gamma_{\mathrm{sp}}}, \quad \frac{\pi \gamma_{\mathrm{sp}}}{\tan \pi \gamma_{\mathrm{sp}}} = -\frac{m_{\mathrm{sp}}^2 - \beta^2}{\beta^2}. \tag{C8}$$

Just such a behavior was conjectured for the QCD light-cone distribution function in the model suggested in Ref. [34]. In $D = 4$, however, the distribution function itself and the exponent 2γ, in particular depend on the renormalization point.

APPENDIX D: PERTURBATIVE CORRECTIONS AND IR REGULARIZATION

The perturbative corrections to the weak decay vertex explicitly depend not only on the gauge but also on the employed infrared regularization, even at $q^2 = 0$ (see, e.g., Ref. [11], Sec. III B). This subtlety is aggravated by severe infrared divergences in the individual diagrams in $D = 2$, which appear when unphysical degrees of freedom are introduced (say, in the covariant gauges). While OPE—applicable for inclusive widths—ensures that the widths are rather insensitive to the infrared contributions, in practical computations it requires carrying out one and the same regularization procedure consistently through all stages of computations. A particular infrared regularization, on the one hand, is built into the 't Hooft equation in the form of the principal value prescription of the Coulomb exchange integral and the self-energy terms. On the other hand, the same infrared regularization was employed in establishing the short-distance nonrenormalization theorem for the weak vertex.

Therefore, it is instructive to obtain the same perturbative correction $\beta^2/(2m_Q^2)$ to the decay width starting with the usual Feynman graphs in the covariant gauge routinely used in four dimensions. To this end we independently calculated the one-loop perturbative corrections to the quark decay width $Q \to q + \phi$ at $m_\phi^2 = 0$, with the decay interaction

$$\mathcal{L}_{\mathrm{weak}} = -\frac{G}{2\pi} \bar{q} Q \phi + \mathrm{H.c.} \tag{D1}$$

In $D = 2$ this is a full analogue of the four-fermion (semileptonic) decay with massless leptons for vectorlike weak currents.

Since all the individual diagrams are too infrared divergent, the computations are simple only in dimensional regularization. All corrections are nontrivial, including virtual vertex corrections, mass, and wave function renormalization and the "real" gluon emission width. The sum of all contributions takes the following form (at $m_q = 0$) in arbitrary dimension:

$$\frac{\delta\Gamma^{\text{1-loop}}}{\Gamma^{\text{tree}}} = C_F \frac{g_s^2}{(4\pi)^{D/2}} m_Q^{D-4} \left\{ \Gamma\left(2 - \frac{D}{2}\right)\left[4\frac{D-3}{D-4} + \frac{(D-1)(D-4)}{D-3}\right] + 3\frac{\Gamma^2(D/2)}{\Gamma(\frac{3}{2}D-2)} \frac{3D^4 - 28D^3 + 109D^2 - 208D + 160}{(D-1)(D-3)(D-4)^2} \right\}.$$

$$(D2)$$

It is easy to see that the poles at $D=2$ and $D=3$ all cancel out (at $D=3$ heavy quark masses acquire infrared logarithmic divergence as well). The remaining pole at $D=4$ is given by the ultraviolet renormalization of the Yukawa coupling. At $D=2$ the above expression reduces to

$$\delta\Gamma^{\text{1-loop}} = C_F \frac{g_s^2}{4\pi m_Q^2} \Gamma^{\text{tree}}, \qquad (D3)$$

which coincides exactly with the result obtained in the light-front formalism and by the summation of the exclusive hadronic widths in the 't Hooft model.

We note that here one has an instructive example: the perturbative calculation of the decay width is IR safe by itself, and the first-order correction is meant to describe the actual decay width with the $1/m_Q^2$ accuracy. Higher orders of the perturbative expansion yield $1/m_Q^4$ and smaller terms, and the uncertainties in summing the perturbative series are exponentially small in m_Q. However, this is not the complete answer already to the order $1/m_Q^2$. This illustrates incompleteness of the purely perturbative expansion, even in that the actual nonperturbative effects are not signaled here by generic intrinsic divergences of the perturbative series.

[1] K. Wilson, Phys. Rev. **179**, 1499 (1969); in *Proceedings of the 1971 International Symposium on Electron and Photon Interactions at High Energies*, edited by N. Mistry (Laboratory of Nuclear Studies, Cornell University, Ithaca, NY, 1972), p. 115; K. Wilson and J. Kogut, Phys. Rep. **12**, 75 (1974); Wilson's ideas were adapted to QCD in Ref. [2].

[2] M. A. Shifman, A. I. Vainshtein, and V. I. Zakharov, Nucl. Phys. **B147**, 385 (1979); **B147**, 448 (1979); V. Novikov, M. Shifman, A. Vainshtein, and V. Zakharov, *ibid.* **B249**, 445 (1985).

[3] I. Bigi, B. Blok, M. Shifman, N. Uraltsev, and A. Vainshtein, in *B Decays*, edited by S. Stone, 2nd ed. (World Scientific, Singapore, 1994), p. 132.

[4] I. Bigi, M. Shifman, and N. G. Uraltsev, Annu. Rev. Nucl. Part. Sci. **47**, 591 (1997).

[5] I. Bigi, N. Uraltsev, and A. Vainshtein, Phys. Lett. B **293**, 430 (1992).

[6] B. Blok and M. Shifman, Nucl. Phys. **B399**, 441 (1993); **B399**, 459 (1993).

[7] This fact was first stated in M. Shifman, Report No. ITEP-86-1134, 1986; M. Voloshin and M. Shifman, Yad. Fiz. **47**, 801 (1988) [Sov. J. Nucl. Phys. **47**, 511 (1988)].

[8] I. Bigi, M. Shifman, N. Uraltsev, and A. Vainshtein, Phys. Rev. D **50**, 2234 (1994).

[9] M. Voloshin and M. Shifman, Yad. Fiz. **41**, 187 (1985) [Sov. J. Nucl. Phys. **41**, 120 (1985)]; Zh. Eksp. Teor. Fiz. **91**, 1180 (1986) [Sov. Phys. JETP **64**, 698 (1986)]; Yad. Fiz. **45**, 463 (1987) [Sov. J. Nucl. Phys. **45**, 292 (1987)]; V. A. Khoze, M. A. Shifman, N. G. Uraltsev, and M. B. Voloshin, Yad. Fiz. **47**, 1687 (1988) [Sov. J. Nucl. Phys. **46**, 112 (1987)].

[10] I. Bigi and N. Uraltsev, Phys. Lett. B **280**, 271 (1992).

[11] I. Bigi, M. Shifman, N. Uraltsev, and A. Vainshtein, Phys. Rev. D **59**, 054011 (1999).

[12] B. Chibisov, R. Dikeman, M. Shifman, and N. Uraltsev, Int. J. Mod. Phys. A **12**, 2075 (1997).

[13] M. Shifman, in *Proceedings of the Workshop on Continuous Advances in QCD*, edited by A. Smilga (World Scientific, Singapore, 1994), p. 249.

[14] References to some earlier studies can be found, e.g., in Refs. [11,25]. Violation of local duality for light flavor hadrons was first addressed in A. Zhitnitsky, Phys. Rev. D **53**, 5821 (1996).

[15] I. Bigi, M. Shifman, N. Uraltsev, and A. Vainshtein, Phys. Rev. D **52**, 196 (1995).

[16] M. Shifman, Nucl. Phys. **B388**, 346 (1992); see also Ref. [6].

[17] B. Grinstein and R. Lebed, Phys. Rev. D **59**, 054022 (1999).

[18] G. 't Hooft, Nucl. Phys. **B75**, 461 (1974).

[19] C. Callan, N. Coote, and D. Gross, Phys. Rev. D **13**, 1649 (1976).

[20] M. Einhorn, Phys. Rev. D **14**, 3451 (1976).

[21] M. Bishary, Ann. Phys. (N.Y.) **129**, 435 (1980); W. Krauss and M. Staudacher, Phys. Lett. B **388**, 808 (1997); M. Einhorn, S. Nussinov, and E. Rabinovici, Phys. Rev. D **15**, 2282 (1977); R. L. Jaffe and P. F. Mende, Nucl. Phys. **B369**, 189 (1992).

[22] M. Burkardt and E. Swanson, Phys. Rev. D **46**, 5083 (1992).

[23] M. Burkardt, Phys. Rev. D **46**, R1924 (1992); **46**, R2751 (1992).

[24] I. Bigi and N. Uraltsev, Nucl. Phys. **B423**, 33 (1994).

[25] B. Blok, M. Shifman, and Da-Xin Zhang, Phys. Rev. D **57**, 2691 (1998).

[26] This observation was first made in a phenomenological context by A. Buras, J.-M. Gerard, and R. Rückl, Nucl. Phys. **B268**, 16 (1986).

[27] B. Grinstein and R. Lebed, Phys. Rev. D **57**, 1366 (1998).

[28] I. Bigi, M. Shifman, N. Uraltsev, and A. Vainshtein, Phys. Rev. D **56**, 4017 (1997).

[29] N. Uraltsev, in *Heavy Flavour Physics: A Probe of Nature's Grand Design*, Proceedings of the International School of Physics ''Enrico Fermi,'' Course CXXXVII, Varenna, 1997, edited by I. Bigi and L. Moroni (IOS Press, Amsterdam, 1998), p. 329, hep-ph/9804275.

[30] A. Le Yaouanc, L. Oliver, O. Pène, and J.-C. Raynal, Phys. Rev. D **52**, 2813 (1995); hep-ph/9504267.

[31] W. Krauss and M. Staudacher, Phys. Lett. B **388**, 808 (1997).

[32] M. Burkardt, Ph.D. thesis, University Erlanger, Germany, 1989 (in German). This contribution contains many useful relations for the 't Hooft model. See also Nucl. Phys. **A504**, 762 (1989).

[33] I. I. Bigi, M. Shifman, N. G. Uraltsev, and A. Vainshtein, Int. J. Mod. Phys. A **9**, 2467 (1994).

[34] D. Dikeman, M. Shifman, and N. G. Uraltsev, Int. J. Mod. Phys. A **11**, 571 (1996).