Precision theory predictions and interpretation of differential distributions in inclusive decays at Belle II

Dissertation zur Erlangung des Doktorgrades an der Fakultät für Mathematik, Informatik und Naturwissenschaften, Fachbereich Physik der Universität Hamburg

Ivan Novikov

2023-06-16

Hamburg
2023

Gutachter/innen der Dissertation:	Dr. Frank Tackmann Prof. Dr. Geraldine Servant
Zusammensetzung der Prüfungskommission:	Dr. Frank Tackmann Prof. Dr. Geraldine Servant Prof. Dr. Sven-Olaf Moch Prof. Dr. Alexander Glazov Prof. Dr. Dieter Horns
Vorsitzender der Prüfungskommission:	Prof. Dr. Sven-Olaf Moch
Datum der Disputation:	2023-06-26
Vorsitzender des Fach-Promotionsausschusses PHYSIK:	Prof. Dr. Günter H. W. Sigl
Leiter des Fachbereichs PHYSIK:	Prof. Dr. Wolfgang J. Parak
Dekan der Fakultät MIN:	Prof. Dr.-Ing. Norbert Ritter

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. |
I hereby declare upon oath that I have written the present dissertation independently and have not used further resources and aids than those stated in the dissertation.

Ort, den | City, date

Unterschrift | Signature

Acknowledgements

This work would have been impossible without the diligent guidance and advice of my supervisor Dr. Frank Tackmann and my colleague Dr. Bahman Dehnadi. They both found a way to combine parenthood and scientific research, and they provided excellent supervision in spite of the tragedies of the last three years. Their patient and thoughtful mentorship will not be forgotten.

I am also thankful to my co-supervisor Dr. Geraldine Servant for her help with the planning and organization of my doctoral studies and proceedings. Thanks should also go to my mentor Dr. Alexander Westphal for advising me on matters of good scientific practice.

My deepest thanks go to Dr. Alexander Glazov, as I could not have undertaken this journey without his unwavering support and encouragement.

I would like to express my appreciation of my colleagues in the DESY collider phenomenology group. I am particularly indebted to Dr. Pedro Cal and Dr. Aditya Pathak for their expertise in the area of effective quantum field theories and for the career advice that they shared with me. I also would like to extend my sincere thanks to the theory group leader Dr. Christophe Grojean for his understanding and support in all organizational matters of my work.

I would be remiss in not mentioning the Belle-II experimental group at DESY. I am particularly grateful to Henrikas Svidras and Tommy Martinov for the productive discussions of some aspects of experimental studies of $B \to X_s\gamma$ and $B \to X_u l\bar{\nu}$ decays, respectively.

Of course, nothing I ever achieved would be possible without the support of my parents, who spared no sacrifice to provide me with the best possible education, and without my grandparents, who inspired me to pursue a career in science.

Finally, I thank my colleague Florian Fabry for helping me translate the summary of this thesis into German.

Summary

This PhD thesis presents improved theoretical predictions for fully-differential decay rates of inclusive $B \to X_s\gamma$ and $B \to X_u l\bar{\nu}$ decays.

The improved predictions for the $B \to X_s\gamma$ photon energy spectrum include all the known singular contributions at N^{3}LO, with the exception of one yet-unknown 3-loop hard function coefficient. In the endpoint region of the spectrum all the singular contributions are resummed at the N^{3}LL accuracy. The remaining, nonsingular contributions are included at NNLO. To combine the singular contributions at N^{3}LO with the nonsingular contributions at NNLO, the yet-unknown 3-loop nonsingular contributions are parametrized in terms of nuisance parameters. The used parametrization avoids the artifacts that arise due to the mismatch of perturbative order of singular and nonsingular contributions. The missing 3-loop hard function coefficient is also treated as a nuisance parameter. The uncertainty related to these missing corrections is estimated and compared to the perturbative truncation uncertainty.

The impact of different definitions of the b-quark mass on the convergence of the perturbative series is studied. The 1S mass scheme, which has been used for $B \to X_s\gamma$ in the past, is demonstrated to break down starting at N^{3}LO. In contrast, the MSR mass scheme is shown to yield much more stable results. The convergence is further improved by an appropriate choice of short-distance definitions of the hadronic parameters λ_1 and ρ_1. In the context of these short-distance mass corrections, a novel kind of correction, which first appears at N^{3}LO, is identified. This correction is formally subleading in the power counting of Soft-Collinear Effective Theory (SCET), but is nevertheless singular in the peak region of the spectrum. Therefore, it cannot be naively accounted for within the previously-used approach. A prescription for the resummation of such terms is developed. Additionally, the impact of a different treatment of higher-order cross-terms in the factorization theorem is studied.

The improved $B \to X_u l\bar{\nu}$ predictions include all singular contributions at NNLO, and the nonsingular corrections are included at NLO. In the endpoint region the singular contributions are resummed at N^{3}LL accuracy. The resummed and fixed-order results are combined using a matching procedure that avoids artificial singularities and yields meaningful results in all regions of the phasespace, with the exception of the deep resonance region, where the inclusive approach is not applicable. The theoretical predictions are compared to the recently published measurements of differential $B \to X_u l\bar{\nu}$ spectra by the Belle collaboration. The agreement is adequate for the used inclusive model, and the differences are discussed.

Zusammenfassung

Diese Doktorarbeit präsentiert verbesserte theoretische Vorhersagen für komplett differentielle Zerfallsraten von inklusiven $B \to X_s\gamma$ und $B \to X_u l\bar{\nu}$ Zerfallskanälen.

Die verbesserten Vorhersagen für das $B \to X_s\gamma$ Photonenenergiespektrum berücksichtigen alle bekannten singulären Beiträge zu nächst-nächst-nächst-führender Ordnung ($\mathrm{N^3LO}$), mit Ausnahme von dem noch unbekannten Drei-Schleifen Koeffizienten der Funktion, die hochenergetische Wechselwirkung beschreibt (harte Funktion). Am Endpunkt des Spektrums sind alle singulären Beiträge bei nächst-nächst-nächst-führender logarithmischer Genauigkeit ($\mathrm{N^3LL}$) resummiert. Die übrigen nicht-singulären Beiträge sind bis zu NNLO berücksichtigt. Um die singulären Beiträge bei $\mathrm{N^3LO}$ mit den nicht-singulären Beiträgen bei NNLO zu kombinieren, sind die noch unbekannten nicht-singulären Drei-Schleifen Beiträge durch Störparameter parametrisiert. Die verwendete Parametrisierung vermeidet die Artefakten, die durch die nicht übereinstimmenden Ordnungen entstehen. Der fehlende Drei-Schleifen Koeffizient der harten Funktion wird auch als ein Störparameter behandelt. Die Unsicherheit, die mit diesen fehlenden Korrekturen verbunden ist, wird abgeschätzt und mit der Trunkierungsunsicherheit verglichen.

Der Einfluss von verschiedenen Definitionen der b-quark Masse auf die Konvergenz der Störungsreihe wird untersucht. Es wird gezeigt, dass das 1S Schema, die für den Zerfall $B \to X_s\gamma$ in der Vergangenheit gebraucht wurde, ab $\mathrm{N^3LO}$ versagt. Im Gegensatz dazu führt das MSR Schema zu sehr stabilen Ergebnissen. Die Konvergenz wird noch durch eine angemessene Auswahl von Definitionen der hadronischen Parameter λ_1 und ρ_1 verbessert. Im Zusammenhang mit diesen Massekorrekturen wird eine neue Art der Korrektur, die erst zu $\mathrm{N^3LO}$ erscheint, identifiziert. Diese Korrektur ist formell nicht-führend im Expansionsparameter der Weich-Kollinearen Effektiven Theorie (SCET). Allerdings ist sie singulär in den Endpunkten des Spektrums und kann deshalb nicht auf die vorher genutzte Weise berücksichtigt werden. Eine Vorschrift für die Resummierung solcher Korrekturen wird entwickelt. Außerdem wird der Einfluss von verschiedenen Konzepten zur Behandung von Kreuztermen höherer Ordnung im Faktorisierungstheorem studiert.

Die verbesserten $B \to X_u l\bar{\nu}$ Vorhersagen berücksichtigen alle singulären Beiträge bis zu NNLO, und die nicht-singuläre Beiträge werden bis zu NLO berechnet. In der Endpunkt Region werden die singulären Beiträge bis zu $\mathrm{N^3LL}$ resummiert. Die resummierten und nicht-resummierten Ergebnisse werden durch eine Vorschrift vereinigt, die künstliche Singularitäten vermeidet und zu sinnvollen Ergebnissen in allen Regionen des Phasenraums führt. Die einzige Ausnahme ist die Resonanz-Region, in dem ein inklusiver Ansatz nicht anwendbar ist. Die theoretischen Berechnungen werden mit den vor kurzen veröffentlichten Messungen der differentiellen $B \to X_u l\bar{\nu}$ Spektren der Belle Kollaboration verglichen. Die Übereinstimmung ist dem benutzten Modell angemessen, wobei die Abweichnungen diskutiert werden.

Prior publications

- B. Dehnadi, I. Novikov and F. J. Tackmann, "The photon energy spectrum in $B \to X_s\gamma$ at N^{n}LL$'$", arXiv:2211.07663 (preprint)

Contents

1. Introduction

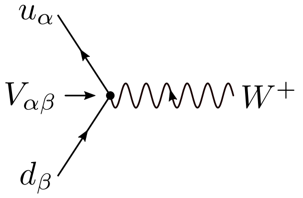
Our understanding of microscopic physics is incomplete. Although the best fundamental theory of particle physics, — the Standard Model (SM), — has been highly successful in describing the observations at high-energy collider experiments, the spectrum of elementary particles and interactions postulated by the Standard Model exhibits unexplained patterns: the three generations of fermions and the hierarchy between their masses, the almost-diagonal structure of the CKM matrix, etc. Furthermore, the CP violation present in the Standard Model interactions appears too small to explain the observed matter-antimatter asymmetry, and the nominal version of the Standard Model does not account for neutrino masses and oscillations. These open questions prompt theorists to speculate about the possible extensions of the Standard Model, the so-called *new physics* or beyond-standard-model (BSM) physics. The search for deviations from predictions of the Standard Model that could hint at new physics has been one of the key goals of collider experiments of this century.

The Standard Model is a gauge theory with a $\mathrm{U}(1) \times \mathrm{SU}(2) \times \mathrm{SU}(3)$ gauge symmetry and an additional Higgs field, which couples to the gauge bosons via gauge interaction and to the three generations of fermions via Yukawa interaction. The Higgs field has a nontrivial self-interaction potential with a "Mexican hat" shape. As a result, in the minimal-energy state the Higgs field acquires a non-zero expectation value $\langle h \rangle = v/\sqrt{2}$, the so-called Higgs condensate. Although the Higgs potential is invariant under SU(2), each individual minimal point of the Higgs potential is not. Consequently, the SU(2) gauge symmetry is spontaneously broken, and is not evident in the physical vacuum. The unbroken Lagrangian of the Standard Model does not include explicit mass terms for the fermions and gauge bosons, because such terms are forbidden by gauge symmetry. However, because these particles couple to the Higgs field, and the Higgs field has a nonzero expectation value in the physical vacuum, the potential energy of their interaction with the Higgs condensate can be interpreted as the rest mass. This is the essence of the Higgs mechanism.

In particular, the masses of the fermions arise from the Yukawa coupling. However, the isospin doublets of quarks that couple to the W boson do not correspond directly to quarks with definite mass. The standard flavor labels of quarks, — u, d, s, c, b, t, — refer to the states with definite mass, while the interaction eigenstates are given by a superposition of these mass eigenstates, and vice versa. As a result, an interaction of a quark with a W boson mixes the three generations of quarks. The amplitude of this mixing is parametrized by a unitary 3×3 matrix called the Cabibbo-Kobayashi-Maskawa

(CKM) matrix:

$$V_{\text{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}, \tag{1.1}$$



$$V_{\text{CKM}}^{\text{magnitudes}} = \begin{pmatrix} 973.73 \pm 0.31 & 224.3 \pm 0.8 & 3.82 \pm 0.2 \\ 221 \pm 4 & 975 \pm 6 & 40.8 \pm 1.4 \\ 8.6 \pm 0.2 & 41.5 \pm 9 & 1014 \pm 29 \end{pmatrix} \cdot 10^{-3} \;\; [1]. \tag{1.2}$$

Eq. (1.2) shows the magnitudes of the elements of the CKM matrix, determined experimentally without assuming unitarity of the CKM matrix [1]. The Standard Model provides no explanation for the observation that the CKM matrix is almost diagonal.

In the Standard Model the CKM matrix arises from the Yukawa interactions after diagonalization of the quark mass terms. This procedure does not fix the complex phases of the six flavors of quark fields. The CKM matrix is invariant under the coherent rotation $q_f \rightsquigarrow e^{i\phi} q_f$ of all six flavors of quarks by the same phase ϕ. The remaining complex phases can be used to fix 5 complex phases of the CKM matrix. Unitarity imposes 9 additional real constraints on the elements of the CKM matrix. As a result, the CKM matrix can be parametrized with 4 real numbers, which are fundamental parameters of the Standard Model. These 4 numbers are usually chosen as 3 rotation angles and one complex phase δ, which is the only source of CP violation in the Standard Model.

Unitarity of the CKM matrix is an experimentally testable prediction of the Standard Model. The unitarity constraints can be visualized as so-called *unitarity triangles* in the complex plane. There is one unitarity triangle for each pair of distinct rows or columns of the CKM matrix. The most popular one, conventionally referred to as *the* unitarity triangle, is shown in figure 1.1. It visualizes the constraint on the product of first and third columns of the CKM matrix:

$$1 + \frac{V_{ud} V_{ub}^*}{V_{cd} V_{cb}^*} + \frac{V_{td} V_{tb}^*}{V_{cd} V_{cb}^*} = 0. \tag{1.3}$$

Unitarity of the CKM matrix implies that the position of the top vertex of the unitarity triangle must be consistent with all experimental constraints simultaneously. The angle $\gamma := \arg\left(-\frac{V_{ud} V_{ub}^*}{V_{cd} V_{cb}^*}\right)$ is approximately equal to the CP-violating complex phase δ, which is why this particular unitarity triangle is convenient for visualization of CP-violation.

As a consequence of confinement, quarks do not exist as free particles at macroscopic distances. They can only be studied as constituents of hadrons, with the exception of the t quark, which decays too quickly to form hadrons. For example, the CKM matrix elements $|V_{cd}|$ and $|V_{cs}|$ can be determined from measurements of decays of D mesons, which contain a constituent c-quark, and $|V_{ub}|$ and $|V_{cb}|$ can be extracted from decays

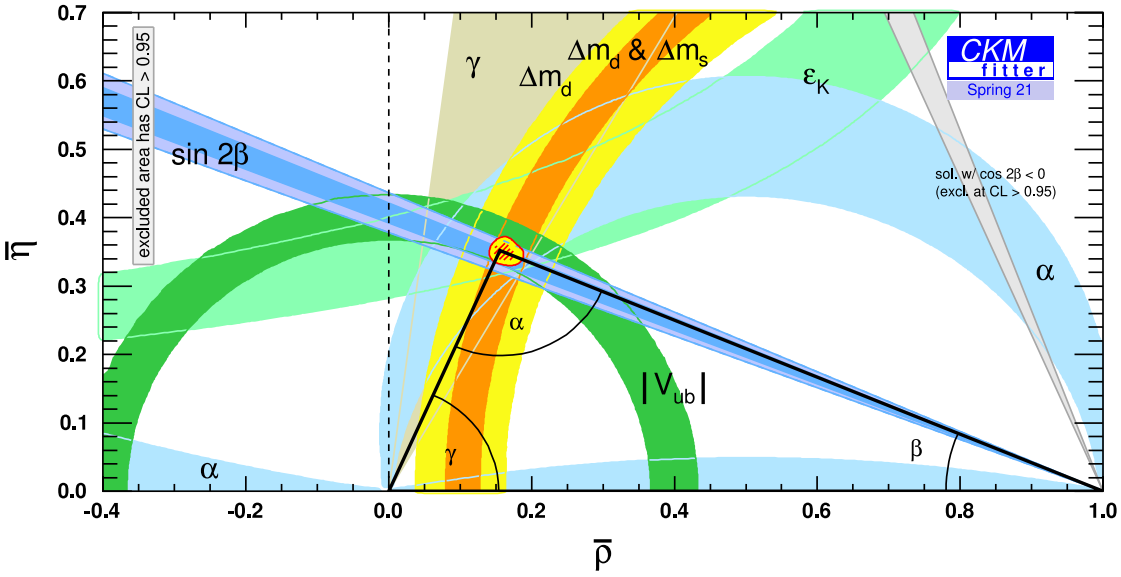


Figure 1.1.: The unitarity triangle in the complex plane $\bar{\rho} + i\bar{\eta} = -\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}$. The shown experimental constraints and the method used to combine them are presented in ref. [2]. Points in the colored regions have p-value $p > 4.55\%$, corresponding to 2σ confidence level, and the points inside the red-dashed area have p-value $p > 31.73\%$, corresponding to 1σ confidence interval.

of B mesons, with a constituent b-quark. Together, these measurements constrain the small side of the unitarity triangle, given by $\left|\frac{V_{ud}V_{ub}}{V_{cd}V_{cb}}\right|$. An improved determination of $|V_{ub}|$ and $|V_{cb}|$ from B-meson decay would increase the precision of this constraint.

Most precise determinations of these CKM matrix elements have been carried out at the so-called B factories, — dedicated e^+e^- collider experiments designed to produce and study B-mesons, — such as Belle/Belle-II and BaBar. These experiments carry out most of the measurements with the center-of-mass energy tuned to the $\Upsilon(4S)$-resonance mass $m_{\Upsilon(4S)} \approx 10.58\,\text{GeV}$. Because $\Upsilon(4S)$ resonance decays predominantly into $B\bar{B}$ pairs, this energy tuning maximizes the number of produced $B\bar{B}$ pairs. Belle-II and BaBar use asymmetric beam energies to produce boosted $B\bar{B}$ pairs, enabling measurements of B-meson lifetimes and CP violation in $B-\bar{B}$ mixing. Although usually not considered a B-factory, the LHCb experiment at the Large Hadron Collider also provides competitive measurements of B-meson decays.

The main determinations of $|V_{ub}|$ and $|V_{cb}|$ are based on measurements of semileptonic B-meson decays $B \to X_{c,u}l\bar{\nu}$. These determinations can be classified as *exclusive* or *inclusive*. The *exclusive* determinations study semileptonic B-meson decays with some specific hadrons $X_{c,u}$ in the final state, while the *inclusive* determinations examine the decay rates summed over all possible species of hadrons $X_{c,u}$ in the final state. Theoretically, predictions of exclusive decay rates rely on so-called *form factors*, which parametrize the amplitude of the transition from B-meson state to some hadronic final

state. These form factors are nonperturbative, but can be determined using in Lattice-QCD calculations.

Theoretical descriptions of inclusive decays take advantage of quark-hadron duality to eliminate the sum over final hadronic states. This inclusive approximation, however, does not describe the production of resonances, — single-particle stable or quasistable hadronic states, — that appear as peaks in the invariant-mass m_X spectrum at relatively small m_X. Although near resonances the local quark-hadron duality is violated, the inclusive approximation is still valid for the description of decay rates integrated over the resonance region, as the resonance effects "average out" for more inclusive observables and the quark-hadron duality is restored. The matrix elements with B-meson states are further simplified using local or non-local operator-product expansion (OPE) in powers of $\Lambda_{\mathrm{QCD}}/m_q$, where m_q is the mass of the heavy c or b quark. This expansion is also called the *Heavy Quark Expansion*. In the case of $B \to X_c l\bar{\nu}$ decay, local OPE can be used to rewrite the differential decay rate in terms of perturbative coefficients and a small set of nonperturbative hadronic parameters [3]. In the case of $B \to X_u l\bar{\nu}$, however, local OPE is not always applicable in the experimentally interesting region, and a non-local version of operator-product expansion is used instead. This case will be discussed in this work in much more detail.

The Monte-Carlo simulations employed by experiments such as Belle-II and BaBar use a so-called *hybrid model* to describe the $B \to X_u l\bar{\nu}$ decay. A hybrid model combines predictions of exclusive models in the resonance region with predictions of the inclusive model in the continuum region [4]. This highlights the fact that the exclusive and inclusive models are complementary, and both are necessary for a complete description of the semileptonic decays.

Both exclusive and inclusive approaches have their advantages and disadvantages. Of course, as long as the assumptions underlying each analysis method are valid, the two approaches should yield compatible results. The reality, however, is more complicated: there has been a persistent tension between the results of inclusive and exclusive determinations of both $|V_{cb}|$ and $|V_{ub}|$ CKM matrix elements, as can be seen in figures 1.2 and 1.3. Clearly such a state of affairs is not satisfactory, and a more thorough investigation of these semileptonic decays is warranted.

$|V_{ub}|$ is the smallest CKM matrix element, and is consequently known with largest relative uncertainty. The rare $B \to X_u l\bar{\nu}$ decay is uniquely suited for the measurement of $|V_{ub}|$, as it is one of the few experimentally accessible processes that is sensitive to $|V_{ub}|$ at leading order in the strong coupling constant α_s (see figure 1.4). Furthermore, the influence of new-physics effects, if there are any, is expected to be small. However, the experimental signal of this decay is obscured by large $B \to X_c l\bar{\nu}$ background: Because $|V_{cb}/V_{ub}| \sim 10$, and the transition probabilities are proportional to the CKM matrix elements squared, in the bulk of the phasespace the decay $B \to X_u l\bar{\nu}$ is suppressed by a factor ~ 100 relative to the $B \to X_c l\bar{\nu}$ decay, which is illustrated in figure 1.5. Experimental measurement of the $B \to X_u l\bar{\nu}$ signal on top of such a large background is challenging. However, because the lightest X_u state, a pion π, is lighter than the lightest X_c state, a D meson, in the phasespace region of small invariant mass m_X of the final hadronic state X the $B \to X_c l\bar{\nu}$ decay is kinematically forbidden, but $B \to X_u l\bar{\nu}$

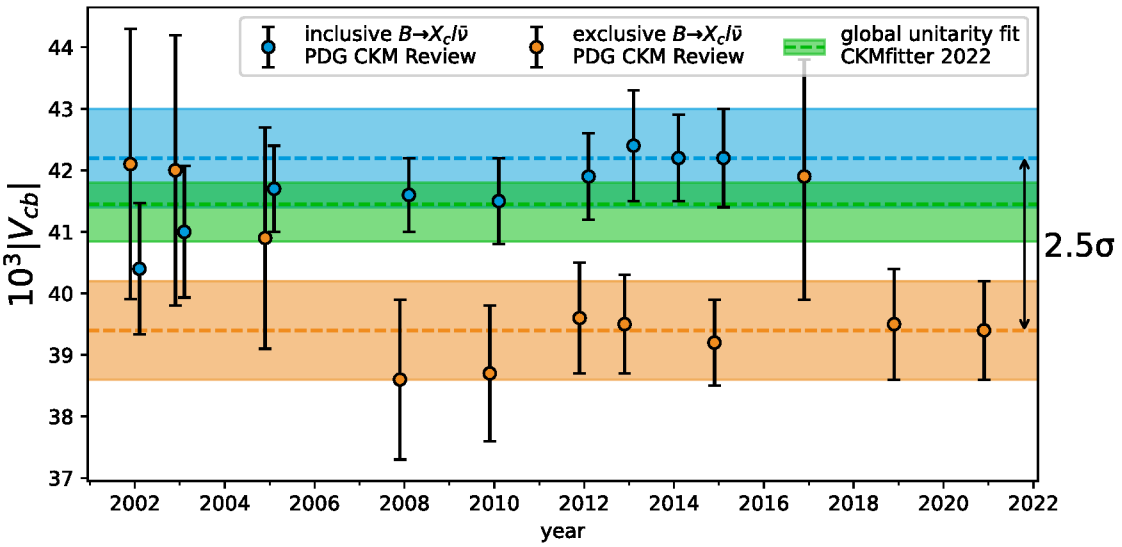


Figure 1.2.: Comparison of inclusive and exclusive determinations of $|V_{cb}|$, as reported by the Particle Data Group reviews in the last 20 years [1]. The blue and orange bands show the results of the latest inclusive and exclusive measurements, respectively. The green band shows the result of the latest CKMfitter global unitarity fit [5].

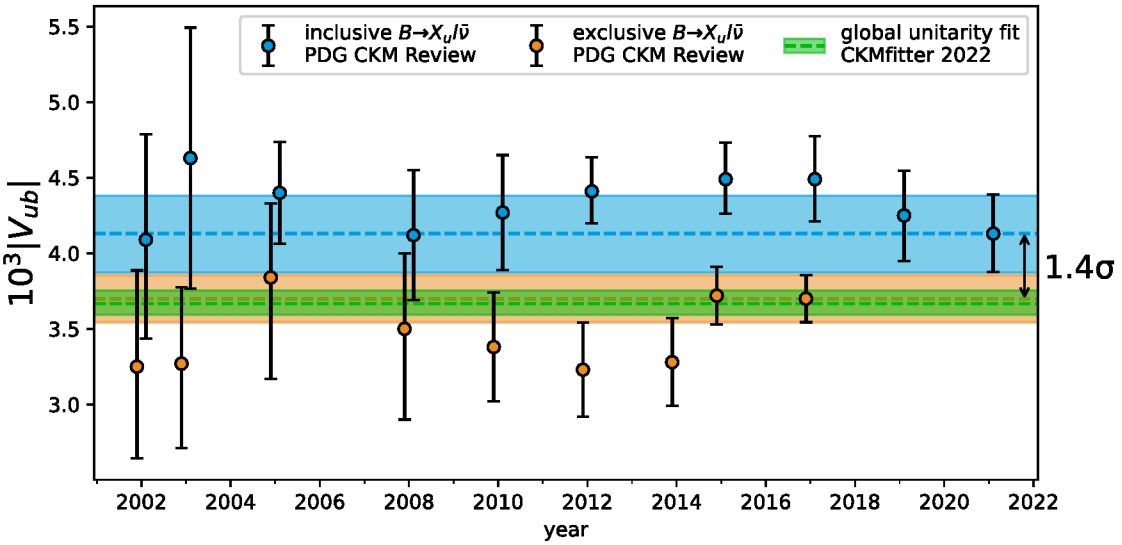


Figure 1.3.: Comparison of inclusive and exclusive determinations of $|V_{ub}|$, as reported by the Particle Data Group reviews in the last 20 years [1]. The blue and orange bands show the results of the latest inclusive and exclusive measurements, respectively. The green band shows the result of the latest CKMfitter global unitarity fit [5].

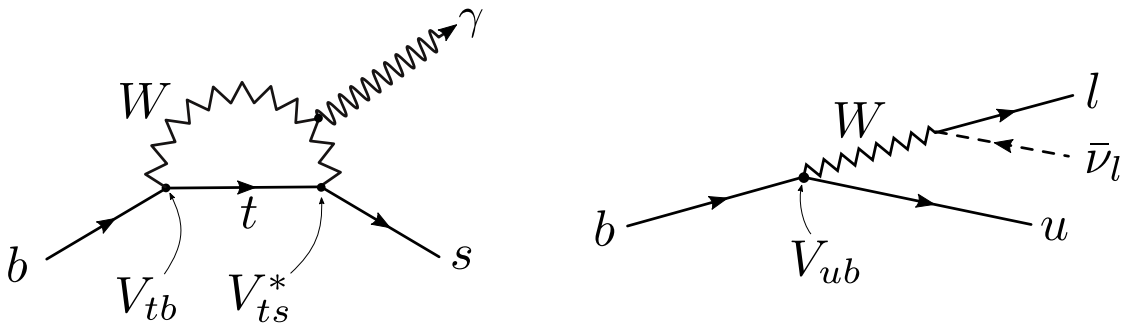


Figure 1.4.: Leading-order Feynman diagrams for the $b \to s\gamma$ (left) and the $b \to ul\bar{\nu}$ (right) transitions.

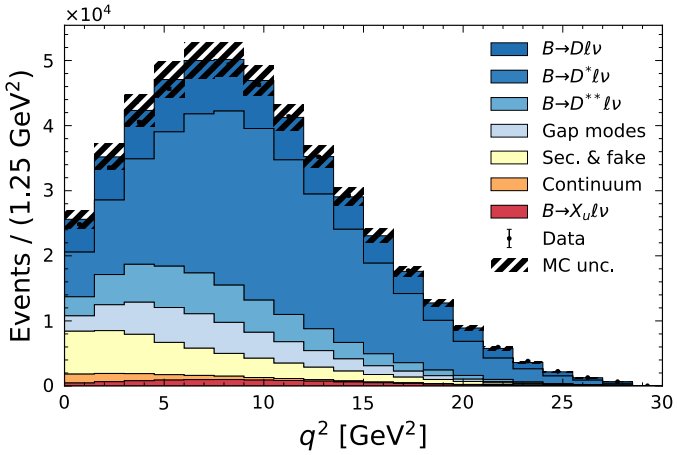


Figure 1.5.: Monte-Carlo simulation of reconstructed $B \to X_u l\bar{\nu}$ signal and background events, with a cut $E_l > 1\,\text{GeV}$ on the energy of the charged lepton. The $B \to X_u l\bar{\nu}$ signal is shown in red, and the $b \to c$ background processes $B \to Dl\nu$, $B \to D^* l\nu$, $B \to D^{**} l\nu$ are shown in blue. Events are binned in the is the reconstructed square-of-invariant-mass q^2 of the lepton pair $l\bar{\nu}$. BDT background suppression is not included. Plot taken from ref. [6].

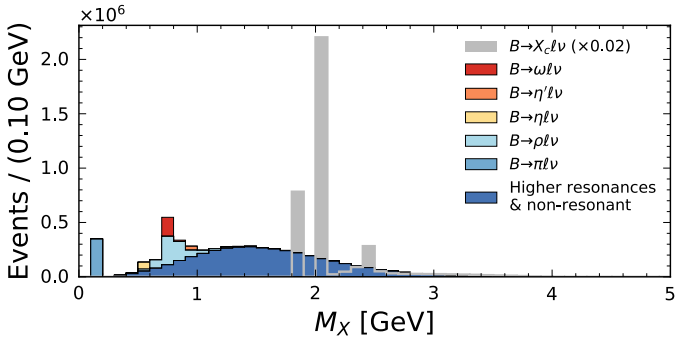


Figure 1.6.: Generator-level simulation of $B \to X_u l\bar{\nu}$ and $B \to X_c l\bar{\nu}$ invariant-mass spectrum. The $B \to X_c l\bar{\nu}$ background is shown in gray and is scaled by a factor 1/50. The colored histograms show different resonant and non-resonant contributions to $B \to X_u l\bar{\nu}$ spectrum. The minimal invariant mass of the $B \to X_c l\bar{\nu}$ final hadronic state is $m_D \approx 1.86\,\text{GeV}$, while for $B \to X_u l\bar{\nu}$ it is $m_\pi \approx 0.135\,\text{GeV}$. Plot taken from ref. [6].

is not. This is illustrated in figure 1.6. By imposing a cut on the invariant mass m_X, the large $B \to X_c l\bar{\nu}$ background can be substantially reduced. Unfortunately, the theoretical description of this endpoint region is more complicated in comparison to the bulk of the phasespace. In particular, local operator-product expansion is not applicable in this endpoint region, and the decay rate is sensitive to nonperturbative effects. On the other hand, Soft-Collinear Effective Theory (SCET) is appropriate for this region. The factorization theorem of SCET disentangles perturbative dynamics at the hard and intermediate energy scales $E \sim m_b$, and $E \sim \sqrt{m_b \Lambda_{\text{QCD}}}$, respectively, and nonperturbative dynamics at the soft energy scale $E \sim \Lambda_{\text{QCD}}$. SCET isolates nonperturbative effects in the so-called *shape functions*. The shape functions encode the Fermi motion of the b-quark inside the B-meson, as well as nonperturbative effects associated with the emission of soft radiation. The expansion of the decay rate in terms of shape functions can be understood as a non-local operator-product expansion. The normalization and first moments of the shape functions are related to the hadronic parameters of the Heavy-Quark Effective Theory (HQET). The shape functions are universal — the same shape functions appear in different B-meson decays, although in different combinations. Although a shape function cannot be calculated, it can be measured in one decay and then used to describe a different decay. At leading power in Λ_{QCD}/m_b there is only one shape function, and it has recently been extracted from measurements of the $B \to X_s\gamma$ photon energy spectrum [7].

The determination of the shape function is based on measurements of the photon energy spectrum in $B \to X_s\gamma$ decay close to the endpoint $E_\gamma = m_B/2$. Because at tree level the $B \to X_s\gamma$ is a two-body decay $b \to s\gamma$, in this approximation the photon energy spectrum would be given by a delta-function $\delta(E_\gamma - m_b/2)$. Perturbative corrections and, more importantly, Fermi motion of the b-quark smear the distribution, but the region $E_\gamma \sim m_b/2$ remains dominant, and the real photon energy spectrum peaks around this

point. The experimental measurements are most precise in this peak region, primarily for the following two reasons: first, the relatively large number of events reduces statistical uncertainty, second, the signal at lower energies is obscured by background photons from ambient processes. The shape of the $B \to X_s\gamma$ photon energy spectrum near the endpoint is dominated by the shape function, and basically *is* the shape function, — up to normalization and perturbative and power corrections. The normalization of the spectrum, on the other hand, can in principle be perturbatively calculated. Because the shape function is normalized to 1, the Standard-Model prediction of the $B \to X_s\gamma$ spectrum normalization can be tested experimentally, in spite of the fact that the shape function has to be fitted to the same measurement. The normalization of the spectrum could be sensitive to new-physics effects, although the current determinations suggest it is compatible with the Standard Model [7].

Although at leading power there is only one shape function, which appears in both $B \to X_s\gamma$ and $B \to X_u l\bar{\nu}$ decays, at subleading power there are multiple shape functions, and they enter $B \to X_s\gamma$ and $B \to X_u l\bar{\nu}$ in different combinations. As a result, subleading shape-function effects in $B \to X_u l\bar{\nu}$ decay cannot be constrained by $B \to X_s\gamma$ measurements as straightforwardly as the leading shape function.

It should be noted that the real $B \to X_s\gamma$ energy spectrum near the endpoint has sharp peaks, corresponding to production of resonances. The inclusive theoretical approach does not describe these resonances, as the quark-hadron duality fails at these points of the spectrum. However, the inclusive predictions are reliable when the energy spectrum is integrated over the resonance region. The presence of these resonances is not evident in the experimental measurements of the $B \to X_s\gamma$ photon energy spectrum due to limited experimental resolution.

The main goal of this work is to improve the theoretical precision of the inclusive $B \to X_s\gamma$ and $B \to X_u l\bar{\nu}$ by incorporating the available next-order perturbative results and to investigate the arising issues of renormalon subtractions and matching between different kinematic regions. The new theoretical models presented in this work have been implemented in the `SCETlib C++` library.

This thesis is organized as follows: The Soft-Collinear Effective Theory (SCET) and the Heavy-Quark Effective Theory (HQET), which form the theoretical foundation for the study of B-meson decays, are briefly reviewed in chapter 2. Chapter 3 is dedicated to the derivation of the leading-power SCET factorization theorem and a discussion of the origin, properties, and treatment of the leading shape function. Different definitions of the b-quark mass and HQET hadronic parameters and their impact on the convergence of perturbative series are investigated in chapter 4. The improved $\mathrm{N^3LL'}$+$\mathrm{N^3LO}(c_k)$ theoretical predictions for the inclusive $B \to X_s\gamma$ decay and corresponding numerical results are presented in chapter 5. In chapter 6 the theoretical predictions for the inclusive $B \to X_u l\bar{\nu}$ decay at $\mathrm{N^3LL}$+NLO are presented and compared with Belle experimental measurements. Finally, chapter 7 summarizes the results of this work.

2. Effective field theories

The notion of an effective field theory is very broad — in the most general sense, an effective field theory is just a field theory ,the Lagrangian of which is a series in some *power counting parameter*, conventionally denoted λ. Most often an effective field theory is constructed to reproduce the dynamics of some full, more fundamental field theory in the regime where the power counting parameter λ is small. In this case, the effective field theory is formulated in terms of degrees of freedom that are relevant for the problem under consideration, rather than degrees of freedom of the full theory. The effect of the remaining degrees of freedom, which are present in the full theory but not in the effective theory, is encoded in the parameters of the Lagrangian of the effective theory.

Often, but not always, the effective theory describes physics in the low-energy limit of the full theory. The power counting parameter in this case is given by the ratio E/Λ of the typical energy E of the process under consideration and the characteristic scale Λ of the short-distance physics. In this context, the low-energy theory is referred to as infrared (IR) effective theory, and the full theory is called its ultraviolet completion (UV-completion).

When the full, more fundamental theory is known, the construction of an effective field theory is called *top-down*. Top-down effective field theories are designed to be compatible with predictions of the full theory in the regime where the effective theory is valid. Their utility comes from the fact that some problems that are too hard to solve in the full theory become tractable in the approximate effective theory. One reason for this simplification is that an effective theory exposes additional symmetries that full theory might have in the considered limit. Another key feature is the separation of dynamics at the different well-separated energy scales, which leads to so-called *factorization theorems*. The factorization theorems, first, disentangle perturbative processes at high-energies and non-perturbative processes at low-energies, and second, provide a systematic way to resum towers of logarithms of the two energy scales that appear in the perturbative series, which improves their convergence.

The simplest example of a top-down effective field theory that is relevant for B-meson decay is the Weak Effective Theory (WET), also known as the Fermi's theory of beta decay. The Weak Effective Theory describes the Standard Model interactions at the energies below the electroweak scale $\sim m_W$. The degrees of freedom in the Weak Effective Theory are the same as in the Standard Model, with the exclusion of the heavy weak gauge bosons $Z, W^{\pm}$, the Higgs boson H^0, and the t-quark.

In contrast to top-down, *bottom-up* refers to the construction of an effective field theory when the full theory is not known, or when the spectrum of particles and interactions in the limit of interest cannot be derived from the Lagrangian of the full theory. Such bottom-up effective field theories are designed to include all possible operators that are

consistent with symmetries of the theory. Usually, bottom-up effective field theories are used to parametrize the effects of short-distance, high-energy physics in the low-energy limit. The prime example of a bottom-up effective theory is the Standard Model Effective Field Theory (SMEFT) [8], an extension of the Standard Model with higher-mass-dimension operators suppressed by powers of Λ_{NP}, — some unknown energy scale of new physics. It provides a systematic framework to parametrize effects of unknown higher-energy physics that is independent of the exact nature of the high-energy interactions. Within this framework, the known, "pure" Standard Model corresponds to leading-power SMEFT.

The canonical example of a bottom-up effective field theory with a known full theory is the chiral perturbation theory (χPT) [9], which is the theory of the low-energy dynamics of QCD. It describes interactions of pions and nucleons at low momentum.

Coefficients of operators in the effective Lagrangian are called the *Wilson coefficients*. The total number of Wilson coefficients, to all orders in the power counting parameter λ, is usually infinite. However, at any finite power of λ the number of Wilson coefficients is finite. Therefore, to derive a prediction that is accurate to some finite order in λ only a limited number of Wilson coefficients needs to be specified. The Wilson coefficients of top-down effective theories can in principle be determined by a so-called *matching* procedure, — by calculating and equating a sufficient number of transition amplitudes in both effective and full theories and solving the resulting constraints for the parameters in the Lagrangian of the effective theory, order-by-order in the power counting parameter λ. An alternative approach to matching that is based on path-integral formalism has been developed in ref. [10]. Matching is usually carried out perturbatively, although the matching of chiral perturbation theory to QCD can only be performed using lattice-QCD methods [11]. When the full theory is not known, or the matching is too complicated, the Wilson coefficients have to be experimentally measured.

Effective field theories are typically not renormalizable in the classical sense, — the divergences that arise from integrals over the whole space, or over all momentum space, are not cancelled by a finite number of counterterms. This is related to the presence of higher-dimension operators in the effective Lagrangian. However, they are renormalizable in the generalized sense, — at any finite order in the power counting parameter λ.

Although effective field theories are usually used as an approximation of some full theory, in general, an effective field theory is a full-fledged theory by itself, and can be formulated and studied without any reference to an underlying full theory. In fact, Fermi's theory of weak interactions and chiral perturbation theory have been invented and productively used before the invention of modern Glashow-Weinberg-Salam model of electroweak interactions and quantum chromodynamics (QCD), respectively. Furthermore, the same effective field theory can approximate different full theories. For example, the chiral perturbation theory emerges as a low-energy effective theory of both QCD and the linear sigma model [9]. To put it another way, a given IR effective theory can have many possible UV completions.

Applications of effective-field-theory methods are numerous and are not limited to collider physics. For example, they have been used to derive quantum corrections to

gravity at low energies and curvatures [12] as well as to study the emission of gravitational waves by binary systems [13]. The latter is an example of a classical, non-quantum effective field theory.

The effective theories that are particularly important in the study of B-meson decays are the Heavy-Quark Effective Theory (HQET) and the Soft-Collinear Effective Theory (SCET). Both HQET and SCET are top-down effective field theories of QCD. The relevant power counting parameter is $\lambda = \frac{\Lambda_{\rm QCD}}{m_b}$, where $m_b \sim 4.7\,\text{GeV}$ is the b-quark mass and $\Lambda_{\rm QCD} \sim 1\,\text{GeV}$ is the confinement scale of strong interaction. The aspects of HQET and SCET that are necessary for the further discussion of B-meson decays are summarized in the following two sections.

2.1. Soft-collinear effective theory (SCET)

Soft-Collinear Effective theory describes soft, — low-energy, — and collinear, — strongly boosted, — degrees of freedom, which is relevant for processes with one or more jets in the final state or with beams in the initial state. Such processes typically also involve wide-angle soft radiation. The necessity of SCET is related to the fact that a strongly boosted QCD parton, — a quark or a gluon, — primarily emits quarks or gluons that are soft or strongly boosted in the same direction. This fact is reflected in the so-called soft and collinear divergences in the amplitudes of processes that involve such emissions. In the practical calculations, such divergences need to be carefully isolated and cancelled, which is a complicated problem in general. SCET provides a systematic approach to this problem. The key idea of SCET is to treat soft and collinear quarks and gluons as if they were different kinds of particles, with separate quantum fields for each soft and collinear sector. SCET is constructed from QCD in a top-down manner. The SCET Lagrangian is derived by rewriting each QCD field as a sum of soft and collinear fields, and expanding in the power counting parameter $\lambda = \frac{\Lambda_{\text{QCD}}}{m_b}$.

Although many effective field theories are designed to reproduce the low-energy limit of the full theory, SCET is an example of a theory that cannot be understood simply as a low-energy limit of QCD. Different components of collinear momenta are of different order in λ, and SCET describes dynamics at different scales — both at low energies and at high energies. Although the degrees of freedom of SCET can be understood as a subset of the degrees of freedom of full QCD, formally SCET has more fields that QCD, and its Lagrangian is constructed out of multiple copies of QCD Lagrangian, — one for each soft and collinear sector. This paradox is resolved by the fact that formally the number of degrees of freedom in both SCET and QCD is infinite, and both the range of all possible momenta in QCD and the range of all possible momenta with a given soft or collinear scaling in SCET, are the same. This subtlety leads to the so-called "zero-bin subtraction" prescription, which applies to the integrals over all collinear momenta, and is designed to avoid double counting of soft and collinear degrees of freedom. There are multiple versions of SCET with different number of soft and collinear sectors, which are designed to describe different processes.

The key feature of SCET is that at leading power its dynamical Lagrangian is *separable*, — it is given by the sum of soft and collinear Lagrangians, with no terms that mix different soft or collinear sectors. This means that the fields in different soft and collinear sectors interact only with fields in the same sector, — there is no interaction between soft and collinear particles, or between collinear particles boosted in different directions. Stated formally, separability of the Lagrangian implies that the Hilbert space of states factorizes into a product of Hilbert spaces of states for each soft and collinear sector. However, beyond the dynamical Lagrangian, the hard interaction operators do mix different soft and collinear sectors. Furthermore, terms in the dynamical SCET Lagrangian at subleading power in λ can also introduce interactions between soft and collinear fields. However, the hard interaction operators and subleading Lagrangian terms can be treated perturbatively, — the perturbation is suppressed either by a small coupling constant of the hard operator, or by the small power counting parameter λ

for subleading Lagrangian terms. As a result, the hard and subleading operators can be treated as operators on the leading-power Hilbert space, which still factorizes. This is the essence of the soft-collinear factorization mechanism of SCET, which enables the derivation of factorization theorems. At leading power, these factorization theorems can be used to rewrite transition probabilities as a product or a convolution of hard, soft, and jet (or beam) functions. The soft and jet functions are given by matrix elements in soft and collinear sectors of SCET, respectively, and the hard function is given by the squared magnitudes of the Wilson coefficients of hard interactions.

In some cases, however, the factorization of different soft and collinear modes is not as simple as presented in the last paragraph. In problems involving particles that are strongly boosted in opposite directions, the power counting analysis identifies the so-called Glauber gluon mode, which can couple to collinear particles in both collinear sectors, and appears at leading power in the power counting parameter λ. When Glauber gluons are accounted for in the SCET Lagrangian, it is no longer separable, and the two collinear sectors do interact via Glauber gluons. Although the contributions of Glauber gluons have been shown to cancel in some specific processes, in general they present an obstacle to factorization, and no general solution exists. Fortunately, because B-meson decays in the endpoint region involve only one jet, the relevant version of SCET has only one collinear sector, and the Glauber modes do not arise.

A detailed review of Soft-Collinear Effective Theory and its formalism can be found in ref. [14]. The following subsections reviews only the most basic features of SCET, which are necessary for the following discussion of B-meson decays.

2.1.1. Lightcone coordinates

The so-called lightcone coordinates is a specific choice of coordinates in Minkowski space that is convenient for analysis of momenta of strongly boosted particles. The coordinates are defined by a choice of two reference lightlike vectors n and $\bar{n}$. One of the two reference momenta is usually chosen to point in the direction of the boosted particle, and the other in the opposite spatial direction. By convention, the vectors n and $\bar{n}$ are picked to satisfy the following normalization and orthogonality conditions:

$$n^2 = \bar{n}^2 = 0, \qquad \bar{n} \cdot n = 2. \tag{2.1}$$

Any 4-vector p can be decomposed into plus, minus, and transverse components:

$$p^+ \coloneqq n \cdot p, \qquad p^- \coloneqq \bar{n} \cdot p, \qquad p^\mu = \frac{1}{2} n^\mu p^- + \frac{1}{2} \bar{n}^\mu p^+ + p_\perp^\mu, \tag{2.2}$$

The momentum p of a particle that is strongly boosted in the direction n has a large p^- component and a small p^+ component, which will be relevant in the next subsection, where the SCET power counting is discussed.

The Dirac gamma matrices γ^μ can be decomposed in the same way:

$$\gamma^\mu = n^\mu \not{\bar{n}}/2 + \bar{n}^\mu \not{n}/2 + \gamma_\perp^\mu. \tag{2.3}$$

The scalar product of two vectors can be written in terms of lightcone coordinates as:

$$p \cdot q = (p^+ q^- + p^- q^+)/2 + p_\perp \cdot q_\perp, \qquad p^2 = p^+ p^- + p_\perp^2. \tag{2.4}$$

It is always possible to choose a coordinate system in which these vectors have the following components:

$$p = (E, p_x, p_y, p_z), \qquad n = (1, 0, 0, 1), \qquad \bar{n} = (1, 0, 0, -1). \tag{2.5}$$

Then

$$p^\pm = E \mp p_z. \tag{2.6}$$

The following notation for lightcone coordinates of a vector p is also used:

$$p = (p^+, p^-, p_\perp). \tag{2.7}$$

2.1.2. Power counting

In the simplest version of SCET, the degrees of freedom, — quarks and gluons, — can be classified as either hard, soft, or collinear in some direction n, according to the characteristic size of their momentum components. The lightcone momentum components $(p^+, p^-, p_\perp)$ of hard, n-collinear, and soft particles are of order:

$$p_{\text{hard}} \sim (1, 1, 1)Q, \qquad p_n \sim (\lambda, 1, \sqrt{\lambda})Q, \qquad p_{\text{soft}} \sim (\lambda, \lambda, \lambda)Q, \tag{2.8}$$

where Q is the relevant hard scale of the problem, which is $Q \sim m_B \sim m_b$ in the case of B-meson decay, and the small power counting parameter is $\lambda \coloneqq \Lambda_{\text{QCD}}/Q$. It should be noted that the hard particles are integrated out in SCET, and are not treated as dynamical degrees of freedom, their effect is encoded in the effective hard interaction vertices and the corresponding Wilson coefficients. Although in general there can be multiple collinear directions, in the context of B-meson decay there is only one jet in the final state, and only one collinear direction as a result. The squared invariant masses of the hard, n-collinear, and soft particles also have a definite power counting:

$$p_{\text{hard}}^2 \sim Q^2, \qquad p_n^2 \sim \lambda Q^2, \qquad p_{\text{soft}}^2 \sim \lambda^2 Q^2. \tag{2.9}$$

The same relative power counting is assigned to the components of the gluon field A^μ of hard, n-collinear, and soft gluons:

$$A_{\text{hard}} \sim (1, 1, 1), \qquad A_n \sim (\lambda, 1, \sqrt{\lambda}), \qquad A_{\text{soft}} \sim (\lambda, \lambda, \lambda). \tag{2.10}$$

The fields $q_{\text{hard}}, \chi_n, q_{\text{s}}$ of hard, n-collinear, and soft quarks are counted as

$$q_{\text{hard}} \sim 1, \qquad \chi_n \sim \sqrt{\lambda}, \qquad q_{\text{s}} \sim \sqrt{\lambda^3}. \tag{2.11}$$

Furthermore, the spin of an n-collinear quark with definite helicity is always assumed to be along the direction of n. The mass of collinear quarks is neglected in SCET, and the Dirac equation-of-motion of a free quark becomes

$$(\not{p} - m)\psi = p^-(\not{n}/2 + \mathcal{O}(\lambda))\psi = 0. \tag{2.12}$$

The spinor components of the gauge-invariant collinear quark field χ_n in SCET are defined such that the equation $\not{n}\chi_n = 0$ holds exactly.

It is convenient to separate the momentum p of a particle into the so-called large *label* momentum p_l and small *residual* momentum p_r: $p = p_l + p_r$. For n-collinear particles with momentum p of order $p \sim (\lambda, 1, \sqrt{\lambda})$ the label momentum is defined to be of order $p_l \sim (0, 1, \sqrt{\lambda})$, while the residual momentum is of order $p_r \sim (\lambda, \lambda, \lambda)$. In particular, the small component p^+ of n-collinear momentum is purely residual. The label momentum of soft particles is zero by definition, their momentum is purely residual. In SCET the label and residual momenta are separately conserved. The full momentum operator $\hat{p}$ can be split into operators that separately measure label and residual momentum: $\hat{p} = \hat{p}_l + \hat{p}_r$.

2.1.3. Collinear quark field

In the context of matching, it is useful to introduce the following definition of the quark field $\chi_{n;\omega}$ with a specific large label momentum component $p_l^- = \omega$ [15]:

$$\chi_{n;\omega}(x) \coloneqq [\delta(\omega - \bar{\mathcal{P}})\chi_n(x)], \tag{2.13}$$

where $\bar{\mathcal{P}} \coloneqq \bar{n} \cdot \mathcal{P}$, and the calligraphic label momentum operator $\mathcal{P}$ is related to the operator $\hat{p}_l$ that measures label momentum as $\mathcal{P} = -\hat{p}_l$. The square brackets in eq. (2.13) indicate that the operator $\bar{\mathcal{P}}$ acts only on the field $\chi_n(x)$, and not on the states to the right. Its action on the field can be defined as the commutator $[\bar{\mathcal{P}}\chi_n(x)] \coloneqq [\bar{\mathcal{P}}, \chi_n(x)]$. The field $\chi_{n;\omega}$ satisfies the following identities:

$$[\bar{n} \cdot \hat{p}_l, \chi_{n;\omega}(x)] = -\omega\chi_{n;\omega}(x), \qquad \int \mathrm{d}\omega \chi_{n;\omega}(x) = \chi_n(x). \tag{2.14}$$

The sign difference between label momentum operators $\mathcal{P}$ and $\hat{p}_l$ is related to the fact that the calligraphic label momentum operator $\mathcal{P}$ is defined such that $[\mathcal{P}\chi_{n;\omega}] \coloneqq [\mathcal{P}, \chi_{n;\omega}] = \omega\chi_{n;\omega}$, while the field operator $\chi_{n;\omega}(x)$ is defined to destroy n-collinear particles and to create n-collinear antiparticles, such that the total change in the large p_l^--component of collinear momentum is $-\omega$. An equivalent definition for the field operator $\chi_{n;\omega}$ is

$$\chi_{n;\omega}(x) = \int \mathrm{d}k \delta(\bar{n} \cdot \hat{p}_l - k + \omega)\chi_n(x)\delta(\bar{n} \cdot \hat{p}_l - k). \tag{2.15}$$

Here and in the following an application $f(\hat{A})$ of a generalized function f to a hermitian operator $\hat{A}$ is defined in terms of eigenvalues of the operator $\hat{A}$:

$$f(\hat{A}) \coloneqq \sum_A |A\rangle f(A) \langle A|, \tag{2.16}$$

where the sum is over the eigenstates of the operator $\hat{A}$.

2.2. Heavy-quark effective theory (HQET)

Heavy-Quark Effective theory describes soft interactions of light degrees of freedom with a nonrelativistic heavy quark Q, which is the b-quark in the case of B-meson decay. HQET is formulated order-by-order in the small power counting parameter $\lambda \coloneqq \frac{\Lambda_{\rm QCD}}{m_Q}$. The light degrees of freedom are assumed to have all momentum components much smaller than the mass m_Q of the heavy quark. As a result, interaction of a heavy quark with soft gluons leaves the velocity of the heavy quark essentially unchanged. At leading power in HQET power counting, i.e. in the limit $m_Q \to \infty$, the heavy quark is treated as a static color charge.

The leading-power HQET exhibits additional symmetries that are not evident in full QCD. First, in pure QCD, without electroweak interactions, the different quark flavors are distinguished only by their mass. In the limit $m_Q \to \infty$, when all heavy-quark masses are infinitely large, the difference between masses of different quark flavors is irrelevant, and as a result all processes are invariant under change of the flavor of a heavy quark. Second, because the chromomagnetic moment is inversely proportional to the mass, in the limit $m_Q \to \infty$ there are no chromomagnetic interactions. Because the chromomagnetic interactions are the only spin-dependent interactions in pure QCD, all processes in the leading-power HQET are invariant under the change of spin of the heavy quark.

Interactions that violate these flavor and spin symmetries are taken into account as subleading corrections in $\frac{\Lambda_{\rm QCD}}{m_Q}$. Because these corrections are suppressed by $\frac{\Lambda_{\rm QCD}}{m_Q}$ they necessarily involve processes at the QCD hadronization scale $\Lambda_{\rm QCD}$, and therefore the origin of such corrections is nonperturbative. A key feature of HQET is that it provides a method to systematically keep track of such nonperturbative corrections in computations involving hadrons with a heavy constituent quark.

The most basic features of HQET, which are necessary for the further discussion, are summarized in the following subsection. A more detailed review of Heavy-Quark Effective Theory can be found in ref. [16].

2.2.1. Heavy quark field

The momentum p_b of a heavy quark b can be separated into large label momentum $m_b v$ and small residual momentum k: $p_b = m_b v + k$. The velocity vector v is defined such that $v^2 = 1$, and in the rest frame of the heavy quark its components are $v = (1, 0, 0, 0)$. In the context of B-meson decay the reference vector v is related to the lightcone reference vectors n and $\bar{n}$ as $2v = n + \bar{n}$. The residual momentum k is of order $k \sim \lambda m_b \sim \Lambda_{\rm QCD}$. At leading power in HQET power counting the heavy quark interacts only with soft gluons, with momenta of order $k \sim \lambda m_b \sim \Lambda_{\rm QCD}$.

The Dirac equation of motion of a heavy quark becomes:

$$(\not{p} - m_b)\psi = m_b(\not{v} - 1 + \mathcal{O}(\lambda))\psi = 0. \tag{2.17}$$

Correspondingly, the spinor components of the HQET field b_v of the heavy quark b are defined such that the equality $\not{v} b_v = b_v$ is exact. The heavy-quark field operator b_v

destroys a heavy quark with label momentum $m_b v$, but does not create a corresponding antiquark. As a result, the action of label momentum operator $\hat{p}_l$ on the heavy-quark field b_v is $[\hat{p}_l, b_v(x)] = -m_b v b_v(x)$.

The heavy-quark field $b_v(x)$ is defined with an additional phase shift $e^{im_b v\cdot x}$ that takes out the label momentum $m_b v$. As a result, the usual identities for a local operator $A(x)$ and momentum operator $\hat{p}$:

$$A(x) = e^{i\hat{p}x} A(0) e^{-i\hat{p}x}, \qquad [\hat{p}, A(x)] = -i\partial A(x), \tag{2.18}$$

do not hold for the operator $b_v(x)$. Instead, it satisfies similar identities, but with the residual momentum operator $\hat{p}_r$:

$$b_v(x) = e^{i\hat{p}_r x} b_v(0) e^{-i\hat{p}_r x}, \qquad [\hat{p}_r, b_v(x)] = -i\partial b_v(x). \tag{2.19}$$

The leading-power HQET Lagrangian is

$$\mathcal{L}_{\text{HQET}} = \bar{b}_v (iv\cdot D - \delta m_b) b_v, \tag{2.20}$$

where $iD = i\partial + gA$ is the covariant derivative. The residual mass term δm_b depends on the definition of the b-quark mass that was used to define the field b_v. The simplest choice $\delta m_b = 0$ corresponds to pole scheme, but other choices are also valid, as long as the residual mass term is of order $\delta m_b \sim \lambda m_b \sim \Lambda_{\text{QCD}}$. This definition of the b-quark mass will be discussed in much more detail in chapter 4.

In the context of matching it is also useful to define the gauge-invariant heavy quark field $\mathcal{H}_v^n$:

$$\mathcal{H}_v^n(x) \coloneqq Y_n^\dagger(x) b_v(x), \tag{2.21}$$

where $Y_n(x)$ is the soft Wilson line, which defined as [17]:

$$Y_n(x) \coloneqq \mathbf{P} \exp\left[ig \int_{-\infty}^{0} \mathrm{d}s\, n\cdot A_{\text{soft}}(x+sn) \right], \tag{2.22}$$

where $\mathbf{P}$ denotes the path-ordering of operators. The soft Wilson line satisfies the identities

$$[n\cdot i\partial Y_n(x)] = -g(n\cdot A_{\text{soft}}(x)) Y_n(x), \qquad Y_n(x)^\dagger Y_n(x) = 1, \tag{2.23}$$

and therefore also satisfies the following operator identity for any generalized function f:

$$Y_n^\dagger(x) f(n\cdot i\partial) Y_n(x) = f(n\cdot iD), \tag{2.24}$$

where, again, function applied to a hermitian operator should be understood in terms of eigenvalues of the operator.

3. Leading-power factorization theorem

In this section the SCET factorization formula is derived for the inclusive B-meson decays $B \to X_s\gamma$ and $B \to X_u l\bar{\nu}$. The factorization is valid at leading power in the power counting parameter $\lambda = \frac{\Lambda_{\rm QCD}}{m_b}$ and the electromagnetic coupling $\alpha_{\rm em} = \frac{e^2}{4\pi}$, but it holds to all orders in the strong coupling constant α_s.

The first step is the factorization of strong and electroweak subprocesses. The Weak Effective Theory (WET) describes both strong and electroweak interactions at energies below the W-boson mass m_W. However, at leading order in the electromagnetic coupling $\alpha_{\rm em}$ the electromagnetic interaction is included as a small perturbation on top of a theory free from electroweak interactions, i.e. in the limit $\alpha_{\rm em} \to 0$. In the limit $\alpha_{\rm em} \to 0$ the QCD and electroweak sectors of the Weak Effective Theory do not interact, and the WET Lagrangian $\mathcal{L}$ is separable into QCD and electroweak Lagrangians $\mathcal{L}_{\rm QCD}$ and $\mathcal{L}_{\rm EW}$. As a result, the Fock space of states factorizes into QCD and electroweak subspaces, the states $|X\rangle$ factorize into QCD and electroweak states $|X^{\rm QCD}\rangle$ and $|X^{\rm EW}\rangle$, and interaction operators $\mathcal{L}_{\rm int}$ factorize into QCD and electroweak currents $J_\mu^{\rm QCD}$ and $(J^{\rm EW})^\mu$:

$$|X\rangle = |X_{\rm QCD}\rangle \otimes |X_{\rm EW}\rangle, \qquad\qquad \mathcal{L}_{\rm int} = J^\mu_{\rm QCD} \otimes (J_{\rm EW})_\mu, \tag{3.1}$$

where the symbol $\otimes$ denotes the tensor product, and the indices μ indicate that the current operators belong to some representation of the Lorentz group. Although the index μ can be assumed to be a vector index for both $B \to X_u l\bar{\nu}$ and $B \to X_s\gamma$ decays, this specific choice of representation is not necessary for the following discussion.

Similarly, the initial B-meson state is assumed to belong purely to the QCD sector, because at leading order in $\alpha_{\rm em}$ the additional electroweak interactions with the b-quark and spectator partons can be neglected:

$$|B\rangle = |B\rangle_{\rm QCD} \otimes |0\rangle_{\rm EW}. \tag{3.2}$$

As a result, the amplitude of the transition from the B-meson state to the final state X factorizes into a product of QCD and electroweak amplitudes:

$$\langle X|\mathcal{L}_{\rm int}|B\rangle = \langle X_{\rm QCD}|J^\mu_{\rm QCD}|B\rangle\langle X_{\rm EW}|(J_{\rm EW})_\mu|0\rangle. \tag{3.3}$$

At leading power in $\alpha_{\rm em}$ the electroweak amplitude is trivial to evaluate. The momenta of the B-meson initial state $|B\rangle$, the electroweak final state $|X_{\rm EW}\rangle$, and the QCD final state $|X_{\rm QCD}\rangle$ will be denoted $m_B v$, q, and p_X respectively. Momentum conservation is then written as $m_B v = q + p_X$. The goal of the following discussion is to derive a formula for the differential decay rate as a function of the momentum q.

The decay rate is proportional to the transition amplitude squared, summed over all possible hadronic final states X_{QCD} and over spins of the final electroweak state X_{EW}:

$$\mathrm{d}\Gamma(q) \propto \sum_{\text{spin}} \sum_{X_{\mathrm{QCD}}} \delta(p_X + q - m_B v) |\langle X|\mathcal{L}_{\mathrm{int}}(0)|B\rangle|^2 = L^{\mu\nu}(q) W_{\mu\nu}(m_B v - q), \tag{3.4}$$

where the electroweak tensor $L^{\mu\nu}$ is defined as

$$L^{\mu\nu}(q) \coloneqq \sum_{\text{spin}} \langle 0|J^{\dagger\mu}_{\mathrm{EW}}(0)|X_{\mathrm{EW}}\rangle\langle X_{\mathrm{EW}}|J^{\nu}_{\mathrm{EW}}(0)|0\rangle, \tag{3.5}$$

and the hadronic tensor $W_{\mu\nu}$ is

$$W^{\mu\nu}(p) \coloneqq \sum_{X_{\mathrm{QCD}}} \delta(p_X - p) \langle B|J^{\dagger\mu}_{\mathrm{QCD}}(0)|X_{\mathrm{QCD}}\rangle\langle X_{\mathrm{QCD}}|J^{\nu}_{\mathrm{QCD}}(0)|B\rangle. \tag{3.6}$$

Here the sum $\sum_{X_{\mathrm{QCD}}}$ includes summation over the number of hadrons in the final state, their species, and spin configurations, and it also includes integrations over their momenta, with a Lorentz-invariant integration measure. The symbol (0) after a current indicates that all field operators are taken at position $x = 0$. Consequently, the conservation of momenta is encoded entirely in the δ-function $\delta(p_X + q - m_B v)$. It should be noted that in general the electroweak and hadronic tensors $L_{\mu\nu}, W_{\mu\nu}$ are not real-valued, but they are hermitian in the sense that $W^*_{\mu\nu} = W_{\nu\mu}$ and $L^*_{\mu\nu} = L_{\nu\mu}$.

As already mentioned, the electroweak tensor $L_{\mu\nu}$ is trivial to calculate. The hadronic tensor $W_{\mu\nu}$, on the other hand, cannot be simplified without additional assumptions. The primary assumption of the inclusive approach is the local quark-hadron duality, which states that a sum over hadronic states with a definite momentum is equivalent to a sum over quark and gluon states with the same momentum:

$$\sum_{X_{\mathrm{QCD}}} \delta(p_{X_{\mathrm{QCD}}} - p)|X_{\mathrm{QCD}}\rangle\langle X_{\mathrm{QCD}}| \approx \sum_{X} \delta(p_X - p)|X\rangle\langle X| =: \delta(\hat{p} - p). \tag{3.7}$$

The assumption of local quark-hadron duality holds only if the invariant mass $m_X \coloneqq \sqrt{p_X^2}$ is much larger than the masses of the hadrons in the final state $|X_{\mathrm{QCD}}\rangle$. In particular, inclusive predictions which are based on eq. (3.7) do not describe resonances, — sharp peaks in the invariant-mass spectrum, which correspond to masses of specific hadrons in the final state. This is a fundamental limitation of the inclusive approach. Nevertheless, the approximation in eq. (3.7) becomes more reliable when integrated over the momentum q. In particular, although inclusive models do not describe resonant structures, their predictions are valid when integrated over a sufficiently large region of the phasespace, even if that region includes the resonance region.

Assuming local quark-hadron duality, the hadronic tensor can be simplified as follows:

$$\begin{aligned} W^{\mu\nu}(m_B v - q) &\approx \langle B|J^{\dagger\mu}_{\mathrm{QCD}}(0)\delta(\hat{p} - m_B v + q) J^{\nu}_{\mathrm{QCD}}(0)|B\rangle \\ &= (2\pi)^{-4}\,\mathrm{Disc}_{q_0} \int \mathrm{d}^4x e^{iqx} \langle B|\,\mathbf{T}\, J^{\dagger\mu}_{\mathrm{QCD}}(0) J^{\nu}_{\mathrm{QCD}}(x)|B\rangle, \end{aligned} \tag{3.8}$$

where the symbol **T** denotes the time-ordering of operators, and the discontinuity operator Disc is defined in appendix A.1.

In the SCET approximation the full QCD current J^μ_{QCD} is matched onto the SCET current. For the heavy-to-light transition, such as $b \to s$ or $b \to u$, the SCET current has the form $\bar{\chi}_{n;\omega}\Gamma^\mu_i \mathcal{H}^n_v$, and the matching is [18]:

$$J^\mu_{\text{QCD}} \to \int \mathrm{d}\omega C^i(\omega) \bar{\chi}_{n;\omega}\Gamma^\mu_i \mathcal{H}^n_v. \tag{3.9}$$

Here $\chi_{n;\omega}$ and $\mathcal{H}^n_v$ are gauge-invariant fields of an n-collinear and a heavy quark, defined in eq. (2.13) and eq. (2.21), respectively. The $C^i(\omega)$ is the hard matching coefficient, and Γ^μ_i is some gamma matrix structure. A summation over the index i is implied. The integration over the variable ω encodes the fact that the hard matching coefficient C^i depends on the large label momentum component of the n-collinear quark. [1] The soft and collinear degrees of freedom in SCET do not interact, and as a result the states and operators factorize into soft and collinear sectors. In fact, the SCET current in eq. (3.9) is already factorized, because $\bar{\chi}_{n;\omega}$ belongs to the collinear sector, and $\mathcal{H}^n_v$ belongs to the soft sector, because heavy quarks in HQET interact only with soft gluons.

Because the B-meson state $|B\rangle$ includes only the heavy quark and light degrees of freedom with residual momenta of order $\sim \Lambda_{\text{QCD}}$, it does not include any collinear particles, and therefore belongs purely to the soft sector:

$$|B\rangle = |B\rangle_{\text{soft}} \otimes |0\rangle_n. \tag{3.10}$$

The momentum operator $\hat{p}$ can also be separated into soft and n-collinear momentum operators: $\hat{p} = \hat{p}_{\text{soft}} + \hat{p}_n$, which measure momentum in the soft and collinear sectors separately. Both $\hat{p}_{\text{soft}}$ and $\hat{p}_n$ measure the full momentum, which is the sum of label and residual momenta. For this reason it is convenient to further factorize the soft momentum operator $\hat{p}_{\text{soft}}$ into soft label and soft residual momentum operators: $\hat{p}_{\text{soft}} = \hat{p}_{\text{soft};l} + \hat{p}_{\text{soft};r}$. Because the B-meson momentum is defined as $m_B v$ exactly, with no additional residual momentum, the B-meson state $|B\rangle$ is an eigenstate of the soft label and residual momentum operators: $\hat{p}_{\text{soft};l}|B\rangle = m_B v|B\rangle$ and $\hat{p}_{\text{soft};r}|B\rangle = 0$. Since $[\hat{p}_{\text{soft};l}, \mathcal{H}^n_v(x)] = -m_b v \mathcal{H}^n_v(x)$ the soft state $\mathcal{H}^n_v|B\rangle$ that appears in the hadronic tensor in eq. (3.8) is also an eigenstate of the soft label momentum: $\hat{p}_{\text{soft};l}\mathcal{H}^n_v|B\rangle = (m_B - m_b)v\mathcal{H}^n_v|B\rangle$. As a result, the δ-function in eq. (3.8) can be simplified to:

$$\delta(\hat{p} - m_B v + q) \to \delta(\hat{p}_n + \hat{p}_{\text{soft};r} - p) = \int \mathrm{d}^4 k \delta(\hat{p} + k - p)_n \otimes \delta(\hat{p}_r - k)_{\text{soft}}, \tag{3.11}$$

where the momentum p is $p \coloneqq m_b v - q = p^+ \bar{n}/2 + p^- n/2$. The left operator in the tensor product in eq. (3.11) acts on the n-collinear states and measures the full momentum in the collinear sector, i.e. the sum of both label and residual momenta. The right operator acts on the soft states and measures only the residual momentum in the soft sector.

[1] The hard matching coefficients C^i do not depend on the transverse component of label momentum p_l of the collinear quark because the lightcone vector n is always chosen such that $(p_l)_\perp = 0$.

After these simplifications, the expression for the hadronic tensor $W^{\mu\nu}$ in eq. (3.8) becomes:

$$W^{\mu\nu} = (\bar{\Gamma}_i^\mu)^{s_1}{}_{s_2}(\Gamma_j^\nu)^{s_3}{}_{s_4} \int \mathrm{d}\omega_1 \mathrm{d}\omega_2 \mathrm{d}^4k C^{i*}(\omega_1) C^j(\omega_2) \\ \cdot \langle 0|\chi^{s_2}_{n;\omega_1;a}(0)\delta(\hat{p}+k-p)\bar{\chi}_{n;\omega_2;b;s_3}(0)|0\rangle \langle B|\bar{\mathcal{H}}^n_{v;a;s_1}(0)\delta(\hat{p}_r-k)\mathcal{H}^{n;s_4}_{v;b}(0)|B\rangle, \tag{3.12}$$

where the color indices a, b and the Dirac bispinor indices s_1, s_2, s_3, s_4 are shown explicitly. Here $\bar{\Gamma}_i^\mu := \gamma^0\Gamma_i^\mu\gamma^0$ is the Dirac adjoint of the matrix Γ_i^μ.

Because both $|B\rangle$ and $|0\rangle$ states are invariant under SU(3) color transformations, both the collinear and soft matrix elements must be color singlets, i.e. proportional to δ_{ab}.

In the SCET region $p^+ \ll p^-$, therefore p^- is the large label component of momentum and p^+ is the small residual momentum. The δ-function on the right side of eq. (3.11) sets the collinear label momentum equal to the label component of the momentum p, which is equal to $np^-/2$. This fact fixes the label ω on the collinear fields $\chi_{n;\omega}$, which can be shown using eq. (2.15):

$$\begin{aligned}\langle 0|\chi_{n;\omega_1}(0)\delta(\hat{p}+k-p)\bar{\chi}_{n;\omega_2}(0)|0\rangle &= \int \mathrm{d}k_1\mathrm{d}k_2 \langle 0|\delta(\bar{n}\cdot\hat{p}_l-k_1+\omega_1)\chi_n(0) \\ &\cdot \delta(\bar{n}\cdot\hat{p}_l-k_1)\delta(\hat{p}+k-p)\delta(\bar{n}\cdot\hat{p}_l-k_2)\bar{\chi}_n(0)\delta(\bar{n}\cdot\hat{p}_l-k_2+\omega_2)|0\rangle \\ &= \int \mathrm{d}k_1\mathrm{d}k_2 \langle 0|\delta(\omega_1-k_1)\chi_n(0)\delta(p^- -k_1)\delta(\hat{p}+k-p)\delta(p^- -k_2)\bar{\chi}_n(0)\delta(\omega_2-k_2)|0\rangle \\ &= \delta(p^- -\omega_1)\delta(p^- -\omega_2)\langle 0|\chi_n(0)\delta(\hat{p}+k-p)\bar{\chi}_n(0)|0\rangle. \end{aligned} \tag{3.13}$$

Here the momentum k is residual, and the residual component of the momentum p is $p^+\bar{n}/2$. However, the collinear matrix element can depend only on the k^+-component of the residual momentum k [17]. This is related to the fact that the all derivatives in the n-collinear SCET Lagrangian have the form $n\cdot\partial$, and therefore the residual momentum k enters the collinear Feynman rules only as $k^+ = n\cdot k$. Therefore, the collinear matrix element can be rewritten as:

$$\langle 0|\chi_n^{s_2}(0)\delta(\hat{p}+k-p)\bar{\chi}_{n;s_3}(0)|0\rangle = \langle 0|\chi_n^{s_2}(0)\delta(\hat{p}+k^+\bar{n}/2-p)\bar{\chi}_{n;s_3}(0)|0\rangle, \tag{3.14}$$

as a result, the integration over the residual momentum k reduces to integration over just its k^+-component:

$$\int \mathrm{d}^4k \delta(\hat{p}+k^+\bar{n}/2-p)_n \otimes \delta(\hat{p}_r-k)_{\text{soft}} = \int \mathrm{d}\omega \delta(\hat{p}+\omega\bar{n}/2-p)_n \otimes \delta(n\cdot\hat{p}_r-\omega)_{\text{soft}}. \tag{3.15}$$

After these simplifications, the hadronic tensor is:

$$W^{\mu\nu} = C^{i*}(p^-)C^j(p^-)(\bar{\Gamma}_i^\mu)^{s_1}{}_{s_2}(\Gamma_j^\nu)^{s_3}{}_{s_4} \int \mathrm{d}\omega \frac{1}{N_c} \langle 0|\chi_n^{s_2}(0)\delta(\hat{p}+\omega\bar{n}/2-p)\bar{\chi}_{n;s_3}(0)|0\rangle \\ \cdot \langle B|\bar{\mathcal{H}}^n_{v;s_1}(0)\delta(n\cdot\hat{p}_r-\omega)\mathcal{H}_v^{n;s_4}(0)|B\rangle. \tag{3.16}$$

The spinor structures can be simplified as follows: Because $p = p^+\bar{n}/2 + p^-n/2$, the momentum $k^+\bar{n}/2 - p$ is invariant under rotations around the $n - \bar{n}$ axis, i.e. under rotations that preserve both vectors n and $\bar{n}$. The vacuum state $|0\rangle$ is of course invariant under rotations, and therefore the collinear matrix element in eq. (3.14) must be invariant under such rotations. As a spinor matrix with indices s_2, s_3 it can therefore be written as a linear combination of matrices $1, \not{n}, \not{\bar{n}}, \not{n}\not{\bar{n}}$. Because $\not{n}\chi_n = 0$ and $\bar{\chi}_n\not{n} = 0$, the collinear matrix element must be proportional to $\not{\bar{n}}$:

$$\langle 0|\chi_n(0)\delta(\hat{p} + k^+\bar{n}/2 - p)\bar{\chi}_n(0)|0\rangle = \frac{\not{n}}{8}\langle 0|\,\mathrm{tr}[\not{\bar{n}}\chi_n(0)\delta(\hat{p} + k^+\bar{n}/2 - p)\bar{\chi}_n(0)]|0\rangle, \quad (3.17)$$

where the trace tr is only over spinor indices.

The soft matrix element is also invariant under rotations that preserve vectors n and $\bar{n}$, and therefore also $v = (n + \bar{n})/2$. Taking into account the identities $(\not{v} - 1)\mathcal{H}_v^n$ and $\bar{\mathcal{H}}_v^n(\not{v} - 1)$, it is straightforward to derive the following:

$$\langle B|\bar{\mathcal{H}}_{v;s_1}^n(0)\delta(n\cdot\hat{p}_r - \omega)\mathcal{H}_v^{n;s_4}(0)|B\rangle = \frac{(1+\not{v})^{s_4}{}_{s_1}}{4}\langle B|\bar{\mathcal{H}}_v^n(0)\delta(n\cdot\hat{p}_r - \omega)\mathcal{H}_v^n(0)|B\rangle. \quad (3.18)$$

The following matrix element is called the quark jet function J, or simply the jet function:

$$J(p^2) \coloneqq \frac{(2\pi)^3}{N_c p^-}\langle 0|\,\mathrm{tr}\left[\frac{\not{\bar{n}}}{4}\chi_n(0)\delta(\hat{p} - p)\bar{\chi}_n(0)\right]|0\rangle, \quad (3.19)$$

where $p = (\bar{n}p^+ + np^-)/2$, $p^2 = p^+p^-$, and the trace tr is over both spinor and color indices. The jet function describes the collinear emissions in the final-state jet. Because a single collinear sector of SCET is in some sense equivalent to a boosted version of full QCD, this collinear matrix element can also be written as a QCD matrix element [19]:

$$\begin{aligned} J(p^2) &= \frac{(2\pi)^3}{N_c p^-}\langle 0|\,\mathrm{tr}\left[\frac{\not{\bar{n}}}{4}W_n^\dagger(0)\psi(0)\delta(\hat{p} - p)\bar{\psi}W_n(0)\right]|0\rangle \\ &= \frac{1}{\pi N_c p^-}\,\mathrm{Im}\, i\langle 0|\,\mathrm{tr}\left[\frac{\not{\bar{n}}}{4}W_n^\dagger(0)\psi(0)\delta(\hat{p} - p)\bar{\psi}W_n(0)\right]|0\rangle, \end{aligned} \quad (3.20)$$

where W_n is the n-collinear Wilson line, and ψ is the quark field operator in full QCD. The quark jet function J has been calculated to 3-loop order [19–22]. Its perturbative series and the RGE equation are presented in appendix E.2.

The soft function S is defined as the following matrix element [23]:

$$\begin{aligned} S(\omega) &\coloneqq \langle B|\bar{\mathcal{H}}_v^n(0)\delta(n\cdot\hat{p}_r - \omega + \delta)\mathcal{H}_v^n(0)|B\rangle \\ &= \langle B|\bar{\mathcal{H}}_v^n(0)\delta(in\cdot\partial + \omega - \delta)\mathcal{H}_v^n(0)|B\rangle \\ &= \langle B|\bar{b}_v(0)\delta(in\cdot D + \omega - \delta)b_v(0)|B\rangle, \end{aligned} \quad (3.21)$$

where $\delta \coloneqq m_B - m_b$. Here in the first equality the residual momentum operator was rewritten as $n\cdot\hat{p}_r = -in\cdot\partial$, because $[\hat{p}_r, b_v(x)] = -i\partial b_v(x)$. In the second equality the identities $\mathcal{H}_v^n = Y_n^\dagger b_v$ and $Y_n^\dagger(x)f(n\cdot i\partial)Y_n(x) = f(n\cdot iD)$ were used. The $Y_n(x)$ is the

soft Wilson line, defined in eq. (2.22), and $iD = i\partial + gA_{\text{soft}}$ is the covariant derivative. Note that the B-meson state $|B\rangle$ in eq. (3.21) is defined in full QCD, not in HQET [23].

Because the soft function is defined as a matrix element between B-meson states, it is fundamentally nonperturbative. The soft function encodes both the emission of wide-angle soft radiation and effects related to the structure of the B meson, such as Fermi motion of the decaying b-quark in the B-meson.

In terms of the jet and the soft function, the hadronic tensor $W^{\mu\nu}$ can be written as:

$$W^{\mu\nu} = (2\pi)^{-3} C^{i*}(p^-) C^j(p^-) \,\mathrm{tr}\left[\frac{1+\not{v}}{2}\bar{\Gamma}_i^\mu \frac{\not{n}}{4}\Gamma_j^\nu\right] \int \mathrm{d}\omega S(\omega) p^- J(p^-(p_X^+ - \omega)), \qquad (3.22)$$

where $p_X^+ = p^+ + \delta = p^+ + m_B - m_b = m_B - q^+$. Both the soft function $S(\omega)$ and the jet function $p^- J(p^-\omega)$ have support $\omega \geq 0$. The product $C^{i*}(p^-)C^j(p^-)$ of the Wilson coefficients is usually called the hard function. Once loop corrections are included, the hard, jet, and soft functions $C^{i*}(p^-)C^j(p^-)$, J, S depend on the renormalization scale μ, but their scale dependence cancels in the product in eq. (3.22) order-by-order, such that the hadronic tensor $W^{\mu\nu}$ is scale-independent, up to higher-order perturbative corrections.

3.1. Shape function

The soft function cannot be calculated from first principles, but it is possible to study perturbative corrections to the soft function, its renormalization and dependence on the renormalization scale μ. For this reason, it is convenient to further factorize the soft function S into the so-called *shape function* F and the *partonic* soft function C_0 [23]:

$$S(\omega,\mu) = (C_0 \otimes F)(\omega) = \int \mathrm{d}k C_0(\omega - k, \mu) F(k). \tag{3.23}$$

The convolution symbol $\otimes$ is defined in appendix B. Here both $C_0(k)$ and $F(k)$ have support $k \geq 0$ [23]. In order to distinguish the partonic soft function C_0 from the soft function S the latter will sometimes be referred to as the *hadronic* soft function.

It should be noted that the factorization in eq. (3.23) is not a rigorous result of SCET, but just an ansatz, a convenient way to parametrize the nonperturbative function S.

The partonic soft function C_0 is defined as a matrix element similar to the soft function in eq. (3.21), but with a b-quark state instead of a B-meson state:

$$C_0(\omega,\mu) \coloneqq \langle b|\bar{b}_v(0)\delta(in\cdot D - \omega)b_v(0)|b\rangle. \tag{3.24}$$

The partonic soft function can be calculated perturbatively, and is currently known up to 3-loop order [24, 25]. Its perturbative series and the RGE equation are given in appendix E.3. All the nonperturbative dynamics of the soft function S are encoded in the shape function F. An additional advantage of the factorization in eq. (3.23) is that all dependence on the renormalization scale μ is in the perturbative contribution C_0, while the shape function F is defined to be independent of the renormalization scale μ and the perturbative order.

In the local-OPE region, where $\omega \gg \Lambda_{\mathrm{QCD}}$, the right-hand side of eq. (3.23) can be expanded in powers of $\Lambda_{\mathrm{QCD}}/\omega$:

$$S(\omega,\mu) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \frac{\partial^n C_0(\omega,\mu)}{\partial\omega^n} M_n, \tag{3.25}$$

where $M_n \sim \Lambda_{\mathrm{QCD}}^n$ are the moments

$$M_n \coloneqq \int \mathrm{d}k k^n F(k). \tag{3.26}$$

As a result, the leading nonperturbative corrections in the local-OPE region are encoded in the first few moments of the shape function F. On the other hand, in the SCET region, where $\omega \sim \Lambda_{\mathrm{QCD}}$, the exact shape of the shape function F is relevant, and the moments expansion is not applicable.

As discussed in ref. [23], an important feature of the factorization in eq. (3.23) is that it provides a common description of the nonperturbative effects across these different kinematic regions, incorporating all available perturbative information in the limit $\omega \gg \Lambda_{\mathrm{QCD}}$ without having to explicitly carry out an expansion in $\Lambda_{\mathrm{QCD}}/\omega$, whose precise

region of validity would be unclear. In other words, all perturbative corrections to the moments of the soft function $S(\omega, \mu)$ are encoded in the perturbative function C_0, while the shape function F is a genuinely nonperturbative object, which needs to be determined experimentally.

The first few moments of the shape function F are given in terms of HQET hadronic parameters:

$$M_0 = 1, \qquad M_1 = \delta, \qquad M_2 = \delta^2 - \frac{\lambda_1}{3}, \qquad M_3 = \delta^3 - \lambda_1 \delta + \frac{\rho_1}{3}, \tag{3.27}$$

where $\delta \coloneqq m_B - m_b$ and the HQET hadronic parameters λ_1 and ρ_1 are usually defined as [26]:

$$\lambda_1 = \langle B|\bar{b}_v (iD)^2 b_v|B\rangle, \qquad \rho_1 = \langle B|\bar{b}_v (iD_\mu)(v\cdot iD)(iD^\mu) b_v|B\rangle. \tag{3.28}$$

However, the b-quark mass m_b and the hadronic parameters λ_1 and ρ_1 defined above are not entirely physical, and can suffer from a renormalon ambiguity of order $\sim \Lambda_{\text{QCD}}^n$, which will be discussed in much more detail in chapter 4. The shape function F and the parameters m_b, λ_1, ρ_1 can instead be defined in a short-distance scheme, in which case the definitions of λ_1 and ρ_1 include an additional correction in comparison to eq. (3.28). The moments of the shape function F in a short-distance scheme are still given by eq. (3.27), except that the parameters δ, λ_1, ρ_1 are defined in a short-distance scheme.

The shape function cannot be calculated from first principles, but it can be extracted from experimental measurements of $B \to X_s\gamma$ photon energy spectrum. Such an extraction has been carried out in ref. [7]. However, the shape function determined in ref. [7] is not exactly the leading-power shape function F, but a combination of the leading shape function F and some subleading shape functions that enter the $B \to X_s\gamma$ spectrum. Following ref. [7] this combination of leading and subleading shape function will be denoted with a calligraphic symbol $\mathcal{F}$. The relative difference between F and $\mathcal{F}$ is a power correction of order Λ_{QCD}/m_b, and formally can be neglected in leading-power SCET. Although the impact of subleading corrections has not been investigated in this work, the impact of these subleading shape functions is known at leading order in α_s [27], and the modified shape function $\mathcal{F}$ effectively takes it into account. The moments of the modified shape function $\mathcal{F}$ include additional power-suppressed corrections in comparison to the moments shown in eq. (3.27). These corrections are given in appendix D in ref. [7].

A general method for parametrizing $\mathcal{F}(k)$ via a systematic expansion around a given base model has been developed in ref. [23]. This method was used for example in ref. [7]. Since this work is focused on the perturbative corrections, the precise form of $\mathcal{F}(k)$ is not relevant. To illustrate and compare the perturbative corrections, the shape function model only needs to be reasonably realistic. The following exponential base model used in refs. [7, 23] will be used for this purpose:

$$\mathcal{F}(k) = \frac{1}{\lambda} Y\Big(\frac{k}{\lambda}\Big) \qquad \text{with} \qquad Y(x) = \frac{(p+1)^{p+1} x^p}{p!} e^{-(p+1)x}. \tag{3.29}$$

Its normalization and first moment are

$$M_0 = \int_0^\infty \mathrm{d}k \mathcal{F}(k) = 1, \qquad M_1 = \int_0^\infty \mathrm{d}k k \mathcal{F}(k) = \lambda. \tag{3.30}$$

With $\lambda \approx m_B - m_b \approx 0.6\,\mathrm{GeV}$ this basic model already provides a reasonably good description of the experimental measurements [7].

Note that evaluating the NNLO soft function in a short-distance scheme involves taking n derivatives of $\mathcal{F}(k)$. Therefore, at $\mathrm{N^3LL'}$, which needs the $\mathrm{N^3LO}$ soft function, the parameter p is required to satisfy $p \geq 4$ to ensure that the soft function vanishes for $\omega \to 0$, which in turn is required for the $B \to X_s\gamma$ photon energy spectrum to vanish at the kinematic endpoint $E_\gamma \to m_B/2$.

4. The renormalon problem

A major challenge in B physics is to parametrize nonperturbative effects in such a way that their extraction from experimental measurements is stable with, and ideally independent of, the perturbative order. Achieving this stability is not trivial due to infrared sensitivity of the involved perturbative series and the resulting ambiguity in the asymptotic series of perturbative QCD, which is known as the renormalon problem. The renormalon problem manifests itself in practice as poor convergence of the perturbative series even at low orders. Since physical, measurable quantities are independent of the perturbative order, the large perturbative corrections at each order are compensated by corresponding large changes in the extracted value of some associated nonperturbative parameter. In other words, the renormalon ambiguity in the perturbative series is compensated order-by-order by an equal and opposite renormalon ambiguity in the nonperturbative parameter.

Conceptually, to resolve this issue, the renormalon must be identified and subtracted from both the perturbative quantity (C) and the associated parameter (p), such that both become renormalon-free and perturbatively stable. To give a simple toy example:

$$C - p = (C - \delta p) - (p - \delta p) =: \widehat{C} - \widehat{p}. \tag{4.1}$$

On the left-hand side, the renormalon only cancels between C and p. On the right-hand side, the so-called residual term δp is a perturbative series in α_s that contains the renormalon. Its specific choice defines a specific so-called short-distance scheme. The renormalon then cancels within each of the parenthesis defining the short-distance objects $\widehat{C}$ and $\widehat{p}$, which are now separately free of the renormalon.

4.1. Soft function

In reality, the structure is of course more complicated than the above simple toy example. In the case of B-meson decays, the leading-power perturbative series that suffers from renormalon ambiguities is that of the partonic soft function $C_0(\omega, \mu)$, whose renormalons are cancelled by the nonperturbative object $\mathcal{F}(k)$. Its leading renormalon ambiguity of order $\sim \Lambda_{\rm QCD}$ is due to the pole mass definition of the b-quark mass $m_b^{\rm pole}$, which explicitly enters the definition of the soft function $S(\omega, \mu)$ in eq. (3.21) and henceforth shows up in all the moments of $\mathcal{F}(k)$. At subleading power, the jet and hard functions involve the pole-mass renormalon as well, which will be discussed in section 5.3. Furthermore, the hadronic parameter λ_1, which first appears in the second moment of $\mathcal{F}(k)$, has a subleading $\mathcal{O}(\Lambda_{\rm QCD}^2)$ renormalon ambiguity [28, 29]. Similarly, the hadronic parameter ρ_1, which first appears in the third moment, is expected to have an $\mathcal{O}(\Lambda_{\rm QCD}^3)$ renormalon.

The parameters in a generic short-distance scheme are defined as:

$$\widehat{m}_b = m_b^{\text{pole}} - \delta m_b, \qquad \widehat{\lambda}_1 = \lambda_1 - \delta\lambda_1, \qquad \widehat{\rho}_1 = \rho_1 - \delta\rho_1. \tag{4.2}$$

The subtraction terms δm_b, $\delta\lambda_1$, and $\delta\rho_1$ are defined to cancel the renormalon in their respective parameter, such that the short-distance parameters on the left-hand side of eq. (4.2) are renormalon-free. Borrowing the language from the pole-mass scheme, the HQET hadronic parameters λ_1 and ρ_1 defined in eq. (3.28) are referred sometimes to as "pole scheme" parameters.

The construction of the short-distance partonic soft function $\widehat{C}_0(\omega,\mu)$ with the appropriate renormalon subtractions is derived in ref. [23]. Up to N^{3}LO:

$$\begin{aligned}\widehat{C}_0(\omega,\mu) &= \left[1 - \frac{\delta\lambda_1}{6}\frac{\mathrm{d}^2}{\mathrm{d}\omega^2} - \frac{\delta\rho_1}{18}\frac{\mathrm{d}^3}{\mathrm{d}\omega^3} + \cdots\right] e^{\delta m_b \frac{\mathrm{d}}{\mathrm{d}\omega}} C_0(\omega,\mu) \\ &= \left[1 + \delta m_b \frac{\mathrm{d}}{\mathrm{d}\omega} + \frac{1}{2}\left(\delta m_b^2 - \frac{\delta\lambda_1}{3}\right)\frac{\mathrm{d}^2}{\mathrm{d}\omega^2}\right. \\ &\quad \left. + \frac{1}{6}\left(\delta m_b^3 - \delta m_b\delta\lambda_1 - \frac{\delta\rho_1}{3}\right)\frac{\mathrm{d}^3}{\mathrm{d}\omega^3} + \cdots\right] C_0(\omega,\mu).\end{aligned} \tag{4.3}$$

The original, pole-scheme partonic soft function $C_0(\omega,\mu)$ is given in appendix E.3. Importantly, for the renormalons to cancel on the right-hand side, it must always be fully expanded to a given fixed order in α_s, including the δm_b, $\delta\lambda_1$, $\delta\rho_1$ series, their products with each other and with C_0. With these subtractions both $\widehat{C}_0(\omega,\mu)$ and $\mathcal{F}(k)$ become renormalon-free, up to yet higher-order renormalons. In particular, the moments of $\mathcal{F}(k)$ are then given by the short-distance parameters $\widehat{m}_b$, $\widehat{\lambda}_1$, $\widehat{\rho}_1$, as shown in eq. (3.27). In this case the shape function $\mathcal{F}(k)$ is also defined in a short-distance scheme. To evaluate the convolution integral $\widehat{C}_0 \otimes \mathcal{F}$, integration by parts is used to move all the derivatives in the eq. (4.3) to act on the shape function $\mathcal{F}$ [23]. The renormalon subtractions significantly improve the perturbative convergence of the soft function compared to the pole scheme, which will be demonstrated numerically in section 5.6.3.

In general, the subtraction terms $\delta m_b, \delta\lambda_1, \delta\rho_1$ depend on the scale, leading to a similar scale dependence of the short-distance parameter, which is not explicitly shown. The scale dependence can be explicit, as for example for the $\overline{\text{MS}}$ mass, in which case it is usually governed by an associated RGE equation. The scale dependence can also be only internal, e.g. in the 1S or the MSR mass schemes, in which case δm_b (and also $\widehat{m}_b$) is formally scale-independent (with only the usual scale dependence from truncating the perturbative series that is cancelled by higher orders). In either case, the residual terms must be expanded at the same scale μ, i.e. in terms of the same $\alpha_s(\mu)$, that is used for the perturbative series whose renormalon is supposed to be subtracted to ensure that the renormalon actually cancels. In the case of partonic soft function $\widehat{C}_0(\omega,\mu_S)$ the relevant scale is the soft scale μ_S.

The HQET power counting for the residual mass term δm_b can be determined from the HQET Lagrangian:

$$\mathcal{L}_{\text{HQET}} = \bar{b}_v(iv\cdot D - \delta m_b)b_v. \tag{4.4}$$

A suitable short-distance scheme for the bottom-quark mass should respect the power counting of residual soft momentum $\delta m_b \sim \alpha_s v \cdot D \sim \alpha_s k \sim \alpha_s \Lambda_{\rm QCD}$ [30]. Note that the perturbative series for δm_b starts at NLO and therefore scales like α_s. Therefore, a low-scale mass scheme, such as the 1S scheme [31–33] or the MSR scheme [34] with $R \sim \Lambda_{\rm QCD}$, are applicable.

4.2. MSR mass scheme

The MSR mass scheme is a short-distance mass scheme designed to subtract the pole-mass renormalon by introducing an infrared cutoff scale R as follows:

$$\delta m_b(R) \coloneqq m_b^{\text{pole}} - m_b^{\text{MSR}}(R) = R \sum_{n=1}^{\infty} a_n^{\text{MSR}} \Big[\frac{\alpha_s(R)}{4\pi}\Big]^n, \tag{4.5}$$

where the coefficients a_n^{MSR} are determined by matching the MSR mass scheme onto the $\overline{\text{MS}}$ mass scheme at $\overline{m}_b(\overline{m}_b)$. The infrared scale R controls the energy cutoff of self-energy diagrams which are absorbed into the mass definition [34, 35]. This work uses the so-called "natural" MSR mass definition [36], where the series coefficients $a_n^{\text{MSR}} = a_n^{\overline{\text{MS}}}(n_l, n_h = 0)$.

The MSR mass is a natural extension of the $\overline{\text{MS}}$ mass for $R \leq \overline{m}_b(\overline{m}_b)$. Because the definition of the MSR mass depends on the infrared scale R, it can be thought of as a continuous family of short-distance mass schemes, parametrized by the parameter R. In the limit where $R \to \overline{m}_b(\overline{m}_b)$ the MSR mass approaches the $\overline{\text{MS}}$ mass. In the opposite limit $R \to 0$, the MSR mass formally approaches the pole mass. However, when taking this limit one encounters the Landau pole of the coupling constant. This issue is related to the pole-mass renormalon and cannot be addressed unambiguously.

Values of the MSR mass at different R-scales are related by the so-called R-evolution equation [34], whose solution resums logarithms $\ln(R_1/R_0)$ in the perturbative correction between $m_b^{\text{MSR}}(R_1)$ and $m_b^{\text{MSR}}(R_0)$. The R-evolution can be used to obtain the MSR mass value at a low R from the $\overline{\text{MS}}$ mass and vice versa.

In contrast to the 1S scheme, the infrared scale R of the MSR scheme is an external parameter. At the infrared scale R the mass correction δm_b^{MSR} is of order $\delta m_b^{\text{MSR}} \sim R\alpha_s$. Therefore, the appropriate value of R-scale for the renormalon subtraction in the soft function is close to the soft scale, and is chosen to be $R = 1\,\text{GeV}$. It will be shown in section 5.6.3 that this ensures a proper cancellation of the renormalon in the soft function. In principle, it is possible to pick a different value for the R scale in different kinematic regions by taking the R-scale to depend on the phasespace point [37], as long as the R-evolution is consistently used to relate the shape functions at different R-scales. This can be used to enforce $R \sim \mu_S$ over the whole spectrum and to eliminate logarithms $\ln(\mu_S/R)$ in the series of $\delta m_b(R)$. Ref. [38] implements this R-evolution setup for an analogous soft function for the thrust distribution in jet production. In the case of $B \to X_s\gamma$, such a setup seems to not lead to a significant improvement in convergence, so for simplicity in a fixed value of R will be used in the following.

4.3. HQET parameters λ_1 and ρ_1

The short-distance definitions for the HQET hadronic parameters λ_1 and ρ_1 will be defined analogously to the "invisible" scheme for λ_1 [23]. The main feature of the "invisible" scheme is that the subtraction series $\delta\lambda_1^{\text{inv}}$ starts at order $\mathcal{O}(\alpha_s^2)$, rather than $\mathcal{O}(\alpha_s)$, which is the case for example in the kinetic scheme [39, 40]. Refs. [28, 29] have

shown that the absence of the $\mathcal{O}(\alpha_s)$ correction is a consequence of Lorentz invariance, and the presence of an $\mathcal{O}(\alpha_s)$ correction in the kinetic scheme is an artifact of the used non-Lorentz-invariant UV regulator. Furthermore, ref. [23] has presented numerical evidence that using the kinetic scheme for λ_1 leads to an over-subtraction in the soft function at $\mathcal{O}(\alpha_s)$.

Renormalon ambiguities in the soft function are studied by examining the singularities of the Borel transform of its perturbative series. It is well-known that the order $\mathcal{O}(\Lambda_{\rm QCD})$ and $\mathcal{O}(\Lambda_{\rm QCD}^2)$ ambiguities associated with the pole mass $m_b^{\rm pole}$ and the "pole" λ_1 correspond to singularities at $u = 1/2$ [41] and $u = 1$ [29] in the Borel plane. Although the existence of the $u = 3/2$ renormalon has not been established, the HQET parameter ρ_1 is conjectured to have a renormalon ambiguity of order $\mathcal{O}(\Lambda_{\rm QCD}^3)$. The renormalon subtractions δm_b and $\delta\lambda_1$ start at order $\mathcal{O}(\alpha_s)$ and $\mathcal{O}(\alpha_s^2)$, respectively. Expecting the corrections at each derivative in eq. (4.3) to be of the same order, it seems natural to define a short-distance scheme for the parameter ρ_1 such that its renormalon subtraction $\delta\rho_1$ starts at order $\mathcal{O}(\alpha_s^3)$:

$$\begin{aligned}
\delta\lambda_1(R_\lambda,\mu) &= R_\lambda^2\frac{\alpha_s^2(\mu)}{\pi^2}\left[\delta\lambda_1^{(2)} + \frac{\alpha_s(\mu)}{\pi}\Big(\delta\lambda_1^{(3)} + \delta\lambda_1^{(2)}\beta_0\ln\frac{\mu}{R_\lambda}\Big)\right] + \mathcal{O}(\alpha_s^4),\\
\delta\rho_1(R_\rho,\mu) &= R_\rho^3\frac{\alpha_s^3(\mu)}{\pi^3}\delta\rho_1^{(3)} + \mathcal{O}(\alpha_s^4).
\end{aligned} \tag{4.6}$$

The infrared scales R_λ, R_ρ are chosen to be 1 GeV. Note that the hadronic parameters λ_1 and ρ_1 have mass dimension 2 and 3, respectively, and therefore $\delta\lambda_1 \propto R_\lambda^2$ and $\delta\rho_1 \propto R_\rho^3$. The value of the coefficient $\delta\lambda_1^{(2)}$ is chosen as in ref. [23], according to the invisible scheme — $\delta\lambda_1^{(2)} = \pi^2/3 - 1$. The value of the coefficient $\delta\lambda_1^{(3)}$ in the invisible scheme, is, however, not known.

Usually, one defines a short-distance scheme to all orders by exploiting the perturbative series of some physical and thus renormalon-free quantity. However, this is not strictly necessary, since after all the main goal of the renormalon subtractions is to obtain a stable perturbative result. Thus, in this work the "invisible" scheme for $\widehat{\lambda}_1$ and $\widehat{\rho}_1$ at $\mathcal{O}(\alpha_s^3)$ is defined by choosing numerical values for $\delta\lambda_1^{(3)}$ and $\delta\rho_1^{(3)}$ such that the resulting soft function at different perturbative orders manifests good convergence, i.e. that the size of scale variations reduces when including higher-order corrections, that the resulting uncertainty bands at different orders have reasonable overlaps, that the peak position for the soft function remains stable, and finally that it remains positive at small k and approaches zero with similar slopes at different orders. Following this procedure, a satisfactory convergence for the soft function is found, see figure 5.12, by taking

$$\delta\lambda_1^{(3)} = 16, \qquad\qquad \delta\rho_1^{(3)} = -3. \tag{4.7}$$

In principle, one could instead consider the perturbative convergence of the $B \to X_s\gamma$ spectrum. However, since it also receives contributions from the jet and hard functions, the dependence on the soft function is washed out in the spectrum. Therefore, it is better to use the convergence of the soft function to define the short-distance scheme.

This is also the most natural choice, since the soft function is the object containing the renormalons to be subtracted. The approach taken here is somewhat similar to the "shape-function" scheme used in ref. [21], where the second and third moments of the perturbative shape function with some, largely arbitrary, hard cutoff is used to define the short-distance parameters. (Similarly, the short-distance b-quark mass is defined based on the first moment.) The disadvantage of the "shape-function" approach is that it yields $\delta\lambda, \delta\rho_1 \sim \alpha_s$, which leads to massive oversubtraction, similar to the kinetic scheme.

5. Inclusive decay $B \to X_s\gamma$ at $\mathrm{N^3LL'}$+$\mathrm{N^3LO}(c_k)$

Following the setup of ref. [7], the $B \to X_s\gamma$ photon energy spectrum can be written as

$$\frac{\mathrm{d}\Gamma}{\mathrm{d}E_\gamma} = 2\Gamma_0 \left(\frac{2E_\gamma}{\widehat{m}_b}\right)^3 \int \mathrm{d}k \widehat{P}(k) \mathcal{F}(m_B - 2E_\gamma - k) + \mathcal{O}\left(\frac{\Lambda_{\mathrm{QCD}}}{\widehat{m}_b}\right). \tag{5.1}$$

As in the previous chapter, symbols with a hat "$\widehat{\ }$" denote objects defined in a short-distance scheme. Here Γ_0 is the total $B \to X_s\gamma$ decay rate at leading order in both α_s and $\Lambda_{\mathrm{QCD}}/\widehat{m}_b$, which is given by

$$\Gamma_0 = \frac{G_F^2 \widehat{m}_b^5}{8\pi^3} \frac{\alpha_{\mathrm{em}}}{4\pi} |V_{tb}V_{ts}^*|^2. \tag{5.2}$$

The overall factor $(2E_\gamma/\widehat{m}_b)^3$ in eq. (5.1) has a purely kinematic origin. It arises from the photon phase space integration and derivative operators in the photon field strength tensor of O_7.

The function $\widehat{P}$ denotes the perturbatively calculable contribution, which is given by [7]:

$$\widehat{P}(k) = \left|C_7^{\mathrm{incl}}\right|^2 \Big[W_{77}^{\mathrm{s}}(k) + W_{77}^{\mathrm{ns}}(k)\Big] + 2\,\mathrm{Re}(C_7^{\mathrm{incl}}) \sum_{i\neq 7} \mathcal{C}_i W_{7i}^{\mathrm{ns}}(k) + \sum_{i,j\neq 7} \mathcal{C}_i \mathcal{C}_j W_{ij}^{\mathrm{ns}}(k). \tag{5.3}$$

The coefficients C_7^{incl}, and $\mathcal{C}_i$ are defined as Wilson coefficients in a special effective theory without the heavy c- and b-quarks. They are determined by matching the Weak Effective Theory at the scale $\sim \widehat{m}_b$, and are therefore defined in terms of the Wilson coefficient C_i of the Weak Effective Theory. In turn, the coefficients C_i can be determined by matching the Standard Model or some extension of it.

At leading order the $B \to X_s\gamma$ decay is governed by the electromagnetic operator O_7 in the Weak Effective Theory:

$$O_7 = \frac{e}{16\pi^2} \overline{m}_b \bar{s} \sigma_{\mu\nu} F^{\mu\nu} P_R b. \tag{5.4}$$

Beyond the leading order other operators contribute as well. However, the contribution of the operator O_7 remains dominant. Consequently, the coefficient C_7^{incl} is dominated by the coefficient C_7 of this operator, but C_7^{incl} is defined in such a way that it also includes contributions of all non-C_7 operators that give rise to C_7-like contributions [7, 42], hence

the notation. In particular, C_7^{incl} is the coefficient of all *singular* contributions, i.e. of the contributions that are dominant around the peak of the spectrum $E_\gamma \sim m_b/2$.

The term $W_{77}^{\mathrm{s}}(k)$ in eq. (5.3) includes all the *singular* contributions, which at a fixed order in α_s are proportional to $\delta(k)$ and $\ln^n(k/\widehat{m}_b)/k$. These are the contributions that are described by leading-power SCET, and consequently have a well-understood factorization and resummation structure [17, 43].

Note that, in agreement with ref. [7], when considering these contributions in SCET the overall kinematic factor $(2E_\gamma/\widehat{m}_b)^3$ is kept exactly, and is not expanded in small $k/\widehat{m}_b = (\widehat{m}_b - 2E_\gamma)/\widehat{m}_b$.

The terms $W_{ij}^{\mathrm{ns}}(k)$ in eq. (5.3) contain the *nonsingular* contributions. In the peak region, where $k/\widehat{m}_b$ is small, they are suppressed by at least a factor $k/\widehat{m}_b$ relative to the singular terms. On the other hand, away from the peak region the singular and nonsingular contributions are of similar magnitude. The perturbative series of the nonsingular contributions W_{ij}^{ns} starts at order $\mathcal{O}(\alpha_s)$.

Because the 77-like contributions W_{77}^{s} and W_{77}^{ns} are dominant in the experimentally interesting peak region the rest of this chapter will be focused on them. Because the remaining non-77-like contributions are purely nonsingular, it would be straightforward to include them as a small correction, as was for example done in ref. [7].

The remaining terms $+\mathcal{O}(\Lambda_{\mathrm{QCD}}/\widehat{m}_b)$ in eq. (5.1) denote the subdominant resolved and unresolved contributions [7], which are not included in this work.

To sum up, the perturbative contributions $\widehat{P}$ are separated into singular and nonsingular contributions. The singular contributions can be derived, factorized and systematically resummed using SCET, while the nonsingular contributions do not appear in leading-power SCET. On the other hand, both singular and nonsingular contributions can be derived using full QCD and local operator-product expansion (OPE). The nonsingular contributions are defined in such a way, that together with the singular contributions they reproduce the results of full-QCD-plus-local-OPE order-by-order in α_s.

Resummation of the singular contribution is expected to improve the accuracy of the perturbative results in the peak region. However, away from the peak region the resummation is not appropriate, and fixed-order results must be used instead. To smoothly transition between resummation and fixed-order regimes, the renormalization scales μ in the perturbative series are taken to be functions of the photon energy E_γ, which are called the *profile functions*. These functions will be discussed in section 5.4.

The nonperturbative function $\mathcal{F}(k)$ in eq. (5.1) contains both the leading shape function and a combination of subleading shape functions that appear at tree level in $B \to X_s\gamma$. The shape function is discussed in section 3.1. The perturbative contributions $\widehat{P}$ and the shape function $\mathcal{F}$ are factorized in eq. (5.1). This factorization enables a coherent description of the spectrum in both peak and tail regions. In the tail region only the first few moments of $\mathcal{F}$ are relevant, while in the peak region its full form is important.

5.1. Singular contributions

The singular contributions $W_{77}^{\mathrm{s}}(k)$ are the leading contributions to the spectrum in the limit $k \ll m_b$. Their well-known factorization theorem [17, 43] enables their systematic resummation to all orders in perturbation theory. Following ref. [23], the SCET factorization theorem for the singular contributions is:

$$W_{77}^{\mathrm{s}}(k) = \hat{h}_s(\widehat{m}_b, \mu_H)\widehat{U}_H(\widehat{m}_b, \mu_H, \mu_J)\times \\ \times \int \mathrm{d}\omega \mathrm{d}\omega' \widehat{m}_b \hat{J}(\widehat{m}_b(k-\omega), \mu_J)\widehat{U}_S(\omega-\omega', \mu_S, \mu_J)\widehat{C}_0(\omega', \mu_S), \tag{5.5}$$

where $\hat{h}_s$, $\hat{J}$, and $\widehat{C}_0$ are the hard, jet, and partonic soft functions, respectively. The hard and soft evolution kernels, $\widehat{U}_H$ and $\widehat{U}_S$, evolve the hard and soft functions from their characteristic hard and soft scales μ_H and μ_S to the common scale μ, which is chosen to be at the jet scale $\mu = \mu_J$. These evolution kernels sum the logarithms of the form $\ln(\mu_H/\mu_J)$ and $\ln(\mu_J/\mu_S)$. Since all ingredients are evolved to the jet scale, the jet evolution kernel $\widehat{U}_J(p^2, \mu_J, \mu_J) = \delta(p^2)$ drops out. As already mentioned, a hat "$\widehat{\ }$" indicates that an object is defined in a renormalon-free short-distance scheme, Explicit results for the perturbative ingredients are collected in appendix E.

The hard, jet, and partonic soft functions $\hat{h}_s(\mu_H)$, $\hat{J}(\mu_J)$, $\widehat{C}_0(\mu_S)$ in eq. (5.5) are evaluated at fixed order. Normally, the product $\hat{h}_s \times \hat{J} \otimes \widehat{C}_0$ of their series is expanded strictly to the given order in α_s, counting the strong coupling constant at different scales as $\alpha_s(\mu_H) \sim \alpha_s(\mu_J) \sim \alpha_s(\mu_S)$. With this prescription, all higher-order cross terms in the product are dropped. In this case, the strict fixed-order expansion of W_{77}^{s} in terms of a common $\alpha_s(\mu)$ can be reproduced simply by setting all renormalization scales to be equal $\mu_H = \mu_J = \mu_S = \mu$. This is different to refs. [7, 23], where only the product $\hat{J} \otimes \widehat{C}_0$ is expanded in α_s, while the hard function is kept unexpanded as an overall multiplicative factor, which results in keeping certain higher-order cross terms in the fixed-order limit. The impact of the different treatments of these higher-order cross terms is studied in section 5.6.2.

To resum the singular corrections using eq. (5.5) to $\mathrm{N^3LL'}$ order, the fixed-order boundary conditions $\hat{h}_s(\mu_H)$, $\hat{J}(\mu_J)$, $\widehat{C}_0(\mu_S)$ need to be included to $\mathcal{O}(\alpha_s^3)$, and the evolution factors $\widehat{U}_H(\mu_H, \mu_J)$ and $\widehat{U}_S(\mu_J, \mu_S)$ need to account for the 3-loop noncusp and 4-loop cusp anomalous dimensions, as well as the 4-loop QCD β-function. The relation between orders of different fixed-order and resummation ingredients is explained in more detail in appendix D. The jet and soft functions have been computed up to three loops in refs. [19, 20, 22] and refs. [20, 24, 25], respectively.

Regarding the hard function, its 3-loop anomalous dimension is also known via the consistency relation with the jet and soft anomalous dimensions, it can be found in ref. [25]. The full hard function is currently only known up to NNLO [18, 23]. Using the RGE equation of the hard function, at $\mathrm{N^3LO}$ all the logarithmic terms in the hard function can be expressed in terms of the known anomalous dimensions and lower-order coefficients. The result is presented in appendix E.1. Thus, the only missing ingredient to obtain W_{77}^{s} at full $\mathrm{N^3LL'}$ is the finite, nonlogarithmic 3-loop constant of the hard

function, h_3. This constant appears in the (pole-scheme) hard function series at $\mu = m_b$ as

$$h_s(m_b, \mu_H = m_b) = 1 + \frac{\alpha_s(m_b)}{\pi} h_1 + \frac{\alpha_s^2(m_b)}{\pi^2} h_2 + \frac{\alpha_s^3(m_b)}{\pi^3} h_3 + \mathcal{O}(\alpha_s^4). \tag{5.6}$$

In this work the unknown constant h_3 is treated as a theory nuisance parameter:

$$h_3 = 0 \pm 80, \tag{5.7}$$

where the range of variation is estimated using the Padé approximation

$$h_3 \sim \frac{h_2^2}{|h_1|} \sim \frac{19.3^2}{4.55} \sim 80. \tag{5.8}$$

In section 5.6 the impact of the constant h_3 on the theoretical predictions will be shown to be minor.

The fixed-order series of the singular contribution up to $\mathrm{N^3LO}$ is

$$\begin{aligned} W_{77}^{\mathrm{s}}(\widehat{m}_b x) = \frac{C_F}{\widehat{m}_b} \Bigg\{ & w_{77}^{\mathrm{s}(0)}(x) + \frac{\alpha_s(\mu)}{\pi} \big[w_{77}^{\mathrm{s}(1)}(x) + \Delta w_{77}^{\mathrm{s}(1)}(\mu, x) \big] \\ & + \frac{\alpha_s^2(\mu)}{\pi^2} \left[w_{77}^{\mathrm{s}(2)}(x) + \frac{\beta_0}{2} w_{77}^{\mathrm{s}(1)}(x) \ln \frac{\mu}{\widehat{m}_b} + \Delta w_{77}^{\mathrm{s}(2)}(\mu, x) \right] \\ & + \frac{\alpha_s^3(\mu)}{\pi^3} \left[w_{77}^{\mathrm{s}(3)}(x) + \left(\beta_0 w_{77}^{\mathrm{s}(2)}(x) + \frac{\beta_1}{8} w_{77}^{\mathrm{s}(1)}(x) \right) \ln \frac{\mu}{\widehat{m}_b} \right. \\ & \left. + \frac{\beta_0^2}{4} w_{77}^{\mathrm{s}(1)}(x) \ln^2 \frac{\mu}{\widehat{m}_b} + \Delta w_{77}^{\mathrm{s}(3)}(\mu, x) \right] + \mathcal{O}(\alpha_s^4) \Bigg\}. \end{aligned} \tag{5.9}$$

The dependence of $W_{77}^{\mathrm{s}}(\widehat{m}_b x)$ on the renormalization scale μ cancels order-by-order. Numerically, the generalized functions $w^{\mathrm{s}(i)}$ are:

$$\begin{aligned} w_{77}^{\mathrm{s}(0)}(x) &= 0.75\delta(x), \\ w_{77}^{\mathrm{s}(1)}(x) &= -4.54\delta(x) - 1.75\mathcal{L}_0(x) - 1.00\mathcal{L}_1(x), \\ w_{77}^{\mathrm{s}(2)}(x) &= (-30.5 + 3.01 n_f)\delta(x) + (5.94 + 0.316 n_f)\mathcal{L}_0(x) + (12.4 - 0.181 n_f)\mathcal{L}_1(x) \\ &\quad + (7.63 - 0.250 n_f)\mathcal{L}_2(x) + 0.667\mathcal{L}_3(x), \\ w_{77}^{\mathrm{s}(3)}(x) &= (88.2 + 0.75 h_3 - 0.0269 n_f - 0.0309 n_f^2)\delta(x) \\ &\quad + (138 - 9.16 n_f - 0.00330 n_f^2)\mathcal{L}_0(x) \\ &\quad + (68.7 - 8.75 n_f + 0.121 n_f^2)\mathcal{L}_1(x) + (-25.1 - 0.815 n_f + 0.00694 n_f^2)\mathcal{L}_2(x) \\ &\quad + (-43.0 + 3.16 n_f - 0.0648 n_f^2)\mathcal{L}_3(x) \\ &\quad + (-6.53 + 0.278 n_f)\mathcal{L}_4(x) - 0.222\mathcal{L}_5(x). \end{aligned} \tag{5.10}$$

The logarithmic plus-distributions are defined in appendix C.1. Note that at this order the unknown 3-loop constant h_3 appears only in the $\delta(x)$ coefficient, so $w_{77}^{\mathrm{s}(3)}(x)$ is completely known for $x > 0$.

The correction terms $\Delta w_{77}^{\mathrm{s}(n)}(\mu, x)$ in eq. (5.9) arise from switching to the short-distance b-quark mass $\widehat{m}_b$, and their μ-dependence separately cancels among them order by order. They are discussed in more detail in section 5.3, and their explicit expressions are provided in eq. (5.22).

5.2. Nonsingular contributions

The nonsingular contribution $W_{77}^{\mathrm{ns}}(k)$ is included at fixed order. It is obtained by subtracting the fixed-order singular terms from the full fixed-order result for $\mathrm{d}\Gamma/\mathrm{d}E_\gamma$, which is known up to $\mathcal{O}(\alpha_s^2)$ [44–46]. Its perturbative expansion up to $\mathrm{N^3LO}$ is

$$\begin{aligned} W_{77}^{\mathrm{ns}}(\widehat{m}_b x) = \frac{C_F}{\widehat{m}_b(1-x)^3}\Bigg\{ & \frac{\alpha_s(\mu_{\mathrm{ns}})}{\pi} w_{77}^{\mathrm{ns}(1)}(x) \\ & + \frac{\alpha_s^2(\mu_{\mathrm{ns}})}{\pi^2}\left[w_{77}^{\mathrm{ns}(2)}(x) + \frac{\beta_0}{2} w_{77}^{\mathrm{ns}(1)}(x) \ln\frac{\mu_{\mathrm{ns}}}{\widehat{m}_b} + \Delta w_{77}^{\mathrm{ns}(2)}(\mu_{\mathrm{ns}}, x)\right] \\ & + \frac{\alpha_s^3(\mu_{\mathrm{ns}})}{\pi^3}\left[w_{77}^{\mathrm{ns}(3)}(x) + \left(\beta_0 w_{77}^{\mathrm{ns}(2)}(x) + \frac{\beta_1}{8} w_{77}^{\mathrm{ns}(1)}(x)\right)\ln\frac{\mu_{\mathrm{ns}}}{\widehat{m}_b} \right. \\ & \left. \qquad + \frac{\beta_0^2}{4} w_{77}^{\mathrm{ns}(1)}(x) \ln^2\frac{\mu_{\mathrm{ns}}}{\widehat{m}_b} + \Delta w_{77}^{\mathrm{ns}(3)}(\mu_{\mathrm{ns}}, x)\right]\Bigg\}. \end{aligned} \tag{5.11}$$

The overall $1/(1-x)^3$ factor is included by convention. Explicit expressions for $w_{77}^{\mathrm{ns}(1)}(x)$ and $w_{77}^{\mathrm{ns}(2)}(x)$ are given in eq. (S21) in ref. [7]. The function $w_{77}^{\mathrm{ns}(3)}(x)$ is currently unknown. The $\Delta w_{77}^{\mathrm{ns}(n)}(\mu_{\mathrm{ns}}, x)$ terms arise from switching to the short-distance mass $\widehat{m}_b$ and are given in eq. (5.23).

In the peak region of the spectrum, where $k \ll \widehat{m}_b$, the nonsingular contributions are power-suppressed by at least $k/\widehat{m}_b$ relative to the singular contributions. Therefore, the two can be considered as independent perturbative series, which are treated separately from each other. In particular, it is consistent to include the nonsingular only at fixed order, while the singular contributions are resummed. On the other hand, in the tail region, where $k \sim \widehat{m}_b$, the separation into singular and nonsingular becomes artificial and only the full result given by their sum, $W_{77}^{\mathrm{full}} = W_{77}^{\mathrm{s}} + W_{77}^{\mathrm{ns}}$, is meaningful. This is reflected by the fact that there are typically large numerical cancellations between the singular and nonsingular contributions in the limit $k \to \widehat{m}_b$, which will be shown explicitly in section 5.4. Consequently, the W_{77}^{s} and W_{77}^{ns} must be included using the same perturbative expansion in this limit, i.e. at the same scale and the same perturbative order, to ensure that the cancellations between them can take place and the proper full result is recovered. This has important ramifications. First, since the full and nonsingular results are only known at fixed order, it is essential to turn off the resummation for W_{77}^{s} for $k \sim \widehat{m}_b$ such that it also reduces to its fixed-order result. Second, the $\mathrm{N^nLL'}$ resummation reduces to the fixed $\mathcal{O}(\alpha_s^n)$ singular result, so consistently matching it to fixed order requires including the nonsingular to $\mathcal{O}(\alpha_s^n)$.

Therefore, at $\mathrm{N^3LL'}$ the nonsingular contributions W_{77}^{ns} need to be included at order $\mathcal{O}(\alpha_s^3)$, which means that the unknown nonsingular function $w_{77}^{\mathrm{ns}(3)}(x)$ needs to be

parametrized. In this work, a parametrization of the nonsingular function $w_{77}^{\mathrm{ns}(3)}(x)$ is constructed based on its expected asymptotic behavior in the $x \to 0$ and $x \to 1$ limits. In particular, the singular-nonsingular cancellations in the limit $x \to 1$ at $\mathcal{O}(\alpha_s^3)$ need to be accounted for, which basically implies that $w_{77}^{\mathrm{ns}(3)}$ and $w_{77}^{\mathrm{s}(3)}$ are not independent.

To begin with, the nonsingular function $w_{77}^{\mathrm{ns}(3)}(x)$ can be separated into a "correlated" and an "uncorrelated" piece:

$$w_{77}^{\mathrm{ns}(3)}(x) = w_{\mathrm{cor}}^{\mathrm{ns}(3)}(x) + w_{\mathrm{uncor}}^{\mathrm{ns}(3)}(x). \tag{5.12}$$

The "correlated" term $w_{\mathrm{cor}}^{\mathrm{ns}(3)}(x)$ is designed to completely cancel the singular corrections in the $x \to 1$ limit without disturbing the hierarchy between singular and nonsingular contributions in the $x \to 0$ limit. In this work it is chosen as

$$w_{\mathrm{cor}}^{\mathrm{ns}(3)}(x) \coloneqq -(1-x)^3 w_{77}^{\mathrm{s}(3)}(1). \tag{5.13}$$

Here the 3-loop singular function $w_{77}^{\mathrm{s}(3)}(x)$ is defined in eq. (5.9). The overall factor $(1-x)^3$ simply cancels the overall $1/(1-x)^3$ in eq. (5.11). The remaining "uncorrelated" piece $w_{\mathrm{uncor}}^{\mathrm{ns}(3)}(x)$ can now be considered independent of the singular contribution. It is parametrized as:

$$w_{\mathrm{uncor}}^{\mathrm{ns}(3)}(x) = (1-x)^3 \sum_{k=0}^{5} c_k^{\mathrm{ns}} L^k(x) \qquad \text{with} \qquad L(x) = \frac{1}{4}\frac{w_{77}^{\mathrm{ns}(1)}(x)}{(1-x)^3} - \frac{9}{16}, \tag{5.14}$$

The basic function $L(x)$ of this parametrization is positive for $0 < x < 1$ and has the following asymptotics in the $x \to 0$ and $x \to 1$ limits:

$$L(x) = -\frac{3}{2} - \ln x + \mathcal{O}(x), \qquad\qquad L(x) = 0 + \mathcal{O}(1-x). \tag{5.15}$$

The function $L(x)$ is constructed using $w_{77}^{\mathrm{ns}(1)}(x)$ with the expectation that its powers provide a reasonable guess of the possible shape of the higher-order function $w_{77}^{\mathrm{ns}(3)}(x)$ in the intermediate region $0 < x < 1$. Furthermore, eq. (5.14) incorporates the expectation that at $\mathcal{O}(\alpha_s^3)$ the nonsingular in the limit $x \to 0$ is a degree-5 polynomial in $\ln x$. The five powers of $L(x)$ are included in eq. (5.15) to ensure that the parametrization can probe the full possible logarithmic structure in the small-x limit.

The parameters $c_{0\ldots5}^{\mathrm{ns}}$ are treated as theory nuisance parameters with zero central values and the following variation magnitudes:

$$c_k^{\mathrm{ns}} = 0 \pm \delta c_k^{\mathrm{ns}} \qquad \text{with} \qquad \delta c_{0\ldots5}^{\mathrm{ns}} = (20, 100, 80, 10, 5, 1). \tag{5.16}$$

The range of variations for the last five parameters $c_{1\ldots5}^{\mathrm{ns}}$ is determined based on the observation that in the limit $x \to 0$ the expression $4x w_{77}^{\mathrm{s}(2)}(x)$ provides a good estimate

of the size of logarithmic terms in $w_{77}^{\mathrm{ns}(2)}(x)$ at one and two loops:

$$\begin{aligned}
4xw_{77}^{\mathrm{s}(1)}(x) &= -7.00 - 4.00\ln x,\\
w_{77}^{\mathrm{ns}(1)}(x) &= \quad 3.75 - 4.00\ln x + \mathcal{O}(x),\\
4xw_{77}^{\mathrm{s}(2)}(x) &= 28.8 + 46.7\ln x + 26.5\ln^2 x + 2.67\ln^3 x,\\
w_{77}^{\mathrm{ns}(2)}(x) &= 16.1 + 33.9\ln x + 25.0\ln^2 x + 2.67\ln^3 x + \mathcal{O}(x),\\
4xw_{77}^{\mathrm{s}(3)}(x) &= 406 + 142\ln x - 113\ln^2 x \ - 125\ln^3 x \ - 21.7\ln^4 x \ - 0.889\ln^5 x,\\
&= 259 + \ 95L(x) + 189L^2(x) + 15.5L^3(x) - 15.0L^4(x) + 0.889L^5(x). \quad (5.17)
\end{aligned}$$

In particular, the highest power of $\ln x$ in $w_{77}^{\mathrm{ns}(1,2)}(x)$ is the same as in $4xw_{77}^{\mathrm{s}(1,2)}$, which is why the factor 4 was chosen. The likely explanation for this pattern is that the universal cusp and a set of subleading noncusp anomalous dimensions govern the logarithmic structure at subleading power, similar to the leading-power case. Expecting this pattern to continue at the 3-loop order, the coefficients of $4xw_{77}^{\mathrm{s}(3)}(x)$ are used to estimate the magnitudes $\delta c_{1\ldots5}^{\mathrm{ns}}$ of the nuisance parameters.

The theoretical uncertainty due to the missing 3-loop nonsingular function $w_{77}^{\mathrm{ns}(3)}$ is estimated by varying each nuisance parameter independently in the ranges given above, and adding in quadrature the resulting deviations in predictions. This procedure is discussed in more detail in section 5.5. Of course, this means that the uncertainty necessarily increases by adding more parameters. To make sure the uncertainty estimate is not overly conservative, the variation magnitudes are chosen such that the total uncertainty estimate for the missing 3-loop correction, including the α_s-suppression, does not exceed the size of the 2-loop corrections. To satisfy this constraint, the values for $\delta c_{2,3,4}^{\mathrm{ns}}$ are chosen somewhat smaller than the corresponding coefficients of $4xw_{77}^{\mathrm{s}(3)}(x)$ in eq. (5.17).

It is easy to see from eq. (5.15) that all powers of $L(x)$ approach zero in the far tail of the spectrum. Thus, in this limit the constant term c_0 dominates. For this reason, the size of δc_0^{ns} is estimated by extrapolating the corresponding lower-order corrections using a Padé approximation:

$$w_{x\to1}^{(k)} = \lim_{x\to1}\left[\frac{w_{77}^{\mathrm{ns}(k)}(x)}{(1-x)^3} + w_{77}^{\mathrm{s}(k)}(x)\right], \quad w_{x\to1}^{(3)} = c_0^{\mathrm{ns}} \sim \frac{(w_{x\to1}^{(2)})^2}{w_{x\to1}^{(1)}} \approx \frac{2.92^2}{0.5} \sim 20. \quad (5.18)$$

It should be stressed that the purpose of the parametrization given in eqs. (5.12) and (5.14) is *not* to construct an approximation of the unknown function $w_{77}^{\mathrm{ns}(3)}(x)$. Rather, the goal is first to enable a consistent matching at $\mathrm{N^3LL'}$, which is essentially achieved by the separation in eq. (5.12), and second, to obtain a reliable estimate of the perturbative uncertainty due to the missing 3-loop nonsingular correction $w_{77}^{\mathrm{ns}(3)}(x)$. For this purpose, the estimate only needs to capture the *typical size* of the theory nuisance parameters, up to an $\mathcal{O}(1)$ factor, for which the procedure described above is sufficient. However, since eq. (5.11) includes all known $\mathcal{O}(\alpha_s^3)$ contributions that are predictable from lower

orders, and since the parametrization of the remaining unknown $w_{77}^{\mathrm{ns}(3)}(x)$ does include nontrivial information on its structure, some improvement in the perturbative precision of the predictions beyond $\mathcal{O}(\alpha_s^2)$ is expected. This will be reflected in the size of the resulting uncertainty, as will be shown in section 5.6.

5.3. Subleading δm_b corrections

As discussed in section 4.1, the leading renormalon at leading power comes from the b-quark mass that enters via the argument of the soft function. In addition to the soft function, the b-quark mass also enters the hard and jet functions through the SCET label momentum $p^- \sim m_b$. By default, label momentum conservation sets $p^- = m_b^{\rm pole}$. Formally, choosing a different label $p^- = \widehat{m}_b$ amounts to a power-suppressed effect. For this reason, in ref. [7] the resulting corrections from changing to the $\widehat{m}_b$ scheme could effectively be absorbed into the nonsingular corrections. It will be shown that at $\mathrm{N^3LL'}$ this is no longer viable, and instead both the hard and jet functions will be explicitly switched to a short-distance mass scheme.

To derive the scheme change, consider the partonic function $W(k)$ appearing in eq. (5.3), which can be either the singular, the nonsingular, or the full contribution. It has mass dimension -1 and depends on two dimensionful quantities, k and m_b. Therefore, by dimensional analysis it must have the form

$$W(k) = \frac{1}{m_b} w\Big[\frac{k}{m_b}, \alpha_s(\mu), \ln\frac{\mu}{m_b}\Big] = \frac{1}{m_b} w\Big[\frac{k}{m_b}, \alpha_s(m_b)\Big], \tag{5.19}$$

where all dependence on m_b is made explicit on the right-hand side and $w(x, \alpha_s, L)$ is a scaleless function of its arguments. Since $W(k)$ is defined to be μ-independent, the μ-dependence on the right-hand side of eq. (5.19) is only the internal μ-dependence from $\alpha_s(\mu)$, which cancels order by order. Therefore, the renormalization scale can be set to $\mu = m_b$, which eliminates all logarithms and allows one to track the associated m_b dependence via the dependence on $\alpha_s(m_b)$. After switching the scheme, the μ-dependence can be easily recovered by expanding $\alpha_s(m_b)$ in terms of $\alpha_s(\mu)$.

The partonic rate in the pole scheme is given by eq. (5.19) evaluated at $k = m_b^{\rm pole} - 2E_\gamma$ and $m_b = m_b^{\rm pole}$. Switching to a short-distance scheme amounts to replacing $k \to k + \delta m_b$ and $m_b \to m_b^{\rm pole} = \widehat{m}_b + \delta m_b$ in eq. (5.19) and expanding in δm_b. This gives

$$\begin{aligned}
\widehat{m}_b W(\widehat{m}_b x) &= \frac{1}{1 + \delta m_b/\widehat{m}_b} w\Big(\frac{x + \delta m_b/\widehat{m}_b}{1 + \delta m_b/\widehat{m}_b}, \alpha_s(\widehat{m}_b + \delta m_b)\Big) \\
&= \bigg\{1 + \frac{\delta m_b}{\widehat{m}_b}\frac{\mathrm{d}}{\mathrm{d}x}(1-x) + \frac{1}{2}\frac{\delta m_b^2}{\widehat{m}_b^2}\frac{\mathrm{d}^2}{\mathrm{d}x^2}(1-x)^2 \\
&\quad + \frac{1}{3!}\frac{\delta m_b^3}{\widehat{m}_b^3}\frac{\mathrm{d}^3}{\mathrm{d}x^3}(1-x)^3 + \frac{\delta m_b}{\widehat{m}_b}\beta[\alpha_s(\widehat{m}_b)]\frac{\mathrm{d}}{\mathrm{d}\alpha_s} + \mathcal{O}(\alpha_s^4)\bigg\} w[x, \alpha_s(\widehat{m}_b)],
\end{aligned} \tag{5.20}$$

where for convenience both sides were multiplied by $\widehat{m}_b$, and the variable k was replaced with $\widehat{m}_b x$, such that $\widehat{m}_b W(\widehat{m}_b x)$ is a dimensionless function of x. In the second step the expression was expanded in δm_b, keeping only terms that contribute up to $\mathcal{O}(\alpha_s^3)$, recalling that $\delta m_b \sim \mathcal{O}(\alpha_s)$ and $\beta(\alpha_s) \sim \alpha_s^2$. The derivatives $\mathrm{d}/\mathrm{d}x$ and $\mathrm{d}/\mathrm{d}\alpha_s$ act on everything to their right.

Substituting the explicit α_s expansions for $\delta m_b/\widehat{m}_b$, $\beta(\alpha_s)$, and $w(x, \alpha_s)$, it is straightforward to derive the correction terms $\Delta w_{77}^{\mathrm{s}(n)}$ and $\Delta w_{77}^{\mathrm{ns}(n)}$ appearing in eqs. (5.9) and

(5.11). Writing the α_s expansion of $\delta m_b/\widehat{m}_b$ as

$$\frac{\delta m_b}{\widehat{m}_b} = \frac{\alpha_s(\mu)}{4\pi}\delta_m^{(1)}(\mu) + \Big[\frac{\alpha_s(\mu)}{4\pi}\Big]^2\delta_m^{(2)}(\mu) + \Big[\frac{\alpha_s(\mu)}{4\pi}\Big]^3\delta_m^{(3)}(\mu) + \mathcal{O}(\alpha_s^4), \tag{5.21}$$

the singular correction is found to be

$$\begin{aligned}
\Delta w_{77}^{\mathrm{s}(1)}(\mu,x) &= \frac{1}{4}\delta_m^{(1)}(\mu)\frac{\mathrm{d}}{\mathrm{d}x}w_{77}^{\mathrm{s}(0)}(x),\\
\Delta w_{77}^{\mathrm{s}(2)}(\mu,x) &= \frac{1}{16}\bigg\{\delta_m^{(2)}(\mu)\frac{\mathrm{d}}{\mathrm{d}x} + \frac{1}{2}\big[\delta_m^{(1)}(\mu)\big]^2\frac{\mathrm{d}^2}{\mathrm{d}x^2}\bigg\}w_{77}^{\mathrm{s}(0)}(x)\\
&\quad + \frac{1}{4}\delta_m^{(1)}(\mu)\frac{\mathrm{d}}{\mathrm{d}x}\big[(1-x)w_{77}^{\mathrm{s}(1)}(x)\big],\\
\Delta w_{77}^{\mathrm{s}(3)}(\mu,x) &= \frac{1}{64}\bigg\{\delta_m^{(3)}(\mu)\frac{\mathrm{d}}{\mathrm{d}x} + \delta_m^{(1)}(\mu)\delta_m^{(2)}(\mu)\frac{\mathrm{d}^2}{\mathrm{d}x^2} + \big[\delta_m^{(1)}(\mu)\big]^3\frac{1}{3!}\frac{\mathrm{d}^3}{\mathrm{d}x^3}\bigg\}w_{77}^{\mathrm{s}(0)}(x)\\
&\quad + \frac{1}{16}\bigg\{-2\beta_0\delta_m^{(1)}(\mu) + \Big[\delta_m^{(2)}(\mu) + 2\beta_0\delta_m^{(1)}(\mu)\ln\frac{\mu}{\widehat{m}_b}\Big]\frac{\mathrm{d}}{\mathrm{d}x}(1-x)\\
&\qquad + \frac{1}{2}\big[\delta_m^{(1)}(\mu)\big]^2\frac{\mathrm{d}^2}{\mathrm{d}x^2}(1-x)^2\bigg\}w_{77}^{\mathrm{s}(1)}(x)\\
&\quad + \frac{1}{4}\delta_m^{(1)}(\mu)\frac{\mathrm{d}}{\mathrm{d}x}(1-x)w_{77}^{\mathrm{s}(2)}(x),
\end{aligned} \tag{5.22}$$

and the nonsingular is

$$\begin{aligned}
\frac{\Delta w_{77}^{\mathrm{ns}(2)}(\mu_{\mathrm{ns}},x)}{(1-x)^3} &= \frac{1}{4}\delta_m^{(1)}(\mu_{\mathrm{ns}})\frac{\mathrm{d}}{\mathrm{d}x}\bigg[\frac{w_{77}^{\mathrm{ns}(1)}(x)}{(1-x)^2}\bigg],\\
\frac{\Delta w_{77}^{\mathrm{ns}(3)}(\mu_{\mathrm{ns}},x)}{(1-x)^3} &= \frac{1}{16}\bigg\{-2\beta_0\delta_m^{(1)}(\mu_{\mathrm{ns}}) + \Big[\delta_m^{(2)}(\mu_{\mathrm{ns}}) + 2\beta_0\delta_m^{(1)}(\mu_{\mathrm{ns}})\ln\frac{\mu_{\mathrm{ns}}}{\widehat{m}_b}\Big]\frac{\mathrm{d}}{\mathrm{d}x}(1-x)\\
&\qquad + \frac{1}{2}\big[\delta_m^{(1)}(\mu_{\mathrm{ns}})\big]^2\frac{\mathrm{d}^2}{\mathrm{d}x^2}(1-x)^2\bigg\}\frac{w_{77}^{\mathrm{ns}(1)}(x)}{(1-x)^3}\\
&\quad + \frac{1}{4}\delta_m^{(1)}(\mu_{\mathrm{ns}})\frac{\mathrm{d}}{\mathrm{d}x}\frac{w_{77}^{\mathrm{ns}(2)}(x)}{(1-x)^2}.
\end{aligned} \tag{5.23}$$

There are three sources of δm_b corrections:

1. Shifting the argument $k \to k + \delta m_b$.
2. Changing to $\widehat{m}_b$ in the argument $k/\widehat{m}_b$, which yields the rescaling $x \to x/(1 + \delta m_b/\widehat{m}_b)$.
3. Changing to $\widehat{m}_b$ in the μ-dependence of $\alpha_s(\widehat{m}_b)$.

Considering just the singular contributions and only keeping the first and neglecting the latter two corrections amounts to only keeping the leading-power terms in eq. (5.20):

$$\widehat{m}_b W^{\mathrm{s}}(\widehat{m}_b x) = \bigg[1 + \frac{\delta m_b}{\widehat{m}_b}\frac{\mathrm{d}}{\mathrm{d}x} + \frac{1}{2}\frac{\delta m_b^2}{\widehat{m}_b^2}\frac{\mathrm{d}^2}{\mathrm{d}x^2} + \frac{1}{3!}\frac{\delta m_b^3}{\widehat{m}_b^3}\frac{\mathrm{d}^3}{\mathrm{d}x^3}\bigg]w^{\mathrm{s}}[x,\alpha_s(\widehat{m}_b)]. \tag{5.24}$$

The relevant formal power counting here is $\delta m_b/\widehat{m}_b \sim \lambda \ll 1$, $\mathrm{d}/\mathrm{d}x \sim 1/x \sim \lambda^{-1}$, so all terms on the right-hand side of eq. (5.24) are leading power. These terms are exactly reproduced by the factorized result at fixed order by changing the soft function to the $\widehat{m}_b$ scheme, which involves the analogous shift of its argument. In the numerical implementation, the derivatives $\mathrm{d}/\mathrm{d}(\widehat{m}_b x) = \mathrm{d}/\mathrm{d}k$ are moved via integration by parts to act on the shape function $\mathcal{F}(k)$, so they count as $1/\Lambda_{\mathrm{QCD}}$. Since $\delta m_b \sim \Lambda_{\mathrm{QCD}}$, again, all terms in eq. (5.24) are leading power, counting as $(\delta m_b \mathrm{d}/\mathrm{d}k)^n \sim 1$.

All terms $\sim x^n w^{\mathrm{s}}(x,\alpha_s)$ are thus induced by the second source. By moving the derivatives to act onto the shape function, they can be shown to be explicitly power-suppressed by x and hence nonsingular. For this reason, they could be included as part of the nonsingular correction terms $\Delta w_{77}^{\mathrm{ns}}$, as was done in ref. [7]. In the singular contributions, the k/m_b dependence only appears in logarithms which are factorized into the soft, jet, and hard functions, where m_b corresponds to the large p^- label momentum, which only appears in the hard and jet functions, while the soft function only depends on the small momentum k. Therefore, the associated correction terms can be reproduced by the leading-power factorized result by changing the m_b dependence in the hard and jet functions to the short-distance $\widehat{m}_b$.

Finally, the third source produces the last term in eq. (5.20),

$$\frac{\delta m_b}{\widehat{m}_b}\beta(\alpha_s)\frac{\mathrm{d}}{\mathrm{d}\alpha_s}[w(x,\alpha_s(\widehat{m}_b)] = -\frac{\alpha_s^3 C_F}{8\pi^3}\delta_m^{(1)}\beta_0 w_{77}^{\mathrm{s}(1)}(x) + \mathcal{O}(\alpha_s^4). \tag{5.25}$$

Since it starts at $\mathcal{O}(\alpha_s^3)$ it first appears at $\mathrm{N^3LL'}$. Formally, this term is also subleading power, because $\delta\widehat{m}_b/\widehat{m}_b \sim \lambda$. However, since there is no explicit kinematic suppression by x, the x-dependence itself is still singular $\sim 1/x$, involving $\delta(x)$ and logarithmic plus distributions. Hence, this term cannot simply be absorbed into the nonsingular contributions, but must be properly accounted for in the resummed singular contribution. The $\widehat{m}_b$-dependence in the corresponding fixed-order $\ln(\mu/\widehat{m}_b)$ also corresponds to the p^- label momentum. Therefore, to account for the distributional structure of this contribution and to resum it, the hard and jet functions are consistently rewritten in terms of the short-distance mass $\widehat{m}_b$. The details of this procedure are discussed in the following two subsections.

5.3.1. Hard function in a short-distance scheme

The hard function in a short-distance scheme is obtained by writing the b-quark mass in the pole-scheme hard function in eq. (E.1) in terms of a short-distance mass and expanding the result strictly in powers of $\alpha_s(\mu)$. At $\mathrm{N^3LO}$:

$$\hat{h}_s(\widehat{m}_b,\mu) = h_s(\widehat{m}_b,\mu) - \sum_{n=2}^{\infty}\sum_{m=0}^{2n-3}\Delta H_m^{(n)}(\mu)\left[\frac{\alpha_s(\mu)}{4\pi}\right]^n \ln^m\frac{\mu}{\widehat{m}_b}, \tag{5.26}$$

where the coefficients $\Delta H_m^{(n)}$ are given by

$$\begin{aligned}
\Delta H_0^{(2)}(\mu) &= \delta_m^{(1)}(\mu) H_1^{(1)},\\
\Delta H_1^{(2)}(\mu) &= 2\delta_m^{(1)}(\mu) H_2^{(1)},\\
\Delta H_0^{(3)}(\mu) &= \delta_m^{(2)}(\mu) H_1^{(1)} + \delta_m^{(1)}(\mu) H_1^{(2)} - \frac{1}{2}\big(\delta_m^{(1)}(\mu)\big)^2\Big(H_1^{(1)} + 2H_2^{(1)}\Big),\\
\Delta H_1^{(3)}(\mu) &= 2\delta_m^{(2)}(\mu) H_2^{(1)} + 2\delta_m^{(1)}(\mu) H_2^{(2)} - \big(\delta_m^{(1)}(\mu)\big)^2 H_2^{(1)},\\
\Delta H_2^{(3)}(\mu) &= 3\delta_m^{(1)}(\mu) H_3^{(2)},\\
\Delta H_3^{(3)}(\mu) &= 4\delta_m^{(1)}(\mu) H_4^{(2)}.
\end{aligned} \tag{5.27}$$

Here $H_m^{(n)}$ are the coefficients of the pole-scheme hard function defined in eq. (E.1), and $\delta_m^{(n)}(\mu)$ are defined in eq. (5.21).

Similarly, the RGE equation for the hard function in a short-distance mass scheme is derived by rewriting the pole mass in eq. (E.2) in terms of a short-distance mass:

$$\frac{\mathrm{d}\hat{h}_s(\widehat{m}_b,\mu)}{\mathrm{d}\ln\mu} = \Big\{\Gamma^H\big(\alpha_s(\mu)\big)\ln\frac{\mu}{\widehat{m}_b} + \gamma^H\big(\alpha_s(\mu)\big) + \Delta\gamma^H(\mu)\Big\}\hat{h}_s(\widehat{m}_b,\mu), \tag{5.28}$$

where

$$\begin{aligned}
\Delta\gamma^H(\mu) &= \Gamma^H\big(\alpha_s(\mu)\big)\ln\frac{\widehat{m}_b}{m_b^{\mathrm{pole}}} = -\Gamma^H\big(\alpha_s(\mu)\big)\ln\Big(1+\frac{\delta m_b}{\widehat{m}_b}\Big)\\
&= -\Big[\frac{\alpha_s(\mu)}{4\pi}\Big]^2\Gamma_0^H\delta_m^{(1)}(\mu)\\
&\quad - \Big[\frac{\alpha_s(\mu)}{4\pi}\Big]^3\Big\{\Gamma_1^H\delta_m^{(1)}(\mu) + \Gamma_0^H\Big[\delta_m^{(2)}(\mu) - \frac{1}{2}(\delta_m^{(1)}(\mu))^2\Big]\Big\} + \mathcal{O}(\alpha_s^4),
\end{aligned} \tag{5.29}$$

and the hard cusp anomalous dimension coefficients $\Gamma_n^H = -2\Gamma_n$ are defined in eq. (D.7). Note that the expansion in eqs. (5.26) and (5.28) must be performed in terms of the same $\alpha_s(\mu)$, such that the hard function $\hat{h}_s(\widehat{m}_b,\mu)$ in eq. (5.26) indeed satisfies the RGE in eq. (5.28).

The renormalon associated with the pole mass is expected to cancel in the perturbative series of the hard anomalous dimension. The cusp anomalous dimension is universal and arises in the evolution of many perturbative objects with Sudakov double logarithms, that do not involve the b-quark mass at all. Therefore, the Γ^H series cannot contain the pole-mass renormalon, so the cancellation must happen between γ^H and $\Delta\gamma^H$. For this reason, it is important to consistently expand $\Delta\gamma^H$ in powers of $\alpha_s(\mu)$ to the same order as γ^H.

The all-order solution to the differential equation (5.28) can be written as

$$\widehat{U}_H(\widehat{m}_b,\mu_H,\mu) = U_H(\widehat{m}_b,\mu_H,\mu)\times\Delta U_H(\mu_H,\mu), \tag{5.30}$$

where U_H is the usual hard evolution factor given in eq. (E.6), and the correction factor ΔU_H is given by

$$\Delta U_H(\mu_H,\mu) = \exp\bigg[\int_{\mu_H}^{\mu}\mathrm{d}\ln\mu\,\Delta\gamma^H(\mu)\bigg]. \tag{5.31}$$

Note that in general, the μ-dependence of $\Delta\gamma^H(\mu)$ coming from $\delta_m^{(n)}(\mu)$ can be more involved than for the usual anomalous dimension, such that the μ integral may have to be performed numerically. Using the MSR mass and up to $\mathrm{N^3LL}$, the integral can still be taking using an analytic approximation analogous to the one used for the K and η integrals in eq. (D.8):

$$\begin{aligned}\ln \Delta U_H(\mu_H,\mu) = {}& \frac{\alpha_s(\mu_H)}{4\pi}\frac{\Gamma_0^H}{2\beta_0}\delta^{(1)}_{\rm MSR}(r-1)\\ &+\frac{\alpha_s(\mu_H)}{4\pi}\frac{\Gamma_0^H}{4\beta_0}\biggl\{\frac{\alpha_s(\mu_H)}{4\pi}\biggl[\biggl(\delta^{(2)}_{\rm MSR}-\frac{1}{2}(\delta^{(1)}_{\rm MSR})^2+2\delta^{(1)}_{\rm MSR}\beta_0\ln\frac{\mu_H}{R}\biggr)\\ &+\delta^{(1)}_{\rm MSR}\biggl(\frac{\Gamma_1^H}{\Gamma_0^H}-\frac{\beta_1}{\beta_0}\biggr)\biggr](r^2-1)-\delta^{(1)}_{\rm MSR}(r-1)^2\biggr\},\end{aligned} \tag{5.32}$$

where $r=\alpha_s(\mu)/\alpha_s(\mu_H)$ and

$$\delta^{(n)}_{\rm MSR} \coloneqq \frac{R}{m_b^{\rm MSR}(R)}a_n^{\rm MSR}. \tag{5.33}$$

At NNLL only the first line on the right-hand side of eq. (5.32) is kept.

5.3.2. Jet function in a short-distance scheme

Similar to the hard function in the previous section, the jet function in a short-distance scheme is defined by expressing $m_b^{\rm pole}$ in terms of a short-distance mass $\widehat{m}_b$. Starting with the jet function in the pole scheme given in eq. (E.7) with $s=m_b^{\rm pole}\omega$, the pole mass $m_b^{\rm pole}$ is expanded in terms of $\widehat{m}_b$ and correction δm_b, which is expanded in a series in $\alpha_s(\mu)$. As a result, $m_b^{\rm pole}J(m_b^{\rm pole}\omega,\mu_J)=\widehat{m}_b\hat{J}(\widehat{m}_b\omega,\mu_J)$ order by order. This yields up to $\mathrm{N^3LO}$:

$$\hat{J}(\widehat{m}_b\omega,\mu)=J(\widehat{m}_b\omega,\mu)+\sum_{n=2}^{3}\sum_{m=-1}^{2n-4}\Delta J_m^{(n)}\biggl[\frac{\alpha_s(\mu)}{4\pi}\biggr]^n\frac{1}{\mu^2}\mathcal{L}_m\Bigl(\frac{\widehat{m}_b\omega}{\mu^2}\Bigr). \tag{5.34}$$

The expansion coefficients $\Delta J_m^{(n)}$ read

$$\begin{aligned}
\Delta J_{-1}^{(2)}(\mu) &= \delta_m^{(1)}(\mu)J_0^{(1)},\\
\Delta J_0^{(2)}(\mu) &= \delta_m^{(1)}(\mu)J_1^{(1)},\\
\Delta J_{-1}^{(3)}(\mu) &= \delta_m^{(2)}(\mu)J_0^{(1)}+\delta_m^{(1)}(\mu)J_0^{(2)}-\frac{1}{2}\bigl(\delta_m^{(1)}(\mu)\bigr)^2\Bigl(J_0^{(1)}-J_1^{(1)}\Bigr),\\
\Delta J_0^{(3)}(\mu) &= \delta_m^{(2)}(\mu)J_1^{(1)}+\delta_m^{(1)}(\mu)J_1^{(2)}-\frac{1}{2}\bigl(\delta_m^{(1)}(\mu)\bigr)^2J_1^{(1)},\\
\Delta J_1^{(3)}(\mu) &= 2\delta_m^{(1)}(\mu)J_2^{(2)},\\
\Delta J_2^{(3)}(\mu) &= 3\delta_m^{(1)}(\mu)J_3^{(2)},
\end{aligned} \tag{5.35}$$

where the $\delta_m^{(n)}$ are defined in eq. (5.21) and the $J_m^{(n)}$ coefficients are those of the original pole-scheme jet function as defined in eq. (E.7).

5.4. Profile functions

The resummation of the singular contributions is determined by the choice of the hard, jet, and soft scales μ_H, μ_J, μ_S in the factorization theorem in eq. (5.5). To achieve the proper resummation they must be chosen according to the kinematics relevant in the different regions of the spectrum. In addition, the nonsingular scale $\mu_{\rm ns}$ determines the scale at which the fixed-order nonsingular terms in eq. (5.11) are evaluated.

The choice of renormalization scales in this work follows ref. [7]. The $B \to X_s\gamma$ photon energy spectrum has three parametrically distinct kinematic regions:

- **Shape function (nonperturbative) region:** $\Lambda_{\rm QCD} \sim (m_B - 2E_\gamma) \ll \widehat{m}_b$
  This corresponds to the peak of the spectrum, where the full shape of the shape function is relevant and the soft scale is fixed to the lowest still-perturbative scale $\mu_S = \mu_0 \gtrsim \Lambda_{\rm QCD}$.

- **Shape function OPE region:** $\Lambda_{\rm QCD} \ll (m_B - 2E_\gamma) \ll \widehat{m}_b$
  This corresponds to the transition region between the peak and the far tail. Here the canonical soft scale is $\mu_S \sim m_B - 2E_\gamma$.

- **Local OPE (fixed-order) region:** $\Lambda_{\rm QCD} \ll (m_B - 2E_\gamma) \sim \widehat{m}_b$
  This corresponds to the far tail of the spectrum, which is described by fixed-order perturbation theory. Here, the distinction between singular and nonsingular becomes meaningless and the resummation must be turned off to ensure that singular and nonsingular contributions properly recombine into the correct fixed-order result. This requires that all scales become equal $\mu_S = \mu_J = \mu_H = \mu_{\rm ns} =: \mu_{\rm FO} \sim \widehat{m}_b$.

The canonical value for the hard scale in all regions is $\mu_H \sim \widehat{m}_b$. In the first two regions the SCET resummation is applicable with the canonical scaling $\mu_J \sim \sqrt{\mu_S \mu_H}$.

To account for these different scale hierarchies the common approach of profile functions [23, 38] is used. In this approach, the scales $\mu_S(E_\gamma)$, $\mu_J(E_\gamma)$, $\mu_{\rm ns}(E_\gamma)$ are taken to be functions of the photon energy E_γ. The key advantage of using profile functions is that they provide a smooth transition between the regions that is solely implemented in terms of scale choices, such that the ambiguities in the precise choice of the profile functions are equivalent to scale ambiguities, which by construction are formally beyond the order one is working and reduce at higher orders.

This work uses the same profile functions as in ref. [7]:

$$\begin{aligned}
\mu_H &= e_H \widehat{m}_b, \\
\mu_S(E_\gamma) &= \mu_0 + (\mu_H - \mu_0) f_\theta\biggl(\frac{E_1 - E_\gamma}{E_1 - E_2}\biggr), \\
\mu_J(E_\gamma) &= \bigl[\mu_S(E_\gamma)\bigr]^{(1-e_J)/2} \mu_H^{(1+e_J)/2}, \\
\mu_{\rm ns}(E_\gamma) &= \bigl[\mu_S(E_\gamma)\bigr]^{(1-e_{\rm ns})/4} \mu_H^{(3+e_{\rm ns})/4},
\end{aligned} \tag{5.36}$$

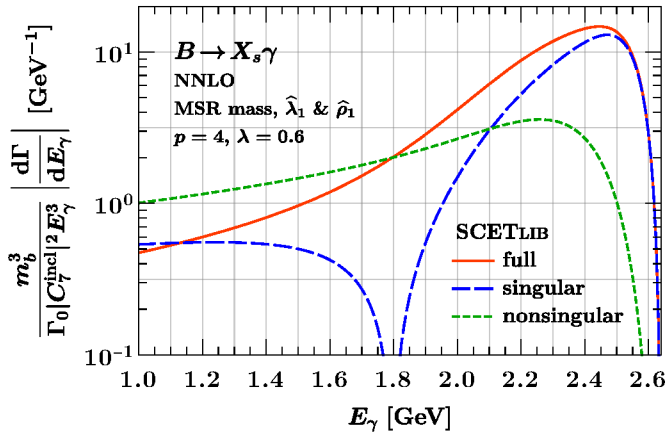


Figure 5.1.: Comparison of the absolute values of the singular, nonsingular, and full contributions to the $B \to X_s\gamma$ photon energy spectrum at NNLO in the MSR mass scheme.

where the quadratic easing function f_θ is defined as

$$f_\theta(x) \coloneqq \begin{cases} 0 & x \le 0, \\ 2x^2 & 0 \le x \le 1/2, \\ 1 - 2(1-x)^2 & 1/2 \le x \le 1, \\ 1 & 1 \le x. \end{cases} \tag{5.37}$$

The function $f_\theta(x)$ is 0 for $x \le 0$, is 1 for $x \ge 1$, and smoothly interpolates between these two values in the region $0 \le x \le 1$.

Figure 5.1 shows the absolute values of the singular and nonsingular contributions, as well as the full result at NNLO, without resummation. The relative magnitude of singular and nonsingular contributions is examined to pick the transition points E_1 and E_2 between the different parametric regions. In the region where $E_\gamma \gtrsim E_1 = 2.2\,\mathrm{GeV}$ the singular contributions clearly dominate. The nonperturbative shape-function effects are important in this region. In the region where $E_\gamma \lesssim E_2 = 1.6\,\mathrm{GeV}$ there are large cancellations between the singular and nonsingular contributions. This corresponds to the fixed-order region, where the separation into singular and nonsingular is artificial and only their sum is meaningful. Therefore, the resummation of the singular contributions must be turned off to not disturb the delicate cancellation between singular and nonsingular, and to recover the correct full result. Since there is little space between E_1 and E_2, the transition region between them effectively coincides with the intermediate shape function OPE region.

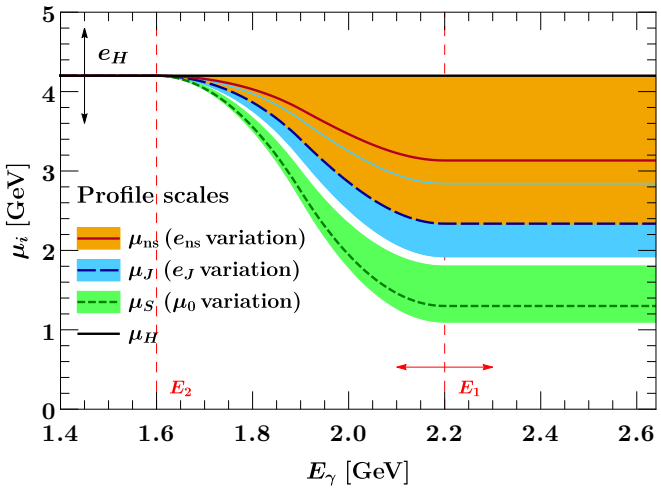


Figure 5.2.: Profile functions used for the hard, jet, soft, and nonsingular scales $\mu_H, \mu_J, \mu_S, \mu_{\rm ns}$. The bands show the individual ranges for the jet, soft, and nonsingular scales. The black and red arrows indicate the variations for the parameter e_H and the transition point E_1.

To summarize, the following values are used for the profile function parameters [7]:

$$\begin{aligned} e_H &\in \{1, 1/2, 2\}, & \mu_0 &\in \max(1, e_H) \times \{1.3, 1.1, 1.8\}\,\mathrm{GeV}, \\ e_J &\in \{0, -1/3, +1/3\}, & e_{\rm ns} &\in \{0, -1, +1\}, \\ E_1 &\in \{2.2, 2.1, 2.3\}\,\mathrm{GeV}, & E_2 &= 1.6\,\mathrm{GeV}. \end{aligned} \tag{5.38}$$

For each parameter, the first value in the set is the central value and the next two are the variations that will be used to assess perturbative uncertainties. This procedure will be discussed in detail in section 5.5.

The central profile functions and variations of individual profile function parameters are illustrated in figure 5.2. Since the nonsingular contributions are treated at fixed order, a priori there is no canonical choice of the nonsingular scale $\mu_{\rm ns}$ beyond the fixed-order region $E_\gamma \geq E_2$. In practice, it is picked as the geometric mean of the hard and central jet scales to account for the fact that the nonsingular terms have some sensitivity to scales below $\widehat{m}_b$. The nonsingular scale $\mu_{\rm ns}$ is varied up to the hard and down to the central jet scales as shown in figure 5.2.

Note that the individual scales are not independent of each other but are parametrized in such a way that their relative hierarchy is preserved for each variation of the profile function parameters. For example, they all depend on μ_H, such that varying μ_H up and down (by varying e_H) simultaneously moves the other scales up and down accordingly. In particular, all scales always merge into a common value for $E_\gamma \leq E_2$ to properly turn off the resummation. Similarly, μ_J and $\mu_{\rm ns}$ depend on μ_S, such that varying μ_0 not only varies μ_S but also moves μ_J and $\mu_{\rm ns}$ up and down accordingly to preserve the hierarchy between them. The variations for μ_J and $\mu_{\rm ns}$ parametrized by e_J and $e_{\rm ns}$ correspond to

small deviations from the default hierarchy. This also means that the μ_0, e_J, and e_{ns} variations smoothly turn off between E_1 and E_2 together with the resummation itself, such that below E_2 only the overall μ_H variation remains, corresponding to the usual fixed-order scale variation.

5.5. Perturbative uncertainties

The theoretical uncertainties can be classified as perturbative or parametric. The parametric uncertainties arise from the uncertainty in the input parameters, such as C_7^{incl}, $|V_{tb}V_{ts}^*|^2$, m_b, $\mathcal{F}(k)$. These are not considered in the following. The presented theoretical predictions for the $B \to X_s\gamma$ spectrum are normalized by dividing out the overall factor $\Gamma_0\big|C_7^{\mathrm{incl}}\big|^2$, so the associated uncertainty drops out. The uncertainty due to the shape function $\mathcal{F}(k)$ and the b-quark mass is also ignored. Although these parameters do affect the shape of the spectrum, in a comparison with experimental $B \to X_s\gamma$ measurements these parameters would be fitted to the data.

This work is focused on the perturbative results and their perturbative uncertainties, which arise from missing higher-order corrections. The sources of perturbative uncertainty can be divided into two categories:

- **Scale variations:** Scale variation uncertainties are estimated by a suitable set of variations of the profile functions discussed in section 5.4, and are further subdivided into the following three groups: The resummation uncertainty Δ_{resum} is obtained by taking the maximum envelope of all $3^3 = 27$ variations of the profile function parameters e_H, e_J, μ_0, which corresponds to scale variations in the resummed singular contributions, together with corresponding correlated variations in the nonsingular contributions. Note also that despite its name, Δ_{resum} reduces to the overall fixed-order scale variation in the fixed-order region. The nonsingular uncertainty Δ_{ns} is determined by varying the parameter e_{ns}, which determines the central value of the nonsingular scale in the resummation regions. The matching uncertainty Δ_{match} comes from varying the transition point E_1, which marks the start of the transition region. These three sources are added in quadrature:

  $$\Delta_{\mathrm{profile}} = \Delta_{\mathrm{resum}} \oplus \Delta_{\mathrm{ns}} \oplus \Delta_{\mathrm{match}}, \tag{5.39}$$

  where the symbol $\oplus$ denotes addition in quadrature: $x \oplus y \coloneqq \sqrt{x^2 + y^2}$.

- **Theory nuisance parameter variations:** The uncertainties Δ_{h_3} and $\Delta_{c^{\mathrm{ns}}}$ are estimated by varying the nuisance parameters h_3 and c_k^{ns}, respectively, within the ranges given in sections 5.1 and 5.2, where $\Delta_{c^{\mathrm{ns}}} \coloneqq \sqrt{\sum_{k=0}^{5} \Delta^2_{c_k^{\mathrm{ns}}}}$. Note that by definition the central values of all nuisance parameters are zero.

  The nuisance parameters at $\mathrm{N^3LL'}$+$\mathrm{N^3LO}(c_k)$ are introduced in such a way that the scale dependence cancels at this order, i.e. all terms that are predicted by scale dependence are correctly included. This means that at $\mathrm{N^3LL'}$+$\mathrm{N^3LO}(c_k)$ the scale variations estimate the uncertainty due to the missing next order $\mathrm{N^nLL'}$+$\mathrm{N^nLO}$,

while the nuisance parameters capture the uncertainty due to the missing $\mathcal{O}(\alpha_s^3)$ ingredients at $\mathrm{N^3LL'}$+$\mathrm{N^3LO}(c_k)$.

The total perturbative uncertainty is obtained by adding all sources in quadrature,

$$\Delta_{\text{total}} = \Delta_{\text{profile}} \oplus \Delta_{h_3} \oplus \Delta_{c^{\text{ns}}}. \tag{5.40}$$

The nuisance parameter uncertainties $\Delta_{h_3} \oplus \Delta_{c^{\text{ns}}}$ only contribute at the highest order $\mathrm{N^3LL'}$+$\mathrm{N^3LO}(c_k)$, while at lower orders the total uncertainty is simply the scale variation uncertainty $\Delta_{\text{total}} = \Delta_{\text{profile}}$.

Note that in ref. [7] the perturbative uncertainty was estimated from scale variations by taking the envelope of all $3^5 = 243$ profile function variations, corresponding to simultaneous variations of the profile function parameters $e_H, \mu_0, e_J, e_{\text{ns}}, E_1$. In this work the scale variations have been refined by separating conceptually different sources of perturbative uncertainties, which leads to an overall more consistent picture of the resulting uncertainties when including the new highest order at $\mathrm{N^3LL'}$. In part, this becomes possible because it is not possible to reexpand the fixed-order hard function against the product of the fixed-order jet and soft functions. It has been checked that the total Δ_{profile}, estimated as described above, is comparable in size to an alternative uncertainty estimate obtained by taking the maximum envelope of all $3^5 = 243$ variations and excluding a small set of obvious outliers.

5.6. Numerical results

The results presented in this section by default use the following setup: The values used for input parameters are summarized in table 5.1. The $B \to X_s\gamma$ spectrum is always divided by the overall normalization factor $\Gamma_0 \left|C_7^{\text{incl}}\right|^2$, and therefore the numerical value of $\Gamma_0 \left|C_7^{\text{incl}}\right|^2$ is irrelevant. The shape function is modelled by the function in eq. (3.29) with $\lambda = 0.6\,\text{GeV}$ and $p = 4$. The impact of changing p and λ is presented later in this section. The finite-charm-mass corrections are neglected, and the perturbative ingredients are evaluated in QCD with $n_f = 4$ massless quark flavors. The 4-loop running of α_s is used for all results, which is sufficient for resummation at $\mathrm{N^3LL}$. The MSR scheme for the b-quark mass and HQET hadronic parameters λ_1 and ρ_1 are switched to a short-distance scheme as discussed in section 4.3. The impact of the short-distance mass scheme is discussed in section 5.6.3. Throughout this section, the colored bands always show the perturbative uncertainties obtained from scale variations Δ_{profile}.

5.6.1. Main results

To begin with, figure 5.3 shows the absolute value of the resummed singular contributions in comparison to the full result at $\mathrm{NNLL'}$ (left panel) and $\mathrm{N^3LL'}$ (right panel). The resummed contribution is indeed dominant in the peak region of the spectrum, but it decreases rapidly in the tail and eventually changes sign at $E_\gamma \lesssim 1.8\,\text{GeV}$, where the resummation is getting turned off. In this region only the full matched result is meaningful, which remains positive and slowly approaches zero in the far tail region.

Parameter	Value
m_B	5.279 GeV
$\alpha_s(\mu = 4.7\,\mathrm{GeV})$	0.2155
$m_b^{\mathrm{MSR}}(R = 1\,\mathrm{GeV})$	4.7 GeV
λ	0.6 GeV
p	4

Table 5.1.: The used numerical values of input parameters.

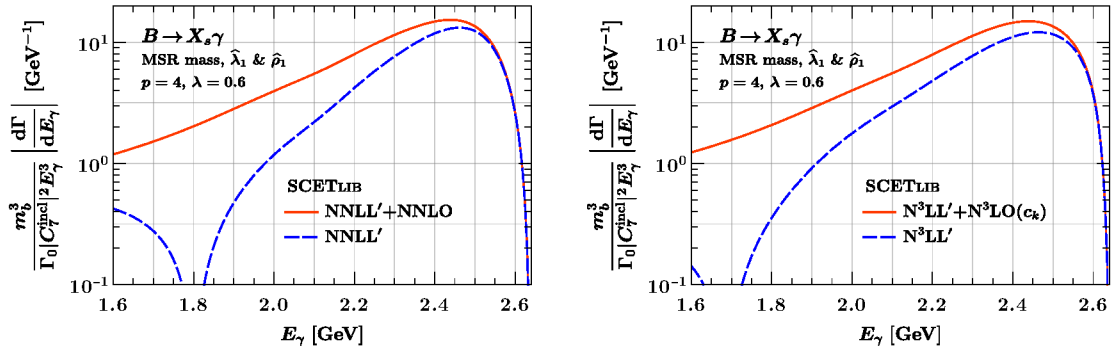


Figure 5.3.: Comparison of the absolute value of resummed contribution to the full $B \to X_s\gamma$ photon energy spectrum in the MSR mass scheme. The left panel shows the 2-loop predictions, while the right panel shows the 3-loop predictions. The overall factor $\Gamma_0 \left|C_7^{\mathrm{incl}}\right|^2 (E_\gamma/\widehat{m}_b)^3$ is divided out (see eq. (5.1)).

The main predictions for the $B \to X_s\gamma$ photon energy spectrum at different perturbative orders are presented in figure 5.4. In addition to the colored bands, which show the scale variation uncertainty $\Delta_{\mathrm{profile}}$, the gray dashed line shows $\Delta_{\mathrm{profile}} \oplus \Delta_{h_3}$, and the black solid line shows Δ_{total}, as defined in eq. (5.40). The first column shows the results for the spectrum itself, and the second column shows the relative deviations from the central value at the highest order, i.e. the deviation from the red solid line on the left. In the first row the same value for the shape-function parameter $p = 4$ is used at all orders, whereas in the second row different values for the parameter p are used at each order. Figure 5.5 shows the breakdown of the relative perturbative uncertainties into the individual contributions at NNLL′+NNLO and $\mathrm{N^3LL'}$+$\mathrm{N^3LO}(c_k)$. As already mentioned in section 3.1, the choice $p = 4$ ensures that the spectrum at the highest order $\mathrm{N^3LL'}$ still vanishes in the limit $E_\gamma \to m_B/2$. In practice, when fitting to the experimental data, the fit will always fix the precise shape near the endpoint to that of the data irrespective of the perturbative order, while the perturbative differences get moved, at least partially, into the fit result for $\mathcal{F}$. The comparison of spectrum predictions at different orders with the same fixed model for the shape function $\mathcal{F}$ can be used to directly assess the perturbative convergence of the spectrum predictions. However, when the same value of the parameter p is used, the spectrum at lower orders vanishes correspondingly faster,

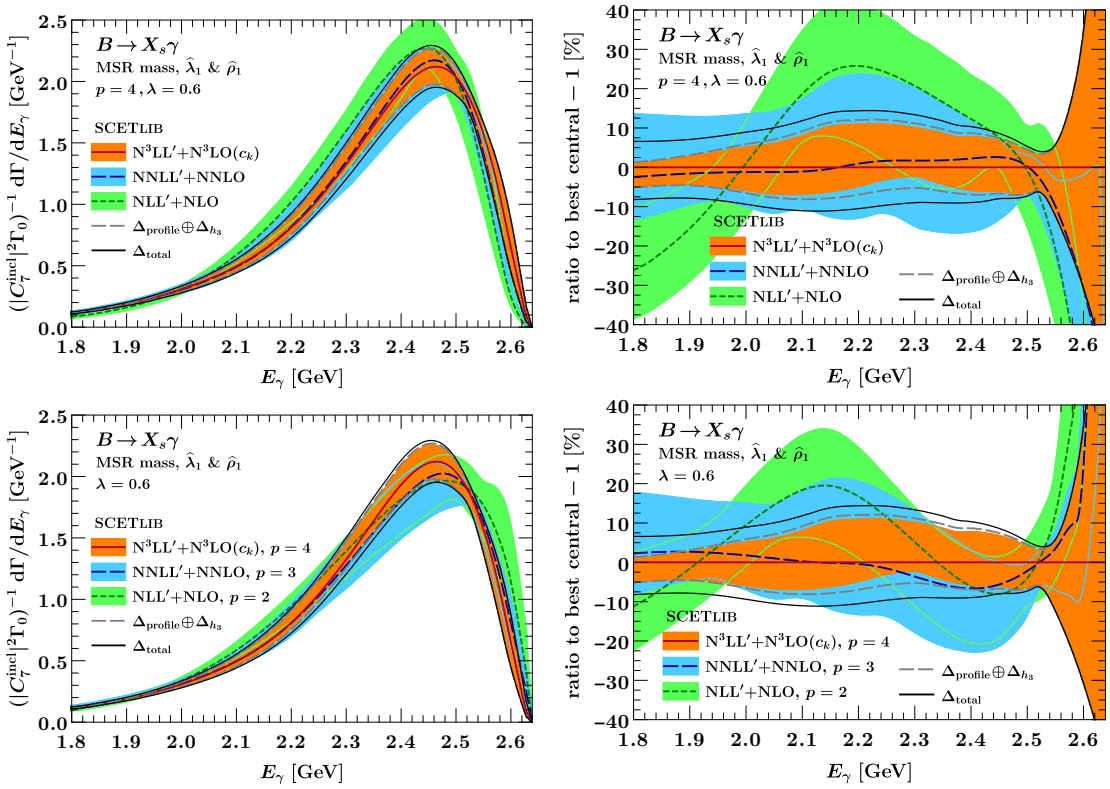


Figure 5.4.: The $B \to X_s\gamma$ spectrum at different perturbative orders. The photon energy spectrum itself is shown on the left, while the relative deviations from the highest-order central value are shown on the right. In the top row, the same value for the shape function parameter $p = 4$ is used at each order, while in the bottom row successive values for each order are used. The colored bands show the scale variation uncertainty $\Delta_{\mathrm{profile}}$. The dashed gray line the uncertainty due to h_3, in addition to the scale variation uncertainty. The solid black line further accounts for the uncertainty related to the nonsingular nuisance parameters, corresponding to the total perturbative uncertainty at $\mathrm{N^3LL'{+}N^3LO}(c_k)$.

i.e. quadratically at NNLL′ and cubically at NLL′, which also affects to some extent the shape of the spectrum near its peak. For this reason, the bottom row figure 5.4 also shows an alternative order comparison, where the model is changed at each order by using successively lower values for p at the lower orders, such that the spectrum vanishes linearly at each order.

The results in figures 5.4 and 5.5 manifest a good perturbative convergence, especially between $\mathrm{NNLL'{+}NNLO}$ and $\mathrm{N^3LL'{+}N^3LO}(c_k)$. The relative uncertainties are under good control over the entire E_γ range, except at the very endpoint, where the spectrum vanishes, and the relative uncertainties necessarily blow up. Apart from the very end-

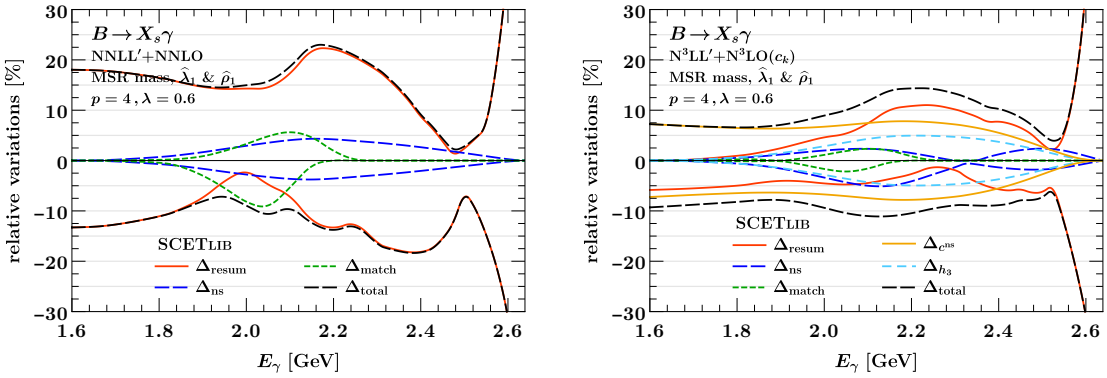


Figure 5.5.: Breakdown of the perturbative uncertainty for the $B \to X_s\gamma$ spectrum into its components at NNLL′+NNLO (left panel) and $\mathrm{N^3LL'}$+$\mathrm{N^3LO}(c_k)$ (right panel).

point, the total uncertainty at $\mathrm{N^3LL'}$+$\mathrm{N^3LO}(c_k)$ is at most 15% and well below that over the most of the E_γ range.

The uncertainties at $\mathrm{N^3LL'}$+$\mathrm{N^3LO}(c_k)$ are substantially reduced in comparison to NNLL′+NNLO, even accounting for the fact that not all 3-loop perturbative ingredients are known. As expected, the higher-order uncertainties estimated from scale variations Δ_{resum}, Δ_{ns}, Δ_{match} are much reduced. Also recall that in the fixed-order regime in the tail region the scale variations contributing to the resummation uncertainty Δ_{resum} turn into the usual fixed-order scale variations. The uncertainty Δ_{h_3} is visible but subdominant. Across the entire peak of the spectrum, the uncertainty $\Delta_{c^{\mathrm{ns}}}$ from the missing 3-loop nonsingular corrections is at most comparable to the other sources thanks to the power suppression of the nonsingular. As expected, in the tail region below $E_\gamma \lesssim E_1 = 2.1\,\mathrm{GeV}$ this uncertainty starts to take over and becomes the dominant uncertainty. This demonstrates that the $\mathrm{N^3LL'}$ resummation increases the precision of the theoretical predictions even in the absence of the full $\mathrm{N^3LO}$ result. Furthermore, in the fixed-order tail region the total uncertainty, dominated by $\Delta_{c^{\mathrm{ns}}}$, is still reduced in comparison to the scale-variation based estimate at NNLL′+NNLO. As discussed at the end of section 5.2, this is anticipated and justified because of the additional nontrivial perturbative information included at $\mathrm{N^3LO}(c_k)$.

The effect of varying the shape function model parameters p and λ is illustrated in figure 5.6. As expected, the position and the height of the peak depend on the model parameters. Figure 5.6 shows that the value of p controls how fast the spectrum vanishes towards the endpoint $E_\gamma \to m_B/2$. The value of λ determines the width of shape function, which as a result controls the width of the peak of the spectrum. It should be stressed that these results are meant as just an additional illustration. In particular, the differences shown in figure 5.6 are not to be interpreted as an additional theoretical uncertainty in the predictions. As mentioned before, the actual shape of $\mathcal{F}$ needs to be determined by fitting to the experimental data.

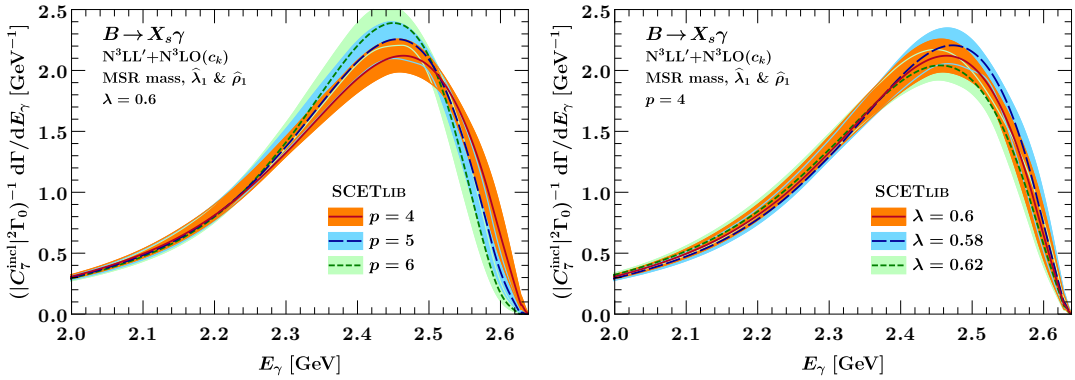


Figure 5.6.: The $B \to X_s\gamma$ spectrum using different values for the shape-function parameters p (left panel) and λ (right panel) at $\mathrm{N^3LL'{+}N^3LO}(c_k)$. The remaining theory parameters are set to their default values.

5.6.2. Different treatment of higher-order singular cross terms

By default, in the product $\hat{h}_s \times \hat{J} \otimes \widehat{C}_0$ the perturbative series of the hard, jet, and partonic soft functions are expanded against each other, and all cross terms that are beyond the working perturbative order are dropped. This approach ensures that in the far-tail region, where the resummation is switched off, the sum of singular and nonsingular contributions reproduces the full fixed-order results. However, in ref. [7] it was observed that in the 1S mass scheme and up to NNLL′+NNLO the perturbative convergence in the resummation region is substantially improved by keeping the hard function unexpanded, and treating it instead as an overall factor. With this alternative treatment, the hard function affects only the normalization of the spectrum, but not its shape, which receives relatively large corrections from the jet and soft functions. The disadvantage is that this approach has the risk of generating unphysically large higher-order corrections in the fixed-order regime. In ref. [7] it was checked that this does not happen in the region of interest.

This subsection is dedicated to the examination of the differences between these two approaches — expanding the hard function against soft and jet functions, which is the default choice, or keeping it unexpanded.

A priori, the additional higher-order terms induced by keeping the hard function unexpanded can easily spoil the delicate cancellation between singular and nonsingular contributions in the far tail of the spectrum. In order to compensate for this effect, an additional constant term is added, which cancels these additional higher-order terms in the fixed-order regime, but amounts to only a power-suppressed correction in the peak region:

$$\Big[W_{77}^{\mathrm{s}}(k) + W_{77}^{\mathrm{ns}}(k)\Big] \to \Big[W_{77}^{\mathrm{s,unexp}}(k) - W_{77,\mathrm{FO}}^{\mathrm{s,unexp}}(\widehat{m}_b)\Big] + \Big[W_{77}^{\mathrm{ns}}(k) + W_{77,\mathrm{FO}}^{\mathrm{s,exp}}(\widehat{m}_b)\Big], \quad (5.41)$$

where the singular contribution $W_{77,\mathrm{FO}}^{\mathrm{s,unexp}}(\widehat{m}_b)$ with unexpanded hard function is subtracted, and the same contribution $W_{77,\mathrm{FO}}^{\mathrm{s,exp}}(\widehat{m}_b)$ with the hard functions expanded out is

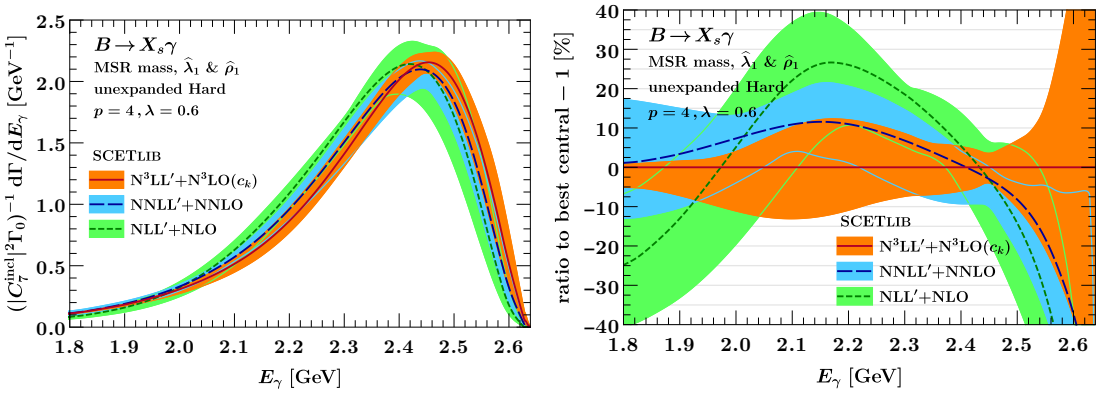


Figure 5.7.: $B \to X_s\gamma$ spectrum with unexpanded hard function.

added. Both of these terms are evaluated at fixed order and at $k = \widehat{m}_b$, corresponding to the tail limit $x = 1$, so in the peak region they are indeed power suppressed. This prescription allows one to take advantage of keeping the hard function unexpanded in the peak region while avoiding unphysically large corrections from higher-order cross terms in the fixed-order region, since they are explicitly removed in the $x \to 1$ limit.

The numerical results using this prescription are shown in figure 5.7. These results are in the MSR mass scheme and use short-distance schemes for $\widehat{\lambda}_1$ and $\widehat{\rho}_1$. These results are compared to the default setup, which uses the expanded hard function, in figure 5.8. Here one can see that, overall, both scenarios lead to somewhat compatible results. Nevertheless, the choice of expanding the hard function or not clearly has a large impact at lower orders (which are expected to be more sensitive to the treatment of higher-order cross terms). Keeping the hard function unexpanded leads to a rather unnatural reduction of scale variations in the peak region of the spectrum. This behavior is quite dramatic at NNLL′+NNLO, where the uncertainty band with an unexpanded hard function (the gray band) barely captures the central line from the results with expanded hard function (solid blue line), and its size is almost as large as the scale variations at $\mathrm{N^3LL'{+}N^3LO}(c_k)$. This is also visible in figure 5.7, where the blue band suspiciously narrows down in the peak region of the spectrum and competes with the orange band. These results show that the default approach taken in this work, which is to expand the hard function series, is preferable. For this reason, the main numerical results, which are presented in section 5.6.1, the fixed-order series of the hard, jet, and soft functions are fully expanded against each other, since it yields an overall more consistent picture of perturbative uncertainties and convergence.

5.6.3. Impact of short-distance schemes: 1S vs MSR mass schemes

As already explained in chapter 4, the right choice of short-distance scheme for the b-quark mass plays a key role in stabilizing the predictions at different orders in perturbation theory. In this section the numerical impact of different schemes for m_b, λ_1, and ρ_1 is investigated. It will be shown that the 1S mass scheme, which was successfully used

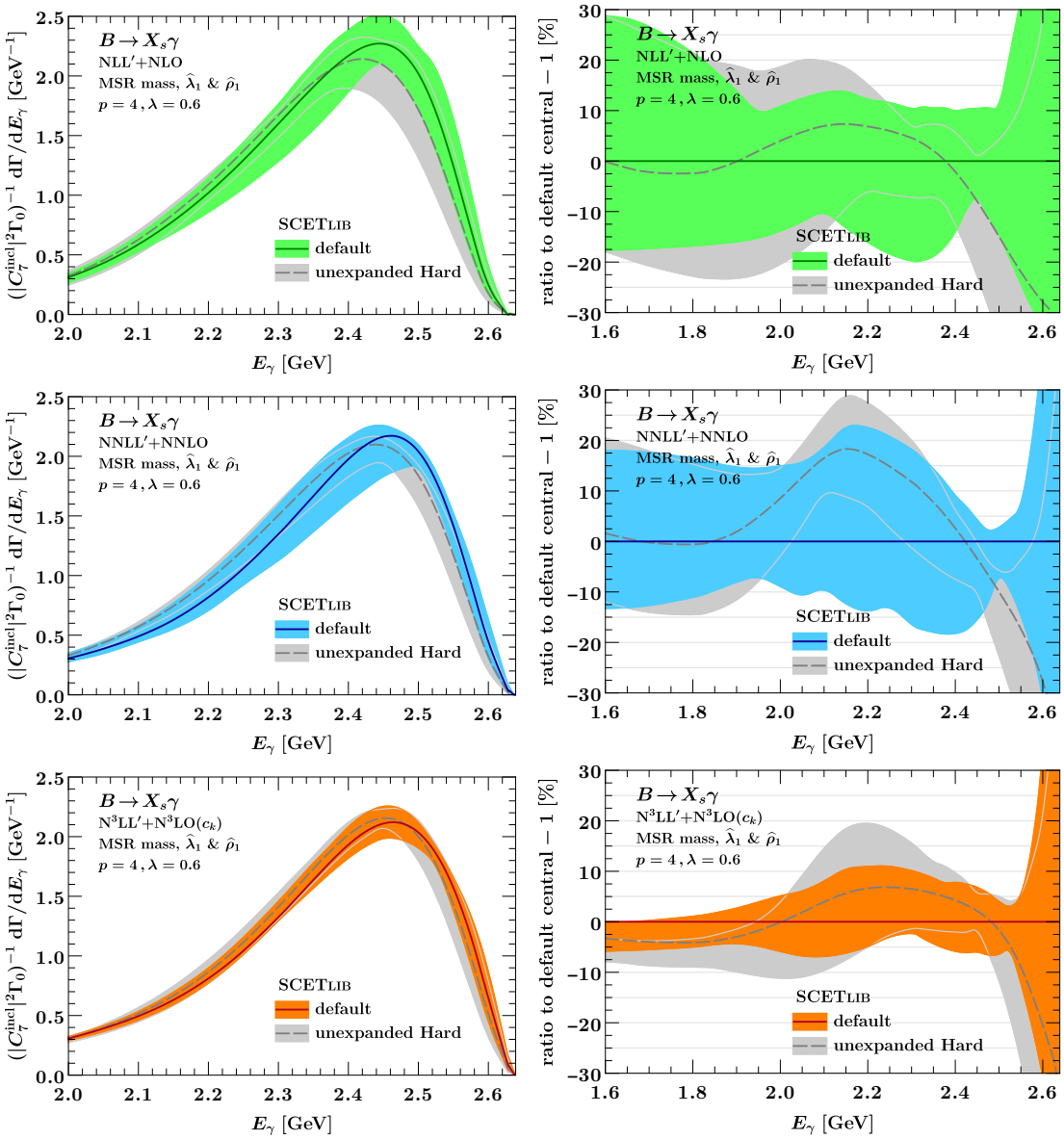


Figure 5.8.: Comparison of $B \to X_s\gamma$ spectrum with expanded and unexpanded hard function. The colored and gray bands at each order display the spectrum with expanded and unexpanded hard function, respectively.

in previous works [7,23] up to NNLL′+NNLO, starts to break down at $\mathrm{N^3LL'}$, while the MSR mass scheme yields convergent, stable results.

Except for the different scheme choices, the results presented in this subsection use the default settings, as discussed at the beginning of section 5.6. The predictions for the soft function and for the photon energy spectrum are shown in figure 5.9 for the pole mass scheme, in figures 5.10 and 5.11 for the 1S mass scheme, and in figure 5.12

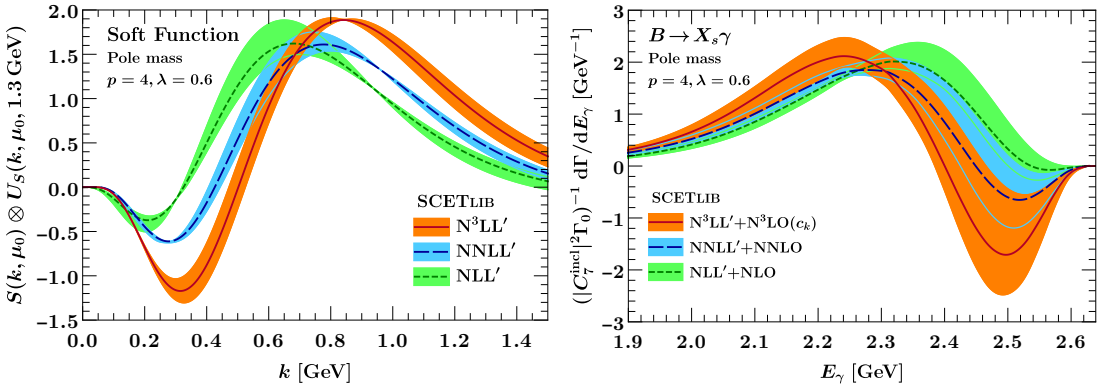


Figure 5.9.: The soft function (left panel) and the $B \to X_s\gamma$ photon energy spectrum (right panel) in the pole mass scheme. Small values of k in the soft function correspond to large E_γ in the spectrum. The pole-mass renormalon leads to the large negative dip at small k and large E_γ and an unstable position of the peak.

for the MSR mass scheme. In these plots the soft function is shown on the left, and the $B \to X_s\gamma$ photon energy spectrum is shown on the right. The shown soft function is $S(k, \mu_0) \otimes_k U_S(k, \mu_0, 1.3\,\mathrm{GeV})$, i.e. the soft evolution kernel $U_S(k, \mu_0, 1.3\,\mathrm{GeV})$ is used to evolve the soft function to the fixed scale $\mu = 1.3\,\mathrm{GeV}$. In this way, the μ_0-dependence cancels up to higher-order corrections. The scale μ_0 is varied within the ranges shown in eq. (5.38) to estimate the perturbative uncertainty in the soft function.

Figure 5.9 clearly shows that the soft function in the pole scheme suffers from a sizable renormalon ambiguity, which is intrinsic to the pole scheme, leading to a large negative deep before the peak. Moreover, the peak position varies significantly from one order to another, which reflects the instabilities in the first moments of the shape function. These features are also visible in the corresponding results for the photon energy spectrum. This behaviour of the pole scheme was already observed in ref. [23] up to NNLL′, and the results shown here demonstrate that these problems get even worse at $\mathrm{N^3LL'}$.

The predictions in the 1S mass scheme are presented in figure 5.10. Although the differential decay rate up to NNLL′ is somewhat stable, the prediction fails dramatically at $\mathrm{N^3LL'}$. In particular, the uncertainty band from scale variation is completely out of control. Adopting short-distance schemes for λ_1 and ρ_1 slightly improves the spectrum up to NNLL′+NNLO, but the picture at $\mathrm{N^3LL'}$+$\mathrm{N^3LO}(c_k)$ does not change. It also does not seem to be possible to substantially improve the convergence by adjusting the values of the short-distance subtraction coefficients $\delta\lambda_1^{(2)}$, $\delta\lambda_1^{(3)}$, and $\delta\rho_1^{(3)}$. Keeping the hard function unexpanded in the 1S scheme somewhat improves the picture up to NNLL′+NNLO, but it does not help at all with the bad behaviour at $\mathrm{N^3LL'}$+$\mathrm{N^3LO}(c_k)$.

The results in the MSR scheme are presented in figure 5.12. In comparison to the 1S scheme, the results are much more stable. In the first row, where only the b-quark mass is switched to the MSR scheme, the spectrum is still is very sensitive to the behavior

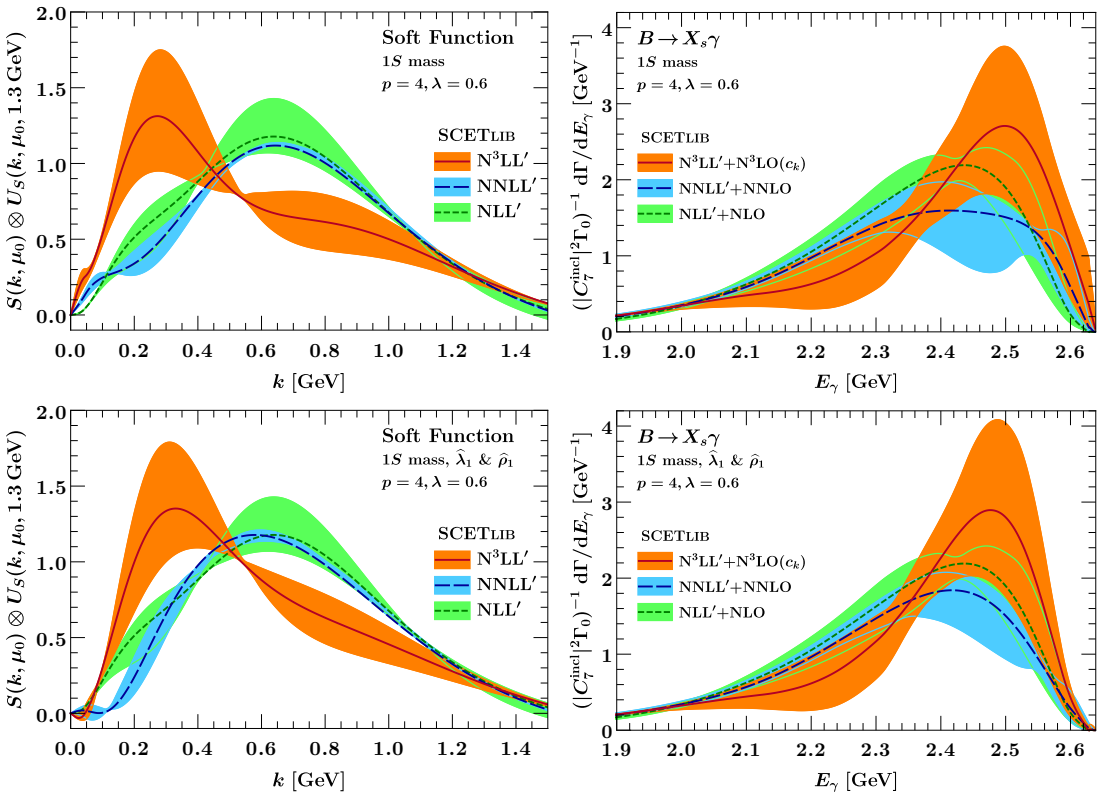


Figure 5.10.: The soft function (left panel) and the $B \to X_s\gamma$ photon energy spectrum (right panel) in the 1S mass scheme. Small values of k in the soft function correspond to large E_γ in the spectrum. The top panels use the pole scheme for λ_1 and ρ_1, while the bottom panels use a short-distance scheme. In both cases the spectrum, and in particular the position of its peak, are stable at NLL′+NLO and NNLL′+NNLO, but at $\mathrm{N^3LL'}$+$\mathrm{N^3LO}(c_k)$ the predictions start breaking down.

of the soft function at small k. This sensitivity is reflected in the uncertainty estimates from scale variation. As can be seen in the bottom row of figure 5.12, the subtractions of the subleading renormalons present in λ_1 and ρ_1 lead to a substantial improvement in the peak region of the spectrum, and the results exhibit excellent convergence between all orders. These results clearly lead to the conclusion that the MSR scheme is indeed a much more suitable mass scheme for the $B \to X_s\gamma$ spectrum beyond NNLL′.

To understand the reason for the breakdown of the 1S scheme at $\mathrm{N^3LL'}$ it is useful to examine the relation between the pole and the 1S mass schemes up to the 3-loop order:

$$m_b^{\mathrm{pole}} = m_b^{\mathrm{1S}} + R^{\mathrm{1S}}(\mu) \sum_{n=1}^{3} \sum_{m=0}^{n-1} c_{nm}^{\mathrm{1S}} \left[\frac{\alpha_s(\mu)}{4\pi}\right]^n \left[\ln \frac{\mu}{R^{\mathrm{1S}}(\mu)}\right]^m . \tag{5.42}$$

where $R^{\mathrm{1S}}(\mu) = C_F m_b^{\mathrm{1S}} \alpha_s(\mu)$ is the built-in infrared cutoff scale of the 1S scheme. The

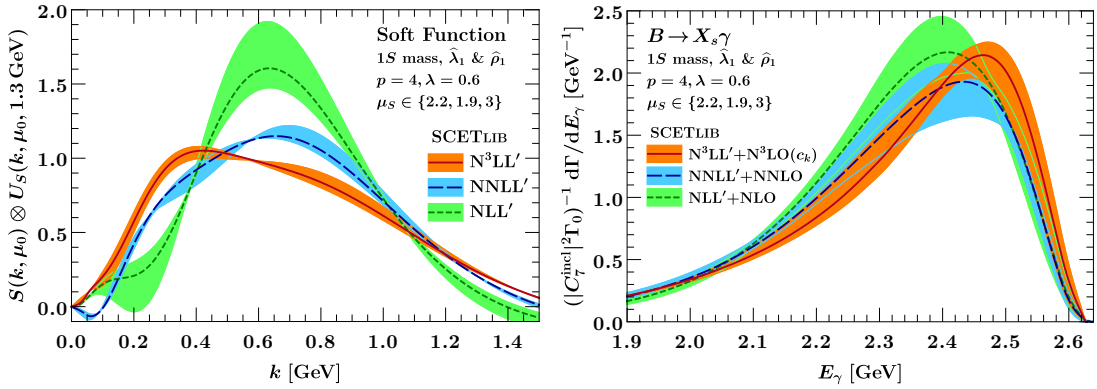


Figure 5.11.: The soft function (left panel) and the $B \to X_s\gamma$ photon energy spectrum (right panel) in the 1S mass scheme. Small values of k in the soft function correspond to large E_γ in the spectrum. A short-distance scheme is used for the hadronic parameters λ_1 and ρ_1. In contrast to the other plots, larger values are used for the soft scale: $\mu_0 \in \{2.2, 1.9, 3\}$ GeV. The larger soft scale amends the breakdown of 1S scheme at $\mathrm{N^3LL'}$+$\mathrm{N^3LO}(c_k)$, but is not compatible with the SCET scale hierarchy.

numerical values of the coefficients c_{nm} are

$$\begin{aligned}
c_{10}^{\mathrm{1S}} &= 2.0944,\\
c_{20}^{\mathrm{1S}} &= 135.438 - 10.2393 n_f,\\
c_{21}^{\mathrm{1S}} &= 92.1534 - 5.58505 n_f,\\
c_{30}^{\mathrm{1S}} &= 11398.2 - 1372.75 n_f + 38.9677 n_f^2,\\
c_{31}^{\mathrm{1S}} &= 7766.02 - 1077.92 n_f + 33.5103 n_f^2,\\
c_{32}^{\mathrm{1S}} &= 3041.06 - 368.614 n_f + 11.1701 n_f^2.
\end{aligned} \tag{5.43}$$

In contrast to the MSR scheme, where the R scale is a free parameter, in the 1S scheme the infrared scale $R^{\mathrm{1S}}(\mu)$ depends on the renormalization scale μ via the coupling constant. Hence its size increases when decreasing the scale μ, e.g. at the hard scale its value is $R^{\mathrm{1S}}(4.75\,\mathrm{GeV}) = 1.36\,\mathrm{GeV}$, whereas at the soft scale it is $R^{\mathrm{1S}}(\mu_S = 1.3\,\mathrm{GeV}) = 2.40\,\mathrm{GeV}$, which is almost twice the size of the soft scale itself. Such a large infrared scale violates the power counting of HQET, which is used to describe the heavy quarks in the B meson with the residual soft momenta $k \sim \Lambda_{\mathrm{QCD}}$ in the peak region.

The mismatch between the size of R^{1S} and $\mu_S \sim k$ not only breaks the power counting of the EFT description of the decay rate, but also spoils the renormalon subtraction in the soft function. This can be demonstrated numerically by considering the mass correction $\Delta m_b := m_b^{\mathrm{MSR}}(R) - m_b^{\mathrm{1S}}$ between the 1S and MSR schemes, which cannot involve a renormalon, because both 1S and MSR are short-distance schemes. Numerically, the

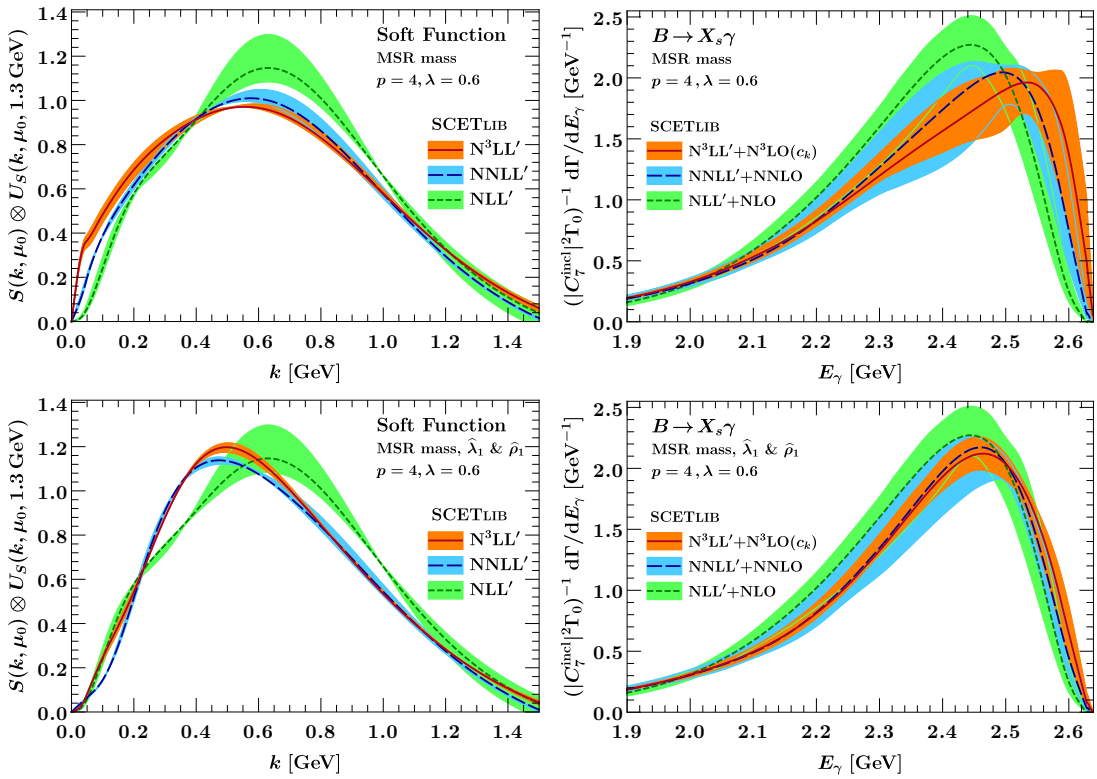


Figure 5.12.: The soft function (left panel) and the $B \to X_s\gamma$ photon energy spectrum (right panel) in the MSR mass scheme. Small values of k in the soft function correspond to large E_γ in the spectrum. The top panels use the pole scheme for λ_1 and ρ_1, while the bottom panels use a short-distance scheme. The convergence is significantly better than in the other mass schemes, and is further improved by adoption of short-distance schemes for λ_1 and ρ_1.

series for this mass correction at different renormalization scales is:

$$\begin{aligned}
\Delta m_b(R=\mu)\Big|_{\mu=4.2\,\mathrm{GeV}} &= -0.35\varepsilon - 0.12\varepsilon^2 - 0.04\varepsilon^3 \ [\mathrm{GeV}],\\
\Delta m_b(R=\mu)\Big|_{\mu=1.93\,\mathrm{GeV}} &= -0.15\varepsilon - 0.06\varepsilon^2 + 0.02\varepsilon^3 \ [\mathrm{GeV}],\\
\Delta m_b(R=\mu)\Big|_{\mu=1.3\,\mathrm{GeV}} &= -0.06\varepsilon - 0.06\varepsilon^2 + 0.10\varepsilon^3 \ [\mathrm{GeV}],
\end{aligned} \tag{5.44}$$

where $\varepsilon := 1$ is an auxiliary variable that is used to indicate the perturbative order of the corrections. The perturbative series in the first line seems to show a rather good convergence for Δm_b at the hard scale. However, in this regime the perturbative expansion for the 1S scheme contains large logarithms of the form $\ln(\mu/R^{\mathrm{1S}}(\mu))|_{\mu=4.7\,\mathrm{GeV}} \sim \ln(4.7/1.36)$. These logarithms are suppressed by $R^{\mathrm{1S}}(\mu = m_b^{\mathrm{1S}})/m_b^{\mathrm{1S}} = 1.36/4.7$ and are therefore harmless in the fixed-order expansion. For example, in the context of calculating the total decay rate, it is well-known that the 1S scheme provides a good description

for the b-quark mass.

It is possible to define a natural scale for the 1S scheme, which will be denoted $\mu_{1\mathrm{S}}$, at which all logarithms of the form $\ln(\mu_{1\mathrm{S}}/R^{1\mathrm{S}}(\mu_{1\mathrm{S}}))$ are resummed. This scale corresponds to the fixed point of the $R^{1\mathrm{S}}$ scale, where $R^{1\mathrm{S}}(\mu_{1\mathrm{S}}) = \mu_{1\mathrm{S}}$, which yields the numerical value of $\mu_{1\mathrm{S}} = 1.93\,\mathrm{GeV}$. The perturbative series for $\Delta m_b(\mu_{1\mathrm{S}})$ is shown in the second line of eq. (5.44). It shows the same good convergence as the first line, but with overall smaller corrections and a change of sign in the $\mathcal{O}(\varepsilon^3)$ coefficient.

Finally, the last line in eq. (5.44) shows the perturbative series for $\Delta m_b(\mu_S)$ at the soft scale $\mu_S = 1.3\,\mathrm{GeV}$. As one can see, the resulting series exhibits no convergence and breaks down at the 3-loop order. This behavior vividly explains the failure of the 1S mass scheme when used at the soft scale to remove the renormalon in the soft function.

Another interesting conclusion from the discussion above is that one can retain the use of the 1S scheme as soon as the actual soft scale in the problem is roughly of the same order as $R^{1\mathrm{S}} = \mu_{1\mathrm{S}} \sim 1.93\,\mathrm{GeV}$. To examine this hypothesis, figure 5.11 shows the results for the soft function and the photon energy spectrum in the 1S mass scheme, where the soft scale is chosen to have larger values, $\mu_S \in \{2.2, 1.9, 3\}\,\mathrm{GeV}$. Indeed the resulting spectrum exhibits a significant improvement at all orders compared to figure 5.10, and in particular the $\mathrm{N^3LL'}$+$\mathrm{N^3LO}(c_k)$ result is now much more well-behaved. However, in practice this setup is not really ideal since the soft scale is now much larger than Λ_{QCD} and the default soft scale $\mu_0 = 1.3\,\mathrm{GeV}$, which leads to rather large unresummed logarithms in the soft function. Consequently, the results in figure 5.11 do not reach the same level of stability as those in the MSR scheme in figure 5.12.

6. Inclusive decay $B \to X_u l \bar{\nu}$ at $\mathrm{N^3LL{+}NLO}$

6.1. Kinematics

The following kinematic variables will be used in the discussion of the semileptonic $B \to X_u l \bar{\nu}$ decay: The 4-momentum of the decaying B meson is $m_B v$, where v is a 4-vector with the normalization $v^2 = 1$, so that $v = (1, 0, 0, 0)$ in the rest frame of the B meson. The symbol q denotes the 4-momentum of the leptonic system, which is the sum of 4-momentum p_l of the charged lepton and the 4-momentum p_ν of the neutrino, and p_X the momentum of the remaining hadronic system X_u. Then the momentum conservation is written as $m_B v = p_X + q$. The lightcone coordinates of these 4-momenta are denoted by

$$q_+ \coloneqq n \cdot q, \qquad q_- \coloneqq \bar{n} \cdot q, \qquad p_X^+ \coloneqq n \cdot p_X, \qquad p_X^- \coloneqq \bar{n} \cdot p_X. \tag{6.1}$$

The reference lightcone momentum n is chosen in the direction of the hadronic final state X_u, such that $p_X = \frac{n}{2} p_X^- + \frac{\bar{n}}{2} p_X^+$, with $p_X^+ \leq p_X^-$. The lightcone coordinates are defined in the section 2.1.1. The energy of the charged lepton in the rest frame of the decaying B meson is denoted by $E_l \coloneqq v \cdot p_l$. In these variables, the kinematically allowed region of the phasespace is defined by the following constraints:

$$0 \leq q^- \leq 2E_l \leq q^+ \leq m_B, \tag{6.2}$$

and is shown in figure 6.1. These kinematic constraints, however, do not account for the

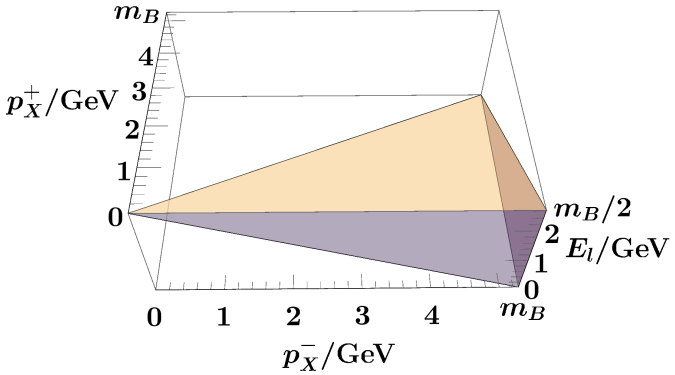


Figure 6.1.: Kinematically allowed phasespace for the semileptonic $B \to X_u l \bar{\nu}$ decay in coordinates (p_X^-, p_X^+, E_l).

fact that the invariant mass $m_X = \sqrt{p_X^+ p_X^-}$ of the final hadronic cannot be smaller than

the mass m_π of the lightest charged meson — the pion π. Because the theoretical model presented here is inclusive and does not describe the resonant structure of the spectrum, the constraint $m_X \geq m_\pi$ is neglected.

Following ref. [27], the notation $\bar{q}_\pm := q_\pm - 2E_l$ will also be used. Note that in the kinematically allowed region $\bar{q}_+$ is positive, but $\bar{q}_-$ is negative.

6.2. Decay rate and structure functions

The triple-differential decay rate in variables (p_X^+, p_X^-, E_l) can be written in terms of the hadronic tensor $W_{\mu\nu}$ and the leptonic tensor $L^{\mu\nu}$:

$$\frac{\mathrm{d}^3\Gamma}{\mathrm{d}p_X^+ \mathrm{d}p_X^- \mathrm{d}E_l} = \frac{G_F^2 |V_{ub}|^2}{8\pi^3}(q_+ - q_-) W_{\mu\nu} L^{\mu\nu}. \tag{6.3}$$

The leptonic tensor is defined as

$$L^{\alpha\beta} := (p_l)_\lambda (p_\nu)_\rho \,\mathrm{tr}[\gamma^\rho \gamma^\alpha \gamma^\lambda \gamma^\beta P_L] = 2(2p_l^{(\alpha} p_\nu^{\beta)} - g^{\alpha\beta}(p_l \cdot p_\nu) + i\varepsilon^{\alpha\beta\lambda\rho}(p_l)_\lambda (p_\nu)_\rho), \tag{6.4}$$

where the subscript $_\nu$ on the neutrino momentum p_ν should not be confused with the Lorentz index ν.

The hadronic tensor is defined as

$$W_{\mu\nu} := \frac{(2\pi)^3}{2m_B} \sum_{X_u} \int \mathrm{d}\Pi_{X_u} \delta(p_X + q - m_B v) \langle B | J_\mu^\dagger(0) | X_u \rangle \langle X_u | J_\nu(0) | B \rangle \tag{6.5}$$

$$\approx \frac{(2\pi)^3}{2m_B} \langle B | J_\mu^\dagger(0) \delta(\hat{p} + q - m_B v) J_\nu(0) | B \rangle = \frac{\langle B | T_{\mu\nu} | B \rangle}{2m_B}, \tag{6.6}$$

where

$$T_{\mu\nu}(q) := \frac{1}{2\pi} \,\mathrm{Disc}_{q^0} \int \mathrm{d}^4 x e^{iqx} \,\mathbf{T}[J_\mu^\dagger(0) J_\nu(x)]. \tag{6.7}$$

The sum $\sum_{X_u}$ is over all species of hadrons that can originate from a u-quark, and $\int \mathrm{d}\Pi_{X_u}$ is the integration over their Lorentz-invariant phasespace. The symbol $\mathbf{T}$ denotes the time-ordering of operators, and the current is $J^\mu(x) := \bar{u}(x)\gamma^\mu P_L b(x)$.

The discontinuity operator Disc is defined in appendix A.1. The hadronic tensor $W^{\mu\nu}$ is not real, but it is hermitian, in the sense that $W_{\mu\nu}^* = W_{\nu\mu}$.

The first equality sign in eq. (6.6) is written as an approximation to emphasize the fact that this step relies on local quark-hadron duality, which can fail to hold in some regions of the phasespace. This limitation will be further discussed in section 6.3.

It depends on the 4-momentum q. However, because the 4-momentum q can be written as $q = (n(m_B - p_X^-) + \bar{n}(m_B - p_X^+))/2$, the hadronic tensor $W^{\mu\nu}$ depends only on the lightcone coordinates $p_X^\pm$, and does not depend on charged-lepton energy E_l.

Following ref. [27], the hadronic tensor $W^{\mu\nu}$ is decomposed into 5 real structure functions $W_{1\ldots5}$:

$$W^{\mu\nu} =: -\frac{1}{2}(g - i\varepsilon)_\perp^{\mu\nu} W_1 - \frac{1}{2}(g + i\varepsilon)_\perp^{\mu\nu} W_2 - n^{(\mu}\bar{n}^{\nu)} W_3 + n^\mu n^\nu W_4 + \bar{n}^\mu \bar{n}^\nu W_5. \tag{6.8}$$

Here $(g \pm i\varepsilon)_\perp^{\mu\nu} \coloneqq g_\perp^{\mu\nu} \pm i\varepsilon_\perp^{\mu\nu}$. The sixth possible basis tensor, proportional to $n^{[\mu}\bar{n}^{\nu]}$, is not included, because $n^{[\mu}\bar{n}^{\nu]}L_{\mu\nu} = 0$, even for nonzero lepton or neutrino masses.

Inversely, the structure functions $W_{1\ldots5}$ are defined as the following projections of the hadronic tensor:

$$W_1 \coloneqq -\frac{1}{2}(g_\perp^{\mu\nu} + i\varepsilon_\perp^{\mu\nu})W_{\mu\nu}, \qquad W_2 \coloneqq -\frac{1}{2}(g_\perp^{\mu\nu} - i\varepsilon_\perp^{\mu\nu})W_{\mu\nu},$$
$$W_3 \coloneqq -\frac{1}{2}n^{(\mu}\bar{n}^{\nu)}W_{\mu\nu}, \qquad W_4 \coloneqq \frac{1}{4}\bar{n}^\mu\bar{n}^\nu W_{\mu\nu}, \qquad W_5 \coloneqq \frac{1}{4}n^\mu n^\nu W_{\mu\nu}. \tag{6.9}$$

Here the transverse metric $g_\perp^{\mu\nu}$ and the transverse Levi-Civita tensor $\varepsilon_\perp^{\mu\nu}$ are defined as:

$$g_\perp^{\mu\nu} \coloneqq g^{\mu\nu} - \frac{n^\mu\bar{n}^\nu}{2} - \frac{\bar{n}^\mu n^\nu}{2}, \qquad \varepsilon_\perp^{\mu\nu} \coloneqq \frac{1}{2}\varepsilon^{\mu\nu\lambda\rho}n_\lambda\bar{n}_\rho. \tag{6.10}$$

Both the hadronic tensor $W^{\mu\nu}$ and the structure functions W_i have mass dimension -1: $W^{\mu\nu} \sim W_i \sim m^{-1}$.

In terms of these structure functions and lightcone variables, the $B \to X_u l\bar{\nu}$ decay rate can be written as

$$\frac{\mathrm{d}^3\Gamma}{\mathrm{d}p_X^+\mathrm{d}p_X^-\mathrm{d}E_l} = \frac{G_F^2|V_{ub}|^2}{4\pi^3(q_+ - q_-)}$$
$$\times (q_+q_-(\bar{q}_-^2W_1 + \bar{q}_+^2W_2) - 2\bar{q}_+\bar{q}_-(q_+q_-W_3 + q_+^2W_4 + q_-^2W_5)). \tag{6.11}$$

Here and in the following the neutrino and charged-lepton masses are neglected. Although there are 5 structure functions, in the limit of massless leptons only 3 E_l-independent combinations of them appear in the decay rate, — $q_+q_-W_1$, $q_+q_-W_2$, and $q_+q_-W_3 + q_+^2W_4 + q_-^2W_5$. In principle, this fact can be used to reduce the number of structure functions to 3 by a judicious choice of basis tensors, as was for example done in ref. [47]. In this work, however, the tensor decomposition used in eq. (6.8) is preferred, because it results in somewhat simpler expressions for the structure functions W_i, with a more uniform asymptotic behavior. This choice of tensor decomposition will be further motivated in section 6.4.

Note that because $|\bar{q}_\pm| < q_+ - q_-$ the differential decay rate vanishes in the limit $(q_+ - q_-) \to 0$, and the factor $(q_+ - q_-)$ in the denominator in eq. (6.11) does not lead to a divergence.

Because the structure functions W_i do not depend on the charged-lepton energy E_l, the integration over E_l is easy to perform analytically:

$$\frac{\mathrm{d}^2\Gamma}{\mathrm{d}p_X^+\mathrm{d}p_X^-} \coloneqq \int_{q_-/2}^{q_+/2}\mathrm{d}E_l\frac{\mathrm{d}^3\Gamma}{\mathrm{d}p_X^+\mathrm{d}p_X^-\mathrm{d}E_l}$$
$$= \frac{G_F^2|V_{ub}|^2}{24\pi^3}(q_+ - q_-)^2(q_+q_-(W_1 + W_2 + W_3) + q_+^2W_4 + q_-^2W_5) \tag{6.12}$$

6.3. Power counting in different kinematic regions

The methods that are used to study the hadronic tensor $W^{\mu\nu}$ and the structure functions W_i are different depending on the hierarchy of sizes of kinematic variables, which is different in the different regions of the phasespace [27]. The distinguished kinematic regions, — the *local OPE* region, the *SCET* region, and the *resonance* region, — are illustrated in figure 6.2.

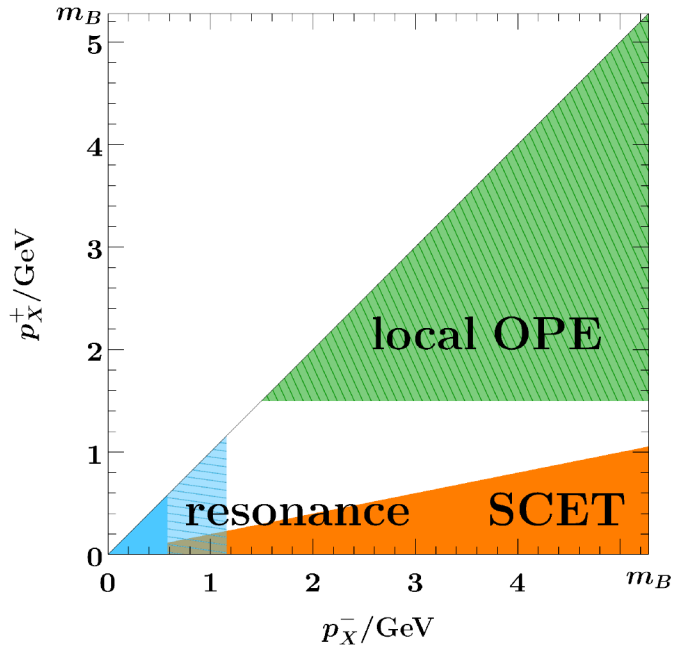


Figure 6.2.: Regions of the $B \to X_u l\bar{\nu}$ phasespace with distinct kinematic hierarchies, projected onto the (p_X^-, p_X^+) plane. The green, orange, and light-blue regions denote the local-OPE, SCET, and resonance regions, respectively. The solid light-blue triangle in the lower-left corner shows the deep resonance region, where $p^- = p_X^- - m_B + m_b$ is negative.

6.3.1. Local-OPE region

The local operator-product expansion (local-OPE) region is characterized by the hierarchy $\Lambda_{\rm QCD} \ll p_X^+ \lesssim p_X^- \lesssim m_B$. In this region an expansion in small parameter $\frac{\Lambda_{\rm QCD}}{p_X^+} \sim \frac{\Lambda_{\rm QCD}}{m_b}$ can be used. This expansion $\frac{\Lambda_{\rm QCD}}{m_b}$ is called the *heavy-quark expansion* (HQE), and it can be carried out systematically within the framework of Heavy-Quark Effective Theory (HQET) [3]. The operator $T_{\mu\nu}$, which defined in eq. (6.7), includes contributions of currents separated by the distance x, i.e. $T_{\mu\nu}$ is a *non-local* operator. However, the power counting of HQET can be used to demonstrate that in the local-OPE

region the typical spacetime separation of fields that are probed by the operator $T_{\mu\nu}$ is $x \sim \frac{1}{m_b}$. This is much smaller than the characteristic size of the B meson $\sim \frac{1}{\Lambda_{\rm QCD}} \gg \frac{1}{m_b}$. As a result, the non-local operator $T_{\mu\nu}$ can be expanded in terms of *local* operators $\mathcal{O}_i$, i.e. operators that involve only fields and derivatives of fields at a single point:

$$T_{\mu\nu}(q) = \sum_i C_{\mu\nu;i}^{\text{local OPE}}(q)\mathcal{O}_i. \tag{6.13}$$

The higher-dimension operators in this expansion are suppressed by inverse powers of m_b, and are usually formulated in terms of HQET field operators [3]. This is the essence of local operator-product expansion.

The matching coefficients $C_{\mu\nu;i}^{\text{local OPE}}$ can be determined perturbatively by calculating a sufficient number of partonic matrix elements of both sides of eq. (6.13). For example, the leading matching coefficient $C_{\mu\nu;0}^{\text{local OPE}}$ can be found by taking a b-quark with momentum $m_b v$ as the state. The only local operator at leading order in the HQE is $\mathcal{O}_0 \coloneqq \bar{b}\not{v}b$. The normalization of states is such that $\langle b|\mathcal{O}_0|b\rangle = 2m_b$ and $\langle B|\mathcal{O}_0|B\rangle = 2m_B$. Therefore

$$2m_b C_{\mu\nu;0}^{\text{local OPE}}(q) = \langle b|T_{\mu\nu}(q)|b\rangle + \mathcal{O}\left(\frac{\Lambda_{\rm QCD}}{m_b}\right). \tag{6.14}$$

As a result, at leading power in the HQE the hadronic tensor can be approximated as

$$W_{\mu\nu} = \frac{\langle B|T_{\mu\nu}(q)|B\rangle}{2m_B} = C_{\mu\nu;0}^{\text{local OPE}} + \mathcal{O}\left(\frac{\Lambda_{\rm QCD}}{m_b}\right). \tag{6.15}$$

The matrix element in eq. (6.14) is similar to the hadronic tensor, except that it is defined with b-quark instead of B-meson states. The tensor $C_{\mu\nu;0}^{\text{local OPE}}$ can be calculated perturbatively in full QCD. It has been calculated at NLO in ref. [48]. The NNLO results in the large-β_0 approximation have been computed in ref. [49], but they are not included in this work. The full calculation at NNLO has been carried out in ref. [50]. However, ref. [50] presents only some partially-integrated spectra of $B \to X_u l \bar{\nu}$ decay and their moments, which are not sufficient to reconstruct the full NNLO structure of the tensor $C_{\mu\nu;0}^{\text{local OPE}}$. Neither the analytic results nor the numerical code for the NNLO corrections are publicly available.

6.3.2. SCET region

The SCET region is defined by the hierarchy $\Lambda_{\rm QCD} \sim p_X^+ \ll p_X^- \lesssim m_B$. The condition $p_X^+ \ll p_X^-$ means that the final hadronic state X_u is a strongly boosted jet. The underlying b-quark decay happens at the energy scale of the order of the b-quark mass m_b, and the emission of this final state jet is accompanied by collinear and wide-angle soft radiation. As a result, SCET is the appropriate effective theory in this kinematic region, and the ratio p_X^+/p_X^- is a small expansion parameter. The factorization theorem of SCET factorizes the B-meson matrix element in the definition of the hadronic tensor in eq. (6.6) into hard, jet, and soft functions:

$$W_{\mu\nu}^{\rm SCET}(p^+, p^-) = H_{\mu\nu}(p^-) \cdot (p^- J(p^-) \otimes S)(p_X^+). \tag{6.16}$$

The jet function $p^- J(p^-)$ and the hard function $H_{\mu\nu}(p^-)$ depend on the label momentum $p^- = p_X^- - m_B + m_b$, which in the case of $B \to X_s\gamma$ was kinematically fixed to $p^- = m_b$. Other than that, the jet J and the soft S functions are the same as in the case of $B \to X_s\gamma$. The hard function $H_{\mu\nu}$ is given by

$$\begin{aligned} H^{\mu\nu}(p^-) &= \sum_{i,j=1}^{3} C_i(p^-) C_j(p^-) \,\mathrm{tr}\left[\bar{\Gamma}_j^\mu \frac{\bar{n}\!\!\!/}{4} \Gamma_j^\nu \frac{1+v\!\!\!/}{2}\right] \\ &= \frac{1}{4}\big(h^\mu h^\nu - C_1^2 (g^{\mu\nu} - i\varepsilon_\perp^{\mu\nu})\big), \end{aligned} \tag{6.17}$$

where

$$\Gamma_i^\mu \coloneqq P_R(\gamma^\mu \delta_{i1} + v^\mu \delta_{i2} + n^\mu \delta_{i3}), \qquad \bar{\Gamma}_i^\mu \coloneqq \gamma^0 (\Gamma_i^\mu)^\dagger \gamma^0, \tag{6.18}$$

and

$$h^\mu = (C_1 + C_3) n^\mu + C_2 v^\mu. \tag{6.19}$$

Here the $C_i(p^-)$ are the Wilson coefficients that arise from matching the $b \to u$ currents J_μ between QCD and SCET. They have been calculated up to NNLO [51–53]. The finite-charm-mass corrections at this order are also known [53], but they are not included in this work. Together with the known soft and jet function, the NNLO hard function enables the $B \to X_u l\bar{\nu}$ spectrum predictions in the SCET region at NNLO. The accuracy of such predictions is further improved by resummation at N^{3}LL order.

It is interesting to compare the local-OPE and the SCET expansions. Because the B meson does not contain any highly energetic partons, the B-meson state is assumed to belong to the soft sector of the factorized Fock space, and it therefore appears in the definition of the (hadronic) soft function in eq. (3.21). In contrast to local OPE, the (hadronic) soft function is defined as a B-meson matrix element of a non-local operator. As a result, the nonperturbative effects in SCET are parametrized by an entire function F, rather than by individual HQET parameters λ_1, ρ_1, etc., as in the local-OPE region. However, the current operators in the soft function are separated only along the lightcone direction n of the final-state jet X_u. In comparison, the unexpanded tensor $T^{\mu\nu}$, defined in eq. (6.7), includes currents with spacetime separation in all directions. In this way, the SCET approximation can be understood as a local operator-product expansion that is carried out in all directions but the lightcone direction n. This is related to the power counting in SCET: The relative power counting of momentum components is $(p_X^+, p_X^-, p_\perp) \sim (\lambda, 1, \sqrt{\lambda})$, where $\lambda = \Lambda_{\mathrm{QCD}}/m_B$ is the small expansion parameter of SCET. Because $p_X \cdot x = \frac{p_X^- x^+}{2} + \frac{p_X^+ x^-}{2} + p_\perp \cdot x_\perp$, the typical spacetime separation x of currents in $T^{\mu\nu}$ also has the same power counting $(x^+, x^-, x_\perp) \sim (\lambda, 1, \sqrt{\lambda})$. Therefore, $x^+ \ll x_\perp \ll x^-$, and the relevant operators can be assumed to be local with respect to small separations x^+ and $x_\perp$. As a result, an additional local operator-product expansion of the soft function S would recover the most singular terms of the local OPE. Such an additional local operator-product expansion of the soft function S corresponds to the

moments expansion of the leading-power shape function F [23]:

$$F(k) = \sum_{i=0}^{\infty} \frac{(-1)^n}{n!} \delta^{(n)}(k) M_n[F], \qquad M_n[F] \coloneqq \int_0^{+\infty} \mathrm{d}k k^n F(k). \tag{6.20}$$

This moments expansion is also sometimes referred to as the *multipole expansion*. When the shape function F is approximated by its first moment $F(k) \approx \delta(k - m_B + m_b)$, the SCET result $W_{\mu\nu}^{\mathrm{SCET}}$ reproduces the most singular terms of the leading-power local-OPE result $C_{\mu\nu;0}^{\text{local OPE}}$. The most singular terms of $C_{\mu\nu;0}^{\text{local OPE}}$ behave as $\mathcal{L}_n(p^+, \mu)$ in the limit $p^+ = p_X^+ - m_B + m_b \to 0$. This is the correspondence between SCET and local-OPE results.

6.3.3. Resonance region

Finally, the kinematic hierarchy in the *resonance* region is $\Lambda_{\mathrm{QCD}} \sim p_X^+ \lesssim p_X^- \ll m_B$. This means that the invariant mass $m_X = \sqrt{p_X^+ p_X^-}$ of the final hadronic state X_u is small, of order $m_X \sim \Lambda_{\mathrm{QCD}}$. Consequently, the resonant structure of the spectrum, — i.e. the presence of individual resonances, — π, ρ, ω, η, etc., — is important. This means that the local quark-hadron duality does not hold in this region, and therefore that the inclusive approach, which is used throughout this work, is not applicable. However, the inclusive approximation can still be used to predict decay rates integrated over a sufficiently large phasespace region, that may include the resonance region.

6.4. Matching of different kinematic regions

It is desirable to have a triple-differential theoretical prediction that can be used in all regions of $B \to X_u l\bar{\nu}$ phasespace. The predictions need to reduce to local-OPE and SCET predictions in the corresponding regions. They cannot be expected to be reliable in the resonance region, where the local quark-hadron duality breaks down, but they need to produce meaningful results when integrated over a sufficiently large phasespace region that includes the resonance region. These goals are achieved by combining the local-OPE and SCET predictions as described in the following.

To begin with, the *resummed* contribution $W_{\mu\nu}^{\mathrm{resum}}$ is defined as the SCET prediction $W_{\mu\nu}^{\mathrm{SCET}}$ with the shape function $\mathcal{F}$ approximated by its first moment $\mathcal{F}(k) \approx \delta(k - m_B + m_b)$. Equivalently, it is the perturbative part of the SCET prediction, which factorizes it as

$$W_{\mu\nu}^{\mathrm{SCET}}(p_X^+, p_X^-) =: \int_0^{p_X^+} \mathrm{d}k \mathcal{F}(k) W_{\mu\nu}^{\mathrm{resum}}(p_X^+, p_X^- - m_B + m_b). \tag{6.21}$$

The *singular* contribution $W_{\mu\nu}^{\mathrm{s}}$ is obtained from the resummed contribution $W_{\mu\nu}^{\mathrm{resum}}$ in the case when the resummation is turned off, which can be achieved by setting all

renormalization scales to the same value $\mu_H = \mu_J = \mu_S = \mu$. As already mentioned, the singular contribution $W^{\rm s}_{\mu\nu}$ corresponds to the most singular terms in the local-OPE prediction $C^{\text{local OPE}}_{\mu\nu;0}$. The remaining terms are called the *nonsingular* contribution $W^{\rm ns}_{\mu\nu} \coloneqq C^{\text{local OPE}}_{\mu\nu;0} - W^{\rm s}_{\mu\nu}$. Theses contributions, — $W^{\rm resum}_{\mu\nu}, W^{\rm s}_{\mu\nu}, C^{\text{local OPE}}_{\mu\nu;0}, W^{\rm ns}_{\mu\nu}$, — are generalized functions of partonic kinematic variables $p^+ = p_X^+ - m_B + m_b$ and $p^- = p_X^- - m_B + m_b$. In these variables, the kinematic constraints $0 \le p_X^+ \le p_X^- \le m_B$ become $-(m_B - m_b) \le p^+ \le p^- \le m_b$, although these functions are inherently zero when $p^+ < 0$ or $p^- < 0$. This subregion $p^+ \ge 0$ of the full $B \to X_u l\bar{\nu}$ phasespace corresponds to the phasespace of a decaying free b-quark, and will therefore be called the *partonic* phasespace. It is convenient to parametrize the partonic phasespace using the dimensionless kinematic variables $x \coloneqq p^+/p^-$, $y \coloneqq p^-/m_b$, then $0 \le x \le 1$ and $0 \le y \le 1$. The partonic phasespace and the dimensionless coordinates x, y are illustrated in figure 6.3. The limit $y = 1$, or, equivalently, $p^- = m_b$, corresponds to kinematics of $B \to X_s\gamma$, where the role of the variable x is $x = 1 - \frac{2E_\gamma}{m_b}$. For both $B \to X_s\gamma$ and $B \to X_u l\bar{\nu}$ the limit $x \to 0$ is the singular, SCET limit, which can be referred to as *peak* region. Analogously, the limit $x \to 1$ is the *tail* region. The limit $y \to 0$ corresponds to the resonance region. In these dimensionless variables, the linear combinations of

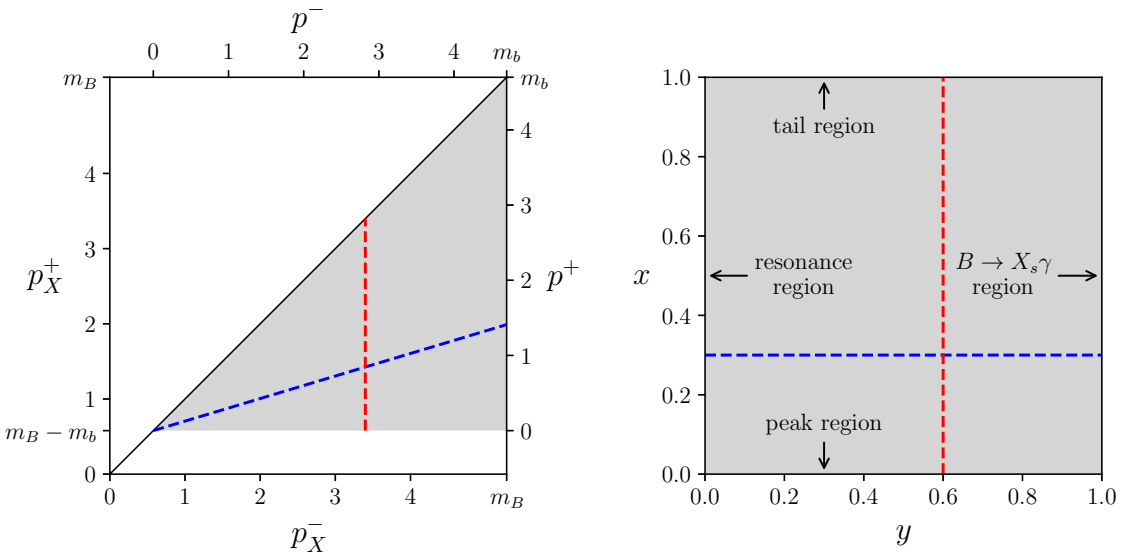


Figure 6.3.: The partonic phasespace as a subregion of the full, hadronic $B \to X_u l\bar{\nu}$ phasespace. The left panel shows the hadronic phasespace in the coordinates p_X^-, p_X^+ and in the coordinates p^-, p^+. The shaded regions on both panels show the partonic phasespace, i.e. the kinematically allowed region in the decay of a free b-quark. The right panel shows the partonic phasespace in coordinates x, y. The dashed blue and red lines illustrate lines $x = 0.3$ and $y = 0.6$, respectively.

nonsingular structure functions $W_i^{\rm ns}$ at NLO can be written rather compactly:

$$\begin{aligned}
p^-(W_1^{\rm ns} + 2W_4^{\rm ns}) &= \frac{\alpha_s C_F}{8\pi}\Big(y^2 + 8 - 8(3-x)L_1(x)\Big) + \mathcal{O}(\alpha_s^2),\\
p^-(W_1^{\rm ns} - 2W_4^{\rm ns}) &= \frac{\alpha_s C_F}{8\pi} y\Big(8 - y + 4(2+4x-xy)L_2(x)\Big) + \mathcal{O}(\alpha_s^2),\\
p^-(W_2^{\rm ns} + 2W_5^{\rm ns}) &= \frac{\alpha_s C_F}{8\pi}\Big(2 + x\big(y^2 + 8 + 8(1+x)L_1(x)\big)\Big) + \mathcal{O}(\alpha_s^2),\\
p^-(W_2^{\rm ns} - 2W_5^{\rm ns}) &= \frac{\alpha_s C_F}{8\pi} yx\Big(-4 + y + 2(4+2x-xy)\big(1 + 2xL_2(x)\big)\Big) + \mathcal{O}(\alpha_s^2),\\
p^- W_3^{\rm ns} &= \frac{\alpha_s C_F}{8\pi} y\Big(2 - (3+3x-xy)\big(1 + 2xL_2(x)\big)\Big) + \mathcal{O}(\alpha_s^2).
\end{aligned} \tag{6.22}$$

The factor p^- on the left-hand side compensates the mass dimension -1 of the structure functions $W_i^{\rm ns}$. Here and in the following the functions L_n are defined as

$$L_n(x) \coloneqq \frac{1}{(1-x)^{n+1}}\left(\ln x + \sum_{k=1}^{n}\frac{(1-x)^k}{k}\right), \tag{6.23}$$

and they have the following asymptotics:

$$L_n(x) = \ln x + \mathcal{O}(x) \quad \text{as } x \to 0, \quad L_n(1-x) = -\frac{1}{n+1} + \mathcal{O}(1-x) \quad \text{as } x \to 1. \tag{6.24}$$

The structure functions $W_i^{\rm s,ns}$ for the singular and nonsingular parts of the hadronic tensor $W_{\mu\nu}$ are defined analogously to the full structure functions W_i, using the same tensor projections as in eq. (6.8). For comparison, the singular structure functions $W_i^{\rm s}$ at NLO are

$$\begin{aligned}
p^-(W_1^{\rm s} + 2W_4^{\rm s}) &= \mathcal{L}_{-1}(x) - \frac{\alpha_s C_F}{\pi}\bigg(\mathcal{L}_{-1}(x)\Big(1 + 2\zeta_2 + \mathrm{Li}_2(1-y) + \frac{3}{2}\ln y\Big)\\
&\qquad + \frac{7}{4}\mathcal{L}_0(x) + \mathcal{L}_1(x)\bigg) + \mathcal{O}(\alpha_s^2),\\
p^-(W_1^{\rm s} - 2W_4^{\rm s}) &= -\frac{\alpha_s C_F}{4\pi}\mathcal{L}_{-1}(x)\big(1 + 2yL_0(y)\big) + \mathcal{O}(\alpha_s^2),\\
p^- W_2^{\rm s} &= 0,\\
p^- W_5^{\rm s} &= \mathcal{O}(\alpha_s^2),\\
p^- W_3^{\rm s} &= -\frac{\alpha_s C_F}{8\pi}\mathcal{L}_{-1}(x)\big(1 + yL_1(y)\big) + \mathcal{O}(\alpha_s^2).
\end{aligned} \tag{6.25}$$

The distributions $\mathcal{L}_n$ are defined in appendix C.1. The NNLO singular contributions are known, but are not shown here for brevity. They are given in appendix E.4. The structure function W_2 is purely nonsingular at all orders in α_s, and the series for the singular structure function $W_5^{\rm s}$ starts at order $\mathcal{O}(\alpha_s^2)$.

It is remarkable that at NLO all the nonsingular structure functions $W_i^{\rm ns}$ are polynomials in y, while the singular structure functions involve the dilogarithm $\mathrm{Li}_2(1-y)$. This

is related to the fact that all the virtual contributions are included in the hard function, and give rise to exclusively singular corrections. As a result, only single-real-emission diagrams contribute to the nonsingular contributions at NLO. These diagrams do not include loops at this order, and the resulting contributions involve only functions of low transcendentality level. Of course, at the next order virtual-real diagrams do give rise to nonsingular contributions, and therefore $W_i^{\rm ns}$ cannot be expected to be just polynomials of y at NNLO.

It is interesting to consider the asymptotics of singular $W_i^{\rm s}$ and nonsingular $W_i^{\rm ns}$ structure functions in different limits. As expected, in the peak region $x \to 0$ at any fixed y the nonsingular structure functions behave as $\sim \ln^n x$ and either tend to a constant or diverge at most logarithmically. Therefore, in this region they are suppressed by one power of x relative to singular contributions $W_i^{\rm s}$.

In the tail region $x \to 1$ both singular and $W_i^{\rm s}$ and nonsingular $W_i^{\rm ns}$ structure functions are finite, and are of the same order. In this region there's some numerical cancellation between singular and nonsingular structure functions:

$$\begin{aligned} p^-(W_1^{\rm s} + 2W_4^{\rm s}) &= \frac{\alpha_s C_F}{4\pi}\left(-7 - \frac{9}{3}(1-x)\right) + \mathcal{O}((1-x)^2), \\ p^-(W_1^{\rm ns} + 2W_4^{\rm ns}) &= \frac{\alpha_s C_F}{4\pi}\left(8 + \frac{y^2}{2} + \frac{14}{3}(1-x)\right) + \mathcal{O}((1-x)^2). \end{aligned} \tag{6.26}$$

The other combinations of singular structure functions $W_i^{\rm s}$ vanish at $x > 0$.

In the resonance region, as $y \to 0$ the coefficients of $\mathcal{L}$-distributions in the singular structure functions $W_i^{\rm s}$ diverge as $\ln^n y$. Curiously, the nonsingular structure functions at NLO have a finite limit as $y \to 0$ at any fixed x. On the other hand, this limit is different for different x, so the two-dimensional limit $\lim_{(p^+,p^-)\to(0,0)} W_i^{\rm ns}$ does not exist for $i = 1, 2, 3, 4$. Nevertheless, at NNLO and higher orders the nonsingular structure functions $W_i^{\rm ns}$ are expected to diverge as $\ln^n y$ in the limit $y \to 0$. In any case, both singular and nonsingular contributions are integrable in the resonance region, in the sense that for any two-dimensional region Ω which includes the point $p^+ = p^- = 0$ the integrals

$$\int_\Omega \mathrm{d}p^+ \mathrm{d}p^- W_i^{\rm resum}(p^+, p^-) \qquad \text{and} \qquad \int_\Omega \mathrm{d}p^+ \mathrm{d}p^- W_i^{\rm ns}(p^+, p^-) \tag{6.27}$$

are convergent.

There is some freedom in the definition of singular and nonsingular contributions: The expression for the triple-differential decay rate in eq. (6.11) includes kinematic factors involving $q_+ = m_B - p_X^+ = m_b - p^+$. In the SCET expansion these factors can be either expanded in small p^+/p^-, or kept exactly. This choice is inconsequential in the peak region $x \to 0$, where the singular and the nonsingular are parametrically separated. However, different choices lead to different definitions of nonsingular structure functions $W_i^{\rm ns}$, with different behavior in the tail $x \to 1$ region. Therefore, this could have an impact on the matching procedure.

This issue is related to the definition of structure functions W_i. In eq. (6.11) all kinematic factors that originate from phasespace integrations and the leptonic tensor

$L_{\mu\nu}$ are kept exactly. Furthermore, the chosen basis tensors in the hadronic tensor decomposition in eq. (6.8) do not depend on kinematic variables, although in general they could. As a result, the only object that is expanded in the SCET region is the hadronic tensor $W_{\mu\nu}$ itself. This is a deliberate choice, which has some advantages and disadvantages. One advantage is that the expressions for the nonsingular structure functions $W_i^{\rm ns}$ in eq. (6.22) are relatively simple. Another feature is that in the tail region $x \to 1$ both singular and nonsingular structure functions are constants of the same order of magnitude as their sum. When the basis tensors in the hadronic tensor decomposition in eq. (6.8) depend on p^+, this is typically not the case. For example, if the factors of q_+^n that occur in eq. (6.11) are absorbed into the definition of structure functions, the sum of singular and nonsingular structure functions $W_i^{\rm s} + W_i^{\rm ns}$ vanishes as $(1-x)^n$ as $x \to 1$, while the singular and nonsingular structure functions are of order ~ 1. This means that there would be a delicate cancellation between singular and nonsingular in the tail region, much stronger than the one shown in eq. (6.26). This is undesirable. As already mentioned, the main disadvantage of the approach adopted in this work is that all 5 structure functions need to be matched, while it is well-known that only 3 linear combinations of them appear in the final decay rate formula in eq. (6.11). However, if the structure functions were defined as these linear combinations, as they are for example in ref. [47], the nonsingular structure functions $W_i^{\rm ns}$ would be noticeably more complicated. In particular, they would no longer be just polynomials in y at NLO.

In order to have a coherent description of the spectrum over the entire phasespace, the resummed and nonsingular contributions, as well as nonperturbative effects of the shape function, need to be accounted for. To this end, the sum of resummed and nonsingular contributions will be called the *partonic tensor* $W_{\mu\nu}^{\rm partonic}$:

$$W_{\mu\nu}^{\rm partonic}(p^+, p^-) \coloneqq \theta(p^- - p^+)\big(W_{\mu\nu}^{\rm resum}(p^+, p^-) + W_{\mu\nu}^{\rm ns}(p^+, p^-)\big). \tag{6.28}$$

Although the local-OPE result $C_{\mu\nu;0}^{\text{local OPE}}$ is inherently zero in the region where $p^+ > p^-$, the SCET result $W_{\mu\nu}^{\rm resum}$ formally has support in the region $0 \le p^+ < +\infty$. In other words, the SCET result $W_{\mu\nu}^{\rm resum}$ formally does not include a θ-function $\theta(p^- - p^+)$. This is a consequence of the fact that in the SCET power counting $p^+ \ll p^-$, and therefore the phasespace border $p^+ = p^-$ is not visible in SCET. Nevertheless, the partonic tensor $W_{\mu\nu}^{\rm partonic}$ is expected to have support only within the partonic phasespace $0 \le p^+ \le p^- \le m_b$. In particular, it is expected to reproduce the local-OPE result $C_{\mu\nu;0}^{\text{local OPE}}$ when the resummation is turned off. For this reason, an additional θ-function $\theta(p^- - p^+)$ was included in the definition of the partonic tensor $W_{\mu\nu}^{\rm partonic}$ in eq. (6.28).

Finally, the full hadronic tensor $W_{\mu\nu}$ is approximated by the convolution of the shape function $\mathcal{F}$ with the partonic tensor $W_{\mu\nu}^{\rm partonic}$:

$$W_{\mu\nu}(p_X^+, p_X^-) \approx \int_0^{p_X^+} \mathrm{d}k \mathcal{F}(k) W_{\mu\nu}^{\rm partonic}(p_X^+ - k, p_X^- - m_B + m_b). \tag{6.29}$$

Here the shape function $\mathcal{F}$ is convolved with both the resummed $W_{\mu\nu}^{\rm resum}$ and the nonsingular $W_{\mu\nu}^{\rm resum}$ contributions. This former is derived in SCET, while the latter is just

an ansatz. This defines a prescription for the matching of SCET and local-OPE regions. In the limit $p^- = m_b$ this prescription agrees with the one used for $B \to X_s\gamma$.

If the resummation is turned off, in the local-OPE region the shape function can be approximated by a delta-function $\mathcal{F}(k) \approx \delta(k - m_B + m_b)$. In this case, the eq. (6.29) reduces to the local-OPE result $C^{\text{local OPE}}_{\mu\nu;0}$, up to small power corrections, which are related to higher moments of the shape function $\mathcal{F}(k)$. The only noteworthy artifact of the matching in eq. (6.29) is that the approximated hadronic tensor $W_{\mu\nu}$ is formally zero in the deep resonance region, where $p_X^- < m_B - m_b$, or, equivalently, $p^- < 0$. However, when eq. (6.29) is integrated over a sufficiently large region, which may include the resonance region, the local-OPE predictions are nevertheless reproduced:

$$\begin{aligned}
&\int_0^{p^+_{X\text{cut}}} \mathrm{d}p_X^+ \int_0^{p^-_{X\text{cut}}} \mathrm{d}p_X^- W_{\mu\nu}(p_X^+, p_X^-) \\
&\approx \int_0^{p^-_{X\text{cut}}} \mathrm{d}p_X^- \int_0^{p^+_{X\text{cut}}} \mathrm{d}p_X^+ W^{\text{partonic}}_{\mu\nu}(p_X^+ - m_B + m_b, p_X^- - m_B + m_b) \\
&= \int_0^{p^-_{\text{cut}}} \mathrm{d}p^- \int_0^{p^+_{\text{cut}}} \mathrm{d}p^+ W^{\text{partonic}}_{\mu\nu}(p^+, p^-),
\end{aligned} \tag{6.30}$$

where $p^\pm_{\text{cut}} = p^\pm_{X\text{cut}}$.

6.5. Profile functions and perturbative uncertainty

The factorization theorem of SCET provides an elegant way to resum singular contributions $W^{\text{s}}_{\mu\nu}$. The resummed results are expected to be more reliable in the SCET region, because resummation exponentiates towers of logarithms $\ln(p^-/p^+)$, which are expected to be large in the SCET region, where $p^+ \ll p^-$. Outside of the SCET region, however, this kind of resummation is not appropriate. In fact, resummation of singular terms in the local-OPE region would spoil the delicate cancellation between singular and non-singular terms that was shown in eq. (6.26). The local-OPE predictions would not be reproduced in this case.

Although the theoretical predictions for the hadronic tensor $W_{\mu\nu}$ depend on the renormalization scale μ, this dependence is an artifact of the truncation of perturbative series, and decreases at higher perturbative orders. The SCET factorization ingredients, — the hard function $H_{\mu\nu}$, the jet function J, and the soft function S, — individually do depend on the renormalization scale. In the SCET approach, resummation is achieved by evaluating fixed-order perturbative series of each of these ingredients at the appropriate scale μ_H, μ_J, μ_S, respectively, and solving the RGE equations to bring all ingredients to the common renormalization scale μ. The dependence of the resummed result

$$W^{\text{resum}}_{\mu\nu}(\mu_H, \mu_J, \mu_S, \mu) \tag{6.31}$$

on the scales μ_H, μ_J, μ_S cancels order-by-order in perturbation theory. If the RGE equations are solved exactly, which is in principle possible, the dependence on the common scale μ cancels exactly. In practice, however, the RGE equations are solved approximately [54], leading to some residual dependence on the scale μ. The appropriate scales in the SCET region are suggested by the SCET power counting: $\mu_H \sim m_b$, $\mu_S \sim \Lambda_{\rm QCD}$, $\mu_J \sim \sqrt{\mu_H \mu_S}$. On the other hand, if all these renormalization scales are set to the same value $\mu = \mu_H = \mu_J = \mu_S$, the evolution kernels vanish and the resummed results reduce to fixed-order results. The nonsingular terms $W^{\rm ns}_{\mu\nu}(\mu_{\rm ns})$ can be calculated using a fixed-order series, the renormalization scale of which is denoted $\mu_{\rm ns}$. The dependence of $W^{\rm ns}_{\mu\nu}(\mu_{\rm ns})$ on $\mu_{\rm ns}$ cancels order-by-order.

To have a coherent description of the $B \to X_u l\bar{\nu}$ decay rate over the entire phasespace it is desirable to use resummed predictions in the SCET region, to use non-resummed, fixed-order results in the local-OPE region, and to gradually turn the resummation off in the transition between these two regimes. The resummation of singular terms is turned on when the scales μ_H, μ_J, μ_S, μ are separated according to their SCET hierarchy, and is turned off when they are set to the same value. Therefore, the smooth transition between resummed and non-resummed predictions can be achived by picking the renormalization scales $\mu_H, \mu_J, \mu_S, \mu, \mu_{\rm ns}$ to be functions of the phasespace point. These functions are called the *profile functions*.

Because the hadronic tensor $W_{\mu\nu}$ depends only on the kinematic variables p_X^+ and p_X^-, and does not depend on the charged-lepton energy E_l, the profile functions are chosen to be functions of only (p_X^+, p_X^-) as well. The choice of the renormalization scales, or, equivalently, the profile functions, is to some extent arbitrary, but it is motivated by the relevant energy scales. The hard scale $\mu_H \sim m_b$ is chosen to be the same over the entire phasespace. For simplicity, the jet scale μ_J and nonsingular scale $\mu_{\rm ns}$ are chosen to be functions of hard and soft scales μ_H, μ_S. Explicitly,

$$\mu_H = e_H m_b, \qquad \mu_J = \mu_H^{(1+e_J)/2} \mu_S^{(1-e_J)/2}, \qquad \mu_{\rm ns} = \mu_H^{(3+e_{\rm ns})/4} \mu_S^{(1-e_{\rm ns})/4}, \tag{6.32}$$

with some parameters $e_H, e_J, e_{\rm ns}$. This choice ensures that the relative hierarchy of the hard μ_H, jet μ_J, and soft μ_S scales is always preserved. The common scale μ is chosen to always coincide with the jet scale μ_J, which simplifies the resummation a little, because the jet function evolution drops out. The soft scale μ_S in the resummation regime should be chosen as close as possible to $\Lambda_{\rm QCD}$, while remaining in the perturbative regime. In the fixed-order regime the soft scale μ_S should be equal to the hard scale μ_H, which collapses the scale hierarchy to a single scale. Therefore, the soft scale goes between some minimal scale μ_0 in the resummation regime and the hard scale μ_H in the fixed-order regime:

$$\mu_S(p_X^+, p_X^-) = \mu_H + (\mu_0 - \mu_H) f_{\rm resum}(p_X^+, p_X^-), \tag{6.33}$$

where the resummation activation function $f_{\rm resum}$ returns 1 in the resummation region, returns 0 in the fixed-order region, and smoothly interpolates the two regions. The resummation and fixed-order regions are assumed to be determined by inequalities $p_X^+ < p_{X;\rm resum}^+(p_X^-)$ and $p_X^+ > p_{X;\rm FO}^+(p_X^-)$, respectively. Here $p_X^{+\rm resum}$ and $p_X^{+\rm FO}$ are

functions of p_X^-, which are for simplicity chosen to be polynomials with 2 coefficients:

$$p_X^{+\text{resum}}(p_X^-) \coloneqq p_X^{\text{resum}} + e^{\text{resum}} p_X^-, \qquad p_X^{+\text{FO}}(p_X^-) \coloneqq p_X^{\text{FO}} + e^{\text{FO}} p_X^-. \tag{6.34}$$

To interpolate between the resummation and fixed-order regimes, the easing function f_θ is used along each line of constant p_X^-:

$$f_{\text{resum}}(p_X^+, p_X^-) \coloneqq f_\theta \left(\frac{p_X^+ - p_X^{+\text{resum}}(p_X^-)}{p_X^{+\text{FO}}(p_X^-) - p_X^{+\text{resum}}(p_X^-)} \right). \tag{6.35}$$

The quadratic easing function $f_\theta(x)$ is 0 for $x \leq 0$, is 1 for $x \geq 1$, and smoothly interpolates between these two values in the region $0 \leq x \leq 1$. It is defined in eq. (5.37).

In the region where the singular and nonsingular contributions $W^{\text{s}}_{\mu\nu}, W^{\text{ns}}_{\mu\nu}$ are similar in size the resummation should be turned off, so that the local-OPE predictions are reproduced. On the other hand, in the region where the singular terms $W^{\text{s}}_{\mu\nu}$ are dominant they should be resummed. Therefore, the fractional contribution $\mathrm{d}\Gamma^{\text{s}}/(\mathrm{d}\Gamma^{\text{s}} + \mathrm{d}\Gamma^{\text{ns}})$ of the singular terms $W^{\text{s}}_{\mu\nu}$ to the decay rate as a function of (p_X^+, p_X^-) can be examined to determine where the resummation should and should not be used. This fractional contribution is shown in figure 6.4. This exercise was performed using fixed-order NLO results with $\mu = m_b$ in the MSR mass scheme with the default shape function model with $p = 4$ and $\lambda = 0.6\,\text{GeV}$ (see section 3.1). The transition lines $p_X^{+\text{resum}}, p_X^{+\text{FO}}$ are chosen to lie roughly where the ratio $\mathrm{d}\Gamma^{\text{s}}/(\mathrm{d}\Gamma^{\text{s}} + \mathrm{d}\Gamma^{\text{ns}})$ is constant $3/4$ and $-4/5$, respectively. The corresponding values of parameters $e^{\text{resum}}, p_X^{\text{resum}}, e^{\text{FO}}, p_X^{\text{FO}}$ are shown in eq. (6.36). The profile functions with the chosen central values of profile function parameters are illustrated in figure 6.5. At the phasespace boundary $p_X^- = m_B$ these profile functions are completely analogous to the profile functions used for $B \to X_s\gamma$, as described in section 5.4, except with different position of boundaries of resummation and fixed-order regimes.

Variations of profile function parameters induce variations in the chosen renormalization scales, which in turn result in variations of theoretical predictions. The latter will be taken as an estimate of the perturbative uncertainty. A feature of the profile-function approach is that the hierarchy between hard, jet, and soft scales is preserved for all variations.

The central values and variations of the profile function parameters are chosen as follows:

$$\begin{aligned}
e_H &\in \{1, 1/2, 2\}, \\
e_J &\in \{0, -1/3, +1/3\}, \\
e_{\text{ns}} &\in \{0, -1, +1\}, \\
\mu_0 &\in \max(1, e_H) \cdot \{1.3, 1.1, 1.8\}\,\text{GeV}, \\
(e^{\text{resum}}, p_X^{\text{resum}}) &\in \{(0.044, 0.67\,\text{GeV}), (0.011, 0.64\,\text{GeV}), (0.077, 0.70\,\text{GeV})\}, \\
(e^{\text{FO}}, p_X^{\text{FO}}) &\in \{(0.24, 0.83\,\text{GeV}), (0.21, 0.80\,\text{GeV}), (0.27, 0.86\,\text{GeV})\}.
\end{aligned} \tag{6.36}$$

The first value in each set is the central, default value. The central values and variations of parameters $e_H, e_J, \mu_0, e_{\text{ns}}$ are the same as in the case of $B \to X_s\gamma$.

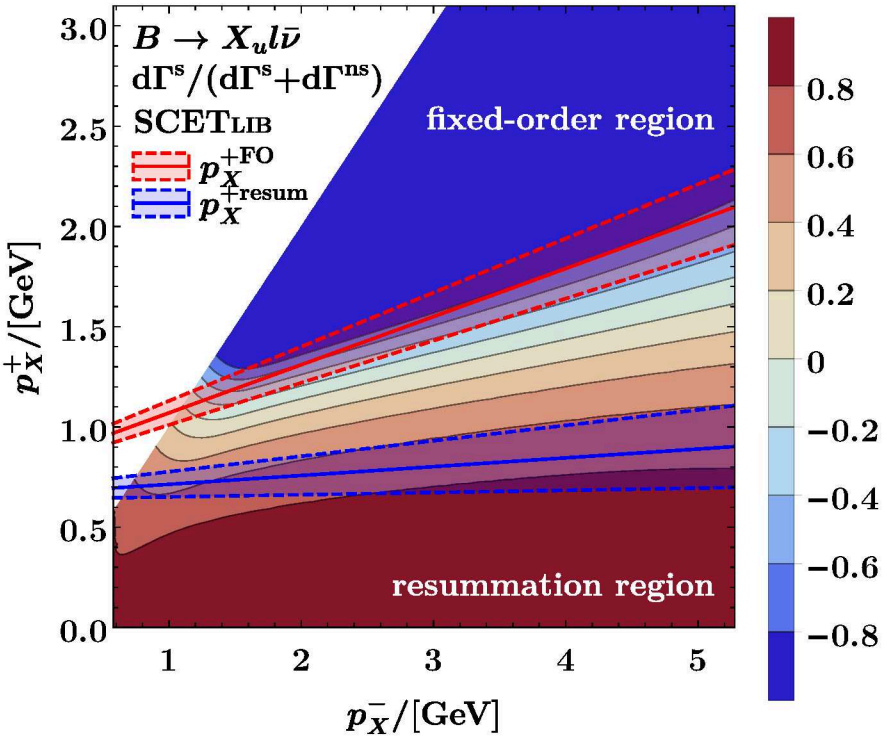


Figure 6.4.: Fractional contribution $d\Gamma^s/(d\Gamma^s + d\Gamma^{ns})$ of singular terms in different regions of the phasespace and the chosen boundaries of fixed-order and resummation regions. The thin black lines show contours of constant fractional contribution $d\Gamma^s/(d\Gamma^s+d\Gamma^{ns})$ of singular terms. The solid red and blue lines show the chosen boundaries $p_X^{+\mathrm{FO}}$, $p_X^{+\mathrm{resum}}$ of fixed-order and resummation regions, respectively. The dashed red and blue lines show variations of these boundaries.

The variations of transition lines $p_X^{+\mathrm{resum}}$ and $p_X^{+\mathrm{FO}}$ are chosen as linear combinations of their default positions in the (p_X^+, p_X^-) plane. The magnitude of the these variations was chosen to correspond to a $\pm 0.2\,\mathrm{GeV}$ variation of p_X^+ at $p_X^- = m_B$. This value of $0.2\,\mathrm{GeV}$ is analogous to $\pm 0.1\,\mathrm{GeV}$ variation of transition point E_1 that was used for the $B \to X_s\gamma$ profile functions (see eq. (5.38)). The corresponding parameters p_X^{resum} and e^{resum}, and similarly p_X^{FO} and e^{FO}, are varied not independently but in a correlated way. These variations of transition lines $p_X^{+\mathrm{resum}}$ and $p_X^{+\mathrm{FO}}$ are illustrated in figure 6.4.

The perturbative uncertainty is estimated as the sum in quadrature of resummation, nonsingular, and matching uncertainties:

$$\Delta_{\mathrm{total}} \coloneqq \Delta_{\mathrm{resum}} \oplus \Delta_{\mathrm{ns}} \oplus \Delta_{\mathrm{match}}. \tag{6.37}$$

The symbol $\oplus$ denotes addition in quadrature: $x \oplus y \coloneqq \sqrt{x^2 + y^2}$. Each of the uncertainty components, — resummation Δ_{resum}, nonsingular Δ_{ns}, and matching Δ_{match}, —

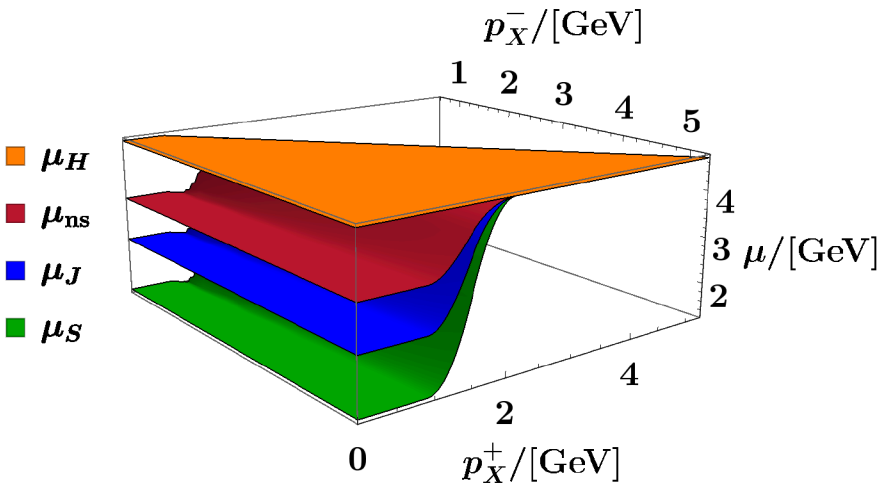


Figure 6.5.: Central profile functions for $B \to X_u l\bar{\nu}$.

is found by taking the envelope of theoretical predictions resulting from variations of profile function parameters (e_H, e_J, μ_0), e_{ns}, and $(e^{\mathrm{resum}}, p_X^{\mathrm{resum}}, e^{\mathrm{FO}}, p_X^{\mathrm{FO}})$, respectively. Excluding the central parameter values, this corresponds to $(3^3-1)+(3-1)+(3^2-1) = 36$ variations.

6.6. Numerical results

Unless stated otherwise, the results presented in this section have been calculated using the following setup: The b-quark mass is defined in the MSR mass scheme with $R = 1\,\mathrm{GeV}$, and at NNLO the invisible scheme is used for the hadronic parameter λ_1 (see section 4.3). The default shape function model with $p = 4$, $\lambda = 0.6\,\mathrm{GeV}$ is used (see section 3.1). The numerical value for the b-quark mass in a short-distance scheme is $\widehat{m}_b = 4.7\,\mathrm{GeV}$, and the B-meson mass is $m_B = 5.279\,\mathrm{GeV}$. The calculations are carried out in theory with $n_f = 4$ flavors of massless quarks. The evolution of the strong coupling constant α_s is performed using an approximate solution that is accurate to N^{3}LL, and the same solution for α_s running is used for all presented results, regardless of their order. The reference value for the strong coupling constant is $\alpha_s(4.7\,\mathrm{GeV}) = 0.2155$. The presented integrated results assume a cut on the charged-lepton energy $E_l > 1\,\mathrm{GeV}$, which is typical for experimental measurements of $B \to X_u l\bar{\nu}$ decay [55].

As already mentioned, although the resummed contributions are known at N^{3}LL, the nonsingular contributions are known only at NLO. The results at N^{3}LL+NLO are valid in the SCET region, where the nonsingular contributions are power-suppressed. In the local-OPE region these results reduce to the sum of singular terms at NNLO and nonsingular terms at NLO. Although their sum formally reproduces the local-OPE results at NLO, the remaining $\mathcal{O}(\alpha_s^2)$ singular terms are not as small as one might hope. As was already shown in eq. (6.26), in the local-OPE region there's a large

cancellation between the singular and the nonsingular contributions. Because in the local-OPE region the singular contributions by themselves are not meaningful, they turn out to be unnaturally large in comparison to the full local-OPE results. As a result, the N^{3}LL+NLO predictions are not reliable outside of the SCET region, not even to NLO. This will be demonstrated numerically in this section.

Figure 6.6 shows the single-differential spectra $\mathrm{d}\Gamma/\mathrm{d}p_X^-$ and $\mathrm{d}\Gamma/\mathrm{d}q^2$. Plots on the

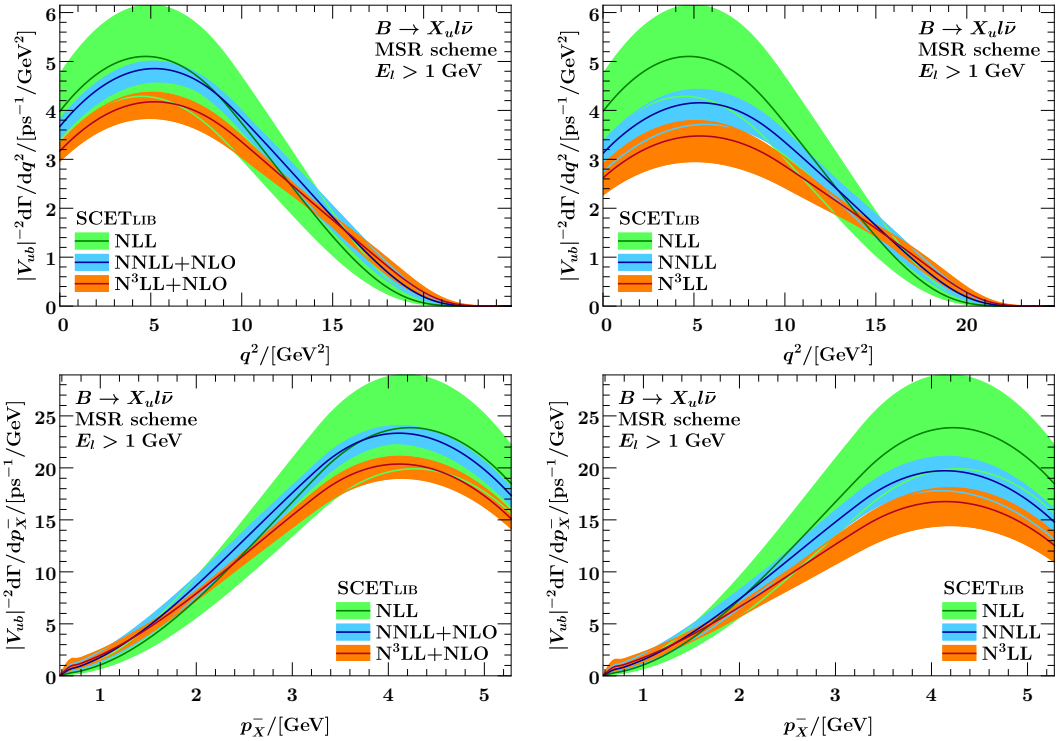


Figure 6.6.: Single-differential $B \to X_u l\bar{\nu}$ spectra $\mathrm{d}\Gamma/\mathrm{d}q^2$ and $\mathrm{d}\Gamma/\mathrm{d}p_X^-$. The presented predictions are integrated over the remaining kinematic variables, with a cut $E_l > 1\,\mathrm{GeV}$ on the charged-lepton energy. The right panels show only the resummed singular contribution, while panels on the left show the full result, which includes both resummed and nonsingular contributions.

right include both the resummed singular and the nonsingular contributions, while plots on the right show only the singular contribution. Because these spectra are integrated over p_X^+, they are not sensitive to the exact shape of the shape function, but depend only on its first moments. The $\mathrm{d}\Gamma/\mathrm{d}p_X^-$ and $\mathrm{d}\Gamma/\mathrm{d}q^2$ spectra are adequately described by fixed-order local-OPE predictions, and resummation of singular contributions is not necessary. The NLL and NNLL+NLO results agree within their uncertainties, but there's a clear tension between the predictions at N^{3}LL+NLO and the lower-order results. To understand the cause of this tension it is useful to examine the singular contributions to these spectra, which are shown on the right side of figure 6.6. The NLL results are the same, because there are no nonsingular corrections at this order. However, the

NNLL+NLO and N^{3}LL+NLO predictions have larger uncertainties, and the singular and full predictions at these two orders are significantly different. This demonstrates the importance of nonsingular corrections to these spectra.

Figure 6.7 compares the theoretical predictions for $\mathrm{d}\Gamma/\mathrm{d}p_X^-$ and $\mathrm{d}\Gamma/\mathrm{d}q^2$ with measurements by the Belle experiment [55]. The theoretical results are averaged over

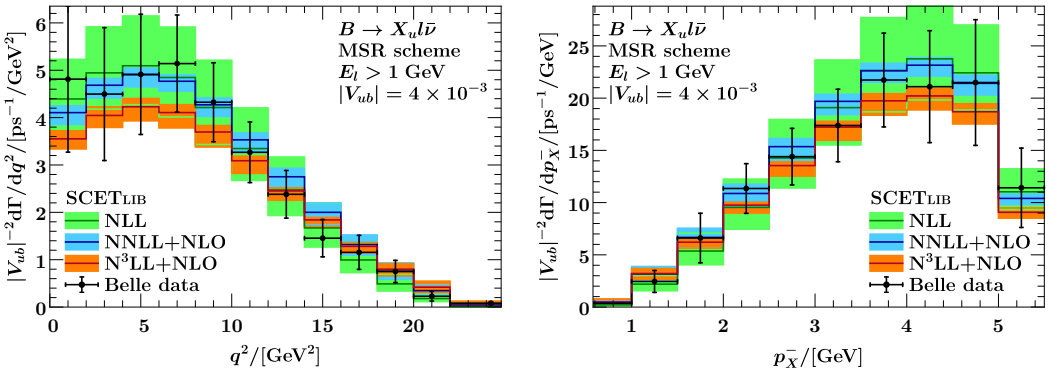


Figure 6.7.: Comparison of theoretical predictions of single-differential $B \to X_u l\bar{\nu}$ spectra $\mathrm{d}\Gamma/\mathrm{d}q^2$ and $\mathrm{d}\Gamma/\mathrm{d}p_X^-$ to Belle measurements [55]. The presented predictions are integrated over the remaining kinematic variables, with a cut $E_l > 1\,\mathrm{GeV}$ on the charged-lepton energy. For comparison, the experimental results are divided by $|V_{ub}|^2$ with $|V_{ub}| = 4 \times 10^{-3}$.

each bin. For comparison, the experimental measurements are divided by $|V_{ub}|^2$ with $|V_{ub}| = 4 \times 10^{-3}$. Overall, the experiment and theory are in agreement, even with the suboptimal predictions at N^{3}LL+NLO. This agreement implies that the chosen normalization $|V_{ub}| = 4 \times 10^{-3}$ is adequate.

Single-differential spectra $\mathrm{d}\Gamma/\mathrm{d}E_l$ and $\mathrm{d}\Gamma/\mathrm{d}p_X^+$ are shown in figure 6.8. Again, plots on the left inlcude both resummed singular and nonsingular contributions, while plots on the right show only the singular contribution. Again, because only singular terms contribute at NLL, predictions at this order are the same on the left and on the right. In contrast to the $\mathrm{d}\Gamma/\mathrm{d}q^2$ and $\mathrm{d}\Gamma/\mathrm{d}p_X^-$ spectra, the $\mathrm{d}\Gamma/\mathrm{d}E_l$ and $\mathrm{d}\Gamma/\mathrm{d}p_X^+$ spectra in their respective endpoint regions are sensitive to the shape of the shape function.

The endpoint region of the $\mathrm{d}\Gamma/\mathrm{d}E_l$ is close to the maximal charged-lepton energy $E_l = m_B/2 \simeq 2.64\,\mathrm{GeV}$, which corresponds to the right side of the top plots in figure 6.8. Away from the endpoint region, the convergence pattern of $\mathrm{d}\Gamma/\mathrm{d}E_l$ predictions is similar to the $\mathrm{d}\Gamma/\mathrm{d}p_X^-$ and $\mathrm{d}\Gamma/\mathrm{d}q^2$: the NLL and NNLL+NLO predictions agree within uncertainties, and the N^{3}LL+NLO prediction is clearly an outlier. Similarly, just the singular contributions at different orders do not agree with each other away from the endpoint region, as can be seen on the top-right plot in figure 6.8. This demonstrates once again that the nonsingular contributions in this region of $\mathrm{d}\Gamma/\mathrm{d}E_l$ spectrum need to be accounted for.

The endpoint region of the $\mathrm{d}\Gamma/\mathrm{d}p_X^+$ is near $p_X^+ = 0\,\mathrm{GeV}$, and it can be seen on the left side of the bottom plots in figure 6.8. The predictions in this region have a similar shape

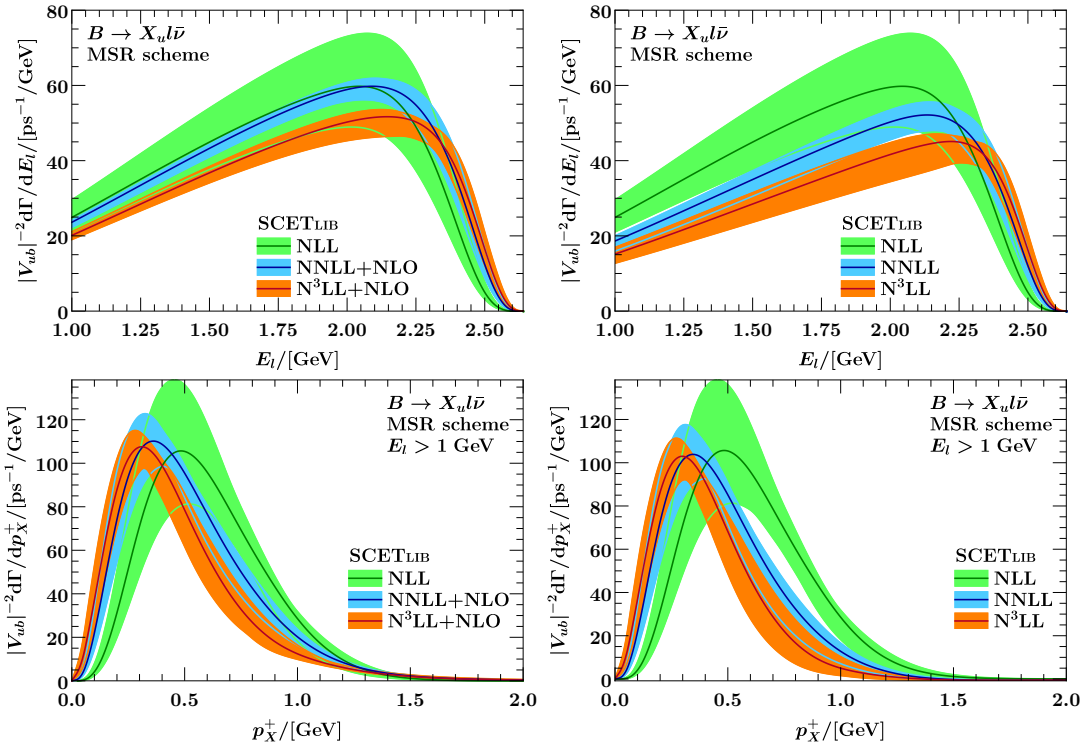


Figure 6.8.: Single-differential $B \to X_u l\bar{\nu}$ spectra $\mathrm{d}\Gamma/\mathrm{d}E_l$ and $\mathrm{d}\Gamma/\mathrm{d}p_X^+$. The presented predictions are integrated over the remaining kinematic variables. The predictions for $\mathrm{d}\Gamma/\mathrm{d}p_X^+$ are integrated over all charged-lepton energies $E_l > 1\,\mathrm{GeV}$. The right panels show only the resummed singular contribution, while panels on the left show the full result, which includes both resummed and nonsingular contributions.

and peak values, although the position of the peak of the distribution is a little different at different orders. In the peak region of the $\mathrm{d}\Gamma/\mathrm{d}p_X^+$ spectrum the singular contributions are dominant and the nonsingular contributions are a relatively small correction. As a result, the full and singular predictions, shown on bottom-left and bottom-right in figure 6.8, are similar in this region. In the peak region the N^{3}LL+NLO predictions are reliable and are expected to be more accurate than predictions at NNLL+NLO. Indeed, the bottom plots in figure 6.8 show some reduction in theoretical uncertainty at the peak of the spectrum between N^{3}LL+NLO (orange band) and NNLL+NLO (blue band).

These theoretical predictions of $\mathrm{d}\Gamma/\mathrm{d}E_l$ and $\mathrm{d}\Gamma/\mathrm{d}p_X^+$ spectra are compared to the Belle measurements [55] in figure 6.9. The `SCETlib` predictions, calculated in this work, are presented in the top row, while the bottom row shows analogous comparison plots from ref. [55]. Some tensions between theory and experiment are clearly visible. For this reason, it is interesting to compare predictions of the model used in this work with predictions of other theoretical models. The latter are compared with the data in the bottom row in figure 6.9. The BLNP [56] and DFN [48] are inclusive models that rely

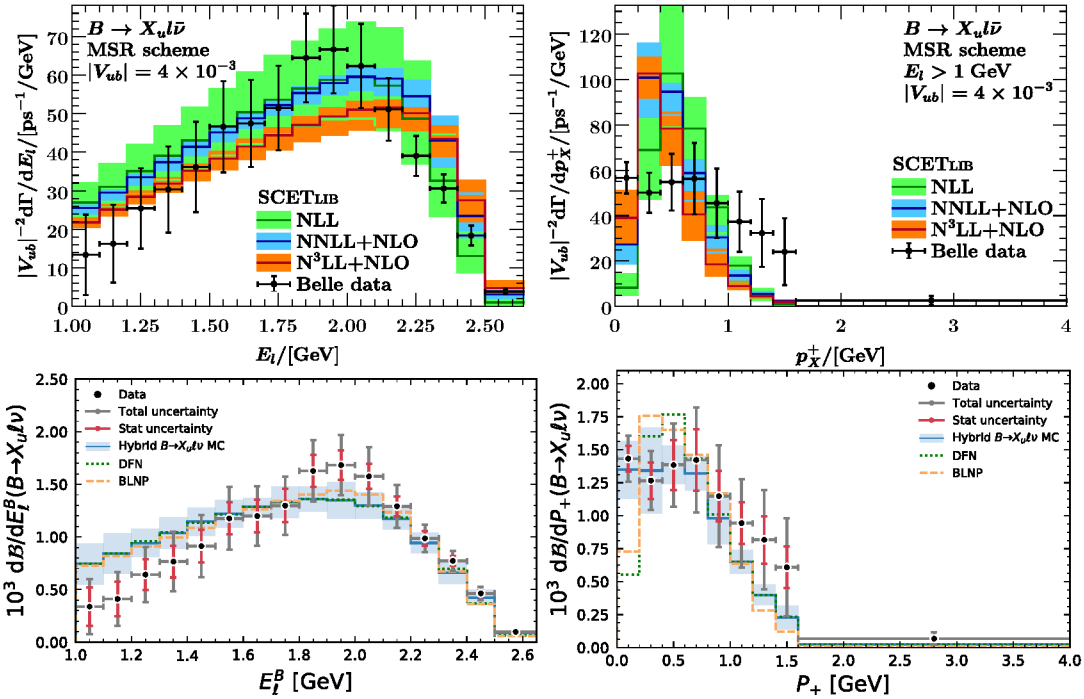


Figure 6.9.: Comparison of theoretical predictions of single-differential $B \to X_u l\bar{\nu}$ spectra $\mathrm{d}\Gamma/\mathrm{d}E_l$ and $\mathrm{d}\Gamma/\mathrm{d}p_X^+$ to the Belle measurements [55]. The results are integrated over the remaining kinematic variables, and for $\mathrm{d}\Gamma/\mathrm{d}p_X^+$ are integrated over all charged-lepton energies $E_l > 1\,\mathrm{GeV}$. The top row shows theoretical predictions derived in this work, and the experimental results on these plots are divided by $|V_{ub}|^2$ with $|V_{ub}| = 4 \times 10^{-3}$. Plots on the bottom row are from ref. [55].

on the assumption of local quark-hadron duality, similarly to this work. The hybrid model [57] combines the predictions of the inclusive models and exclusive models. The comparison of top and bottom rows in figure 6.9 reveals that the discrepancies between theory and data are qualitatively similar between the model used in this work and the other inclusive models.

In the $E_l < 1.5\,\mathrm{GeV}$ region of the lepton-energy spectrum $\mathrm{d}\Gamma/\mathrm{d}E_l$ all theory predictions are larger than the measurements. This deviation is consistent across the first five bins. However, ref. [55] notes that the measurements in the region $1\,\mathrm{GeV} < E_l < 1.8\,\mathrm{GeV}$ are strongly correlated with each other and only weakly correlated with measurements in the other regions of the spectrum. As a result, the statistical significance of this discrepancy is not as large as implied by the plots on the left side in figure 6.9. In the region $1.8\,\mathrm{GeV} < E_l < 2.1\,\mathrm{GeV}$ the measurements are larger than all theory predictions. However, in the region $E_l > 2.2\,\mathrm{GeV}$ the measurements are smaller than predictions derived in this work, but are in agreement with the other inclusive models. This discrepancy could be due to the different shape-function model, the shape of which is important in

this region of the spectrum. Another possible cause of this difference is the subleading shape functions, the effect of which is known at leading order in α_s [27]. The effect of subleading shape functions is included in the BLNP inclusive model [56], but not in this work.

Similarly, for the spectrum $\mathrm{d}\Gamma/\mathrm{d}p_X^+$ the right side of figure 6.9 shows a clear difference between the results of this work, predictions of the other inclusive models, and experimental measurements. All inclusive theoretical models predict a peak in the $\mathrm{d}\Gamma/\mathrm{d}p_X^+$ distribution at $p_X^+ \simeq 0.5\,\mathrm{GeV}$. In comparison to the other inclusive models, the `SCETlib` results show a sharper, higher and narrower, peak. This is most likely explained by the different shape function model, as the shape of the peak in the $\mathrm{d}\Gamma/\mathrm{d}p_X^+$ is almost completely determined by the shape function. In the endpoint region $p_X^+ < 0.6\,\mathrm{GeV}$ all inclusive predictions are smaller than the measurements in the first bin, and larger than the measurements in the second and third bin. On the other hand, predictions of the hybrid model agree with the data very well in this region. This is a clear sign that the tension between the inclusive theory and the experiment in this region is due to the resonant structure of the $\mathrm{d}\Gamma/\mathrm{d}p_X^+$. Finally, in the region $1\,\mathrm{GeV} < p_X^+ < 1.6\,\mathrm{GeV}$ all theoretical models, — both inclusive and hybrid, — consistently predict smaller branching fraction than observed. The cause of this difference is unknown.

The contribution of the resonances that is visible in the peak region of the $\mathrm{d}\Gamma/\mathrm{d}p_X^+$ spectrum is even more obvious in the invariant-mass spectra $\mathrm{d}\Gamma/\mathrm{d}m_X$ and $\mathrm{d}\Gamma/\mathrm{d}m_X^2$, which are shown in figure 6.10. None of the inclusive models can adequately describe these spectra below $m_X \lesssim 1.4\,\mathrm{GeV}$, while the hybrid model, the predictions of which are shown on the bottom row in figure 6.10, is in good agreement with the data in this region. This unambigously demonstrates the violation of quark-hadron duality, which is assumed in all inclusive models.

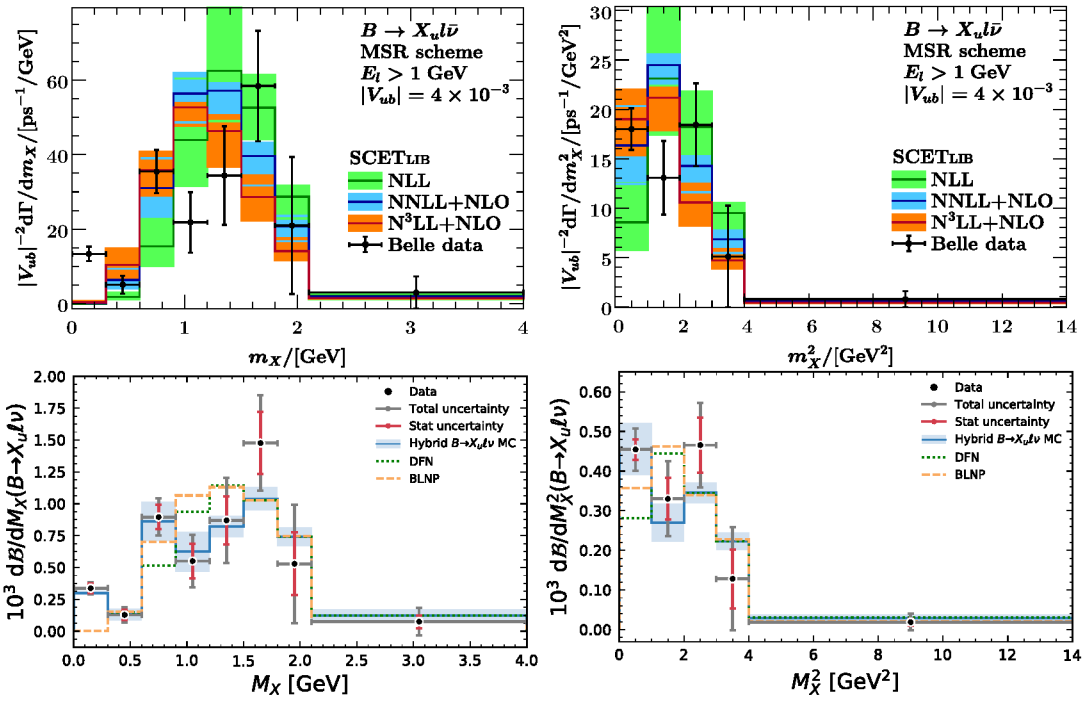


Figure 6.10.: Comparison of theoretical predictions of single-differential $B \to X_u l\bar{\nu}$ invariant-mass spectra $\mathrm{d}\Gamma/\mathrm{d}m_X$ and $\mathrm{d}\Gamma/\mathrm{d}m_X^2$ to the Belle measurements [55]. The results are integrated over the remaining kinematic variables with a cut $E_l > 1\,\mathrm{GeV}$ on the energy of the charged lepton. The top row shows theoretical predictions derived in this work, and the experimental results on these plots are divided by $|V_{ub}|^2$ with $|V_{ub}| = 4 \times 10^{-3}$. Plots on the bottom row are from ref. [55].

7. Conclusions

The main result of this work is the improved theoretical predictions for the inclusive $B \to X_s\gamma$ and $B \to X_u l\bar{\nu}$ differential decay rates. Several improvements to the state-of-the-art theory have been made.

First, all the available 3-loop corrections to the $B \to X_s\gamma$ photon energy spectrum have been taken into account. The few missing corrections have been parametrized in terms of nuisance parameters, and their size has been estimated based on an extrapolation from the known lower-order corrections. Variations of the nuisance parameters provide an estimate of the uncertainty related to these missing 3-loop contributions. A further advantage of the nuisance-parameter approach is that it exposes the correlation of the theoretical uncertainty in the different regions of the spectrum. In a potential future fit the values of the nuisance parameters could be constrained by the experimental data.

Second, the impact of different short-distance mass schemes has been investigated. The presented numerical results illustrate the well-known fact that the pole mass suffers from a renormalon ambiguity. The associated convergence issue is evident in both the $B \to X_s\gamma$ photon energy spectrum predictions and the soft function. The results in the 1S short-distance mass scheme are more reasonable, and, taken together with the invisible scheme for the hadronic parameter λ_1, are stable up to and including NNLO [23]. At N^{3}LO, however, the 1S mass scheme clearly oversubtracts. This fact is clearly visible not only in both the photon energy spectrum and the soft function in the 1S scheme, but also in the perturbative series for the mass correction $m_b^{\rm MSR} - m_b^{\rm 1S}$. This failure at N^{3}LO is explained by the observation that the intrinsic scale $R^{\rm 1S}$ of the 1S scheme becomes too large at scales below $\mu \sim 2\,\text{GeV}$. The predictions in the MSR mass scheme are much more stable across different perturbative orders, which can be attributed to the fact that its intrinsic scale is a parameter, which can be set to the appropriately low infrared scale $R \sim 1\,\text{GeV}$. The stability of the predictions in the MSR scheme is further improved by tuning the definitions of the HQET hadronic parameters λ_1 and ρ_1, which are expected to have subleading renormalon ambiguities.

Third, the short-distance mass corrections δm_b in the hard and jet functions have been taken into account in the $B \to X_s\gamma$ spectrum predictions. Because these corrections are suppressed by $\delta m_b/m_b \sim \Lambda_{\rm QCD}/m_b$ and are formally subleading, previous studies have treated them as small nonsingular corrections. However, starting at N^{3}LO a contribution appears that is suppressed by $\delta m_b/m_b$, but is nevertheless singular, in the sense that it behaves as $\delta(x)$ or $\ln^n x/x$ in the peak region. Because this contribution arises from the mass correction δm_b in the hard and jet functions, whose resummation structure is well-understood, this subleading singular contribution can be resummed as well. The resummation of this contribution is achieved automatically if all δm_b corrections in the hard and jet functions are consistently taken into account, as was done in this work.

Fourth, different treatments of higher-order cross-terms in the product of hard, jet, and soft functions have been investigated. Previous studies have observed that, at NNLO, keeping all terms of the hard function series in the product of hard, jet, and soft functions seems to reduce the scale variation uncertainty [23]. The results presented in this work show that this is true at NNLO, but not at $\mathrm{N^3LO}$. Furthermore, the reduced scale variation uncertainty at NNLO appears to underestimate the next-order contribution. As expected, as the theory uncertainty reduces at higher perturbative orders, so does the difference between the two ways of treating these higher-order terms.

As a result, the uncertainty of the theoretical predictions for the $B \to X_s\gamma$ photon energy spectrum at $\mathrm{N^3LL'}$+$\mathrm{N^3LO}(c_k)$ has been reduced compared to NNLL′+NNLO, even accounting for the uncertainty due to the yet-unknown 3-loop corrections.

The presented improved $B \to X_u l\bar{\nu}$ predictions include all singular contributions at NNLO, resummed at $\mathrm{N^3LL}$ accuracy. The developed procedure for matching resummed and fixed-order contributions avoids artificial singularities and negative decay rates. Although the fixed-order NNLO calculations have been carried out [50], the full results of the calculations have not been published. As a result, the highest accuracy level of the presented predictions is $\mathrm{N^3LL}$+NLO, and the nonsingular corrections are included only at NLO. A numerical examination of these results shows a reduction of theoretical uncertainty at $\mathrm{N^3LL}$+NLO in the SCET region. However, outside of the SCET region the NNLO singular results by themselves are not reliable, and the NNLO nonsingular results are needed at this level of accuracy. The theoretical predictions are compared with the available Belle measurements and with other theoretical models. As expected, the theoretical predictions for inclusive spectra $\mathrm{d}\Gamma/\mathrm{d}q^2$ and $\mathrm{d}\Gamma/\mathrm{d}p_X^-$ agree with the experimental data. The discrepancies between data and theory in the other observables are investigated. The presented theoretical results could be further improved by accounting for nonsingular corrections at NNLO and effects of subleading shape functions. Derivation of the analytic NNLO nonsingular corrections is an ongoing project of our working group.

In the future, the improved theoretical predictions presented in this work could be used in a combined fit of $B \to X_s\gamma$ and $B \to X_u l\bar{\nu}$ measurements, which would simultaneously constrain $|V_{ub}|$ and leading shape-function effects. The results presented in this thesis are an important step towards a more reliable determination of $|V_{ub}|$.

Appendices

A. Notation and conventions

The following conventions are used throughout this thesis: The signature of the metric tensor is $(+,-,-,-)$. The sign of the Levi-Civita tensor is chosen such that:

$$\varepsilon^{0123} \coloneqq 1 \tag{A.1}$$

Symmetrization and antisymmetrization of indices are defined as:

$$A^{(\mu\nu)} \coloneqq \frac{1}{2}(A^{\mu\nu} + A^{\nu\mu}), \qquad A^{[\mu\nu]} \coloneqq \frac{1}{2}(A^{\mu\nu} - A^{\nu\mu}). \tag{A.2}$$

The tensor $\sigma^{\mu\nu}$ is defined as

$$\sigma^{\mu\nu} \coloneqq i\gamma^{[\mu}\gamma^{\nu]} = \frac{i}{2}(\gamma^\mu\gamma^\nu - \gamma^\nu\gamma^\mu), \tag{A.3}$$

where γ^μ are the standard gamma matrices for Dirac bispinors in 4 dimensions. The chirality matrix γ_5 and the left and right chirality projectors $P_{L,R}$ are defined as

$$\gamma_5 = i\gamma^0\gamma^1\gamma^2\gamma^3, \qquad P_L = \frac{1-\gamma_5}{2}, \qquad P_R = \frac{1+\gamma_5}{2}. \tag{A.4}$$

Unless states otherwise, all quantum states and operators are in Heisenberg picture.

The symbols C_F and C_A denote the quadratic Casimir elements of the fundamental and adjoint representations of SU(3), respectively, which are equal to

$$C_F = \frac{4}{3}, \qquad C_A = 3. \tag{A.5}$$

A.1. Discontinuity operator

The discontinuity operator Disc measures the discontinuity of a complex function across a branch cut:

$$\operatorname{Disc}_z f(z) \coloneqq \lim_{\varepsilon\to+0}(f(z+i\varepsilon) - f(z-i\varepsilon)). \tag{A.6}$$

If the function f satisfies the condition $f(z^*) = (f(z))^*$, then its discontinuity is purely imaginary:

$$\operatorname{Disc}_z f(z) = \lim_{\varepsilon\to+0} 2i \operatorname{Im} f(z+i\varepsilon). \tag{A.7}$$

Alternatively, if the function satisfies the condition $f(z^*) = -(f(z))^*$, then its discontinuity is real:

$$\mathrm{Disc}_z\, f(z) = \lim_{\varepsilon \to +0} 2\,\mathrm{Im}\, i f(z + i\varepsilon). \tag{A.8}$$

For this reason, the discontinuity across a branch cut is usually written as an imaginary part Im of some expression.

Furthermore, the notation with Im is used to denote the discontinuity even in cases when the discontinuity does not just equal the imaginary part. For example, the $B \to X_u l \bar{\nu}$ hadronic tensor $W^{\mu\nu}$, defined eq. (6.6), is commonly written as an imaginary part Im:

$$\mathrm{Disc}_{q^0} \int \mathrm{d}^4 x e^{iqx} \langle B|\, \mathbf{T}\, J_\mu^\dagger(0) J_\nu(x) |B\rangle \approx 2\,\mathrm{Im}\, i \int \mathrm{d}^4 x e^{iqx} \langle B|\, \mathbf{T}\, J_\mu^\dagger(0) J_\nu(x) |B\rangle. \tag{A.9}$$

The notation with Im is misleading in this case, because the hadronic tensor has an imaginary part that is not related to its discontinuity across a branch cut.

B. Fourier-style convolution

The symbol $\otimes$ denotes the Fourier-style convolution:

$$(f \otimes g)(x) \coloneqq \int_{-\infty}^{+\infty} f(t)g(x-t)\mathrm{d}t. \tag{B.1}$$

When it is not obvious, the subscript x on the convolution symbol $\otimes_x$ indicates the convolution variable x.

This convolution corresponds to a simple pointwise product in Fourier space. This fact immediately implies that the convolution $\otimes$ is commutative and associative.

For sufficiently regular functions one can use integration by parts to move derivatives:

$$\partial(f \otimes g) = (\partial f) \otimes g = f \otimes (\partial g). \tag{B.2}$$

as well as to shift the argument by a finite constant:

$$f(x+a) = (e^{a\partial} f)(x), \qquad e^{a\partial}(f \otimes g) = (e^{a\partial} f) \otimes g = f \otimes (e^{a\partial} g). \tag{B.3}$$

Integral up to some point x can be written as a convolution with a θ-function:

$$(\theta \otimes f)(x) = \int_{-\infty}^{+\infty} f(t)\theta(x-t)\mathrm{d}t = \int_{-\infty}^{x} f(t)\mathrm{d}t \tag{B.4}$$

C. Plus-distributions

Consider a function f with support $x \geq 0$, such that $xf(x)$ is integrable around $x = 0$, i.e. for small ε the integral $\int_0^\varepsilon xf(x)\mathrm{d}x$ is finite. Then the plus-distribution $[f]_+^a$ is defined as [23]:

$$[f]_+^a(x) \coloneqq \lim_{\varepsilon\to 0} \partial_x \left[\theta(x-\varepsilon)\int_a^x \mathrm{d}t f(t)\right] = \lim_{\varepsilon\to 0}\left[\theta(x-\varepsilon)f(x) - \delta(x-\varepsilon)\int_\varepsilon^a \mathrm{d}t f(t)\right], \tag{C.1}$$

and $[f]_+ \coloneqq [f]_+^1$. The plus-distributions are useful, because the plus-distribution $[f(x)]_+^a$ is integrable in the neighbourhood of $x = 0$ even if the function $f(x)$ itself is not:

$$(\theta \otimes [f]_+^a)(x) = \lim_{\varepsilon\to 0}\left[\int_\varepsilon^x f(t)\mathrm{d}t - \int_\varepsilon^a \mathrm{d}t f(t)\right] = \int_a^x f(x)\mathrm{d}x. \tag{C.2}$$

This type of integral of a plus-distribution is called a "cumulant integral", or just a "cumulant". The values $f(x)$ of a plus-distribution at points $x > 0$ are referred to as its "spectrum".

A convolution of a regular function $g(x)$ with support $x \geq 0$ with a plus-distribution $[f(x)]_+^a$ can be written in terms of its spectrum and cumulant:

$$([f]_+^a \otimes g)(x) = \int_0^x f(t)[g(x-t) - g(x)]\mathrm{d}t + g(x)\int_a^x f(t)\mathrm{d}t. \tag{C.3}$$

Note that the first integral converges because for small t the integrand

$$f(t)[g(x-t) - g(x)] = -f(t)tg'(x) + \mathcal{O}(t^2 f(t)) \tag{C.4}$$

is integrable.

C.1. $\mathcal{L}$-distributions

The distributions $\mathcal{L}_n^a$ with $a > -1$, $n \geq -1$, $\mu > 0$ are defined as in ref. [23] and ref. [58]:

$$\mathcal{L}_n^a(x) \coloneqq \begin{cases} [\theta(x)x^{a-1}\ln^n x]_+ & \text{if } n \geq 0 \\ \delta(x) & \text{if } n = -1 \end{cases}, \tag{C.5}$$

$$\mathcal{L}_n^a(k,\mu) \coloneqq \frac{1}{\mu}\mathcal{L}_n^a(k/\mu), \qquad \mathcal{L}_n \coloneqq \mathcal{L}_n^0, \qquad \mathcal{L}^a \coloneqq \mathcal{L}_0^a \tag{C.6}$$

These distributions arise for example in the series expansion of $\theta(x)x^{-1-\varepsilon}$:

$$\frac{\theta(x)}{x^{1+\varepsilon}} = \frac{1}{\varepsilon}\mathcal{L}_{-1}(x) + \sum_{n=0}^{\infty}\frac{\varepsilon^n}{n!}\mathcal{L}_n(x). \tag{C.7}$$

For $n \geq 0$ these distributions satisfy the following differential relations:

$$\frac{\partial}{\partial a}\mathcal{L}_n^a = \mathcal{L}_{n+1}^a, \qquad \frac{\partial}{\partial \ln\mu}\mathcal{L}_n^a(k,\mu) = -a\mathcal{L}_n^a(k,\mu) - (n+\delta_n^0)\mathcal{L}_{n-1}^a(k,\mu). \tag{C.8}$$

Ref. [58] derives Fourier transforms of $\mathcal{L}_n$ and $\mathcal{L}^a$. $\mathcal{L}$-distributions of different arguments are related by the following identities:

$$\mathcal{L}_n^a(k,\mu) = \frac{1}{\mu}\mathcal{L}_n^a(k/\mu), \qquad \mathcal{L}_n(x,\nu) = \sum_{k=-1}^{n} C_k^n \mathcal{L}_k(x,\mu)\left(\ln\frac{\mu}{\nu}\right)^{n-k}, \tag{C.9}$$

where

$$C_k^n = \binom{n}{k} = \frac{n!}{k!(n-k)!} \text{ for } n,k \geq 0, \quad C_{-1}^n = \frac{1}{n+1} \text{ for } n \geq 0, \quad C_{-1}^{-1} = 1. \tag{C.10}$$

Cumulant integral of an $\mathcal{L}$-distribution with $n \geq 0$ is:

$$\begin{aligned}(\theta \otimes \mathcal{L}_n^a)(x) &= \int_1^x t^{a-1}\ln^n t\,\mathrm{d}t = \partial_a^n \frac{x^a-1}{a} = n!(-a)^{-n-1}\left(1 - x^a\sum_{k=0}^{n}\frac{(-a\ln x)^k}{k!}\right)\\ &= n!x^a(-a)^{-n-1}\sum_{k=n+1}^{\infty}\frac{(-a\ln x)^k}{k!}\\ &= n!x^a(\ln x)^{n+1}\sum_{k=0}^{\infty}\frac{(-a\ln x)^k}{(n+1+k)!}\\ &= (\ln x)^{n+1}\sum_{k=0}^{\infty}\frac{(a\ln x)^k}{(n+1+k)k!}.\end{aligned} \tag{C.11}$$

The last formula is convenient for numerically accurate evaluation when $a\ln x$ is small.

In the special case $a = 0$, $n \geq 0$:

$$(\theta \otimes \mathcal{L}_n)(x) = \frac{(\ln x)^{n+1}}{n+1}. \tag{C.12}$$

Ref. [23] derives convolutions of $\mathcal{L}$-distributions:

$$\mathcal{L}^a \otimes \mathcal{L}^b = \frac{a+b}{ab} B(a,b) \mathcal{L}^{a+b} - \frac{\mathcal{L}^a}{b} - \frac{\mathcal{L}^b}{a} + \frac{B(a,b)-1}{ab} \mathcal{L}_{-1}, \tag{C.13}$$

$$\mathcal{L}^a \otimes \mathcal{L}_n =: \frac{1}{a} \left(\sum_{k=-1}^{n+1} V_k^n(a) \mathcal{L}_k^a - \mathcal{L}_n(x) \right), \tag{C.14}$$

$$\mathcal{L}_m \otimes \mathcal{L}_n =: \sum_{k=-1}^{m+n+1} V_k^{mn} \mathcal{L}_k, \tag{C.15}$$

where

$$B(a,b) \coloneqq \frac{\Gamma(1+a)\Gamma(1+b)}{\Gamma(1+a+b)}, \tag{C.16}$$

and the convolution coefficients V_k^{mn} and $V_k^n(a)$ are related to the coefficients of the Taylor series of $B(a,b)$:

$$V_k^{mn} = \begin{cases} V_k^{nm} & \text{if } n > m, \\ \delta_k^m & \text{if } m \geq n = -1, \\ \frac{\partial_a^{m+1} \partial_b^{n+1}}{(m+1)(n+1)} B(a,b)|_{a=b=0} & \text{if } k = -1, \\ \sum_{p=0}^{m} \sum_{q=0}^{n} \delta_{p+q}^{m+n-k} \binom{m}{p} \binom{n}{q} \partial_a^p \partial_b^q \left(\frac{\partial_a}{p+1} + \frac{\partial_b}{q+1} \right) B(a,b)|_{a=b=0} & \text{if } 0 \leq k < m+n, \\ 0 & \text{if } 0 \leq k = m+n, \\ \frac{1}{m+1} + \frac{1}{n+1} & \text{if } k = m+n+1, \\ 0 & \text{if } k > m+n+1, \end{cases} \tag{C.17}$$

where $m, n, k \geq -1$.

$$V_k^n(a) = \begin{cases} 1 & \text{if } n = k = -1, \\ a & \text{if } n = -1 \text{ and } k = 0, \\ \frac{\partial_b^{n+1}}{n+1} B(a,b)|_{b=0} & \text{if } k = -1 \text{ and } n \geq 0, \\ \binom{n}{k} \left(\partial_b^{n-k} + \frac{a \partial_b^{n-k+1}}{n-k+1} \right) B(a,b)|_{b=0} & \text{if } 0 \leq k < n, \\ 1 + a \partial_b B(a,b)|_{b=0} & \text{if } k = n \geq 0, \\ \frac{a}{n+1} & \text{if } k = n+1 \geq 1, \\ 0 & \text{if } k > n+1. \end{cases} \tag{C.18}$$

For any given $N \in \mathbb{N}$, $a \in \mathbb{R}$ all nonzero coefficients $V_k^n(a)$ with $n \leq N$ can be expressed in terms of $\partial_b B(a,b)|_{b=0}, \ldots, \partial_b^{N+1} B(a,b)|_{b=0}$. The derivative $\partial_b^n B(a,b)|_{b=0}$ is given by the Faà di Bruno's formula:

$$\partial_b^n B(a,b)|_{b=0} = \partial_b^n e^{\ln B(a,b)}|_{b=0} = B_n(\partial_b \ln B(a,b), \ldots, \partial_b^n \ln B(a,b))|_{b=0}, \tag{C.19}$$

where B_n is the n-th exponential Bell polynomial, and the identity $B(a,0)=1$ was used. The derivative $\partial_b^n \ln B(a,b)|_{b=0}$ is

$$\partial_b^n \ln B(a,b)|_{b=0} = \begin{cases} \psi^{(n-1)}(1) - \psi^{(n-1)}(1+a) & \text{if } n \geq 1, \\ 0 & \text{otherwise,} \end{cases} \tag{C.20}$$

where $\psi^{(n)}$ is the polygamma function of order n. To avoid loss of precision during numerical evaluation, for small values of a the following series expansion is used instead:

$$\partial_b^n \ln B(a,b)|_{b=0} = (-1)^{n+1} \sum_{k=0}^{\infty} (n+k)! \zeta(n+k+1) \frac{(-a)^{k+1}}{(k+1)!}, \tag{C.21}$$

where ζ is the Riemann zeta function.

Exponential Bell polynomials can be efficiently evaluated using the recurrence relation:

$$B_{n+1}(x_1,\ldots,x_{n+1}) = \sum_{i=0}^{n} \binom{n}{i} B_i(x_1,\ldots,x_i) x_{n-i+1}, \qquad B_0 = 1. \tag{C.22}$$

This is the algorithm implemented in `SCETlib`.

C.2. $\mathcal{V}$-distributions

The distributions $\mathcal{V}_a(k,\mu)$ with $a > -1$ are defined as:

$$\mathcal{V}_a(x) \coloneqq \frac{e^{-\gamma_E a}}{\Gamma(1+a)} (a\mathcal{L}^a(x) + \delta(x)), \qquad \mathcal{V}_a(k,\mu) \coloneqq \frac{1}{\mu} \mathcal{V}_a(k/\mu), \tag{C.23}$$

where γ_E is the Euler-Mascheroni constant.

The $\mathcal{V}$-distributions satisfy the following identities:

$$\lim_{a\to 0} \mathcal{V}_a = \mathcal{L}_{-1}, \qquad \frac{\partial}{\partial a} \mathcal{V}_a = \mathcal{L}_0 \otimes \mathcal{V}_a. \tag{C.24}$$

The convolutions of $\mathcal{V}$-distributions are

$$\mathcal{V}_a \otimes \mathcal{V}_b = \mathcal{V}_{a+b}, \qquad \mathcal{V}_a \otimes \mathcal{L}_n = \frac{e^{-\gamma_E a}}{\Gamma(1+a)} \sum_{k=-1}^{n+1} V_k^n(a) \mathcal{L}_k^a. \tag{C.25}$$

D. Order of resummed results

Perturbative results without resummation have the form of a series in some small expansion parameter α_s. In the context of QCD phenomenology the small expansion parameter α_s is the strong coupling constant. Assuming that the approximated quantity is itself of order $\mathcal{O}(1)$, the "n-times-next to leading order" $\mathrm{N^nLO}$ refers to an approximation with perturbative error of order $\mathcal{O}(\alpha_s^{n+1})$. The cases $n = 0, 1, 2$ are denoted LO, NLO, NNLO, respectively.

It should be noted that all leading-order contributions to $B \to X_s\gamma$ and $B \to X_u l\bar{\nu}$ are purely singular, and the nonsingular corrections start at order $\mathcal{O}(\alpha_s)$. For this reason, the first nonzero term in the perturbative series of nonsingular corrections is referred to as the NLO correction, not LO.

In QCD the strong coupling constant $\alpha_s(\mu)$ and coefficients of perturbative series depend on the renormalization scale μ. In the *fixed-order* counting the coefficients of the perturbative series are treated as order $\mathcal{O}(1)$ quantities, regardless of the renormalization scale μ. However, such counting may be inappropriate in presence of multiple well-separated physical energy scales. If the scales μ_1 and μ_0 are well-separated, such that the relative difference $(\alpha_s(\mu_1) - \alpha_s(\mu_0))/\alpha_s(\mu)$ in the strong coupling constant α_s is not small, it is more appropriate to count logarithms $\ln \frac{\mu_1}{\mu_0}$ as parametrically large, of order $\mathcal{O}(\alpha_s^{-1})$. This counting is justified by the following estimate [14]. The QCD β-function is:

$$\beta\big(\alpha_s(\mu)\big) \coloneqq \frac{\partial \alpha_s(\mu)}{\partial \ln \mu} = -2\alpha_s(\mu) \sum_{n=0}^{\infty} \beta_n \left[\frac{\alpha_s(\mu)}{4\pi}\right]^{n+1}. \tag{D.1}$$

Throughout this work the β-function is defined in the $\overline{\mathrm{MS}}$ scheme. The β-function coefficients β_n up to β_3 have been calculated in refs. [59–62]. The β-function is of order $\mathcal{O}(\alpha_s^2)$, and therefore

$$\ln \frac{\mu_1}{\mu_0} = \int_{\alpha_s(\mu_0)}^{\alpha_s(\mu_1)} \frac{\mathrm{d}\alpha}{\beta(\alpha)} \sim \frac{\alpha_s(\mu_1) - \alpha_s(\mu_0)}{\beta(\alpha_s)} \sim \frac{\alpha_s}{\alpha_s^2} \sim \alpha_s^{-1}. \tag{D.2}$$

For example, in the context of B-meson decay, the strong coupling constant is numerically almost 2 times larger at the soft scale $\mu_S \approx 1.3\,\mathrm{GeV}$ in comparison to the hard scale $\mu_H \approx 4.7\,\mathrm{GeV}$. The purpose of resummation is to properly account for corrections enhanced by such large logarithms.

If the perturbatively calculable quantity $C(\mu)$ obeys an RG evolution equation with

anomalous dimension γ,

$$\frac{\partial \ln C(\mu)}{\partial \ln \mu} = \gamma\big(\alpha_s(\mu)\big), \tag{D.3}$$

then resummation can be achieved by rewriting $C(\mu)$ as

$$C(\mu_1) = C(\mu_0) \exp\left(\int_{\mu_0}^{\mu_1} \gamma\big(\alpha_s(\mu)\big) \frac{\mathrm{d}\mu}{\mu}\right) = C(\mu_0) \exp\left(\int_{\alpha_s(\mu_0)}^{\alpha_s(\mu_1)} \frac{\gamma(\alpha)}{\beta(\alpha)} \mathrm{d}\alpha\right), \tag{D.4}$$

and calculating both the $C(\mu_0)$ and the exponent argument $\int_{\mu_0}^{\mu_1} \gamma(\mu)\mathrm{d}\mu/\mu$ perturbatively. This procedure is sometimes called "exponentiation of large logarithms", and the approach is called "RGE-improved perturbation theory". If the large logarithms $\ln \frac{\mu_1}{\mu_0}$ are counted as $\mathcal{O}(\alpha_s^{-1})$, and $\gamma(\mu)$ is of order $\mathcal{O}(1)$ with such counting, then

$$\int_{\mu_0}^{\mu_1} \gamma\big(\alpha_s(\mu)\big)\mathrm{d}\mu/\mu \sim (\ln \mu_1 - \ln \mu_0)\gamma(\mu) \sim \ln \frac{\mu_1}{\mu_0} \sim \alpha_s^{-1}. \tag{D.5}$$

In other words, the leading term in the perturbative series for $\int_{\mu_0}^{\mu_1} \gamma(\mu)\mathrm{d}\mu/\mu$ is of order $\mathcal{O}(\alpha_s^{-1})$. In contrast, here the quantity $C(\mu)$ itself is assumed to be of order $\mathcal{O}(1)$. To calculate $C(\mu_1)$ with perturbative error $\mathcal{O}(\alpha_s^n)$, both the factor $C(\mu_0)$ and the exponent argument $\int_{\mu_0}^{\mu_1} \gamma(\mu)\mathrm{d}\mu/\mu$ must be calculated to the same order $\mathcal{O}(\alpha_s^n)$. The latter requires a calculation of the anomalous dimension $\gamma(\alpha_s)$ with error at most $\mathcal{O}(\alpha_s^{n+1})$, and of the β-function with error not larger than $\mathcal{O}(\alpha_s^{n+3})$. A result with "$n$-times-next to leading-logarithmic" $\mathrm{N^nLL}$ accuracy is an approximation with perturbative error of order $\mathcal{O}(\alpha_s^n) = \mathcal{O}\left(\alpha_s^{n+k} \ln^k \frac{\mu_1}{\mu_0}\right)$, for any k. Therefore, the fixed-order part $C(\mu_0)$ of such a calculation has $\mathrm{N^{n-1}LO}$ accuracy. Cases $n = 0, 1, 2$ are denoted LL, NLL, NNLL.

The anomalous dimension $\gamma(\mu)$ can usually be separated into noncusp and cusp anomalous dimensions γ_{noncusp} and Γ_{cusp}:

$$\gamma\big(\alpha_s(\mu)\big) = \gamma_{\text{noncusp}}\big(\alpha_s(\mu)\big) + c\Gamma_{\text{cusp}}\big(\alpha_s(\mu)\big) \ln \frac{\mu}{\mu_0}, \tag{D.6}$$

where

$$\begin{aligned} \gamma_{\text{noncusp}}\big(\alpha_s(\mu)\big) &= \sum_{n=0}^{\infty} \gamma_n \Big[\frac{\alpha_s(\mu)}{4\pi}\Big]^{n+1}, \\ \Gamma_{\text{cusp}}\big(\alpha_s(\mu)\big) \coloneqq \Gamma^q_{\text{cusp}}\big(\alpha_s(\mu)\big) &= \sum_{n=0}^{\infty} \Gamma_n \Big[\frac{\alpha_s(\mu)}{4\pi}\Big]^{n+1}. \end{aligned} \tag{D.7}$$

The constants c and μ_0 and the noncusp anomalous dimension γ_{noncusp} are different for different quantities C, but the quark cusp anomalous dimension Γ_{cusp} is universal. The quark cusp anomalous dimension coefficients Γ_n up to Γ_4 have been calculated in refs. [63–66]. Because $\ln \frac{\mu}{\mu_0} \sim \alpha_s^{-1}$, the cusp anomalous dimension Γ_{cusp} needs to be

approximated to one order higher in comparison to the noncusp anomalous dimension γ_{noncusp}. As a result, an $\text{N}^\text{n}\text{LL}$ calculation needs to account for all perturbative coefficients up to and including γ_{n-1}, Γ_n, β_n. In particular, the noncusp anomalous dimension γ_{noncusp} can be neglected at LL.

The solutions to the RGE equation can be written in terms of the following standard integrals:

$$\begin{aligned} K_\Gamma(\mu_0,\mu) &= \int_{\alpha_s(\mu_0)}^{\alpha(\mu)} \frac{\mathrm{d}\alpha}{\beta(\alpha)} \Gamma_{\text{cusp}}(\alpha) \int_{\alpha_s(\mu_0)}^{\alpha} \frac{\mathrm{d}\alpha'}{\beta(\alpha')}, \\ \eta_\Gamma(\mu_0,\mu) &= \int_{\alpha_s(\mu_0)}^{\alpha_s(\mu)} \frac{\mathrm{d}\alpha}{\beta(\alpha)} \Gamma_{\text{cusp}}(\alpha), \\ K_{\gamma_{\text{noncusp}}}(\mu_0,\mu) &= \int_{\alpha_s(\mu_0)}^{\alpha_s(\mu)} \frac{\mathrm{d}\alpha}{\beta(\alpha)} \gamma_{\text{noncusp}}(\alpha). \end{aligned} \tag{D.8}$$

In principle, these integrals can be taken analytically [67]. For simplicity, for numerical calculations the standard analytic approximations [54] are used instead. These approximate solutions are derived by taking the integrals after expanding the numerators in α_s.

In a different situation, if the scales μ_0 and μ_1 are close to each other, and the logarithm $\ln\frac{\mu_1}{\mu_0}$ is not large and is instead counted as $\mathcal{O}(1)$, the exponent argument $\int_{\mu_0}^{\mu_1}\gamma(\mu)\mathrm{d}\mu/\mu$ in eq. (D.4) is of order $\mathcal{O}(1)$ as well. As a result, the same number of terms need to be accounted for in the perturbative series of the factor $C(\mu_0)$ and in the exponent argument $\int_{\mu_0}^{\mu_1}\gamma(\mu)\mathrm{d}\mu/\mu$. The primed order $\text{N}^\text{n}\text{LL}'$ refers to such an RGE-improved result, the perturbative error of which is of order $\mathcal{O}(\alpha_s^{n+1})$ if the logarithms $\ln\frac{\mu_1}{\mu_0}$ are counted as $\mathcal{O}(1)$, and is of order $\mathcal{O}(\alpha_s^n)$ if $\ln\frac{\mu_1}{\mu_0}$ is counted as $\mathcal{O}(\alpha_s^{-1})$. The fixed-order part $C(\mu_0)$ of such a calculation has $\text{N}^\text{n}\text{LO}$ accuracy, and an $\text{N}^\text{n}\text{LL}$ solution is used for the exponent argument $\int_{\mu_0}^{\mu_1}\gamma(\mu)\mathrm{d}\mu/\mu$.

Usually only some contributions can be resummed, and the remaining contributions are included at a fixed order. Such combined results, where resummed contributions at $\text{N}^{\text{n}+1}\text{LL}$ or $\text{N}^\text{n}\text{LL}'$ are matched to the remaining fixed-order results at $\text{N}^\text{m}\text{LO}$, are labelled as $\text{N}^{\text{n}+1}\text{LL}+\text{N}^\text{m}\text{LO}$ and $\text{N}^\text{n}\text{LL}'+\text{N}^\text{m}\text{LO}$, respectively. Normally $m = n$, but in chapter 6 theory predictions at $\text{N}^3\text{LL}+\text{NLO}$ are considered as well.

Finally, in chapter 5, the special notation $\text{N}^3\text{LL}'+\text{N}^3\text{LO}(c_k)$ is introduced to emphasize that some perturbative ingredients at this order are not known, but are only parametrized by nuisance parameters c_k.

E. Perturbative ingredients

E.1. $B \to X_s\gamma$ hard function in pole scheme

In the pole mass scheme, the hard function can be written down as

$$h_s(m_b^{\text{pole}},\mu) = \sum_{n=0}^{\infty}\sum_{m=0}^{2n} H_m^{(n)}\left(\frac{\alpha_s(\mu)}{4\pi}\right)^n \ln^m(\mu/m_b^{\text{pole}}). \tag{E.1}$$

It satisfies the following RGE equation:

$$\frac{\partial h_s(m_b^{\text{pole}},\mu)}{\partial \ln\mu} = \left(\gamma^H\big(\alpha_s(\mu)\big) + \Gamma^H\big(\alpha_s(\mu)\big)\ln\frac{\mu}{m_b^{\text{pole}}}\right) h_s(m_b^{\text{pole}},\mu), \tag{E.2}$$

where $\Gamma^H \coloneqq -2\Gamma_{\text{cusp}}$ and $\gamma^H = 2(\gamma^q + \gamma^Q)$ are the hard cusp and noncusp anomalous dimensions, respectively. The noncusp anomalous dimension coefficients $\gamma_n^H = 2(\gamma_n^q + \gamma_n^Q)$ are known to 3-loop order [25]. The nonlogarithmic coefficients of the hard function in eq. (E.1) are known up to 2-loop order [18, 23]:

$$\begin{aligned}
H_0^{(0)} &= 1,\\
H_0^{(1)} &= -C_F\left(12 + \frac{\pi^2}{6}\right),\\
H_0^{(2)} &= 16C_F\Bigg\{3.88611C_F + 5.89413C_A - \left(\frac{7859}{3456} + \frac{109\pi^2}{576} + \frac{13\zeta_3}{48}\right)\beta_0\\
&\qquad + \frac{3563}{1296} - \frac{29\pi^2}{108} - \frac{\zeta_3}{6}\Bigg\},
\end{aligned} \tag{E.3}$$

The 3-loop coefficient $H_0^{(3)}$ is currently unknown and is treated as a nuisance parameter as discussed in section 5.1, where $h_n \coloneqq H_0^{(n)}/4^n$.

The coefficients of the logarithmic terms are determined by the RGE in eq. (E.2). By substituting eq. (E.1) into eq. (E.2), one obtains a recurrence relation that expresses them in terms of the anomalous dimensions and lower-order nonlogarithmic coefficients:

$$H_m^{(n)} = \frac{1}{m}\Bigg\{\sum_{j=0}^{t_1}\big(\gamma_j^H + 2(n-j-1)\beta_j\big)H_{m-1}^{(n-j-1)} + \theta(m\geq 2)\sum_{j=0}^{t_2}\Gamma_j^H H_{m-2}^{(n-j-1)}\Bigg\}, \quad m\geq 1 \tag{E.4}$$

with summation limits $t_1 = \lfloor n-(m+1)/2\rfloor$ and $t_2 = \lfloor n-m/2\rfloor$. The condition $\theta(m\geq 2)$ indicates that the second sum is present only if $m \geq 2$.

The explicit expressions for the logarithmic coefficients up to the 3-loop order read:

$$
\begin{aligned}
H_1^{(1)} &= \gamma_0^H, \\
H_2^{(1)} &= \frac{1}{2}\Gamma_0^H, \\
H_1^{(2)} &= H_0^{(1)}\big(2\beta_0 + \gamma_0^H\big) + \gamma_1^H, \\
H_2^{(2)} &= \frac{1}{2}\Big\{2\beta_0\gamma_0^H + \big(\gamma_0^H\big)^2 + H_0^{(1)}\Gamma_0^H + \Gamma_1^H\Big\}, \\
H_3^{(2)} &= \frac{1}{6}\Gamma_0^H\big(2\beta_0 + 3\gamma_0^H\big), \\
H_4^{(2)} &= \frac{1}{8}\big(\Gamma_0^H\big)^2, \\
H_1^{(3)} &= H_0^{(2)}\big(4\beta_0 + \gamma_0^H\big) + H_0^{(1)}\big(2\beta_1 + \gamma_1^H\big) + \gamma_2^H, \\
H_2^{(3)} &= \frac{1}{2}\Big\{2\beta_1\gamma_0^H + 4\beta_0\gamma_1^H + 2\gamma_0^H\gamma_1^H + H_0^{(2)}\Gamma_0^H \\
&\quad + H_0^{(1)}\Big(8\beta_0^2 + 6\beta_0\gamma_0^H + \big(\gamma_0^H\big)^2 + \Gamma_1^H\Big) + \Gamma_2^H\Big\}, \\
H_3^{(3)} &= \frac{1}{6}\Big\{8\beta_0^2\gamma_0^H + \big(\gamma_0^H\big)^3 + \big(2\beta_1 + 3\gamma_1^H\big)\Gamma_0^H + 3\gamma_0^H\big(H_0^{(1)}\Gamma_0^H + \Gamma_1^H\big) \\
&\quad + \beta_0\Big(6\big(\gamma_0^H\big)^2 + 8H_0^{(1)}\Gamma_0^H + 4\Gamma_1^H\Big)\Big\}, \\
H_4^{(3)} &= \frac{1}{24}\Gamma_0^H\big(8\beta_0^2 + 20\beta_0\gamma_0^H + 6\big(\gamma_0^H\big)^2 + 3H_0^{(1)}\Gamma_0^H + 6\Gamma_1^H\big), \\
H_5^{(3)} &= \frac{1}{24}\big(\Gamma_0^H\big)^2\big(4\beta_0 + 3\gamma_0^H\big), \\
H_6^{(3)} &= \frac{1}{48}\big(\Gamma_0^H\big)^3.
\end{aligned}
\tag{E.5}
$$

The all-order solution of the RGE in eq. (E.2) is given by

$$
\begin{aligned}
h_s(m_b^{\text{pole}}, \mu) &= h_s(m_b^{\text{pole}}, \mu_H) U_H(m_b^{\text{pole}}, \mu_H, \mu), \\
U_H(m_b, \mu_H, \mu) &= \exp\left[-2K_\Gamma(\mu_H, \mu) - 2\eta_\Gamma(\mu_H, \mu)\ln\frac{\mu_H}{m_b} + K_{\gamma^H}(\mu_H, \mu)\right].
\end{aligned}
\tag{E.6}
$$

E.2. Jet function in pole scheme

The perturbative series for the renormalized jet function is

$$
J(s,\mu) = \sum_{n=0}^{\infty}\sum_{m=-1}^{2n-1} J_m^{(n)}\left[\frac{\alpha_s(\mu)}{4\pi}\right]^n \frac{1}{\mu^2}\mathcal{L}_m\Big(\frac{s}{\mu^2}\Big). \tag{E.7}
$$

The jet function is normalized such that $J_{-1}^{(n)} = 1$. Explicit expressions for the coefficients $J_m^{(n)}$ up to 3-loop order can be found in refs. [19, 68].

The jet function obeys the following RGE equation:

$$\frac{\mathrm{d}J(s,\mu)}{\mathrm{d}\ln\mu} = \left\{\Gamma^J[\alpha_s(\mu)]\frac{1}{\mu^2}\mathcal{L}_0\Big(\frac{s}{\mu^2}\Big) + \gamma^J[\alpha_s(\mu)]\delta(s)\right\} \otimes_s J(s,\mu) \tag{E.8}$$

where $\Gamma^J(\alpha_s) = -2\Gamma_{\text{cusp}}(\alpha_s)$. However, because in this work the common renormalization scale μ is always chosen equal to the jet scale μ_J, the jet evolution kernel drops out.

E.3. Partonic soft function

The partonic soft function $C_0(\omega,\mu)$ is pole scheme in defined in eq. (3.24). Its perturbative series is

$$C_0(\omega,\mu) = \sum_{n=0}^{\infty}\sum_{m=-1}^{2n-1} S_m^{(n)}\left[\frac{\alpha_s(\mu)}{4\pi}\right]^n \frac{1}{\mu}\mathcal{L}_m\Big(\frac{\omega}{\mu}\Big). \tag{E.9}$$

The coefficients $S_m^{(n)}$ can be found in ref. [25].

The RGE equation for the soft function is

$$\frac{\mathrm{d}C_0(\omega,\mu)}{\mathrm{d}\ln\mu} = \left\{\Gamma^S[\alpha_s(\mu)]\frac{1}{\mu}\mathcal{L}_0\Big(\frac{\omega}{\mu}\Big) + \gamma^S[\alpha_s(\mu)]\delta(\omega)\right\} \otimes_\omega C_0(\omega,\mu), \tag{E.10}$$

where $\Gamma^S(\alpha_s) = 2\Gamma_{\text{cusp}}(\alpha_s)$ and the noncusp anomalous dimension coefficients γ^S are known to 3-loop order [25].

The soft function RGE equation can be solved to obtain a recurrence relation for the coefficients $S_m^{(n)}$ with $m \geq 0$:

$$S_m^{(n)} = -\frac{1}{m+\delta_{m0}}\left\{\sum_{j=0}^{t_1}\big[\gamma_j^S + 2(n-j-1)\beta_j\big]S_{m-1}^{(n-j-1)} + \sum_{j=0}^{t_2}\sum_{i=t_3}^{2(n-j)-3}\Gamma_j^S V_{m-1}^{0i}\, S_i^{(n-j-1)}\right\}, \tag{E.11}$$

where the summation limits are $t_1 = \lfloor n-1-m/2\rfloor$, $t_2 = \lfloor n - \max(m+1,2)/2\rfloor$, and $t_3 = \max(m,1)-2$. It is easy to check that eq. (E.11) reproduces the explicit results to 3-loop order in ref. [25].

The all-order solution of the soft RGE equation is [23, 69–71]

$$\begin{aligned} C_0(\omega,\mu) &= C_0(\omega,\mu_S) \otimes_\omega \widehat{U}_S(\omega,\mu_S,\mu), \\ \widehat{U}_S(\omega,\mu_S,\mu) &= \exp\Big[-2K_\Gamma(\mu_S,\mu) + K_{\gamma^S}(\mu_S,\mu)\Big]\mathcal{V}\big(2\eta_\Gamma(\mu_S,\mu),\mu_S,\omega\big), \end{aligned} \tag{E.12}$$

where the distribution $\mathcal{V}$ is defined in appendix C.2, K_Γ, η_Γ, K_γ are defined in eq. (D.8), and γ_E is the Euler-Mascheroni constant.

E.4. Singular contribution to $B \to X_u l\bar{\nu}$ decay

The singular contribution $W^{\rm s}_{\mu\nu}(p^+, p^-)$ to $B \to X_u l\bar{\nu}$ decay, which is defined in section 6.4, can be written in terms of the dimensionless variable $x = p^+/p^-$ up to NNLO as follows:

$$
\begin{aligned}
&p^- W^{\rm s}_{\mu\nu}(p^- x, p^-) = H_{\mu\nu}(p^-) \cdot (p^-)^2 (J \otimes C_0)(p^- x) = H^{(0)}_{\mu\nu} \mathcal{L}_{-1}(x) \\
&+ \frac{\alpha_s(p^-)}{\pi} \left\{ \left[\frac{7}{4} \left(1 - \frac{\pi^2}{6} \right) C_F H^{(0)}_{\mu\nu} + H^{(1)}_{\mu\nu}(p^-) \right] \mathcal{L}_{-1}(x) - \frac{7}{4} C_F H^{(0)}_{\mu\nu} \mathcal{L}_0(x) - C_F H^{(0)}_{\mu\nu} \mathcal{L}_1(x) \right\} \\
&+ \left[\frac{\alpha_s(p^-)}{\pi} \right]^2 \left\{ \left[\left(\left\{ -\frac{313\zeta_3}{144} + \frac{50521}{10368} - \frac{1259\pi^2}{1728} + \frac{5\pi^4}{288} \right\} C_A \right. \right. \right. \\
&\qquad \left. + \left\{ \frac{11\zeta_3}{8} + \frac{205}{128} - \frac{47\pi^2}{48} + \frac{401\pi^4}{5760} \right\} C_F + \left\{ -\frac{\zeta_3}{36} - \frac{4073}{2592} + \frac{73\pi^2}{432} \right\} n_f T_F \right) C_F H^{(0)}_{\mu\nu} \\
&\qquad + \frac{7}{4} \left(1 - \frac{\pi^2}{6} \right) C_F H^{(1)}_{\mu\nu}(p^-) + H^{(2)}_{\mu\nu}(p^-) \Big] \mathcal{L}_{-1}(x) \\
&\quad + \left[\left(\left\{ \frac{\zeta_3}{4} - \frac{905}{288} + \frac{17\pi^2}{72} \right\} C_A + \left\{ -\frac{\zeta_3}{2} - \frac{101}{32} + \frac{11\pi^2}{32} \right\} C_F \right. \right. \\
&\qquad \left. + \left\{ \frac{85}{72} - \frac{\pi^2}{18} \right\} n_f T_F \right) C_F H^{(0)}_{\mu\nu} - \frac{7}{4} C_F H^{(1)}_{\mu\nu}(p^-) \Big] \mathcal{L}_0(x) \\
&\quad + \left[C_F H^{(0)}_{\mu\nu} \left(\left\{ \frac{95}{144} + \frac{\pi^2}{12} \right\} C_A + \left\{ \frac{21}{16} + \frac{\pi^2}{8} \right\} C_F - \frac{13 n_f T_F}{36} \right) - C_F H^{(1)}_{\mu\nu}(p^-) \right] \mathcal{L}_1(x) \\
&\quad + C_F \left[\frac{11 C_A}{8} + \frac{21 C_F}{8} - \frac{n_f T_F}{2} \right] H^{(0)}_{\mu\nu} \mathcal{L}_2(x) + \frac{C_F^2}{2} H^{(0)}_{\mu\nu} \mathcal{L}_3(x) \Big\} + \mathcal{O}(\alpha_s^3).
\end{aligned}
\tag{E.13}
$$

The factor p^- on the left-hand side was included to compensate for the mass dimension -1 of the singular contribution $W^{\rm s}_{\mu\nu}$. Here the symbols $H^{(i)}_{\mu\nu}$ are defined as the following coefficients in the perturbative series of the hard function $H_{\mu\nu}$:

$$
H_{\mu\nu}(p^-) =: H^{(0)}_{\mu\nu} + \frac{\alpha_s(p^-)}{\pi} H^{(1)}_{\mu\nu}(p^-) + \left[\frac{\alpha_s(p^-)}{\pi} \right]^2 H^{(2)}_{\mu\nu}(p^-) + \mathcal{O}(\alpha_s^3). \tag{E.14}
$$

The hard function $H_{\mu\nu}$ itself is defined in eq. (6.17). Note that the leading coefficient $H^{(0)}_{\mu\nu}$ does not depend on p^-.

In the above formulae the renormalization scale is set to $\mu = p^-$ in order to suppress terms with logarithms $\ln \frac{\mu}{p^-}$ for compactness of presentation. Because the singular contributions $W^{\rm s}_{\mu\nu}$ are scale-independent order-by-order, the full scale dependence can be easily recovered by expanding the strong coupling constant α_s as follows:

$$
\frac{\alpha_s(p^-)}{\pi} = \frac{\alpha_s(\mu)}{\pi} + \left(\frac{\alpha_s(\mu)}{\pi} \right)^2 \frac{\beta_0}{2} \ln \frac{\mu}{p^-} + \mathcal{O}(\alpha_s^3). \tag{E.15}
$$

Bibliography

[1] P. D. Group, R. L. Workman, V. D. Burkert, V. Crede, E. Klempt, U. Thoma et al., *Review of Particle Physics*, *Progress of Theoretical and Experimental Physics* **2022** (2022) [https://academic.oup.com/ptep/article-pdf/2022/8/083C01/45434166/ptac097.pdf].

[2] J. Charles et al., *Current status of the Standard Model CKM fit and constraints on $\Delta F = 2$ New Physics*, *Phys. Rev. D* **91** (2015) 073007 [1501.05013].

[3] T. Mannel, D. Moreno and A. A. Pivovarov, *NLO QCD corrections to inclusive $b \to c\ell\bar{\nu}$ decay spectra up to $1/m_Q^3$*, *Phys. Rev. D* **105** (2022) 054033 [2112.03875].

[4] C. Ramirez, J. F. Donoghue and G. Burdman, *Semileptonic $b \to u$ decay*, *Phys. Rev. D* **41** (1990) 1496.

[5] CKMfitter group, "Updated results on the CKM matrix (preliminary)." http://ckmfitter.in2p3.fr/www/results/plots_spring21/num/ckmEval_results_spring21.pdf.

[6] Belle collaboration, L. Cao et al., *Measurements of Partial Branching Fractions of Inclusive $B \to X_u \ell^+ \nu_\ell$ Decays with Hadronic Tagging*, *Phys. Rev. D* **104** (2021) 012008 [2102.00020].

[7] SIMBA collaboration, F. U. Bernlochner, H. Lacker, Z. Ligeti, I. W. Stewart, F. J. Tackmann and K. Tackmann, *Precision Global Determination of the $B \to X_s\gamma$ Decay Rate*, *Phys. Rev. Lett.* **127** (2021) 102001 [2007.04320].

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

[9] A. Pich, *Introduction to chiral perturbation theory*, *AIP Conf. Proc.* **317** (1994) 95 [hep-ph/9308351].

[10] T. Cohen, X. Lu and Z. Zhang, *Functional Prescription for EFT Matching*, *JHEP* **02** (2021) 228 [2011.02484].

[11] S. Aoki et al., *Review of lattice results concerning low-energy particle physics*, *Eur. Phys. J. C* **77** (2017) 112 [1607.00299].

[12] J. F. Donoghue, *General relativity as an effective field theory: The leading quantum corrections*, *Phys. Rev. D* **50** (1994) 3874 [gr-qc/9405057].

[13] R. A. Porto, *The effective field theorist's approach to gravitational dynamics*, *Phys. Rept.* **633** (2016) 1 [1601.04914].

[14] T. Becher, A. Broggio and A. Ferroglia, *Introduction to Soft-Collinear Effective Theory*, vol. 896. Springer, 2015, 10.1007/978-3-319-14848-9, [1410.1892].

[15] C. W. Bauer and I. W. Stewart, *Invariant operators in collinear effective theory*, *Phys. Lett. B* **516** (2001) 134 [hep-ph/0107001].

[16] A. V. Manohar and M. B. Wise, *Heavy quark physics*, *Heavy Quark Physics* (2000) .

[17] C. W. Bauer, D. Pirjol and I. W. Stewart, *Soft collinear factorization in effective field theory*, *Phys. Rev. D* **65** (2002) 054022 [hep-ph/0109045].

[18] C. W. Bauer, S. Fleming, D. Pirjol and I. W. Stewart, *An Effective field theory for collinear and soft gluons: Heavy to light decays*, *Phys. Rev. D* **63** (2001) 114020 [hep-ph/0011336].

[19] R. Brüser, Z. L. Liu and M. Stahlhofen, *Three-Loop Quark Jet Function*, *Phys. Rev. Lett.* **121** (2018) 072003 [1804.09722].

[20] C. W. Bauer and A. V. Manohar, *Shape function effects in $B \to X_s\gamma$ and $B \to X_u l\bar{\nu}$ decays*, *Phys. Rev. D* **70** (2004) 034024 [hep-ph/0312109].

[21] S. W. Bosch, B. O. Lange, M. Neubert and G. Paz, *Factorization and shape function effects in inclusive B meson decays*, *Nucl. Phys. B* **699** (2004) 335 [hep-ph/0402094].

[22] T. Becher and M. Neubert, *Toward a NNLO calculation of the $\bar{B} \to X_s\gamma$ decay rate with a cut on photon energy. II. Two-loop result for the jet function*, *Phys. Lett. B* **637** (2006) 251 [hep-ph/0603140].

[23] Z. Ligeti, I. W. Stewart and F. J. Tackmann, *Treating the b quark distribution function with reliable uncertainties*, *Phys. Rev. D* **78** (2008) 114014 [0807.1926].

[24] T. Becher and M. Neubert, *Toward a NNLO calculation of the $\bar{B} \to X_s\gamma$ decay rate with a cut on photon energy: I. Two-loop result for the soft function*, *Phys. Lett. B* **633** (2006) 739 [hep-ph/0512208].

[25] R. Brüser, Z. L. Liu and M. Stahlhofen, *Three-loop soft function for heavy-to-light quark decays*, *JHEP* **03** (2020) 071 [1911.04494].

[26] T. Mannel, *Higher order 1/m corrections at zero recoil*, *Phys. Rev. D* **50** (1994) 428 [hep-ph/9403249].

[27] F. J. Tackmann, *Full-phase-space twist expansion in semileptonic and radiative B-meson decays*, *Phys. Rev. D* **72** (2005) 034036 [hep-ph/0503095].

[28] G. Martinelli, M. Neubert and C. T. Sachrajda, *The Invisible renormalon*, *Nucl. Phys. B* **461** (1996) 238 [hep-ph/9504217].

[29] M. Neubert, *Exploring the invisible renormalon: Renormalization of the heavy quark kinetic energy*, *Phys. Lett. B* **393** (1997) 110 [hep-ph/9610471].

[30] A. F. Falk, M. Neubert and M. E. Luke, *The Residual mass term in the heavy quark effective theory*, *Nucl. Phys. B* **388** (1992) 363 [hep-ph/9204229].

[31] A. H. Hoang, Z. Ligeti and A. V. Manohar, *B decay and the Upsilon mass*, *Phys. Rev. Lett.* **82** (1999) 277 [hep-ph/9809423].

[32] A. H. Hoang, Z. Ligeti and A. V. Manohar, *B decays in the upsilon expansion*, *Phys. Rev. D* **59** (1999) 074017 [hep-ph/9811239].

[33] A. H. Hoang and T. Teubner, *Top quark pair production close to threshold: Top mass, width and momentum distribution*, *Phys. Rev. D* **60** (1999) 114027 [hep-ph/9904468].

[34] A. H. Hoang, A. Jain, I. Scimemi and I. W. Stewart, *Infrared Renormalization Group Flow for Heavy Quark Masses*, *Phys. Rev. Lett.* **101** (2008) 151602 [0803.4214].

[35] A. H. Hoang, *The Top Mass: Interpretation and Theoretical Uncertainties*, in *7th International Workshop on Top Quark Physics*, 12, 2014, 1412.3649.

[36] A. H. Hoang, A. Jain, C. Lepenik, V. Mateu, M. Preisser, I. Scimemi et al., *The MSR mass and the $\mathcal{O}(\Lambda_{\rm QCD})$ renormalon sum rule*, *JHEP* **04** (2018) 003 [1704.01580].

[37] M. Butenschoen, B. Dehnadi, A. H. Hoang, V. Mateu, M. Preisser and I. W. Stewart, *Top Quark Mass Calibration for Monte Carlo Event Generators*, *Phys. Rev. Lett.* **117** (2016) 232001 [1608.01318].

[38] R. Abbate, M. Fickinger, A. H. Hoang, V. Mateu and I. W. Stewart, *Thrust at N^3LL with Power Corrections and a Precision Global Fit for $\alpha_s(mZ)$*, *Phys. Rev. D* **83** (2011) 074021 [1006.3080].

[39] A. Czarnecki, K. Melnikov and N. Uraltsev, *Complete $\mathcal{O}(\alpha_s^2)$ corrections to zero recoil sum rules for $B \to D^*$ transitions*, *Phys. Rev. D* **57** (1998) 1769 [hep-ph/9706311].

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

[41] 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** (1994) 301 [hep-ph/9402364].

[42] K. S. M. Lee, Z. Ligeti, I. W. Stewart and F. J. Tackmann, *Extracting short distance information from $b \to s l^+ l^-$ effectively*, *Phys. Rev. D* **75** (2007) 034016 [hep-ph/0612156].

[43] G. P. Korchemsky and G. F. Sterman, *Infrared factorization in inclusive B meson decays*, *Phys. Lett. B* **340** (1994) 96 [hep-ph/9407344].

[44] I. R. Blokland, A. Czarnecki, M. Misiak, M. Slusarczyk and F. Tkachov, *The Electromagnetic dipole operator effect on $\bar{B} \to X_s \gamma$ at $\mathcal{O}(\alpha_s^2)$*, *Phys. Rev. D* **72** (2005) 033014 [hep-ph/0506055].

[45] K. Melnikov and A. Mitov, *The Photon energy spectrum in $B \to X_s \gamma$ in perturbative QCD through $\mathcal{O}(\alpha_s^2)$*, *Phys. Lett. B* **620** (2005) 69 [hep-ph/0505097].

[46] H. M. Asatrian, T. Ewerth, A. Ferroglia, P. Gambino and C. Greub, *Magnetic dipole operator contributions to the photon energy spectrum in $\bar{B} \to X_s \gamma$ at $\mathcal{O}(\alpha_s^2)$*, *Nucl. Phys. B* **762** (2007) 212 [hep-ph/0607316].

[47] K. S. M. Lee and F. J. Tackmann, *Nonperturbative m_X cut effects in $B \to X_s l^+ l^-$ observables*, *Phys. Rev. D* **79** (2009) 114021 [0812.0001].

[48] F. De Fazio and M. Neubert, *B —> X(u) lepton anti-neutrino lepton decay distributions to order alpha(s)*, *JHEP* **06** (1999) 017 [hep-ph/9905351].

[49] P. Gambino, E. Gardi and G. Ridolfi, *Running-coupling effects in the triple-differential charmless semileptonic decay width*, *JHEP* **12** (2006) 036 [hep-ph/0610140].

[50] M. Brucherseifer, F. Caola and K. Melnikov, *On the $O(\alpha_s^2)$ corrections to $b \to X_u e \bar{\nu}$ inclusive decays*, *Phys. Lett. B* **721** (2013) 107 [1302.0444].

[51] H. M. Asatrian, C. Greub and B. D. Pecjak, *NNLO corrections to anti-B —> X(u) l anti-nu in the shape-function region*, *Phys. Rev. D* **78** (2008) 114028 [0810.0987].

[52] M. Beneke, T. Huber and X. Q. Li, *Two-loop QCD correction to differential semi-leptonic b —> u decays in the shape-function region*, *Nucl. Phys. B* **811** (2009) 77 [0810.1230].

[53] G. Bell, *NNLO corrections to inclusive semileptonic B decays in the shape-function region*, *Nucl. Phys. B* **812** (2009) 264 [0810.5695].

[54] G. Billis, F. J. Tackmann and J. Talbert, *Higher-Order Sudakov Resummation in Coupled Gauge Theories*, *JHEP* **03** (2020) 182 [1907.02971].

[55] Belle collaboration, L. Cao et al., *Measurement of Differential Branching Fractions of Inclusive $B \to X_u \ell^+ \nu_\ell$ Decays*, *Phys. Rev. Lett.* **127** (2021) 261801 [2107.13855].

[56] B. O. Lange, M. Neubert and G. Paz, *Theory of charmless inclusive B decays and the extraction of V(ub)*, *Phys. Rev. D* **72** (2005) 073006 [hep-ph/0504071].

[57] Belle collaboration, M. T. Prim et al., *Search for $B^+ \to \mu^+ \nu_\mu$ and $B^+ \to \mu^+ N$ with inclusive tagging*, *Phys. Rev. D* **101** (2020) 032007 [1911.03186].

[58] M. A. Ebert and F. J. Tackmann, *Resummation of transverse momentum distributions in distribution space*, *Journal of High Energy Physics* **2017** (2017) [1611.08610].

[59] O. V. Tarasov, A. A. Vladimirov and A. Y. Zharkov, *The Gell-Mann-Low Function of QCD in the Three Loop Approximation*, *Phys. Lett. B* **93** (1980) 429.

[60] S. A. Larin and J. A. M. Vermaseren, *The Three loop QCD Beta function and anomalous dimensions*, *Phys. Lett. B* **303** (1993) 334 [hep-ph/9302208].

[61] T. van Ritbergen, J. A. M. Vermaseren and S. A. Larin, *The Four loop beta function in quantum chromodynamics*, *Phys. Lett. B* **400** (1997) 379 [hep-ph/9701390].

[62] M. Czakon, *The Four-loop QCD beta-function and anomalous dimensions*, *Nucl. Phys. B* **710** (2005) 485 [hep-ph/0411261].

[63] G. P. Korchemsky and A. V. Radyushkin, *Renormalization of the Wilson Loops Beyond the Leading Order*, *Nucl. Phys. B* **283** (1987) 342.

[64] S. Moch, J. A. M. Vermaseren and A. Vogt, *The Three loop splitting functions in QCD: The Nonsinglet case*, *Nucl. Phys. B* **688** (2004) 101 [hep-ph/0403192].

[65] J. M. Henn, G. P. Korchemsky and B. Mistlberger, *The full four-loop cusp anomalous dimension in $\mathcal{N} = 4$ super Yang-Mills and QCD*, *JHEP* **04** (2020) 018 [1911.10174].

[66] A. von Manteuffel, E. Panzer and R. M. Schabinger, *Cusp and collinear anomalous dimensions in four-loop QCD from form factors*, *Phys. Rev. Lett.* **124** (2020) 162001 [2002.04617].

[67] M. A. Ebert, *Analytic results for Sudakov form factors in QCD*, *JHEP* **02** (2022) 136 [2110.11360].

[68] J. Gaunt, M. Stahlhofen, F. J. Tackmann and J. R. Walsh, *N-jettiness Subtractions for NNLO QCD Calculations*, *JHEP* **09** (2015) 058 [1505.04794].

[69] C. Balzereit, T. Mannel and W. Kilian, *Evolution of the light cone distribution function for a heavy quark*, *Phys. Rev. D* **58** (1998) 114029 [hep-ph/9805297].

[70] M. Neubert, *Renormalization-group improved calculation of the B —> X(s) gamma branching ratio*, *Eur. Phys. J. C* **40** (2005) 165 [hep-ph/0408179].

[71] S. Fleming, A. H. Hoang, S. Mantry and I. W. Stewart, *Top Jets in the Peak Region: Factorization Analysis with NLL Resummation*, *Phys. Rev. D* **77** (2008) 114003 [0711.2079].