Kolya: an open-source package for inclusive semileptonic *B* decays

Matteo Fael[1]⋆, Ilija S. Milutin[2]† and K. Keri Vos[3,4]‡

1 Theoretical Physics Department, CERN, 1211 Geneva, Switzerland
2 Theoretische Physik 1, Center for Particle Physics Siegen,
Universität Siegen, D-57068 Siegen, Germany
3 Gravitational Waves and Fundamental Physics (GWFP),
Maastricht University, Duboisdomein 30, NL-6229 GT Maastricht, the Netherlands
4 Nikhef, Science Park 105, NL-1098 XG Amsterdam, the Netherlands

⋆ matteo.fael@cern.ch   † Ilija.Milutin@uni-siegen.de   ‡ k.vos@maastrichtuniversity.nl

Abstract

We introduce the code `kolya`, an open-source tool for phenomenological analyses of inclusive semileptonic *B* meson decays. It contains a library to compute predictions for the total rate and various kinematic moments within the framework of the heavy quark expansion, utilizing the so-called kinetic scheme. The library currently includes power corrections up to $1/m_b^5$. All available QCD perturbative corrections are implemented via interpolation grids for fast numerical evaluation. We also include effects from new physics parameterised as Wilson coefficients of dimension-six operators in the weak effective theory below the electroweak scale. The library is interfaced to CRunDec for easy evaluation of the quark masses and strong coupling constant at different renormalization scales. The library is developed in Python and does not require compilation. It can be used in an interactive Jupyter notebook session.

1 Introduction

Measurements of semileptonic B decays lie at the core of the Belle II and LHCb physics program in the upcoming years. Thanks to relatively large rates and clean experimental signatures, inclusive and exclusive semileptonic decays with a $b \to c l \bar{\nu}_l$ transition ($\ell = e, \mu$) offer a clean avenue for the determinations of $|V_{cb}|$, the element of the Cabibbo-Kobayashi-Maskawa matrix (CKM) which parameterizes the strength of the weak interaction among bottom and charm quarks in the Standard Model of particle physics.

Inclusive determinations of $|V_{cb}|$ exploit that the semileptonic rate Γ_{sl} and moments of kinematic spectra can be described with good precision using the heavy quark expansion (HQE) [1–3]. In the HQE, these observables are expressed as a series of non-perturbative HQE elements proportional to

increasing powers of the inverse bottom quark mass times the QCD scale parameter, Λ_{QCD}/m_b. In addition, each order in the HQE also receives corrections expressed as a series expansion in the strong coupling constant, α_s, which can be systematically calculated in perturbative QCD.

This paper documents the first release of the open-source code `kolya` [4]. It consists of a Python library which computes the prediction for the total rate and lepton-energy, hadronic invariant mass and dilepton invariant mass kinematic moments within the framework of the HQE utilizing the so-called kinetic scheme [5–8]. The `kolya` code supersedes and extends the code developed for the fit of q^2 moments [9] measured by Belle [10] and Belle II [11]. We also include effects from new physics (NP) studied in Ref. [12]. These are parameterised as Wilson coefficients of dimension-six operators in the weak effective theory below the electroweak scale [13, 14].

Several building blocks necessary for the prediction of Γ_{sl} and the moments to high orders in α_s and the $1/m_b$ expansion have been presented over the last 30 years (see Sec. 3.2 for an exhaustive list of references). The `kolya` library provides the first comprehensive open-source framework in which all available corrections are implemented and validated. A schematic overview of the perturbative corrections implemented for the total rate and the moments is given in Tab. 1. This document accompanies the first release of `kolya` and details the specifics of the code. Although this paper represents a reference for future analyses of Belle II measurements of inclusive $B \to X_c l \bar{\nu}_l$ decays and gives a first outlook of `kolya` with basic examples to try in a Jupyter notebook, it is not meant to be a review article on semileptonic decays. To obtain a deeper understanding of the scientific part, the user is referred to e.g. Refs. [15–17]. The software `kolya` complements in scope several other open-source packages in HEP, in particular `flavio` [18], `EOS` [19], `HEPfit` [20], `HAMMER` [21] and `SuperIso` [22].

This article is structured as follows. In Sec. 2, we present the definitions of observables. Their implementation in the code is discussed in Sec. 3 where we discuss various ingredients implemented and quote the original references from which the material was obtained. The definition of the effective Hamiltonian parametrizing NP effects is given in Sec. 4. Section 5 focuses on basic usage of the code, illustrating the installation, the classes implemented in the code, use of the code for calculating Γ_{sl} and the moments together with details about our validation of the code. We close in Sec. 6 with an outlook.

2 Definitions

We focus on the the inclusive semileptonic decay of a B meson,

$$\overline{B}(p_B) \to X_c(p_X) \, l(p_l) \, \bar{\nu}_l(p_\nu) \text{ with } l = e, \mu, \tag{1}$$

in its rest frame where $p_B = (M_B, \vec{0})$. We assume that the charged lepton and the neutrino are massless. Let $q = p_l + p_\nu$ be the sum of the neutrino and the charged lepton momenta. $p_X = p_B - q$ is the total momentum of the hadronic system X_c. The energy of e or μ in the rest frame of the B meson is E_l.

In the HQE framework, predictions can be made for various differential rates with respect to the total leptonic energy $q_0 = E_l + E_\nu$, the charged lepton's energy E_l and q^2. However, these predictions cannot be directly compared to data on a point-by-point basis, as the phase space region allowed at the parton level is smaller than the physical one. Moreover, close to the endpoint, the $1/m_b$-suppressed terms contain singularities. Consequently, the inclusive process $\overline{B} \to X_c l \bar{\nu}_l$ must be compared to theoretical predictions of integrated quantities, such as the total rate

$$\Gamma_{\text{sl}} = \int \frac{\mathrm{d}^3\Gamma}{\mathrm{d}q^2 \, \mathrm{d}q_0 \, \mathrm{d}E_l} \mathrm{d}q^2 \, \mathrm{d}q_0 \, \mathrm{d}E_l, \tag{2}$$

or partial decay rate with a lower cut on q^2 or E_l:

$$\Delta\Gamma_{\text{sl}}(q^2_{\text{cut}}) = \int_{q^2 \geq q^2_{\text{cut}}} \frac{d\Gamma}{dq^2}\, dq^2 , \qquad\qquad \Delta\Gamma_{\text{sl}}(E_{\text{cut}}) = \int_{E_l \geq E_{\text{cut}}} \frac{d\Gamma}{dE_l}\, dE_l. \tag{3}$$

The moments of normalized differential distributions can be also calculated in the framework of the HQE. They are defined by

$$\langle (O)^n \rangle_{\text{cut}} = \int_{\text{cut}} (O)^n \frac{d\Gamma}{dO}\, dO \bigg/ \int_{\text{cut}} \frac{d\Gamma}{dO}\, dO , \tag{4}$$

where O strands for E_l, q^2 or M_X^2, and $d\Gamma/dO$ is the spectrum for the observable O. The term "cut" in the integrals refers to a selection threshold on q^2 or on E_l. Theoretically, the dependence of the moments on such a lower cut offers additional insights into the HQE parameters, improving their extraction in the global fits. Experimentally, the full spectrum is often not measurable due to detector acceptance. For instance, at the B-factories electrons are selected with $E_l \geq E_{\text{cut}}$ and $E_{\text{cut}} \simeq 0.5$ GeV to suppress the background.

Starting from $n = 2$, one normally compares data for centralized moments. For the charged lepton energy spectrum, the centralized moments are

$$\ell_1(E_{\text{cut}}) = \langle E_l \rangle_{E_l \geq E_{\text{cut}}} , \qquad\qquad \ell_n(E_{\text{cut}}) = \left\langle (E_l - \langle E_l \rangle)^n \right\rangle_{E_l \geq E_{\text{cut}}} \quad \text{for } n \geq 2. \tag{5}$$

For the hadronic invariant mass spectrum, they are given by

$$h_1(E_{\text{cut}}) = \langle M_X^2 \rangle_{E_l \geq E_{\text{cut}}} , \qquad\qquad h_n(E_{\text{cut}}) = \left\langle (M_X^2 - \langle M_X^2 \rangle)^n \right\rangle_{E_l \geq E_{\text{cut}}} \quad \text{for } n \geq 2 . \tag{6}$$

The first moment is the mean value of the differential distribution over the considered range and the second moment is its variance. Higher moments, e.g. $n = 3$ and $n = 4$, are also measured. In addition, the measurement of the q^2 moments, defined by

$$q_1(q^2_{\text{cut}}) = \langle q^2 \rangle_{q^2 \geq q^2_{\text{cut}}} , \qquad\qquad q_n(q^2_{\text{cut}}) = \left\langle (q^2 - \langle q^2 \rangle)^n \right\rangle_{q^2 \geq q^2_{\text{cut}}} \quad \text{for } n \geq 2 . \tag{7}$$

was suggested in Ref. [23]. It was shown that q^2 moments are invariant under reparametrization and therefore depending on a reduced set of HQE parameters (which we will introduce in Sec. 3), like the total rate (see Refs. [23, 24]). The definition of the q^2 moments do not incorporate a the cut $E_l > E_{\text{cut}}$ since it would break reparametrization invariance (RPI). In order to preserve RPI, the selection $q^2 > q^2_{\text{cut}}$ was proposed. This also implies an indirect constraint on the energy of the charged lepton through

$$E_l \geq \frac{M_B^2 + q^2_{\text{cut}} - M_D^2 - \lambda^{1/2}(M_B^2, q^2_{\text{cut}}, M_D^2)}{2M_B} , \tag{8}$$

where $\lambda(a, b, c) = a^2 + b^2 + c^2 - 2ab - 2ac - 2bc$ is the Källén function. Therefore, a cut on q^2 can effectively exclude low-energy electrons in the experimental analysis, just as a cut on E_l would.

3 Implementation in the SM

3.1 Building blocks

We use the heavy quark expansion (HQE) and express the total semileptonic width $\Gamma_{\rm sl}$ and the kinematic moments as a double expansion in $1/m_b$ and α_s. While working with the HQE, it is often advantageous to consider dimensionless quantities normalized w.r.t. the bottom quark mass m_b. We will denoted them with a "hat($\hat{}$)": e.g. $\hat{q}^2 = q^2/m_b^2, \hat{E}_l = E_l/m_b$.

As a starting point, we define the following building blocks:

$$Q_{ij} = \frac{1}{\Gamma_0} \int_{\rm cut} dE_l \, dq_0 \, dq^2 \, (q^2)^i (q_0)^j \frac{d^3\Gamma}{dq^2 \, dq_0 \, dE_l} \,, \tag{9}$$

where $q_0 = (E_l + E_\nu) = v \cdot q$ is the total leptonic energy with $v = p_B/M_B$ and q^2 is the leptonic invariant mass. Schematically, we write

$$
\begin{aligned}
Q_{ij} = (m_b)^{2i+j} \bigg[& Q_{i,j}^{(0)} + Q_{i,j}^{(1)} \frac{\alpha_s(\mu_s)}{\pi} + Q_{i,j}^{(2)} \left(\frac{\alpha_s(\mu_s)}{\pi} \right)^2 + \frac{\mu_\pi^2}{m_b^2} \left(Q_{i,j,\pi}^{(0)} + Q_{i,j,\pi}^{(1)} \frac{\alpha_s(\mu_s)}{\pi} \right) \\
& \frac{\mu_G^2(\mu_b)}{m_b^2} \left(Q_{i,j,G}^{(0)} + Q_{i,j,G}^{(1)} \frac{\alpha_s(\mu_s)}{\pi} \right) + \frac{\rho_D^3(\mu_b)}{m_b^3} \left(Q_{i,j,D}^{(0)} + Q_{i,j,D}^{(1)} \frac{\alpha_s(\mu_s)}{\pi} \right) \\
& + \frac{\rho_{LS}^3(\mu_b)}{m_b^3} \left(Q_{i,j,LS}^{(0)} + Q_{i,j,LS}^{(1)} \frac{\alpha_s(\mu_s)}{\pi} \right) + O\left(\frac{1}{m_b^4} \right) \bigg],
\end{aligned}
\tag{10}
$$

where

$$\Gamma_0 = \frac{m_b^5 G_F^2 A_{\rm ew}}{192\pi^3} |V_{cb}|^2 \,. \tag{11}$$

The factor $A_{\rm ew} = 1.01435$ stems from short-distance radiative corrections at the electroweak scale [25]. $\alpha_s \equiv \alpha_s^{(n_f)}(\mu_s)$ is the strong coupling constant taken with n_f active quarks and at the renormalization scale μ_s. To leading order in $1/m_b$, the heavy B meson decay coincides with the decay of a free bottom computed in perturbative QCD. Starting from $O(1/m_b^2)$ the predictions depend on a set of HQE parameters: non-perturbative matrix elements of local operators. These are denoted by $\mu_\pi^2, \mu_G^2, \rho_D^3, \rho_{LS}^3$. The tree-level expressions are known also to higher orders in $1/m_b$ (see Refs. [23, 26–28]). They are implemented in kolya up to $1/m_b^5$. However, in (10) they are omitted to keep a compact notation. The explicit definitions of the HQE parameters up to $1/m_b^5$ are reported in Appendix A. The HQE parameters in (10) are quoted in what we refer to as the "historical" basis employed in e.g. [29, 30]. For RPI quantities, like q^2 moments, it is however useful to work in the RPI basis, which has a reduced number of parameters [23, 31]. The differences between these two bases are detailed in Ref.[31, 32].

The functions denoted by Q_{ij} are the fundamental building blocks necessary to assemble the predictions for the centralized moments q_i in (7) and h_i in (6). They all depend on the mass ratio

$$\rho \equiv \frac{m_c}{m_b} \,, \tag{12}$$

where m_c and m_b refer to the on-shell masses of the charm and bottom quark. In (9), the subscript "cut" refers to certain restrictions in the phase-space integration. For the prediction of q_i in (7), we

apply the cut $q^2 > q_{\text{cut}}^2$ so that various build blocks in (10) depend on ρ and $\hat{q}_{\text{cut}}^2$: $Q_{ij} \to Q_{ij}(\rho, \hat{q}_{\text{cut}}^2)$. For the hadronic moments h_n in (6), the restriction is on the electron energy $E_l > E_{\text{cut}}$, so that the building blocks are functions of ρ and $\hat{E}_{\text{cut}}$: $Q_{ij} \to Q_{ij}(\rho, \hat{E}_{\text{cut}})$ (see Eq. (18) for the relation between h_n and Q_{ij}).

The QCD corrections depend also on the renormalization scale μ_s of the strong coupling constant starting at $O(\alpha_s^2)$. The functions $Q_{i,j,G}, Q_{i,j,D}$ and $Q_{i,j,LS}$ depend on μ_b, the scale at which the Wilson coefficients of the HQET Lagrangian are matched onto QCD, starting from $O(\alpha_s)$.

To construct the centralized electron energy moments in (5), we consider the moments of the charged-lepton energy $E_l = p_l \cdot v$ within the HQE:

$$
\begin{aligned}
L_i = & \frac{1}{\Gamma_0} \int_{E_l \geq E_{\text{cut}}} \mathrm{d}E_l \, \mathrm{d}q_0 \, \mathrm{d}q^2 \, (E_l)^i \frac{\mathrm{d}^3 \Gamma}{\mathrm{d}q^2 \, \mathrm{d}q_0 \, \mathrm{d}E_l} \\
= & (m_b)^i \Bigg[L_i^{(0)} + L_i^{(1)} \frac{\alpha_s(\mu_s)}{\pi} + L_i^{(2)} \left(\frac{\alpha_s(\mu_s)}{\pi} \right)^2 + \frac{\mu_\pi^2}{m_b^2} \left(L_{i,\pi}^{(0)} + L_{i,\pi}^{(1)} \frac{\alpha_s(\mu_s)}{\pi} \right) \\
& + \frac{\mu_G^2(\mu_b)}{m_b^2} \left(L_{i,G}^{(0)} + L_{i,G}^{(1)} \frac{\alpha_s(\mu_s)}{\pi} \right) + \frac{\rho_D^3(\mu_b)}{m_b^3} \left(L_{i,D}^{(0)} + L_{i,D}^{(1)} \frac{\alpha_s(\mu_s)}{\pi} \right) \\
& + \frac{\rho_{LS}^3(\mu_b)}{m_b^3} \left(L_{i,LS}^{(0)} + L_{i,LS}^{(1)} \frac{\alpha_s(\mu_s)}{\pi} \right) + O\left(\frac{1}{m_b^4} \right) \Bigg],
\end{aligned}
\tag{13}
$$

where in this case we allow for a cut on E_l only. All functions L_i depend on ρ and the cut E_{cut}: $L_i \to L_i(\rho, \hat{E}_{\text{cut}})$.

The total semileptonic rate corresponds to

$$
\Gamma_{\text{sl}} = \frac{m_b^5 G_F^2 A_{\text{ew}}}{192 \pi^3} |V_{cb}|^2 Q_{0,0}(\rho, 0) = \Gamma_0 \, L_0(\rho, 0),
\tag{14}
$$

with no cut applied, namely $E_{\text{cut}} = q_{\text{cut}}^2 = 0$. For the partial decay width, we similarly have

$$
\Delta\Gamma_{\text{sl}}(E_{\text{cut}}) = \Gamma_0 \, L_0(\rho, \hat{E}_{\text{cut}}), \qquad\qquad \Delta\Gamma_{\text{sl}}(q_{\text{cut}}^2) = \Gamma_0 \, Q_{0,0}(\rho, \hat{q}_{\text{cut}}^2).
\tag{15}
$$

The ratios defined in (4) correspond to

$$
\langle (q^2)^n \rangle = \frac{Q_{n,0}}{Q_{0,0}}, \qquad\qquad \langle E_l^n \rangle = \frac{L_n}{L_0}.
\tag{16}
$$

The centralized moments are obtained by inserting the double expansions of (10) or (13) into (4-7) and re-expanding in α_s and $1/m_b$ up to the relevant order. To assemble the M_X^2 moments, we express the hadronic invariant mass in terms of the parton level quantities in the B rest frame:

$$
M_X^2 = (M_B v - q)^2 = M_B^2 + q^2 - 2M_B q_0.
\tag{17}
$$

The moments of M_X^2 are obtained as linear combinations of the mixed moments $Q_{i,j}$:

$$
\begin{aligned}
M_n = & \frac{1}{\Gamma_0} \int_{E_l \geq E_{\text{cut}}} \mathrm{d}E_l \, \mathrm{d}q_0 \, \mathrm{d}q^2 \, (M_B^2 - 2M_B q_0 + q^2)^n \frac{\mathrm{d}^3 \Gamma}{\mathrm{d}E_l \, \mathrm{d}q_0 \, \mathrm{d}q^2} \\
= & \sum_{i=0}^{n} \sum_{j=0}^{i} \binom{n}{i} \binom{i}{j} (M_B^2)^{n-i} (-2M_B)^{i-j} Q_{j,i-j},
\end{aligned}
\tag{18}
$$

$\Gamma_{\rm sl}$	tree	α_s	α_s^2	α_s^3
Partonic		[34]	[35–38]	[39]
μ_π^2, μ_G^2	[1, 2]	[40–43]		
ρ_D^3, ρ_{LS}^3	[44]	[45]		
$1/m_b^4, 1/m_b^5$	[23, 26–28]			

$q_n(q_{\rm cut}^2)$	tree	α_s	α_s^2	
Partonic		[45, 46]	[47]	
μ_G^2, μ_π^2	[1, 2]	[41, 42]		
ρ_D^3, ρ_{LS}^3	[44]	[45]		
$1/m_b^4, 1/m_b^5$	[23, 28]			

$\ell_n(E_{\rm cut}), h_n(E_{\rm cut})$	tree	α_s	$\alpha_s^2\beta_0$	α_s^2
Partonic		[46, 48, 49]	[46]	[33][*]
μ_G^2, μ_π^2	[1, 2]	[42, 50]		
ρ_D^3, ρ_{LS}^3	[44]			
$1/m_b^4, 1/m_b^5$	[26–28]			

Table 1: Schematic overview of the perturbative corrections implemented for the rate $\Gamma_{\rm sl}$, the q^2 moments, the E_l and M_X^2 moments. (*) The α_s^2 corrections to h_n and ℓ_n are only available for several ρ and $E_{\rm cut}$ values in [33].

and $\langle (M_X^2)^n \rangle = M_n/M_0$.

    In `kolya`, we first implement all building blocks introduced above, corresponding to the on-shell scheme for both m_b and m_c. We collect in the Tab. 1 the list of references from which the various building blocks are retrieved. The implementation of the building blocks is described in Sec. 3.2. The implementation of the NNLO corrections to E_l and M_X moments, based on the results published in Ref. [33], requires a dedicated discussion in sections 3.3 and 3.4. In Sec. 3.5, we perform a scheme change to the kinetic scheme to obtain the final prediction for the total rate and the centralized moments.

3.2 Analytic expressions and grids for QCD corrections

In `kolya`, the tree level expressions up to $O(1/m_b^5)$ (see Refs. [23, 26–28]) are implemented in an exact analytic form. For example, the tree-level expressions at leading order in $1/m_b$ for $L_i^{(0)}(\rho, \hat{E}_{\rm cut})$ in (13) are coded in Python as follows.

```python
from numba import jit
import math

@jit(cache=True, nopython=True)
def L_0(i,elcuthat,r,dEl,dr):
    """ Tree level (partonic) for El moments and their derivatives """
```

```
y    = 2*elcuthat
logy = math.log((1-y)/r**2)
# tree level function
if (dEl == 0 and dr == 0 and i==0):
    return (1-8*r**2-6*r**4+12*logy*r**4+4*r**6-r**8
        -(2*r**6)/(-1+y)**2-(6*r**4*(1
        +r**2))/(-1+y)
        -2*r**4*(-3+r**2)*y+2*(-1+r**2)*y**3+y**4)
if (dEl == 0 and dr == 0 and i==1):
    return (3*logy*r**4*(3+r**2)+(7-75*r**2-180*r**4
        +120*r**6-15*r**8+3*r**10)/20
        -r**6/(-1+y)**2-(r**4*(3+5*r**2))/(-1+y)
        +6*r**4*y-(r**4*(-3+r**2)*y**2)/2
        +(3*(-1+r**2)*y**4)/4+(2*y**5)/5)
...
```

where the arguments of L_0 refer to the moment $i \in [0, \ldots, 4]$, the normalized electron energy cut $\hat{E}_{\text{cut}} = E_{\text{cut}}/m_b$ (elcuthat) and the mass ratio $\rho = m_c/m_b$ indicated by (r). The additional two arguments (dEl,dr) are positive integers referring to the derivatives of $L_i(\rho, \hat{E}_{\text{cut}})$ w.r.t. to $\hat{E}_{\text{cut}}$ or ρ. These derivatives are required when expressing the predictions in the kinetic scheme (see detailed discussion in Sec. 3.5). The tree level expressions up to $O(1/m_b^3)$ for the moments are implemented in the files Q2moments_SM.py, Elmoments_SM.py and MXmoments_SM.py. The power corrections of order $1/m_b^4$ and $1/m_b^5$ are given in separate files Q2moments_HO.py, Elmoments_HO.py and MXmoments_HO.py. For the q^2 moments, the higher power corrections are in Q2moments_HO_RPI.py

In order to use kolya for phenomenological applications, for instance to perform a fit, it is important to ensure adequate speed for the numerical evaluations. To this end, our implementation utilizes the library Numba [51], which translates Python functions to optimized machine code at runtime using the standard LLVM compiler library. The functions decorated with @jit, as shown in the example above, are compiled to machine code "just-in-time" for execution at native machine-code speed. Numba-compiled routines in Python approach the speed of C or FORTRAN.

Let us now discuss the implementation of the QCD corrections in kolya. For Γ_{sl}, the functions in (10) depend only on the mass ratio ρ and we implement the NLO corrections up to $1/m_b^3$ using the analytic expressions given in Ref. [45]. At NNLO in the free bottom quark approximation, there are asymptotic expansions available either in the limit $\rho \to 0$ [35, 36] or $\delta = 1 - \rho \to 0$ [37]. Recently, analytic expressions for the NNLO corrections written in terms of iterated integrals were presented in Ref. [38]. In kolya we use the expressions expanded in terms of $\delta = 1 - \rho$ up to δ^{46} given in Ref. [38]. For the third-order correction to Γ_{sl} we implement the asymptotic expansion up to δ^{12} computed in Ref. [39]. The total rate is implemented up to $1/m_b^3$ in TotalRate_SM.py, the higher power corrections are given in TotalRate_HO.py and TotalRate_HO_RPI.py.

We use interpolation grids to implements most of the QCD corrections for the moments. Specifically, we use grids for all NLO corrections, the NNLO corrections to the q^2 moments and the so-called BLM corrections (of order $\alpha_s^2 \beta_0$) [52] to the E_l and M_X moments.

Analytic expressions for the differential rate $d\Gamma/dq^2$ for a free bottom quark are available at NLO from Ref. [53] and at NNLO from Ref. [47]. The NLO corrections to leading order in $1/m_b$ were computed also for the triple differential rate in Ref. [46, 48]. The NLO corrections to μ_G^2 and ρ_D^3 for the q^2 spectrum have been presented in Ref. [45]. By making use of reparametrization invariance [24],

one can also show that in the on-shell scheme

$$-\frac{1}{2}Q_{i0}^{(n)}(\rho, q_{\text{cut}}^2) = Q_{i0,\pi}^{(n)}(\rho, q_{\text{cut}}^2), \qquad\qquad -Q_{i0,G}^{(n)}(\rho, q_{\text{cut}}^2) = Q_{i0,LS}^{(n)}(\rho, q_{\text{cut}}^2), \qquad (19)$$

to all orders $n \geq 0$ in the perturbative expansion. Therefore, the functions Q_{ij} entering (10) can be computed for any q_{cut}^2 and ρ via:

$$Q_{i0}(\rho, \hat{q}_{\text{cut}}^2) = \frac{1}{\Gamma_0} \int_{\hat{q}_{\text{cut}}^2}^{(1-\rho)^2} (\hat{q}^2)^i \frac{d\Gamma}{d\hat{q}^2} \, d\hat{q}^2 . \qquad (20)$$

The differential rate $d\Gamma/d\hat{q}^2$ at higher orders is expressed in terms of functions defined via iterated integrals like harmonic polylogarithms (HPLs) [54] and generalized polylogarithms (GPLs) [55, 56]. It is not convenient to integrate the differential rate numerically "on-the-fly" since there are several HPLs and GPLs whose evaluation (for instance with GiNaC [57]) is time-consuming. For this reason, we opt to implement all higher QCD corrections for the moments not in an exact form, but through Chebyshev two-dimensional grids. The functions implemented via grids are $Q_{ij}^{(1)}, Q_{ij}^{(2)}, Q_{ij,\pi}^{(1)} Q_{ij,G}^{(1)}, Q_{ij,D}^{(1)}, Q_{ij,LS}^{(1)}$ in (10) and $L_{ij}^{(1)}, L_{ij}^{(2)}, L_{ij,\pi}^{(1)} L_{ij,G}^{(1)}$ in (13).

Let us briefly review here how a generic function $f(x)$ can be discretized on a grid consisting of the so-called Chebyshev points (for more details see e.g. Ref. [58]). The idea is to evaluate $f(x)$ in n points x_n corresponding to the zeros of the Chebyshev polynomial $T_n(x)$ of degree n:

$$x_k = \cos\left(\frac{\pi(k-1/2)}{n}\right) \text{ for } k = 1, \dots, n . \qquad (21)$$

If $f(x)$ is an arbitrary function defined on the domain $x \in [-1, 1]$, we calculate the coefficients c_j, with $j = 0, \dots, n-1$ given by

$$c_j = \frac{2}{n} \sum_{k=1}^{n} f(x_k) T_j(x_k), \qquad (22)$$

which can be employed to construct the polynomial

$$\tilde{f}(x) = -\frac{1}{2}c_0 + \sum_{k=1}^{n-1} c_k T_k(x) \approx f(x), \qquad (23)$$

approximating $f(x)$ in the interval $[-1, 1]$. In particular $\tilde{f}(x) = f(x)$ for all n zeros of $T_n(x)$. In case the function to interpolate $f(y)$ is defined between two arbitrary limits, e.g. $y \in [a, b]$, we apply the variable transformation

$$x = \frac{y - \frac{1}{2}(b+a)}{\frac{1}{2}(b-a)}, \qquad (24)$$

and then perform the interpolation in x as before.

In our setup, the functions to interpolate depend on ρ and $\hat{q}_{\text{cut}}^2$ (or ρ and $\hat{E}_{\text{cut}}$) and can be obtained via two consecutive one-dimensional Chebyshev interpolations. First, we discretize the interval $\rho \in [1/6, 1/3]$ (relevant for the phenomenology) in n_ρ points distributed according to (21). Then for each $\rho_k \in \{\rho_1, \dots, \rho_{n_\rho}\}$, we discretize $\hat{q}_{\text{cut}}^2$ or $\hat{E}_{\text{cut}}$ into further n_{cut} points within the allowed range: $\hat{q}_{\text{cut}}^2 \in [0, (1-\rho)^2]$ or $\hat{E}_{\text{cut}} \in [0, (1-\rho^2)/2]$. An example of how the discretization is performed is shown in Fig. 1, for a grid in ρ and $\hat{q}_{\text{cut}}^2$ with $n_\rho = n_{\text{cut}} = 10$.
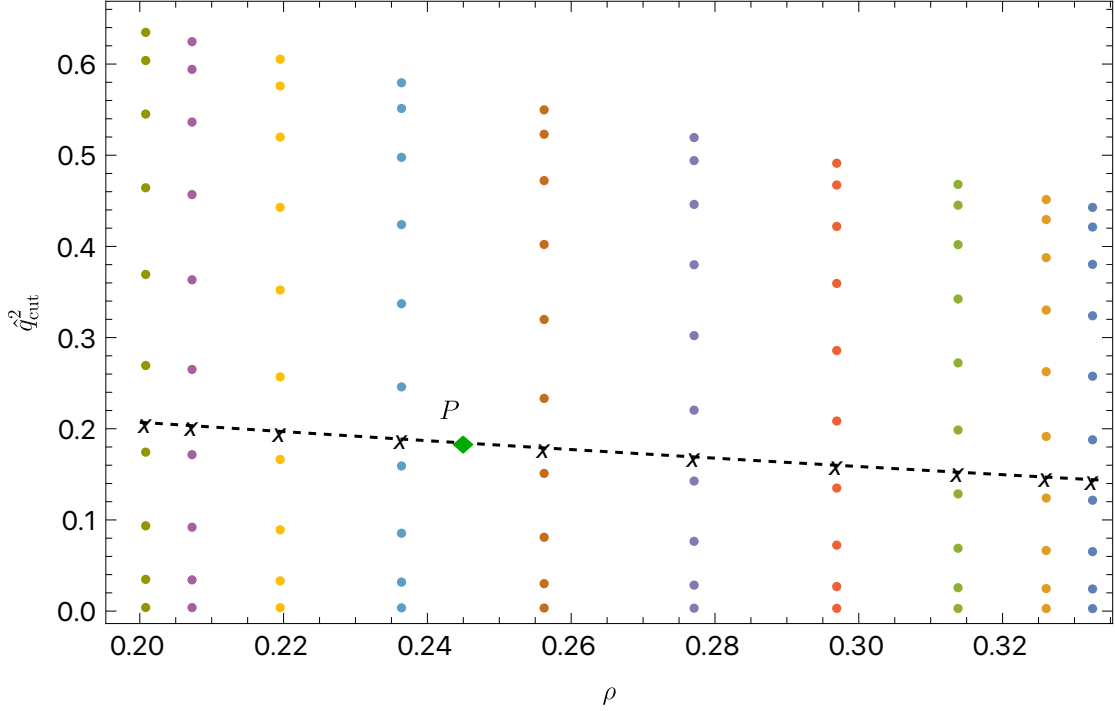
Figure 1: Example of how we discretize functions depending on two variables using Chebyshev nodes. In the plot, we consider the grid for Q_{ij} depending on ρ and $\hat{q}_{\text{cut}}^2$.

To estimate the function at a new point $P = (\rho_P, \hat{q}_P^2)$ (the green diamond in Fig. 1), we proceed as follows. For each $\rho_k, k = 1, \dots, n_\rho$, we calculate $Q(\rho_k, \hat{q}_P^2 (1 - \rho_P)^2 / (1 - \rho_k)^2)$ using one-dimensional interpolations in the variable $\hat{q}_{\text{cut}}^2$. These values calculated at fixed ρ are shown by black crosses in Fig. 1. Afterwards, we use them as nodes for a second interpolation this time along the ρ direction, as displayed by the black dashed line in Fig. 1. The second interpolation yields the final estimate for $f(\rho_P, \hat{q}_P^2)$.

For the implementation of the QCD corrections to h_i and ℓ_i, which depend on ρ and E_{cut}, we also use Chebyshev interpolation grids. At NLO, it is possible to write the differential rate $d\Gamma/dE_l$ in a closed analytic form at NLO [49] for a free quark. To compute $Q_{ij}(\rho, \hat{E}_{\text{cut}})$ and $L_i(\rho, \hat{E}_{\text{cut}})$ at $O(\alpha_s)$ we perform the one variable integration, as for instance

$$L_i(\rho, \hat{E}_{\text{cut}}) = \frac{1}{\Gamma_0} \int_{\hat{E}_{\text{cut}}}^{(1-\rho^2)/2} (\hat{E}_l)^i \, \frac{d\Gamma}{d\hat{E}_l} \, d\hat{E}_l \, ,$$

$$Q_{ij}(\rho, \hat{E}_{\text{cut}}) = \frac{1}{\Gamma_0} \int_{\hat{E}_{\text{cut}}}^{(1-\rho^2)/2} \frac{dQ_{ij}}{d\hat{E}_l} \, d\hat{E}_l \, . \tag{25}$$

In the last equation, we define

$$\frac{dQ_{ij}}{d\hat{E}_l} = \int (q^2)^i (q_0)^j \, \frac{d^3\Gamma}{dq^2 \, dq_0 \, dEl} \, dq^2 dq_0 \tag{26}$$

which can also be computed analytically up to NLO following Ref. [49, 59].

At order $1/m_b^2$, the NLO corrections $Q_{ij}(\rho, \hat{E}_{\text{cut}})$ and $L_i(\rho, \hat{E}_{\text{cut}})$ can be calculated from the triple differential distributions given in Refs. [41, 42] by performing the phase space integration numerically as described in Ref. [60]. The NLO corrections to ρ_D^3 and ρ_{LS}^3 are not known at the moment.

The values of the coefficients c_j in (22) for all grids are stored in the directory `grids` as multidimensional arrays. The routines which perform the interpolation of the NLO and NNLO corrections at $O(1/m_b^0)$ are implemented in `NLOpartonic.py` and `NNLOpartonic.py`. The routines for the NLO corrections to the power-suppressed terms are given in `NLOpw.py`.

We validate the grid implementation by generating 100 random points in the two-dimensional plane $(\rho, \hat{q}_{\text{cut}}^2)$ or $(\rho, \hat{E}_{\text{cut}})$. For each point, we compare the approximation provided by the grids and high-precision evaluations obtained with Mathematica. We verify that the two estimates differ by less than 10^{-5} for the points considered.

3.3 NNLO corrections to the lepton energy moments

The NNLO corrections to the ℓ_i moments are not known in a closed form. As discussed, the BLM corrections are implemented through interpolation grids. The remaining "non-BLM" corrections are only known for specific values of ρ and E_{cut} from Ref. [33]. Their functional form can be obtained from a two-dimensional fit to these data points. In order to perform this fit, we write:

$$
\ell_n(\rho, E_{\text{cut}}) = (m_b)^n \Bigg[Y_n^{(0)} + Y_n^{(1)} \frac{\alpha_s(\mu_s)}{\pi} + \Big(\beta_0 Y_n^{(2,\text{BLM})} + Y_n^{(2,\text{nonBLM})} \Big) \left(\frac{\alpha_s(\mu_s)}{\pi} \right)^2
$$

$$
+ O\left(\frac{1}{m_b}, \left(\frac{\alpha_s(\mu_s)}{\pi} \right)^3 \right) \Bigg], \tag{27}
$$

where the E_{cut}- and ρ-dependence of $Y_n^{(i)}$ is implied and $Y_n^{(0)}$ is the partonic contribution without any α_s or $1/m_b$ corrections. In terms of the building-blocks defined in (13), we can write the non-BLM terms as:

$$
Y_1^{(2,\text{nonBLM})} = \frac{L_1^{(2,\text{nonBLM})}}{L_0^{(0)}} - \frac{L_0^{(2,\text{nonBLM})} L_1^{(0)}}{(L_0^{(0)})^2} ,
$$

$$
Y_2^{(2,\text{nonBLM})} = \frac{L_2^{(2,\text{nonBLM})}}{L_0^{(0)}} - 2 \frac{L_1^{(2,\text{nonBLM})} L_1^{(0)}}{(L_0^{(0)})^2} + L_0^{(2,\text{nonBLM})} \left(2 \frac{(L_1^{(0)})^2}{(L_0^{(0)})^3} - \frac{L_2^{(0)}}{(L_0^{(0)})^2} \right) ,
$$

$$
Y_3^{(2,\text{nonBLM})} = \frac{L_3^{(2,\text{nonBLM})}}{L_0^{(0)}} - 3 \frac{L_2^{(2,\text{nonBLM})} L_1^{(0)}}{(L_0^{(0)})^2} + 3 L_1^{(2,\text{nonBLM})} \left(2 \frac{(L_1^{(0)})^2}{(L_0^{(0)})^3} - \frac{L_2^{(0)}}{(L_0^{(0)})^2} \right)
$$

$$
+ L_0^{(2,\text{nonBLM})} \left(-6 \frac{(L_1^{(0)})^3}{(L_0^{(0)})^4} + 6 \frac{L_1^{(0)} L_2^{(0)}}{(L_0^{(0)})^2} - \frac{L_3^{(0)}}{(L_0^{(0)})^2} \right) . \tag{28}
$$

Ref. [33] gives the $L_n^{(2,\text{nonBLM})}$ terms at $\rho = \{0.20, 0.22, 0.24, 0.25, 0.26, 0.28\}$ and $y = \{0, 0.1, \ldots, 0.7\}$, with $y \equiv 2\hat{E}_{\text{cut}}$. From these,[1] the non-BLM contributions $Y_n^{(2,\text{nonBLM})}$ to the ℓ_n moments are obtained

---

[1] Note that the $L_i^{(n)}$ defined in Ref. [33] are normalized to the total partonic rate without cut while we only normalize to Γ_0 defined in (14).

by combining with the tree-level building blocks $L_n^{(0)}$. In Ref. [61], these non-BLM contributions are studied in detail and compared to the effect of their BLM counterparts. We fit the values for $Y_n^{(2,\text{nonBLM})}$ assuming the following polynomial ansatz

$$Y_n(\rho, y) = \sum_{i=1}^{5} (a_{n,i} + b_{n,i}\rho)(y + \rho^2 - 1)^i , \qquad (29)$$

for each moment n. Following Ref. [61], we only include one power of ρ in our ansatz, but keep up terms up y^5 in our interpolating fit. In addition, the ansatz is chosen to ensure that the non-BLM corrections vanish at the end point $y = 1 - \rho^2$ [33]. We stress that our approach differs from [61] as we first construct $Y_n^{(2,\text{nonBLM})}$ in each available (ρ, y) point and then perform the analysis. Fitting first the $L_n^{(2,\text{nonBLM})}$ and then combining them resulted in strongly oscillating functions due to accidental cancellations. Fitting directly Y_n, we find

$$Y_1^{(2,\text{nonBLM})}(\rho, y) =$$

$$(72.57\rho - 25.62)(y + \rho^2 - 1)^5 + (177.60\rho - 64.65)(y + \rho^2 - 1)^4$$

$$+ (157.19\rho - 59.27)(y + \rho^2 - 1)^3 + (62.69\rho - 24.32)(y + \rho^2 - 1)^2$$

$$+ (11.25\rho - 4.35)(y + \rho^2 - 1) ,$$

$$Y_2^{(2,\text{nonBLM})}(\rho, y) =$$

$$(12.61 - 44.47\rho)(y + \rho^2 - 1)^5 + (32.28 - 112.34\rho)(y + \rho^2 - 1)^4$$

$$+ (29.36 - 100.36\rho)(y + \rho^2 - 1)^3 + (11.07 - 36.89\rho)(y + \rho^2 - 1)^2$$

$$+ (1.42 - 4.55\rho)(y + \rho^2 - 1) , \qquad (30)$$

and

$$Y_3^{(2,\text{nonBLM})}(\rho, y) =$$

$$(32.23\rho - 7.28)(y + \rho^2 - 1)^5 + (79.34\rho - 17.72)(y + \rho^2 - 1)^4$$

$$+ (67.06\rho - 14.67)(y + \rho^2 - 1)^3 + (21.52\rho - 4.49)(y + \rho^2 - 1)^2$$

$$+ (1.71\rho - 0.289)(y + \rho^2 - 1) . \qquad (31)$$

These functions are implemented in `kolya`. We do not assign any uncertainty to these fit functions.

In Fig. 2, we show our fit results for $Y_{1,2,3}^{(2,\text{nonBLM})}$, as a function of y for fixed $\rho = 0.22$ (solid blue line). Due to the fit ansatz, we observe a light oscillatory behavior as a function of y. In black, we show the constructed data points at fixed y obtained from [33]. The fit uncertainty is given by the red dotted line, which represents the 90% C.L. interval of the fit. Since we only have data points up to $y = 0.7$ and impose that the contribution vanishes at the endpoint, we notice large uncertainties towards higher y values. In a typical phenomenological analysis, missing higher order terms (here α_s^3) terms would be accounted for by varying the scale of α_s. The blue bland corresponds to the effect of such a scale variation from $[\alpha_s(m_b^{\text{kin}}/2), \alpha_s(2m_b^{\text{kin}})]$. For Y_1, we observe that the α_s variation covers
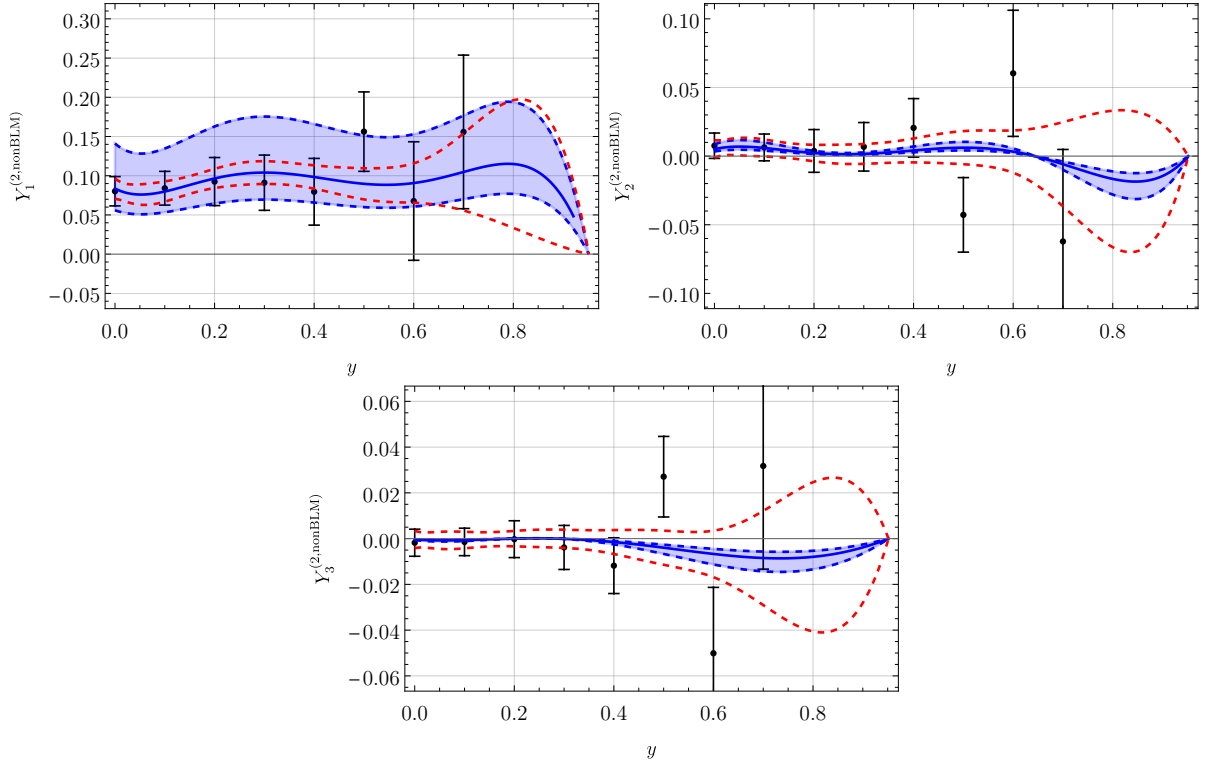
Figure 2: Results of our fitted non-BLM contributions $Y_n^{(2,\text{nonBLM})}$ (solid blue line) as a function of $y = 2E_{\text{cut}}/m_b$ for $\rho = 0.22$ as discussed in the text. Data points are constructed from [33].

the data points and fit uncertainty. For the higher moments, we observe that the fit uncertainty is higher than the α_s variation for large y. However, given the smallness of the BLM contributions to these moments, our default setting is to not include an additional uncertainty for these corrections.

3.4 NNLO corrections to the hadronic invariant mass moments

The nonBLM corrections to the hadronic moments can be obtained from Ref. [46]. However, as for the lepton energy moments, the additional "non-BLM" contributions are only known numerically for several values of ρ and E_{cut} [33]. In Ref. [33], building blocks H_{ij} are defined as

$$H_{ij}(\rho, \hat{E}_{\text{cut}}) = \frac{1}{\Gamma_0 L_0^{(0)}(\rho, 0)} \int_{E_{\text{cut}}} d\hat{E}_l \, d\hat{q}_0 \, d\hat{q}^2 \left(\frac{m_x^2 - m_c^2}{m_b^2} \right)^i \left(\frac{E_h}{m_b} \right)^j \frac{d^3\Gamma}{d\hat{E}_l \, d\hat{q}_0 \, d\hat{q}^2} \,, \tag{32}$$

where $L_0^{(0)}(\rho, 0)$ is defined in (13) and where m_x and E_x are partonic invariant mass and energy. A linear combination of these H_{ij} corresponds to our building blocks Q_{ij} defined in (18). This then allows us to calculate the non-BLM contributions to h_1 by combining different H_{ij} contributions. However, not all the Q_{ij} moments necessary to construct the hadronic moments h_2 and h_3 are calculated. In terms of the definition of [33], the non-BLM contributions of the combinations

$$H_{20}, \, H_{30}, \, H_{21} \,,$$

are missing.

In order to determine the effect of the non-BLM terms, we write

$$h_n(y,\rho) = X_n(y,\rho) + m_b^2 \left(\frac{\alpha_s(\mu_s)}{\pi}\right)^2 \left[\beta_0 X_n^{(2,\text{BLM})}(y,\rho) + X_n^{(2,\text{nonBLM})}(y,\rho)\right] , \tag{33}$$

where $X_n^{(2,\text{nonBLM})}$ is the non-BLM contribution, $X_n^{(2,\text{BLM})}$ the BLM contribution and X_n contains all other contributions. For h_1, we can now proceed as for the lepton energy moments. Here, we use

$$X_1^{(2,\text{nonBLM})}(\rho,y) = \sum_{i=0}^{5}(a_{1,i} + b_{1,i}\rho)y^i , \tag{34}$$

as our fit ansatz, where we do not require that the first moment vanishes at the end point. Using the data points from Ref. [33] (to be conservative, we assume a 1% uncertainty on all data points for which no uncertainty was given), we then find

$$\begin{aligned}
X_1^{(2,\text{nonBLM})}(y,\rho) = &\ y^5(33.13 - 97.49\rho) + y^4(224.82\rho - 71.16) \\
&+ y^3(49.01 - 160.90\rho) + y^2(37.72\rho - 10.92) \\
&+ y(0.73 - 2.56\rho) + 2.08\rho - 0.92 .
\end{aligned} \tag{35}$$

In Fig. 3, we show $X_1^{(2,\text{nonBLM})}$ as a function of y. As for the lepton energy moments, the red dotted line shows the fit uncertainty, which now diverges close to the end point. However, as usually the experimental data does not have lepton energy cuts $y > 0.8$ this feature does not pose a great problem. In addition, we notice the small uncertainty on the data points, clearly much smaller than the effect of the α_s variation (shown by the blue band). As such, we do not implement an additional fit uncertainty on (35) and implement this function into kolya. A similar fit was done in [62], where also ρ^2 terms were taken into account. We do not notice an improvement in the fit when including such terms and therefore use our minimal fit ansatz presented above.

For the h_2 and h_3 moments, we currently do not have sufficient information to perform a fit. However, for $y = 0$, we can use the expression for Q_{ij} from [59], where analytic expression in terms of $\delta = 1 - \rho$ were given. From these, we can then also determine the non-BLM effect over the BLM contribution. We find stable results for the different X_n. Using $\rho = 0.25$ for illustration, we have

$$\frac{X_1^{(2,\text{nonBLM})}}{\beta_0 X_1^{(2,\text{BLM})}} = -0.237 , \quad \frac{X_2^{(2,\text{nonBLM})}}{\beta_0 X_2^{(2,\text{BLM})}} = -0.227 , \quad \frac{X_3^{(2,\text{nonBLM})}}{\beta_0 X_3^{(2,\text{BLM})}} = -0.241 . \tag{36}$$

We do not quote an uncertainty on these values, which could be obtained from [59] by consider the effect of the highest power in the δ expansion. At the moment, we do not have more information for the X_2 and X_3 contributions at nonzero values of the lepton cut y. In [61], the non-BLM/BLM ratios of the available relevant combinations of H_{ij} moments were calculated. For the values in [33], it was found that this ratio is rather independent of y. This may indicate that also the missing terms are constant when considering ratios of non-BLM versus BLM corrections. In [61], this constant behavior was therefore assumed for the missing $H_{20,21,30}^{(2)}$ moments.[2] On the other hand, we note that the

---

[2]Using [59], we can calculate for the first time also the ratio of $H_{20}^{(2,\text{non}-\text{BLM})}/H_{20}^{(2,BLM)}$ and similarly for {21} and {30}. We confirm within uncertainties the ratios quoted in [61].
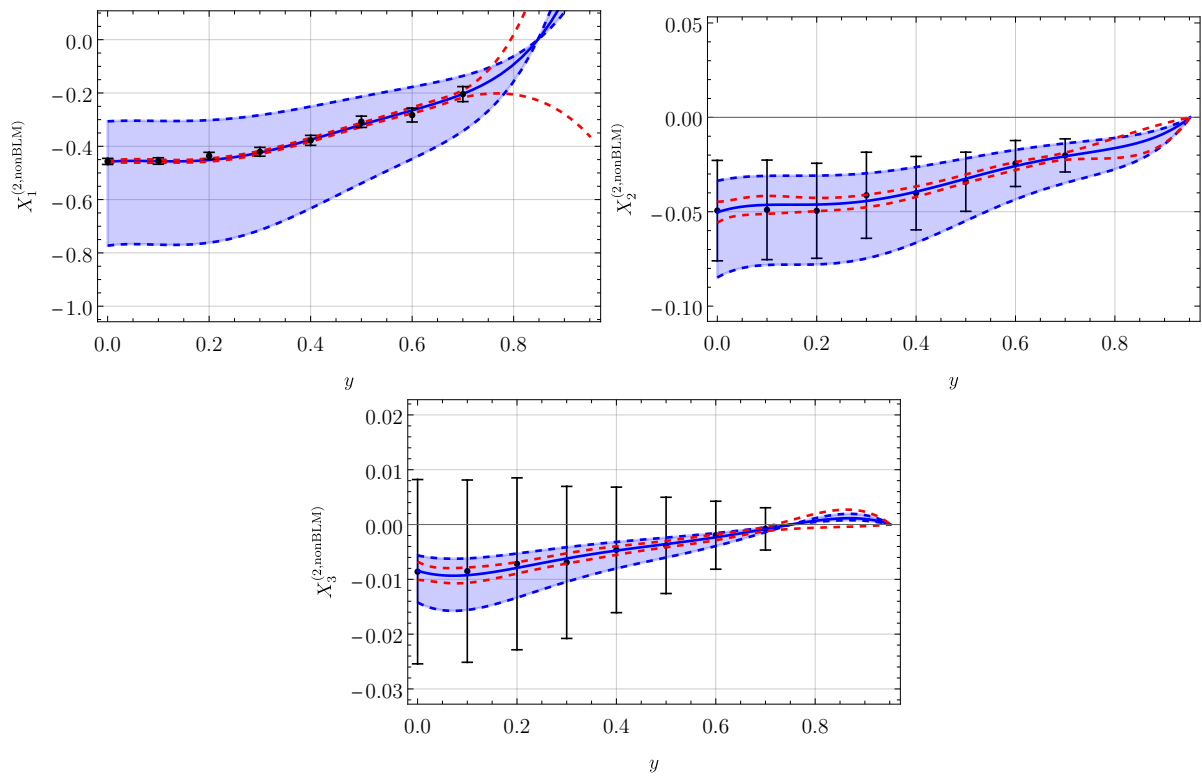
Figure 3: The non-BLM contributions $X_n^{(2,\text{nonBLM})}$ as a function of $y = 2E_{\text{cut}}/m_b$ for $\rho = 0.22$ as described in the text.

analytic results in [47] for the q^2 moments with a q^2 cut do not show a constant non-BLM over BLM behavior.

Nevertheless, to further study h_2 and h_3, we follow [61] and assume constant non-BLM over BLM ratios for the missing moments. We then find the constructed data points for $X_2^{(2,\text{nonBLM})}$ and $X_3^{(2,\text{nonBLM})}$ given in Fig. 3. In blue, we present the fit to the data using a similar ansatz as in (29) (ensuring also that the contribution vanishes at the end point). We note that the fit uncertainty (dotted red line) is very small and that also the α_s variation (blue band) is much smaller than the uncertainty on the data points. In addition, the assumption for the missing moments may bias our predictions. Therefore, we decide for the moment not to include any non-BLM corrections to the h_2 and h_3 hadronic mass moments.

3.5 Kinetic and $\overline{\text{MS}}$ scheme

The predictions for inclusive decayse must be formulated using a short-distance mass scheme for the charm and the bottom quarks, to ensure the cancellation of the leading renormalon divergence [63, 64]. In kolya, we adopt the kinetic scheme [5] for the bottom quark mass and the HQE parameters. In the kinetic scheme, we substitute the bottom quark pole mass is favour of the kinetic mass thorugh the relation

$$m_b^{\text{pole}} = m_b^{\text{kin}}(\mu) + [\overline{\Lambda}(\mu)]_{\text{pert}} + \frac{[\mu_\pi^2(\mu)]_{\text{pert}}}{2m_b^{\text{kin}}(\mu)} + O\left(\frac{1}{m_b^2}\right), \tag{37}$$

where the Wilsonian cutoff μ is a scale chosen such that $\Lambda_{\text{QCD}} \ll \mu \ll m_b$. The last two terms in (37) labelled by "pert" are the HQE parameters calculated in pertubative QCD with the Small Velocity sum rules [65]. The analytic expressions up to $O(\alpha_s^3)$ are given in Refs. [6, 8]. We stress that in our implementation, we follow [8] and include also the decoupling effects of the charm quark in the kinetic scheme. We show the effect of this decoupling numerically in Sec. 3.6. In the kinetic scheme, also the HQE parameters in (10) and (13) are redefined in the following way:

$$\mu_\pi^2(0) = \mu_\pi^2(\mu) - [\mu_\pi^2(\mu)]_{\text{pert}}, \qquad\qquad \rho_D^3(0) = \rho_D^3(\mu) - [\rho_D^3(\mu)]_{\text{pert}}. \tag{38}$$

These two definitions apply in principle to μ_π^2 and ρ_D^3 in the $m_b \to \infty$ limit. The analyses from Refs. [9, 29] define the HQE parameters at finite m_b and both setup do not include terms of order μ^3 from the redefinition of μ_π^2 and μ_G^2. Ref. [29] and [9] employ different operator basis, both implemented in kolya and referred as the "historical basis" and the "RPI" basis.

For the charm mass, we use the $\overline{\text{MS}}$ scheme. This is related to the pole charm mass via

$$m_c^{\text{pole}} = \overline{m}_c(\mu_c) \left[1 + \frac{\alpha_s(\mu_c)}{\pi} \left(\frac{4}{3} + L \right) \right.$$
$$\left. + \left(\frac{\alpha_s(\mu_c)}{\pi} \right)^2 \left(\frac{37}{24}L^2 + \frac{415}{72}L + \frac{779}{96} + \frac{\pi^2}{6} - \frac{\zeta_3}{6} + \frac{1}{9}\pi^2\log(2) \right) \right] + O(\alpha_s^3), \tag{39}$$

where $L = \log(\mu_c^2/\overline{m}_c^2(\mu_c))$.

In order to implement the scheme change for the quark masses and the HQE parameters, we replace m_b^{pole} and m_c^{pole} in the expressions for the centralized moments using (37) and (39) and expand the formulas as a series in α_s and $1/m_b$.

Hard coding these scheme changes would lead to huge expressions for the moments. To illustrate this, we write the mass conversion formulas schematically as follows:

$$m_c^{\text{pole}} = \overline{m}_c(1 + \alpha_s \delta_{m_c,1} + \alpha_s^2 \delta_{m_c,2} + \alpha_s^3 \delta_{m_c,3}) \,,$$
$$m_b^{\text{pole}} = m_b^{\text{kin}}(1 + \alpha_s \delta_{m_b,1} + \alpha_s^2 \delta_{m_b,2} + \alpha_s^3 \delta_{m_b,3}) \,. \tag{40}$$

The coefficients appearing at order α_s^n from the bottom and charm mass scheme conversion formulas are denoted by $\delta_{m_b,n}$ and $\delta_{m_c,n}$, respectively. Let us now consider a simple example with a function $f(\rho)$ depending only on the mass ratio. The scheme change for $f(\rho)$ up to $O(\alpha_s)$ reads

$$f(\rho) = f\left(\frac{m_c^{\text{pole}}}{m_b^{\text{pole}}}\right) = f\left(\overline{m}_c(1 + \alpha_s \delta_{m_c,1})\middle/ m_b^{\text{kin}}(1 + \alpha_s \delta_{m_b,1})\right)$$

$$= f\left(\frac{\overline{m}_c}{m_b^{\text{kin}}}\right) + \alpha_s f'\left(\frac{\overline{m}_c}{m_b^{\text{kin}}}\right)\frac{\overline{m}_c}{m_b^{\text{kin}}}(\delta_{m_c,1} - \delta_{m_b,1}) + O(\alpha_s^2) \,. \tag{41}$$

The function $f(\rho)$ and its derivative are the building blocks entering in `kolya`. This has several advantages with respect to hard coding the expression in.

In fact for the centralized moments, the role of $f(\rho)$ is played by the building blocks denoted by Q_{ij} and L_i on the r.h.s. of (10) and (13). They appear multiple times after the re-expansion of the ratios in (4) and the scheme change. Therefore, we find more convenient to calculate first all building blocks in (10) and (13) (and the necessary derivatives) and cache the results at a given value of ρ and the cut. Afterwards, we assemble the centralized moments using expressions similar to (41). This approach yields code with a much smaller size and improved evaluation time.

The coefficients $\delta_{m_b,n}$ and $\delta_{m_c,n}$ which enter in the scheme change are implemented up to $O(\alpha_s^3)$ in the file `schemechange_KINMS.py`. This allows in principle to easily adopt a different mass scheme from our default one, by changing the expression for $\delta_{m_b,n}$ and $\delta_{m_c,n}$ to the required mass scheme in a separate file.

3.6 Numerical results for lepton energy and hadronic mass moments

The accuracy of our moment predictions is summarized in Tab. 1. As the NNLO corrections to the q^2 moments are known analytically [47], and a detail discussion was given recently in that reference, we do not discuss these here in detail. However, since our implementation for the lepton energy and hadronic mass moments depends on our fit for the non-BLM contributions, it is interesting to give the relative contributions to the moments. These contributions are given in Tab. 2 for $E_{\text{cut}} = 1$ GeV. For the HQE parameters, we use $\mu_\pi^2 = 0.4$ GeV^2, $\mu_G^2 = 0.35$ GeV^2, $\rho_{LS}^3 = -0.15$ GeV^3, $\rho_D^3 = 0.2$ GeV^3. For the other input parameters, we employ $m_b^{\text{kin}}(1 \text{ GeV}) = 4.563$ GeV, $\overline{m}_c(3 \text{ GeV}) = 0.989$ GeV, $\alpha_s(m_b^{\text{kin}}) = 0.2182$. In Tab. 2, we also explicitly give the effect of including the charm decoupling in the kinetic scheme conversion. This contribution is labelled as Δ_c. Our results are in good agreement with Ref. [61] up to the non-BLM corrections. The Δ_c effects were not included in that reference. Currently, these contributions are automatically included at NNLO in `kolya`, and they cannot be separated from other contributions. We observe that Δ_c has a 0.1% effect and is of similar in size but opposite in sign as the non-BLM correction. Overall, the α_s^2 contributions have a 0.05% effect.

We provide additional numerical examples and comparisons with literature in the GitLab repository, in the Jupyter notebook `example-reproduce_literature.ipynb`.

Moment	tree	α_s	$\alpha_s^2 \beta_0$	α_s^2 scheme	α_s^2 non-BLM	$\alpha_s^2 \Delta_c$
ℓ_1 [GeV]	1.5650	1.5521	1.5540	1.5459	1.5480	1.5465
ℓ_2 [GeV^2]	0.0895	0.0870	0.0881	0.0861	0.0867	0.0863
ℓ_3 [GeV^3]	-0.0018	-0.0003	0.0004	0.0006	-0.0006	-0.0006
h_1 [GeV^2]	4.166	4.331	4.304	4.417	4.381	4.403
h_2 [GeV^4]	0.609	0.818	1.001	0.987	-	0.990
h_3 [GeV^6]	5.071	4.810	4.487	4.641	-	4.640

Table 2: Contributions of the different orders to the moments at $E_{\text{cut}} = 1$ GeV for $\mu_\pi^2 = 0.4$ GeV^2 , $\mu_G^2 = 0.35$ GeV^2 , $\rho_{LS}^3 = -0.15$ GeV^3 , $\rho_D^3 = 0.2$ GeV^3. For the other input parameters, we employ $m_b^{\text{kin}}(1 \text{ GeV}) = 4.563$ GeV, $\overline{m}_c(3 \text{ GeV}) = 0.989$ GeV, $\alpha_s(m_b^{\text{kin}}) = 0.2182$.

4 Extension to physics beyond the SM

In kolya, we also implement NP effects in $b \to c l \bar{\nu}_l$ decays following [12]. In order to parametrize effects beyond the SM, we use the weak effective theory (WET), an effective field theory valid below the EW scale written as an expansion in powers of the inverse electroweak scale $G_F = 1/(\sqrt{2}v^2)$. The effective Hamiltonian relevant for inclusive semileptonic B decays is

$$\mathcal{H}_{\text{eff}} = \frac{4 G_F V_{cb}}{\sqrt{2}} \left[\left(1 + C_{V_L}\right) \mathcal{O}_{V_L} + \sum_{i=V_R, S_L, S_R, T} C_i \mathcal{O}_i \right], \tag{42}$$

where we consider only the operators of dimension six contributing to the differential rate at tree level:

$$\mathcal{O}_{V_{L(R)}} = \left(\bar{c} \gamma_\mu P_{L(R)} b\right)\left(\bar{\ell} \gamma^\mu P_L \nu_\ell\right),$$
$$\mathcal{O}_{S_{L(R)}} = \left(\bar{c} P_{L(R)} b\right)\left(\bar{\ell} P_L \nu_\ell\right),$$
$$\mathcal{O}_T = \left(\bar{c} \sigma_{\mu\nu} P_L b\right)\left(\bar{\ell} \sigma^{\mu\nu} P_L \nu_\ell\right). \tag{43}$$

We define $\sigma^{\mu\nu} = \frac{i}{2}[\gamma^\mu, \gamma^\nu]$ and $P_{L(R)} = 1/2(1 \mp \gamma_5)$. Since only the operator O_{V_L} appears in the SM, we have explicitly factored it out in Eq. (42), ensuring that the NP Wilson coefficients vanish in the SM. We assume that all C_i are real-valued and that neutrinos are exclusively left-handed.

Note that below the EW scale, the expansion parameter is $1/v$ formally. However, from the point of view of the SMEFT [66], NP effects are associated to an additional expansion parameter Λ, the scale of new physics above EW scale. At tree-level, the matching between the SMEFT operators and the operators in the effective Hamiltonian (42) are given in Ref. [67]. From the SMEFT prospective, the Wilson coefficients in (42) are not of order one but additionally suppressed by the small ratio $(v/\Lambda)^2$.

The NP contributions to the differential rate of $B \to X_c l \bar{\nu}_l$ have been presented in Ref. [12]. The

NP effects for the moments defined in (4) can be written schematically in the following way

$$
\begin{aligned}
\langle O \rangle = {} & \xi_{\text{SM}} + |C_{V_R}|^2\, \xi_{\text{NP}}^{\langle V_R, V_R \rangle} + |C_{S_L}|^2\, \xi_{\text{NP}}^{\langle S_L, S_L \rangle} + |C_{S_R}|^2\, \xi_{\text{NP}}^{\langle S_R, S_R \rangle} + |C_T|^2\, \xi_{\text{NP}}^{\langle T, T \rangle} \\
& + \text{Re}((C_{V_L} - 1)C_{V_R}^*)\, \xi_{\text{NP}}^{\langle V_L, V_R \rangle} + \text{Re}(C_{S_L} C_{S_R}^*)\, \xi_{\text{NP}}^{\langle S_L, S_R \rangle} + \text{Re}(C_{S_L} C_T^*)\, \xi_{\text{NP}}^{\langle S_L, T \rangle} \\
& + \text{Re}(C_{S_R} C_T^*)\, \xi_{\text{NP}}^{\langle S_R, T \rangle} .
\end{aligned}
\tag{44}
$$

The terms denoted by ξ are the various interference terms between different operators. They depend on m_b, m_c, the HQE parameters and E_{cut} or q^2_{cut}. Ref. [12] provides results for the power corrections at tree level up to $O(1/m_b^3)$ and NLO perturbative corrections to leading order in the power expansion $(1/m_b^0)$. In (44), Wilson coefficients C_i are considered much smaller than one. We therefore expand up to quadratic terms when preparing the predictions for the centralized moments. We note that the term $C_{V_L}\xi_{\text{NP}}^{\langle V_L \rangle}$ do not appear because of the normalization. It affects only the branching ratio, which however only results in a rescaling of V_{cb}.

The first term ξ_{SM} in (44) corresponds to the SM prediction, whose implementation has been described in the previous sections. The additional NP contributions generated by the effective operators have been implemented in `kolya` including power corrections up to $1/m_b^3$ and NLO perturbative QCD corrections at partonic level. The implementation closely follows the methods described for the SM case. Namely, we implement the tree-level contributions in an exact form, while for the QCD corrections, we generate interpolation grids for their fast evaluation. The NP contributions to the moments are implemented in `Q2moments_NP.py`, `Elmoments_NP.py` and `MXmoments_NP.py`. The NP extension of the total rate is in `TotalRate_NP.py`.

5 Usage of the library

5.1 Installation

The software `kolya` requires Python version 3.6 or above and runs on Linux and Mac. The code is released under the GNU GPL v3 license. To download the package, clone the master branch of the Git repository via

```
$: git clone https://gitlab.com/vcb-inclusive/kolya.git
```

Afterwards, change the directory to the `kolya` directory and install it with `pip3` in the following way:

```
$: cd kolya
$: pip3 install .
```

The dependencies will be automatically downloaded and installed during the setup. To get started, just import the package into a Python shell or a Jupyter notebook:

```
>>> import kolya
```

Note that the first time `kolya` is loaded, several functions are translated from Python to optimized machine code by Numba and cached. This stage may take from several seconds up to a few minutes.

5.2 Parameter classes

The library contains classes to store various real-valued variables. One class is dedicated to the physical parameters like heavy quark masses and the strong coupling constant, one for the HQE parameters, and one for the Wilson coefficients in the NP extension. Dimensionful quantities, like the quark masses, are given in units of GeV. The values of the physical parameters are stored in an object of `parameters.physical_parameters` class

```
>>> par = kolya.parameters.physical_parameters()
```

The new object `par` contains information about M_B, $m_b^{\mathrm{kin}}(\mu_{\mathrm{kin}})$, $\overline{m}_c(\mu_c)$ and $\alpha_s^{(4)}(\mu_s)$. The bottom quark mass in the kinetic scheme and its scale are given in the variables `mbkin` and `scale_mbkin`, while the charm mass in the $\overline{\mathrm{MS}}$ at renormalization scale μ_c corresponds to the variables `mcMS` and `scale_mcMS`. The strong coupling constant $\alpha_s^{(4)}$ and its renormalization scale scale μ_s correspond to the class variables `alphas` and `scale_alphas`. The class initializes also the renormalization scales of the HQE parameters μ_G^2, ρ_D^3 and ρ_{LS}^3 through the variables `scale_muG`, `scale_rhoD` and `scale_rhoLS`. At present, these are set equal to $m_b^{\mathrm{kin}}(\mu_{\mathrm{kin}})$.

The values stored in the object `par` can be shown with the `show` method:

```
>>> par.show()
bottom mass:       mbkin( 1.0  GeV)      =   4.563  GeV
charm mass:        mcMS( 3.0  GeV)       =   0.989  GeV
coupling constant: alpha_s( 4.563  GeV) =   0.2182
```

where the current default values are based on the latest version of the FLAG 21 review [68] as of February 2024 [69]. Values different from the default ones can be set during initialization. For instance, we can initialize the $\overline{m}_c(2\,\mathrm{GeV}) = 1.094$ GeV as follows:

```
>>> par = kolya.parameters.physical_parameters(mcMS=1.094,scale_mcMS=2.0)
>>> par.show()
bottom mass:       mbkin( 1.0  GeV)      =   4.563  GeV
charm mass:        mcMS( 2.0  GeV)       =   1.094  GeV
coupling constant: alpha_s( 4.563  GeV) =   0.2182
```

The example above shows that during initialization, the values of `mcMS` and `scale_mcMS` must be set consistently. The following command

```
>>> par = kolya.parameters.physical_parameters(mcMS=1.094)
```

would initialize the charm mass to the (unphysical) value $\overline{m}_c(3\,\mathrm{GeV}) = 1.094$ GeV. In order to set the quark masses at scales different from the default ones in a consistent way, we include the method FLAG2024. For instance, we set the quark masses at a scale $\mu_{\mathrm{kin}} = \mu_c = 2$ GeV in the following way:

```
>>> par = kolya.parameters.physical_parameters()
>>> par.FLAG2024(scale_mcMS=2.0, scale_mbkin=2.0)
>>> par.show()
bottom mass:       mbkin( 2.0  GeV)      =   4.295730717092438  GeV
charm mass:        mcMS( 2.0  GeV)       =   1.0940623249384822  GeV
coupling constant: alpha_s( 4.563  GeV) =   0.21815198098622618
```

Internally, the bottom and quark masses are recalculated using `CRunDec` [70–72] using the default values from Ref. [69]. The scale of the strong coupling constant can be modified in a similar way:

```
>>> par = kolya.parameters.physical_parameters()
>>> par.FLAG2024(scale_alphas=3.0)
>>> par.show()
bottom mass:       mbkin( 1.0  GeV)  =  4.56266484311551  GeV
charm mass:        mcMS( 3.0  GeV)   =  0.989  GeV
coupling constant: alpha_s( 3.0  GeV) = 0.2531150801276913
```

Also in this case, we internally use `CRunDec` to evaluate $\alpha_s^{(4)}(3\,\text{GeV})$.

The values of the HQE parameters in the historical basis (sometimes referred to as the "perp" basis in literature) are stored into an object of the class `parameters.HQE_parameters`. By default, their values are set to zero unless explicitly initialized:

```
>>> hqe = kolya.parameters.HQE_parameters(
          muG = 0.306,
          rhoD = 0.185,
          rhoLS = -0.13,
          mupi = 0.477)
>>> hqe.show()
mupi  =   0.477  GeV^2
muG   =   0.306  GeV^2
rhoD  =   0.185  GeV^3
rhoLS =  -0.13  GeV^3
```

where the values up to $1/m_b^3$ can be visualized all at once with `show()`. Since there are several operators at order $1/m_b^4$ and $1/m_b^5$, we do not print them by default, however, their values can be inspected with the additional option `show(flagmb4=1)` and `show(flagmb5=1)`. We introduce the class `parameters.HQE_parameters_RPI` for the HQE parameters in the RPI basis:

```
>>> hqe = kolya.parameters.HQE_parameters_RPI(
          muG = 0.306,
          rhoD = 0.185,
          mupi = 0.477)
```

The parameter classes `LSSA_HQE_parameters` and `LSSA_HQE_parameters_RPI` contain numerical values for the HQE parameters in the historical and RPI basis, respectively, obtained using the "lowest-lying state saturation ansatz" (LLSA). The LLSA approximates the higher-order HQE parameters by expressing them through the bottom quark mass m_b, the HQE parameters $(\mu_\pi^2)^\perp$ and $(\mu_G^2)^\perp$, and the excitation energies $\epsilon_{1/2}$ and $\epsilon_{3/2}$. For further details on the LLSA, we refer to Ref. [73], and for the expressions of the higher-order HQE parameters and their LLSA values, we refer to Refs. [28, 32].

The Wilson coefficients of the effective dimension-six operators defined in (42) are initialized via the class `parameters.WCoefficients` and their values are inspected with the method `show()`

```
>>> wc = kolya.parameters.WCoefficients(SL=0.1,SR=-0.05)
>>> wc.show()
C_{V_L} =  0
C_{V_R} =  0
```

```
C_{S_L} =  0.1
C_{S_R} =  -0.05
C_{T}   =  0
```

The Wilson coefficients $C_{V_L}, C_{V_R}, C_{S_L}, C_{S_R}$ and C_T are denoted by VL, VR, SL, SR and T respectively. By default, they are initialized to zero.

5.3 Moment predictions

We implemented the first four centralized moments of the q^2 spectrum and the first three moments of E_l and M_X^2. To evaluate them, we first need to initialize three objects for the physical parameters, the HQE parameters, and the Wilson coefficients:

```
>>> par = kolya.parameters.physical_parameters()
>>> hqe = kolya.parameters.HQE_parameters(
            muG = 0.306,
            rhoD = 0.185,
            rhoLS = -0.13,
            mupi = 0.477)
>>> wc = kolya.parameters.WCoefficients()
```

The prediction for the q^2 moments receive as inputs the value of q_{cut}^2 expressed in GeV^2, and the three objects par, hqe and wc. The functions Q2moments.moment_n_KIN_MS(q2cut, par, hqe, wc), where n= 1, 2, 3, 4, return the value of each q^2 moment. For example, to evaluate $q_1(q_{cut}^2 = 8.0\,\text{GeV}^2)$ type

```
>>> q2cut = 8.0    #GeV^2
>>> kolya.Q2moments.moment_1_KIN_MS(q2cut, par, hqe, wc)
8.996406491856465 #GeV^2
```

The result is provided in the respective powers of GeV^{2n}. The suffix KIN and MS refers to the scheme for bottom (kinetic) and charm ($\overline{\text{MS}}$) masses. By default, the evaluation considers power corrections up to $1/m_b^3$. The corrections of order $1/m_b^4$ and $1/m_b^5$ can be included by setting the optional arguments flagmb4=1 and flagmb5=1. For instance, we can set the higher-order HQE parameters $m_1 = 0.1\,\text{GeV}^4$ and $r_1 = 0.1\,\text{GeV}^5$ and compare the predictions up to order $1/m_b^5$ in the following way:

```
>>> hqe.m1=0.1 #GeV^4
>>> hqe.r1=0.1 #GeV^5
>>> kolya.Q2moments.moment_1_KIN_MS(q2cut, par, hqe, wc)
8.996406491856465
>>> kolya.Q2moments.moment_1_KIN_MS(q2cut, par, hqe, wc, flagmb4=1)
8.966491804532277
>>> kolya.Q2moments.moment_1_KIN_MS(q2cut, par, hqe, wc, flagmb4=1, flagmb5=1)
8.75600814954277
```

Setting flagmb4=0 and flagmb5=0 eliminates all terms of order $1/m_b^4$ and $1/m_b^5$, respectively. We note that at these orders also mixing terms proportional to $\mu_G^2\mu_\pi^2$ or $\mu_G^2\rho_D^3$ enter which can only be excluded by putting these two flags to zero. Therefore, putting these flags to zero does not have the same effect as simply setting all the $1/m_b^4$ and $1/m_b^5$ HQE element to zero.

Concerning the perturbative corrections, these are all included by default in the numerical evaluation (see Tab. 1 for the current orders in α_s implemented). For cross-checks with the literature and the study of their impact, the NNLO corrections can be switched off via the optional argument `flag_includeNNLO=0` (default `flag_includeNNLO=1`). NLO corrections to the power-suppressed terms can be excluded with `flag_includeNLOpw=0`.

Moreover, the option `flag_DEBUG=1` will print a report of the various contributions coming from the higher-order QCD corrections:

```
>>> kolya.Q2moments.moment_1_KIN_MS(8.0, par, hqe, wc, flag_DEBUG=1)
Q2moment n. 1 LO =   9.148659808170105
Q2moment n. 1 NLO = api * -1.319532010835962
Q2moment n. 1 NNLO = api^2 * -9.616956902561078
Q2moment n. 1 NLO pw = api * -0.7873907726673756
Q2moment n. 1 NNLO from NLO pw = api^2 * 8.39048437244325
```

The contributions denoted by `NLO` and `NNLO` are the coefficients in front of $\alpha_s(\mu_s)/\pi$ and $(\alpha_s(\mu_s)/\pi)^2$ to leading order in $1/m_b$. The term `NLO pw` corresponds to the overall NLO correction in the terms of order $1/m_b^2$ and $1/m_b^3$. In the kinetic scheme, the inclusion of the NLO corrections to the power-suppressed terms induces also an additional $O(\alpha_s^2)$ contribution to leading order in $1/m_b$, which is reported in the last line.

The predictions for the E_l and M_X^2 moments follow a similar syntax. The first argument passed to the function corresponds to the value of the cut E_{cut} in units of GeV. For instance, the first moments of E_l and M_X^2 for $E_{\text{cut}} = 1.0$ GeV are evaluated as follows:

```
>>> par = kolya.parameters.physical_parameters()
>>> hqe = kolya.parameters.HQE_parameters(
            muG = 0.306,
            rhoD = 0.185,
            rhoLS = -0.13,
            mupi = 0.477)
>>> wc = kolya.parameters.WCoefficients()
>>> elcut=1.0 #GeV
>>> kolya.Elmoments.moment_1_KIN_MS(elcut, par, hqe, wc)
1.5498126092840554 #GeV
>>> kolya.MXmoments.moment_1_KIN_MS(elcut, par, hqe, wc)
4.348386230761794 #GeV^2
```

Higher moments ℓ_2, ℓ_3, h_2, h_3 are computed in a similar way by replacing `moment_1` with `moment_2` or `moment_3`. The result for $\ell_n(E_{\text{cut}})$ is in GeV^n, while for $h_n(E_{\text{cut}})$ the result is in GeV^{2n}.

By default, the moments are calculated using the HQE parameters as defined in the historical basis. The q^2 moments and the total rate can also be calculated using the RPI basis adopted in Ref. [9]. The predictions in the RPI basis are obtained by passing the optional argument `flag_basisPERP=0`. In this case, the HQE parameters must be passed to the function through an object of the class `HQE_parameters_RPI`:

```
>>> par = kolya.parameters.physical_parameters()
>>> hqeRPI = kolya.parameters.HQE_parameters_RPI(
            muG = 0.38,
```

```
          rhoD = 0.03,
          mupi = 0.43)
>>> wc = kolya.parameters.WCoefficients()
>>> kolya.Q2moments.moment_1_KIN_MS(8.0, par, hqeRPI, wc, flag_basisPERP=0)
9.350141389366614
```

The RPI basis is supported only for q_n moments and the total rate since reparametrization invariance reduces the number of HQE parameters only for them. Similar to before, the $1/m_b^4$ and $1/m_b^5$ corrections in the RPI basis can be included by using the optional arguments `flagmb4=1` and `flagmb5=1`. For the ℓ_n and h_n, we stick to the historical basis.

5.4 Branching ratio prediction

To obtain the branching ratio $\mathrm{Br}(B \to X_c l \bar{\nu}_l)$ or the total semileptonic width Γ_{sl}, three objects for the physical parameters, the HQE parameters, and the Wilson coefficients must be initialized, as discussed for the moments in the previous subsection. For the total rate Γ_{sl} defined in (2), type

```
>>> Vcb = 42.2e-3
>>> kolya.TotalRate.TotalRate_KIN_MS(Vcb, par, hqe, wc)
4.4016320941077224e-14 #GeV
```

where the first argument is the value of $|V_{cb}|$ and the result is expressed in GeV. For the evaluation of the branching ratio, we use

```
>>> Vcb = 42.2e-3
>>> kolya.TotalRate.BranchingRatio_KIN_MS(Vcb, par, hqe, wc)
0.10555834162102022
```

We obtain the branching ratio by dividing Γ_{sl} by the average lifetime of the $B^\pm$ and B_0 mesons.

The partial width $\Delta\Gamma_{\mathrm{sl}}(E_{\mathrm{cut}})$ with cut on E_l is obtained with

```
>>> Vcb = 42.2e-3
>>> elcut = 1.0 # GeV
>>> kolya.DeltaBR.DeltaRate_KIN_MS(Vcb, elcut, par, hqe, wc)
3.47993610329532e-14 #GeV
```

where the first argument is the value of $|V_{cb}|$, the second argument the value of E_{cut} in GeV. The result is reported in GeV. The corresponding value for the branching ratio is given by

```
>>> Vcb = 42.2e-3
>>> elcut = 1.0 # GeV
>>> kolya.DeltaBR.DeltaBR_KIN_MS(Vcb, elcut, par, hqe, wc)
0.08345456325227786
```

The function for the branching ratio allows the optional arguments `flagmb4=1` and `flagmb5=1` to include the power corrections of order $1/m_b^4$ and $1/m_b^5$. The predictions in the RPI basis are obtained by passing the optional argument `flag_basisPERP=0`. For cross-checks with the literature and the study of the impact of QCD corrections, the NNLO and N3LO corrections to the total rate can be switched off via the optional arguments `flag_includeNNLO=0` and `flag_includeN3LO=0` (by default, all these corrections are included). Moreover, the effects arising from the NLO corrections to the power-suppressed terms can be excluded with `flag_includeNLOpw=0`.

6 Outlook & Conclusion

In this document, we have presented the first version of the open-source library `kolya`, corresponding to the release 1.0. In this release, we have implemented the predictions in the HQE for the total rate and the moments of q^2, E_l and M_X^2. Currently, this is sufficient for comparison with published experimental results by B factories. We included all higher order corrections in α_s and $1/m_b$ which are available at this specific point in time and are summarized in Tab. 1.

On the GitLab repository, we provide, additionally, interactive tutorials running as a Jupyter notebook and validation notebooks which demonstrate how the library can reproduce the results available in the literature. The library is open source, so code contributions and improvements are very welcome. In particular, new higher-order corrections can be implemented like

- QED effects calculated in Ref. [74],

- exact results for the NNLO corrections to E_l and M_X^2 moments with a lower cut E_{cut},

- renormalization group evolution of the HQE parameters to NLO,

- the NLO corrections in the coefficients of ρ_D^3 and ρ_{LS}^3 for the E_l and M_X^2 moments.

Additional observables can play an important role in better improving the extraction of the HQE parameters or have an important role in testing the SM. These new observables may include

- forward-backward asymmetries A_{FB} and q^2, E_l and M_X^2 moments for forward and backward events [75, 76],

- the ratio $R_X = \Gamma_{B \to X_c \tau \bar{\nu}_\tau}/\Gamma_{B \to X_c l \bar{\nu}_l}$,

- the lifetime of B mesons within the HQE,

- predictions for the decay into charmless final states $B \to X_u l \bar{\nu}_l$.

Finally, `kolya` could also be extended to include predictions for inclusive D decays discussed in detail in [77].

Acknowledgments

We thank F. Bernlochner and M. Prim for ongoing collaboration and suggestions and P. Gambino for providing the results of Refs. [41, 42, 46] in electronic form. We also thank D. Straub for discussion about the Python package `python-rundec` [78], which provides a wrapper around `CRunDec` [70–72]. The work of M.F. is supported by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 101065445 - PHOBIDE. The work of I.S.M. was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under grant 396021762 – TRR 257 "Particle Physics Phenomenology after the Higgs Discovery". K.K.V. acknowledges support from the project "Beauty decays: the quest for extreme precision" of the Open Competition Domain Science which is financed by the Dutch Research Council (NWO).

A   Definition of the HQE elements

Here, we define the HQE parameters both in the historical and the RPI basis up to $1/m_b^5$. The conversions between these two bases can be found in Ref. [28].

A.1   Historical basis

The HQE matrix elements in the historical basis, denoted by "$\perp$", are defined through the spacial covariant derivatives $iD_\mu^\perp = g_{\mu\nu}^\perp iD^\nu$, where

$$g_{\mu\nu}^\perp = g_{\mu\nu} - v_\mu v_\nu \,, \tag{45}$$

as in Refs. [27, 79]. We will employ here the notation $\langle \bar{b}_v \dots b_v \rangle \equiv \langle B(v)|\bar{b}_v \dots \bar{b}_v|B(v)\rangle$. At $1/m_b^2$, we have

$$2m_B(\mu_\pi^2)^\perp = -\langle \bar{b}_v (iD^\rho)(iD^\sigma) b_v \rangle \, g_{\rho\sigma}^\perp \,,$$

$$2m_B(\mu_G^2)^\perp = \frac{1}{2}\langle \bar{b}_v \left[(iD^\rho), (iD^\sigma) \right](-i\sigma^{\alpha\beta}) b_v \rangle \, g_{\rho\alpha}^\perp g_{\sigma\beta}^\perp \,, \tag{46}$$

where $\gamma^\mu \gamma^\nu = g^{\mu\nu} + (-i\sigma^{\mu\nu})$. At $1/m_b^3$, we have

$$2m_B(\rho_D^3)^\perp = \frac{1}{2}\langle \bar{b}_v \left[(iD^\rho), \left[(iD^\sigma), (iD^\lambda) \right] \right] b_v \rangle \, g_{\rho\lambda}^\perp v_\sigma \,,$$

$$2m_B(\rho_{LS}^3)^\perp = \frac{1}{2}\langle \bar{b}_v \left\{ (iD^\rho), \left[(iD^\sigma), (iD^\lambda) \right] \right\}(-i\sigma^{\alpha\beta}) b_v \rangle \, g_{\rho\alpha}^\perp g_{\lambda\beta}^\perp v_\sigma \,. \tag{47}$$

The nine HQE parameters at $1/m_b^4$ were first introduced in Ref. [27]. We list them here:

$$2m_B m_1 = \langle \bar{b}_v (iD^\rho)(iD^\sigma)(iD^\lambda)(iD^\delta) b_v \rangle \frac{1}{3} \left(g_{\rho\sigma}^\perp g_{\lambda\delta}^\perp + g_{\rho\lambda}^\perp g_{\sigma\delta}^\perp + g_{\rho\delta}^\perp g_{\sigma\lambda}^\perp \right) \,,$$

$$2m_B m_2 = \langle \bar{b}_v \left[(iD^\rho), (iD^\sigma) \right]\left[(iD^\lambda), (iD^\delta) \right] b_v \rangle \, g_{\rho\delta}^\perp v_\sigma v_\lambda \,,$$

$$2m_B m_3 = \langle \bar{b}_v \left[(iD^\rho), (iD^\sigma) \right]\left[(iD^\lambda), (iD^\delta) \right] b_v \rangle \, g_{\rho\lambda}^\perp g_{\sigma\delta}^\perp \,,$$

$$2m_B m_4 = \langle \bar{b}_v \left\{ (iD^\rho), \left[(iD^\sigma), \left[(iD^\lambda), (iD^\delta) \right] \right] \right\} b_v \rangle \, g_{\sigma\lambda}^\perp g_{\rho\delta}^\perp \,,$$

$$2m_B m_5 = \langle \bar{b}_v \left[(iD^\rho), (iD^\sigma) \right]\left[(iD^\lambda), (iD^\delta) \right](-i\sigma^{\alpha\beta}) b_v \rangle \, g_{\alpha\rho}^\perp g_{\beta\delta}^\perp v_\sigma v_\lambda \,,$$

$$2m_B m_6 = \langle \bar{b}_v \left[(iD^\rho), (iD^\sigma) \right]\left[(iD^\lambda), (iD^\delta) \right](-i\sigma^{\alpha\beta}) b_v \rangle \, g_{\alpha\sigma}^\perp g_{\beta\lambda}^\perp g_{\rho\delta}^\perp \,,$$

$$2m_B m_7 = \langle \bar{b}_v \left\{ \{(iD^\rho), (iD^\sigma)\}, \left[(iD^\lambda), (iD^\delta) \right] \right\}(-i\sigma^{\alpha\beta}) b_v \rangle \, g_{\sigma\lambda}^\perp g_{\alpha\rho}^\perp g_{\beta\delta}^\perp \,,$$

$$2m_B m_8 = \langle \bar{b}_v \left\{ \{(iD^\rho), (iD^\sigma)\}, \left[(iD^\lambda), (iD^\delta) \right] \right\}(-i\sigma^{\alpha\beta}) b_v \rangle \, g_{\rho\sigma}^\perp g_{\alpha\lambda}^\perp g_{\beta\delta}^\perp \,,$$

$$2m_B m_9 = \langle \bar{b}_v \left[(iD^\rho), \left[(iD^\sigma), \left[(iD^\lambda), (iD^\delta) \right] \right] \right](-i\sigma^{\alpha\beta}) b_v \rangle \, g_{\rho\beta}^\perp g_{\lambda\alpha}^\perp g_{\sigma\delta}^\perp \,. \tag{48}$$

Finally, eighteen more parameters are present at $1/m_b^5$, as defined in Ref. [27]:

$$2m_B r_1 = \langle \bar{b}_v (iD_\mu)(ivD)^3 (iD^\mu) b_v \rangle \,,$$

$$2m_B r_2 = \langle \bar{b}_v (iD_\mu)(ivD)(iD^\mu)(iD)^2 b_v \rangle \,,$$

$$2m_B r_3 = \langle \bar{b}_v (iD_\mu)(ivD)(iD_\nu)(iD^\mu)(iD^\nu) b_v \rangle \,,$$

$$2m_B r_4 = \langle \bar{b}_v (iD_\mu)(ivD)(iD)^2 (iD^\mu) b_v \rangle \,,$$

$$2m_B r_5 = \langle \bar{b}_v (iD)^2 (ivD)(iD)^2 b_v \rangle \,,$$

$$2m_B r_6 = \langle \bar{b}_v (iD_\mu)(iD_\nu)(ivD)(iD^\nu)(iD^\mu) b_v \rangle \,,$$

$$2m_B r_7 = \langle \bar{b}_v (iD_\mu)(iD_\nu)(ivD)(iD^\mu)(iD^\nu) b_v \rangle \,,$$

$$2m_B r_8 = \langle \bar{b}_v (iD_\alpha)(ivD)^3 (iD_\beta)(-i\sigma^{\alpha\beta}) b_v \rangle \,,$$

$$2m_B r_9 = \langle \bar{b}_v (iD_\alpha)(ivD)(iD_\beta)(iD)^2 (-i\sigma^{\alpha\beta}) b_v \rangle \,,$$

$$2m_B r_{10} = \langle \bar{b}_v (iD_\mu)(ivD)(iD^\mu)(iD_\alpha)(iD_\beta)(-i\sigma^{\alpha\beta}) b_v \rangle \,,$$

$$2m_B r_{11} = \langle \bar{b}_v (iD_\mu)(ivD)(iD_\alpha)(iD^\mu)(iD_\beta)(-i\sigma^{\alpha\beta}) b_v \rangle \,,$$

$$2m_B r_{12} = \langle \bar{b}_v (iD_\alpha)(ivD)(iD_\mu)(iD_\beta)(iD^\mu)(-i\sigma^{\alpha\beta}) b_v \rangle \,,$$

$$2m_B r_{13} = \langle \bar{b}_v (iD_\mu)(ivD)(iD_\alpha)(iD_\beta)(iD^\mu)(-i\sigma^{\alpha\beta}) b_v \rangle \,,$$

$$2m_B r_{14} = \langle \bar{b}_v (iD_\alpha)(ivD)(iD)^2 (iD_\beta)(-i\sigma^{\alpha\beta}) b_v \rangle \,,$$

$$2m_B r_{15} = \langle \bar{b}_v (iD_\alpha)(iD_\beta)(ivD)(iD)^2 (-i\sigma^{\alpha\beta}) b_v \rangle \,,$$

$$2m_B r_{16} = \langle \bar{b}_v (iD_\mu)(iD_\alpha)(ivD)(iD_\beta)(iD^\mu)(-i\sigma^{\alpha\beta}) b_v \rangle \,,$$

$$2m_B r_{17} = \langle \bar{b}_v (iD_\alpha)(iD_\mu)(ivD)(iD^\mu)(iD_\beta)(-i\sigma^{\alpha\beta}) b_v \rangle \,,$$

$$2m_B r_{18} = \langle \bar{b}_v (iD_\mu)(iD_\alpha)(ivD)(iD^\mu)(iD_\beta)(-i\sigma^{\alpha\beta}) b_v \rangle \,. \tag{49}$$

A.2 RPI basis

The RPI HQE matrix elements up to $1/m_b^4$ have been determined in Ref. [31]. We list them here:

$$2m_B\mu_\pi^2 = -\langle \bar{b}_v\, (iD)^2\, b_v\rangle\,,$$

$$2m_B\mu_G^2 = \langle \bar{b}_v\, (iD_\alpha)(iD_\beta)(-i\sigma^{\alpha\beta})\,b_v\rangle\,,$$

$$2m_B\tilde{\rho}_D^3 = \frac{1}{2}\langle \bar{b}_v\left[(iD_\mu),\left[\left((ivD)+\frac{1}{2m_b}(iD)^2\right),(iD^\mu)\right]\right]b_v\rangle\,,$$

$$2m_B r_G^4 = \langle \bar{b}_v\left[(iD_\mu),(iD_\nu)\right]\left[(iD^\mu),(iD^\nu)\right]b_v\rangle\,,$$

$$2m_B\tilde{r}_E^4 = \langle \bar{b}_v\left[\left((ivD)+\frac{1}{2m_b}(iD)^2\right),(iD_\mu)\right]\left[\left((ivD)+\frac{1}{2m_b}(iD)^2\right),(iD^\mu)\right]b_v\rangle\,,$$

$$2m_B s_B^4 = \langle \bar{b}_v\left[(iD_\mu),(iD_\alpha)\right]\left[(iD^\mu),(iD_\beta)\right](-i\sigma^{\alpha\beta})\,b_v\rangle\,,$$

$$2m_B\tilde{s}_E^4 = \langle \bar{b}_v\left[\left((ivD)+\frac{1}{2m_b}(iD)^2\right),(iD_\alpha)\right]\left[\left((ivD)+\frac{1}{2m_b}(iD)^2\right),(iD_\beta)\right](-i\sigma^{\alpha\beta})\,b_v\rangle,$$

$$2m_B s_{qB}^4 = \langle \bar{b}_v\left[(iD_\mu),\left[(iD^\mu),\left[(iD_\alpha),(iD_\beta)\right]\right]\right](-i\sigma^{\alpha\beta})\,b_v\rangle\,, \tag{50}$$

We note that we choose to use μ_π^2 instead of μ_3 as in Ref. [31], in order to avoid factors of $1/\mu_3$ in the centralized moments. Furthermore, we note that $\tilde{\rho}_D^3$, $\tilde{r}_E^4$, and $\tilde{s}_E^4$ contain their so-called RPI-completion terms as described in Refs. [28, 31]. In Ref. [28], the RPI matrix elements at $1/m_b^5$ have been determined to be:

$$2m_B X_1^5 = \langle \bar{b}_v\left[(ivD),\left[(ivD),(iD_\mu)\right]\right]\left[(ivD),(iD^\mu)\right]b_v\rangle\,,$$

$$2m_B X_2^5 = \langle \bar{b}_v\left[(ivD),\left[(iD_\mu),(iD_\nu)\right]\right]\left[(iD^\mu),(iD^\nu)\right]b_v\rangle\,,$$

$$2m_B X_3^5 = \langle \bar{b}_v\left[(iD_\mu),\left[(ivD),(iD_\nu)\right]\right]\left[(iD^\mu),(iD^\nu)\right]b_v\rangle$$

$$2m_B X_4^5 = \langle \bar{b}_v\left[(iD_\mu),\left[(iD_\nu),\left[(iD^\mu),\left[(ivD),(iD^\nu)\right]\right]\right]\right]b_v\rangle\,,$$

$$2m_B X_5^5 = \langle \bar{b}_v\left[(ivD),\left[(ivD),(iD_\alpha)\right]\right]\left[(ivD),(iD_\beta)\right](-i\sigma^{\alpha\beta})\,b_v\rangle\,,$$

$$2m_B X_6^5 = \langle \bar{b}_v\left[(ivD),\left[(iD_\mu),(iD_\alpha)\right]\right]\left[(iD^\mu),(iD_\beta)\right](-i\sigma^{\alpha\beta})\,b_v\rangle\,,$$

$$2m_B X_7^5 = \langle \bar{b}_v\left[(iD_\mu),\left[(ivD),(iD_\alpha)\right]\right]\left[(iD^\mu),(iD_\beta)\right](-i\sigma^{\alpha\beta})\,b_v\rangle\,,$$

$$2m_B X_8^5 = \langle \bar{b}_v\left[(iD_\mu),\left[(ivD),(iD_\alpha)\right]\right]\left[(iD^\mu),(iD_\beta)\right](-i\sigma^{\alpha\beta})\,b_v\rangle\,,$$

$$2m_B X_9^5 = \langle \bar{b}_v\left[(iD_\mu),\left[(ivD),(iD^\mu)\right]\right]\left[(iD_\alpha),(iD_\beta)\right](-i\sigma^{\alpha\beta})\,b_v\rangle\,,$$

$$2m_B X_{10}^5 = \langle \bar{b}_v\left[(iD_\mu),\left[(ivD),(iD^\mu)\right]\right]\left[(iD_\alpha),(iD_\beta)\right](-i\sigma^{\alpha\beta})\,b_v\rangle\,. \tag{51}$$

References

[1] A. V. Manohar and M. B. Wise, *Inclusive semileptonic B and polarized Λ_b decays from QCD*, Phys. Rev. D **49**, 1310 (1994), doi:10.1103/PhysRevD.49.1310, hep-ph/9308246.

[2] B. Blok, L. Koyrakh, M. A. Shifman and A. I. Vainshtein, *Differential distributions in semileptonic decays of the heavy flavors in QCD*, Phys. Rev. D **49**, 3356 (1994), doi:10.1103/PhysRevD.50.3572, [Erratum: Phys.Rev.D 50, 3572 (1994)], hep-ph/9307247.

[3] I. I. Y. Bigi, M. A. Shifman, N. G. Uraltsev and A. I. Vainshtein, *QCD predictions for lepton spectra in inclusive heavy flavor decays*, Phys. Rev. Lett. **71**, 496 (1993), doi:10.1103/PhysRevLett.71.496, hep-ph/9304225.

[4] Fael, Matteo and Milutin, Ilija and Vos, K. Keri, *Kolya 1.0*, https://doi.org/10.5281/zenodo.10818194 (2024).

[5] I. I. Y. Bigi, M. A. Shifman, N. Uraltsev and A. I. Vainshtein, *High power n of m(b) in beauty widths and n=5 —> infinity limit*, Phys. Rev. D **56**, 4017 (1997), doi:10.1103/PhysRevD.56.4017, hep-ph/9704245.

[6] A. Czarnecki, K. Melnikov and N. Uraltsev, *NonAbelian dipole radiation and the heavy quark expansion*, Phys. Rev. Lett. **80**, 3189 (1998), doi:10.1103/PhysRevLett.80.3189, hep-ph/9708372.

[7] M. Fael, K. Schönwald and M. Steinhauser, *Kinetic Heavy Quark Mass to Three Loops*, Phys. Rev. Lett. **125**(5), 052003 (2020), doi:10.1103/PhysRevLett.125.052003, 2005.06487.

[8] M. Fael, K. Schönwald and M. Steinhauser, *Relation between the $\overline{\text{MS}}$ and the kinetic mass of heavy quarks*, Phys. Rev. D **103**(1), 014005 (2021), doi:10.1103/PhysRevD.103.014005, 2011.11655.

[9] F. Bernlochner, M. Fael, K. Olschewsky, E. Persson, R. van Tonder, K. K. Vos and M. Welsch, *First extraction of inclusive V_{cb} from q^2 moments*, JHEP **10**, 068 (2022), doi:10.1007/JHEP10(2022)068, 2205.10274.

[10] R. van Tonder *et al.*, *Measurements of q^2 Moments of Inclusive $B \to X_c \ell^+ \nu_\ell$ Decays with Hadronic Tagging*, Phys. Rev. D **104**(11), 112011 (2021), doi:10.1103/PhysRevD.104.112011, 2109.01685.

[11] F. Abudinén *et al.*, *Measurement of lepton mass squared moments in $B \to Xc\ell\nu^-\ell$ decays with the Belle II experiment*, Phys. Rev. D **107**(7), 072002 (2023), doi:10.1103/PhysRevD.107.072002, 2205.06372.

[12] M. Fael, M. Rahimi and K. K. Vos, *New physics contributions to moments of inclusive $b \to c$ semileptonic decays*, JHEP **02**, 086 (2023), doi:10.1007/JHEP02(2023)086, 2208.04282.

[13] J. Aebischer, M. Fael, C. Greub and J. Virto, *B physics Beyond the Standard Model at One Loop: Complete Renormalization Group Evolution below the Electroweak Scale*, JHEP **09**, 158 (2017), doi:10.1007/JHEP09(2017)158, 1704.06639.

[14] E. E. Jenkins, A. V. Manohar and P. Stoffer, *Low-Energy Effective Field Theory below the Electroweak Scale: Operators and Matching*, JHEP **03**, 016 (2018), doi:10.1007/JHEP03(2018)016, [Erratum: JHEP 12, 043 (2023)], 1709.04486.

[15] P. Gambino and C. Schwanda, *Inclusive semileptonic fits, heavy quark masses, and V_{cb}*, Phys. Rev. D **89**(1), 014022 (2014), doi:10.1103/PhysRevD.89.014022, 1307.4551.

[16] J. Dingfelder and T. Mannel, *Leptonic and semileptonic decays of B mesons*, Rev. Mod. Phys. **88**(3), 035008 (2016), doi:10.1103/RevModPhys.88.035008.

[17] M. Fael, M. Prim and K. K. Vos, *Inclusive $B \to X_c \ell \bar{\nu}_\ell$ and $B \to X_u \ell \bar{\nu}_\ell$ decays: current status and future prospects*, Eur. Phys. J. ST **233**(2), 325 (2024), doi:10.1140/epjs/s11734-024-01090-w.

[18] D. M. Straub, *flavio: a Python package for flavour and precision phenomenology in the Standard Model and beyond* (2018), 1810.08132.

[19] D. van Dyk *et al.*, *EOS: a software for flavor physics phenomenology*, Eur. Phys. J. C **82**(6), 569 (2022), doi:10.1140/epjc/s10052-022-10177-4, 2111.15428.

[20] J. De Blas *et al.*, HEPfit: *a code for the combination of indirect and direct constraints on high energy physics models*, Eur. Phys. J. C **80**(5), 456 (2020), doi:10.1140/epjc/s10052-020-7904-z, 1910.14012.

[21] F. U. Bernlochner, S. Duell, Z. Ligeti, M. Papucci and D. J. Robinson, *Das ist der HAMMER: Consistent new physics interpretations of semileptonic decays*, Eur. Phys. J. C **80**(9), 883 (2020), doi:10.1140/epjc/s10052-020-8304-0, 2002.00020.

[22] A. Arbey and F. Mahmoudi, *SuperIso Relic v3.0: A program for calculating relic density and flavour physics observables: Extension to NMSSM*, Comput. Phys. Commun. **182**, 1582 (2011), doi:10.1016/j.cpc.2011.03.019.

[23] M. Fael, T. Mannel and K. Keri Vos, *V_{cb} determination from inclusive $b \to c$ decays: an alternative method*, JHEP **02**, 177 (2019), doi:10.1007/JHEP02(2019)177, 1812.07472.

[24] A. V. Manohar, *Reparametrization Invariance Constraints on Inclusive Decay Spectra and Masses*, Phys. Rev. D **82**, 014009 (2010), doi:10.1103/PhysRevD.82.014009, 1005.1952.

[25] A. Sirlin, *Current Algebra Formulation of Radiative Corrections in Gauge Theories and the Universality of the Weak Interactions*, Rev. Mod. Phys. **50**, 573 (1978), doi:10.1103/RevModPhys.50.573, [Erratum: Rev.Mod.Phys. 50, 905 (1978)].

[26] B. M. Dassinger, T. Mannel and S. Turczyk, *Inclusive semi-leptonic B decays to order $1/m_b^4$*, JHEP **03**, 087 (2007), doi:10.1088/1126-6708/2007/03/087, hep-ph/0611168.

[27] T. Mannel, S. Turczyk and N. Uraltsev, *Higher Order Power Corrections in Inclusive B Decays*, JHEP **11**, 109 (2010), doi:10.1007/JHEP11(2010)109, 1009.4622.

[28] T. Mannel, I. S. Milutin and K. K. Vos, *Inclusive semileptonic $b \to c\ell\bar{\nu}$ decays to order $1/m_b^5$*, JHEP **02**, 226 (2024), doi:10.1007/JHEP02(2024)226, 2311.12002.

[29] M. Bordone, B. Capdevila and P. Gambino, *Three loop calculations and inclusive Vcb*, Phys. Lett. B **822**, 136679 (2021), doi:10.1016/j.physletb.2021.136679, 2107.00604.

[30] G. Finauri and P. Gambino, *The q^2 moments in inclusive semileptonic B decays* (2023), 2310.20324.

[31] T. Mannel and K. K. Vos, *Reparametrization Invariance and Partial Re-Summations of the Heavy Quark Expansion*, JHEP **06**, 115 (2018), doi:10.1007/JHEP06(2018)115, 1802.09409.

[32] T. Mannel, I. S. Milutin, R. Verkade and K. K. Vos, *Quark-Hadron Duality Violations and Higher-Order $1/m_b$ Corrections in Inclusive Semileptonic B Decays* (2024), 2407.01473.

[33] S. Biswas and K. Melnikov, *Second order QCD corrections to inclusive semileptonic $b \to X_c \ell \bar{\nu}_\ell$ decays with massless and massive lepton*, JHEP **02**, 089 (2010), doi:10.1007/JHEP02(2010)089, 0911.4142.

[34] Y. Nir, *The Mass Ratio m_c/m_b in Semileptonic B Decays*, Phys. Lett. B **221**, 184 (1989), doi:10.1016/0370-2693(89)91495-0.

[35] A. Pak and A. Czarnecki, *Mass effects in muon and semileptonic $b \to c$ decays*, Phys. Rev. Lett. **100**, 241807 (2008), doi:10.1103/PhysRevLett.100.241807, 0803.0960.

[36] A. Pak and A. Czarnecki, *Heavy-to-heavy quark decays at NNLO*, Phys. Rev. D **78**, 114015 (2008), doi:10.1103/PhysRevD.78.114015, 0808.3509.

[37] M. Dowling, J. H. Piclum and A. Czarnecki, *Semileptonic decays in the limit of a heavy daughter quark*, Phys. Rev. D **78**, 074024 (2008), doi:10.1103/PhysRevD.78.074024, 0810.0543.

[38] M. Egner, M. Fael, K. Schönwald and M. Steinhauser, *Revisiting semileptonic B meson decays at next-to-next-to-leading order*, JHEP **09**, 112 (2023), doi:10.1007/JHEP09(2023)112, 2308.01346.

[39] M. Fael, K. Schönwald and M. Steinhauser, *Third order corrections to the semi-leptonic $b \to c$ and the muon decays* (2020), 2011.13654.

[40] T. Becher and M. Neubert, *Toward a NNLO calculation of the anti-B —> X(s) gamma decay rate with a cut on photon energy. II. Two-loop result for the jet function*, Phys. Lett. B **637**, 251 (2006), doi:10.1016/j.physletb.2006.04.046, hep-ph/0603140.

[41] A. Alberti, T. Ewerth, P. Gambino and S. Nandi, *Kinetic operator effects in $\bar{B} \to X_c l \nu$ at $O(\alpha_s)$*, Nucl. Phys. B **870**, 16 (2013), doi:10.1016/j.nuclphysb.2013.01.005, 1212.5082.

[42] A. Alberti, P. Gambino and S. Nandi, *Perturbative corrections to power suppressed effects in semileptonic B decays*, JHEP **01**, 147 (2014), doi:10.1007/JHEP01(2014)147, 1311.7381.

[43] T. Mannel, A. A. Pivovarov and D. Rosenthal, *Inclusive weak decays of heavy hadrons with power suppressed terms at NLO*, Phys. Rev. D **92**(5), 054025 (2015), doi:10.1103/PhysRevD.92.054025, 1506.08167.

[44] M. Gremm and A. Kapustin, *Order 1/m(b)**3 corrections to B –> X(c) lepton anti-neutrino decay and their implication for the measurement of Lambda-bar and lambda(1)*, Phys. Rev. D **55**, 6924 (1997), doi:10.1103/PhysRevD.55.6924, hep-ph/9603448.

[45] T. Mannel, D. Moreno and A. A. Pivovarov, *NLO QCD Corrections to Inclusive $b \to c \ell \bar{\nu}$ Decay Spectra up to $1/m_Q^3$* (2021), 2112.03875.

[46] V. Aquila, P. Gambino, G. Ridolfi and N. Uraltsev, *Perturbative corrections to semileptonic b decay distributions*, Nucl. Phys. B **719**, 77 (2005), doi:10.1016/j.nuclphysb.2005.04.031, hep-ph/0503083.

[47] M. Fael and F. Herren, *NNLO QCD corrections to the q^2 spectrum of inclusive semileptonic B-meson decays*, JHEP **05**, 287 (2024), doi:10.1007/JHEP05(2024)287, 2403.03976.

[48] M. Trott, *Improving extractions of $|V_{cb}|$ and m_b from the hadronic invariant mass moments of semileptonic inclusive B decay*, Phys. Rev. D **70**, 073003 (2004), doi:10.1103/PhysRevD.70.073003, hep-ph/0402120.

[49] Fael, Matteo and Herren, Florian and Schönwald, Kay, in preparation.

[50] T. Becher, H. Boos and E. Lunghi, *Kinetic corrections to $B \to X_c \ell \bar{\nu}$ at one loop*, JHEP **12**, 062 (2007), doi:10.1088/1126-6708/2007/12/062, 0708.0855.

[51] https://numba.pydata.org/.

[52] S. J. Brodsky, G. P. Lepage and P. B. Mackenzie, *On the Elimination of Scale Ambiguities in Perturbative Quantum Chromodynamics*, Phys. Rev. D **28**, 228 (1983), doi:10.1103/PhysRevD.28.228.

[53] M. Jezabek and J. H. Kuhn, *QCD Corrections to Semileptonic Decays of Heavy Quarks*, Nucl. Phys. B **314**, 1 (1989), doi:10.1016/0550-3213(89)90108-9.

[54] E. Remiddi and J. A. M. Vermaseren, *Harmonic polylogarithms*, Int. J. Mod. Phys. A **15**, 725 (2000), doi:10.1142/S0217751X00000367, hep-ph/9905237.

[55] A. B. Goncharov, *Multiple polylogarithms, cyclotomy and modular complexes*, Math. Res. Lett. **5**, 497 (1998), doi:10.4310/MRL.1998.v5.n4.a7, 1105.2076.

[56] A. B. Goncharov, *Multiple polylogarithms and mixed Tate motives* (2001), math/0103059.

[57] C. W. Bauer, A. Frink and R. Kreckel, *Introduction to the GiNaC framework for symbolic computation within the C++ programming language*, J. Symb. Comput. **33**, 1 (2002), doi:10.1006/jsco.2001.0494, cs/0004015.

[58] W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, *Numerical Recipes in FORTRAN: The Art of Scientific Computing* (1992).

[59] M. Fael, K. Schönwald and M. Steinhauser, *A first glance to the kinematic moments of $B \to X_c \ell \nu$ at third order* (2022), 2205.03410.

[60] A. Alberti, P. Gambino, K. J. Healey and S. Nandi, *The Inclusive Determination of $|V_{cb}|$*, Nucl. Part. Phys. Proc. **273-275**, 1325 (2016), doi:10.1016/j.nuclphysbps.2015.09.212.

[61] P. Gambino, *B semileptonic moments at NNLO*, JHEP **09**, 055 (2011), doi:10.1007/JHEP09(2011)055, 1107.3100.

[62] P. Gambino and J. F. Kamenik, *Lepton energy moments in semileptonic charm decays*, Nucl. Phys. B **840**, 424 (2010), doi:10.1016/j.nuclphysb.2010.07.019, 1004.0114.

[63] M. Beneke and V. M. Braun, *Heavy quark effective theory beyond perturbation theory: Renormalons, the pole mass and the residual mass term*, Nucl. Phys. B **426**, 301 (1994), doi:10.1016/0550-3213(94)90314-X, hep-ph/9402364.

[64] I. I. Y. Bigi, M. A. Shifman, N. G. Uraltsev and A. I. Vainshtein, *The Pole mass of the heavy quark. Perturbation theory and beyond*, Phys. Rev. D **50**, 2234 (1994), doi:10.1103/PhysRevD.50.2234, hep-ph/9402360.

[65] I. I. Y. Bigi, M. A. Shifman, N. G. Uraltsev and A. I. Vainshtein, *Sum rules for heavy flavor transitions in the SV limit*, Phys. Rev. D **52**, 196 (1995), doi:10.1103/PhysRevD.52.196, hep-ph/9405410.

[66] B. Grzadkowski, M. Iskrzynski, M. Misiak and J. Rosiek, *Dimension-Six Terms in the Standard Model Lagrangian*, JHEP **10**, 085 (2010), doi:10.1007/JHEP10(2010)085, 1008.4884.

[67] J. Aebischer, A. Crivellin, M. Fael and C. Greub, *Matching of gauge invariant dimension-six operators for $b \to s$ and $b \to c$ transitions*, JHEP **05**, 037 (2016), doi:10.1007/JHEP05(2016)037, 1512.02830.

[68] Y. Aoki *et al.*, *FLAG Review 2021*, Eur. Phys. J. C **82**(10), 869 (2022), doi:10.1140/epjc/s10052-022-10536-1, 2111.09849.

[69] http://flag.unibe.ch/2021/Media?action=AttachFile&do=get&target=FLAG_webupdate_February2024.pdf.

[70] K. G. Chetyrkin, J. H. Kuhn and M. Steinhauser, *RunDec: A Mathematica package for running and decoupling of the strong coupling and quark masses*, Comput. Phys. Commun. **133**, 43 (2000), doi:10.1016/S0010-4655(00)00155-7, hep-ph/0004189.

[71] B. Schmidt and M. Steinhauser, *CRunDec: a C++ package for running and decoupling of the strong coupling and quark masses*, Comput. Phys. Commun. **183**, 1845 (2012), doi:10.1016/j.cpc.2012.03.023, 1201.6149.

[72] F. Herren and M. Steinhauser, *Version 3 of RunDec and CRunDec*, Comput. Phys. Commun. **224**, 333 (2018), doi:10.1016/j.cpc.2017.11.014, 1703.03751.

[73] J. Heinonen and T. Mannel, *Improved Estimates for the Parameters of the Heavy Quark Expansion*, Nucl. Phys. B **889**, 46 (2014), doi:10.1016/j.nuclphysb.2014.09.017, 1407.4384.

[74] D. Bigi, M. Bordone, P. Gambino, U. Haisch and A. Piccione, *QED effects in inclusive semi-leptonic B decays* (2023), 2309.02849.

[75] S. Turczyk, *Additional Information on Heavy Quark Parameters from Charged Lepton Forward-Backward Asymmetry*, JHEP **04**, 131 (2016), doi:10.1007/JHEP04(2016)131, 1602.02678.

[76] F. Herren, *The forward-backward asymmetry and differences of partial moments in inclusive semileptonic B decays* (2022), 2205.03427.

[77] M. Fael, T. Mannel and K. K. Vos, *The Heavy Quark Expansion for Inclusive Semileptonic Charm Decays Revisited*, JHEP **12**, 067 (2019), doi:10.1007/JHEP12(2019)067, 1910.05234.

[78] Straub, David, https://github.com/DavidMStraub/rundec-python.

[79] P. Gambino, K. J. Healey and S. Turczyk, *Taming the higher power corrections in semileptonic B decays*, Phys. Lett. B **763**, 60 (2016), doi:10.1016/j.physletb.2016.10.023, 1606.06174.