The Thomas Jefferson National Accelerator Facility
Theory Group Preprint Series

Additional copies are available from the authors.

The Southeastern Universities Research Association (SURA) operates the Thomas Jefferson National Accelerator Facility for the United States Department of Energy under contract DE-AC05-84ER40150.

JLAB-THY-98-03

revised 28 August 1998

Duality in Inclusive Semileptonic Heavy Quark Decay

Nathan Isgur

Jefferson Lab

12000 Jefferson Avenue, Newport News, Virginia 23606

Abstract

I identify a source of Λ_{QCD}/m_Q corrections to the assumption of quark-hadron duality in the application of heavy quark methods to inclusive heavy quark semileptonic decays $Q \to q\ell\bar{\nu}_\ell$. These corrections could substantially affect the accuracy of such methods in practical applications and in particular compromise their utility for the extraction of the Cabibbo-Kobayashi-Maskawa matrix element V_{cb}.

Although the classic application of heavy quark symmetry is in the exclusive semileptonic decays of heavy quarks [1], there has also been substantial work on using heavy quark effective theory (HQET) [2] to systematically improve decay predictions for inclusive semileptonic decay rates of heavy hadrons induced by an underlying $Q \to q\ell\bar{\nu}_\ell$ decay [3–5]. In these inclusive applications, decays are treated in an operator product expansion (OPE) which leads *via* HQET to a $1/m_Q$ expansion in which the leading term is free quark decay, $1/m_Q$ terms appear to be absent, and terms of order $1/m_Q^2$ take into account such effects as the difference $\bar{\Lambda} \equiv m_B - m_b$ between m_B (the mass of a $\bar{B}$ meson) and m_b (the mass of a b quark), the nonzero kinetic energy $(-\lambda_1/2m_b)$ of the heavy quark, small residual spin-dependent interactions $(2\lambda_2/m_b \equiv m_{B^*} - m_B)$ of the heavy quark, *etcetera*. Although these calculations have become very sophisticated [4,5], it is widely appreciated [6,7] that there remains a basic unproved hypothesis in their derivation: the assumption of quark-hadron duality. It is the accuracy of this assumption that I want to call into question here.

While supposedly valid for any semileptonic decay $Q \to q\ell\bar{\nu}_\ell$ of a heavy quark Q, recent applications have centered around the hope that this approach offers an alternative to the classic exclusive methods for determining V_{cb}, and I will accordingly focus most of my remarks on the case $b \to c\ell\bar{\nu}_\ell$ where both quarks are heavy. In inclusive $b \to c\ell\bar{\nu}_\ell$ decays, which materialize as $\bar{B} \to X_c\ell\bar{\nu}_\ell$, about 65% of the X_c spectrum is known to be due to the very narrow ground states D and D^*. The relatively narrow $s_\ell^{\pi_\ell} = \frac{3}{2}^+$ states [8] $D_2^*(2460)$ and $D_1(2420)$ account for perhaps another 5% of the rate, and it may be assumed that the remaining rate involves decays to higher mass resonances (quarkonia and hydrids) and continua [9]. The HQET-based inclusive calculations predict continuous X_c spectra which are assumed to be dual to the true hadronic spectrum (see Fig. 1).

A picture like Fig. 1 might lead one to dismiss the duality approximation since the inclusive spectrum clearly does not meet the usual requirement that it be far above the resonance region [10]. *I.e.*, normally the accuracy of quark-hadron duality would be determined by a parameter Λ_{QCD}/E where the relevent energy scale E is the mean hadronic excitation energy $\Delta m_{X_c} \equiv \bar{m}_{X_c} - m_D$. However, as first explained by Shifman and Voloshin [11,12], this

is *not* the expansion variable in this case: duality for heavy-to-heavy semileptonic decays sets in *at threshold* since even as $\delta m \equiv m_b - m_c$ (and therefore Δm_{X_c}) approaches zero, as $m_b \to \infty$ the heavy recoiling c quark has an energy much greater than Λ_{QCD} so that its final state interactions can be ignored. In the small velocity (SV) limit, it *must* therefore hadronize with unit probability (up to potential Λ_{QCD}/m_Q corrections) as D and D^*. This "cannonball" approximation is in fact an essential part of the physical basis of the HQET expansion in $1/m_Q$. Thus the issue is not whether duality holds in semileptonic heavy quark decays, but rather how accurately it holds.

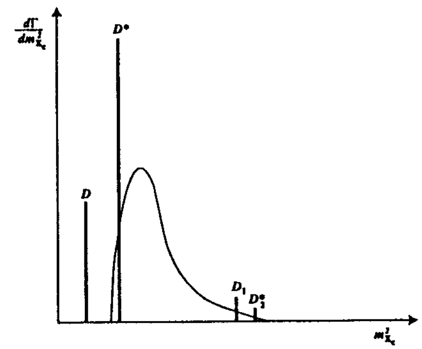

Fig. 1: A sketch for $b \to c$ semileptonic decay of the continuous inclusive recoil spectrum of the OPE calculations (smooth curve) compared to the known hadronic spectrum (shown as individual resonance lines).

Let me begin my discussion of the accuracy of inclusive methods with an overview. Up to *caveats* regarding the unknown accuracy of the assumption of duality, the OPE indicates that inclusive calculations should be accurate up to corrections of order Λ_{QCD}^2/m_Q^2. Here I will identify a source of duality-violation which leads to Λ_{QCD}/m_Q corrections. It is revealed by considerations of a Bjorken sum rule [13] which may be viewed as an extension of Shifman-Voloshin duality to arbitrary recoils. Bjorken's sum rule guarantees that, as $m_b \to \infty$, duality will be enforced locally in the semileptonic decay Dalitz plot of rate versus $w-1$ and

E_ℓ (where $w \equiv v \cdot v'$ is the usual heavy quark double-velocity variable and E_ℓ is the lepton energy). For regions of the Dalitz plot for which $w-1$ is not large (and in $b \to c$ decay nearly the whole Dalitz plot satisfies this condition), the Bjorken sum rule explicitly relates the loss of total rate from the "elastic" $s_\ell^{\pi_\ell} = \frac{1}{2}^-$ channels, as the Isgur-Wise function falls, to the turn-on of the production of $s_\ell^{\pi_\ell} = \frac{1}{2}^+$ and $\frac{3}{2}^+$ states [14].

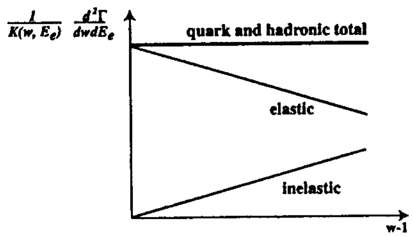

Fig. 2: The exact compensation by inelastic channels of the fall of the elastic rate in the linear region as $m_b \to \infty$.

In particular, in this region the Isgur-Wise function may be taken to be linear:

$$\xi(w) \simeq 1 - \rho^2(w-1) \equiv 1 - \left[\frac{1}{4} + \rho^2_{dyn}\right](w-1) \quad . \tag{1}$$

In that case, for fixed $r \equiv m_c/m_b$, as $m_b \to \infty$ inelastic $s_\ell^{\pi_\ell} = \frac{1}{2}^+$ and $\frac{3}{2}^+$ channels open up to give a semileptonic rate that would exactly and locally compensate in the Dalitz plot the loss of rate from the elastic channels due to ρ^2_{dyn}. *I.e.*, if

$$\frac{d^2\Gamma^{inclusive}_{quark}}{dwdE_\ell} = K(w, E_\ell) \tag{2}$$

then [13,14]

$$\frac{d^2\Gamma^{inclusive}_{hadron}}{dwdE_\ell} = K(w,E_\ell)\left(\frac{w+1}{2}|\xi(w)|^2 + 2(w-1)\left[\sum_m |\tau_{\frac{1}{2}}^{(m)}(1)|^2 + 2\sum_p |\tau_{\frac{3}{2}}^{(p)}(1)|^2\right]\right) \tag{3}$$

as $m_b \to \infty$ and for $\rho^2(w-1)_{max} = \rho^2\frac{(1-r)^2}{2r} << 1$. Since according to (1)

$$\left(\frac{w+1}{2}\right)|\xi(w)|^2 \simeq 1 - 2\rho^2_{dyn}(w-1) \quad , \tag{4}$$

we see explicitly the compensation of the dynamical part of the fall of the elastic channels by the onset of inelastic states (of all sorts) with $s_\ell^{\pi_\ell} = \frac{1}{2}^+$ and $\frac{3}{2}^+$. This situation is sketched in Figure 2; if it were applicable to $b \to c$ decays, then quark-hadron duality would be exact.

Having established conditions for its validity as $m_b \to \infty$, it is easy to see why one should be concerned about quark-hadron duality for $b \to c$ decays. For fixed r, $w-1$ lies in the fixed range from 0 to $(1-r)^2/2r$, and as $m_b \to \infty$ any given hadronic threshold collapses to the point $w = 1$. However, for finite m_b there is a gap in $w-1$ in which the rate to the elastic $\frac{1}{2}^-$ channels falls by Λ_{QCD}/m_Q terms but the potentially compensating excited state channels $\frac{1}{2}^+$ and $\frac{3}{2}^+$ are not yet kinematically allowed. More precisely, if $m_{D^{**}}$ is the mass of a generic charmed inelastic state, then $t_m^{**} = (m_B - m_{D^{**}})^2$ would be the threshold in t for this state, corresponding to a value of $w-1$ in the quark-decay Dalitz plot of

$$\frac{t_m - t_m^{**}}{2m_b m_c} \simeq (1-r)\frac{\Delta}{m_c} \tag{5}$$

where $t_m \equiv (m_B - m_D)^2 \simeq (m_b - m_c)^2$ and $\Delta \equiv m_{D^{**}} - m_D$. Since $\Delta \simeq 500$ MeV and $(w-1)_{max} \simeq 0.6$, this region covers more than one third of the Dalitz plot and the compensation is very substantially delayed: see Figure 3. Eqs. (5) and (1) show that this effect is of order Λ_{QCD}/m_Q, seemingly at odds with the OPE result.

Despite this apparent contradiction, there is actually no inconsistency: the OPE result that the leading corrections to the inclusive rate are of order Λ^2_{QCD}/m^2_Q can still be valid as derived in the limit of large energy release in the $b \to c$ transition, while Λ_{QCD}/m_Q effects can arise for energy releases of the order of Λ_{QCD} due to a finite radius of convergence of the OPE. The main purpose of this paper is indeed to call attention to this effect, and to use the quark model to estimate its importance.

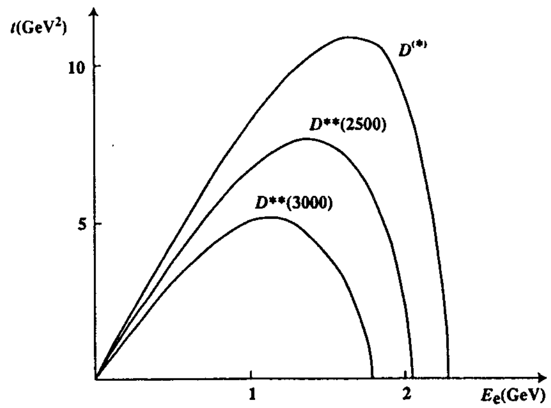

Fig. 3: An overlay of the Dalitz plots for $\bar{B} \to D^{(*)} e\bar{\nu}_e$, $\bar{B} \to D^{**}(2500) e\bar{\nu}_e$, and $\bar{B} \to D^{**}(3000) e\bar{\nu}_e$. The $D^{(*)}$ mass is taken as the hyperfine average of the D and D^* masses; the two D^{**} masses are chosen for illustrative purposes.

The basic issues can be most easily exposed by considering [4] spinless quarks coupled to a scalar field ϕ of mass μ, and by studying the decay $b \to c\phi$ with weak coupling constant g. Differential semileptonic decay rates have a more complex spin structure, but otherwise correspond to the case $\mu = \sqrt{t}$; total semileptonic rates correspond to a weighted average over kinematically allowed μ but, as we shall see below, this averaging does not change the essentials of the problem. In our simplified case

$$\Gamma(b \to c\phi) = \frac{g^2 p_{cb}}{8\pi m_b^2} \tag{6}$$

where $p_{fi} \equiv [(m_i - m_f)^2 - \mu^2]^{1/2}[(m_i + m_f)^2 - \mu^2]^{1/2}/2m_i$ is the momentum of ϕ from the two-body decay of mass m_i into masses m_f and μ.

To compare Eq. (6) with a hadronic world, we can use the quark model framework of Refs. [15,16] where

$$\Gamma(B \to D^{(n)}\phi) = \frac{g^2 p_{D^{(n)}B}}{8\pi m_B^2}\Big(\frac{m_{D^{(n)}} m_B}{m_c m_b}\Big)|\xi^{(n)}(\vec{v}_{D^{(n)}B})|^2 \tag{7}$$

with

$$\xi^{(n)}(\vec{v}_{D^{(n)}B}) \equiv \int d^3p\, \phi^*_{D^{(n)}}(\vec{p})\phi_B(\vec{p} - m_d\vec{v}_{D^{(n)}B}) \tag{8}$$

and where

$$\vec{v}_{D^{(n)}B} = \vec{p}_{D^{(n)}B}/m_{D^{(n)}} \tag{9}$$

is the recoil velocity of $D^{(n)}$. These results may be used to estimate

$$R \equiv \frac{\sum_n \Gamma(B \to D^{(n)}\phi)}{\Gamma(b \to c\phi)} \quad . \tag{10}$$

Using simple harmonic oscillator wavefunctions in (8) gives

$$\xi_{DB}(\vec{v}) = \Big[\frac{\beta_D\beta_B}{\bar{\beta}_{DB}^2}\Big]^{3/2} exp\Big(-\frac{m_d^2 v^2}{4\bar{\beta}_{DB}^2}\Big) \tag{11}$$

and

$$\xi^i_{D^{**}B}(\vec{v}) = \frac{m_d v^i}{\sqrt{2}\beta_B}\Big[\frac{\beta_D\beta_B}{\bar{\beta}_{DB}^2}\Big]^{5/2} exp\Big(-\frac{m_d^2 v^2}{4\bar{\beta}_{DB}^2}\Big) \tag{12}$$

where $\bar{\beta}_{DB}^2 \equiv \frac{1}{2}(\beta_B^2 + \beta_D^2)$, ξ_{DB} is the $B \to D$ form factor, and $\xi^i_{D^{**}B}$ is the form factor for transitions into the lowest $\ell = 1$ excited state with $m_\ell = i$. If we define an order Λ_{QCD}/m_Q expansion parameter

$$\epsilon \equiv \frac{m_d}{\mu_-} = \frac{m_d(m_b - m_c)}{m_b m_c} \tag{13}$$

and the "scaled energy release"

$$T^* \equiv \frac{m_b - m_c - \mu}{\Delta} \tag{14}$$

where $\Delta \equiv m_{D^{**}} - m_D$, then for small T^* (*i.e.*, small velocities)

$$|\xi_{DB}|^2 = 1 - \epsilon T^* + O(\epsilon^2) \tag{15}$$

and, defining $|\xi_{D^{**}B}|^2 \equiv \sum_i |\xi^i_{D^{**}B}|^2$,

$$|\xi_{D^{**}B}|^2 = \epsilon(T^* - 1) + O(\epsilon^2) \quad . \tag{16}$$

ince in this limit

$$\frac{p_{D^{(n)}B}}{p_{cb}} = \left[\frac{m_{D^{(n)}} m_b}{m_c m_B}\right]^{1/2} \left(\frac{T^* - 1}{T^*}\right)^{1/2} , \tag{17}$$

e can obtain a model [4] for R by truncating the sum over n after the first D^{**}:

$$R_1^{D^{**}} \equiv \frac{\Gamma(B \to D\phi) + \Gamma(B \to D^{**}\phi)}{\Gamma(b \to c\phi)} \tag{18}$$

$$= [1 + \frac{3}{2}\epsilon - \epsilon T^*]\theta(T^*) + \epsilon \frac{(T^* - 1)^{3/2}}{T^{*1/2}} \theta(T^* - 1) , \tag{19}$$

herein we have made the weak binding approximation $\beta << m_d$ and shown explicitly the vo thresholds at $T^* = 0$ and $T^* = 1$. As $T^* \to \infty$, the complete expression for R will isplay a tower of states (each with their appropriate threshold factors) being produced ith a strength given by a power series in T^* with coefficients which are power series in eneralized ϵ-coefficients of order Λ_{QCD}/m_Q. For duality to be valid, each term T^{*n} for > 0 must have zero coefficient. The OPE further states that the constant term T^{*0} is of ie form $1 + O(\Lambda^2_{QCD}/m^2_Q)$.

While extreme, this truncation of the complete expression for R has the properties that:

1. At $T^* \to \infty$, it is of the form $1 + O(\epsilon^2) + O(\epsilon/T^*)$ as required by the OPE.

2. There are no other terms of order 1, ϵ, or ϵT^* possible beyond those shown: a more ccurate treatment of $\Gamma(B \to D\phi)$ could only generate ϵ^2, $\epsilon^2 T^*$, $\epsilon^2 T^{*2}$, ... terms; a more ccurate treatment of $\Gamma(B \to D^{**}\phi)$ could only generate $\epsilon^2 T^*$, $\epsilon^2 T^{*2}$, ... terms; and all igher states first make a contribution at order $\epsilon^2 T^{*2}$ or higher. Conversely, we note that if, r example, $\epsilon^2 T^{*2}$ terms are retained, they must all cancel exactly or the requirements of ie OPE would be violated as $T^* \to \infty$.

3. As $\Delta m \equiv m_b - m_c \to 0$, $R_1^{D^{**}} \to 1 + O(\frac{\Lambda_{QCD}\Delta m}{m_b^2})$ as required [17].

4. Near $T^*_{max} \equiv m_b - m_c$, ϵT^*_{max} is in general large. This observation corresponds in the sual language of heavy quark symmetry to the statement that the natural scale of the slope ^{2} of the Isgur-Wise function is of order unity. It is also consistent with the experimental bservation that $|\xi_{DB}|^2$ has dropped to less than half its value between $w - 1 = 0$ and its

maximum value. Given this, the extension of Eq. (19) to higher orders in T^* will require a "conspiracy" of the entire spectrum of possible hadronic final states. Nevertheless, we can use Eq. (19) across the full range of T^* as an indicator of the Λ_{QCD}/m_Q effects arising from the order 1 and order T^* terms in the expansion of R; this corresponds to a "best case" assumption that duality is locally perfect for the terms T^{*n} with $n > 1$.

While this simple example clearly demonstrates the existence of the claimed duality-violating Λ_{QCD}/m_Q effects for finite T^*, it remains to discuss their quantitative importance. Before doing so, I will introduce a number of simple variants of this prototypical model. The first corrects an idiosyncracy of the simple harmonic oscillator model: in it the first D^{**} state saturates the Bjorken sum rule, *i.e.*, completely compensates the $-\epsilon T^*$ elastic term. In contrast, in the ISGW model where harmonic oscillator wavefunctions are simply used as a variational basis, $\beta_{D^{**}} \neq \beta_D$ and this saturation does not occur. This is typical of the general case where (in the narrow resonance approximation) the total resonant P-wave term is of the form

$$\frac{\epsilon}{T^{*1/2}} \sum_n f_n (T^* - t_n^*)^{3/2} \theta(T^* - t_n^*) \tag{20}$$

with $\sum_n f_n = 1$ and t_n^* being the threshold for channel n. As $T^* \to \infty$, these contributions automatically cancel the $-\epsilon T^*$ term from the elastic form factor, and constrain the $O(\epsilon)$ correction:

$$R^{D^{**}}_{1+2+...} = [1 + \frac{3}{2}\epsilon \bar{t}^* - \epsilon T^*]\theta(T^*) + \frac{\epsilon}{T^{*1/2}} \sum_n f_n (T^* - t_n^*)^{3/2} \theta(T^* - t_n^*) , \tag{21}$$

where

$$\bar{t}^* = \sum_n f_n t_n^* \tag{22}$$

is the weighted average threshold position. Note that since some T_n^* exceed T^*_{max}, $R^{D^{**}}_{1+2+...}$ cannot heal to unity in the physical decay region.

As described above, both $R_1^{D^{**}}$ and $R^{D^{**}}_{1+2+...}$ are "best case" truncations which assume exact cancellations of $\epsilon^2 T^*$, $\epsilon^2 T^{*2}$, ... terms. While sufficient for the purposes of this study, I note that it is straightforward to recursively "construct duality" to the required order in ϵ to any finite order in T^*. For example, for a simple harmonic oscillator spectrum one can easily construct

$$\begin{aligned} R^{ho} = \frac{exp(-\epsilon T^*)}{T^{*1/2}} \Big(& [1+\frac{3}{2}\epsilon]T^{*1/2}\theta(T^*) \\ & +\epsilon[1+\frac{5}{2}\epsilon+\frac{35}{16}\epsilon^2+\frac{35}{32}\epsilon^3+\frac{385}{1024}\epsilon^4+...](T^*-1)^{3/2}\theta(T^*-1) \\ & +\frac{1}{2!}\epsilon^2[1+\frac{7}{2}\epsilon+\frac{21}{4}\epsilon^2+\frac{77}{16}\epsilon^3+...](T^*-2)^{5/2}\theta(T^*-2) \\ & +\frac{1}{3!}\epsilon^3[1+\frac{9}{2}\epsilon+\frac{297}{32}\epsilon^2+...](T^*-3)^{7/2}\theta(T^*-3) \\ & +\frac{1}{4!}\epsilon^4[1+\frac{11}{2}\epsilon+...](T^*-4)^{9/2}\theta(T^*-4) \\ & +\frac{1}{5!}\epsilon^5[1+...](T^*-5)^{11/2}\theta(T^*-5)+...\Big) \quad , \end{aligned}$$

where the ellipses denote terms of order $\epsilon^6 T^{*n}$ with $1 \leq n \leq 5$ and all terms of order $\epsilon^m T^{*m}$ and higher with $m > 5$. This truncated expansion is accurate even at $T^* = 5$ up to corrections of order Λ_{QCD}^2/m_Q^2; as we will see below, higher values of T^* are probably not physically relevant.

The models just introduced are all based on the duality of $b \to c\phi$ to a tower of $c\bar{d}$ resonances. While the thresholds associated with such towers are a source of duality-violating Λ_{QCD}/m_Q corrections which must be a cause for concern in comparing inclusive calculations with experiment, I am even more concerned about processes which could give a nonperturbative high-mass tail to the recoil mass distribution. The convergence to unity of R in the former case would be controlled by an expansion in $1/T^* \sim \Delta/(m_b - m_c - \mu)$, *i.e.*, T must be large compared to the single resonance scale Δ. Since $\Delta/T_{max} \sim 1/6$, there are reasons to be cautious, but perhaps not alarmed. However, I believe that there is another "harder" effective scale in low energy hadron structure which I identify with the constituent quark size r_q. While, like Δ, $1/r_q$ must be "of order Λ_{QCD}", empirically it is much larger: the constituent quark model makes sense only if the quarks are small relative to hadronic radii. Quantitative estimates from spectroscopy [18] indicate that this scale is indeed $1/r_q \sim 2$ GeV. This is in turn potentially very dangerous for duality when T_{max} is only of order 3 GeV as in $b \to c$ decays.

Let me mention two concrete examples of how this scale could lead to a high-mass tail to the recoil mass distribution which would delay the compensation required for duality. If the glue around a constituent quark is indeed very compact, then high recoil momenta of the c quark are required if it is to be left behind, *i.e.*, before the current c quark can carry away all of the energy of the underlying $b \to c\phi$ transition. This effect would lead to a convergence problem for the resonance contributions to R associated with the excitation of hybrid mesons. The second example is probably more dangerous: such a scale could lead to the nonperturbative production of high-mass nonresonant states [19], *e.g.*, an $X_c Y$ mass spectrum for $B \to X_c Y \phi$ extending from $X_c Y$ threshold up to masses of order $m_{X_c} + m_Y + 1/r_q \sim 5$ GeV $\sim m_B$. We should therefore be concerned that a substantial fraction of nonresonant production is unavailable to participate in the compensation required for duality to be realized in $b \to c$ decays.

For a crude estimate of the effects of a high-mass hybrid contribution, I take a simple two-component resonance model consisting of "normal" $c\bar{d}$ resonances with $\bar{t}^*_{c\bar{d}}$ and $c\bar{d}$ hybrids with $\bar{t}^*_{hybrid}$ substantially larger. If we assume that the latter are responsible for a fraction κ of ρ^2_{dyn}, then we would have

$$\begin{aligned} R^{hybrid} = & [1+\frac{3}{2}\epsilon\bar{t}^* - \epsilon T^*]\theta(T^*) \\ & +(1-\kappa)\epsilon\frac{(T^*-\bar{t}^*_{c\bar{d}})^{3/2}}{T^{*1/2}}\theta(T^*-\bar{t}^*_{c\bar{d}}) \\ & +\kappa\epsilon\frac{(T^*-\bar{t}^*_{hybrid})^{3/2}}{T^{*1/2}}\theta(T^*-\bar{t}^*_{hybrid}) \quad , \end{aligned} \tag{23}$$

with $\bar{t}^* = (1-\kappa)\bar{t}^*_{c\bar{d}} + \kappa\bar{t}^*_{hybrid}$. Since both experimentally and theoretically the hybrid mass spectrum begins about 1.5 GeV above the ground state, this provides a minimum value for $\bar{t}^*_{hybrid}$. Since their "hard" production mechanism will raise their mean T^* above this minimum, the effects of their postponed onset could be serious.

As already implied, I believe that the postponed nonresonant contributions are an even more serious cause for concern. While the model (23) for the hybrids could also be used as a template for a crude model for nonresonant states (their contribution will in first order be controlled by a parameter $\bar{t}^*_{nr}$ which would replace $\bar{t}^*_{hybrid}$ in Eq. (23)), a more appropriate model would be

$$\begin{aligned} R^{nr} &= [1 + \frac{3}{2}\epsilon\bar{t}^* - \epsilon T^*]\theta(T^*) \\ &+(1-\lambda)\epsilon\frac{(T^* - \bar{t}^*_{cd})^{3/2}}{T^{*1/2}}\theta(T^* - \bar{t}^*_{cd}) \\ &+\lambda\epsilon\int_{T^*_{min}}^{T^*} dt^*\rho(t^*)\frac{(T^* - t^*)^{3/2}}{T^{*1/2}} \quad , \end{aligned} \tag{24}$$

where λ is the fraction of ρ^2_{dyn} due to nonresonant states and $\rho(t^*)$ is the appropriate normalized spectral function ($\int_{T^*_{min}}^{\infty} dt^*\rho(t^*) = 1$) which begins at T^*_{min} but drops off very slowly with a scale determined by $1/r_q$. In this situation, $\bar{t}^* = (1-\lambda)\bar{t}^*_{cd} + \lambda\int_{T^*_{min}}^{\infty} dt^*\rho(t^*)t^*$.

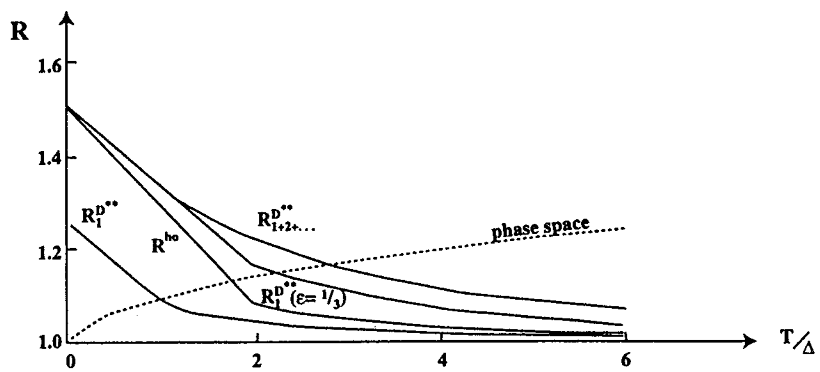

Fig. 4: Four resonance models of the approach to duality: (a) $R_1^{D^{**}}$ (with the baseline value $\epsilon = 1/6$), (b) $R_1^{D^{**}}$ (with $\epsilon = 1/3$ corresponding to $T^*_{max} = 3$), (c) $R_{1+2+...}^{D^{**}}$ (with $t^*_n = n$ and $f_n = (\frac{1}{2})^n$ so that $\bar{t}^* = 2$), and (d) R^{ho}.

I now turn to rough quantitative estimates of duality-violating effects using the models I have presented. Since experimentally $\rho^2 \sim 1$, $(w-1)_{max} \sim 1/2$ and $(m_b - m_c)/\Delta \sim 6$, I will

take $\epsilon = 1/6$ and $T^*_{max} = 6$ as a realistic baseline for all of the following examples. However, in keeping with the main message of this paper that inclusive results must be interpreted cautiously until duality-violating effects are better understood, I will choose pessimistic values for other parameters of the models and for variations around these baseline values.

Figure 4 shows four different examples of resonance compensation: $R_1^{D^{**}}$ (with both the baseline value $\epsilon = 1/6$ and for $\epsilon = 1/3$ corresponding to $T^*_{max} = 3$, *i.e.*, to using $\Delta_{eff} = 2\Delta$ for the mean location of the P-wave strength), $R_{1+2+...}^{D^{**}}$ (with $t^*_n = n$ and $f_n = (\frac{1}{2})^n$ so that $\bar{t}^* = 2$), and R^{ho}. These examples show that duality-violation of the order of 10% could easily arise in $b \to c$ decays from the delayed onset of resonances with a scale $\Delta \simeq 500$ MeV.

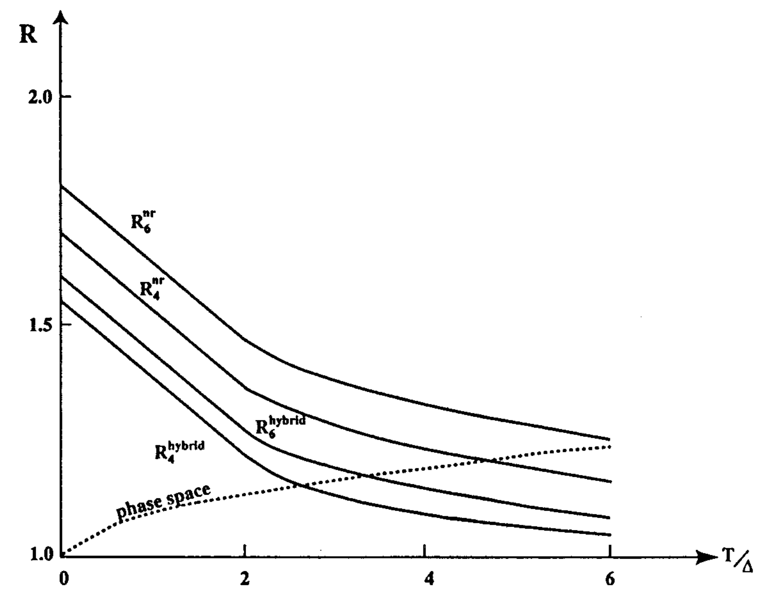

Fig. 5: Four examples of the effects of a nonperturbative high mass tail on the recoil mass spectrum: (a) R^{hybrid} (with $\kappa = 1/10$, $\bar{t}^*_{cd} = 2$, and $\bar{t}^*_{hybrid} = 4$), (b) As in (a), but with $\bar{t}^*_{hybrid} = 6$, (c) R^{nr} (with $\lambda = 1/5$, $\bar{t}^*_{cd} = 2$, $T^*_{min} = 2$, and $s = 4$), and (d) As in (c) but with $s = 6$.

Figure 5 shows the potentially even more dangerous effects of there being a second scale in this problem much larger than Δ. Here I once again give four examples: R^{hybrid} (with $\kappa = 1/10$, $\bar{t}^*_{c\bar{d}} = 2$, and $\bar{t}^*_{hybrid} = 4$ and 6) and R^{nr} (with $\lambda = 1/5$, $\bar{t}^*_{c\bar{d}} = 2$, and

$$\rho(t^*) = \frac{1}{s} exp(\frac{T^*_{min} - t^*}{s})\theta(t^* - T^*_{min}) \tag{25}$$

with $T^*_{min} = 2$ and $s = 4$ and 6). Note that the choices made for $\bar{t}^*_{hybrid}$ and s are based on the hypothetical scale $1/r_q \sim 2$ GeV. The choices $\kappa = 1/10$ and $\lambda = 1/5$ are based on estimates [9] of the strengths of these contributions, but values of this order are certainly reasonable (*e.g.*, λ is a $1/N_c$ effect). These examples show that it is difficult to avoid the conclusion that nonperturbative high mass contributions to ρ^2 could lead to substantial duality-violating Λ_{QCD}/m_Q effects in $b \to c$ decays.

Although our main focus has been on heavy-to-heavy transitions, the physics issues raised here (if not their explicit forms) are also relevant for $Q \to q\ell\bar{\nu}_\ell$ transitions. Before concluding, let me therefore point out a simple application of the OPE to inclusive heavy-to-light transitions where it seems certain to me that they will fail: Cabibbo-forbidden charm decays. (Even though such decays might be an unimportant application of the inclusive calculations in practice, they provide a valid theoretical testing ground for their accuracy.) In particular, consider the $c \to d\bar{\ell}\nu_\ell$ decays of the D^0 and D^+. They will be dominated by the channels $D^0 \to \pi^-\bar{\ell}\nu_\ell$ and $\rho^-\bar{\ell}\nu_\ell$ and by $D^+ \to \pi^0\bar{\ell}\nu_\ell$, $\eta\bar{\ell}\nu_\ell$, $\eta'\bar{\ell}\nu_\ell$, $\rho^0\bar{\ell}\nu_\ell$, and $\omega\bar{\ell}\nu_\ell$. Since the OPE corrections in the D^0 and D^+ are *identical*, their Cabibbo-forbidden semileptonic partial widths and spectral distributions are predicted to be identical. However, simple isospin symmetry implies that $\Gamma(D^+ \to \pi^0\bar{\ell}\nu_\ell) = \frac{1}{2}\Gamma(D^0 \to \pi^-\bar{\ell}\nu_\ell)$, so the inclusive Cabibbo-forbidden rates can only be equal if $\Gamma(D^+ \to \eta\bar{\ell}\nu_\ell) + \Gamma(D^+ \to \eta'\bar{\ell}\nu_\ell) = \Gamma(D^+ \to \pi^0\bar{\ell}\nu_\ell)$. In many models this latter relation would be true if $m_\eta = m_{\eta'} = m_\pi$, since it is rather natural for the squares of matrix elements to satisfy its analogue. However, with real phase space factors, this relation is typically badly broken. Since Cabibbo-forbidden decays, like their Cabibbo-allowed counterparts, will receive little P-wave compensation, I expect this prediction to fail.

Finally, I note that the duality-violating effects I have highlighted here will have an effect on the long-standing $\bar{B}$ semileptonic branching ratio puzzle [20]. Since the hadronic mass distribution in $b \to c\bar{u}d$ is weighted toward higher masses than the leptonic mass distribution in $b \to c\ell\bar{\nu}_\ell$, the ratio of these two rates will be changed.

In summary, I have shown here that hadronic thresholds lead to Λ_{QCD}/m_Q violations of duality in $b \to c$ decays which do not explicitly appear in the operator product expansion. Since such violations cannot appear as the $b \to c$ energy release $T \to \infty$, there are "conspiracies" (*i.e.*, sum rules) which relate hadronic thresholds and transition form factors. As emphasized by Bigi, Uraltsev, Shifman, Vainshtein, and others [4,5,7], these relations tend to compensate the otherwise extremely large Λ_{QCD}/m_Q effects even at small T. In this paper I have displayed several models of such hadronic compensation mechanisms which indicate that these duality-violating Λ_{QCD}/m_Q effects could nevertheless be very substantial. While the examples I have selected are perhaps pessimistic, they indicate that these effects must be better understood before inclusive methods can be applied with confidence to heavy quark semileptonic decays.

ACKNOWLEDGEMENTS

I am very grateful to number of colleagues who examined and commented upon a draft version of this paper, including Adam Falk, Mark Wise, Zoltan Ligeti, Iain Stewart, Matthias Neubert, and Richard Lebed.

I am particularly indebted to Ikaros Bigi, Misha Shifman, Nikolai Uraltsev, and Arkady Vainshtein who replied with a detailed and very pedagogical explanation of their understanding of the effects discussed in this paper. I certainly learned more from this exchange than they did, and substantial errors in the draft version were corrected as a result.

REFERENCES

[1] N. Isgur and M.B. Wise, Phys. Lett. **B232**, 113 (1989); Phys. Lett. **B237**, 527 (1990). For an overview of Heavy Quark Symmetry and additional references see N. Isgur and M.B. Wise, "Heavy Quark Symmetry" in *B Decays* (Revised 2nd edition), ed. S. Stone (World Scientific, Singapore, 1994), p. 231.

[2] H. Georgi, Phys. Lett. **B240**, 447 (1990).

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

[4] I.I. Bigi, N.G. Uraltsev, and A.I. Vainshtein, Phys. Lett. **B293**, 430 (1992);**297**, 477(E)(1993); I.I. Bigi, M.A. Shifman, N.G. Uraltsev, and A.I. Vainshtein, Phys. Rev. Lett. **71**, 496 (1993); Int. J. Mod. Phys. **A9**, 2467 (1994); B. Blok, L. Koyrakh, M.A. Shifman, and A.I. Vainshtein, Phys. Rev. **D49**, 3356 (1994);**50** 3572(E)(1994); I.I. Bigi and N.G. Uraltsev, Z. Phy. **C62**, 623 (1994); M.A. Shifman, N.G. Uraltsev, and A.I. Vainshtein, Phys. Rev. **D51**, 2217 (1995); 3149(E) (1995); B. Chibisov, R.D. Dikeman, M.A. Shifman, and N.G. Uraltsev, Int. J. Mod. Phys. **A12**, 2075 (1997); and see also [20] for a review.

[5] T. Mannel, Nuc. Phys. **B413**, 396 (1994); A.F. Falk, M. Luke, and M.J. Savage, Phys. Rev. **D49**, 3367 (1994); **D53**, 2491 (1996); **D53**, 6316 (1996); A.V. Manohar and M.B. Wise, Phys. Rev. **D49**, 1310 (1994); M. Neubert, Phys. Rev. **D49**, 3392 (1994); **49**, 4623 (1994); T. Mannel and M. Neubert, Phys. Rev. **D50**, 2037 (1994); M. Luke and M. Savage, Phys. Lett. **B321**, 88 (1994); A.F. Falk, Z. Ligeti, M. Neubert, and Y. Nir, Phys. Lett. **B326**, 145 (1994); C.G. Boyd, B. Grinstein, and A.V. Manohar, Phys. Rev. **D54**, 2081 (1996).

[6] H. Georgi, private communication (1991).

[7] See, in addition to Refs. [4], C.G. Boyd, B. Grinstein, and A.V. Manohar in Ref. [5]; M. Neubert, Nuc. Phys. B(Proc. Suppl.) **59**, 101 (1997); H.J. Lipkin, Phys. Lett. **B308**, 105 (1993); Nucl Phys. **A560**, 548 (1993).

[8] Here $s_\ell^{\pi_\ell}$ refers to the spin s_ℓ and parity π_ℓ of the light (ℓ) degrees of freedom which determine the spectroscopy of heavy-light systems; see N. Isgur and M.B. Wise, Phys. Rev. Lett. **66**, 1130 (1991).

[9] N. Isgur, Phys. Rev. **D54**, 5896 (1996).

[10] As argued in Refs. [15,16] by considering the free recoil of the final state quark c against a static spectator $\bar{d}$ quark, Δm_{X_c} should be roughly $[(m_c+m_d)^2+m_d m_b^{-1}(m_b-m_c)^2]^{\frac{1}{2}}-(m_c+m_d)\simeq 0.2$ GeV, consistent with the data and the inclusive models shown in Figure 1, and certainly not leading to a small expansion parameter $\Lambda_{QCD}/\Delta m_{X_c}$ for $b\to c$ decays.

[11] M. B. Voloshin and M. A. Shifman, Sov. J. Nucl. Phys. **45**, 292 (1987); M. B. Voloshin and M. A. Shifman, Yad. Fiz. **47**, 801 (1988); Sov. J. Nucl. Phys. **47**, 511 (1988); M. A. Shifman in *Proceedings of the 1987 International Symposium on Lepton and Photon Interactions at High Energies*, Hamburg, West Germany, 1987, edited by W. Bartel and R. Rückl, Nucl. Phys. B (Proc. Suppl.) **3**, 289 (1988)

[12] N. Isgur, Phys. Rev. D**40**, 109 (1989); see also Ref. [17].

[13] J.D. Bjorken, in *Results and Perspectives in Particle Physics*, Proceedings of the 4^{th} Rencontre de Physique de la Valleé d'Aoste, La Thuile, Italy, 1990, edited by M. Greco (Editions Frontières, Gif-sur-Yvette, France, 1990); J. D. Bjorken, in *Proceedings of the XXVth International Conference on High Energy Physics*, Singapore, 1990, edited by K.K. Phua and Y. Yamaguchi (World Scientific, Singapore, 1992).

[14] N. Isgur and M. B. Wise, Phys. Rev. D **43**, 819 (1991).

[15] D. Scora and N. Isgur, Phys. Rev. D **52**, 2783 (1995); N. Isgur, D. Scora, B. Grinstein, and M. B. Wise, Phys. Rev. D **39**, 799 (1989); B. Grinstein, M.B. Wise, and N. Isgur, Caltech Report No. CALT-68-1311 and University of Toronto Report No. UTPT-85-37, 1985 (unpublished); B. Grinstein, M. B. Wise, and N. Isgur, Phys. Rev. Lett. **56**, 298 (1986).

[16] D. Scora and N. Isgur, Phys. Rev. **D52**, 2783 (1995).

[17] The validity of the inclusive expansion in $1/m_Q$ has been studied in the Shifman-Voloshin limit in C.G. Boyd, B. Grinstein, and A.V. Manohar, Phys. Rev. **D54**, 2081 (1996), who demonstrate that the inclusive and exclusive rates do indeed agree through order $\delta m/m_Q$, $\lambda_1/m_Q\delta m$, and $\lambda_2/m_Q\delta m$ in this limit. It should be noted that they did *not* find agreement through order λ_1/m_Q^2 and λ_2/m_Q^2, consistent with our picture, although their analysis did not uncover the correction being discussed here because D^{**} production is of order $\delta m^2/m_Q^2$ in the Shifman-Voloshin limit. We also note in passing that Shifman-Voloshin duality through order $\delta m/m_Q$ was first discussed in the context of the quark model before the discovery of heavy quark symmetry in Ref. [12], where identical results were obtained (though expressed in term of the parameter $x_m \equiv (m_Q^2 - m_q^2)/2m_Q^2$ instead of $\delta m/m_Q$).

[18] For one example of a quantitative estimate, see the Appendix of P. Geiger and N. Isgur, Phys. Rev. D44, 799 (1991), where $r_q \leq 0.1$ fm is obtained. For a simple qualitative argument, consider the ratio of the $a_2 - b_1$ and $\rho - \pi$ mass splittings: the P-wave hyperfine splittings are only suppressed as observed if the constituent quarks are much smaller than hadrons.

[19] N. Isgur, "Nonresonant Semileptonic Decays of Heavy Quarks", in preparation.

[20] I. Bigi, B. Blok, M. Shifman, N. Uraltsev, and A. Vainshtein, "Non-Leptonic Decays of Beauty Hadrons - from Phenomenology to Theory" in *B Decays* (Revised 2nd edition), ed. S. Stone (World Scientific, Singapore, 1994), p. 132.