Heavy flavor decays, OPE, and duality in the two-dimensional 't Hooft model

I. Bigi,[1] M. Shifman,[2] N. Uraltsev,[1,2,3] and A. Vainshtein[2,4]
[1]*Department of Physics, University of Notre Dame du Lac, Notre Dame, Indiana 46556*
[2]*Theoretical Physics Institute, University of Minnesota, Minneapolis, Minnesota 55455*
[3]*Petersburg Nuclear Physics Institute, Gatchina, St. Petersburg 188350, Russia**
[4]*Budker Institute of Nuclear Physics, Novosibirsk 630090, Russia*
(Received 7 May 1998; published 1 February 1999)

The 't Hooft model (two-dimensional QCD in the limit of a large number of colors) is used as a laboratory for exploring various aspects of heavy quark expansion in the nonleptonic and semileptonic decays of heavy flavors. We perform a complete operator analysis and construct the operator product expansion (OPE) up to terms $\mathcal{O}(1/m_Q^4)$, inclusively. The OPE-based predictions for the inclusive widths are then confronted with the ''phenomenological'' results, obtained by summation of all open exclusive decay channels, one by one. The summation is carried out analytically, by virtue of the 't Hooft equation. The two alternative expressions for the total widths match. We comment on the recent claim in the literature of a $1/m_Q$ correction to the total width which would be in clear conflict with the OPE result. The issue of duality violations both in the simplified setting of the 't Hooft model and in actual QCD is discussed. The amplitude of oscillating terms is estimated. [S0556-2821(99)05503-4]

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

I. OVERVIEW

The development of heavy quark theory started in the 1980s has essentially been completed. While at the early stages the main emphasis was placed on the symmetry aspects (the so-called heavy quark symmetry), the present (mature) stage deals with dynamical aspects. A formalism based on Wilson's operator product expansion (OPE) [1] has been developed and applied to many cases of practical interest, in particular to inclusive decays of heavy flavor hadrons. The theory of such decays is at a rather advanced stage now (see [2] and references therein). Calculations we could not even dream of several years ago have become possible.

The decays of heavy flavor hadrons H_Q are shaped by nonperturbative dynamics. While QCD at large distances is not yet solved, considerable progress has been achieved in this problem. The width of an inclusive transition $H_Q \to f$ is expressed through an OPE. The nonperturbative effects are then parametrized through expectation values of various local operators $\mathcal{O}_i$ built from the quark and/or gluon fields. Observable quantities, such as semileptonic and nonleptonic widths of heavy hadrons H_Q, are then given by

$$\Gamma_{H_Q}=\frac{1}{M_{H_Q}}\sum_i \ \mathrm{Im}\, c_i(\mu)\langle H_Q|\mathcal{O}_i(\mu)|H_Q\rangle, \qquad (1)$$

where c_i are the OPE coefficients, and μ stands for a normalization point separating out soft contributions [which are lumped into the matrix elements $\langle H_Q|\mathcal{O}_i(\mu)|H_Q\rangle$] from the hard ones (which belong to the coefficient functions c_i).

There are many subtle and interrelated issues, both conceptual and technical, associated with the operator product expansion in QCD.

(1) Equation (1) represents an expansion in powers of $1/m_Q$ with m_Q being the Q quark mass and the coefficients c_i scaling like $(1/m_Q)^{d_i-2}$ for an operator with dimension d_i. In Γ_{H_Q} there are two sources for contributions depending on powers of $1/m_Q$, namely higher-dimensional *operators* and higher-order terms in the expansion of their *expectation values*. In addition every coefficient c_i is a series in the running coupling $\alpha_s(m_Q)\propto 1/\log m_Q$, of which only a few terms are known for a given coefficient c_i. This immediately raises a grave concern: how can we retain terms suppressed by powers of $1/m_Q$ without a complete summation of the parametrically larger powers of $1/\log m_Q$ in the leading coefficient?

(2) Although the normalization point μ conceptually represents a straightforward ''book-keeping'' device for separating hard and soft contributions, it is technically difficult to actually carry out such a program since no user-friendly definition of what is soft and hard exists in QCD. So far, the vast majority of all discussions related to the introduction of μ are conducted in a hand-waving manner.

(3) It is quite conceivable that there are *hard non*perturbative contributions in the coefficient functions: $c_i^{nonpert} \sim (\Lambda_{\rm QCD}/m_Q)^\delta$ with δ being some positive number. The possible size of such contributions is essentially unknown. A related problem is the convergence (or divergence) of the perturbative series for the coefficient functions c_i.

(4) Truncating the series (1) at some finite order introduces an error estimated by so-called exponential terms, which in Euclidean domain appear as expressions of the type $\exp[-(m_Q/\Lambda_{\rm QCD})^k]$. In order to obtain Γ_{H_Q} we analytically continue from the Euclidean domain, where the OPE is well defined and the coefficients c_i are real, to the Minkowski domain where they acquire an imaginary part. Such analytic continuation is implicit in Eq. (1) and is based on the assumption of smoothness. Under analytic continuation the exponential terms convert themselves into *oscillating* terms of the type $\cos[(m_Q/\Lambda_{\rm QCD})^k]$ [3]; the expansion (1) does not

*Permanent address.

account for them. It can thus be understood on general grounds that duality violation is described—or at least modeled—by oscillating expressions. To which degree those are suppressed by powers of $1/m_Q$ depends on details of the strong interactions and the specifics of the process.

All these questions are circumvented in the so-called *practical version* of OPE [4] routinely used so far in all instances when there is need in numerical predictions. This version is admittedly approximate, however. The questions formulated above are legitimate; they deserve to attract theorists' attention, and continue to cause confusion in the literature. They have to be addressed also because they are emerging as a major source of the uncertainties in quantitative predictions; these problems have specifically been suspected to underlie phenomenological difficulties encountered recently, e.g., a relatively short lifetime of b-flavored baryons and a relatively small semileptonic branching ratio of b-flavored mesons.

We find it useful and instructive to study all these issues in models that while retaining basic features of QCD—most notably quark confinement—are simpler without being trivial and can be solved dynamically. QCD defined in one time and one space dimension—hereafter referred to as $1+1$ QCD—is especially suitable for this purpose: with the Coulomb potential necessarily growing linearly in two dimensions, quark confinement is built in. Likewise the theory is superrenormalizable, i.e., very simple in the ultraviolet domain. There are no logarithmically divergent ''tails'' in the Feynman graphs. As a result, the book-keeping of OPE (separation of the hard and soft parts) becomes simple, and all subtle aspects in the construction of the OPE can be studied in a transparent environment.

In particular, the perturbative contributions in the coefficients c_i become an expansion in g^2/m_Q^2 (where g is the gauge coupling in $1+1$ QCD). They are thus power-suppressed in the same way as the higher-dimensional operators; the first problem formulated above therefore does not arise here. Without the logarithmic UV tails the second problem becomes tractable. Concerning the third problem it is easy to see that in $1+1$ QCD nonperturbative corrections cannot generate power suppressed terms in the coefficients c_i. For the leading operator $\bar{Q}Q$ we will find its coefficient function to all orders of perturbation theory (in the limit of $N_c\to\infty$), demonstrating the convergence of the perturbative series. At the same time, the divergence of the condensate expansion in high orders will become manifest indirectly, through the occurrence of oscillating terms in Γ_{H_Q}, which appear with suppression factor $(1/m_Q)^9$ in the case at hand. Thus *all the four* problems formulated above will be answered.

We will perform our explicit calculations for $1+1$ QCD in the limit of a large number of colors N_c—the famous 't Hooft model [5–7]. For $N_c\to\infty$ only planar diagrams contribute in QCD; $1+1$ QCD has the additional special feature that one can choose a gauge such that there are no gluon self-interactions. Then only planar ladder diagrams have to be considered, and we have an exactly solvable theory in our hands. All hadronic matrix elements of interest are therefore calculable. This enables us to describe every given transition in two complementary ways: we can confront the OPE-based expression with a ''phenomenological'' representation for the same process obtained by saturating the rate by exclusive hadronic channels.

We want to take advantage of these unique features of the 't Hooft model to illustrate all crucial elements of heavy quark theory and the theory of inclusive heavy flavor decays in particular. One should keep in mind that heavy quark theory, as we know it now, is merely an adaptation of the general OPE-based approach. Some of the questions to be discussed below can therefore be actually formulated in a wider setting.

The 't Hooft model has been exploited as a theoretical laboratory for testing various analytic QCD methods in applied problems before. Heavy quark symmetry and heavy flavor decays were analyzed in Refs. [8–10]. The model was used recently for discussing general aspects of OPE (convergence of the OPE series, exponential terms violating duality, and so on) [11,12].

In Ref. [10] heavy flavor inclusive widths were calculated numerically, by adding the exclusive channels one by one. It was found that the inclusive width Γ_{H_Q} approaches its asymptotic (partonic) value, and the sum over the exclusive hadronic states converges rapidly. At the same time, small deviations from the asymptotic value observed in the numerical analysis [10] were claimed to be a signal of $1/m_Q$ corrections in the total width, in contradiction with the OPE-based result.

In this work we treat the very same problem, inclusive heavy flavor decays in $1+1$ QCD, *analytically*. We first develop a technique perfectly parallel to that in four-dimensional QCD [2]. It includes such elements as a complete operator analysis and the construction of the transition operator. Unlike four-dimensional QCD, the coefficient functions for the leading operator are exactly calculable (in the limit $N_c\to\infty$). Moreover, all relevant expectation values of the local operators involved in the problem are calculable too. We get a complete prediction through order $1/m_Q^4$.

Then we carry out a ''hadronic calculation'' of the same width, by saturating all open decay modes, using the 't Hooft equation [5]. By comparing the phenomenological representation of the total width with the OPE-based formula, we are able to identify, term-by-term, the subsequent terms of the heavy quark expansion. The situation actually turns out to be simpler than one could expect *a priori*:

(1) In the $1/m_Q$ expansion for the inclusive width corrections of the order $(1/m_Q)^2$, $(1/m_Q)^3$ and $(1/m_Q)^4$ to the parton width come only from the leading operator $\bar{Q}Q$, i.e., from the expansions of its OPE coefficient $c_{\bar{Q}Q}$ and its expectation value $\langle H_Q|\bar{Q}Q|H_Q\rangle$. Operators of higher dimension contribute to the total width first at order $(1/m_Q)^5$.

(2) The perturbative series in g^2/m_Q^2 for the OPE coefficient of the operator $\bar{Q}Q$ is completely defined by the one-loop renormalization of heavy quark mass. The result can be formulated in terms of the light-cone gauge formalism as the absence of renormalization.

These results are based on a general operator analysis. On the ''phenomenological'' side, we use a sum rule, which is a consequence of the 't Hooft equation, to show that the total width is determined by a quantity coinciding with the matrix element of the $c_{\bar{Q}Q}\bar{Q}Q$ term in the OPE expansion (1), through order $(1/m_Q)^4$. Thus, we observe a *perfect match* between the expression derived from the OPE and from adding up all relevant hadronic channels up to high order power corrections.

After testing the validity of OPE, we exploit results obtained *en route* in order to discuss the issue of oscillating contributions related to the high-order tails in the OPE series that are factorially divergent. Due to the simplicity of the model we can estimate them reliably. A nonmonotonous duality-violating component of the width for large m_Q is suppressed by high power of $1/m_Q$ which we determined. Implications of our analysis for real QCD are briefly discussed.

The remainder of the paper is organized as follows: after formulating the problem in Sec. II we construct the OPE and calculate the coefficients in Sec. III; after establishing the match between the OPE-based result for the inclusive width and the sum rules for the same width resulting from the 't Hooft equation through order $(1/m_Q)^4$ in Sec. IV, we discuss an appearance of oscillating terms in the order $(1/m_Q)^9$ and the duality violations they cause in Sec. V; in the same section we discussed along similar lines a possible pattern of the violation of the local duality for τ decays in 1+3 dimensions; in Sec. VI we comment on the paper [10] and analyze effects due to nonvanishing masses of light quarks; Sec. VII presents a general discussion and conclusions.

II. PRELIMINARIES

We start by formulating the problem and introducing our notation and conventions.

In two-dimensional QCD the Lagrangian looks superficially the same as in four dimensions

$$\mathcal{L}_{1+1} = -\frac{1}{4g^2} G^a_{\mu\nu} G^a_{\mu\nu} + \sum \bar{\psi}_i (i\not{D} - m_i)\psi_i ,$$

$$iD_\mu = i\partial_\mu + A^a_\mu T^a; \tag{2}$$

T^a denote generators of SU(N_c) in the fundamental representation, $G^a_{\mu\nu}$ the gluon field strength tensor and ψ_i the quark field (i is a flavor index) with a mass m_i; g the gauge coupling constant.

One has to keep the following peculiarities in mind: g carries dimension of mass as does $\bar{\psi}\psi$. The field strength $G^a_{\mu\nu}$ on the other hand has dimension M^2 in our normalization, just as in four-dimensional QCD. With the theory being superrenormalizable no (infinite) renormalization is needed; observables like the total width Γ_{H_Q} can be expressed in terms of the *bare* masses m_i and *bare* coupling g appearing in the Lagrangian. Anticipating the large N_c limit we will use a parameter β instead of g where

$$\beta^2 = \frac{g^2}{2\pi}\left(N_c - \frac{1}{N_c}\right) \quad \text{with} \quad \lim_{N_c\to\infty} \beta^2 = \text{finite}. \tag{3}$$

This dimensionful quantity β, which—in contrast to m_i—provides an intrinsic mass unit for the 't Hooft model, can be seen as the analog of $\Lambda_{\rm QCD}$ of four-dimensional QCD.

We need at least two quarks denoted by Q and q with masses m_Q and m_q, respectively, to realize heavy flavor transitions $Q\to q$. For quark masses we impose

$$m_Q - m_q \gg \beta, \tag{4}$$

where both $m_q \neq 0$ and $m_q = 0$ are allowed for. Condition (4) guarantees that the inclusive methods of Ref. [2] are applicable since it makes the energy release[1] in the weak decay large relative to the intrinsic scale β. We will actually employ the dimensionless ratio β/m_Q as our expansion parameter.

Next we need to introduce a flavor-changing weak interaction; we choose it to be of the current-current form:

$$\mathcal{L}^V_{\rm weak} = -\frac{G}{\sqrt{2}}(\bar{q}\gamma_\mu Q)(\bar{\psi}_a\gamma^\mu\psi_b). \tag{5}$$

Here G is an analog of the Fermi coupling constant; it is dimensionless in two dimensions. The fields $\psi_{a,b}$ can be either the light quark or the lepton fields to describe nonleptonic or semileptonic decays, respectively. In 1+1 dimensions the axial current reduces to the vector one. The most general current-current interaction contains an additional term where the vector currents are contracted via the antisymmetric $\epsilon_{\mu\nu}$ instead of $g_{\mu\nu}$. For the total width—our main focus here—such an additional term is of no importance. The product of scalar densities, on the other hand, is inequivalent to that of vector densities; we will briefly discuss it, but mainly focus on the $V\times V$ interaction (5).

For $N_c\to\infty$ factorization holds; i.e., the transition amplitude can be written as the product of matrix elements of the currents $\bar{q}\gamma_\mu Q$ and $\bar{\psi}_a\gamma^\mu\psi_b$. For the inclusive widths which are discussed below the property of factorization can be expressed as follows:

$$M_{H_Q}\Gamma_{H_Q} = \text{Im}\int d^2x\, i\langle H_Q|T\{\mathcal{L}_{\rm weak}(x)\mathcal{L}^\dagger_{\rm weak}(0)\}|H_Q\rangle$$
$$= G^2\int d^2x\, \text{Im}\Pi_{\mu\nu}(x)\, \text{Im}T^{\mu\nu}(x), \tag{6}$$

where $\Pi_{\mu\nu}(x)$ and $T_{\mu\nu}(x)$ are defined as

$$\Pi_{\mu\nu}(x) = i\langle 0|T\{\bar{\psi}_a(x)\gamma_\mu\psi_b(x)\bar{\psi}_b(0)\gamma_\nu\psi_a(0)\}|0\rangle, \tag{7}$$

$$T^{\mu\nu}(x) = i\langle H_Q|T\{\bar{q}(x)\gamma^\mu Q(x)\bar{Q}(0)\gamma_\nu q(0)\}|H_Q\rangle. \tag{8}$$

---

[1]It can hardly be overemphasized that it is the size of the energy release rather than of m_Q that controls the reliability of the expansion in four dimensions as well.

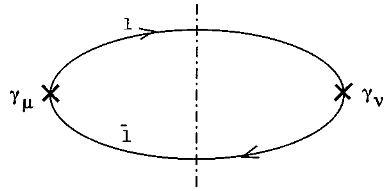

FIG. 1. Polarization operator for lepton current.

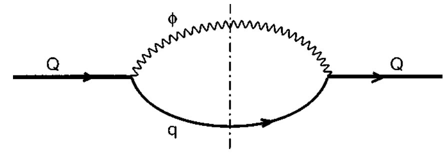

FIG. 2. Transition operator in the leading order. The wavy line of massless ϕ field substitutes the propagation of lepton pair.

This factorization follows from the fact that at $N_c \to \infty$ there is no communication between $\bar{\psi}_a \gamma_\mu \psi_b$ and $\bar{q} \gamma^\mu Q$ currents: any gluon exchange brings in a suppression factor $1/N_c^2$.

The only difference between the semileptonic and nonleptonic widths resides in $\Pi_{\mu\nu}(x)$, in the first case ψ_a are leptonic fields while in the second case they are the quark fields. At $m_\psi = 0$ we get one and the same $\Pi_{\mu\nu}(x)$ (up to the overall normalization factor N_c) as we will show shortly. For this reason at $m_\psi = 0$ the distinction between the nonleptonic and semileptonic cases is actually immaterial. At $m_\psi \neq 0$ quark and lepton polarization tensors $\Pi_{\mu\nu}(x)$ become different. This difference is proportional to powers of m_ψ. It will be discussed in Sec. VI. For the time being we will treat ψ's as massless leptons.

The one-loop graph determining $\Pi_{\mu\nu}$ is depicted in Fig. 1. For a massless fermion ψ we get the well-known expression

$$\Pi_{\mu\nu}(q) = \int d^2x e^{iqx} \Pi_{\mu\nu}(x) = -\frac{1}{\pi}\left(\frac{q_\mu q_\nu}{q^2} - g_{\mu\nu}\right). \quad (9)$$

This expression obtained from a one-loop graph is known to be exact. If ψ is a lepton field, this statement is trivial. If ψ is the quark field all gluon insertions inside the loop automatically vanish due to special properties of the two-dimensional γ matrices.[2] Thus, at $m_\psi = 0$ the only distinction between $\psi_{a,b}$ being quark rather than lepton fields is an overall factor N_c on the right-hand side of Eq. (9).

A remarkable feature of Eq. (9) is the occurrence of the pole at $q^2 = 0$, which is specific for the vector interaction. This means that a pair of massless leptons produced by the vector current is equivalent to one massless boson, whose coupling is proportional to its momentum q_μ. In the case of the quark fields, it is known [6] from the early days of the 't Hooft model that the vector current $\bar{\psi}_a \gamma^\mu \psi_b$ produces from the vacuum only one massless meson, the pion. This is readily seen by inspecting the 't Hooft equation [5].

For all computational purposes the vector current $\bar{\psi}_a \gamma^\mu \psi_b$ in Eq. (5) can thus be substituted by $\epsilon^{\mu\nu} \partial_\nu \phi / \sqrt{\pi}$ where ϕ denotes a pseudoscalar massless noninteracting field,

$$\tilde{\mathcal{L}}_{\text{weak}}^V = -\frac{G}{\sqrt{2\pi}} \bar{q} \gamma_\mu Q \epsilon^{\mu\nu} \partial_\nu \phi. \quad (10)$$

In other words, the problem is formulated as the inclusive decay of the heavy quark Q into a lighter quark q plus a sterile boson ϕ. More exactly, we deal with the decays of a Q containing hadron H_Q into a q containing final hadronic state X_q plus ϕ.

Let us pause here for two remarks. (i) The fact that the interaction vertex of the massless field ϕ involves $\epsilon^{\mu\nu}$, see Eq. (10), is most obvious when $\psi_{a,b}$ are quark fields. For in this case ϕ is the pion, as mentioned above, and the pion is coupled to the vector current $\bar{\psi}_a \gamma_\mu \psi_b$ obviously through $\epsilon^{\mu\nu}$. The case of the leptonic fields $\psi_{a,b}$ is indistinguishable; therefore, the coupling of ϕ is the same. (ii) To keep the analysis to be presented below as clean and transparent as possible we want to be free of annihilation and Pauli interference contributions to the total width [at least through $\mathcal{O}(1/m_Q^4)$]. This is readily achieved by assuming throughout the paper that the spectator light quark q_{sp} in H_Q is distinct from q.

In the leading approximation the transition operator is determined by the diagram of Fig. 2, where the wavy line corresponds to the ϕ quantum. A straightforward calculation yields for the transition operator

$$\hat{T}_0 = c_{\bar{Q}Q}^0 \bar{Q}Q; \quad 2\,\text{Im}\, c_{\bar{Q}Q}^0 = \Gamma_Q = \frac{G^2}{4\pi} \cdot \frac{m_Q^2 - m_q^2}{m_Q}, \quad (11)$$

where Γ_Q is the decay width for a free quark Q as evaluated in the parton model. This parton expression will serve as reference in analyzing the $(1/m_Q)^n$ corrections to the total width Γ.

III. OPERATOR PRODUCT EXPANSION FOR INCLUSIVE WIDTHS

A. Catalog of operators

The $1/m_Q$ expansion for inclusive widths of heavy flavor hadrons is constructed from the Lorentz invariant weak transition operator [13]

$$\hat{T}(Q \to Q) = \int d^2x\, iT\{\mathcal{L}_{\text{weak}}(x) \mathcal{L}_{\text{weak}}^\dagger(0)\} = \sum c_i(\mu) \mathcal{O}_i(\mu). \quad (12)$$

The local operators $\mathcal{O}_i$ are ordered according to their dimensions. The leading one is $\bar{Q}Q$ with dimension $d_{\bar{Q}Q} = 1$. Higher operators have dimensions $d_i > 1$. By dimensional counting the corresponding coefficients are proportional to $(1/m_Q)^{(d_i - 2)}$. The ratio of the coefficients $c_i / c_{\bar{Q}Q}$ is proportional to $m_Q^{-d_i + d_{\bar{Q}Q}}$.

[2] Namely, one uses the fact that $\gamma^\alpha \gamma^\mu \gamma_\alpha = 0$ and any odd number of γ matrices reduces to one.

The coefficients c_i are determined in perturbation theory as a series in β^2/m_Q^2. It is crucial that these coefficients are saturated by the domain of virtual momenta $\sim m_Q$ and are infrared stable by construction. (All infrared contributions reside in the matrix elements of the operators $\mathcal{O}_i$.) At this point we should mention a drastic distinction between four- and two-dimensional QCD. In four dimensions the expansion parameter for the coefficients is the running coupling $\alpha_s(m_Q)$; nonperturbative contributions to the coefficients coming from distances $\sim 1/m_Q$ could in principle show up in the form $\exp(-C/\alpha_s(m_Q)) \sim (\Lambda_{\rm QCD}/m_Q)^\delta$ where δ is some unknown positive index, not necessarily integer. In two-dimensional QCD such terms cannot appear: an analog of the exponential term above would be $\exp(-Cm_Q^2/\beta^2)$.

A note concerning the choice of the normalization point μ: we will imply that

$$m_Q \gg \mu \gg \beta. \tag{13}$$

In this range there is no real dependence on μ, so we will suppress the argument μ both in the coefficient functions and operators.

The coefficient functions are not the only source of a m_Q dependence. The matrix elements of the operators $\mathcal{O}_i$ contain an implicit m_Q dependence too. (We recall that in our formalism, unlike HQET [14], the fields of which the operators $\mathcal{O}_i$ are built are the standard Heisenberg operators, rather than asymptotic in m_Q.) In particular, for the leading operator $\bar{Q}Q$ we have the following relation:

$$\int d^2x \bar{Q}Q = \int d^2x \left\{ \bar{Q}\gamma_0 Q + \bar{Q}\frac{\pi_\mu \pi^\mu + \frac{i}{2}\sigma^{\mu\nu}G_{\mu\nu}}{2m_Q^2} Q \right\}, \tag{14}$$

where $\pi_\mu = iD_\mu - g_{\mu 0} m_Q$ and the integration over x allows us to omit terms which are total derivatives.

In the rest frame of the hadron H_Q the expectation value of $\bar{Q}\gamma_0 Q$ counts the number of Q quarks,

$$\frac{1}{2M_{H_Q}} \langle H_Q | \bar{Q}\gamma_0 Q | H_Q \rangle = 1. \tag{15}$$

The factor $1/2M_{H_Q}$ will be present in all matrix elements; it corresponds to a relativistic normalization of the states,

$$\langle H_Q(\vec{p}\,') | H_Q(\vec{p}) \rangle = 2E_{H_Q} \delta(\vec{p}\,' - \vec{p}). \tag{16}$$

From relation (14) the matrix element of $\bar{Q}Q$ is therefore unity, up to a quadratic correction:

$$\frac{1}{2M_{H_Q}} \langle H_Q | \bar{Q}Q | H_Q \rangle = 1 + \mathcal{O}\left(\frac{1}{m_Q^2}\right). \tag{17}$$

Moreover relation (14) provides an operator form for $1/m_Q^2$ corrections. They come from $\bar{Q}\pi_\mu\pi^\mu Q/2m_Q^2$ which equals $\bar{Q}D_1^2 Q/2m_Q^2$ up to $1/m_Q^4$ corrections. Notice, that the operator $\bar{Q}D_1^2 Q$ is Lorentz noncovariant and cannot enter directly into the OPE for the total width but, as we see, enters indirectly through the matrix element of the operator $\bar{Q}Q$.

Equation (14) contains also the "chromomagnetic" operator and it looks as if this operator contributes to $1/m_Q^2$ corrections. It is not, however, the case. Indeed, this operator can be rewritten as follows:

$$\begin{aligned}\mathcal{O}_G &= \frac{i}{2}\bar{Q}\sigma^{\mu\nu}G_{\mu\nu}Q = \frac{i}{2}\bar{Q}\gamma_5 \epsilon^{\mu\nu}G_{\mu\nu}Q \\ &= \frac{i}{4}\bar{Q}\gamma_5\epsilon^{\mu\nu}G_{\mu\nu}(1-\gamma^0)Q + \frac{i}{4}\bar{Q}(1-\gamma^0)\gamma_5\epsilon^{\mu\nu}G_{\mu\nu}Q,\end{aligned} \tag{18}$$

where the relation $\sigma^{\mu\nu} = \gamma_5 \epsilon^{\mu\nu}$ is used. Taking advantage of the non-relativistic equations of motion to replace $(1-\gamma^0)Q$ by $(1/m_Q)\gamma^1 iD_1 Q$ we get (up to total derivatives)

$$\begin{aligned}\mathcal{O}_G &= -\frac{1}{2m_Q}\bar{Q}(D^\mu G_{\mu\nu})\gamma^\nu Q + \mathcal{O}\left(\frac{1}{m_Q^2}\right) \\ &= \frac{g^2}{2m_Q}\bar{Q}\gamma_\mu t^a Q \sum_q \bar{q}\gamma^\mu t^a q + \mathcal{O}\left(\frac{1}{m_Q^2}\right),\end{aligned} \tag{19}$$

where t^a stand for the generators of the color group $SU(N_c)$. Thus, the operator $\mathcal{O}_G$ reduces to a four-fermion operator $\mathcal{O}_{4q}$ with coefficient g^2/m_Q.

The absence of operators with the gluon field strength tensor $G_{\mu\nu}$ in the OPE is a specific feature of two-dimensional QCD. The physical reason for the reducibility of the gluonic operators is the absence of real gluons in two dimensions. A particular consequence of Eq. (19) is that in Eq. (14) the chromomagnetic operator generates $1/m_Q^3$ terms only.

Thus we come to the following representation for the matrix element of the leading operator $\bar{Q}Q$:

$$\begin{aligned}\frac{\langle H_Q|\bar{Q}Q|H_Q\rangle}{2M_{H_Q}} &= 1 - \frac{1}{2m_Q^2}\frac{\langle H_Q|\bar{Q}(-D_1^2)Q|H_Q\rangle}{2M_{H_Q}} \\ &\quad + \frac{g^2}{2m_Q^3}\frac{\langle H_Q|\bar{Q}\gamma_\mu t^a Q\sum_q \bar{q}\gamma^\mu t^a q|H_Q\rangle}{2M_{H_Q}} \\ &\quad + \mathcal{O}\left(\frac{1}{m_Q^4}\right).\end{aligned} \tag{20}$$

Let us proceed further with the operator analysis. The first subleading operator is the dimension-two four-fermion operator of the type

$$\mathcal{O}_{4q} = \bar{Q}\Gamma_1 Q \bar{q}\Gamma_2 q, \tag{21}$$

where $\Gamma_{1,2}$ denote color and spinor matrices. This is in distinction with 1+3 QCD where the first subleading operator

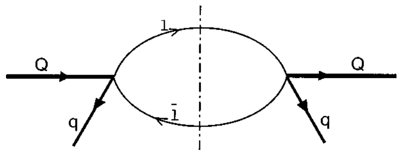

FIG. 3. Four-fermion operators in the leading order.

was $\bar{Q}\sigma^{\mu\nu}G_{\mu\nu}Q$. On dimensional grounds the operator $\mathcal{O}_{4q}$, if present, could produce a *linear* $1/m_Q$ correction in the total decay width. To this end the corresponding coefficient must arise in zeroth order in the coupling g^2 (see the diagram in Fig. 3).

For the $V\times V$ weak coupling of Lagrangian (10) this graph vanishes identically in the imaginary part, Im $c_{4q}=0$, provided the leptons are massless, see discussion in Sec. II. (If the fields propagating in the loop are massive, the operator $\mathcal{O}_{4q}$ appears. The corresponding modifications are considered in Sec. VI.)

However, in the case of a scalar-scalar weak interaction of the form

$$\mathcal{L}^S_{\rm weak}=-\frac{G_S}{\sqrt{2}}(\bar{q}Q)(\bar{\psi}_a\psi_b) \tag{22}$$

the graph of Fig. 3 is not zero, and gives the following contribution to the transition operator:

$$\mathrm{Im}\,T_S=\frac{G_S^2}{4}(\bar{Q}q)(\bar{q}Q). \tag{23}$$

With this contribution the total width takes the form (we put $m_q=0$ for simplicity):

$$\Gamma^S_{H_Q}=\frac{G_S^2 m_Q}{16\pi}\left(1+\frac{4\pi}{m_Q}\frac{\langle H_Q|(\bar{Q}q)(\bar{q}Q)|H_Q\rangle}{\langle H_Q|\bar{Q}Q|H_Q\rangle}\right). \tag{24}$$

Let us emphasize that, unlike four-dimensional QCD where no operator can induce a $1/m_Q$ contribution to the total width [15,16], this can happen in two dimensions. The vanishing of Im c_{4q} for the $V\times V$ weak Lagrangian in the leading order is a specific dynamical feature of this particular Lorentz structure of the weak interaction. Note at this point that the argumentation presented in Ref. [15] was not sufficient to prove the absence of $1/m_Q$ corrections in real QCD, see the discussion in Sec. V A.

In Sec. III B 3 we will show that Im c_{4q} vanishes not only in the leading order in strong coupling but also in the order g^2. The first nonvanishing contribution to Im c_{4q} comes in the order g^4 what leads to $1/m_Q^5$ corrections to the width.

Next in the list comes the dimension-three operator containing six quark fields:

$$(\bar{Q}\Gamma_1 Q)(\bar{q}_a\Gamma_2 q_b)(q_c\Gamma_3 q_d).$$

Along the same line of reasoning as for four-fermion operators we show that six-fermion operators appear only in the order g^6. It therefore contributes in order $1/m_Q^8$. For multifermion operators every extra $\bar{q}q$ pair brings in an extra $1/m_Q^3$ suppression.

In summary: to a quite high accuracy the operator $\bar{Q}Q$ is the only one to contribute.

B. Calculating coefficients

1. Light-cone gauge and nonrenormalization theorem

The calculations are most conveniently done in the light-cone gauge. This technology can be traced back to the pioneering work of ’t Hooft [5], and has been well studied in the literature. In this formalism the energy-momentum vector is described by

$$p^{\pm}=\frac{1}{\sqrt{2}}(E\pm p), \tag{25}$$

so that the mass-shell condition becomes $p^2=2p_+p_-=m^2$.

Let us write down the Lagrangian of the model in the light-cone formalism:

$$\mathcal{L}=\sum_i \chi_i^\dagger\left[i\partial_+-\frac{m_i^2}{2i\partial_-}\right]\chi_i + \frac{g^2}{2}\left(\sum_i \chi_i^\dagger t^a\chi_i\right)\frac{1}{\partial_-^2}\left(\sum_k \chi_k^\dagger t^a\chi_k\right). \tag{26}$$

In this formalism two-component quark q_i fields are expressed via the one-component fermionic fields χ_i,

$$q_i=\frac{1}{2^{1/4}}\begin{pmatrix}\chi_i\\ \dfrac{m_i}{\sqrt{2}i\partial_-}\chi_i\end{pmatrix} \tag{27}$$

(in the basis where $\gamma_5=\gamma^0\gamma^1$ is diagonal). With the gauge fixed by $A_-=0$, the A_+ component is expressed in terms of the quark fields.

The weak interaction (10) takes the form:

$$\tilde{\mathcal{L}}^V_{\rm weak}=-\frac{G}{\sqrt{2\pi}}\left\{\chi_q^\dagger\chi_Q\partial_+\phi - \frac{m_q m_Q}{2}\left[\frac{1}{i\partial_-}\chi_q\right]^\dagger\left[\frac{1}{i\partial_-}\chi_Q\right]\partial_-\phi\right\}. \tag{28}$$

A remarkable simplification occurs due to ϕ carrying lightlike momentum q_μ: $q^2=2q_+q_-=0$. We can satisfy this condition by choosing the ‘‘spatial’’ component of the momentum $q_-=0$, i.e., $\partial_-\phi=0$. (This means that the ϕ quantum is a left-mover.) Thus, on the ϕ ‘‘mass shell’’ the second term (containing $\partial_-\phi$) in Eq. (28) vanishes and the weak $Qq\phi$ coupling takes a simple form,

$$\tilde{\mathcal{L}}^V_{\rm weak}=-\frac{G}{\sqrt{2\pi}}\chi_q^\dagger\chi_Q\partial_+\phi. \tag{29}$$

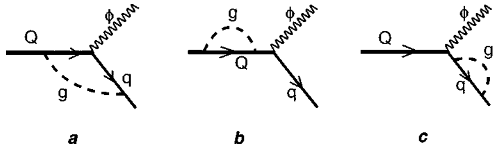

FIG. 4. Radiative corrections to the weak coupling $Qq\phi$. Dashed lines denote the gluons.

Here we come to a very important point. In the 't Hooft model nonrenormalization theorem for flavor nondiagonal currents at $q_-=0$ arises: *all* corrections to the weak vertex at this point vanish. In the light-cone formalism the kinematical point $q_-=0$ looks analogous to the zero recoil point in four-dimensional heavy quark theory. The analogy is a superficial one, however.[3] In $1+3$ QCD flavor nondiagonal currents at the zero recoil are renormalized by radiative and power corrections.

To prove the theorem let us consider the gluon corrections to the weak coupling (29). To order g^2 the relevant graphs are depicted in Fig. 4.

The Feynman rules can be read off from Eqs. (26), (29). In particular, the weak $Qq\phi$ vertex is

$$V=-i\frac{G}{\sqrt{2\pi}}q_+ . \tag{30}$$

The graph 4(a) gives rise to the following expression:

$$\Delta V=V\cdot i\frac{\beta^2}{4\pi}\int d^2k\frac{1}{k_-^2}\times\frac{1}{(p_q+k)_+-\dfrac{m_q^2-i\epsilon}{2(p_q+k)_-}}\cdot\frac{1}{(p_Q+k)_+-\dfrac{m_Q^2-i\epsilon}{2(p_Q+k)_-}}. \tag{31}$$

The integration over k_+ can easily be performed by the residue method. It is clear then that the integration over k_+ produces a nonzero result only in the case of opposite signs of $(p_q+k)_-$ and $(p_Q+k)_-$ (the poles should be on the different sides of the integration path). For $q_-=(p_Q-p_q)_-=0$ these signs are certainly the same, $(p_q+k)_-=(p_Q+k)_-$, and there is no correction to the vertex.

To finish up with the g^2 correction to the weak coupling we need to add graphs (b) and (c) of Fig. 4 containing the self-energies Σ_Q and Σ_q of Q and q quarks.

$$\Sigma_Q(p_+,p_-)=-i\frac{\beta^2}{4\pi}\int d^2k\frac{1}{k_-^2}\cdot\frac{1}{p_++k_+-\dfrac{m_Q^2-i\epsilon}{2(p_-+k_-)}}. \tag{32}$$

The integration contains a single pole only as a function of k_+. Unlike the previously considered vertex correction, though, the integration over k_+ gives a nonzero result because the integral over the large semicircle in the complex plane of k_+ does not vanish. The integration over k_- requires an infrared regularization.[4] Following 't Hooft [5] we define the integration in Eq. (32) by putting a symmetric ultraviolet cutoff K for the k_+ integration, and a symmetric infrared cutoff λ for the k_- integration,

$$|k_+|<K, \quad |k_-|>\lambda. \tag{33}$$

Then at $|p_+|\ll K$ the result for Σ is

$$\Sigma_Q(p_+,p_-)=\left[\frac{\beta^2}{2p_-}-\frac{\beta^2}{2\lambda}\epsilon(p_-)\right]\theta(|p_-|-\lambda). \tag{34}$$

The independence of Σ on p_+ means that no Z factor appears. The first term corresponds to a shift in the quark masses,

$$m_Q^2\to m_Q^2-\beta^2, \quad m_q^2\to m_q^2-\beta^2. \tag{35}$$

The second term produces a (noncovariant) shift in the reference point for the light-cone energy on mass shell,

$$p_+=\frac{m_Q^2-\beta^2}{2p_-}+\frac{\beta^2}{2\lambda} \quad (p_->\lambda). \tag{36}$$

This shift produces no effect on the widths. One-loop radiative corrections thus do not affect $Qq\phi$ transitions besides the mass shift given by Eq. (35).

Moreover, it stays true for higher loops as well within the 't Hooft model. For in the limit $N_c\to\infty$ there are no fermion loop insertions into the gluon propagators. Then the higher loop corrections to the vertex, as well as to the self-energy, vanish in the way discussed above since the integration over k_+ yields zero.

Notice that the nonrenormalization theorem we derive within the 't Hooft model is a stronger statement than the one about zero recoil in four-dimensional QCD where radiative and power corrections break the nonrenormalization of flavor nondiagonal currents.

2. The leading coefficient $c_{\bar{Q}Q}$

Now it is simple to account for higher orders in the coefficient $c_{\bar{Q}Q}$ of the leading operator $\bar{Q}Q$. To zeroth order in g^2 this coefficient was determined in Sec. II, see Fig. 2 and Eq. (11). As just discussed higher loop corrections merely shift the quark masses, Eq. (35), and therefore we get the coefficient $c_{\bar{Q}Q}$ in all orders,

[3] In particular, for heavy-to-heavy transitions at $q_-=0$ there is no dominance of the ground state production in the 't Hooft model.

[4] OPE ensures that the dependence on the infrared regularization disappears in the width as long as it is the same at all stages of the calculation.

$$2\,\mathrm{Im}\,c_{\bar{Q}Q}=\frac{G^2}{4\pi}\cdot\frac{m_Q^2-m_q^2}{\sqrt{m_Q^2-\beta^2}}. \qquad (37)$$

Combining this result with Eq. (20) and with the suppression of the four-fermion operators, see Sec. III B 3, we conclude that (i) there is no $1/m_Q$ corrections in the total width, much in the same way as in actual QCD [15,16]; (ii) corrections $1/m_Q^2, 1/m_Q^3$ and $1/m_Q^4$ to the total width are associated exclusively with the operator $\bar{Q}Q$.

In Sec. IV we will prove a stronger statement: irrespective of the explicit form of these corrections, the hadronic saturation yields exactly the same result for the total width as the contribution of $c_{\bar{Q}Q}\bar{Q}Q$ in OPE.

The expression (37) for $c_{\bar{Q}Q}$ refers to a low normalization point, $\beta\ll\mu\ll m_Q$. In order to calculate the matrix element of $\bar{Q}Q$ over H_Q we will need to express $\bar{Q}Q$ in terms of χ_Q. It goes without saying that the operator $\bar{Q}Q$ must be taken at the same normalization point. Then, the resulting series in β^2/m_Q^2 cancels against a similar expansion coming from the operator $\bar{Q}Q$; there is no β dependence in the product $c_{\bar{Q}Q}\bar{Q}Q$. This can be seen by rewriting $\bar{Q}Q$ in terms of the unrenormalized one-component field χ_Q and mass m_Q:

$$\bar{Q}Q=\chi_Q^{\dagger}\frac{m_Q}{i\partial_-}\chi_Q. \qquad (38)$$

In evolving down to μ, higher orders lead to the substitution $m_Q\to\sqrt{m_Q^2-\beta^2}$ in this relation as well. With the quark mass substitution being the only effect of the radiative corrections we have

$$2\,\mathrm{Im}\,c_{\bar{Q}Q}\bar{Q}Q=\frac{G^2}{4\pi}(m_Q^2-m_q^2)\chi_Q^{\dagger}\frac{1}{i\partial_-}\chi_Q. \qquad (39)$$

The statement that the product $c_{\bar{Q}Q}\bar{Q}Q$ is renormalization group invariant is trivial, of course. A nontrivial part of the result is encoded in Eq. (39), which is valid to all orders in g. One could obtain this result by doing calculations at $\mu\gg m_Q$ when the mass of the Q quark coincides with its "bare" value m_Q, at $\mu=m_Q$, or at $\mu\ll m_Q$, when a nonlogarithmic evolution of the $\bar{Q}Q$ operator and its coefficient functions must be taken into account, the outcome is the same, see Eq. (39). To make contact with the 't Hooft equation (i.e., to calculate $\bar{Q}Q$ in terms of the 't Hooft wave function defined for bare quantities) we will need Eq. (39) at the ultraviolet cutoff. Note that it can be conveniently rewritten as

$$2\,\mathrm{Im}\,c_{\bar{Q}Q}\bar{Q}Q=\Gamma_Q\chi_Q^{\dagger}\frac{m_Q}{i\partial_-}\chi_Q. \qquad (40)$$

In Sec. IV B we will find the matrix element of $\chi_Q^{\dagger}(m_Q/i\partial_-)\chi_Q$ and show that the corresponding expression for the total width coincides with the one obtained through the hadronic saturation.

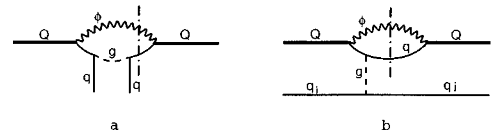

FIG. 5. Four-fermion operators in g^2 order.

3. Four-fermion and multifermion operators

As discussed above for the current-current weak interactions four-fermion operators do not arise to zeroth order in the strong coupling. Diagrams generating four-fermion operators in the g^2 order are shown in Fig. 5.

The light-cone gauge turns out to be again a convenient tool to use. Let us start for illustration with the simple one-loop diagram of Fig. 2. Feynman rules in the light-cone gauge, see Sec. III B 1, lead to the following expression:

$$\mathrm{Im}\,i\int d^2k\,\delta(2k_+k_-)\frac{k_+^2}{(p_Q-k)_+-\dfrac{m_q^2-i\epsilon}{2(p_Q-k)_-}}. \qquad (41)$$

Here $\delta(2k_+k_-)$ appears from the cut of the ϕ propagator, the cutting of the q quark propagator done by the taking an imaginary part. The integration over k_- is immediate due to the $k_-=0$ root of delta function (the other root, $k_+=0$, gives the same contribution, so just a factor 2)

$$\mathrm{Im}\,i\int dk_+k_+\frac{1}{(p_Q-k)_+-\dfrac{m_q^2-i\epsilon}{2(p_Q)_-}}=\pi\left[(p_Q)_+-\frac{m_q^2}{2(p_Q)_-}\right]. \qquad (42)$$

It means that the term

$$\hat{T}_0=\mathrm{const}\,\chi_Q^{\dagger}\left[i\partial_+-\frac{m_q^2}{2i\partial_-}\right]\chi_Q \qquad (43)$$

appears in the transition operator. In the zeroth order in the 't Hooft coupling the equation of motion for the χ_Q field is

$$\partial_+\chi_Q=m_Q^2/(2i\partial_-)\chi_Q. \qquad (44)$$

In the rest frame $i\partial_-\to p_-=m_Q/\sqrt{2}$, and we reproduce Eq. (11).

Let us now apply the same technique to the loop part in Fig. 5(b) in the limit of vanishing gluon momentum. Integrating over k_- we get

$$\mathrm{Im}\,i\int dk_+k_+\frac{1}{\left[(p_Q-k)_+-\dfrac{m_q^2-i\epsilon}{2(p_Q)_-}\right]^2}=\pi. \qquad (45)$$

It produces the term

$$\hat{T}_1=\mathrm{const}\,\chi_Q^{\dagger}A_+\chi_Q \qquad (46)$$

in the transition operator, with the same overall factor as in Eq. (43). Summing up $\hat{T}_0$ and $\hat{T}_1$ results in the substitution of

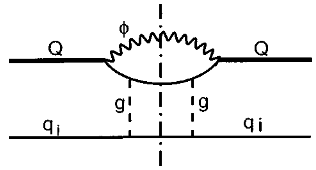

FIG. 6. Four-fermion operators in g^4 order.

$i\partial_+$ by $iD_+=i\partial_+ + A_+$. In this order the equation of motion (44) also should be modified by the same substitution. The net result is that no change in OPE coefficients is produced by the loop in Fig. 5(b).

Note that it is true not only for vanishing momentum of gluon field but for any soft field as well, in other words, terms with derivatives of A_μ do not appear in the transition operator in one-loop order [17]. Note also that if extra gluons are emitted from the loop we get zero for the diagram. Indeed, it increases the power in the integrand of Eq. (45) and the integral over the large semicircle in the complex plane of k_+ vanishes. It is the reason why diagrams of the type of Fig. 5(b) give no rise to multifermion operators.

To finish up with multifermion operators we need to account for diagrams of the type given in Fig. 5(a). The Feynman expression for this diagram after integrating over k_- is

$$\mathrm{Im}\, i\int \mathrm{d}k_+ k_+ \frac{1}{\left[(p_Q-k)_+ - \dfrac{m_q^2 - i\epsilon}{2(p_Q)_-}\right]^2}\,\frac{1}{[(p_Q-p_q)_-]^2}$$

$$=\frac{\pi}{[(p_Q-p_q)_-]^2}. \tag{47}$$

It is simple to check then that the diagram 5(a) cancels out against similar diagrams where the gluon exchange is between q and Q quarks. In case of extra gluon insertions in one loop (relevant for six-fermion and higher dimension operators) we get a vanishing result right away.

Thus, we proved that four-fermion and multifermion operators do not arise at the level of one loop. They show up at the level of the second loop; see Fig. 6 for four-fermion operator. The dimensional counting reveals then that four-fermion operators give $1/m_Q^5$ correction to the total width.

C. OPE representation for inclusive width

Putting everything together we get the OPE representation for the inclusive width:

$$\Gamma_{H_Q}=\frac{G^2}{4\pi}\,\frac{m_Q^2-m_q^2}{\sqrt{m_Q^2-\beta^2}}\left[\frac{\langle H_Q|\bar{Q}Q|H_Q\rangle}{2M_{H_Q}}+\mathcal{O}\left(\frac{1}{m_Q^5}\right)\right]$$

$$=\Gamma_Q\left[\frac{\left\langle H_Q\left|\chi_Q^\dagger\dfrac{m_Q}{i\partial_-}\chi_Q\right|H_Q\right\rangle}{2M_{H_Q}}+\mathcal{O}\left(\frac{1}{m_Q^5}\right)\right]. \tag{48}$$

We have thus obtained a very simple result.

(1) The partonic expression Γ_Q represents the asymptotic term for $m_Q\to\infty$.

(2) There is no $1/m_Q$ contribution in OPE as long as the weak interactions are of the $V\times V$ type.

(3) Through order $1/m_Q^4$ only a single operator contributes, $\bar{Q}Q$.

(4) The leading correction $\sim\mathcal{O}(1/m_Q^2)$ enters through the expectation value $(1/M_{H_Q})\langle H_Q|\chi_Q^\dagger(m_Q/i\partial_-)\chi_Q|H_Q\rangle$.

The second part of Eq. (48) has been written in terms of the light-cone operators to provide a way of rewriting the matrix element in terms of the 't Hooft wave function of the hadron H_Q. The OPE result for the inclusive width can be recast in terms of the sum over exclusive hadronic channels. This will be proven next.

IV. MATCH BETWEEN OPE-BASED EXPRESSIONS AND HADRONIC SATURATION

A. Exclusive widths via the 't Hooft wave function

With 1+1 QCD describing manifestly confining dynamics, its spectrum consists of mesonic quark-antiquark bound states. In the $N_c\to\infty$ limit these mesons are stable in regard to strong decays. The masses and the light-cone wave functions $\varphi(x)$ (with $x\epsilon[0,1]$ meaning a portion of momentum carried by the quark) of these mesons can be determined as eigenfunctions and eigenvalues of the 't Hooft equation. In particular, the initial state $H_Q=[Q\bar{q}_{sp}]$ is the ground state in the sector with the heavy quark Q and the spectator antiquark $\bar{q}_{sp}$. Its wave function φ_{H_Q} satisfies the following equation:

$$M_{H_Q}^2\varphi_{H_Q}(x)=\left[\frac{m_Q^2-\beta^2}{x}+\frac{m_{sp}^2-\beta^2}{1-x}\right]\varphi_{H_Q}(x)$$

$$-\beta^2\int_0^1 dy\,\frac{\varphi_{H_Q}(y)}{(y-x)^2}, \tag{49}$$

where m_{sp} denotes the mass of the spectator antiquark and the integral is understood in the principal value prescription. The solutions to the equation 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:

$$\frac{\pi\gamma_0}{\tan\pi\gamma_0}=-\frac{m_Q^2-\beta^2}{\beta^2},\qquad \frac{\pi\gamma_1}{\tan\pi\gamma_1}=-\frac{m_{sp}^2-\beta^2}{\beta^2}. \tag{50}$$

The masses M_n and wave functions φ_n of final mesons $h_n=[q\bar{q}_{sp}]_n$ are defined by the same 't Hooft equation with m_Q substituted by m_q:

$$M_n^2\varphi_n(x)=\left[\frac{m_q^2-\beta^2}{x}+\frac{m_{sp}^2-\beta^2}{1-x}\right]\varphi_n(x)-\beta^2\int_0^1 dy\,\frac{\varphi_n(y)}{(y-x)^2}. \tag{51}$$

The functions φ_n form a complete basis, i.e.,

$$\sum_n \varphi_n(x)\varphi_n(y)=\delta(x-y). \tag{52}$$

Let us find out now how an exclusive width Γ_n of $H_Q \to h_n \phi$ decay is expressed via wave functions φ_{H_Q}, φ_n of initial and final mesons,

$$\Gamma_n = \frac{1}{2M_{H_Q}} \cdot \frac{1}{(M_{H_Q}^2 - M_n^2)} \left| \langle h_n \phi | \frac{G}{\sqrt{2\pi}} \chi_q^\dagger \chi_Q \phi q_+ | H_Q \rangle \right|^2$$
$$= \frac{G^2}{4\pi} \frac{M_{H_Q}^2 - M_n^2}{M_{H_Q}} \left[\frac{\langle h_n | \chi_q^\dagger \chi_Q | H_Q \rangle}{\sqrt{2} M_{H_Q}} \right]^2, \qquad (53)$$

where the factor $1/(M_{H_Q}^2 - M_n^2)$ is the Lorentz invariant phase space (LIPS) of the two-particle final state and for the matrix element we have used Eq. (29) in the kinematics where the ϕ momentum q_μ is

$$q_- = 0, \quad q_+ = \frac{1}{\sqrt{2} M_{H_Q}} (M_{H_Q}^2 - M_n^2). \qquad (54)$$

It is simple then to write the matrix element of $\chi_q^\dagger \chi_Q$ in terms of the 't Hooft wave functions,

$$\Gamma_n = \frac{G^2}{4\pi} \frac{M_{H_Q}^2 - M_n^2}{M_{H_Q}} \left| \int_0^1 dx \varphi_n(x) \varphi_{H_Q}(x) \right|^2. \qquad (55)$$

B. Sum rules

Using the completeness condition (52) we can derive sum rules for these partial widths by weighing them with powers of $M_{H_Q}^2 - M_n^2$. The first one is

$$\frac{4\pi M_{H_Q}}{G^2} \sum_{n=0}^{\infty} \frac{\Gamma_n}{M_{H_Q}^2 - M_n^2} = \sum_{n=0}^{\infty} \left| \int_0^1 dx \varphi_n(x) \varphi_{H_Q}(x) \right|^2$$
$$= \int_0^1 dx \varphi_{H_Q}^2(x) = 1. \qquad (56)$$

Note that the sum runs over *all* states h_n including those unaccessible in the real decays of H_Q, i.e., with masses $M_n > M_{H_Q}$. These transitions are still measurable by the process of inelastic lepton scattering off the H_Q meson. This sum rule is an analog of the first Bjorken sum rule and was discussed in [18].

To get the next sum rules let us multiply Eq. (49) by $\varphi_n(x)$ and Eq. (51) by $\varphi_{H_Q}(x)$, respectively. After integrating over x and subtracting we find

$$(M_{H_Q}^2 - M_n^2) \int_0^1 dx \varphi_n(x) \varphi_{H_Q}(x)$$
$$= (m_Q^2 - m_q^2) \int_0^1 \frac{dx}{x} \varphi_n(x) \varphi_{H_Q}(x). \qquad (57)$$

Two more sum rules then arise:

$$\frac{4\pi M_{H_Q}}{G^2} \sum_{n=0}^{\infty} \Gamma_n = (m_Q^2 - m_q^2) \int_0^1 \frac{dx}{x} \varphi_{H_Q}^2(x), \qquad (58)$$

$$\frac{4\pi M_{H_Q}}{G^2} \sum_{n=0}^{\infty} \Gamma_n (M_{H_Q}^2 - M_n^2) = (m_Q^2 - m_q^2)^2 \int_0^1 \frac{dx}{x^2} \varphi_{H_Q}^2(x). \qquad (59)$$

The second and third sum rules differ from the first one in two aspects: they depend on the quark masses m_Q and m_q explicitly and the integral over the wave function is not fixed by a normalization condition; it can, however, be calculated in the 't Hooft model. One should note that while the integrand $\varphi_{H_Q}^2(x)/x^2 \sim x^{-\beta^2/m_Q^2}$ is singular at $x=0$, it is still integrable, since $\beta^2/m_Q^2 \ll 1$. Note also that expanding the second sum rule (58) in $1/m_Q$ to the linear order we reproduce the corresponding sum rule of Ref. [18].

The sum rules above provide us with detailed information on the saturation of the sums over the final hadronic state. The quantity

$$w_n = \frac{4\pi M_{H_Q}}{G^2} \frac{\Gamma_n}{M_{H_Q}^2 - M_n^2} \qquad (60)$$

can be interpreted as a normalized probability of producing the state n. Indeed $\Sigma w_n = 1$ according to the first sum rule (56). Then the sum rule (58) implies

$$\langle M_{H_Q}^2 - M_n^2 \rangle = (m_Q^2 - m_q^2) \int_0^1 \frac{dx}{x} \varphi_{H_Q}^2(x). \qquad (61)$$

If both masses m_{sp} and m_q are smaller or of the order of β we conclude that

$$\langle M_n^2 \rangle = \left\langle \frac{1}{x} - 1 \right\rangle m_Q^2 + \mathcal{O}(\beta^2) \sim \beta m_Q. \qquad (62)$$

Here we have anticipated the result for $\langle 1/x \rangle$ from Eq. (71) in the next subsection, in conjunction with Eq. (70). The reason for $\langle M_n^2 \rangle \sim \beta m_Q$ is clear on the physical grounds: in the partonic approximation the final state is formed by the quark q with the momentum $m_Q/2$ and by the spectator antiquark $\overline{q_s p}$ with the momentum of order β.

The sum rule (59) [after subtracting the square of the second sum rule (58)] determines the dispersion

$$\langle M_n^4 \rangle - \langle M_n^2 \rangle^2 = \left\langle \frac{1}{x^2} \right\rangle - \left\langle \frac{1}{x} \right\rangle^2 \sim \langle M_n^2 \rangle^2. \qquad (63)$$

What about higher moments? It is not difficult to see that the next one, $\Sigma_{n=0}^{\infty} \Gamma_n (M_{H_Q}^2 - M_n^2)^2$, is a divergent sum because $\varphi_{H_Q}^2(x)/x^3$ would no longer be integrable. It defines the asymptotics of Γ_n at large n,

$$\Gamma_n \propto \frac{1}{M_n^6}. \qquad (64)$$

Let us recall [5] that $M_n^2=\pi^2\beta^2 n$ for high excitations, $n \gg 1$. We will see in the next subsection that the asymptotics (64) matches the contribution of the four-fermion operators in OPE.

C. Matching

Armed with exact sum rules (56), (58), (59) we are well prepared to verify a perfect match between the OPE result and the expression obtained by summing over hadronic final states. From the second sum rule (58) we have for the total semileptonic width Γ_{H_Q} the following relation:

$$\Gamma_{H_Q}=\frac{G^2}{4\pi}\cdot\frac{m_Q^2-m_q^2}{M_{H_Q}}\int_0^1\frac{dx}{x}\varphi_{H_Q}^2(x)-\sum_{M_n>M_{H_Q}}\Gamma_n. \quad (65)$$

The second term is actually positive [Γ_n defined by Eq. (53) is negative at $M_n>M_{H_Q}$]. Using Eq. (64) its size can be estimated

$$-\sum_{M_n>M_{H_Q}}\Gamma_n\propto\frac{1}{m_Q^4}. \quad (66)$$

Thus we have derived from the 't Hooft equation the following result for the inclusive width:

$$\Gamma_{H_Q}=\Gamma_Q\left[\frac{m_Q}{M_{H_Q}}\int_0^1\frac{dx}{x}\varphi_{H_Q}^2(x)+\mathcal{O}\left(\frac{1}{m_Q^5}\right)\right]. \quad (67)$$

This expression coincides with the OPE result of Eq. (48) as seen by rewriting the matrix element in Eq. (48) in terms of the ground state wave function $\varphi_{H_Q}(x)$,

$$\frac{\langle H_Q|\chi_Q^\dagger\frac{m_Q}{i\partial_-}\chi_Q|H_Q\rangle}{2M_{H_Q}}=\frac{m_Q}{M_{H_Q}}\int_0^1\frac{dx}{x}\varphi_{H_Q}^2(x), \quad (68)$$

where, besides normalization factors, we have used also the substitution $i\partial_-\to xM_{H_Q}/\sqrt{2}$.

This completes the proof of the perfect matching between OPE and the hadronic saturation through the order $1/m_Q^4$. Let us stress that the matrix element (68) given by the integral over the ground state wave function is implicitly m_Q dependent; its leading term is 1 followed by $1/m_Q^2$ and higher terms (see the discussion in Sec. III A). In the 't Hooft model one can, of course, evaluate the matrix element explicitly although it is not relevant to our main objective—probing the quark-hadron duality.

The absence of $1/m_Q$ corrections was demonstrated in Sec. III A by operator methods. Let us show now that the same statement can be derived from the 't Hooft equation as well. To this end we use the approach of Ref. [18] to the heavy quark limit, generalizing it to include $1/m_Q^2$ corrections. Instead of x, the appropriate variable for the large mass limit is

$$t=(1-x)m_Q, \quad \varphi(x)=\sqrt{m_Q}\phi(t). \quad (69)$$

An expansion of the 't Hooft equation in $1/m_Q$ (after a substitution of the new variable) and the virial theorem lead to the following relation for the meson mass:

$$\frac{M_{H_Q}^2}{m_Q^2}=1+2\frac{\langle t\rangle}{m_Q}-\frac{\beta^2}{m_Q^2}+\mathcal{O}\left(\frac{1}{m_Q^3}\right), \quad (70)$$

where averaging is over the H_Q meson wave function $\phi(t)$. As compared with Ref. [19] we have added a β^2/m_Q^2 term. Its origin is simple: it accounts for the renormalization of the heavy quark mass. Note that $\langle t\rangle\sim\beta$ provided $m_{sp}\lesssim\beta$.

We also need a similar expansion for the integral entering Eq. (67),

$$\int_0^1\frac{dx}{x}\varphi_{H_Q}^2(x)=1+\frac{\langle t\rangle}{m_Q}+\frac{\langle t^2\rangle}{m_Q^2}+\mathcal{O}\left(\frac{1}{m_Q^3}\right). \quad (71)$$

Substituting Eqs. (70), (71) in Eq. (67) we have

$$\frac{\Gamma_{H_Q}}{\Gamma_Q}=1+\frac{1}{2m_Q^2}(\beta^2-\langle t^2\rangle+\langle t\rangle^2)+\mathcal{O}\left(\frac{1}{m_Q^3}\right). \quad (72)$$

This expansion should be compared with the operator representation (20). The $1/m_Q$ expansion of $c_{\bar{Q}Q}\overline{Q}Q$ produces the same $1/m_Q^2$ term; the part $\propto\beta^2$ comes from the expansion of $c_{\bar{Q}Q}$; see Eq. (37).

Note that without the β^2 term the correction to 1 is negative, i.e.,

$$\frac{\Gamma_{H_Q}}{\Gamma_Q}-1<\frac{\beta^2}{2m_Q^2}. \quad (73)$$

One more comment about $1/m_Q^5$ terms. In the OPE approach they are due to the four-fermion operators generated by the graph in Fig. 6. Although the corresponding OPE coefficients are not calculated, the consideration above shows that the contribution of the four-fermion operators is dual to the sum of Γ_n with $M_n>M_{H_Q}$ for final state mesons, i.e., channels kinematically inaccessible in the decay.

V. VIOLATIONS OF DUALITY

A. Global and local duality

Having established a perfect match between the OPE prediction for the total width and the result of the saturation by exclusive decay modes, through $\mathcal{O}(1/m_Q^4)$, we must now turn to the issue of where the OPE-based prediction is supposed to fail. The failure usually goes under the name of "duality violations," a topic under intense scrutiny in the current literature. The definition of what duality violation is varies from publication to publication. Quite often, the researchers in the field stick to a vague notion of deviations "of certain rates for processes involving hadrons from the underlying partonic rates." This is, for instance, the convention of Ref. [10] where duality is understood as the coincidence with the parton-model prediction. If so, any nonperturbative contribution to the given rate would be interpreted as a ''duality violation,'' which does not make much sense to us.

We must precisely define what is meant by duality and its violations. Assume that a certain process is amenable to calculations within OPE. This means that an appropriate Euclidean quantity can be chosen, and the OPE series can be constructed. This series presents the quantity of interest as an expansion in an inverse large parameter, e.g., $1/Q^2$ or $1/E$. The very same quantity can be expressed as a dispersion integral over the imaginary part defined in Minkowski space. In e^+e^- annihilation the imaginary part coincides with $R(e^+e^-)$, in the transition amplitudes for heavy flavors the imaginary part reduces to semileptonic spectral densities, etc.

In order to treat the nonleptonic decays in the same vein one can introduce a spurion in the weak vertex, carrying a momentum q. In other words, let us substitute the weak Lagrangian by

$$\mathcal{L}_{\rm weak}(x)\to S(x)\mathcal{L}_{\rm weak}(x), \tag{74}$$

where $S(x)$ is a spurion field. Now let us consider the forward amplitude $\mathcal{A}(s)$ of the process

$$S(q)+H_Q(p)\to \text{light hadrons}\to S(q)+H_Q(p) \tag{75}$$

as a function of $s=(p+q)^2$. This variable s plays the same role as $s=q^2$ in e^+e^- annihilation. The total nonleptonic width is given by $\mathrm{Im}\,\mathcal{A}$ at $s=M_{H_Q}^2$. We are free to consider $\mathrm{Im}\,\mathcal{A}(s)$ in the complex s plane where it has two cuts: at $s>0$ and at $s<-2M_{H_Q}^2+2q^2$ (the second cut is due to u channel). Choosing a reference point s_0 far away from both cuts but closer to the first one we can express $\mathcal{A}(s_0)$ as a dispersion integral over the discontinuity across the cuts. On the other hand at the very same point s_0 one can apply the operator product expansion for calculating $\mathcal{A}(s_0)$ in terms of matrix elements of local operators, $\langle H_Q|\mathcal{O}_i|H_Q\rangle$. This gives sum rules which allow us to determine $\mathrm{Im}\,\mathcal{A}(s)$ at large s, in particular, at $s=M_{H_Q}^2$. This is fully analogous to what one does in e^+e^- annihilation for $R(s)$. In both cases smoothness is assumed (of course, in e^+e^- annihilation all positive values of s are accessible and one can check this assumption while in the case of nonleptonic decays M_{H_Q} is fixed). Needless to say that the total semileptonic widths can be treated along the same lines, the only difference is the presence of leptons in the intermediate states.

In the semileptonic decays OPE allows to predict, additionally, various distributions in the lepton momenta. This is, probably, the reason why it is usually claimed that the status of duality is more solid in the semileptonic decays. To show that it is not the case let us consider the semileptonic decays with the light quark in the final state. The OPE-based predictions for the spectral distributions are valid almost everywhere; they fail only in the end-point domain [15]. For this reason the total semileptonic widths cannot be obtained by integrating over the spectrum if we want a prediction which includes the linear in $1/m_Q$ corrections; this is why the argumentation in Ref. [15] was not sufficient to prove the absence of $1/m_Q$ corrections. Nevertheless, the statement of absence of such corrections in the total semileptonic widths can certainly be justified by the procedure described above for the nonleptonic widths. Thus, we see that the theoretical status of all these processes—e^+e^- annihilation, semileptonic and nonleptonic decays of heavy flavors—is basically the same.

By performing an appropriate expansion of the dispersion integral we obtain sum rules relating certain moments of the imaginary part of transition amplitude to matrix elements of consecutive terms in the OPE series constructed in the Euclidean domain. The predictions obtained in this way will be referred to as *global duality*. Taken at their face value, they are exact, to the extent we can calculate the coefficient functions and matrix elements of the operators involved in OPE. No additional assumptions are made. The predictions obtained in this way are consequences of fundamental QCD. Therefore, it does not make any sense to speak about violations of the global duality. One can only speak of the precision of calculation of the coefficient functions and determination of the matrix elements.

Unfortunately the term *global duality* is often used in a loose and ambiguous sense. It is applied indiscriminantly to integrals over the spectral densities with the weight functions chosen *ad hoc*. Our definition is narrower: it refers only to those specific integrals which emerge from the dispersion representation.

The notion of *local duality* on the other hand requires further assumptions. Assume that we want to predict imaginary parts (spectral densities) point by point, at large energies (or q^2). *If* one assumes that the spectral densities at the given energy are smooth, then from the moment integrals we can certainly predict the densities themselves. This amounts to an analytic continuation of the OPE series (truncated in a certain way), term by term from the Euclidean to Minkowski domain, with the subsequent calculation of the imaginary parts of each individual term in the series. The prediction obtained in this way is evidently a smooth function of the parameters. We then compare this prediction with the quantity measured in terms of hadronic contributions. The difference between the OPE-based smooth result and the experimental hadronic measurement is referred to as the *duality violation* meaning the violation of *local duality*.

By its nature the OPE results are series in powers of $\Lambda_{\rm QCD}/E$ and do not account for terms like $\exp[-(E/\Lambda_{\rm QCD})^k]$ (in the Euclidean domain). Although such terms are due to large distances, a signal of their appearance could show up in the short distance OPE series in the form of a factorial divergence of the series in higher dimensions. The situation is reminiscent of that in the perturbative expansion. The divergent α_s series (e.g., due to infrared renormalons) give rise to terms $\exp(-C/\alpha_s)$ although such terms can appear even in the absence of renormalons (for instance, as the quark condensate).

In other words, the OPE construction accounts properly for short distance singularities while the exponential terms are due to large distances being nonsingular at short distances. Thus, the duality violation is something we do not see in the (truncated) OPE series. The duality-violating terms are exponential in the Euclidean domain and oscillating (like $\sin[(E/\Lambda_{\rm QCD})^k]$) in the Minkowski domain.

From this standpoint there is no distinction between, say, the total e^+e^- annihilation cross section or the semileptonic rates of heavy flavors, on the one hand, and the nonleptonic rates of heavy flavors, on the other. Sometimes it is claimed that the former processes are "pure" while the latter are "impure;" it is even asserted that "duality follows from OPE in the first case while it has no theoretical justification in the second case." We assert that a duality violating exponential-oscillating component, associated with the neglected tails of OPE, is present in all processes, and the only physically meaningful question is its magnitude, as a function of large parameters (e.g., a momentum transfer or m_Q) and specific details of the process under consideration.

B. Oscillating terms in 't Hooft model

The appearance of duality violations in the form of oscillating terms is evident in the 't Hooft model where the spectral density is formed by zero-width discrete states. Indeed, each time a new decay channel opens $d\Gamma_{H_Q}/dm_Q$ experiences a jump (Γ_{H_Q} is continuous) , so that immediately above threshold $d\Gamma_{H_Q}/dm_Q$ is larger than the smooth OPE curve, in the middle between two successive thresholds it crosses the smooth prediction, and immediately below the next threshold $d\Gamma_{H_Q}/dm_Q$ is lower than the OPE-based expectation.

The amplitude of oscillations can be estimated as follows. Let us present the total width Γ_{H_Q} as

$$\Gamma_{H_Q}=\sum_{n=0}^{n=\infty}\Gamma_n-\sum_{M_n>M_{H_Q}}\Gamma_n. \tag{76}$$

Widths Γ_n are exclusive widths of two-body decays, $H_Q\to\phi+h_n$, where h_n is the nth excited $q\bar{q}_{sp}$ state with the mass M_n. For $M_n>M_{H_Q}$ widths Γ_n are not, of course, physical ones for H_Q decays but they are well defined. We have used this presentation in Sec. IV C where it was shown that the first term can be related to the wave function of H_Q state on one hand and to the matrix element of $\bar{Q}Q$ operator on the other. It is clear that the first term in Eq. (76) is a smooth function of m_Q and contains no nonanalytic terms we are going after; they are in the second one.

Thus we need to to know Γ_n for M_n in the vicinity of M_{H_Q}. For $M_n\gg M_{H_Q}$ we found in Sec. IV B that $\Gamma_n\propto M_n^{-6}$, see Eq. (64). To extrapolate to the vicinity of M_{H_Q} we account for the threshold factor, $\Gamma_n\propto(M_{H_Q}^2-M_n^2)/M_n^8$. The coefficient in this dependence can be fixed by duality of the second term in Eq. (76) to four-fermion operators $\mathcal{O}_{4q}$, discussed in Sec. IV C,

$$-\sum_{M_n>M_{H_Q}}\Gamma_n=c_{4q}\frac{\langle H_Q|\mathcal{O}_{4q}|H_Q\rangle}{2M_{H_Q}}. \tag{77}$$

We did not calculate coefficients c_{4q} but we found that up to numerical factor

$$c_{4q}\sim\frac{G^2}{2\pi}\frac{\beta^4}{m_Q^4}. \tag{78}$$

Assuming m_{sp}, $m_q\lesssim\beta$ we estimate $\langle H_Q|\mathcal{O}_{4q}|H_Q\rangle/2M_{H_Q}\sim\beta$. Then our model for Γ_n in the range $M_n\gtrsim M_{H_Q}$ is

$$\Gamma_n=3\pi G^2\beta^7\frac{M_{H_Q}^2-M_n^2}{M_n^8}. \tag{79}$$

The sum in Eq. (77) was approximated by the integral using $M_n^2=\pi^2\beta^2 n$.

Now we can evaluate the sum (77) with the better accuracy accounting for small nonanalytical terms. The result is

$$-\sum_{M_n>M_{H_Q}}\Gamma_n=\Delta\Gamma^{\rm OPE}+\Delta\Gamma^{\rm osc},$$

$$\Delta\Gamma^{\rm OPE}=\frac{G^2\beta}{2\pi}\left[\left(\frac{\beta}{M_{H_Q}}\right)^4+\frac{\pi^4}{2}\left(\frac{\beta}{M_{H_Q}}\right)^8\right],$$

$$\Delta\Gamma^{\rm osc}=\frac{3}{2}\pi^3G^2\beta\left(\frac{\beta}{M_{H_Q}}\right)^8\left[x(1-x)-\frac{1}{6}\right], \tag{80}$$

where

$$x=\text{fractional part of }\left(\frac{M_{H_Q}^2}{\pi^2\beta^2}\right),\quad x\in[0,1). \tag{81}$$

The smooth part $\Delta\Gamma^{\rm OPE}$ in Eq. (80) is given by same OPE term of Eq. (77) we discussed above plus higher in the power of $1/m_Q$ corrections . The part $\Delta\Gamma^{\rm osc}$, nonanalytic in $M_{H_Q}^2$, oscillates with period $\pi^2\beta^2$, see its plot in Fig. 7.

The amplitude of oscillation is

$$\left|\frac{\Delta\Gamma^{\rm osc}}{\Gamma_Q}\right|_{\max}\sim\frac{3\pi^4}{2}\left(\frac{\beta}{M_{H_Q}}\right)^9. \tag{82}$$

Note that the derivative $d(\Delta\Gamma^{\rm osc})/dm_Q$ contains discontinuities at thresholds, the amplitude of oscillations is larger for the derivative,

$$\left|\frac{d(\Delta\Gamma)^{\rm osc}/dm_Q}{d\Gamma_Q/dm_Q}\right|_{\max}\sim12\pi^2\left(\frac{\beta}{M_{H_Q}}\right)^7. \tag{83}$$

The oscillations under discussion cannot be produced by any truncated OPE series; they are not seen in the OPE. Thus the estimate (82) gives the actual scale of the expected duality violations in the problem at hand. Of course, this estimate is obtained within a specific model for Γ_n. The gross features of the result are independent of the model, however. They are determined only by the fact that the resonances have zero widths.

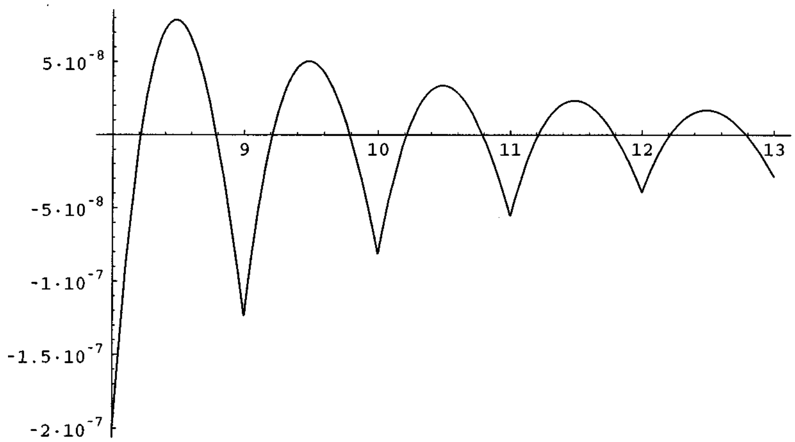

FIG. 7. Oscillations in Γ_{H_Q}. The ratio $\Delta\Gamma^{\rm osc}/G^2\beta$ is presented as a function of $M^2_{H_Q}/\pi^2\beta^2$.

It should be noted that the limit $N_c\to\infty$ presents a scenario which maximizes duality violations. In this limit the thresholds open ''abruptly,'' right at the position of the resonances, since the resonance widths vanish. In the real world of finite N_c the highly excited states have finite widths, and this effect, on its own, smears the hadron-saturated cross sections dynamically. If in the zero width approximation the oscillating duality violating component is suppressed only by powers of a large parameter ($1/m_Q$ in the case at hand), switching on finite widths will further suppress the oscillating component exponentially; see Sec. V D and Ref. [12] for further details.

C. Lessons

From the considerations above we conclude that (i) oscillating terms violating local duality are definitely present in the total decay rate (considered as a function of m_Q), (ii) amplitude of oscillations is $\sim\mathcal{O}(1/m_Q^9)$, i.e., strongly suppressed, and (iii) if we could average over m_Q in a sufficiently large interval the power suppression of the oscillations would turn into exponential suppression (in actual QCD, with $N_c=3$, the finite resonance widths do a similar job). Then it is perfectly legitimate to consider the OPE-based predictions beyond $1/m_Q^9$.

With this understanding in mind we now turn to the discussion of duality violations in actual four-dimensional QCD.

D. τ decays in 1+3 dimensions

Let us discuss a quantity of practical interest in 1+3 dimensions along similar lines, namely the normalized hadronic τ width R_τ:

$$R_\tau\equiv\frac{\Gamma(\tau^-\to\nu_\tau+\text{hadrons})}{\Gamma(\tau^-\to\nu_\tau e^-\overline{\nu}_e)}. \tag{84}$$

It can be expressed in terms of spectral densities ρ_V and ρ_A in the vector and axial-vector channels, respectively,

$$R_\tau=\int_0^{M_\tau^2}\frac{ds}{M_\tau^2}\left(1-\frac{s}{M_\tau^2}\right)^2\left(1+2\frac{s}{M_\tau^2}\right)[\rho_V(s)+\rho_A(s)]$$
$$=\frac{I_0(M_\tau^2)}{M_\tau^2}-3\frac{I_2(M_\tau^2)}{M_\tau^6}+2\frac{I_3(M_\tau^2)}{M_\tau^8}, \tag{85}$$

where the moments I_n are defined as

$$I_n(M)=\int_0^{M^2}ds s^n[\rho_V(s)+\rho_A(s)]. \tag{86}$$

While τ decays represent a simpler dynamical problem than the weak decays of heavy flavor hadrons we have to simplify it further still before we can arrive at some definite conclusions. To estimate the oscillating contribution to R_τ which constitutes duality violation that cannot be seen in a truncated OPE we consider the limiting cases of M_τ and N_c large. We will show that, adopting the resonance model motivated by two-dimensional QCD, for $N_c\to\infty$ and M_τ large, yet finite, the duality violation in R_τ scales as $1/M_\tau^6$; for $M_\tau\to\infty$ with N_c large, though finite, the oscillating term is suppressed exponentially.

Our consideration will be admittedly illustrative. One should not take literally the numbers we will obtain for many reasons: first of all the τ mass is not much larger than the spacing between the resonances; second, N_c is not large enough to warrant the zero width approximation. Still we believe that the consideration is instructive in a qualitative aspect.

For large N_c the spectrum of 1+3 QCD is expected to consist of an infinite comb of narrow resonances—in complete analogy to the 't Hooft model [20]. To keep the closest parallel to it we further assume that the high excitations in a given channel (e.g., the vector channel) are equally spaced in m^2. This agrees with the general expectation of a stringlike realization of confinement leading to asymptotically linear Regge trajectories. The masses of the excited states in, say, the ρ channel are then given[5] by $m_n^2=m_\rho^2+2n/\alpha'$ [21], with α' being the slope of the Regge trajectory (for a review see [22]). Experimentally one finds $2/\alpha'\simeq 2$ GeV^2. For large values of s the spectral densities for both the vector and axial-vector channels will approach the form:

[5]In other, less QCD-friendly scenarios, one obtains instead $m_n^2=m_\rho^2+n/\alpha'$. The distinctions between these two scenarios are irrelevant for our discussion.

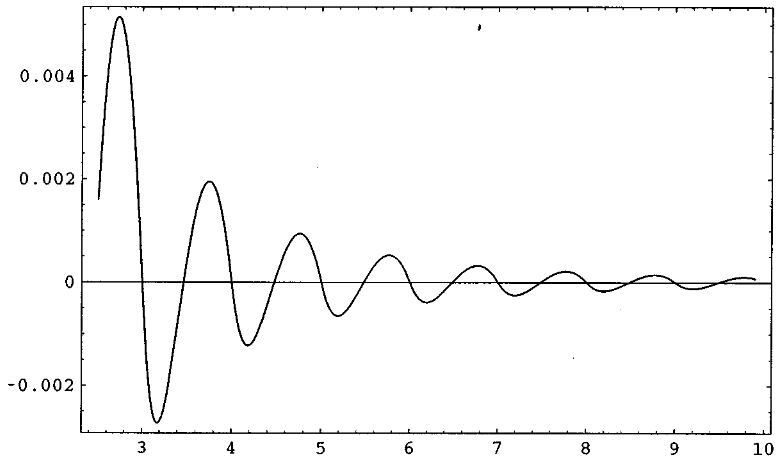

FIG. 8. Oscillations in R_τ. The plot of $\delta R^{\rm osc}/R_\tau^0$ is presented as a function of M_τ^2/σ^2.

$$\rho_V(s)=\rho_A(s)=N_c\cdot\sum_{n=1}^{\infty}\delta\left(\frac{s}{\sigma^2}-n\right);\quad \sigma^2=\frac{2}{\alpha'},\quad (87)$$

where a special notation σ^2 is introduced for $2/\alpha'$. Equation (87) is clearly not expected to hold at moderate and small values of s where the vector and axial-vector channels are drastically different and the resonances are not equidistant. However, details of the spectral densities at small s play no role in duality violation.

The contribution of any particular resonance of mass M_k to R_τ, according to Eq. (85), is given by a simple polynomial in $1/M_\tau^2$ [times the step function $\theta(M_\tau^2-M_k^2)$]. Therefore, variations of properties of resonances below a certain fixed mass μ change only the regular terms of the $1/M_\tau^2$ expansion, but have no impact on the oscillatory component. From Eq. (85) it is clear that such variations change only coefficients of the $1/M_\tau^2, 1/M_\tau^6$ and $1/M_\tau^8$ terms. It will be clear in what follows that the formal OPE for R_τ exactly reproduces these three expansion coefficients as well.

The spectral density in Eq. (87) is dual to the parton model result; i.e., it coincides with it after averaging over energy,

$$\langle\rho_{V,A}(s)\rangle=N_c\,. \quad (88)$$

Thus, the asymptotic prediction for R_τ at $M_\tau^2\to\infty$ is

$$R_\tau^0=N_c\,. \quad (89)$$

The sum over resonances in R_τ is easily calculated analytically: for the spectral density of Eq. (87) it is

$$R_\tau=R_\tau^{\rm OPE}+\delta R^{\rm osc},$$

$$\frac{R_\tau^{\rm OPE}}{N_c}=1-\frac{\sigma^2}{M_\tau^2}+\frac{1}{30}\left(\frac{\sigma^2}{M_\tau^2}\right)^4,$$

$$\frac{\delta R^{\rm osc}}{N_c}=-x(1-x)(1-2x)\left(\frac{\sigma^2}{M_\tau^2}\right)^3+\left[x^2(1-x)^2-\frac{1}{30}\right]\times\left(\frac{\sigma^2}{M_\tau^2}\right)^4, \quad (90)$$

where

$$x=\text{fractional part of }\left(\frac{M_\tau^2}{\sigma^2}\right),\quad x\in[0,1).$$

We presented the result as a sum of two functions of M_τ^2, the first one, $R_\tau^{\rm OPE}$, is a smooth function expandable in $1/M_\tau^2$. The second one, $\delta R^{\rm osc}$, oscillates with the period σ^2; its average vanishes; see the plot of $\delta R^{\rm osc}/R_\tau^0$ in Fig. 8.

Let us show now that $R_\tau^{\rm OPE}$ coincides with the OPE prediction in the model. Power corrections can be presented in the following way:

$$R_\tau^{\rm OPE}=N_c+\frac{\tilde I_0}{M_\tau^2}-3\frac{\tilde I_2}{M_\tau^6}+2\frac{\tilde I_3}{M_\tau^8}, \quad (91)$$

where the ‘‘condensates’’ $\tilde I_n$ are

$$\tilde I_n=\int_0^\infty ds\, s^n[\rho_V(s)+\rho_A(s)-2N_c]. \quad (92)$$

This integral representations for the ‘‘condensates’’ $\tilde I_n$ follows from Eqs. (85), (86) if one assumes that the spectral densities approach their asymptotic limits faster than any power of $1/s$. In the model at hand, with the comblike spectral density, the integral representation (92) requires regularization. As a regularization one can introduce the weight factor $\exp(-\epsilon s)$, taking the limit $\epsilon\to 0$ at the end. With this regularization, $R_\tau^{\rm OPE}$ from Eq. (90) is reproduced.

To justify the procedure it is instructive to make one step back and follow more literally the original OPE procedure. Namely, our primary object of interest is the Euclidean polarization operator

$$\Pi(Q^2)=\frac{1}{\pi}\int ds\frac{\rho_V(s)+\rho_A(s)}{s+Q^2}. \tag{93}$$

The original meaning of $\widetilde{I}_n$ is the coefficients in the asymptotic expansion of $\Pi(Q^2)$ at $Q^2\to\infty$:

$$\pi\Pi(Q^2)=-2N_c\ln Q^2+\text{const}+\sum_{n=0}^{\infty}(-1)^n\frac{\widetilde{I}_n}{(Q^2)^{n+1}}. \tag{94}$$

The comb of the model of Eq. (87) was considered in Ref. [3],

$$\pi\Pi(Q^2)=2N_c\sum_{k=1}^{\infty}\frac{1}{(Q^2/\sigma^2)+n}+\text{const}$$
$$=-2N_c\left[\psi\left(\frac{Q^2}{\sigma^2}\right)+\frac{\sigma^2}{Q^2}\right]+\text{const}, \tag{95}$$

where ψ is Euler's ψ function. The asymptotic expansion of $\Pi(Q^2)$ in the model takes the form

$$\pi\Pi(Q^2)=-2N_c\left[\ln\frac{Q^2}{\sigma^2}+\frac{1}{2}\frac{\sigma^2}{Q^2}-\sum_{n=1}^{\infty}\frac{B_{2n}}{2n}\left(\frac{\sigma^2}{Q^2}\right)^{2n}\right]+\text{const}, \tag{96}$$

where B_n stand for the Bernoulli numbers. Taking the coefficients of $1/Q^2, 1/Q^6$ and $1/Q^8$ terms from this equation we find the consistency with R_τ^{OPE} of Eq. (90). (The term $1/Q^6$ is absent and $1/Q^8$ is defined by $B_4=-1/30$.)

Note that the spectral density (87) is not literally "QCD-compatible:" the corresponding $\Pi(Q^2)$ has the $1/Q^2$ nonperturbative correction forbidden in QCD [4]. This can be easily cured by adding the resonance with $n=0$ with a half weight, which just would amount to adding $N_c\sigma^2/Q^2$ to $\pi\Pi(Q^2)$. Since this obviously does not change δR^{osc} at any value of M_τ, this is inessential for us.

Let us discuss now the duality-violating $\delta R^{\text{osc}}(M_\tau^2)$ (see Fig. 8). Its dominant component scales as $1/M_\tau^6$. It is intriguing to note that the very same scaling law was obtained in Ref. [23] from totally different considerations invoking instantons. This oscillating component vanishes at M_τ corresponding to the new thresholds, and at one point in the middle between the successive resonances; only the second derivative of R_τ has a jump at the thresholds,

$$\frac{1}{R_\tau}\left(\frac{d}{dM_\tau^2}\right)^2 R_\tau\Bigg|_{x=1}^{x=0}=6\frac{\sigma^2}{M_\tau^6}. \tag{97}$$

This is the consequence of the threshold factor $(1-s/M_\tau^2)^2$ in Eq. (85). One power in it is just the two-body phase space factor $|\vec{p}|/M_\tau$. In $1+1$ dimensions one would have instead $1/|\vec{p}|$, the threshold singularity, and the duality-violating component would be enhanced correspondingly.

The average of δR^{osc} vanishes while the amplitude of oscillations amounts to

$$\left|\frac{\delta R^{\text{osc}}}{R_\tau}\right|_{\max}=\frac{1}{3\sqrt{12}}\left(\frac{\sigma^2}{M_\tau^2}\right)^3. \tag{98}$$

It is interesting that, in spite of the fact that $M_\tau^6\delta R^{\text{osc}}$ is given by a polynomial between the thresholds, the whole function is well approximated by

$$-\frac{1}{3\sqrt{12}}\left(\frac{\sigma^2}{M_\tau^2}\right)^3\sin\left(2\pi\frac{M_\tau^2}{\sigma^2}\right). \tag{99}$$

Note that the numerical coefficient in Eq. (99) is rather small, compared, e.g., to Eq. (97). This suppression is related to the fact that the characteristic scale is $\sigma^2/2\pi$ rather then σ^2.

Taking our estimate of the oscillation amplitude at its face value and using the actual value of the τ mass in Eq. (98) we find $\delta R^{\text{osc}}/R_\tau\sim 3\%$. It is clear that this is a very crude estimate given the fact that in the actual τ decay $\sigma^2/M_\tau^2\sim 2/3$ and we deal with one oscillation at most.

In the real world with $N_c=3$ we expect a further suppression of deviations from duality due to the nonvanishing widths of the resonances naturally smearing out the amplitude of the oscillation. A rough estimate of this effect can be given in close analogy to Ref. [12]. Let us introduce a dimensionless constant B representing the width-to-mass ratio:

$$\frac{\Gamma_n}{m_n}=\frac{B}{N_c}(1+\mathcal{O}(1/N_c)); \tag{100}$$

i.e., B stays finite for large N_c. One actually guesses to estimate $B\sim 0.5$. Then we infer (see [12] for details)

$$\frac{\Delta R_\tau}{R_\tau^0}=\frac{1}{3\sqrt{12}}\left(\frac{\sigma^2}{M_\tau^2}\right)^3\exp\left(-\frac{2\pi BM_\tau^2}{N_c\sigma^2}\right). \tag{101}$$

The power-suppressed oscillations eventually turn into exponentially suppressed, although at a larger energy.

VI. NONVANISHING m_ψ AND COMMENTS ON THE LITERATURE

The work [10] stimulated our interest in the 't Hooft model as a laboratory for exploring heavy quark expansions in inclusive decays, and the implementation of duality. The authors of Ref. [10] compared the decay width of a heavy flavor hadron in the parton approximation with the result obtained by summing over all exclusive transition rates for $H_Q\to h_i h_j$. A systematic excess of the total width Γ_{H_Q} over its parton value Γ_Q was observed and was fitted to be

$$\frac{\Gamma_{H_Q}-\Gamma_Q}{\Gamma_Q}\sim\frac{0.15}{m_Q}. \tag{102}$$

The authors interpreted this access as a violation of duality. According to our understanding (see Sec. V A) it should rather be called breaking of OPE.

Our analytical treatment does not support such a conclusion. In particular, we have proved the absence of linear in $1/m_Q$ corrections not only by the OPE method but by the direct analysis of the 't Hooft equation (see Sec. IV). Although we tried to closely follow the analysis of Ref. [10], still there is a difference of kinematical nature. We considered the ψ fields (leptons or quarks) to be massless, while $m_\psi = m_q \neq 0$ in Ref. [10]. The choice of $m_\psi = 0$ allows us to limit ourselves to the point $q^2=0$ where great simplifications occur, in particular, only the massless pion state is produced in the $\psi\bar{\psi}$ channel. Analytic solution of the problem turns out to be possible.

We have checked that the analytical expression for the triple overlap integral of Ref. [10] reduces at $q^2=0$ to a simple overlap integral (55) of two wave functions. The simple structure of Eq. (55) combined with the completeness of the wave functions is sufficient to prove a perfect match between OPE and hadronic saturation, at the level of $1/m_Q^4$.

To make full contact between our results and those of Ref. [10] we must now consider the impact of $m_\psi \neq 0$. Needless to say that at $m_\psi \neq 0$ the parton result for Γ_Q changes,

$$\Gamma_Q(m_\psi \neq 0) = \Gamma_Q(m_\psi = 0)\left[1 - \frac{2m_\psi^2}{m_Q^2} + \mathcal{O}\left(\frac{m_\psi^4}{m_Q^4}\right)\right]. \tag{103}$$

This effect was certainly included in the analysis of Ref. [10]. It is also accounted for in OPE as a change of the coefficient of the operator $\bar{Q}Q$.

The leading effect due to $m_\psi \neq 0$ is linear in m_ψ, however. Thus, one can wonder whether it produces $1/m_Q$ corrections. In the next subsection we will show that the linear in m_ψ corrections to the total width are suppressed as $1/m_Q^3$. Moreover, this is a leading effect which produces a distinction between semileptonic and nonleptonic total widths; all other effects which differentiate them are suppressed as $1/m_Q^4$. It is clear then that the $1/m_Q$ violation of duality claimed in Ref. [10] if it would be present in the nonleptonic width must have been present in the semileptonic width as well.

Corrections $1/m_Q^3$ come also from quadratic in m_ψ terms in OPE. As it was mentioned above these corrections do not differentiate semileptonic and nonleptonic widths. We estimate them in Sec. VI B. We also estimate in Sec. VI C effects of novanishing m_ψ for violations of local duality. Overall, we conclude that the OPE approach shows that the nonvanishing m_ψ results in $1/m_Q^3$ corrections which are numerically small and cannot explain the alleged deviation (102).

On the hadronic saturation side we have checked that the triple overlap integral of Ref. [10] is expandable in m_ψ, and the leading correction is quadratic in $m_{\psi,q}$. Assuming that $m_\psi \lesssim \beta$ and performing an *analytic* summation of the widths using the expressions of Ref. [10] for the amplitudes in conjunction with the sum rules for the hadronic polarization tensor $\mathrm{Im}\,\Pi(q^2)$ similar to those in Eqs. (56), (58), (59), we obtained exactly the same m_ψ^2/m_Q^2 correction as in Eq. (103). This holds for the inclusive width smeared over a small interval of masses as discussed in Sec. VI C. Moreover, we checked the exact matching with the OPE at the level $1/m_Q^3$.

A. Linear in m_ψ corrections

All effects due to $m_\psi \neq 0$ reside in $\Pi_{\mu\nu}$; see Eq. (7). The linear in m_ψ part in OPE for $\Pi_{\mu\nu}$ is

$$\Delta\Pi_{\mu\nu}(q) = \frac{4m_\psi}{q^2}\langle 0|\bar{\psi}\psi|0\rangle\left(\frac{q_\mu q_\nu}{q^2} - g_{\mu\nu}\right). \tag{104}$$

It is easy to see that this effect is nothing but a shift of the pion ($\phi = [\psi\bar{\psi}]_0$) mass from zero,

$$\mu_\phi^2 = -\frac{4\pi}{N_c} m_\psi \langle 0|\bar{\psi}\psi|0\rangle. \tag{105}$$

Indeed, comparing Eq. (9) and Eq. (104) we see that the latter is the first order term in μ_ϕ^2 expansion of the pion pole $1/(q^2-\mu_\phi^2)$ in $\Pi_{\mu\nu} = -(1/\pi)(q_\mu q_\nu - q^2 g_{\mu\nu})/(q^2-\mu_\phi^2)$. Thus, in order to take into account linear in m_ψ effects in the transition operator all one needs to do is to calculate the decay $Q \to \phi + q$ with the nonvanishing pion mass. The result is

$$\frac{\Delta\Gamma_{H_Q}^{(1)}}{\Gamma_Q} = \mu_\phi^2 \frac{2m_Q m_q}{(m_Q^2 - m_q^2)^2} = -\frac{8\pi}{N_c} m_\psi \langle 0|\bar{\psi}\psi|0\rangle \frac{m_Q m_q}{(m_Q^2-m_q^2)^2}. \tag{106}$$

At large m_Q it falls off as $1/m_Q^3$. An extra suppression m_q/m_Q is specific for two dimensions.

As for the numerical value of the correction (106) we will substitute $\langle \bar{\psi}\psi \rangle = -N_c\beta/\sqrt{12}$ in the chiral limit [11,19]. For the value of $m_\psi = m_q = 0.56\beta$ adopted in Ref. [10] we get

$$\frac{\Delta\Gamma_{H_Q}^{(1)}}{\Gamma_Q} = 2.27\left(\frac{\beta}{m_Q}\right)^3. \tag{107}$$

B. Quadratic in m_ψ corrections

The analysis of the previous subsection refers to the operator $\bar{Q}Q\bar{\psi}\psi$ in the OPE for the total width. Another operator generated due to $m_\psi \neq 0$ is the four-fermion operator of the type $\bar{Q}Q\bar{q}q$. It appears from the graph of Fig. 3 with lepton lines substituted by ψ lines. Calculation of this graph is a simple exercise. The result for the corresponding correction to the width can be presented in the following form:

$$\Delta\Gamma_{H_Q}^{(2)} = N_c \frac{G^2 m_\psi^2}{m_Q^2} \frac{\langle H_Q|\bar{Q}\gamma_5 q \bar{q}\gamma_5 Q|H_Q\rangle}{2M_{H_Q}}. \tag{108}$$

Up to the factor N_c the same contribution appears in the semileptonic width. In difference with the operator $\bar{Q}Q\bar{\psi}\psi$

the large N_c limit does not allow us to factorize the matrix element of Eq. (108). We will use the factorization to get an estimate

$$\langle H_Q|\bar{Q}\gamma_5 q\bar{q}\gamma_5 Q|H_Q\rangle \sim -\frac{1}{2N_c}\langle 0|\bar{q}q|0\rangle\langle H_Q|\bar{Q}Q|H_Q\rangle. \tag{109}$$

As in the previous subsection we use the chiral limit value $\langle\bar{q}q\rangle = -N_c\beta/\sqrt{12}$ of the VEV [11,19] and $m_\psi = m_q = 0.56\beta$ for the numerical estimate

$$\frac{\Delta\Gamma^{(2)}_{H_Q}}{\Gamma_Q} \sim 0.57\left(\frac{\beta}{m_Q}\right)^3. \tag{110}$$

C. Local duality violations at $m_\psi \neq 0$

Here we discuss the impact of new thresholds opening in the spectral density of the vector $\bar{\psi}\gamma_\mu\psi$ currents, as m_Q increases. At $m_\psi = 0$ the duality in the $\langle\bar{\psi}\gamma_\mu\psi,\bar{\psi}\gamma_\nu\psi\rangle$ correlation function is perfect, since, due to bosonization, the theoretical expression for this correlator reduces to exactly one massless state propagation with a known coupling constant. If $m_\psi \neq 0$ then higher $[\psi\bar{\psi}]$ mesonic states appear in the imaginary part of $\langle\bar{\psi}\gamma_\mu\psi,\bar{\psi}\gamma_\nu\psi\rangle$ with residues proportional to m_ψ^2.

To calculate the dependence of the total width Γ_{H_Q} on m_Q near thresholds, i.e., near

$$M_{H_Q} = M_n^{[\psi\bar{\psi}]} + M_k^{[q\bar{q}_{sp}]}, \tag{111}$$

we need to find the exclusive width $\Gamma_{nk}(H_Q \to [\psi\bar{\psi}]_n + [q\bar{q}_{sp}]_k)$. We did it in Sec. V B for $m_\psi = 0$ when only the $n=0$ massless ϕ state is produced in the $[\psi\bar{\psi}]$ channel.

Now, the highly excited states can appear in this channel. They are pseudoscalar ones corresponding to even values of n (scalar states are not produced by the conserved current $\bar{\psi}\gamma^\mu\psi$). Moreover, with nonvanishing $M_n^{[\psi\bar{\psi}]}$ the amplitude of transition to scalar $[q\bar{q}_{sp}]$ states (odd k) is not proportional to $|\vec{p}|$. Thus, near the threshold (111) the exclusive width of decay into the pair of the pseudoscalar $[\psi\bar{\psi}]$ and the scalar $[q\bar{q}_{sp}]$ is singular because the factor $1/|\vec{p}|$ in the phase space that explode at thresholds. Therefore, exactly at threshold, Γ_{nk} is infinite. It goes without saying that it cannot coincide with the smooth OPE prediction near the threshold. These spikes are clearly visible on the plots of Ref. [10].

To maximize the exclusive width Γ_{nk} for the decay $H_q \to [\psi\bar{\psi}]_n + [q\bar{q}_{sp}]_k$ near the threshold (111) we choose the range of $M_n^{[\psi\bar{\psi}]}$ and $M_k^{[q\bar{q}_{sp}]}$ close to the corresponding partonic configuration. In terms of partons we deal with the decay $Q \to \psi + \bar{\psi} + q$. If we fix the square of the effective mass of $\psi\bar{\psi}$ pair $M^2_{\psi\bar{\psi}}$ then the momentum of the light q is $(m_Q^2 - M^2_{\psi\bar{\psi}})/(2m_Q)$. For hadrons $M_n^{[\psi\bar{\psi}]}$ close to $M^2_{\psi\bar{\psi}}$ and the characteristic hadronic mass squared in the $q\bar{q}_{sp}$ system is $M^2_{[q\bar{q}_{sp}]} \sim \beta(m_Q^2 - M^2_{\psi\bar{\psi}})/m_Q$. Then, it is simple to check that the threshold range is

$$M_{H_Q} - M_n^{[\psi\bar{\psi}]} \sim \beta, \quad M_k^{[q\bar{q}_{sp}]} \sim \beta. \tag{112}$$

Thus, it is just a few first scalar excitations in the $q\bar{q}_{sp}$ system that are relevant. If $M_n^{[\psi\bar{\psi}]}$ and $M_k^{[q\bar{q}_{sp}]}$ fall outside the kinematics indicated above then the exclusive width Γ_{nk} is additionally suppressed.

Let us recall that the bulk contribution into the total width is provided by the ground state in the $\psi\bar{\psi}$ channel and $M_k^{[q\bar{q}_{sp}]} \lesssim \sqrt{m_Q\beta}$. What is under discussion is a small tail of highly excited $\psi\bar{\psi}$ states which determines the oscillating component.

The relevant matrix element is

$$\begin{aligned}\langle[\psi\bar{\psi}]_n[q\bar{q}_{sp}]_k|(\bar{q}\gamma_\mu Q)(\bar{\psi}\gamma^\mu\psi)|H_Q\rangle \\ = \langle[\psi\bar{\psi}]_n|\bar{\psi}\gamma^\mu\psi|0\rangle\langle[q\bar{q}_{sp}]_k|\bar{q}\gamma_\mu Q|H_Q\rangle.\end{aligned} \tag{113}$$

It is easy to see that the residue $\langle[\psi\bar{\psi}]_n|\bar{\psi}\gamma^\mu\psi|0\rangle$ scales as $\sqrt{N_c}m_\psi\beta/M_n^{[\psi\bar{\psi}]}$, while the transition amplitude $\langle[q\bar{q}_{sp}]_k|\bar{q}\gamma_\mu Q|H_Q\rangle$ scales as $\sqrt{m_Q\beta}$. Assembling all factors together we find that near the threshold

$$\Gamma_{nk}(H_Q \to [\psi\bar{\psi}]_n + [q\bar{q}_{sp}]_k) \sim G^2 N_c \frac{m_\psi^2\beta^3}{m_Q^3|\vec{p}|} \sim \Gamma_Q \frac{m_\psi^2\beta^3}{m_Q^4|\vec{p}|}. \tag{114}$$

This equation describes both the singularity at thresholds and deviation from local duality between the thresholds. The spatial momentum $|\vec{p}|$ is

$$|\vec{p}| \approx \sqrt{2M_k^{[q\bar{q}_{sp}]}(M_{H_Q} - M_n^{[\psi\bar{\psi}]} - M_k^{[q\bar{q}_{sp}]})}. \tag{115}$$

In the middle between the thresholds in $M_n^{\psi\bar{\psi}}$ the value of $|\vec{p}| \sim (\beta^3/m_Q)^{1/2}$, and

$$\Gamma_{nk}(H_Q \to [\psi\bar{\psi}]_n + [q\bar{q}_{sp}]_k) \sim \Gamma_Q \frac{m_\psi^2\beta^{3/2}}{m_Q^{7/2}}. \tag{116}$$

This estimate gives the amplitude of the oscillating component. It is applicable also to the threshold spikes provided these spikes are averaged over the intervals of M_{H_Q} less but comparable with the period of oscillations in M_{H_Q} dependence ($\sim \pi^2\beta^2/M_{H_Q}$).

Therefore, we conclude that the violation of local duality dies off as $1/m_Q^{7/2}$. This effect is by far the largest duality violating contribution. The occurrence of a relatively weak suppression is due to (a) the zero resonance width approximation and (b) the singular nature of the two-body phase space in two dimensions. Both features have no parallel in actual QCD.

In summary, we identified two leading effects that are responsible for deviations from the parton formula—one is associated with the four-fermion operator in OPE, appearing due to $m_\psi \neq 0$, the second is the additional duality violating component that was absent at $m_\psi = 0$. The first, inclusive one, dies off as $1/m_Q^3$; the second (exclusive) at least as $1/m_Q^{7/2}$. We do not see any room for $1/m_Q$ deviations, even oscillating.

VII. DISCUSSION AND CONCLUSIONS

The situation we encounter in the 't Hooft model is very instructive. The model is readily treatable, which allows one to advance quite far in constructing the OPE series. It is super-renormalizable, thus providing an especially clean environment for testing various subtle aspects of OPE. The perturbative series for the coefficient functions in the large N_c limit converges. We find, with satisfaction, that all general statements regarding OPE are fully confirmed.

The model also clearly exhibits the breaking of local duality by oscillating terms. These oscillations are related to the exponential terms in the Euclidian domain and not seen in OPE. Because of zero meson widths in the large N_c limit they are suppressed only by powers of $1/m_Q$ which we have determined.

We note that in the t' Hooft model the local duality of the OPE predictions in the inclusive heavy flavor decays (both semileptonic and nonleptonic) holds much better than in $R(e^+e^-)$, the generally recognized classical laboratory for applications of OPE. It is in contrast to the opinion often expressed in the literature that OPE is not applicable in the inclusive heavy quark widths. Moreover, the numerical computations of Ref. [10] suggest that OPE width approaches the (smeared) hadronic ones at a few percent level very soon, right after a few first channels are open.

In actual QCD, already the first excited states are broad enough and inconspicuous, leave alone high excitations. When a finite resonance width is introduced, it immediately leads to dynamical smearing of the spectral densities, ensuring an exponential suppression of the oscillating duality violating terms (see Sec. V D and Ref. [12]). Thus, in terms of actual QCD we are still very far from the solution of this extremely important problem; the exercise performed gives us some kind of an upper bound.

As previously, in actual QCD, we have to rely on models while estimating the exponential (oscillating) terms not seen in OPE. The choice is not large—only two models were suggested previously. One of them is an instanton-based model [23], another is a resonance-based model [12], close in spirit to estimates we have presented above. The instanton-based model is simple and predictive, but it apparently lacks the sophistication inherent to the phenomenon in actuality. In particular, it predicts an oscillating component $\sim \sin m_Q$, rather than $\sim \sin m_Q^2$ as would be natural from the resonance point of view. Thus, the 't Hooft model teaches us that the instanton-based estimates cannot be fully true. On the other hand, the resonance-based model, which works satisfactorily in the limit of infinitely narrow resonances, does not give a full answer as to how strongly the oscillating component is suppressed when the finite resonance widths are switched on. It is clear that further steps in developing the existing or engineering new models are needed.

This work presents the first estimate of the duality violations, from the resonance-related considerations based on a $1/N_c$ expansion, in the practically important problem of the hadronic τ decays. Although not fully conclusive, the results are very encouraging, and call for expansion of these ideas in other processes. This is an obvious task for the future.

ACKNOWLEDGMENTS

We are grateful to R. Jaffe, A. Kaidalov, and M. Voloshin for useful discussions. This work was supported in part by the DOE under the grant number DE-FG02-94ER40823, by the NSF under the grant number PHY96-05080, and by the RFFI under the grant number 96-15-96764.

[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 (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. [4].

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

[3] M. Shifman, in *Proceedings of the Workshop on Continuous Advances in QCD*, edited by A. Smilga (World Scientific, Singapore, 1994), p. 249 (hep-ph/9405246); in *Particles, Strings and Cosmology*, Proc. PASCOS-95, edited by J. Bagger *et al.* (World Scientific, Singapore, 1996), p. 69 (hep-ph/9505289); see also Ref. [23].

[4] 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).

[5] G. 't Hooft, Nucl. Phys. **B75**, 461 (1974) [reprinted in G. 't Hooft, *Under the Spell of the Gauge Principle* (World Scientific, Singapore, 1994), p. 443]; see also Ref. [7] for a discussion of subtleties in the regularization of the 't Hooft model.

[6] C. Callan, N. Coote, and D. Gross, Phys. Rev. D **13**, 1649 (1976); M. Einhorn, *ibid.* **14**, 3451 (1976); M. Einhorn, S. Nussinov, and E. Rabinovici, *ibid.* **15**, 2282 (1977); I. Bars and M. Green, *ibid.* **17**, 537 (1978).

[7] F. Lenz, M. Thies, S. Levit, and K. Yazaki, Ann. Phys. (N.Y.) **208**, 1 (1991).

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

[9] B. Grinstein and P. Mende, Nucl. Phys. **B425**, 451 (1994).

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

[11] A. Zhitnitsky, Phys. Lett. **165B**, 405 (1985); Phys. Rev. D **53**, 5821 (1996).
[12] B. Blok, M. Shifman, and Da-Xin Zhang, Phys. Rev. D **57**, 2691 (1998).
[13] M. Voloshin and M. Shifman, Yad. Fiz. **41**, 187 (1985) [Sov. J. Nucl. Phys. **41**, 120 (1985)]; Zh. Éksp. Teor. Fiz. **91**, 1180 (1986) [Sov. Phys. JETP **64**, 698 (1986)].
[14] E. Eichten and B. Hill, Phys. Lett. B **234**, 511 (1990); H. Georgi, *ibid.* **240**, 447 (1990).
[15] J. Chay, H. Georgi, and B. Grinstein, Phys. Lett. B **247**, 399 (1990).
[16] I. Bigi, N. Uraltsev, and A. Vainshtein, Phys. Lett. B **293**, 430 (1992); **297**, 477(E) (1993).
[17] V. Novikov, M. Shifman, A. Vainshtein, and V. Zakharov, Fortschr. Phys. **32**, 585 (1985).
[18] M. Burkardt, Phys. Rev. D **46**, R2751 (1992).
[19] M. Burkardt, in *Proceedings of Quantum Infrared Physics 1994* (Paris, France, 1994), p. 438 (hep-ph/9409333).
[20] G. 't Hooft, Nucl. Phys. **B72**, 461 (1974); E. Witten, *ibid.* **B149**, 285 (1979).
[21] A. Dubin, A. Kaidalov, and Yu. A. Simonov, Phys. Lett. B **323**, 41 (1994).
[22] A. Kaidalov, Sov. Phys. Usp. **14**, 600 (1972); P. Collins, *An Introduction to Regge Theory and High Energy Physics* (Cambridge University Press, Cambridge, England, 1977).
[23] B. Chibisov, R. Dikeman, M. Shifman, and N. Uraltsev, Int. J. Mod. Phys. A **12**, 2075 (1997).