Theoretical Physics Institute
University of Minnesota

TPI-MINN-94/12-T
UMN-TH-1250-94
CERN-TH.7250/94
UND-HEP-94-BIG 05

Sum Rules for Heavy Flavor Transitions in the SV Limit

I. Bigi [a,b], M. Shifman [c], N.G. Uraltsev [a,c,d], A. Vainshtein [c,e]

[a] *TH Division, CERN, CH-1211 Geneva 23, Switzerland*[*]
[b] *Dept.of Physics, Univ. of Notre Dame du Lac, Notre Dame, IN 46556, U.S.A.*[†]
[c] *Theoretical Physics Institute, Univ. of Minnesota, Minneapolis, MN 55455*
[d] *St.Petersburg Nuclear Physics Institute, Gatchina, St.Petersburg 188350, Russia*[†]
[e] *Budker Institute of Nuclear Physics, Novosibirsk 630090, Russia*

e-mail addresses:
bigi@cernvm.cern.ch, shifman@vx.cis.umn.edu, vainshte@vx.cis.umn.edu

Abstract

We show how previously obtained sum rules for the weak decays of heavy flavor hadrons can be derived as the moments of spectral distributions in the small velocity (SV) limit. This systematic approach allows to determine corrections to these sum rules, to obtain new sum rules and it provides us with a transparent physical interpretation; it also opens a new perspective on the notion of the heavy quark mass. Applying these sum rules we derive a *lower* bound on the deviation of the exclusive form factor $F_{B\to D^*}$ from unity at zero recoil; likewise we give a field-theoretical derivation of a previously formulated inequality between the expectation value for the kinetic energy operator of the heavy quark and for the chromomagnetic operator. We analyze how the known results on nonperturbative corrections must be understood when one takes into account the renormalization point dependence of the low energy parameters.

HEP-PH-9405410

[*] During the academic year 1993/94
[†] Permanent address

1 Introduction

The theory of preasymptotic effects in inclusive weak decays of heavy flavor hadrons has been developed from the early eighties on [1] – [8] and is entering now a rather mature stage [9] – [18]. In recent papers [14, 15, 16] it was in particular shown how the effects due to the motion of the heavy quarks inside the hadrons can be incorporated in a systematic way, namely through distribution functions which crucially depend on the ratio $\gamma = m_q/m_Q$ where m_Q and m_q are the masses of the initial and final state quarks, respectively. Among other things, it was mentioned that the formalism developed in Ref. [14] automatically ensures the Bjorken sum rule [19, 20, 21] in the small velocity (SV) limit [22] (Sect. 4 of Ref. [14]). In the present work we discuss in more detail the sum rules of Refs. [22], [19] – [20], as well as the so-called optical sum rules derived later by Voloshin [23]. Quantum mechanical approach to sum rules has been considered by Lipkin [24], the so-called Mössbauer sum rules.

These are the three main observations of our paper:

• We will demonstrate that these apparently isolated sum rules represent merely moments of observable spectral distributions. Their physical meaning becomes absolutely transparent within the formulation of the problem suggested in Ref. [14]. Actually many crucial elements are already included, implicitly and explicitly, in Ref. [14], so our task is to combine them. This approach allows one to get new sum rules and to obtain corrections to the old ones in a systematic and comprehensive way.

• The analysis of the sum rules will also give us an opportunity to discuss the notion of the heavy quark mass from the point of view complementary to recent investigations [25, 26]. It will be shown that the key theoretical parameter $\overline{\Lambda}(\mu)$ is directly related to quantities measurable in inclusive semileptonic decays of B mesons in a certain kinematical regime. The relation obtained makes absolutely explicit the fact that $\overline{\Lambda}$ does *not* exist as a universal constant, as had been previously believed. Any consistent treatment must deal with a μ dependent function, $\overline{\Lambda}(\mu)$, where μ is a normalization point.

• One can use the sum rules to derive a lower bound on the deviation of the exclusive form factor $F_{B\to D^*}$ from unity at zero recoil. The lower bound is essentially determined by the average value of the chromomagnetic operator μ_G^2, familiar from the previous studies. We also obtain a new field-theoretic derivation of the previously formulated inequality $\mu_\pi^2 > \mu_G^2$ (μ_π^2 is the average value of the kinetic energy operator) whose validity was questioned by some authors through the suggestion that the original line of reasoning was purely quantum-mechanical.

The organization of the paper is as follows. In Sect. 2 all basic elements of our approach are demonstrated in the simplified model where the heavy quarks are deprived of their spins and the external "weak" current considered is scalar. Sect. 3 is devoted to real QCD; here we find a lower bound on $1 - F^2_{B\to D^*}$ at zero recoil and present expressions for $\overline{\Lambda}(\mu)$ in terms of the differential distributions observable in semileptonic B meson decays. Sect. 4 addresses practical implications of the "running" of basic low energy parameters of the effective theory, and in Sect. 5 we briefly summarize the main results.

2 The 'Optical' and Other Sum Rules

Inclusive heavy flavor decays are closely related to the case of deep inelastic scattering (DIS). While the latter was in the focus of theoretical investigations in the early days of QCD, heavy flavor decays received marginal attention. Recently it has been shown that many elements of the theory of DIS find their parallels in heavy flavor decays. In this paper we discuss one more aspect with an apparent analogy in the theoretical treatments, namely the sum rules.

Let us recall that the standard analysis of DIS [27] proceeds as follows. One starts from the operator product expansion (OPE) for the T product

$$\hat{T}_{\mu\nu} = i \int d^4x e^{-iqx} T\{j_\mu^\dagger(x) j_\nu(0)\} \tag{1}$$

where j_μ is the electromagnetic or some other current of interest. The average of $\hat{T}_{\mu\nu}$ over nucleon state with momentum p presents a forward scattering amplitude of Compton type. The expansion is performed for large Euclidean q^2 and small (vanishing) values of $\nu = qp$, i.e. parametrically far from the physical cuts. After averaging of the operator product expansion one arrives at a set of predictions for the coefficients of the ν expansion of the Compton amplitude at $\nu = 0$ and given q^2. These coefficients are related, *via* dispersion relations, to the integrals over the imaginary part of the amplitude at hand (the moments of the structure functions).

The strategy, as well as the results obtained, are quite general. At the same time, certain moments play a distinguished role due to the fact that they turn out to be proportional to operators whose matrix elements over the nucleon state are known on general grounds. Relations emerging in this way are called sum rules proper and possess particular names. In the case of unpolarized target we deal with the Adler sum rule for neutrino scattering and the Gross-Lewellyn-Smith sum rule [27].

Conceptually a very similar description can be applied to the weak decays of heavy flavor hadrons. One relates the observable quantities to a non-local transition operator $\hat{T}$, expands the latter into a series of local operators and determines their expectation values between the state H_Q. There exist of course several technical differences relative to the case of DIS: one deals with heavy quark currents, uses the heavy quark mass as the expansion parameter instead of the square of the momentum transfer and forms the expectation values for the heavy flavor hadron H_Q. Furthermore one can predict absolute decay rates, not only their evolution as the scale changes.

To introduce the reader to the range of questions to be considered below in the most straightforward way we first "peel off " all inessential technicalities, like the quark spins, and resort to a simplified model which has been previously discussed in Ref. [14]. It will be a rather simple exercise to return afterwards to standard QCD.

2.1 Description of the model

We consider a toy example where all quarks are spinless; two, denoted by Q and q, couple to a massless real scalar field ϕ:

$$\mathcal{L}_\phi = h\bar{Q}\phi q \; + \; h.c. \; , \tag{2}$$

where h is the coupling constant and $\bar{Q} = Q^\dagger$. The masses of the quarks Q and q are both *large*. Later on we will analyze the SV limit where

$$\Lambda_{QCD} \ll m_Q - m_q \ll m_Q \tag{3}$$

has to hold, but for the time being the ratio $\gamma = m_q/m_Q$ can be arbitrary provided that $m_Q - m_q \gg \Lambda_{QCD}$ [1]. The field ϕ carries color charge zero; the reaction $Q \to q + \phi$ is thus a toy model for the radiative decays of the type $B \to X_s\gamma$.

The total width for the free quark decay $Q \to q + \phi$ is given by the following expression:

$$\Gamma(Q \to q\phi) = \frac{h^2 E_0}{8\pi m_Q^2} \equiv \Gamma_0 \tag{4}$$

where

$$E_0 = \frac{m_Q^2 - m_q^2}{2m_Q}. \tag{5}$$

As explained in Refs. [1] – [14] the theory of preasymptotic effects in the inclusive decays is based on introducing the transition operator,

$$\hat{T} = i \int d^4x \, e^{-iqx} T\{\bar{Q}(x)q(x)\, , \, \bar{q}(0)Q(0)\}. \tag{6}$$

Then the energy spectrum of the ϕ particle in the inclusive decay is obtained from $\hat{T}$ in the following way:

$$\frac{d\Gamma}{dE} = \frac{h^2 E}{4\pi^2 M_{H_Q}} \mathrm{Im}\, \langle H_Q|\hat{T}|H_Q\rangle \tag{7}$$

where H_Q denotes a hadron built from the heavy quark Q and the light cloud (including the light antiquark). If not stated otherwise H_Q will denote the ground state in a given channel. Moreover, one can (and must) apply the Wilson operator product expansion [28] (OPE) to express the non-local operator $\hat{T}$ through an infinite series of local operators with calculable coefficients.

[1] To reach the SV limit in the $B \to X_c e\nu$ transitions there is no need to assume that $m_b - m_c \ll m_c$. The small velocity regime for the c quark can be ensured by adjusting q^2 appropriately.

2.2 OPE and predictions for observable quantities

In the Born approximation the transition operator has the form (Fig. 1)

$$\hat{T} = -\int d^4x (x|\bar{Q}\frac{1}{(P_0 - q + \pi)^2 - m_q^2} Q|0) \ , \tag{8}$$

$$iD_\mu \equiv m_Q v_\mu + \pi_\mu$$

which is particularly suitable for constructing the OPE by expanding eq.(8) in powers of π (see Ref. [14] where all notations have been introduced). The operators appearing as a result of this expansion are ordered according to their dimension. The leading operator $\bar{Q}Q$ has dimension 2 (let us remind that the scalar Q field has dimension 1 in contrast to the real quark fields of dimension 3/2 which leads in particular to different normalization factors; we still use relativistic normalization). Its expectation value for the state H_Q reduces to unity to leading order in $1/m_Q$; this contribution gives rise to the parton result eq.(4). Beyond the leading approximation it takes the form [14]

$$\frac{1}{2M_{H_Q}}\langle H_Q|\bar{Q}Q|H_Q\rangle = \frac{1}{2m_Q}\left(1 - \frac{1}{2m_Q^2}\langle H_Q|\bar{Q}\vec{\pi}^2 Q|H_Q\rangle + ...\right) ; \tag{9}$$

thus, one gets a correction of order $1/m_Q^2$.

Corrections of the same order come from higher-dimensional operators in the expansion of the transition operator. There are no relevant operators of next dimension 3 (more exactly, as was first noted in Ref. [4] in the framework of HQET, they vanish because of the equations of motion[2]). The only operator of dimension four in our toy model has the form $\bar{Q}\vec{\pi}^2 Q$ and it generates a $1/m_Q^2$ correction after taking the matrix element over H_Q. We will not discuss corrections of order $1/m_Q^3$ or higher in this paper.

After simple algebra one finds

$$\frac{1}{\pi}\mathrm{Im}\,\langle\hat{T}\rangle = \left(\frac{\langle\bar{Q}Q\rangle}{2m_Q} - \frac{\langle\bar{Q}\vec{\pi}^2 Q\rangle}{12m_Q^3}\right)\delta(E - E_0) -$$

$$\frac{E_0\langle\bar{Q}\vec{\pi}^2 Q\rangle}{12m_Q^3}\delta'(E - E_0) + \frac{E_0^2\langle\bar{Q}\vec{\pi}^2 Q\rangle}{12m_Q^3}\delta''(E - E_0) + ... \tag{10}$$

where operators of higher dimension have been ignored and we have used a shorthand notation for the expectation value over H_Q: $\langle ...\rangle \equiv \langle H_Q|...|H_Q\rangle$.

[2]This statement is sometimes erroneously interpreted as a proof for the absence of a term linear in $1/m_Q$ in the total width, see below. As a matter of fact the authors of Ref. [4] believed that the linear term *may* appear in the total width from the overall normalization of $\langle H_Q|\bar{h}_v h_v|H_Q\rangle$, as explicitly stated, e.g., on page 404 of Ref. [4]. Anyway this question cannot be addressed purely in HQET *per se* because applying effective theory requires here explicit discussion of matching the coefficient functions to full QCD, which, in turn, depends on the operator structure in both theories, see corresponding discussion in Ref. [25].

The expansion of Im $\hat{T}$ into local operators generates more and more singular terms at the point where the ϕ spectrum would be concentrated in the approximation of the free quarks. The physical spectrum on the other hand is a smooth function of E. In principle, one could derive a smooth spectrum by summing up an infinite set of operators to all orders (for more detail see Refs. [14, 15, 16]). There is no need to carry out this summation here, however, since we are interested only in integral characteristics (we will discuss certain sum rules), and as far as they are concerned the expansion in eq. (10) is perfectly legitimate.

At first, we calculate the total width by substituting eq. (10) into eq. (7) and integrating over E. As a matter of fact, the result has already been given in Ref. [14],

$$\Gamma = \int dE \frac{d\Gamma}{dE} = \Gamma_0 \left(1 - \frac{\mu_\pi^2}{2m_Q^2} + ...\right). \tag{11}$$

where the integration runs from 0 to the physical boundary E_0^{phys}, expressed in the hadron masses

$$E_0^{phys} = \frac{M_Q^2 - M_q^2}{2M_Q}, \quad M_Q \equiv M_{H_Q}, \tag{12}$$

and the same convention for M_q. We use here the standard notation for the expectation value of $\vec{\pi}^2$,

$$\mu_\pi^2 = \langle H_Q|\bar{Q}\vec{\pi}^2 Q|H_Q\rangle. \tag{13}$$

The expression in the brackets in eq. (11) is nothing but the (corrected) matrix element of the operator $\bar{Q}Q$; the only possible effective operator $\bar{Q}\vec{\pi}^2 Q$ of dimension 4 drops out in Γ on general grounds [14] being not a Lorentz scalar. The meaning of this term in $\langle \bar{Q}Q\rangle$ is quite transparent: it reflects time dilation for the moving quark, and the coefficient $(-1/2)$ could therefore have been guessed from the very beginning. The absence of a correction of order $1/m_Q$ in the total width in the toy example under consideration is a manifestation of the general theorem established in Refs. [6, 7] and discussed in more detail recently in Ref. [25].

As we will see shortly, eq. (11) treated in the SV limit is equivalent to two results simultaneously: that of Ref. [22] and the Bjorken sum rule [19, 20], with appropriate corrections due to terms which were not considered in Refs. [19, 20, 22]. All sum rules analogous to that of eq. (11) below will generically be referred to as *the first sum rule*.

Eq. (11) is the first example of the sum rules we will be dealing with throughout the paper. Its derivation (as well as that of all similar sum rules) intuitively is perfectly clear – the integrated spectrum of the decay obtained at the quark-gluon level using the OPE is equated with the integrated physical spectrum, saturated by the genuine hadronic states. We will elaborate on the justification for this procedure in Sect. 3.1 where, among other things, we discuss the accuracy one may expect from relations of this type; the general idea lying behind all such relations is widespread in QCD. For the moment we just adopt the heuristic attitude outlined above without submerging into further, less pragmatic, questions.

Next, we calculate the average energy of the ϕ particle. The corresponding sum rule (see Ref. [29]) in the SV limit is just a version of Voloshin's optical sum rule!

To be more exact, one considers

$$I_1 = \int_0^{E_0^{phys}} dE\,(E_0^{phys} - E)\frac{d\Gamma}{dE}. \tag{14}$$

Notice that in the SV limit $E_0^{phys} - E$ reduces to the excitation energy of the final hadron produced in the decay. The factor $E_0^{phys} - E$ in the integrand eliminates the "elastic" peak, so that the integral is saturated only by the inelastic contributions. Let us note in passing (we will discuss this point later in more detail) that the optical sum rule [23] in its original formulation is actually divergent; a cutoff must be introduced by hand. In our formulation, of course, there is no place for divergences; the expectation value of $E_0^{phys} - E$ is defined *via* convolution with the physical spectrum, and since the ϕ energy in the decay is finite the expectation value of $E_0^{phys} - E$, as well as higher moments, are certainly finite. The decay kinematics provides us with a natural cutoff at the scale of the energy release.

Explicit calculation of I_1 using eq. (10) yields

$$I_1 = \Gamma_0\left(\Delta - \frac{\mu_\pi^2 E_0^{phys}}{2m_Q^2}\right) \tag{15}$$

where Δ is defined as follows

$$\Delta = E_0^{phys} - E_0, \tag{16}$$

see eqs. (5) and (12). The sum rule (15) contains Voloshin's result, again with corrections left aside previously [23, 29]. The sum rules analogous to that of eq. (15) below will be generically referred to as *the second sum rule.*

It is quite evident that the series of sum rules generalizing those of Bjorken and Voloshin can readily be continued further. For the next moment, for instance, we get

$$I_2 = \int_0^{E_0^{phys}} dE\,(E_0^{phys} - E)^2\frac{d\Gamma}{dE} =$$
$$\Gamma_0\left[\Delta^2\left(1 - \frac{\mu_\pi^2}{2m_Q^2}\right) - \frac{\mu_\pi^2 E_0\Delta}{m_Q^2} + \frac{\mu_\pi^2 E_0^2}{3m_Q^2}\right]. \tag{17}$$

Analyzing this sum rule in the SV limit one obtains, in principle, additional information, not included in the results of Refs. [19, 20, 22, 23]. It is worth emphasizing that in eqs. (11), (15) and (17) we have collected all terms through order Λ_{QCD}^2, whereas those of order Λ_{QCD}^3 are systematically omitted. Predictions for higher moments would require calculating terms $\mathcal{O}(\Lambda_{QCD}^3)$ and higher.

2.3 SV limit and the sum rules

We proceed now to discussing the sum rules (11), (15) and (17) in the SV limit, eq. (3). The availability of the extra expansion parameter, $|\vec{v}| \ll 1$ makes the SV limit a very interesting theoretical laboratory. We have already mentioned that in this limit the Bjorken sum rule relates the distortion of the "elastic" peak to the integral over the inelastic contributions. Here we will elaborate on this issue. First, we briefly recall what is already known and then present new results.

For technical reasons we will assume that

$$\Lambda_{QCD} \ll m_Q - m_q \ll \sqrt{m_Q \Lambda_{QCD}}. \tag{18}$$

The inequality on the right hand side is not essential and can easily be lifted; it helps, however, to make all formulae more concise, and we will accept it for a while, thus replacing eq. (3) by the stronger condition of eq. (18). We now supplement the expansion in Λ_{QCD} carried out above by an expansion in $\vec{v}$. The natural hierarchy of parameters in the domain (18) is as follows

$$\vec{v}^2, \;\; |\vec{v}|(\Lambda_{QCD}/m_Q), \;\; \Lambda^2_{QCD}/m_Q^2 \;\; .$$

Terms of order $v^4 \ll \Lambda^2_{QCD}/m_Q^2$ will be omitted.

Expanding the hadron masses in terms of the heavy quark masses m_Q and m_q one finds for scalar quarks

$$M_Q = m_Q + \overline{\Lambda} + \frac{\mu_\pi^2}{2m_Q} + \mathcal{O}(1/m_Q^2), \quad M_q = m_q + \overline{\Lambda} + \frac{\mu_\pi^2}{2m_q} + \mathcal{O}(1/m_q^2) \;\; . \tag{19}$$

Eq. (19) implies, in turn, that

$$\Delta = E_0^{phys} - E_0 = \frac{1}{2} v_0^2 \left(\overline{\Lambda} + \frac{2\overline{\Lambda}^2 + \frac{1}{2}\mu_\pi^2}{M_Q} \right) - v_0 \frac{\mu_\pi^2}{2M_Q} + ... \tag{20}$$

where for convenience we introduced a parameter v_0,

$$v_0 = (M_Q - M_q)/M_Q. \tag{21}$$

This parameter approximately coincides with the velocity of the heavy hadron produced in the transition $Q \to q\phi$. Indeed, if the mass of the produced excited hadron is $M_q + \epsilon$ then its velocity

$$|\vec{v}| = v_0 \frac{1-(v_0/2)}{1-v_0} - \frac{\epsilon}{M_Q} \frac{1 - v_0 + (v_0^2/2)}{(1-v_0)^2} - \frac{\epsilon^2}{2M_Q^2} \frac{1 - 4v_0 - 2v_0^2}{(1-v_0)^3} + \mathcal{O}(\epsilon^3). \tag{22}$$

With all these definitions the sum rules (15) and (17) take the form

$$I_1 = \Gamma_0 \left(\frac{1}{2} v_0^2 \overline{\Lambda} - v_0 \frac{\mu_\pi^2}{M_Q} + ... \right) \tag{23}$$

and

$$I_2 = \Gamma_0\,\frac{1}{3}v_0^2\mu_\pi^2 + ... \tag{24}$$

where ellipses denote the systematically omitted terms of order Λ_{QCD}^3, as well as $\mathcal{O}(v_0^2\Lambda_{QCD}^2)$ terms for I_1 and $\mathcal{O}(v_0^4\Lambda_{QCD}^2)$ ones for I_2. The first term on the right hand side of eq. (23) corresponds to Voloshin's relation, the second term is a correction whose physical meaning will soon become clear.

To interpret the sum rules derived it will be instructive to consider the transition operator off the physical cut. Then, expressing the result in terms of $E - E_0^{phys}$ we get

$$M_Q^{-1}\langle H_Q|\hat{T}|H_Q\rangle = \frac{1}{2m_Q^2}\left\{\left(1-\frac{2\mu_\pi^2}{3m_Q^2}\right)\frac{1}{E-E_0^{phys}} + \right.$$
$$\left.\left(-\Delta+\frac{E_0\mu_\pi^2}{6m_Q^2}\right)\frac{1}{(E-E_0^{phys})^2} + \frac{E_0^2\mu_\pi^2}{3m_Q^2}\frac{1}{(E-E_0^{phys})^3} + ...\right\}. \tag{25}$$

To zeroth order in v or Λ_{QCD} there remains only the first term on the right hand side of eq. (25). The inclusive decay rate is then totally saturated by a single "elastic" channel, the production of the ground state meson containing q. This is the perfect inclusive-exclusive duality noted in Ref. [22]. The peak in the ϕ spectrum obtained in the quark transition $Q \to \phi q$ survives hadronization in this approximation; at the hadronic level $H_Q \to H_q\phi$ we still have the same peak at the same energy. Eqs. (23) and (24) are, of course, trivially satisfied in this case because in the absence of the inelastic contributions both sum rules yield vanishing numbers.

Furthermore, the terms with Λ_{QCD} come with v. To order $\overline{\Lambda}$ there is only one such term, appearing in eq. (23) for I_1. This term shows that the inelastic production *must* already be present at this level. It corresponds to the production of a meson H_q^* with excitation energy $\sim \Lambda_{QCD}$ and a residue $\propto v_0^2$. Then eq. (11) implies that the height of the elastic peak is reduced by v_0^2. A rough model examplifying this picture can be obtained from eq. (25). In this approximation, one can rewrite it as follows, omitting all numerical factors:

$$M_Q^{-1}\langle H_Q|\hat{T}|H_Q\rangle = \frac{1}{2m_Q^2}\left\{(1-v_0^2)\frac{1}{E-E_0^{phys}} + v_0^2\frac{1}{(E-E_0^{phys}+\overline{\Lambda})} + ...\right\}. \tag{26}$$

The first term is the elastic peak while the second is an inelastic contribution. Of course eq. (26) is an illustration and not a unique solution.

Technically, the term $v^2\overline{\Lambda}$ in eq. (23) arises because the "elastic" peak of the quark transition situated at $E = E_0$ is slightly shifted when we pass to the hadronic transition; the genuine hadronic elastic peak is situated at $E = E_0^{phys}$, to the right of the quark peak (see Fig. 2).

Including $\mathcal{O}(\overline{\Lambda}^2)$ contributions we get correction terms in eqs. (11) and (23), and eq. (24) for I_2 becomes non-trivial. The new term in eq. (23) has a simple meaning. In the toy example at hand the excited mesons produced in the decay, at the level of

v^2 are spin-1 mesons, with the vertex proportional to $(\vec{v}\vec{\epsilon})$ where $\vec{\epsilon}$ is the polarization vector and $\vec{v}$ is the velocity of the given meson. This velocity, differs, however, from the quark-level velocity v_0 by terms of order Λ_{QCD}/m_Q, see eq. (22). This rather trivial shift in velocity nicely explains all the sum rules above. Indeed, let us take into account that the physical v^2 reduces to $v_0^2 - v_0(\overline{\Lambda}/M_Q) + (\overline{\Lambda}^2/M_Q^2)$. The height of the inelastic contribution is proportional to the square of the physical velocity, while the size of the inelastic domain is $\sim \overline{\Lambda} + (\overline{\Lambda}^2/m_Q)$. Hence, with our accuracy the right-hand side of eq. (23) is expected to be $\sim v_0^2(\overline{\Lambda} + (\overline{\Lambda}^2/m_Q)) - v_0(\overline{\Lambda}^2/m_Q)$, while the right-hand side of eq. (24) is expected to be $\sim \overline{\Lambda}^2 v_0^2$, in full accord with what we actually have. A qualitative model of saturation now takes the form

$$M_Q^{-1}\langle H_Q|\hat{T}|H_Q\rangle = \frac{1}{2m_Q^2}\left\{\left(1 - \frac{2\mu_\pi^2}{3m_Q^2} - \gamma^{-1}(\Delta - \frac{E_0\mu_\pi^2}{6m_Q^2})\right)\frac{1}{E - E_0^{phys}} + \right.$$
$$\left. \gamma^{-1}\left(\Delta - \frac{E_0\mu_\pi^2}{6m_Q^2}\right)\frac{1}{E - E_0^{phys} + \gamma}\right\} . \tag{27}$$

where

$$\gamma \sim [\overline{\Lambda} + (\overline{\Lambda}^2/m_Q)] \ .$$

2.4 Perturbative gluon corrections

So far we have assigned the gluon field to play the role of a soft medium to incorporate the effects of long distance dynamics and have completely ignored perturbative gluon corrections. Yet those have to be included; among other things the emission of hard gluons generates the spectral density outside the end-point region which is very relevant for our analysis.

In calculating radiative gluon correction we can disregard, in the leading approximation, non-perturbative effects, like the difference between m_Q and M_Q or the 'Fermi' motion of the initial quark. Thus we deal with the decay of the free quark Q at rest into $q + \phi +$ gluon. The virtual gluon contribution merely renormalizes the constant h in the analysis presented above. The analogous renormalization for the spinor case has been calculated in Ref. [22].

The effects from real gluon emission are most simply calculated in the Coulomb gauge, where only the graph shown in Fig. 3 contributes. A straightforward computation yields to leading order in v^2 (or E)

$$\frac{d\Gamma^{(1)}}{dE} = \Gamma_0 \frac{8\alpha_s}{9\pi}\frac{E^3}{E_0 m_Q^2(E_0 - E)}. \tag{28}$$

It is well-known that the logarithmic singularity in the integral over E for the Γ_{tot} is canceled by a contribution of soft virtual gluons into the renormalization of h. For the second sum rule we are going to discuss here, this infrared range is not singular.

Gluon emission obviously contributes to the spectrum in its entire domain $0 < E < E_0$. In this order α_s does not run, of course. To demonstrate its scale dependence one has to carry out a two-loop calculation; it is quite evident, however, that it is $\alpha_s(E - E_0)$ that enters. Therefore, strictly speaking, one cannot apply eq. (28) too close to E_0. Even leaving aside the blowing up of $\alpha_s(E - E_0)$, there exists another reason not to use eq. (28) in the vicinity of E_0: if E is close to E_0 the gluon emitted is soft; such gluons are to be treated as belonging to the soft gluon medium in order to avoid double counting. The separation between soft and hard gluons is achieved by explicitly introducing a normalization point μ. The value of μ should be large enough to justify a smallness of $\alpha_s(\mu)$. On the other hand we would like to choose μ as small as possible. The possible choice is to have μ proportional to Λ_{QCD}, but with a constant of proportionality that is much larger than unity

$$\mu = C\Lambda_{QCD}, \ \ C \gg 1.$$

Then we draw a line: to the left of $E_0 - \mu$ the gluon is considered to be hard, to the right soft. Of course, the consistent introduction of the infrared renormalization point μ requires that the purely perturbative corrections to the weak vertex, infrared convergent in any finite number of loops, have to be calculated using this explicit cutoff as well (see, e.g. the discussion in Ref. [25]) – which is almost never done in practice. The corresponding modifications will be discussed below in Sect.5.

Let us discuss now the sum rule corresponding to the first moment of $E_0^{phys} - E$, i.e. an analog of eq.(23) with radiative corrections now included. Since our main purpose in this section is methodical, we will limit ourselves to the first order in $\overline{\Lambda}$ and the second order in v. A qualitative sketch of how $d\Gamma/dE$ looks like is presented in Fig. 4. Then the prediction for I_1 can obviously be rewritten as

$$I_1 = \int_0^{E_0^{phys}} (E_0^{phys} - E)\frac{d\Gamma}{dE}dE =$$
$$= \Gamma_0 \frac{1}{2} v_0^2 \left[\overline{\Lambda}(\mu) + v_0^{-2} \int_0^{E_0 - \mu} \frac{16\alpha_s}{9\pi} \frac{E^3}{E_0 m_Q^2} dE \right] \tag{29}$$

where by definition

$$\overline{\Lambda}(\mu) = 2\Gamma_0^{-1} v_0^{-2} \int_{E_0^{phys} - \mu}^{E_0^{phys}} \frac{d\Gamma}{dE}(E_0^{phys} - E)\, dE. \tag{30}$$

Without the radiative tail the prediction for I_1 could be obtained by integrating the theoretical expression (10) over a very narrow domain near E_0, which formally corresponds to $\mu = 0$ (cf. eq.(23)). Clearly, there is no way to switch off the radiative corrections in QCD: one has to deal with the perturbative and non-perturbative contributions simultaneously. The introduction of the parameter μ thus becomes mandatory. Eq. (30) then can provide us with one possible physical definition of $\overline{\Lambda}(\mu)$ (among others) relating this quantity to an integral over a physically measurable spectral density. One may rephrase this statement as follows. Since quarks are permanently confined the notion of the heavy quark mass becomes ambiguous. To eliminate this ambiguity one must explicitly specify the procedure of measuring "the heavy quark mass". Any conceivable procedure will necessarily involve a cut-off parameter μ much in the same way as the procedure defined above, and then $\overline{\Lambda}(\mu) = M_Q - m_Q(\mu)$. In the "most inclusive" procedure when one does not try to separate out any kind of effects, one integrates the tail to the kinematical bound $E \simeq m_Q - m_q$ and, therefore, obtains $\overline{\Lambda}$ normalized at the energy release.

Since this question is very important let us look at it from a slightly different angle. It had been widely believed that $\overline{\Lambda}$ can be defined as a universal constant. The standard definition, being applied to our example, would involve three steps: (i) Take the radiative perturbative tail to the left of the shoulder and extrapolate it all the way to the point $E = E_0$; (ii) subtract the result from the measured spectrum; (iii) integrate the difference over dE with the weight function $(E - E_0^{phys})$. The elastic peak drops out and the remaining integral is equal to $\Gamma_0(v^2/2)\overline{\Lambda}$. Yet this procedure cannot be carried out consistently! For there exists no unambiguous way to extrapolate the perturbative tail too close to E_0^{phys}, the end-point of the spectra[3]. Our procedure, with the normalization point μ introduced explicitly, is free from this ambiguity. We will further comment on the issue in Sect. 3.4 where we discuss the possibility of measuring $\overline{\Lambda}(\mu)$ in the inclusive $B \to X_c l\nu$ decays.

In practice, the μ dependence of $\overline{\Lambda}(\mu)$ may turn out to be rather weak. This is the case if the spectral density is such as shown in Fig. 4, where the contribution of the first excitations (lying within Λ_{QCD} from E_0^{phys}) is numerically much larger than the radiative tail representing (at least in the sense of duality) high excitations. It is quite clear that if the physical spectral density resembles that of Fig. 4 and $\mu =$ several units $\times\, \Lambda_{QCD}$ the running $\overline{\Lambda}(\mu)$ is rather insensitive to the particular choice of μ. As known from QCD sum rules it is just this situation which occurs for the standard quark and gluon condensates (the so-called practical version of Wilson's OPE).

Still, even if the μ dependence of $\overline{\Lambda}$ is practically weak, conceptually it is impossible to define $\overline{\Lambda}$ in the limit $\mu \to 0$. Physically it is quite clear from the discussion presented above. This consideration can be thought of as an illustration to a more formal argument presented recently [25, 26].

If one replaces E in the numerator by E_0 (the non-relativistic approximation) one arrives at the formula obtained previously in Ref. [23]. Notice that the total prediction for I_1 is, of course, μ independent. Eq. (29) shows how $\overline{\Lambda}(\mu)$ changes under the variation of the normalization point,

$$\delta\overline{\Lambda} = \delta\mu \frac{16}{9}\frac{\alpha_s(\mu)}{\pi}. \tag{31}$$

The numerical coefficient in front of $\delta\mu$ is slightly different from that found in

[3]The standard procedure effectively corresponds to using the literal perturbative expression for this tail with non-running strong coupling (for one loop calculations), or accounting for the first term in the expansion of $\alpha_s(k)$ in terms of $\alpha_s(m)$ (in the two loop ones).

Ref. [25] (16/9 versus $2\pi/3$). The reason is obvious – we use here a different procedure for defining $\overline{\Lambda}(\mu)$ compared to that suggested in [25]: introducing a gluon mass λ "switches off" the perturbative tail in a 'soft' rather than 'hard' way at $\mu = \lambda$. The fact itself of the presence of a linear (in μ) renormalization is obviously common for all proper procedures. In other words, the anomalous dimension of $\overline{\Lambda}$ is the same as that of the running mass and in the limit $m_Q \to \infty$ it is not logarithmic, but power-like. This fact can be traced back to the mixing between the operators $\bar{Q}(ivD)Q$ and $\bar{Q}Q$ established in [25]. Numerically $\delta\overline{\Lambda} \sim 0.1\,\mathrm{GeV}$ if μ changes from 1 to 1.5 GeV, i.e. $\delta\overline{\Lambda}(\mu)/\overline{\Lambda}(\mu) \sim 0.2$.

By the same token, using the second sum rule, eq. (24), one can give a physical definition of $\mu_\pi^2(\mu)$,

$$I_2 = \int_0^{E_0^{phys}} (E_0^{phys} - E)^2 \frac{d\Gamma}{dE} dE =$$

$$= \Gamma_0 \frac{1}{3} v_0^2 \left[\mu_\pi^2(\mu) + v_0^{-2} \int_0^{E_0 - \mu} \frac{8\alpha_s}{3\pi} \frac{E^3 (E_0 - E)}{E_0 m_Q^2} dE \right] \tag{32}$$

where by definition [4]

$$\mu_\pi^2(\mu) = 3\Gamma_0^{-1} v_0^{-2} \int_{E_0^{phys} - \mu}^{E_0^{phys}} \frac{d\Gamma}{dE} (E_0^{phys} - E)^2 \, dE. \tag{33}$$

The μ dependence is then obtained as follows:

$$\delta\mu_\pi^2 \sim \frac{4\alpha_s}{3\pi} \delta\mu^2 \sim 0.1\,\mathrm{GeV}^2 \quad \text{if} \quad \delta\mu^2 \sim 1\,\mathrm{GeV}^2 \ .$$

Because the integral of the perturbative tail does not depend on the particular heavy flavor hadron in the initial state, this equation is equivalent to the corresponding power mixing of the kinetic energy operator with the leading one:

$$\frac{d}{d\mu^2} \bar{Q}\,(i\vec{D})^2\,Q \simeq \frac{4\alpha_s(\mu)}{3\pi}\,\bar{Q}Q \tag{34}$$

where the exact coefficient holds for this particular way of introducing the renormalization point.

2.5 Sum rules for the form factors at zero recoil

We continue investigating our toy model with the aim of establishing sum rules for the form factors at zero recoil. These sum rules will allow us to find corrections to the elastic form factor at *zero recoil* in terms of inelastic contributions.

The starting idea is to consider the kinematical point $\vec{q} = 0$ and q_0 close to $\Delta M = M_B - M_D$. In other words we abandon the case $q^2 = 0$ and turn to $q^2 \neq 0$

[4] We hope that this rather clumsy notation will not cause confusion: μ_π^2 is the matrix element of the kinetic energy operator while μ (with no subscript) is a normalization point.

where q is now the momentum transfer caused by some external "weak" current. Since $\vec{q} = 0$ the ground state B meson cannot decay into a P-wave state. Unlike the case of the Bjorken sum rule there is no "external" vector $\vec{v}$. The fact that we are at zero recoil implies that inelastic contributions are suppressed as Λ^2/m^2, i.e. produce effect of the same order of magnitude as the corrections to the elastic form factor.

In the case when $\vec{q} \neq 0$ the strength of the contribution of the excited resonance is proportional to $\vec{q}^2/M_D^2$ [20]. In our kinematics $\vec{q} = 0$ and the only relevant velocity is provided by the primordial motion of the c quark inside D (more exactly, it is the 'difference' between the quark motions in B and D that counts). The limit $\vec{q} \to 0$ is obtained, qualitatively, if instead of the "external" velocity we use that of the primordial motion,

$$\vec{q}^{\,2}/M_D^2 \to \Lambda_{QCD}^2/M_D^2 \ .$$

This analogy nicely illustrates why the residues of the excited resonances are proportional to Λ_{QCD}^2/M_D^2 and match the picture of Fig. 4 where the contribution of the excited resonances is proportional to $\vec{q}^{\,2}/M_D^2$.

Let us sketch how the sum rules at zero recoil emerge technically. To this end we again consider the T product eq. (6) sandwiched between the B meson state. The contribution of a given hadronic state in this amplitude is proportional to

$$\frac{1}{(M_Q - q_0)^2 - M_q^2} = \frac{1}{(\Delta M - q_0)(\Delta M - q_0 + 2M_q)} \ . \tag{35}$$

It is convenient to introduce the variable

$$\epsilon = \Delta M - q_0 \ . \tag{36}$$

We are interested in singularities of the amplitude at small ϵ. The second pole in eq. (35) at $\epsilon = -2M_q$ is a reflection of a distant singularity corresponding to the transition $B + \bar{D} \to \phi^*$ where the asterisk marks the virtual ϕ quantum. We assume, as usual, the hierarchy $\Lambda_{QCD} \ll \epsilon \ll m_{Q,q}$. Correspondingly, we will expand in Λ_{QCD}/ϵ and in ϵ/m. In particular, the factor responsible for the second pole in eq. (35) becomes

$$\frac{1}{\Delta M - q_0 + 2M_q} = \frac{1}{2M_q}(1 - \frac{\epsilon}{2M_q} + ...) \ ;$$

the second and higher order terms in the brackets can be omitted since they lead to a non-singular (at $\epsilon = 0$) expression. The non-singular expressions have no imaginary part and are irrelevant for our purposes.

Next one considers the theoretical expression for the quark level transition operator,

$$-\hat{T} = \bar{Q}\frac{1}{(m_Q - m_q - q_0)(m_Q + m_q - q_0) + (\pi^2 + 2m_Q\pi_0 - 2q_0\pi_0)}Q \ . \tag{37}$$

Our task is to expand the transition operator in Λ_{QCD}/ϵ and in $\epsilon/m_{Q,q}$ and then to compare the terms singular in $1/\epsilon$ in this theoretical expansion with the phenomenological expression obtained in the language of the resonance saturation. A technical point which deserves mentioning right at the beginning is a mismatch in the definitions of ϵ. The theoretical expression (37) is phrased in terms of the quark mass differences without reference to mass differences of mesons. Since we would like to get the sum rules written in terms of the physical excitation energies (measured from the mass of the lowest-lying meson state) we have to express eq. (37) in terms of ΔM rather than Δm before expanding it. As we will see shortly, for our purposes it is necessary to keep all effects through order Λ^3/m^3; those of order Λ^4/m^4 and higher can be neglected.

The expansion for the meson mass in inverse powers of the heavy quark mass for spinless quarks considered here looks especially simple. Namely,

$$M_Q = m_Q + \overline{\Lambda} + \frac{\langle\vec{\pi}^2\rangle}{2m_Q} + \mathcal{O}\left(\frac{\Lambda^3_{QCD}}{m_Q^2}\right) \tag{38}$$

where the following short-hand notation is used,

$$\langle\vec{\pi}^2\rangle \equiv \langle H_Q|\bar{Q}\vec{\pi}^2 Q|H_Q\rangle \ .$$

The third term on the right hand side in eq. (38) represents a $1/m_Q^2$ relative effect. We are interested here in the effects of order $1/m_Q^3$. Generally speaking, $1/m_Q^3$ corrections can arise from two different sources: (i) $1/m_Q^2$ terms in the expansion of the Lagrangian in the effective theory describing relativistic corrections; (ii) the second order in $\vec{\pi}^2/m_Q$ correction to the meson mass. It is rather trivial to show that for spinless quarks $\mathcal{O}(1/m_Q^2)$ terms in the effective Lagrangian are absent, and iteration of the $\vec{\pi}^2/m_Q$ term becomes the only source for the mass correction we are looking for. It is helpful to note that this correction, then, is negative, as a second order perturbation, in accordance with the well-known theorem. Let us call the corresponding parameter μ_3^3,

$$M_Q = m_Q + \overline{\Lambda} + \frac{\langle\vec{\pi}^2\rangle_0}{2m_Q} - \frac{\mu_3^3}{m_Q^2} + \mathcal{O}(m_Q^{-3}) \tag{39}$$

where

$$\mu_3^3 = i m_Q \int d^4x \langle H_Q|T\{\bar{Q}(x)\vec{\pi}^2 Q(x), \bar{Q}(0)\vec{\pi}^2 Q(0)\}|H_Q\rangle' > 0 \tag{40}$$

where the prime indicates the fact that the elastic pole has to be removed from the correlation function on the right-hand side. The subscript '0' in $\langle\vec{\pi}^2\rangle_0$ is to show that it is the value of the matrix element at $m_Q \to \infty$. The expectation value over the physical state H_Q at finite m_Q is different,

$$\langle\vec{\pi}^2\rangle = \langle\vec{\pi}^2\rangle_0 - 4\frac{\mu_3^3}{m_Q} + \mathcal{O}(m_Q^{-2}) \ , \tag{41}$$

so knowing $\langle \vec{\pi}^2 \rangle$ for both H_Q and H_q fixes $\langle \vec{\pi}^2 \rangle_0$ and μ_3^3 separately. With these definitions one has

$$\Delta M = \Delta m \left(1 - \frac{\langle \vec{\pi}^2 \rangle_0}{2m_Q m_q} + \mu_3^3 \frac{(m_Q + m_q)}{m_Q^2 m_q^2} \right) \quad . \tag{42}$$

Substituting eq. (42) in eq. (37) and expanding in Λ/ϵ and $\epsilon/m_{Q,q}$ we arrive at

$$-\langle H_Q|\hat{T}|H_Q\rangle = \frac{1}{\epsilon}\frac{\langle \bar{Q}Q\rangle}{2m_q}\left(1 - \frac{\langle \vec{\pi}^2\rangle}{2m_q^2}\right) + \frac{1}{\epsilon^2}\mu_3^3 \frac{(\Delta m)^2}{2m_Q^2 m_q^3} \tag{43}$$

plus terms of higher order in $1/\epsilon$ and in Λ.

This theoretical expression is to be confronted now with what one obtains from saturating the amplitude at hand with meson poles. From eq. (35) it is not difficult to see that

$$-\langle H_Q|\hat{T}|H_Q\rangle = \sum_{i=0,1,...} \left(\frac{1}{\epsilon}\frac{F_i^2}{2(M_q)_i} + \frac{1}{\epsilon^2}\frac{\epsilon_i^* F_i^2}{2(M_q)_i} + ... \right) \tag{44}$$

where F_i is the form factor for the transition $H_Q \to H_q^i$ induced by the vertex eq. (2),

$$\langle H_q^i|\bar{q}Q|H_Q\rangle = F_i \quad , \tag{45}$$

$(M_q)_i$ is the mass of i-th state ($(M_q)_0 = M_q$ is the mass of ground state) and

$$\epsilon_i^* = (M_q)_i - (M_q)_0 \tag{46}$$

is the excitation energy of the i-th state; all form factors are taken at the zero recoil point, where the meson produced in the transition $Q \to q$ is at rest in the rest frame of H_Q.

Comparing eqs. (44) and (43) we find that

$$\sum_{i=0,1,...} \frac{F_i^2}{2(M_q)_i} = \frac{\langle \bar{Q}Q\rangle}{2m_q}\left(1 - \frac{\langle \vec{\pi}^2\rangle}{2m_q^2}\right) \tag{47}$$

and

$$\sum_{i=1,...} \frac{\epsilon_i^* F_i^2}{2(M_q)_i} = \mu_3^3 \frac{(\Delta m)^2}{2m_Q^2 m_q^3}. \tag{48}$$

Note that the elastic pole gives no contribution in eq. (48). Hence we conclude that the residues of the excited states are proportional to Λ^2. Transferring then all excited states in eq. (47) to the right-hand side we observe that the square of the elastic form factor receives corrections of order Λ^2, both, of local nature and due to excited states. Since all excitation energies and all residues are positive an obvious inequality holds,

$$\sum_{i=1,2,...} \frac{F_i^2}{2(M_q)_i} < \frac{1}{\epsilon_1^*}\mu_3^3 \frac{(\Delta m)^2}{2m_Q^2 m_q^3} \tag{49}$$

where ϵ_1^* is the excitation energy of the first excited state. Therefore

$$\frac{\langle \bar{Q}Q \rangle}{2m_q}\left(1-\frac{\langle \vec{\pi}^2 \rangle}{2m_q^2}\right)-\frac{1}{\epsilon_1^*}\mu_3^3\frac{(\Delta m)^2}{2m_Q^2 m_q^3}<\frac{F_0^2}{2M_q}<\frac{\langle \bar{Q}Q \rangle}{2m_q}\left(1-\frac{\langle \vec{\pi}^2 \rangle}{2m_q^2}\right) \tag{50}$$

where F_0 is the form factor of the "elastic" transition $H_Q \to H_q$.

It is worth noting that the explicit form of the corrections, in particular to the first sum rule eq. (47) and, therefore, the "local" terms $\propto \langle \vec{\pi}^2 \rangle$ in eq. (50) depend on the structure of the "weak" current considered and refer to the case of the scalar vertex. Should we use the vector current, coefficients in sum rules would take a form leading to $F_0 = 1$ at $m_q = m_Q$ in accord with the exact conservation of the vector current for equal masses.

Concluding this section let us mention a convenient computational device. It is helpful to let the initial quark mass go to infinity and retain corrections only in $1/m_q$. In this way one removes nonperturbative corrections originating in the initial state. The results referring to finite m_Q can be simply reconstructed at the very end. On the other hand, in the opposite limit, $m_q \gg m_Q$ one suppresses nonperturbative effects in the final state; in this way it is convenient to obtain relations for static hadronic quantities.

Similar sum rules at zero recoil in real QCD will be discussed in Sect. 3.2.

3 Real QCD

We proceed now to discuss the sum rules emerging in QCD for processes of the type $B \to X_c l\nu$. It is clear that the approximation $m_b - m_c \ll m_{b,c}$ is not suitable in this case; one still can reach the SV limit, however, by using the fact that q^2 is not necessarily zero in this transition (from now on q is the momentum of the lepton pair). Indeed, if q^2 is close to its maximal value,

$$q_{max}^2 = (M_B - M_D)^2 \ ,$$

the D meson velocity is small. At the maximal value of q^2 the velocity vanishes. It is not difficult to show that the velocity $\vec{v} = -\vec{q}/E$ is related to q^2 as follows

$$\frac{1}{\sqrt{1-v^2}} - 1 = \frac{(M_B - M_D)^2 - q^2}{2M_B M_D} \ , \tag{51}$$

whereas at the quark level

$$\frac{1}{\sqrt{1-v^2}} - 1 = \frac{(m_b - m_c)^2 - q^2}{2m_b m_c} \equiv \xi \ . \tag{52}$$

For a sizeable fraction of events measured in the semi-leptonic B decays the values of q^2 are such that these events actually do belong to the SV limit (i.e. v is small). An indirect proof of the relevance of the SV limit to the inclusive semi-leptonic decays of B's comes from the fact that about 65% of the total semi-leptonic rate is given by the "elastic" transitions to D and D^* [30]. The analysis below is carried out under the assumption that the right-hand side in eq. (51) is small. Even though we do not assume that $m_b - m_c \ll m_{b,c}$ both quarks, b and c, will be treated as heavy, $m_{b,c} \gg \Lambda_{QCD}$.

Needless to say that the proximity of q^2 to q^2_{max} can be realized in different ways; for instance, one can merely put $\vec{q} = 0$ – this is especially convenient if we are interested in the zero recoil point – or one can keep $\vec{q} \neq 0$, but small and study the terms proportional to $\vec{q}^{\,2}$. This yields a practically realizable method of measuring $\overline{\Lambda}(\mu)$ in the semileptonic decays $B \to X_c e\nu$. We will consider first the simplest sum rule for the total decay width analogous to eq. (47) in Sect. 2.5. Surprisingly, this analysis produces a lower bound on the deviation from unity in the $B \to D^*$ elastic form factor at zero recoil which does not quite agree with the previous estimates obtained by a different method [31] (see also Ref. [18]). The result is of a paramount importance for experimental determination of $|V_{cb}|$, the CKM matrix element, from the exclusive decay $B \to D^* e\nu$. Then we turn to an analog of Voloshin's sum rule which appears to be a promising tool for extracting $\overline{\Lambda}(\mu)$.

The correct operator definition of $\overline{\Lambda}$ as the quantity relating the mass of the heavy flavour hadron to the heavy quark mass has been obtained in Ref. [14] analyzing the heavy quark distribution function appearing in the SV limit for the final state quark. Using eqs. (75) and (76) of that paper one can express it directly via the temporal distribution function $G(y)$:

$$\overline{\Lambda} = \frac{2}{3}\int \frac{d\epsilon}{\epsilon}\int \frac{dt}{2\pi} e^{it\epsilon} \frac{1}{2M_B} \langle B_{\vec{v}=0}|\bar{b}(t,\vec{x}=0)\; \pi_i \; e^{i\int_0^t A_0(t')dt'}\; \pi_i \, b(0)|B_{\vec{v}=0}\rangle \; . \qquad (53)$$

It is therefore given by the $n = -1$ moment of the temporal distribtuion function G and thus is *not* expressed in terms of a local operator (see also Ref. [25]). Its renormalization point dependence can easily be traced formally through the properties of the path ordered exponent which, being the field operator, requires specification of the normalization point for the gauge fields. In Sect. 2.4 we have illustrated the renormalization point dependence using the saturation of the correlator by intermediate states, which is equivalent to calculating the correlator *via* dispersion relations.

It is worth mentioning that eq.(53) has a transparent meaning in quantum mechanics. Using the quantum mechanical interpretation of the matrix element with the path ordered exponent as explained below in Sect.3.1, we can express the correlator function in eq.(53) as a sum over intermediate states,

$$\langle B|\bar{b}(t,\vec{x}=0)\; \pi_i \; e^{i\int_0^t A_0(t')dt'}\; \pi_i \, b(0)|B\rangle = \frac{1}{2M_B}\sum_n |\langle B|\bar{b}\,\pi_i\, b|n\rangle|^2 \cdot e^{-i(E_n - M_B)t}$$

and inegrating over x we arrive at

$$\frac{\overline{\Lambda}}{2} = \sum_n \; \frac{1}{3}\frac{|\langle B|\bar{b}\,\pi_i\, b|n\rangle|^2}{E_n - M_B}$$

(here we use nonrelativistic normalization of states). It is easy to see that in simple potential description of the heavy hadron as a two body system the sum on the right hand side is given by the half of the mass of the spectator. Indeed,

$$- \sum_n \frac{|\langle B|\vec{\pi}\vec{v}|n\rangle|^2}{E_n - M_B}$$

represents the second order in v correction to the ground state energy, produced by perturbation

$$\delta H = \vec{\pi}\vec{v} \ .$$

On the other hand

$$H + \delta H = H(\vec{\pi} + \vec{v}) - \frac{mv^2}{2} \ .$$

The eigenvalues of the Hamiltonian given by the first term on the right hand side are the same as for the nonperturbed one, which leads to the relation stated above.

The operator definition of $\overline{\Lambda}$ suggested above also helps elucidate the problems assocciated with the attempts to define the pole mass at the level of Λ_{QCD} from a somewhat different perspective. For to determine the mass of a hadron in terms of m_Q specified in some way (say at normalization point $\sim 1\,\mathrm{GeV}$) one needs to evaluate the matrix element in eq. (53) between the particular hadron. The pole mass of the heavy quark then is to be understood as the corresponding matrix element taken over the perturbative states $|Q\rangle$ representing the isolated (but strongly interacting!) quark Q, which persist in the perturbation theory. Such states do not exist in reality, therefore one could only think that the operator object in eq. (53) is not infrared sensitive so that "nonperturbative" effects occuring at the scale $\sim \Lambda_{QCD}$ do not spoil it. On the other hand, purely phenomenologically we know that the matrix elements for the operator in eq. (53) do differ by values of order Λ_{QCD}, say for mesons and baryons. Therefore if the pole mass could unambiguously be defined at the level of Λ_{QCD}, one would have had masses of heavy flavour hadrons degenerate – in a clear conflict with experiment.

3.1 Sum rules at zero recoil: generalities

Our analysis of the $B \to X_c e\nu$ problem at zero recoil will parallel the corresponding consideration carried out in the toy model of Sect. 2.5. The presence of spin is a technicality which can easily be incorporated. As a matter of fact, all formulae necessary for derivation of the first and the second sum rules exist in the literature; we will borrow them from Ref. [10] as well as all relevant notations.

The point $\vec{q} = 0$ represents zero recoil. Then the transition operator

$$\hat{T}_{\mu\nu} = i \int e^{-iqx} dx T\{j_\mu^\dagger(x) j_\nu(0)\} \ ; \qquad (54)$$

for the $b \to c$ transitions,

$$j_\mu = \bar{c}\Gamma_\mu b \ ,$$

can be presented (in the tree approximation) in the form of the following expansion:

$$\hat{T}_{\mu\nu} = \bar{b}\Gamma_\mu (k_0\gamma_0 + m_c + \not{\pi})\frac{1}{(m_c^2 - k_0^2)}\sum_{n=0}^{\infty}\left(\frac{2k_0\pi_0 + \pi^2 + (i/2)\sigma G}{m_c^2 - k_0^2}\right)^n \Gamma_\nu b \tag{55}$$

with

$$k_0 = m_b - q_0.$$

The operator product expansion (55) is justified provided that

$$\Lambda_{QCD} \ll |m_c - k_0|.$$

In other words, the expansion (55) is a series in $\Lambda_{QCD}/(m_c - k_0)$. At the same time, apart from the poles $1/(m_c - k_0)$ it obviously contains powers of $1/(m_c + k_0)$ which develop "distant" singularities at $k_0 = -m_c$. We want these singularities corresponding to the propagation of the antiquark $\bar{c}$ to be indeed distant so that the dispersion integrals we will be dealing with do not stretch up to these $\bar{c}$ containing states. To this end we must impose the second condition on $|m_c - k_0|$, namely

$$|m_c - k_0| \ll m_c\,.$$

Once this condition is imposed we expand $\hat{T}_{\mu\nu}$ in powers of $(m_c - k_0)/m_c$ and $\Lambda_{QCD}/(m_c - k_0)$. The result is then ordered with respect to the powers of $1/(m_c - k_0)$. The terms non-singular in $(m_c - k_0)$ are irrelevant and can be discarded. Each particular power $1/(m_c - k_0)^{n+1}$ in the expansion leads to a sum rule with the weight function $\propto (m_c - k_0)^n$.

Let us sketch the basic elements of the procedure in some detail. We start from a series in $1/(m_c - k_0)$. The next step is averaging of $\hat{T}_{\mu\nu}$ over the B meson state,

$$h_{\mu\nu} = \frac{1}{2M_B}\langle B|\hat{T}_{\mu\nu}|B\rangle. \tag{56}$$

The hadronic tensor $h_{\mu\nu}$ consists of different kinematical structures [4, 10],

$$h_{\mu\nu} = -h_1 g_{\mu\nu} + h_2 v_\mu v_\nu - i h_3 \epsilon_{\mu\nu\alpha\beta} v_\alpha q_\beta + h_4 q_\mu q_\nu + h_5 (q_\mu v_\nu + q_\nu v_\mu). \tag{57}$$

Moreover, the invariant hadronic functions h_1 to h_5 depend on two variables, q_0 and q^2, or q_0 and $|\vec{q}|$. Since we put $\vec{q} = 0$ only one variable survives, and only two of five tensor structures in $h_{\mu\nu}$ are independent.

Each of these hadronic invariant functions satisfies a dispersion relation in q_0,

$$h_i(q_0) = \frac{1}{2\pi}\int \frac{w_i(\tilde{q}_0)d\tilde{q}_0}{\tilde{q}_0 - q_0} \;+\; \text{polynomial} \tag{58}$$

where w_i are observable structure functions,

$$w_i = 2\text{Im}\, h_i.$$

This dispersion representation for the h_i assumes, as usual, that the integral in the right-hand side runs over all cuts that the transition operator may have. The general structure of the cuts in the complex q_0 plane is rather sophisticated; the issue deserves a special discussion since it is not always properly understood.

The structure of the cuts of the functions $h_i(q_0)$ is shown in Fig. 5. The part accessible in the decay channel of the B mesons covers the interval $[0, M_B - M_D]$. The dispersion integral (58) can be written as a sum of two integrals,

$$h_i(q_0) = \frac{1}{2\pi}\int_0^{M_B-M_D} \frac{w_i(\tilde{q}_0)d\tilde{q}_0}{\tilde{q}_0 - q_0} + \frac{1}{2\pi}\int_A \frac{w_i(\tilde{q}_0)d\tilde{q}_0}{\tilde{q}_0 - q_0} \tag{59}$$

where the domain A consists of two subdomains, $q_0 < 0$ (A_1) and $q_0 > M_B + M_D$ (A_2). For real decays we are interested only in the first integral since the second one, rather than describing the $B \to X_c e\nu$ decay, refers to other physical processes. The subdomain A_1 actually describes a similar $b \to c$ amplitude, yet with negative q_0, and can be called the lower cut. The integral over A_2, on the other hand, will be referred to as the integral over the distant cuts. Two kinds of problems are encountered in evaluating the total dispersion integral. The first one emerges due to the fact that for real decay kinematics one has only $q_0 > 0$; therefore, say, for calculating the total width one has not the integral over the whole physical cut, but needs to consider the smaller interval without the subdomain A_1. The corresponding problems of separating the contribution of the same type of intermediate states, but at different values of q_0 are usually referred to as "local" duality. In the context of the present paper this is however not very important. For in our sum rules, from purely theoretical point of view, it does not matter whether a particular transition can be measured in real experiment, or not; e.g. the lower bounds we will discuss rely only on the positivity of the corresponding transition probabilities.

There is generally another complication assocciated with the integral over the distant cuts (subdomain A_2) corresponding to quite different intermediate states. The problem of isolating these contributions can be generically referred to as "global" duality.

Both contributions thus represent a contamination for real decays. Fortunately, this contamination is irrelevant for our analysis.

Indeed, to address the contamination due to the "lower" cut, let us choose the "reference" point of q_0 between the cuts (see Fig. 5), close to $M_B - M_D$,

$$q_0 = M_B - M_D - \epsilon \tag{60}$$

where ϵ is a negative number,

$$-\epsilon \gg \Lambda_{QCD}\ ,\ \ -\epsilon \ll M_D .$$

When calculating the functions h_i at the quark level from the operator product expansion we get a similar dispersion relation for the OPE coefficients (with the meson masses replaced by the quark ones). We then use duality concepts in identifying the physical relevance of these cuts. *Local* duality of QCD means that there is a one-to-one correspondence between the part of the OPE coefficients originating from the lower cut and the corresponding hadronic contribution in the phenomenological (hadronic) representation of h_i. The validity of local duality can be verified by itself by choosing q_0 close to the particular remote cut, the lower one for the case at hand. Therefore, we can systematically discard the contribution of that cut simultaneously, in the theoretical expression for h_i and in the "phenomenological" saturation. In this way we arrive at the relations

$$\int_0^{m_b-m_c} w_i^{quark} \epsilon^n d\epsilon = \int_0^{M_B-M_D} w_i \epsilon^n d\epsilon \equiv E_n, \qquad (61)$$

where ϵ is the same variable as defined in eq. (60), but on the cut it is positive. It represents the excitation energy. The left hand side of eq. (61) includes the perturbative corrections as well as the power-like non-perturbative terms. The local duality we have invoked to discard the contribution from this "lower" cut has an accuracy of the type $\exp\{-\,\mathrm{Const}\cdot m_c/\Lambda_{QCD}\}$. It is worth emphasizing that it is the ratio m_c/Λ_{QCD}, not m_b/Λ_{QCD} that enters. In the real world m_c/Λ_{QCD} numerically is not so large, and since the constant in the exponent is unknown one may be afraid of an insufficient accuracy of the local duality for D's. At present theory provides us with no clues as to the value of the constant in the exponential; the degree of possible violations of the local duality should be established empirically. With this caveat in mind we still believe that heavy quark expansion must work well in $B \to X_c e\nu$.

Analogous analysis can be repeated almost verabatim for the contribution of the distant cuts to address the question of "global duality". This duality is even more transparent physically and is explicit in all perturbative calculations and for calculations in "soft" external fields. On the other hand, in principle, its accuracy is generally determined by the same factor depending on m_c, namely $\exp\{-\,\mathrm{Const}\cdot m_c/\Lambda_{QCD}\}$.

After these more general remarks we return to the concrete calculations of the theoretical part of the sum rules. To find w_i^{quark} we take the discontinuity of the transition operator eq.(55),

$$\frac{1}{i}\mathrm{disc}\,\hat{T}_{\mu\nu} = -2\pi\bar{b}\Gamma_\mu(k_0\gamma_0 + m_c + \not{\pi})\times$$

$$\sum_{n=0}^{\infty}\frac{1}{n!}\delta^{(n)}(k_0 - m_c)\frac{(2k_0\pi_0 + \pi^2 + (i/2)\sigma G)^n}{(m_c + k_0)^{n+1}}\Gamma_\nu b \ . \qquad (62)$$

Using the QCD equations of motion

$$\pi_0 b = -\frac{1}{2m_b}(\pi^2 + \frac{i}{2}\sigma G)b \ ,$$

$$\pi_0 b = -\vec{\pi}\vec{\gamma}\,\gamma_0 b + m_b(\gamma_0 - 1)b \qquad (63)$$

in the leading non-trivial approximation one then has (we suppress for simplicity the indices μ and ν that mark the dependence of the moments E_n on the weak currents considered)

$$\frac{1}{2\pi}E_0 = -\bar{b}\Gamma_\mu \left(\frac{1+\gamma_0}{2} - \frac{\vec{\pi}\vec{\gamma}}{2m_c} + \frac{\pi^2 + \frac{i}{2}\sigma G}{2m_c^2}\right) \Gamma_\nu b + \mathcal{O}(\Lambda^3_{QCD}/m^3)$$

$$\frac{1}{2\pi}(E_1 - E_0^{el}\cdot\delta) = -\bar{b}\Gamma_\mu \left(\frac{1+\gamma_0}{2}(-\pi_0 - \frac{\pi^2+\frac{i}{2}\sigma G}{2m_c}) - \frac{\pi_0^2}{2m_c}\gamma_0 - \frac{\pi_0\gamma_0(\pi^2+\frac{i}{2}\sigma G)}{4m_c^2} + \frac{\vec{\pi}\vec{\gamma}}{2m_c}(\pi_0 + \frac{\pi^2+\frac{i}{2}\sigma G}{2m_c})\right) \Gamma_\nu b + \mathcal{O}(\Lambda^4_{QCD}/m^3)$$

$$\frac{1}{2\pi}E_n = -\bar{b}\Gamma_\mu \left\{\frac{1+\gamma_0}{2}\left((\frac{1}{2m_b} - \frac{1}{2m_c})^2 \cdot (\vec{\pi}^2 - \frac{i}{2}\sigma G)\,(-\pi_0)^{n-2}\,(\vec{\pi}^2 - \frac{i}{2}\sigma G) + (\frac{1}{2m_b} - \frac{1}{2m_c})\frac{\vec{\pi}\vec{\gamma}}{2m_c}\pi_0\,\pi_0^{n-2}\,\frac{i}{2}\sigma G\right)\right\} \Gamma_\nu b + \mathcal{O}(\Lambda^{n+3}_{QCD}/m^3) \qquad (64)$$

where the last equation refers to $n \geq 2$; matrix elements over the initial hadron state are assumed here on the *rhs* , and m generically denotes both m_c and m_b. In the second equation we have introduced E_0^{el}, which is the elastic contribution to E_0 and the mass shift δ

$$\delta = \epsilon - \epsilon_0 = (M_B - M_D) - (m_b - m_c) \qquad (65)$$

that determines the difference in the threshold energy between real hadrons at zero recoil and the quark mass difference, *i.e.* the direct zero recoil analog of Δ in eq.(16). Note that this shift is irrelevant in the first sum rule for E_0, and at $n \geq 2$ it enters only when considering higher order power corrections.

The structure of the solution of eqs.(64) for the excitation function $w(\epsilon)$ can be deduced from the fact that $w(\epsilon)$ is positive (for appropriate $\Gamma_{\mu,\nu}$). The first equation tells us that the sum of all probabilities equals unity up to small corrections $\sim \mathcal{O}(1/m^2)$. On the other hand, higher moments all start with the terms of order $\Lambda^n_{QCD}\cdot(\Lambda_{QCD}/m)^2$. Because the scale for the excitation energies ϵ_i is given by Λ_{QCD} [5] one immediately concludes that the probabilities of transitions to the excited states all scale like $1/m^2$ with the heavy quark masses. To saturate the first sum rule one then needs state(s) which do not contribute to the higher moments E_n and are produced with practically unit probability; the only way to satisfy it is to have the final states, D and D^* with masses

$$M_B - (m_b - m_c)$$

[5] Effectively the same refers even to the thresholds assocciated with the $D^{(*)}$ + pion(s) states in the chiral limit, which strictly speaking have no such excitation gap. This holds true owing to the fact that the corresponding amplitudes are proportional to the pion momentum; in reality they can produce only chiral logs and do not change powers of mass in the analysis, see Ref. [38].

up to corrections that vanish in the limit $m_q, m_Q \to \infty$. Moreover, their transition amplitudes are to be unity up to terms inversely proportional to the *square* of the heavy quark masses – a fact observed originally by M.Shifman and M.Voloshin [22] and now known as Luke's theorem [32]. In this way one obtains the statement of flavour symmetry in the spectrum of hadrons. More specifically, considering the zeroth component of the vector current one selects the transitions to D whereas for $\gamma_i\gamma_5$ vertices only D^* contributes; thus one can derive separate relations for both states and amplitudes.

Considering the second of eqs.(64) corresponding to $n = 1$, one can infer the following: the contribution of excited states leads only to the terms of order Λ^3_{QCD} and they can be discarded for contributions of order Λ^2_{QCD}/m. On the other hand, the *rhs* does contain terms of this order: for example for the vector current ($\Gamma_\mu = \Gamma_\nu = \gamma_0$) they are given by

$$(\frac{1}{2m_c} - \frac{1}{2m_b})\bar{b}(\vec{\pi}^2 - \frac{i}{2}\sigma G)b \ . \tag{66}$$

Because the correction to $E_0 = 1$ starts with terms quadratic in $1/m$, this relation unambiguously determines the $1/m$ shift in the mass of heavy flavor hadrons, B and D in the case under consideration. Therefore the leading, $1/m$ power correction to the effective Lagrangian of heavy quarks is obtained; considering, for example, the transitions induced by the axial current one obtains the known relation between the hyperfine splitting and the matrix element of the chromomagentic operator.

It is then easy to obtain the general expression for the function $w(\epsilon)$ as it emerges *via* its moments E_n. The inelastic part of $w(\epsilon)$ appears at the $1/m^2$ level and is given by the Fourier transform of the time dependent correlation functions of the leading power correction operators in the effective Lagrangian:

$$\epsilon^2 w(\epsilon) \simeq (\frac{1}{2m_b} - \frac{1}{2m_c})^2 \times$$

$$\int \frac{dt}{2\pi} e^{it\epsilon} \langle B|\bar{b}(t, \vec{x} = 0)\ \Gamma_\mu(\vec{\pi}^2 + \vec{\sigma}\vec{B}) \mathrm{T}\, e^{i\int_0^t A_0(t')dt'}\ (\vec{\pi}^2 + \vec{\sigma}\vec{B})\ \Gamma_\nu\ b(0)|B\rangle \tag{67}$$

where B denotes the chromomagnetic field. Let us note that the explicit terms containing $\vec{\pi}\vec{\gamma}$ in eq.(64) generate operators including also the chromoelectric field; in reality they appear to cancel the effects due to the propagation of the lower components of the quark fields caused by the chromoelectric operator contained in the full Lorentz scalar combination σG; in the final expression only the chromomagnetic part survives in the leading approximation.

It is worth clarifying at this point the physical meaning of the path ordered exponents appearing in the OPE in higher orders. Generically we obtain expressions of the form

$$\langle H_b|\bar{b}(x)\ o_1\ \mathrm{P}\, e^{i\int_0^x A_\mu(x')dx'_\mu}\ o_2\ b(0)|H_b\rangle \quad , \quad x^2 > 0 \ . \tag{68}$$

One can rewrite 'P exp' introducing an auxilliary "ultraheavy" quark Q field using the *operator* relation

$$\mathrm{P}\, e^{i\int_0^x A_\mu(x')dx'_\mu} = Q(x)\,\bar{Q}(0)\;\; e^{-im_Q|x|} \tag{69}$$

where the mass $m_Q \to \infty$ and in particular is assumed to be much larger than the renormalization point (ultraviolet cutoff) in the theory (the nonrelativistic normalization of Q is used). Note that m_Q is the *bare* mass of this ultraheavy quark appearing in its bare Lagrangian. The c-number exponential factor can be conveniently removed if one factors out the "mechanical" momentum of this auxilliary Q field to arrive at the fields Q' routinely used in HQET; this is done identifying the heavy quark velocity v with the classical velocity of a particle moving from point 0 to x:

$$Q(z) = e^{-im_Q vz} Q'(z) \quad , \quad v_\mu \equiv \frac{x_\mu}{|x|} \;\; .$$

With these notations any nonlocal matrix element of the type eq. (68) can be rewritten as

$$\langle H_b|\bar{b}(x)\; o_1\; \mathrm{P}\, e^{i\int_0^x A_\mu(x')dx'_\mu}\; o_2\; b(0)|H_b\rangle = \langle H_b|\bar{b}\, o_1 Q'(x)\; \bar{Q}'(0)\, o_2 b(0)|H_b\rangle =$$

$$= \sum_n \langle H_b|\bar{b}\, o_1 Q'(x)|n\rangle \cdot \langle n|\bar{Q}'(0) o_2 b(0)|H_b\rangle \; e^{i(p_{H_b}-p_n)x} \;\; . \tag{70}$$

The second expression in this equation is manifestly gauge invariant. Moreover, in the third representation we have re-expressed it as a sum over the intermediate states that can be created from H_b by the current $\bar{b}oQ$, that is, the corresponding states in the true effective theory with infinitely heavy quark Q instead of b (it clearly does not matter at this point whether the initial, b quark is considered as heavy, or with a finite mass).

Eq. (70) clearly reveals which meaning is to be attributed to nonlocal correlators of operators with path exponents in a quantum mechanical picture. It shows one that he is actually dealing with the correlator of the two currents of the type $\bar{b}Q$ and $\bar{Q}b$. Therefore in the intermediate state one merely has – instead of real beauty hadrons – hadrons propagating that contain an infinitely heavy quark *without* any finite mass corrections. To phrase it differently, we come back to the similar $b \to c$ transitions but effectively remove all $1/m_c^k$ corrections from the propagation between the two vertices.

If one uses this physical interpretation, eq. (67) acquires a very transparent meaning: it is nothing but the second order term in the expansion due to the perturbation given by the operator

$$\left(\frac{1}{2m_b} - \frac{1}{2m_c}\right)\;\; \bar{c}\,(\vec{\pi}^2 + \vec{\sigma}\vec{B})\, b \tag{71}$$

in the effective theory. The Fourier transform of the Green function consists of δ-functions (for discrete intermediate states) residing at energy values ϵ_i defined by the mass differences $M_{B^i} - M_B$ between the ground and excited states. This equation expresses the fact that the transition amplitudes at zero recoil are given by the matrix elements of this operator between the initial and intermediate state. Apart from this non-local effect, there are also local terms renormalizing the weak current, given by the first sum rule. The latter can be viewed as the power terms in the commutator of the currents of the effective theory due to the finite values of the heavy quark masses. It is not difficult to trace the straightforward similarity between the derivation given above and the analysis of the heavy quark distribution function in the SV limit carried out in Ref. [14] that lead to an analogous temporal function. In that case, however, inelastic excitations appeared proportional to the velocity of the final quark, and it was the correlator of the space-like momentum operators $\vec{\pi}$ that emerged.

Let us make a few comments. Obviously, one can, in principle, calculate higher order power corrections to the sum rules eq. (64). It is possible to see that in what concerns the inelastic contributions to the structure function, $w(\epsilon)$ will order by order reproduce the successive terms of the ordinary quantum mechanical perturbation theory corresponding to both next order iterations of the leading terms as well as to perturbations representing subleading power operators in the effective heavy quark Lagrangian.

Similarly, one can consider the sum rules even in the case of nonzero recoil when $\vec{q}$ does not vanish. Most simply it is done in the SV limit $|\vec{q}\,| \ll m_q$ when one can expand in $|\vec{v}| \simeq |\vec{q}\,|/m_q$. In the leading power approximation and keeping terms quadratic in $|\vec{v}|$ it has been done in Ref. [14] and the leading power corrections are incorporated in previous sections. Generally, one can go beyond the small velocity kinematics as long as $|\vec{q}\,|/m_q$ does not scale with the mass. However, in this case one can obtain only a limited number of symmetry relations between different spin amplitudes because the unknown function of the velocity determined by the infinite series of the matrix elements of powers of operators πq appears already in the leading $1/m$ approximation.

To conclude this section, let us say a few words clarifying why this general analysis based on the sum rules for the heavy flavor transitions is thought to be important although it may seem not to lead immediately to new results beyond the picture of standard perturbation theory in quantum mechanics.

First, we shall see in subsequent sections that this approach allows one to obtain useful bounds on the transition amplitudes and on the 'static' matrix elements governing nonperturbative corrections.

Secondly, it demonstrates in a transparent and unambiguous way the necessity to introduce the infrared renormalization point for addressing nonperturbative effects and clarifies both its physical meaning and the qualitative trend of the dependence on μ. This is important in view of the existing negligence of this problem in applications of HQET.

Last, but not least, we clearly see here that such assumptions about QCD as the validity of concepts of global duality discussed above, that sometimes are naively thought to be specific only for the OPE-based approach for inclusive transitions, are in reality the most general requirement inherent in any consistent consideration. For if for some particular physical reason the quark-hadron correspondence used above were modified – for example there were a sizeable "leak" in dispersion integrals from distant cuts, it would immediately lead to new contributions in the *rhs* of the sum rules; this would result in physically observable corrections not contained in the expansions obtained in HQET.

3.2 A bound on the form factor at zero recoil from the sum rules

We consider, for definiteness, transitions of the type $B \to D^*$ and $B \to$ excitations of the vector mesons. Practically they are most important in the exclusive approach to the problem of determination of $|V_{cb}|$. These transitions are induced by the axial-vector current

$$A_\mu = \bar{c}\gamma_\mu\gamma_5 b \ .$$

It is quite obvious that it is most convenient to focus on the spatial component of this current.

The T-product

$$h_{\mu\nu} = i\int d^4x e^{-iqx}\frac{1}{2M_B}\langle B|T\{A^\dagger_\mu(x)A_\nu(0)\}|B\rangle \tag{72}$$

is calculated, up to $\mathcal{O}(\Lambda^2_{QCD})$ terms [10, 11, 12]; this accuracy is sufficient for our purposes. In general the amplitude $h_{\mu\nu}$ is decomposed into a sum of five terms, see eq. (57). For the spatial components of the current we need to consider only h_1.

To get the first sum rule in the zero recoil point we put

$$\vec{q} = 0 \ .$$

In this problem the lowest lying state produced in the decay is D^*; therefore the expansion parameter ϵ is naturally introduced as follows:

$$\epsilon = M_B - M_{D^*} - q_0 \ , \tag{73}$$

analogously to what was done in Sect. 2.5. If our aim is to obtain the first sum rule, to accuracy $\mathcal{O}(\Lambda^2_{QCD})$, the definition (73) can be replaced by a simplified expression,

$$\epsilon \to \Delta m - q_0 \ .$$

As previously, we will expand h_1 in Λ_{QCD}/ϵ and in $\epsilon/m_{Q,q}$.

Eq. (A.1) of Ref. [10] implies that

$$-h_1 = \frac{1}{\epsilon} - \frac{\mu_G^2 - \mu_\pi^2}{2m_b}\left(\frac{1}{3} - \frac{m_c}{m_b}\right)\frac{1}{\epsilon(2m_c+\epsilon)} +$$

$$\left[\frac{4}{3}\mu_G^2 m_b - (\mu_G^2 - \mu_\pi^2)q_0\right] \frac{1}{\epsilon^2(2m_c+\epsilon)m_b} \quad , \tag{74}$$

where μ_G^2 and μ_π^2 parametrize the matrix elements of the chromomagnetic and kinetic energy operators,

$$\mu_G^2 = \frac{1}{2M_{H_Q}}\langle H_Q|\bar{Q}\frac{i}{2}\sigma_{\mu\nu}G^{\mu\nu}Q|H_Q\rangle \;\; ; \;\; \mu_\pi^2 = \frac{1}{2M_{H_Q}}\langle H_Q|\bar{Q}\vec{\pi}^2 Q|H_Q\rangle \;\; .$$

The chromomagnetic parameter is known experimentally:

$$\mu_G^2 \simeq \frac{3}{4}(M_{B^*}^2 - M_B^2) = 0.37\,\mathrm{GeV}^2 \;\; .$$

Expanding in Λ_{QCD}/ϵ and in $\epsilon/m_{b,c}$ and keeping only the term linear in $1/\epsilon$ we find

$$-\,h_1 \rightarrow \frac{1}{\epsilon}\left[1 - \frac{1}{3}\frac{\mu_G^2}{m_c^2} - \frac{\mu_\pi^2 - \mu_G^2}{4}\left(\frac{1}{m_c^2} + \frac{1}{m_b^2} + \frac{2}{3m_c m_b}\right)\right] \quad . \tag{75}$$

The sum rule stemming from eq. (75) obviously takes the form

$$F_{B\to D^*}^2 + \sum_{i=1,2,\ldots} F_{B\to excitations}^2$$

$$= \left[1 - \frac{1}{3}\frac{\mu_G^2}{m_c^2} - \frac{\mu_\pi^2 - \mu_G^2}{4}\left(\frac{1}{m_c^2} + \frac{1}{m_b^2} + \frac{2}{3m_c m_b}\right)\right], \tag{76}$$

where the sum on the left-hand side runs over all excited states and all form factors are taken at the zero recoil point. This is a perfect analog of eq. (47). The form factor $B \to D^*$ at zero recoil is defined as

$$\langle B|A_i|D^*\rangle = \sqrt{4M_B M_{D^*}}\; F_{B\to D^*}\; \epsilon_i \;\; .$$

If all terms $\mathcal{O}(\Lambda_{QCD}^2)$ are switched off higher states cannot be excited at zero recoil – only the elastic $B \to D^*$ transition survives – and we arrive at the well-known result that

$$F_{B\to D^*} = 1, \quad (\text{zero recoil}) \;\; ,$$

the statement of the heavy quark (or Isgur-Wise [33]) symmetry first noted in the SV limit in Ref. [22] (see also [34]). Including $\mathcal{O}(\Lambda_{QCD}^2)$ terms we start exciting higher states; all transition form factors squared are proportional to Λ_{QCD}^2/m^2. Simultaneously the form factor of the elastic transition shifts from unity.

The power correction on the right-hand side is negative. What is crucial is the fact that the contribution of the excited states is strictly positive. Transferring them to the right-hand side we arrive at the following lower bound

$$1 - F_{B\to D^*}^2 > \frac{1}{3}\frac{\mu_G^2}{m_c^2} + \frac{\mu_\pi^2 - \mu_G^2}{4}\left(\frac{1}{m_c^2} + \frac{1}{m_b^2} + \frac{2}{3m_c m_b}\right). \tag{77}$$

Thus, we find that the (absolute value of) the deviation of the elastic form factor $F_{B\to D^*}$ from unity at zero recoil is definitely larger than

$$\frac{M_{B^*}^2 - M_B^2}{8m_c^2} \sim 0.035 \ .$$

As a matter of fact, the second term on the right-hand side of eq. (77) is also positive, $\mu_\pi^2 > \mu_G^2$. We will rederive this inequality within the framework of sum rules itself in the next section. Previously it was obtained in this form by M.Voloshin [35] who extended the quantum-mechanical argument of Ref. [14] applicable at a low normalization point. This inequality is in perfect agreement with the most refined QCD sum rule calculation of μ_π^2 [36], according to which

$$\mu_\pi^2 = 0.54 \pm 0.12 \ \mathrm{GeV}^2 \ .$$

If this estimate is accepted then the second term in eq.1-F amounts to $\sim 1/2$ of the first one, and the lower bound for the deviation becomes $1 - F_{B\to D^*}^2 > 0.1$. The actual deviation is probably larger by ~ 0.1. First, the sum rule derived above neglects perturbative α_s corrections. The first order correction to the elastic form factor was calculated in Ref. [22]. If in zeroth order in α_s the $\bar{b}\gamma_\mu\gamma_5 c$ axial-vector vertex at zero recoil is unity, the first order correction renormalizes it to

$$\eta_A = 1 + \frac{\alpha_s}{\pi}\left(\frac{m_b + m_c}{m_b - m_c}\ln\frac{m_b}{m_c} - \frac{8}{3}\right) .$$

The renormalization-group improvement of this calculation (including the leading and the first subleading logs of the ratio of m_b/m_c) has been carried out in Ref. [37]. For the axial-vector current the perturbative correction is negative, so that unity in eq. (76) is replaced by 0.96. Then, the contribution of the excited states in eq. (76) is strictly positive, and this also reduces $F_{B\to D^*}^2$. This contribution may be as large as, roughly, the power correction on the right-hand side. An estimate of the excited state contribution supporting this statement is given in Ref. [38] where a more detailed numerical discussion of all corrections is given. Notice that in our approach the excited state contribution replaces a non-local contribution of Ref. [39].

We conclude that $1 - F_{B\to D^*}^2$ is definitely larger than 0.1, somewhat beyond the window obtained in Ref. [31]. The phenomenological impact of this observation is discussed in Ref. [38].

Let us note that in order to get the second sum rule analogous to eq. (48) we would need to know $\mathcal{O}(\Lambda_{QCD}^3)$ terms both in the transition operator and in the relation between ΔM and Δm. In $\hat{T}_{\mu\nu}$ these terms are due to the four-quark operator while in $\Delta M - \Delta m$ they come from two sources: a local contribution from the four-quark operator and a non-local one analogous to μ_3^3 of Sect. 2.5. Classification of $\mathcal{O}(\Lambda_{QCD}^3)$ terms in $M - m$ is discussed in Ref. [39].

In a very similar way one can obtain the bound and estimate for the vector form factor of the $B \to D$ transition $F_{B\to D}$ at zero recoil. Here only the timelike component of the current contributes, and for this reason the full semileptonic decay amplitude is proportional to the lepton masses. Therefore this mode is not advantageous; the corresponding formfactor is measurable (in principle) at zero recoil in the $B \to D + \tau\nu_\tau$ decays. Taking $\Gamma_\mu = \Gamma_\nu = \gamma_0$ we obtain for this case the sum rule

$$F^2_{B\to D} + \sum_{i=1,2,\ldots} F^2_{B\to excit} = 1 - \frac{\mu_\pi^2 - \mu_G^2}{4}\left(\frac{1}{m_c} - \frac{1}{m_b}\right)^2 \quad . \tag{78}$$

Perturbative corrections also differ and now look as follows [22]:

$$1 + \frac{\alpha_s}{\pi}\left(\frac{m_b + m_c}{m_b - m_c}\ln\frac{m_b}{m_c} - 2\right) \quad . \tag{79}$$

The corrections in the case of the vector formfactor obviously vanish at $m_b = m_c$, as they have to in view of the exact conservation of the current in this limit. Numerically therefore they are expected to be smaller for vector transitions than for axial ones.

It is worth mentioning that the excitation probabilities entering the sum rules eq.(76) and eq.(78) are generated separately by the axial or the vector current, respectively, but not by the $V-A$ current that directly produces the experimental widths. Actually at zero velocity transfer the axial and vector currents cannot interfere. Therefore for the $V-A$ semileptonic transitions into massless leptons one has just to add to eq.(76) (assuming that no final state identification is attempted) the contribution of the $\bar{c}\gamma_i b$ current. The sum rule for this current is obtained in the next subsection. Combining the two sum rules one gets

$$F^2_{B\to D^*} = \eta_A^2 - \frac{\mu_\pi^2 - \mu_G^2}{3m_b m_c} - \int_{\epsilon > M_{D^*} - M_D} d\epsilon\,\frac{w_1}{2\pi} \tag{80}$$

where the last term representing the inelastic contribution is expressed via the differential semileptonic width at zero recoil:

$$\frac{w_1}{2\pi} = \frac{8\pi^2}{G_F^2|V_{cb}|^2 q_0^2|\vec{q}|}\cdot\frac{d^2\Gamma_{SL}}{d\vec{q}^{\,2}\,dq_0}\Big|_{\vec{q}=0,\; q_0 = M_B - M_D - E} \tag{81}$$

(q is the momentum of the lepton pair in the process). Eq.(80) is much less useful as an upper bound because the main part of the correction to the elastic formfactor is to come from the presently unknown contribution from the excitations.

3.3 Lower bound on μ_π^2

Reversing the line of reasoning used to derive the upper bound on the formfactors, we can exploit the very same idea to get a constraint on the matrix elements μ_G^2 and μ_π^2. At zero recoil there are only two independent structure functions for the correlator of the $V-A$ currents; similar functions can be introduced for other weak vertices as well. Choosing a particular current one projects out a certain combinations of the structure functions. It is important that for the Hermitian conjugated currents in the two weak vertices one gets a definitely positive structure function expressed as a sum over certain transition probabilities. If an appropriate channel is found where the elastic peak is kinematically absent, the theoretical side of the corresponding sum rule will contain no unity term, and start with the leading $1/m^2$ corrections given just by a linear combination of μ_G^2 and μ_π^2. The "phenomenological" side, a sum over the excited states, is positive-definite. In this way we arrive at a constraint on the linear combination of μ_G^2 and μ_π^2 at hand.

It is not difficult to find a specific example. Indeed, let us consider the vector current, $\bar{q}\gamma_\mu Q$, and in particular its *space-like* components. It is obvious that $-(1/3)h_{ii}^{VV} = -h_1^{VV}$ is populated only by the excited states – if the initial state H_Q is the ground state pseudoscalar, the final one, H_q^*, must be the axial-vector meson, non-degenerate with H_Q in the symmetry limit. (In the quark model language one would say that H_q^* is a P-wave state.) Again, using the results of Ref. [10] we find

$$-\,h_1^{VV}(q_0) = \frac{1}{\epsilon}\left[\frac{\mu_\pi^2-\mu_G^2}{4}\left(\frac{1}{m_Q^2}+\frac{1}{m_q^2}-\frac{2}{3m_Q m_q}\right)+\frac{\mu_G^2}{3}\frac{1}{m_q^2}\right] \quad . \tag{82}$$

The expression in the square brackets is equal to the sum over excitations:

$$\left[\frac{\mu_\pi^2-\mu_G^2}{4}\left(\frac{1}{m_Q^2}+\frac{1}{m_q^2}-\frac{2}{3m_Q m_q}\right)+\frac{\mu_G^2}{3}\frac{1}{m_q^2}\right] = \sum_i F^2_{H_Q\to H_q^*} \quad , \tag{83}$$

and, hence, is always positive, for any values of m_Q and m_q. Being interested in the "static" properties of the initial state only, it is convenient, according to the comment in the end of Sect. 2, to consider the theoretical limit $m_q \gg m_Q$ where only initial state effects survive [6]. Requiring the positivity of eq. (82) at $m_q \gg m_Q$ we conclude that

$$\mu_\pi^2-\mu_G^2 = \mathrm{Const}\cdot\sum_i |F_{H_Q\to H_q^*}|^2 \tag{84}$$

and, therefore,

$$\mu_\pi^2 > \mu_G^2 \quad . \tag{85}$$

This is literally the same inequality that has been obtained previously [35] in the quantum-mechanical language along the lines suggested in Ref. [14]. The argument presented above can be viewed as a consistent and transparent field-theoretic reincarnation. According to the general discussion of Sect. 2.4 in reality one is to introduce the cutoff in the "phenomenological" integral of the decay probabilities from the upper side of excitation energies ϵ. It is most important that the integral in the *rhs* of eq. (84)

$$\sum_{\epsilon_i\le\mu} |F_{H_Q\to H_q^*}|^2$$

[6] The very same bound can be obtained considering, say, the correlator of two $i\gamma_5$ currents, or even merely when both vertices are $1-\gamma_0$. In either case one gets directly the difference $\mu_\pi^2-\mu_G^2$ with the coefficients $1/m_c^2$ and $(1/m_c-1/m_b)^2$, respectively.

is positive for *any* normalization point μ, and therefore ensures the validity of the inequality eq. (85) for operators normalized at arbitrary values of μ, provided the renormalization point is consistently introduced in this particular way. For high enough μ the contribution of the excited states is given by perturbative expressions and the μ dependence is explicitly calculable; of course for very large μ the inequality becomes trivial. We shall discuss it in more detail in Sect. 4.

It is instructive to note that the zero recoil sum rule for the $V_i \times V_i$ transitions (or similar ones where there is no elastic peak) provides us with the most direct way to determine the evolution of the kinetic energy operator. One proceeds here along the same way of reasoning as has been outlined in Sect. 2.4 where we considered the small velocity kinematics with $\vec{q} \neq 0$. In this case the elastic peak identically vanishes for purely kinematic reasons. Therefore the only possible impact of introducing the infrared normalization point *via*, say, the gluon mass can emerge (at least to lowest nontrivial order in the strong coupling) from the kinematic constraint on the emitted gluon.

To use this way to determine the evolution of $\vec{\pi}^2$ one needs to calculate the perturbative probability to emit the gluon at zero recoil [7]. The corresponding amplitude is clearly given by the Thompson scattering amplitude of nonrelativistic particle (with obvious modification due to different coupling constants in the case at hand), which is the same for real spinor and hypothetical scalar quarks [8]. The exact calculation of this perturbative contribution to the dispersion integral in the sum rule eq. (83) is trivial; accounting for the explicit factors we obtain

$$\frac{d}{d\mu^2}\bar{Q}\,(i\vec{D})^2\,Q \simeq \frac{4\alpha_s(\mu)}{3\pi}\,\bar{Q}Q \tag{86}$$

where we assumed $m_q = m_Q$ to obtain the exact coefficient [9]. This result coincides with eq. (34).

Throughout this paper we phrased our discussion of real QCD in terms of the transitions where initial states were heavy flavor *mesons*, namely B. It is clear that exactly the same reasoning can be applied for the transitions of heavy baryons, for example when the initial hadron is Λ_b. The matrix elements, of course, are different.

[7]It indeed defines renormalization of only the kinetic operator, not of the chromomagnetic one. To see that one can, for example, consider the sum rule applied to Λ_b decays where the chromomagnetic operator vanishes for any renormalization point. On the other hand, by definition, perturbative mixing does not depend on the particular state and the matrix element of the leading operator $\bar{Q}Q$ is the same in Λ_b as in B. In other words, when taking matrix element in eq. (82) over the perturbative states of quasifree heavy quarks $\langle\mu_G^2\rangle = 0$.

[8]This ensures that such defined renormalization does not depend on the heavy quark spin, which seems to be mandatory in the effective low energy theory.

[9]This is necessary to have the dispersion integral in full QCD well behaved asymptotically; at $m_q \neq m_Q$ the current $\bar{q}\gamma_\mu Q$ is not conserved and asymptotic behavior of the amplitude deteriorates reflecting the nontrivial commutator of currents; the integral over asymptotically large momenta can get then additional μ^2 dependence. From practical viewpoint the change in the coefficient is minor, by a factor 4/3 only in the whole range of the quark mass ratio.

In particular, the expectation value of the chromomagnetic operator vanishes for Λ_b. Moreover, contrary to the meson case no non-trivial lower bound on the kinetic term emerges. Therefore it is naturally to expect smaller deviations from the symmetry limit for both vector and axial formfactors in baryons than for mesons.

An analysis of corrections to $F_{B\to D^*}$ resembling our in spirit, but not technically, has been carried out recently in Ref. [39]. There $1 - F_{B\to D^*}$ is expressed in terms of some local and non-local expectations values; the latter are unknown. In our analysis the role of the non-local expectation values is played by the contribution of the excited states. What is crucial, this contribution is always positive. In [39] $1 - F_{B\to D^*}$ is found to be positive (good!), and a numerical estimate is presented relating $1 - F_{B\to D^*}$ to μ_π^2, a parameter that is somewhat more uncertain than μ_G^2. As a matter of fact, the numerical values of μ_π^2 accepted in [39] under the influence of some recent claims are in contradiction with the inequality (85).

3.4 The second sum rule at $\vec{q} \neq 0$; measuring $\overline{\Lambda}(\mu)$

If at $\vec{q} = 0$ the second sum rule requires the knowledge of $\mathcal{O}(\Lambda_{QCD}^3)$ terms, at $\vec{q} \neq 0$ (i.e. $\Lambda_{QCD} \ll |\vec{q}\,| \ll M_D$) a non-trivial prediction, an analog of Voloshin's sum rule [23], arises to order Λ_{QCD}. Higher order corrections will be briefly discussed later. We need to consider the average value of $q_{0max} - q_0$. This quantity is related to $M_{X_c}^2$, the average invariant mass squared of the hadronic system produced [10]. Indeed,

$$M_{X_c}^2 = M_D^2 + 2M_B(q_{0max} - q_0) \quad , \quad q_{0max} = \frac{M_B^2 - M_D^2 + q^2}{2M_B}. \tag{87}$$

All elements necessary for the derivation of the sum rule are available; we will use the expressions for $h_{\mu\nu}$ obtained in [10] including terms $\mathcal{O}(\Lambda_{QCD}^2)$. The basic idea is the same as that demonstrated in Sect. 2.4 in the toy model: to order Λ_{QCD} in the SV limit the weighted integral over the excited states is proportional to $\overline{\Lambda}(\mu)$. If in the previous section we considered the point of no recoil; now we have to shift from this point and consider terms proportional to the square of the c quark velocity. The SV limit will be ensured by choosing $|\vec{q}\,| \ll M_D$.

Let us assume first that all structure functions in eq. (57) are known separately. Then it is most convenient to consider

$$(-1/3)h_{ii} \equiv -h_1 \; .$$

If we are aimed at effects linear in Λ_{QCD} all μ_π^2 and μ_G^2 corrections can be neglected and (see eq. (A.7) of Ref. [10])

$$-\,h_1 = 2\frac{m_b - q_0}{(m_b + m_c - q_0)(m_b - m_c - q_0) - \vec{q}^{\,2}} \quad , \tag{88}$$

[10]The corresponding analysis of the average invariant hadronic mass presented in Ref. [4] was incorrect, see Ref. [41]. The average invariant mass of the final state hadrons is *not* given directly by local operators at the level of nonperturbative corrections, contrary to claims in Ref. [4].

where the hadronic tensor considered is that induced by the $V - A$ current. The parameter ϵ, the distance from the pole, must be defined now as

$$\epsilon = M_B - M_D - \frac{\vec{q}^{\,2}}{2M_D} - q_0 = m_b - m_c - \frac{\vec{q}^{\,2}}{2m_c} + \frac{\vec{q}^{\,2}}{M_D^2}\frac{\overline{\Lambda}}{2} - q_0 + \dots \ . \tag{89}$$

Notice that working to first order in Λ_{QCD} formally there is no need to differentiate between the masses of D and D^*; this difference actually can give us an idea about the numerical impact of $1/m_c^2$ corrections. If the second order corrections are included it becomes important.

Next, we express the right hand side of eq. (88) in terms of ϵ and expand in $1/\epsilon$ and ϵ/m_q. The term of interest is $\mathcal{O}(\overline{\Lambda}/\epsilon^2)$. There is only one such term in the expansion,

$$-\,h_1 = \dots + \frac{\vec{q}^{\,2}}{M_D^2}\frac{\overline{\Lambda}}{2}\frac{1}{\epsilon^2} + \dots \ . \tag{90}$$

This result is to be confronted with the hadronic state representation for $-h_1$. In this way we obtain

$$\frac{1}{2\pi}\int_{q_{0max}-\mu}^{q_{0max}} (q_{0max} - q_0)\,dq_0\, w_1 = \frac{\vec{q}^{\,2}}{M_D^2}\frac{\overline{\Lambda}(\mu)}{2} \tag{91}$$

where $\vec{q}$ is supposed to be fixed ($|\vec{q}\,| \ll M_q$), and $w_1 = 2\mathrm{Im}\, h_1$ is a structure function that is measurable, in principle.

To reiterate, we introduce a normalization point μ in such a way that all frequencies smaller than μ can be considered as "soft" or inherent to the bound state wave function; at the same time $\alpha_s(\mu)$ has to be sufficiently small for the perturbative expansion in $\alpha_s(\mu)/\pi$ to make sense. We then draw a line at $q_0 = q_{0max} - \mu$ (the picture is similar to that of Fig. 4, with $d\Gamma/dE$ replaced w_1 and E by q_0). The integral eq. (91) taken over the range $q_{0max} - \mu$ to q_{0max} represents $(v^2/2)\overline{\Lambda}(\mu)$ modulo corrections of higher order in v and in Λ_{QCD}. The running mass is then defined as $m_b(\mu) = M_B - \overline{\Lambda}(\mu)$.

Practically it may be not so easy to separate different structure functions from each other. A similar prediction can be given for double differential distribution in the semileptonic decay.

Using eqs. (8), (A.7) and (A.8) from Blok *et al.* [10] one arrives at

$$\int (q_{0max} - q_0)\,dq_0 \frac{d^2\Gamma}{dq_0 dq^2} = \overline{\Lambda}\,\frac{(\Delta M)^2 - q^2}{2M_B^2}\,\frac{d\Gamma}{dq^2} \tag{92}$$

plus terms of the second and higher order in $\overline{\Lambda}$. Here

$$\Delta M = M_B - M_D \ ,$$

$$\frac{d\Gamma}{dq^2} = \frac{G_F^2|V_{cb}|^2}{8\pi^3 m_b}\left\{ q^2\,\frac{m_b^2 + m_c^2 - q^2}{2m_b}|\vec{q}\,| + \frac{2}{3}m_b|\vec{q}\,|^3 \right\} \ , \tag{93}$$

where

$$|\vec{q}\,| = \frac{(m_b^4 + m_c^4 + q^4 - 2m_b^2 q^2 - 2m_b^2 m_c^2 - 2q^2 m_c^2)^{1/2}}{2m_b} = m_c(2\xi + \xi^2)^{1/2} \ . \qquad (94)$$

Equation (92) is valid for arbitrary values of q^2, not necessarily close to the maximal value $q^2_{max} = (\Delta M)^2$. In deriving this expression we have used the fact that to order Λ_{QCD} $\Delta M = \Delta m$ where $\Delta m = m_b - m_c$. In the SV limit, when the right hand side of eq. (51) is small, the results for the radiative corrections obtained in the previous section are directly applicable to the semi-leptonic decays. Indeed, the gluon can be emitted (absorbed) either by the color charge of the c quark or by its magnetic moment. Moreover, there is no interference – the gluon emitted by the magnetic moment has to be absorbed by the magnetic moment. It is not difficult to check that the maximal value of the gluon momentum $k \sim m_c\xi$ and, as long as we are retaining terms of order ξ only, we can disregard the gluon interaction with the magnetic moment. Then we are left with the charge interaction only which is the same for spin-0 bosons and spin-1/2 fermions. As a result, the expression for the radiative correction obtained previously is modified in a minimal way, only due to a slightly different kinematics, and we get

$$\int_{\sqrt{q^2}}^{q_{0max}} (q_{0max} - q_0) dq_0 \frac{d^2\Gamma}{dq_0 dq^2} =$$

$$= \left\{ \overline{\Lambda}(\mu) + \frac{16\alpha_s}{9\pi} \int_{\sqrt{q^2}}^{q_{0max}-\mu} dq_0 \frac{(q_0^2 - q^2)^{3/2}}{(q_{0max}^2 - q^2)^{1/2}} \frac{1}{2\xi} \frac{M_B^2}{M_D^4} \right\} \cdot \ \xi \frac{M_D}{M_B} \frac{d\Gamma^{(0)}}{dq^2} \qquad (95)$$

plus terms of order ξ^2 and higher. In the $\mathcal{O}(\alpha_s)$ term in eq. (95) we do not distinguish between the quark and meson masses.

Thus, the sum rules eq. (91) or eq. (95) can be used to elucidate what is actually meant by the heavy quark mass. This question is rather subtle since the heavy quark mass is a purely theoretical parameter which is not directly measurable. On the other hand, it is a very important parameter, crucial in a wide range of questions.

The sum rules (91) or (95) express $\overline{\Lambda}$ in terms of the integral over the physically observable quantities. Therefore, it is tempting to say that we, thus, have a suitable definition of $\overline{\Lambda}$ and, through this quantity, the heavy quark mass. Of course, both of them depend explicitly on the renormalization point.

In a similar manner one can use an appropriate sum rule to determine the kinetic energy operator. For example, it can be extracted in a model-independent way from the sum rule similar to eq. (33) if the double differential measurements are used and one can select small velocity events:

$$\mu_\pi^2(\mu) = 3\Gamma^{-1} v^{-2} \int_{q_{0_{max}}^{phys} - \mu}^{q_{0_{max}}^{phys}} \frac{d^2\Gamma}{dq_0 dq^2} \, (q_{0_{max}} - q_0)^2 \, dq_0 \qquad (96)$$

where the "quark velocity" v, the normalization Γ and the the maximum energy release $q_{0_{max}}^{phys}$ are now defined as

$$v^{-2} = \frac{(m_b^2 + m_c^2 - q^2)^2}{(m_b^2 + m_c^2 - q^2)^2 - 4m_b^2 m_c^2}$$

$$q_{0_{max}}^{phys} = \frac{M_B^2 + M_D^2 - q^2}{2M_B}$$

$$\Gamma = \int_{q_{0_{max}}^{phys} - \mu}^{q_{0_{max}}^{phys}} \frac{d^2\Gamma}{dq_0 dq^2}\, dq_0 \quad . \tag{97}$$

The integration in eq. (96) must run over the energy release q_0 close to the parton values $q_0 \lesssim m_b - m_c - \mu$ whereas the integration over q^2 must include only the region of small v. The normalization in eq. (97) must be calculated over the same region, of course. This determines the value of μ_π^2 normalized at point μ.

On the contrary, selecting events with minimal q^2 one could try to invoke an approximation neglecting the deviation of the distribution function from the light cone one [14, 17, 18]. However this seems not to be safe enough from the theoretical side. The normalization point dependence would enter then in a more complicated way.

4 Impact of the perturbative evolution of effective operators

In this section we shall briefly discuss the practical modifications that arise in the calculation of nonperturbative effects in heavy flavour decays if one accounts for the normalization point dependence (in the infrared region) of the corresponding operators. Although this question is rather standard, we feel the need to dwell on it in view of the apparent confusion taking place in applications of heavy quark expansions existing in the literature.

It is well understood (see, for example, the recent discussion in Ref. [25]) that the consistent incorporation of nonperturbative effects in the framework of the Wilson OPE procedure requires introduction of the real separation of the contributions of momenta below and above a certain normalization point μ. The transparent physical necessity of such separation was demonstrated above on a few simple examples. The low momentum physics is then attributed to the matrix elements of the operators whereas the high momentum piece enters Wilson coefficients; both therefore depend explicitly on μ in such a way that observables are μ-independent.

This fact is always taken into account when the corresponding operators undergo logarithmic renormalization, for an obvious reason: the Feynman integrals determining the coefficient functions in this case logarithmically diverge and one merely cannot put μ, the infrared regularization, to zero. In calculating the power corrections to heavy flavour decays another situation can typically arise: when the leading operator has zero (logarithmic) anomalous dimension, but there are subleading operators of higher dimension that mix with the lower dimensional operators. It is a rather common situation for the case of zero recoil (or small velocity) transitions and for observables related to the inclusive widths. In this case the coefficient functions of the leading operators possess a safe infrared limit; they are then usually calculated without an infrared cutoff, which is equivalent to setting the normalization point to zero. This procedure is known to lead, from a theoretical perspective, to grave problems even within the perturbative expansion itself in higher orders, which manifest themselves, for example, as infrared renormalons (see, e.g. the recent discussion in Ref. [25] and references therein to the original publications). However to finite number of loops, and in particular for typically existing one or two loop calculations, this kind of problems may not be noticed.

It is obvious, on the other hand, that this naive approach to perturbative calculation of Wilson coefficients contains physically (but not necessary numerically relevant) flaw in that it leads in general to a double counting of some contributions. Let us take, as an example, the typical calculation of the corrections to the zero recoil formfactor of $B \to D^{(*)}$ transition; another example is the calculation of the corrections to the total semileptonic B decays discussed in detail in Ref. [25]. The usual procedure for dealing with such problems is to take, say, known expressions for the one loop perturbative corrections for the particular quantity and add to them nonperturbative corrections expressed in terms of certain matrix elements. In this way, for example, the perturbative correction η_A to the the axial current $\bar{c}\gamma_\mu\gamma_5 b$ was simply added to the nonperturbative contribution of eq. (77). It is clear however that the nonperturbative contribution *per se* takes care of *all* relevant gluon exchanges with momenta below μ; still the one loop Feynman integral for the radiative correction has some (small) piece coming from very low momenta – that strictly speaking must be excluded.

It is natural to expect that, in reality, the extra contribution in the perturbative integral is *numerically* smaller than the nonperturbative ("condensate") effects and, therefore, the error made in such procedure is not very important for a reasonable choice of normalization point μ. This assumption constitutes in fact the conceptual grounds for using the so-called "practical version" of OPE in QCD where such effects are neglected. It is not, however, an obvious *a priori* property of QCD and is to be cross-checked in any new situation. In principle one should calculate the Wilson coefficients by evaluating Feynman integrals with an explicit infrared cutoff in the propagators to avoid these problems.

What if we still use in practice concrete expressions that effectively correspond to $\mu = 0$? Let us consider for example the *rhs* of the sum rule eq. (76) with the perturbative corrections added:

$$F^2_{B\to D^*} + \sum_{i=1,2,...} F^2_{B\to excit} = \eta_A^2 - \frac{1}{3}\frac{\mu_G^2}{m_c^2} - \frac{\mu_\pi^2 - \mu_G^2}{4}\left(\frac{1}{m_c^2} + \frac{1}{m_b^2} + \frac{2}{3m_c m_b}\right) \quad , \qquad (98)$$

and neglect for simplicity the spin of the heavy quark, so that only the kinetic energy operator $\vec{\pi}^2$ will represent non-perturbative corrections. Then, in the consistent procedure, calculating the perturbative correction factor η_A we need to introduce the infrared renormalization point $\mu \ll m_c$. In practice, instead, one can analyze the dependence on μ of the matrix element of the operator $\vec{\pi}^2$ – for the whole expression is μ-independent!

Using only this observation, we can conclude that the dependence of the perturbative factor η_A on μ is quadratic. Say, in the one loop approximation it is to take the form

$$\eta_A(\mu/m_c) = \eta_A(0) + c_{\eta_A}\frac{\alpha_s}{\pi}\frac{1}{m_c^2}\,\mu^2 \;+\mathcal{O}(\mu^3/m_c^3) \;\; ; \tag{99}$$

no term linear in μ can appear because the first subleading operator has a dimension higher by two units than the leading one, $\bar{Q}Q$. Moreover, the coefficient c_{η_A} must match the one determining the powerlike mixing of operator $\bar{Q}\vec{\pi}^2Q$ with $\bar{Q}Q$, eq. (34):

$$\frac{d}{d\mu^2}\bar{Q}\,(i\vec{D})^2\,Q = \beta_\pi(\alpha_s(\mu))\bar{Q}Q \;\simeq c_\pi\frac{\alpha_s(\mu)}{\pi}\,\bar{Q}Q \;\; . \tag{100}$$

The scale dependence of the strong coupling in these equations is not essential in one loop calculations; if higher loops are accounted for, the sum of all Feynman integrals with the particular infrared cutoff μ will automatically have suitable form to give logs of the ratio of m_c/μ necessary to convert $\alpha_s(m_c)$ appearing in the perturbative calculations, into $\alpha_s(\mu)$ (more exactly, into the series in $\alpha_s(\mu)$ entering the renormalization group evolution eq. (100)).

We can now use eq. (99) to correct the initial relation eq. (98) to account for the proper μ dependence of the perturbative term. The result is very simple: μ_π^2 entering it must be merely understood as

$$\mu_\pi^2 \to \mu_\pi^2(\mu) - \mu^2\frac{d\mu_\pi^2}{d\mu^2} \simeq \mu_\pi^2(\mu) - c_\pi\frac{\alpha_s(\mu)}{\pi}\cdot\mu^2 \;\; . \tag{101}$$

The expression on the *rhs* is nothing but a linear (in μ^2) extrapolation of the matrix element from the point μ to $\mu = 0$. Formally, it is independent on μ: the dependence appears in terms proportional to α_s^2. Therefore one can take any value of $\Lambda_{QCD} \ll \mu \lesssim m_c$. The modification for the case of higher loop calculations is trivial: if the perturbative corrections are calculated with an infrared cutoff at zero through order n, the matrix elements of higher order operators must be extrapolated to $\mu = 0$ using the corresponding number of terms in the $\alpha_s(\mu)$ expansion of the β-function $\beta(\alpha_s)$ determining the powerlike mixing of the operator:

$$\mu_\pi^2 \to \mu_\pi^2(\mu) - \sum_{k=1}^{n}\alpha_s^k(\mu)\int_0^{\mu^2} d\lambda^2\;a_k(\lambda^2/\mu^2) \tag{102}$$

if the perturbative expansion of the powerlike evolution equation for the operator takes the form

$$\beta_\pi(\alpha_s(\lambda)) = \sum_{k=1}^{\infty} a_k(\lambda^2/\mu^2)\alpha_s^k(\mu) \;\; . \tag{103}$$

Note that the simple "renormalon" calculation corresponds only to the expansion of the one loop expression for the running coupling $\alpha_s(\lambda)$ in terms of $\alpha_s(\mu)$, therefore it leads to

$$a_k(\lambda^2/\mu^2) = \frac{c_\pi}{\pi}\left(\frac{b}{2\pi}\log\frac{\lambda}{\mu}\right)^{k-1} \quad .$$

The integrals of powers of logs in eq. (102) produce then factorials $k!$ and this generates the $1/\mu^2$ renormalon that would lead to uncertainty $\sim \Lambda_{QCD}^2$ in high loops. Generalization of the rule eq. (102) to the case of operators of a different dimension is straightforward.

The prescription eq. (102) is formally sensible at any *finite* number of loops n; in reality it is only reasonable for one or two loops, depending on the value of the strong coupling $\alpha_s(\mu)$: numerically the series start to blow up and would produce large, though finite values $\gg \Lambda_{QCD}^2$ for the extrapolated matrix elements at large n differing dramatically from order to order. It does not happen at all if one does not attempt to extrapolate to a low normalization point as it is assumed in the original Wilson approach. In practical applications limited to lowest loop accuracy it seems to be one of convenient *technical* ways to formulate the result. It appears that routinely the operators relevant for the heavy quark expansions (including such quantity as the heavy quark mass m_Q, see e.g. [35]) are given just in this form. Needless to say that then it must be specified in each particular case to what order in α_s the concrete value is obtained – for otherwise the result loses any sense at the level of nonperturbative effects. It is also clear that in extrapolating results to $\mu = 0$ one should consistently use the *same* order in perturbation theory for all quantities involved.

Now let us turn from this rather general theoretical discussion to more practical questions related to heavy flavour decays. Up to now non-perturbative effects have been discussed in detail through corrections of order $1/m_Q^3$. Some effects, like invariant mass of the final hadronic state in the decays considered above, have corrections starting at order $1/m_Q$ and are expressed via the parameter $\overline{\Lambda}$. As pointed out in Ref. [25] and illustrated in the present paper, their effects are determined by Feynman integrals which have a linear behavior in the infrared region; the corresponding IR effects are *not* expressed in terms of matrix elements of any local operator.

The inclusive widths of heavy flavour particles *are* obtained from an expansion in local operators, and corrections start with terms scaling like $1/m_Q^2$. These are described by two universal operators – chromomagnetic one $\bar{Q}\,\frac{i}{2}\sigma G\,Q$ and the kinetic energy operator, $\bar{Q}\,(i\vec{D})^2\,Q$. The natural normalization scale for them is given by $\mu \simeq m_Q$, at least if one neglects the mass of the final state quark(s) [11]. One cannot however use directly this high normalization point because then the matrix elements of high dimension operators will scale like m_Q to the corresponding power due to purely perturbative contributions, and instead of an expansion in $1/m_Q$ one would

[11] It is important that for the hyperfine splitting in heavy mesons which allows one to extract experimentally the chromomagnetic matrix elements, the same normalization point emerges.

obtain the suppression of the higher order terms only as some powers of $\alpha_s(m_Q)$. To obtain the real power expansion one has to evolve these operators down to a scale μ which is to be much smaller than m_Q but still much larger than Λ_{QCD}. The expansion one arrives at in this way runs, strictly speaking, in powers of μ/m_Q

The present state of the art in this kind of calculations is limited only to one loop corrections to Wilson coefficients which – apart from the chromomagnetic operator that has a logarithmic renormalization – are calculated (or even typically borrowed from the old QED calculations) without an infrared cutoff. The same refers to the corrections to weak currents used in the present paper where the relevant ultraviolet normalization point is given by m_c. The analysis above suggests therefore that the value of the kinetic energy term μ_π^2 is to be understood in the corresponding expressions as a linear extrapolation to $\mu = 0$. This problem does not arise at all for the chromomagnetic operator O_G whose mixing with the leading one appears only in the next order in $1/m_Q$ due to the fact that it is not a spin singlet; its contribution is absent in the perturbative integral and no double counting occurs for inclusive widths. From practical viewpoint, because the value of μ_π^2 is basically unknown up to now, it does not make a big difference at present to prefer this or an alternative definition. Let us note in passing that it is quite probable that the QCD sum rule estimates determine a similar quantity extrapolated (linearly) to $\mu = 0$ because no *explicit* infrared cutoff in the integrals is introduced; though this question definitely deserves a more careful analysis.

The issue of an accurate understanding of the definition of matrix elements of operators becomes most important when one turns to the real practical bounds on physical observables of the type discusses in this paper. For the extrapolation to the zero renormalization point, for example, of the operator $\bar{Q}\,\vec{\pi}^2\,Q$ implies a subtraction of a positive quantity that, in principle, might have even changed the sign of the matrix element. Even more essentially, in deriving our model-independent lower bound on $F_{B\to D^*}$ we included into the final number the contribution of the perturbative corrections part of which (obviously small though) is already contained in the matrix element μ_π^2. To state it differently, one may be concerned whether the inequality eq. (85) survives the extrapolation of μ_π^2 to a low point as has been assumed in our reasoning. We shall argue now that this effect is too small numerically and cannot upset the bound we used.

To see it, let us consider the reasonably high normalization point $\mu \simeq 1\,\mathrm{GeV}$. Using the explicit estimate of the renormalization point dependence of operator $\bar{Q}\,\vec{\pi}^2\,Q$ in eq. (34) and assuming the one loop value $\alpha_s(\mu) \simeq .36$ one readily obtains that the amount one may need to subtract from $\mu_\pi^2(\mu)$ constitutes at most $0.15\,\mathrm{GeV}^2$, a value that does not exceed the theoretical uncertainties in the existing estimates of μ_π^2. Most probably this number overestimates the real contribution to be subtracted, because approximate duality of the perturbative corrections is expected to start earlier, and, on the other hand, the perturbative corrections are calculated numerically using a smaller value of α_s. The second effect, though formally of higher order in α_s, is too transparent physically to raise doubts that more realistic estimate corresponds to using $\alpha_s(m_c)$ rather than $\alpha_s(1\,\mathrm{GeV})$ above.

At the same time, as emphasized in Sect. 3.2, if the normalization point is introduced *via* the upper bound in the integral over the energy of the excited states, the inequality $\mu_\pi^2 > \mu_G^2$ holds for any normalization point. In other words, this regularization does not violate the positivity of the Pauli operator for the spinor quark. Therefore, it is legitimate to take the normalization point as low as 1 GeV. In this case, obviously, one deals with μ_G^2 normalized at this low point as well, and it is known that the perturbative evolution *increases* its value toward lower μ! In our estimates we took μ_G^2 directly from the hyperfine splitting of B and B^*, therefore that value corresponded to $\mu \simeq 4.5\,\mathrm{GeV}$. Apparently its hybrid log enhancement would safely make up for the relatively insignificant subtraction of the "perturbative" contribution to μ_π^2. Based on these arguments we have stated in the previous paper [38] that the inequality

$$\mu_\pi^2 > \mu_G^2$$

that holds in beauty mesons must survive renormalization effects, in spite of recent claims [42] that it cannot hold true.

It is instructive to trace how this inequality works at different scales μ. Most trivially it is fulfilled when μ is taken parametrically large. Then μ_π^2 contains large *positive* perturbative piece of the order of $\frac{\alpha_s}{\pi}\mu^2$ that grows faster than any possible change in μ_G^2 having no additive renormalization, even if the hybrid anomalous dimension of the latter were negative.

A more interesting consideration emerges when one wants to push μ toward lower values. Using the naive one loop expression for the evolution of μ_G^2 corresponding to the hybrid anomalos dimension $\gamma_G = 3$ one would obtain an arbitrary large value for μ_π^2 which, of course has little sense. The answer to this apparent paradox is rather obvious, especially if one looks at the hypothetic zero recoil excitation curve (for the external current $\bar{c}\gamma_i b$) similar to the one depicted in Fig. 4. The exact evolution of the difference $\mu_\pi^2 - \mu_G^2$, according to the sum rule eq. (83), is given by the decay probability occurring at energy $\epsilon = \mu$; obviously the latter in no way is given at low ϵ by the simple perturbative formulae using the strong coupling α_s with the Landau pole, and rather stays finite at any μ. In other words, the exact evolution of μ_G^2 is to be smooth even when one approaches the strong coupling regime, and no formal contradiction emerges.

It is worth emphasizing here that, of course, the exact form of the β-function for all operators has no definite sense at high orders depending on the used renormalization scheme, therefore in general speculations about the evolution in the strong coupling regime are not very meaningful. However, fixing a particular scheme makes it well defined theoretically, and choosing a physical definition of the cutoff like the one adopted in the present paper allows one to conclude that the corresponding 'exact' evolution must be non-singular.

5 Conclusions

In the present paper we have addressed weak transitions between heavy quarks from the "inclusive" side most suitable theoretically for applying the technique of the Wilson OPE. This analysis represents the natural extension of consideration outlined in Ref. [14] concerning the heavy quark distribution function relevant for the decays in the limit of small velocity for the final state hadron system. It has been demonstrated that a few sum rules discussed so far in the literature are in fact some dispersion relations for the moments of a single distribution function, considered in different orders in $1/m_Q$ and in different kinematics. We have shown how the expansions of HQET can consistently be obtained from QCD using this strategy, and in this way illustrated that such natural assumptions as "global duality", that usually are attributed only to the inclusive width calculations, are in fact necessary ingredients in *any* consistent model-independent treatment and, in particular, are implicitly used in HQET as well.

The analysis of the sum rules proved to be very instructive in elucidating the important fact that has been usually neglected in HQET – the necessity of introducing an explicit infrared normalization point μ ensuring true separation of low and high momentum physics, which *cannot* be set to zero. The consistent application of this approach leads to the fact that all nonperturbative parameters, including $\overline{\Lambda}$, cannot be sensibly defined as universal constants, but rather depend explicitly on the normalization point. This has been previously mentioned in our paper [14] and discussed in detail in Refs. [25, 26]. Here we gave a physical illustration of how it works analyzing possible constructive phenomenological definitions of corresponding quantities in the presence of radiative corrections. In this way we supplemented the previous calculations by estimates of the dependence of the kinetic energy operator $\bar{Q}\,\vec{\pi}^2\,Q$ on the renormalization point.

The fact that such "purely nonperturbative" objects like the pole mass of the heavy quark, $\overline{\Lambda}$, "purely non-pertubative" distribution function of heavy quarks routinely used in HQET are incompatible with the consistent OPE-based approach and are ill-defined theoretically calls for the clarification of how the known results on non-perturbative corrections in HQET must be interpreted. This does not mean of course that the concrete calculations that have been done so far are irrelevant, and we formulated the way in which they are to be understood for a few typical examples.

As a practical application of our sum rules we have derived a model independent lower bound on the deviation of the exclusive axial form factor $F_{B\to D^*}$ of the $B \to D^* + l\,\nu$ decay at zero recoil – a process that for a long time has been believed to give the best theoretical accuracy to determine $|V_{cb}|$, and estimated a reasonable 'central' value, $F_{B\to D^*}(\vec{q}=0) \simeq 0.9$. The deviation appears to be essentially larger than the estimates that had been obtained before from model calculations based on standard HQET, and apparently better agree with quite general expectations about the size of corrections to the Heavy Quark symmetry for charmed particles. On the other hand, the theoretical clarification of the notion of the heavy quark mass made in recent papers [25, 26] suggests that the most accurate theoretically way to determine the CKM matrix elements for heavy quark decays is using the *inclusive* semileptonic widths. These results have been reported in Ref. [38].

It is worth clarifying in this respect that in our estimates of the exclusive form factors of the $b \to c$ transitions we consistently took into account terms through order $1/m_Q^2$ and discarded effects that scale like $1/m_Q^3$. The parameter $1/m_c$ is actually not very small and even the second order corrections are as large as 10% here; therefore one can expect sizeable relative corrections for real form factors due to higher order terms. In particular, this applies to the model-independent upper bound for $F_{B\to D^*}$. Our result is strict in the sense that it holds for corrections through terms of order $1/m_Q^2$ that have been addressed in the literature so far.

One of the sum rules at zero recoil enabled us to rederive a model independent lower bound on the value of the kinetic energy operator in B mesons,

$$\mu_\pi^2 = \frac{1}{2M_B}\langle B|\bar{b}\,\vec{\pi}^2\, b|B\rangle \gtrsim \frac{3}{4}\left(M_B^2 - M_{B^*}\right)$$

in a way that clearly showed its physical relevance even in real QCD, and not only in the approximate framework of quantum mechanical consideration, as it is sometimes stated. Moreover, we believe that using the corresponding sum rule eq. (83) (or similar ones mentioned in Sect. 3.3) that expresses the difference $\mu_\pi^2 - \mu_G^2$ as a dispersion integral over the excited states can provide one with the most promising theoretical method to calculate, by means of QCD sum rules, this matrix element, which is very important for heavy flavor physics.

ACKNOWLEDGEMENTS: N.U. gratefully acknowledges interesting discussions with M.Voloshin about charm and beauty quark masses and the creative atmosphere and hospitality at Theoretical Physics Institute of University of Minnesota, as well as the exchange of ideas related to the subject of this paper with V.Braun. He also thanks T.Mannel, A.Falk and M.Savage for the discussion of standard HQET calculations. This work was supported in part by the National Science Foundation under the grant number PHY 92-13313 and by DOE under the grant number DOE-AC02-83ER40105.

Figure Captions

Fig. 1. The tree graph for the transition operator.

Fig. 2. A qualitative picture of the spectrum $d\Gamma/dE_\phi$ in the $H_Q \to \phi X_q$ with $\mathcal{O}(v^2)$ terms included. The monochromatic line of the quark transition $Q \to \phi q$ (the dashed line at $E = E_0$) is dual to the physical line corresponding to the elastic decay $H_Q \to H_q \phi$ at $E = E_0^{phys}$ plus a shoulder due to the transitions to the excited states $H_Q \to \phi H_q^*$. The height of the shoulder is $\sim v^2$. Hard gluons are neglected.

Fig. 3. The diagram responsible for the one-gluon correction in the energy distribution

$$\frac{d\Gamma(Q \to \phi q + \ gluon)}{dE_\phi}.$$

Fig. 4. A sketch of the energy spectrum $d\Gamma/dE_\phi$ with $\mathcal{O}(\alpha_s)$ radiative tail included.

Fig. 5. The cuts of $h_i(q_0)$ in the complex q_0 plane. Each of the three points, $q_0 \simeq M_B + M_D$, $q_0 = M_B - M_D$ and $q_0 \simeq -(3M_B + M_D)$, marks the beginning of two cuts. Only a part of one cut at $0 < q_0 < M_B - M_D$ is relevant to the decay process under consideration.

References

[1] M. Shifman, M. Voloshin, 1982, see in V. Khoze, M. Shifman, *Uspekhi Fiz. Nauk* **140** (1983) 3 [*Sov. Phys. Uspekhi* **26** (1983) 387].

[2] N. Bilić, B. Guberina, J. Trampetić, *Nucl. Phys.* **B248** (1984) 261;
M. Voloshin, M. Shifman, *Yad. Fiz.* **41** (1985) 187 [*Sov. Journ. Nucl. Phys.* **41** (1985) 120]; *ZhETF* **91** (1986) 1180 [*Sov. Phys. – JETP* **64** (1986) 698].

[3] M. Voloshin, N. Uraltsev, V. Khoze and M. Shifman, *Yad. Fiz.* **46** (1987) 181 [*Sov. Journ. Nucl. Phys.* **46** (1987) 112].

[4] J. Chay, H. Georgi, B. Grinstein, *Phys. Lett.* **B247** (1990) 399.

[5] I.I. Bigi and N.G. Uraltsev, *Phys. Lett.* **280B** (1992) 120.

[6] I. Bigi, N. Uraltsev and A. Vainshtein, *Phys. Lett.* **B293** (1992) 430; (E) B297 (1993) 477.

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

[8] I. Bigi, B. Blok, M. Shifman, N. Uraltsev and A. Vainshtein, *The Fermilab Meeting* Proc. of the 1992 DPF meeting of APS, C. H. Albright *et al.*, Eds. World Scientific, Singapore 1993, vol. 1, page 610.

[9] I. Bigi, M. Shifman, N. Uraltsev and A. Vainshtein, *Phys. Rev. Lett.* **71** (1993) 496.

[10] B. Blok, L. Koyrakh, M. Shifman and A. Vainshtein, *Phys. Rev.* **D49** (1994) 3356.

[11] A. Manohar and M. Wise, *Phys. Rev.* **D49** (1994) 1310.

[12] T. Mannel, *Nucl. Phys.* **B413** (1994) 396.

[13] M. Neubert, *Preprint* CERN-TH.7087/93.

[14] I. Bigi, M. Shifman, N. Uraltsev and A. Vainshtein, TPI-MINN-93/60-T [*Int. Journ. Mod. Phys. A,* ***9*** *(1994)*, to appear].

[15] A. Falk, E. Jenkins, A. Manohar and M. Wise, *Preprint* UCSD/PTH 93-38.

[16] M. Neubert, *Preprint* CERN-TH.7113/93.

[17] I. Bigi, M. Shifman, N. Uraltsev and A. Vainshtein, *Preprint* CERN-TH.7159/94; [to be published in *Phys. Lett. B*].

[18] T. Mannel and M. Neubert, Preprint CERN-TH.7156/94.

[19] J. Bjorken, in *Proceedings of the 4th Rencontres de Physique de la Vallee d'Aosta*, La Thuille, Italy, 1990 ed. M. Greco (Editions Frontières, Gif-Sur-Yvette, France, 1990).

[20] N. Isgur and M. Wise, *Phys. Rev.* **D43** (1991) 819.

[21] J.D. Bjorken, I. Dunietz and J. Taron, *Nucl. Phys.* **B371** (1992) 111.

[22] M. Voloshin and M. Shifman, *Yad. Fiz.* **47** (1988) 801 [*Sov. J. Nucl. Phys.* **47** (1988) 511].

[23] M. Voloshin, *Phys. Rev.* **D46** (1992) 3062.

[24] H. Lipkin, *Phys. Lett.* **B308** (1993) 105; *Nucl. Phys.* **A560** (1993) 548.

[25] I. Bigi, M. Shifman, N. Uraltsev and A. Vainshtein, *Preprint* CERN-TH.7171/94, TPI-MINN-94/4-T [to be published in *Phys. Rev. D*].

[26] M. Beneke and V. Braun, *Preprint* MPI-PhT/94-9.
M. Beneke, V. Braun and V. I. Zakharov, *Preprint* MPI-PhT/94-18.

[27] For a review see e.g. B.L. Ioffe, V.A. Khoze and L.N. Lipatov, *Hard Processes*, North-Holland, Amsterdam, 1984.

[28] K. Wilson, *Phys. Rev.* **179** (1969) 1499;
K. Wilson and J. Kogut, *Phys. Reports* **12** (1974) 75.

[29] A.G. Grozin, *Preprint* BudkerINP 92-97 (1992), Part 1.

[30] H. Albrecht *et al.* (ARGUS Collab.), *Z. Phys.* **C57** (1993) 533;
D.Z. Besson, Talk at 16th International Symposium on Lepton-Photon Interactions, Cornell, August 1993;
S.Stone, "Semileptonic B Decays", Second Edition. *To be published in* **World Scientific**, Singapore.

[31] A. Falk and M. Neubert, *Phys. Rev.* **D47** (1993) 2965.

[32] M.E. Luke, *Phys. Lett.,* **B252** (1990) 447.

[33] N. Isgur and M. Wise, *Phys. Lett.,* **B232** (1989) 113; *Phys. Lett.,* **B237** (1990) 527.

[34] S. Nussinov and W. Wetzel, *Phys. Rev.* **D36** (1987) 130.

[35] M. Voloshin, to be published.

[36] P. Ball and V. Braun, *Phys. Rev.* **D49** (1994) 2472;
for earlier estimates see
M. Neubert, *Phys. Rev.* **D46** (1992) 1076; E. Bagan, P. Ball, V. Braun and H. Dosch, *Phys. Lett.* **B278** (1992) 457.

[37] M. Neubert, *Phys. Rev.* **D46** (1992) 2212.

[38] M. Shifman, N. Uraltsev and A. Vainshtein, *Preprint* TPI-MINN-94/13-T.

[39] T. Mannel, Preprint CERN-TH.7162/94

[40] Ref. [6]; for more details see the preprint version: I. Bigi, N. Uraltsev and A. Vainshtein, Preprint FERMILAB-PUB-92/158-T June 1992.

[41] I. Bigi, N. Uraltsev, Preprint CERN-TH.7020/93 (accepted for publication in *Nucl.Phys.* ***B***).

[42] M. Neubert, unpublished; see also M. Neubert *Phys. Lett.,* **B322** (1994) 419.