The photon energy spectrum in $B \to X_s\gamma$ at N³LL′

Bahman Dehnadi, Ivan Novikov and Frank J. Tackmann

Deutsches Elektronen-Synchrotron DESY,
Notkestr. 85, 22607 Hamburg, Germany

E-mail: bahman.dehnadi@desy.de, ivan.novikov@desy.de,
frank.tackmann@desy.de

ABSTRACT: We present predictions for the photon energy spectrum in inclusive $B \to X_s\gamma$ decays mediated by the electromagnetic penguin operator O_7 to N³LL′. We use soft-collinear effective theory (SCET) to resum the singular contributions in the peak region at large photon energy. In the tail region the resummed predictions are matched to fixed order at N³LO, where we include the known fixed-order contributions for O_7 up to $\mathcal{O}(\alpha_s^2)$. We develop a method to suitably parametrize the still unknown $\mathcal{O}(\alpha_s^3)$ nonsingular corrections in terms of theory nuisance parameters, whose variations provide an estimate of the associated theory uncertainty. In this context, we also study different ways to treat higher-order cross terms in the matching. Another important aspect of our analysis is the short-distance scheme used for the b-quark mass m_b. We find that in the present context, the $1S$ mass scheme, which was previously used up to 2-loop order, fails to work at 3-loop order, because the mass scheme enters at a soft scale much smaller than m_b here, for which the $1S$ scheme was not devised. Using instead the MSR mass scheme with $R \sim 1\,\mathrm{GeV}$, we obtain stable results with good perturbative convergence up to N³LL′.

Contents

1 Introduction

The flavor-changing neutral-current $b \to s\gamma$ transition plays a key role in exploring the flavor sector of the Standard Model (SM) [1–4] and in searches for possible physics beyond the SM [5–7]. A prominent example is the inclusive $B \to X_s\gamma$ decay, whose normalization is sensitive to beyond-SM contributions.

Experimental measurements of $B \to X_s \gamma$ are most sensitive in the peak region at large photon energy E_γ, where as a result also most information on the normalization of the $B \to X_s \gamma$ rate comes from. Furthermore, the shape of the E_γ spectrum is directly sensitive to the b-quark distribution function, known as shape function, which describes the relevant nonperturbative dynamics of the b quark within the B meson [8–10]. Recently, the first global fit exploiting all available experimental information on the $B \to X_s \gamma$ spectrum [11–14] was carried out by the SIMBA collaboration [15], simultaneously extracting the normalization of the $B \to X_s \gamma$ rate, encoded in the effective inclusive Wilson coefficient C_7^{incl}, the b-quark mass m_b, as well as the shape function.

The shape function is a universal object that also enters the description of inclusive $B \to X_u \ell \nu$ [9, 10, 16] and $B \to X_s \ell^+ \ell^-$ [17, 18] decays, where one restricts the phase space to small hadronic invariant masses to suppress the otherwise overwhelmingly large background from $b \to c \ell \nu$ transitions. In particular, it enters in the extraction of the Cabibbo-Kobayashi-Maskawa (CKM) matrix element $|V_{ub}|$ from $B \to X_u \ell \nu$, which is important for overconstraining the flavor sector of the SM as it is one of the few tree-level quantities. Inclusive determinations of $|V_{ub}|$ show some tensions with its determination from exclusive decays as well as the indirect determination from the CKM unitarity [19].

In the analysis of ref. [15], the theoretical predictions for the $B \to X_s \gamma$ spectrum are obtained at NNLL′+NNLO. The final fit results exhibit a similar size of theoretical and experimental uncertainties. On the experimental side, upcoming and future Belle II measurements [20, 21] of $B \to X_s \gamma$ and $B \to X_u \ell \nu$ will further reduce the experimental uncertainties. To benefit from the improved experimental precision, the theoretical predictions have to be improved likewise.

In this work, we take an important step in this direction by extending the resummed predictions to the next order, N³LL′, taking advantage of the recent computation of the 3-loop jet function and heavy-to-light soft function [22, 23]. At this order, the 3-loop hard function, corresponding to the $b \to s\gamma$ form factor, is also needed but currently not known. Therefore, in our numerical results, we treat its unknown nonlogarithmic constant term at $\mathcal{O}(\alpha_s^3)$ as a theory nuisance parameter [24, 25], which we vary as part of our perturbative uncertainties.

To obtain a complete description of the spectrum away from the endpoint, we have to match to the full fixed-order result. At N³LL′, this requires one to perform the matching to N³LO. Since the full fixed-order results at this order are not known, we devise a method to parametrize the missing ingredients in terms of a set of theory nuisance parameters c_k [24, 25] in such a way that the matching can be performed in a consistent manner and the perturbative uncertainties due to the missing ingredients can be estimated. We denote the so-constructed matched result as N³LL′+N³LO(c_k). It provides a description of the $B \to X_s \gamma$ spectrum which benefits from the improved precision at N³LL′ in the peak region while consistently including the known fixed-order results up to $\mathcal{O}(\alpha_s^2)$ [26–28].

Throughout the paper, we focus on the contributions from the electromagnetic penguin operator, O_7, in the electroweak Hamiltonian, which induces the to-be-resummed singular contributions that dominate at large photon energies. At sufficiently high order, operators other than O_7 also produce singular contributions, but these are always O_7-like and are

automatically included via the definition of C_7^{incl}. The purely nonsingular contributions from non-O_7 operators can simply be added to the order they are known as in ref. [15], so we do not discuss them here further.

Extending the resummation order to $\text{N}^3\text{LL}'$ turns out to be more subtle than one might naively expect. We find that the $1S$ mass scheme [29–31], which was used at NNLL' in refs. [15, 32], fails to provide a stable prediction at $\text{N}^3\text{LL}'$. The cause for this failure lies in the intrinsic scale $R^{1S}(\mu_S)$ of the $1S$ scheme, which at the soft scale μ_S becomes large and incompatible with the power counting of HQET. On the other hand, by using the MSR scheme [33] we are able to obtain stable predictions that exhibit good perturbative convergence. Furthermore, at $\text{N}^3\text{LL}'$ it turns out to become necessary to switch to a short-distance mass scheme also in the jet and hard functions, which makes the RGE of the hard function more involved, where m_b plays the role of setting the hard kinematic scale.

The outline of this paper is as follows. In section 2 we present our theory framework and discuss all the ingredients for computing the $B \to X_s\gamma$ photon energy spectrum at $\text{N}^3\text{LL}'+\text{N}^3\text{LO}(c_k)$. We discuss the implementation of the short-distance mass scheme in more detail in section 3. Our methodology for estimating the perturbative uncertainties of our predictions is given in section 4. In section 5, we present our numerical results and discuss alternative choices for the b-quark mass scheme and the treatment of higher-order perturbative terms. We summarize our findings in section 6.

2 Theory framework

2.1 Overview

We follow the setup of ref. [15] and write the $B \to X_s\gamma$ photon energy spectrum as

$$\frac{\mathrm{d}\Gamma}{\mathrm{d}E_\gamma} = 2\Gamma_0 \left(\frac{2E_\gamma}{\widehat{m}_b}\right)^3 \int \mathrm{d}k\, \widehat{P}(k)\, \mathcal{F}(m_B - 2E_\gamma - k) + \mathcal{O}\left(\frac{\Lambda_{\text{QCD}}}{\widehat{m}_b}\right), \qquad (2.1)$$

where $\widehat{m}_b$ denotes the b-quark mass in a short-distance scheme and

$$\Gamma_0 = \frac{G_F^2\, \widehat{m}_b^5}{8\pi^3} \frac{\alpha_{\text{em}}}{4\pi} |V_{tb}V_{ts}^*|^2\,. \qquad (2.2)$$

The function $\widehat{P}(k)$ is perturbatively calculable and corresponds to the partonic $b \to s\gamma$ spectrum with $k \sim \widehat{m}_b - 2E_\gamma$. We write it as [15]

$$\widehat{P}(k) = \left|C_7^{\text{incl}}\right|^2 \left[W_{77}^{\text{s}}(k) + W_{77}^{\text{ns}}(k)\right] + 2\,\text{Re}(C_7^{\text{incl}})\sum_{i\neq 7}\mathcal{C}_i W_{7i}^{\text{ns}}(k) + \sum_{i,j\neq 7}\mathcal{C}_i\mathcal{C}_j W_{ij}^{\text{ns}}(k)\,. \qquad (2.3)$$

The coefficient C_7^{incl} contains by definition all virtual contributions from operators in the electroweak Hamiltonian that give rise to singular contributions [15, 34]. It is dominated by the Wilson coefficient C_7 of the electromagnetic operator O_7. In this paper, we focus on the contributions proportional to $\left|C_7^{\text{incl}}\right|^2$, which are discussed in more detail in the following.

The function $W_{77}^{\text{s}}(k)$ in eq. (2.3) accounts for the contributions to the partonic spectrum $\sim \delta(k)$ and $\ln^n(k/\widehat{m}_b)/k$ that are singular and dominant in the peak of the spectrum

where $k \ll \widehat{m}_b$. They can be resummed to all orders based on their well-known factorization [35, 36]. We will resum them to $\mathrm{N}^3\mathrm{LL}'$, as discussed in more detail in section 2.2. Note that the overall factor $(2E_\gamma/\widehat{m}_b)^3$ in eq. (2.1) 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. As in ref. [15], we factor it out of the singular contributions and keep it unexpanded in the endpoint region.

The function $W_{77}^{\mathrm{ns}}(k)$ in eq. (2.3) contains the remaining nonsingular contributions to the partonic spectrum. They start at $\mathcal{O}(\alpha_s)$ and are accompanied by powers of $k/\widehat{m}_b$ relative to the singular contributions in W_{77}^{s}. Thus, they are power-suppressed in the peak region where $k \ll \widehat{m}_b$. On the other hand, in the tail region where $k \sim \widehat{m}_b$, they are of similar size as the singular contributions. We will include their known results at fixed order to $\mathcal{O}(\alpha_s^2)$, while the currently unknown corrections at $\mathcal{O}(\alpha_s^3)$ are estimated by introducing appropriate theory nuisance parameters, as discussed in section 2.3.

The remaining non-77 contributions W_{i7}^{ns} and W_{ij}^{ns} in eq. (2.3) are purely nonsingular and thus only become relevant in the tail region. Since they do not have singular counterparts, they can be straightforwardly added to the order they are known, as was done in ref. [15]. We neglect them in the following, since our focus here is on the resummation and fixed-order matching of the dominant 77 contributions. Similarly, we neglect the remaining $\mathcal{O}(\Lambda_{\mathrm{QCD}}/\widehat{m}_b)$ terms in eq. (2.1), which contain subdominant resolved and unresolved contributions.

To ensure that eq. (2.1) reproduces the full fixed-order result in the tail with a smooth transition between the peak and tail regimes, we use profile scales to gradually switch off the resummation away from the peak region, as discussed in section 2.4.

The nonperturbative function $\mathcal{F}(k)$ in eq. (2.1) contains the leading shape function as well as the combination of subleading shape functions that appear at tree level in $B \to X_s\gamma$. It is discussed in section 2.5. The partonic spectrum $\widehat{P}$ and the shape function $\mathcal{F}$ are completely factorized in eq. (2.1). This factorization enables a coherent description of the spectrum in both peak and tail region. In the tail region only the first few moments of $\mathcal{F}$ are relevant, while in the peak region its full form is required.

2.2 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 [35, 36] allows us to systematically resum the large logarithmic distributions to all orders in perturbation theory. Here we make use of the SCET-based factorization theorem following ref. [32]

$$
\begin{aligned}
W_{77}^{\mathrm{s}}(k) = {}& \widehat{h}_s(\widehat{m}_b, \mu_H)\, \widehat{U}_H(\widehat{m}_b, \mu_H, \mu_J) \\
& \times \int \mathrm{d}\omega\, \mathrm{d}\omega'\, \widehat{m}_b\, \widehat{J}(\widehat{m}_b(k-\omega), \mu_J)\, U_S(\omega - \omega', \mu_S, \mu_J)\, \widehat{C}_0(\omega', \mu_S)\,,
\end{aligned} \tag{2.4}
$$

where $\widehat{h}_s$, $\widehat{J}$, and $\widehat{C}_0$ are the hard, jet, and partonic soft functions, respectively. The hard and soft evolution kernels, $\widehat{U}_H$ and U_S, evolve the hard and partonic soft functions from their characteristic hard and soft scales, μ_H and μ_S, to the jet scale, μ_J, thereby summing logarithms of the form $\ln(\mu_H/\mu_J)$ and $\ln(\mu_J/\mu_S)$. Since we choose to evolve

everything to the jet scale, the jet evolution kernel $\widehat{U}_J(p^2, \mu_J, \mu_J) = \delta(p^2)$ drops out. The hats indicate that an object is defined in a renormalon-free short-distance scheme, as discussed in more detail in section 3. Explicit results for the perturbative ingredients are collected in appendix A.

In eq. (2.4), $\widehat{h}_s(\mu_H)$, $\widehat{J}(\mu_J)$, $\widehat{C}_0(\mu_S)$ are the boundary conditions for the evolution, which are evaluated at fixed order. When doing so, by default we always strictly reexpand their product to the given order in α_s, i.e., we count $\alpha_s(\mu_H) \sim \alpha_s(\mu_J) \sim \alpha_s(\mu_S)$ and drop all higher-order cross terms in the product of their fixed-order series. By doing so, the strict fixed-order expansion of $W_{77}^{\rm s}$ in terms of a common $\alpha_s(\mu)$ is reproduced simply by taking all scales to be equal $\mu_H = \mu_J = \mu_S = \mu$. This is different to refs. [15, 32], where only the product $\widehat{J} \otimes \widehat{C}_0$ is reexpanded, 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 effect of these differences is studied in section 5.2.

To resum the singular corrections using eq. (2.4) to N³LL′ order, we have to include the fixed-order boundary conditions $\widehat{h}_s(\mu_H)$, $\widehat{J}(\mu_J)$, $\widehat{C}_0(\mu_S)$ to $\mathcal{O}(\alpha_s^3)$ and use the 3-loop noncusp and 4-loop cusp anomalous dimensions as well as the 4-loop beta function in the evolution factors $\widehat{U}_H(\mu_H, \mu_J)$ and $U_S(\mu_J, \mu_S)$. The jet and partonic soft functions have been computed up to three loops in refs. [16, 22, 37] and refs. [16, 23, 38], respectively.

Regarding the hard function, its 3-loop anomalous dimension is also known via the consistency relation with the jet and soft anomalous dimensions and is given in ref. [23]. The full hard function is currently only known up to NNLO [32, 39]. We account for all the logarithmic terms at N³LO, which are determined using the RGE in terms of the known anomalous dimensions and lower-order constant terms. The result is given in appendix A.2. Thus, the only missing ingredient to obtain $W_{77}^{\rm s}$ at full N³LL′ is the finite, nonlogarithmic 3-loop constant of the hard function, h_3, which is defined by expanding the (pole-scheme) hard function at $\mu = m_b$ as

$$h_s(m_b, \mu_H = m_b) = 1 + \frac{\alpha_s(m_b)}{\pi} h_1 + \frac{\alpha_s^2(m_b)}{\pi^2} h_2 + \frac{\alpha_s^3(m_b)}{\pi^3} h_3 + \mathcal{O}(\alpha_s^4). \qquad (2.5)$$

We treat this unknown constant as a theory nuisance parameter,

$$h_3 = 0 \pm 80, \qquad (2.6)$$

where the range of variation is estimated using the Padé approximation

$$h_3 \sim \frac{h_2^2}{|h_1|} \sim \frac{19.3^2}{4.55} \sim 80. \qquad (2.7)$$

We will see in section 5 that it only has a minor impact on the perturbative precision of our results.

For future reference, we write the fixed-order expansion of the singular contribution up to N³LO as

$$
\begin{aligned}
W_{77}^{\mathrm{s}}(\widehat{m}_b x) = \frac{C_F}{\widehat{m}_b} \bigg\{ & w_{77}^{\mathrm{s}(0)}(x) + \frac{\alpha_s(\mu)}{\pi}\big[w_{77}^{\mathrm{s}(1)}(x) + \Delta w_{77}^{\mathrm{s}(1)}(\mu, x) \big] \\
& + \frac{\alpha_s^2(\mu)}{\pi^2}\bigg[w_{77}^{\mathrm{s}(2)}(x) + \frac{\beta_0}{2} w_{77}^{\mathrm{s}(1)}(x) \ln\frac{\mu}{\widehat{m}_b} + \Delta w_{77}^{\mathrm{s}(2)}(\mu, x) \bigg] \\
& + \frac{\alpha_s^3(\mu)}{\pi^3}\bigg[w_{77}^{\mathrm{s}(3)}(x) + \bigg(\beta_0 w_{77}^{\mathrm{s}(2)}(x) + \frac{\beta_1}{8} w_{77}^{\mathrm{s}(1)}(x) \bigg) \ln\frac{\mu}{\widehat{m}_b} \\
& + \frac{\beta_0^2}{4} w_{77}^{\mathrm{s}(1)}(x) \ln^2\frac{\mu}{\widehat{m}_b} + \Delta w_{77}^{\mathrm{s}(3)}(\mu, x) \bigg] + \mathcal{O}(\alpha_s^4) \bigg\}, \qquad (2.8)
\end{aligned}
$$

where we have made the fixed-order μ dependence and its order-by-order cancellation explicit. Numerically, we have

$$
\begin{aligned}
w_{77}^{\mathrm{s}(0)}(x) =\, & 0.75\,\delta(x)\,, \\
w_{77}^{\mathrm{s}(1)}(x) =\, & {-4.54}\,\delta(x) - 1.75\,\mathcal{L}_0(x) - 1.00\,\mathcal{L}_1(x)\,, \\
w_{77}^{\mathrm{s}(2)}(x) =\, & (-30.5 + 3.01\,n_f)\,\delta(x) + (5.94 + 0.316\,n_f)\,\mathcal{L}_0(x) + (12.4 - 0.181\,n_f)\,\mathcal{L}_1(x) \\
& + (7.63 - 0.250\,n_f)\,\mathcal{L}_2(x) + 0.667\,\mathcal{L}_3(x)\,, \\
w_{77}^{\mathrm{s}(3)}(x) =\, & (88.2 + 0.75 h_3 - 0.0269\,n_f - 0.0309\,n_f^2)\,\delta(x) \qquad\qquad (2.9) \\
& + (138 - 9.16\,n_f - 0.00330\,n_f^2)\,\mathcal{L}_0(x) \\
& + (68.7 - 8.75\,n_f + 0.121\,n_f^2)\,\mathcal{L}_1(x) + (-25.1 - 0.815\,n_f + 0.00694\,n_f^2)\,\mathcal{L}_2(x) \\
& + (-43.0 + 3.16\,n_f - 0.0648\,n_f^2)\,\mathcal{L}_3(x) + (-6.53 + 0.278\,n_f)\,\mathcal{L}_4(x) - 0.222\,\mathcal{L}_5(x)\,,
\end{aligned}
$$

where $\mathcal{L}_n(x) \equiv [\ln^n(x)/x]_+$ are the usual plus distributions defined in eq. (A.4). 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. (2.8) 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 3.4, and their explicit expressions are provided in eq. (3.12).

2.3 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 $\mathcal{O}(\alpha_s^2)$ [26–28]. We write its perturbative expansion up to N³LO as

$$
\begin{aligned}
W_{77}^{\mathrm{ns}}(\widehat{m}_b x) = \frac{C_F}{\widehat{m}_b(1-x)^3} \bigg\{ &\frac{\alpha_s(\mu_{\mathrm{ns}})}{\pi}\, w_{77}^{\mathrm{ns}(1)}(x) \\
&+ \frac{\alpha_s^2(\mu_{\mathrm{ns}})}{\pi^2} \left[w_{77}^{\mathrm{ns}(2)}(x) + \frac{\beta_0}{2}\, w_{77}^{\mathrm{ns}(1)}(x)\, \ln\frac{\mu_{\mathrm{ns}}}{\widehat{m}_b} + \Delta w_{77}^{\mathrm{ns}(2)}(\mu_{\mathrm{ns}}, x) \right] \\
&+ \frac{\alpha_s^3(\mu_{\mathrm{ns}})}{\pi^3} \bigg[w_{77}^{\mathrm{ns}(3)}(x) + \left(\beta_0 w_{77}^{\mathrm{ns}(2)}(x) + \frac{\beta_1}{8} w_{77}^{\mathrm{ns}(1)}(x) \right) \ln\frac{\mu_{\mathrm{ns}}}{\widehat{m}_b} \\
&+ \frac{\beta_0^2}{4}\, w_{77}^{\mathrm{ns}(1)}(x) \ln^2\frac{\mu_{\mathrm{ns}}}{\widehat{m}_b} + \Delta w_{77}^{\mathrm{ns}(3)}(\mu_{\mathrm{ns}}, x) \bigg] \bigg\}. \qquad (2.10)
\end{aligned}
$$

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. [15]. The $\mathcal{O}(\alpha_s^3)$ 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. (3.13).

In the peak region for $k \ll \widehat{m}_b$, the nonsingular contributions are power-suppressed by $k/\widehat{m}_b$ relative to the singular. Hence, 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 are being resummed. By contrast, for $k \sim \widehat{m}_b$, the separation into singular and nonsingular becomes ill-defined 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 for $k \to \widehat{m}_b$, as we will see explicitly in section 2.4. Consequently, as already pointed out in refs. [40, 41], 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 $\mathrm{N}^n\mathrm{LL}'$ 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 N³LL′ we need W_{77}^{ns} to $\mathcal{O}(\alpha_s^3)$, which means we have to parametrize the unknown nonsingular function $w_{77}^{\mathrm{ns}(3)}(x)$. We do so by considering its required asymptotic behavior in the $x \to 0$ and $x \to 1$ limits. In particular, for $x \to 1$ we have to account for the singular-nonsingular cancellations at $\mathcal{O}(\alpha_s^3)$, which basically implies that $w_{77}^{\mathrm{ns}(3)}$ and $w_{77}^{\mathrm{s}(3)}$ are not independent. In addition, we want to exploit the parameterization to estimate the perturbative uncertainty due to the missing $w_{77}^{\mathrm{ns}(3)}(x)$.

We begin by separating $w_{77}^{\mathrm{ns}(3)}(x)$ into a "correlated" and an "uncorrelated" piece,

$$
w_{77}^{\mathrm{ns}(3)}(x) = w_{\mathrm{cor}}^{\mathrm{ns}(3)}(x) + w_{\mathrm{uncor}}^{\mathrm{ns}(3)}(x). \qquad (2.11)
$$

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. We define it as

$$
w_{\mathrm{cor}}^{\mathrm{ns}(3)}(x) = -(1-x)^3\, w_{77}^{\mathrm{s}(3)}(1), \qquad (2.12)
$$

where the 3-loop singular function $w_{77}^{\mathrm{s}(3)}(x)$ is defined in eq. (2.8). The overall factor $(1-x)^3$ simply cancels the overall $1/(1-x)^3$ in eq. (2.10). The remaining "uncorrelated" piece $w_{\mathrm{uncor}}^{\mathrm{ns}(3)}(x)$ can now be considered independent of the singular contribution, so we can parametrize it. We do so by expanding it as

$$w_{\mathrm{uncor}}^{\mathrm{ns}(3)}(x) = (1-x)^3 \sum_{k=0}^{5} c_k^{\mathrm{ns}} L^k(x) \qquad \text{with} \qquad L(x) = \frac{1}{4} \frac{w_{77}^{\mathrm{ns}(1)}(x)}{(1-x)^3} - \frac{9}{16}, \qquad (2.13)$$

where the function $L(x)$ is positive for $0 < x < 1$ and has the following asymptotics in the $x \to 0$ and $x \to 1$ limits,

$$L(x) = -\frac{3}{2} - \ln x + \mathcal{O}(x), \qquad\qquad L(x) = 0 + \mathcal{O}(1-x). \qquad (2.14)$$

We construct $L(x)$ 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. (2.13) incorporates the knowledge that at $\mathcal{O}(\alpha_s^3)$ the nonsingular in the limit $x \to 0$ is a degree-5 polynomial in $\ln x$. Therefore we include up to five powers of $L(x)$ to ensure we are able to probe the complete logarithmic structure in the small-x limit.

The parameters $c_{0...5}^{\mathrm{ns}}$ can be treated as theory nuisance parameters with zero central values and the following variation magnitudes:

$$c_k^{\mathrm{ns}} = 0 \pm \delta c_k^{\mathrm{ns}} \qquad \text{with} \qquad \delta c_{0...5}^{\mathrm{ns}} = (20, 100, 80, 10, 5, 1). \qquad (2.15)$$

To determine the range of variations for $c_{1...5}^{\mathrm{ns}}$, we use 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,

$$4x\, w_{77}^{\mathrm{s}(1)}(x) = -7.00 - 4.00 \ln x,$$
$$w_{77}^{\mathrm{ns}(1)}(x) = 3.75 - 4.00 \ln x + \mathcal{O}(x),$$
$$4x\, w_{77}^{\mathrm{s}(2)}(x) = 28.8 + 46.7 \ln x + 26.5 \ln^2 x + 2.67 \ln^3 x,$$
$$w_{77}^{\mathrm{ns}(2)}(x) = 16.1 + 33.9 \ln x + 25.0 \ln^2 x + 2.67 \ln^3 x + \mathcal{O}(x),$$
$$4x\, w_{77}^{\mathrm{s}(3)}(x) = 406 + 142 \ln x - 113 \ln^2 x - 125 \ln^3 x - 21.7 \ln^4 x - 0.889 \ln^5 x,$$
$$= 259 + 95 L(x) + 189 L^2(x) + 15.5 L^3(x) - 15.0 L^4(x) + 0.889 L^5(x). \qquad (2.16)$$

In particular, the highest power of $\ln x$ in $w_{77}^{\mathrm{ns}(1,2)}(x)$ is precisely determined by $4x\, w_{77}^{\mathrm{s}(1,2)}$. The reason is that, similar to the leading-power case, it is expected that the universal cusp and a set of subleading noncusp anomalous dimensions govern the logarithmic structure at subleading power. Thus, we similarly exploit $4x\, w_{77}^{\mathrm{s}(3)}(x)$ for estimating the typical size of $\delta c_{1...5}^{\mathrm{ns}}$.

To assess the uncertainty due to the missing $w_{77}^{\mathrm{ns}(3)}$ we separately vary each nuisance parameter in the ranges given above. The different nuisance parameters are considered as independent such that the resulting individual uncertainties from varying them are added

in quadrature. This is discussed in more detail in section 4. Of course, this means that the uncertainty necessarily increases by adding more parameters. Therefore, to obtain a realistic uncertainty estimate and avoid becoming overly conservative we also put the constraint that the total uncertainty estimate for the missing 3-loop correction does not exceed the size of the 2-loop corrections. To satisfy this constraint, the values for $\delta c_{2,3,4}^{\rm ns}$ are chosen somewhat smaller than the corresponding coefficients of $4x\, w_{77}^{{\rm s}(3)}(x)$ in eq. (2.16).

It is easy to see from eq. (2.14) 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. Therefore, we estimate the size of $\delta c_0^{\rm ns}$ using the Padé approximation from the corresponding lower-order corrections,

$$
w_{x\to 1}^{(k)} = \lim_{x\to 1}\left[\frac{w_{77}^{{\rm ns}(k)}(x)}{(1-x)^3} + w_{77}^{{\rm s}(k)}(x)\right]\,, \quad w_{x\to 1}^{(3)} = c_0^{\rm ns} \sim \frac{(w_{x\to 1}^{(2)})^2}{w_{x\to 1}^{(1)}} \approx \frac{2.92^2}{0.5} \sim 20\,. \quad (2.17)
$$

Finally, we like to stress that the purpose of the parameterization given by eqs. (2.11) and (2.13) is *not* to construct an approximation of the unknown function $w_{77}^{{\rm ns}(3)}(x)$. Rather, the goal is first to enable a consistent matching at ${\rm N^3LL'}$, which is essentially achieved by the separation in eq. (2.11), and second to obtain a reliable estimate of the perturbative uncertainty due to its unknown form. For this purpose, we only need to estimate the *typical size* of the theory nuisance parameters, say within a factor of a few, for which the above considerations are sufficient. However, since eq. (2.10) includes all known $\mathcal{O}(\alpha_s^3)$ contributions that are predictable from lower orders, and since the parameterization of the remaining unknown $w_{77}^{{\rm ns}(3)}(x)$ does include nontrivial information on its structure, we do expect some improvement in the perturbative precision of our predictions beyond $\mathcal{O}(\alpha_s^2)$, which will be reflected in the size of the resulting uncertainty, as we will see in section 5.

2.4 Profile scales

The resummation of the singular contributions is determined by the choices of the hard μ_H, jet μ_J, and soft μ_S scales in the factorization theorem in eq. (2.4). 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. (2.10) are evaluated.

For our scale choices we follow ref. [15]. The E_γ spectrum has three parametrically distinct kinematic regions:

- **Shape function (nonperturbative) region:** $\Lambda_{\rm QCD} \sim (m_B - 2\,E_\gamma) \ll \widehat{m}_b$
  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 region left of the peak, i.e., the transition region between the peak and the far tail. Here the soft scale has canonical scaling $\mu_S \sim m_B - 2E_\gamma$.

- **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 $\mu_S = \mu_J = \mu_H = \mu_{\rm ns} \equiv \mu_{\rm FO} \sim \widehat{m}_b$.

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 we use the common approach of profile scales [32, 42], where $\mu_S(E_\gamma)$, $\mu_J(E_\gamma)$, $\mu_{\rm ns}(E_\gamma)$ are taken as functions of E_γ. The key advantage of using profile scales to connect the descriptions in the different parametric regions is that they provide a smooth transition between the regions that is solely implemented in terms of scale choices, such that the ambiguity in the precise choices of the transition are equivalent to scale ambiguities, which by construction are formally beyond the order one is working and reduce as we go to higher order.

For our numerical analysis we employ the profile scales used in ref. [15], given by

$$\mu_H = e_H \widehat{m}_b \,,$$
$$\mu_S(E_\gamma) = \mu_0 + (\mu_H - \mu_0)\, f_\theta\!\left(\frac{E_1 - E_\gamma}{E_1 - E_2}\right),$$
$$\mu_J(E_\gamma) = \left[\mu_S(E_\gamma)\right]^{(1-e_J)/2} \mu_H^{(1+e_J)/2} \,,$$
$$\mu_{\rm ns}(E_\gamma) = \left[\mu_S(E_\gamma)\right]^{(1-e_{\rm ns})/4} \mu_H^{(3+e_{\rm ns})/4} \,, \tag{2.18}$$

where the function $f_\theta(x)$ provides a smooth transition from $f_\theta(x \le 0) = 0$ to $f_\theta(x \ge 1) = 1$,

$$f_\theta(x) = \begin{cases} 0 & x \le 0 \,, \\ 2\,x^2 & 0 < x \le 1/2 \,, \\ 1 - 2\,(1-x)^2 & 1/2 < x \le 1 \,, \\ 1 & 1 < x \,. \end{cases} \tag{2.19}$$

Figure 1 shows the absolute values of the singular and nonsingular contributions as well as the full result at NNLO (i.e. without resummation), which is used to pick the transition points E_1 and E_2 between the different parametric regions. For $E_\gamma \gtrsim E_1 = 2.2\,{\rm GeV}$, the singular contributions clearly dominate, which also corresponds to the nonperturbative region. For $E_\gamma \lesssim E_2 = 1.6\,{\rm GeV}$, 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 ill-defined and only their sum is meaningful. Therefore, the resummation of the singular must be turned off to ensure that this cancellation is not spoiled by it and the correct full result is recovered. Since there is little space between E_1 and E_2, the transition region in between them effectively coincides with the intermediate shape function OPE region.
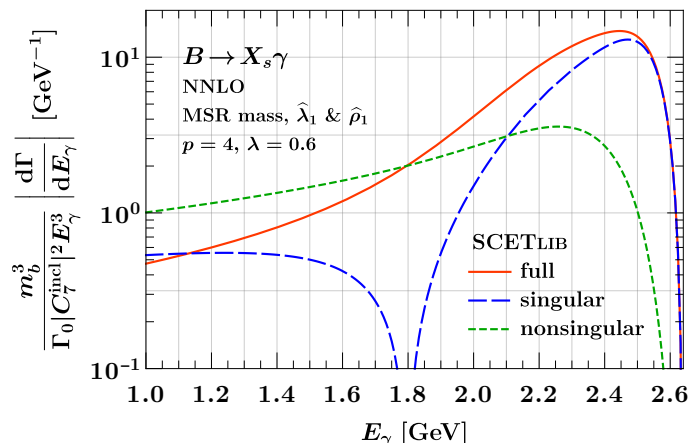
Figure 1. Comparison of the absolute values of the singular, nonsingular, and full contributions to the $B \to X_s \gamma$ photon energy spectrum at fixed NNLO, using the MSR scheme for the b-quark mass and short-distance definitions of the hadronic parameters $\widehat{\lambda}_1$ and $\widehat{\rho}_1$, which are defined in section 3.1.

To summarize we use the following values for the profile scale parameters [15]:

$$
\begin{aligned}
e_H &= \{1, 1/2, 2\}, & \mu_0 &= \max(1, e_H) \times \{1.3, 1.1, 1.8\} \,\text{GeV}, \\
e_J &= \{0, -1/3, +1/3\}, & e_{\text{ns}} &= \{0, -1, +1\}, \\
E_1 &= \{2.2, 2.1, 2.3\} \,\text{GeV}, & E_2 &= 1.6 \,\text{GeV}.
\end{aligned}
\tag{2.20}
$$

For each parameter, the first value in the set is the central value and the next two are the variations that we will use to assess perturbative uncertainties in section 4. The central values for the profile scales along with their individual variation ranges are illustrated in figure 2. Since the nonsingular contributions are treated at fixed order, a priori there is no canonical scaling to guide the choice of the nonsingular scale μ_{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 $\widehat{m}_b$, and this choice is varied up to the hard and down to the central jet scales as shown in figure 2.

Note that the individual scales are not independent of each other but are parametrized in such a way that their relative hierarchies are preserved upon varying the profile 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 μ_{ns} depend on μ_S, such that varying μ_0 not only varies μ_S but also moves μ_J and μ_{ns} up and down accordingly to preserve the hierarchy between them. The variations for μ_J and μ_{ns} parametrized by e_J and e_{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 like the resummation itself, such that below E_2 only the overall μ_H variation remains corresponding to the usual fixed-order scale variation.
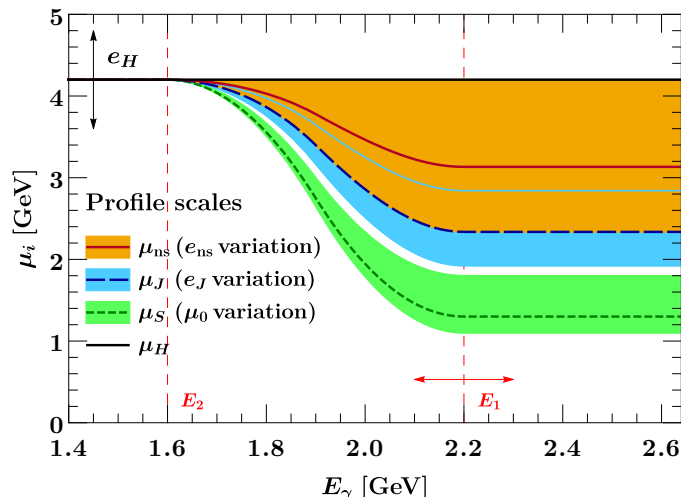
Figure 2. Profile scales used for the hard, jet, soft, and nonsingular scales. The bands show the individual ranges for the jet, soft, and nonsingular scales. The black and red arrows indicate the variations for the parameter e_H and the transition point E_1.

2.5 Shape function

The leading-power factorization theorem for $B \to X_s \gamma$ in eq. (2.4) separates all soft dynamics into the hadronic soft function

$$S(\omega, \mu) \equiv \langle B | \bar{b}_v \delta(iD_+ - \delta + \omega) b_v | B \rangle \,, \tag{2.21}$$

where b_v is the HQET b-quark field, $|B\rangle$ is the full QCD B-meson state, and $\delta = m_B - m_b^{\text{pole}}$. This definition of $S(\omega, \mu)$ is such that it has support for $\omega \geq 0$ [32]. The hadronic soft function contains both perturbative soft radiation as well as the nonperturbative Fermi motion of the b quark inside the B meson. Following ref. [32], we further factorize it as

$$S(\omega, \mu) = \int dk \, \widehat{C}_0(\omega - k, \mu) \, \mathcal{F}(k) \,, \tag{2.22}$$

where the partonic soft function $\widehat{C}_0$ can be calculated in perturbation theory, while the shape function $\mathcal{F}(k)$ is a nonperturbative object. It has support for $k \geq 0$ and peaks around $k \sim \Lambda_{\text{QCD}}$. Following conventional terminology, we will often refer to $S(\omega, \mu)$ as just the soft function.

In the tail region where $\omega \gg \Lambda_{\text{QCD}}$, the right-hand side of eq. (2.22) can be expanded in powers of $\Lambda_{\text{QCD}}/\omega$,

$$S(\omega, \mu) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \frac{d^n \widehat{C}_0(\omega, \mu)}{d\omega^n} M_n \,, \tag{2.23}$$

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

$$M_n \equiv \int dk \, k^n \mathcal{F}(k) \,. \tag{2.24}$$

Hence, in the tail region the leading nonperturbative corrections are encoded in the first few nonperturbative moments of $\mathcal{F}(k)$, and its moment expansion recovers the local-OPE

description of the spectrum. On the other hand, in the peak region where $\omega \sim \Lambda_{\rm QCD}$, the full function $\mathcal{F}(k)$ is needed.

As discussed in ref. [32], an important feature of the factorization in eq. (2.22) is that it provides a common description of the nonperturbative effects across these different kinematic regions, incorporating all available perturbative information in the limit $\omega \gg \Lambda_{\rm QCD}$ without having to explicitly carry out an expansion in $\Lambda_{\rm QCD}/\omega$, whose precise region of validity would be unclear. That is, all perturbative corrections to moments of $S(\omega, \mu)$ are encoded in the perturbative function $\widehat{C}_0(\omega)$, while the shape function $\mathcal{F}(k)$ is a genuinely nonperturbative function that is formally independent of the perturbative order and can be extracted from experimental data.

The first few moments of $\mathcal{F}(k)$ are given in terms of B-meson matrix elements of local HQET operators,

$$
\begin{aligned}
M_0 &= 1 \,, \\
M_1 &= m_B - \widehat{m}_b + \cdots \,, \\
M_2 &= (m_B - \widehat{m}_b)^2 - \frac{\widehat{\lambda}_1}{3} + \cdots \,, \\
M_3 &= (m_B - \widehat{m}_b)^3 - \widehat{\lambda}_1 (m_B - \widehat{m}_b) + \frac{\widehat{\rho}_1}{3} + \cdots \,,
\end{aligned}
\tag{2.25}
$$

The hadronic parameters $\widehat{\lambda}_1$ and $\widehat{\rho}_1$ are matrix elements of local HQET operators defined in a short-distance scheme as discussed in section 3.3. The ellipses denote terms that suppressed by relative powers of $\Lambda_{\rm QCD}/\widehat{m}_b$ arising from subleading shape functions that are typically absorbed into $\mathcal{F}(k)$ [15], but which are not relevant for our purposes here.

A general method for parametrizing $\mathcal{F}(k)$ via a systematic expansion around a given base model has been developed in ref. [32], which was used in ref. [15] to fit $\mathcal{F}(k)$ from data. Since our primary interest in this paper are the perturbative corrections, the precise form of $\mathcal{F}(k)$ is not relevant here. We only need a reasonably realistic model for it in order to illustrate our results numerically. For this purpose, we take the exponential base model used in refs. [15, 32], given by

$$
\mathcal{F}(k) = \frac{1}{\lambda} Y\left(\frac{k}{\lambda}\right) \qquad \text{with} \qquad Y(x) = \frac{(p+1)^{p+1} x^p}{p!} e^{-(p+1)x} \,.
\tag{2.26}
$$

Its normalization and first moment are

$$
M_0 = \int_0^\infty \mathrm{d}k \, \mathcal{F}(k) = 1 \,, \qquad M_1 = \int_0^\infty \mathrm{d}k \, k \, \mathcal{F}(k) = \lambda \,.
\tag{2.27}
$$

By picking $\lambda \approx m_B - \widehat{m}_b \approx 0.6\,\text{GeV}$ this base model already provides a good fit of the experimental measurements [15].

Note that evaluating the N^nLO soft function in a short-distance scheme involves taking n derivatives of $\mathcal{F}(k)$. Therefore, at $\text{N}^3\text{LL}'$, which needs the N^3LO soft function, and assuming integer p, we require $p \geq 4$ to ensure that the soft function vanishes for $\omega \to 0$, which in turn is required for the E_γ spectrum to vanish at the kinematic endpoint $E_\gamma \to m_B/2$.

3 Short-distance schemes

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 or no 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 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,

$$C - p = (C - \delta p) - (p - \delta p) \equiv \widehat{C} - \widehat{p} . \tag{3.1}$$

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 $\widehat{C}$ and $\widehat{p}$, which are now separately free of the renormalon.

3.1 Soft function

In reality, the structure is of course more complicated than the above simple toy example. In our case, 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 $\mathcal{O}(\Lambda_{\mathrm{QCD}})$ is due to the pole mass definition of the b quark, m_b^{pole}, which explicitly enters in the definition of the soft function $S(\omega, \mu)$ in eq. (2.21) and henceforth shows up in all the moments of $\mathcal{F}(k)$. (At subleading power, also the jet and hard functions involve the pole-mass renormalon, which we come back to in section 3.4.) Furthermore, the hadronic parameter λ_1, which first appears in the second moment of $\mathcal{F}(k)$, has a subleading $\mathcal{O}(\Lambda_{\mathrm{QCD}}^2)$ renormalon ambiguity [43, 44]. Similarly, we expect the hadronic parameter ρ_1, which first appears in the third moment, to have an $\mathcal{O}(\Lambda_{\mathrm{QCD}}^3)$ renormalon.

We write the parameters in a generic short-distance scheme as

$$\widehat{m}_b = m_b^{\mathrm{pole}} - \delta m_b , \qquad \widehat{\lambda}_1 = \lambda_1 - \delta\lambda_1 , \qquad \widehat{\rho}_1 = \rho_1 - \delta\rho_1 . \tag{3.2}$$

The residual 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 are renormalon

free. The original[1] HQET parameters λ_1 and ρ_1 are defined in dimensional regularization as [45, 46]

$$\lambda_1 = \langle B| \, \bar{b}_v (iD)^2 b_v \, |B\rangle \,, \qquad \rho_1 = \langle B| \, \bar{b}_v (iD_\mu)(v \cdot iD)(iD^\mu) b_v \, |B\rangle \,. \tag{3.3}$$

The construction of the short-distance partonic soft function $\widehat{C}_0(\omega, \mu)$ with the appropriate renormalon subtractions is derived in detail in ref. [32]. Up to N³LO, we have

$$\begin{aligned}
\widehat{C}_0(\omega, \mu) &= \left[1 - \frac{\delta\lambda_1}{6} \frac{\mathrm{d}^2}{\mathrm{d}\omega^2} - \frac{\delta\rho_1}{18} \frac{\mathrm{d}^3}{\mathrm{d}\omega^3} + \cdots \right] e^{\delta m_b \frac{\mathrm{d}}{\mathrm{d}\omega}} C_0(\omega, \mu) \\
&= \left[1 + \delta m_b \frac{\mathrm{d}}{\mathrm{d}\omega} + \frac{1}{2}\left(\delta m_b^2 - \frac{\delta\lambda_1}{3} \right) \frac{\mathrm{d}^2}{\mathrm{d}\omega^2} \right. \\
&\qquad \left. + \frac{1}{6}\left(\delta m_b^3 - \delta m_b \delta\lambda_1 - \frac{\delta\rho_1}{3} \right) \frac{\mathrm{d}^3}{\mathrm{d}\omega^3} + \cdots \right] C_0(\omega, \mu) \,,
\end{aligned} \tag{3.4}$$

where the original pole-scheme $C_0(\omega, \mu)$ is defined and given in appendix A.4. 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, $\delta\lambda_1$, $\delta\rho_1$ and 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. (2.25). [Note that $\mathcal{F}(k) \equiv \widehat{\mathcal{F}}(k)$ in eq. (2.22) is already the short-distance parameter.] To evaluate the convolution integral $\widehat{C}_0 \otimes \mathcal{F}$, we use integration by parts to move all the derivatives in eq. (3.4) to act on $\mathcal{F}$ [32]. The renormalon subtractions significantly improve the perturbative convergence of the soft function compared to the pole scheme, which we demonstrate numerically in section 5.3.

In general, the residual terms are scale dependent, leading to a similar scale dependence of the short-distance parameter, which we suppress for simplicity in our generic notation. The scale dependence can be explicit, as e.g. for the $\overline{\mathrm{MS}}$ mass, in which case it is usually governed by an associated RGE. It can also be only internal, as e.g. for the $1S$ mass or the MSR mass, 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 our case this is the soft scale μ_S at which we evaluate the fixed-order boundary condition $\widehat{C}_0(\omega, \mu_S)$ in the factorization theorem in eq. (2.4).

In this context, the power counting of the HQET Lagrangian provides a powerful constraint on the generic size of the residual mass term δm_b,

$$\mathcal{L}_{\mathrm{HQET}} = \bar{b}_v \left(iv \cdot D - \delta m_b \right) b_v \,. \tag{3.5}$$

A suitable short-distance scheme for the bottom-quark mass should respect the power counting of residual soft momentum $\delta m_b \sim \alpha_s \, v \cdot D \sim \alpha_s \, k \sim \alpha_s \Lambda_{\mathrm{QCD}}$ [47]. Note that

---

[1]These are often referred to as defined in the "pole scheme" borrowing the language from the pole mass.

the perturbative series for δm_b starts at NLO and therefore scales like α_s. As stated in the previous section, in $B \to X_s\gamma$ the residual momentum of the bottom quark in the B meson scales like $k \sim m_B - 2E_\gamma$, which in the peak region scales like $\Lambda_{\rm QCD}$. Thus, in the peak region one expects that a low-scale mass scheme, such as the $1S$ scheme [29–31] or the MSR scheme [33] with $R \sim \Lambda_{\rm QCD}$, are applicable. In the context of $B \to X_s\gamma$, the $1S$ mass scheme has been extensively discussed in ref. [32], and we remind the reader of the main features of the MSR scheme in the following section.

3.2 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,

$$\delta m_b(R) \equiv m_b^{\rm pole} - m_b^{\rm MSR}(R) = R \sum_{n=1}^{\infty} a_n^{\rm MSR} \left[\frac{\alpha_s(R)}{4\pi} \right]^n , \tag{3.6}$$

where $a_n^{\rm MSR}$ coefficients are determined from matching the MSR mass scheme onto the $\overline{\rm MS}$ mass scheme at $\overline{m}_b(\overline{m}_b)$. The infrared scale R controls the size of self-energy contributions which are absorbed into the mass definition [33, 48]. In our analysis we use the so-called "natural" MSR mass definition, where the series coefficients $a_n^{\rm MSR} = a_n^{\overline{\rm MS}}(n_l, n_h = 0)$, see ref. [49] for more details.

The MSR mass is a natural extension of the $\overline{\rm MS}$ mass for $R \leq \overline{m}_b(\overline{m}_b)$, which interpolates between all short-distance schemes with residual power counting $\delta m_b(R) \sim R\,\alpha_s$. In the limit where $m_b^{\rm MSR}(m_b^{\rm MSR}) \to \overline{m}_b(\overline{m}_b)$ one approaches the $\overline{\rm MS}$ mass, whereas in the opposite limit, $R \to 0$, the MSR mass formally approaches the pole mass. In practice, when taking this limit one encounters the Landau pole of the coupling constant. This issue is deeply 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 [33], whose solution resums logarithms $\ln(R_1/R_0)$ in the perturbative correction between $m_b^{\rm MSR}(R_1)$ and $m_b^{\rm MSR}(R_0)$. The R evolution can be used to obtain the MSR mass value at a low R from the $\overline{\rm MS}$ mass and vice versa.

In contrast to the $1S$ scheme, the infrared scale R of the MSR scheme is an external parameter. We exploit this feature and pick $R = 1\,{\rm GeV}$ to be of the same order as the soft scale in the peak region. In section 5.3 we demonstrate 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 regions of the spectrum by introducing a profile function $R(E_\gamma)$ [50], 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. [42] implements this R-evolution setup for an analogous soft function for the thrust distribution in jet production. In our case the use of R evolution does not lead to a significant improvement in convergence, so for simplicity in our numerical analysis we use a fixed value of R.

3.3 Short-distance schemes for λ_1 and ρ_1

To express the B-meson matrix elements λ_1 and ρ_1 in a short-distance scheme, we use analogous schemes to the "invisible" scheme for λ_1 [32], where $\delta\lambda_1 \propto \alpha_s^2$. The reason for this α_s^2 scaling in the invisible scheme compared to the kinetic scheme [51, 52], where $\delta\lambda_1 \propto \alpha_s$, is that in the invisible scheme one employs Lorentz-invariant UV regulators for regularizing the kinetic energy operator [43, 44], while in the kinetic scheme the regulator is not Lorentz-invariant. Indeed it has been shown in ref. [32] that using the kinetic scheme for λ_1 leads to an over-subtraction of the $u = 1$ renormalon in the soft function. It is worth to note here that there is no analogous study for the $u = 3/2$ renormalon present in ρ_1. Following these features of the invisible scheme for λ_1, we write

$$\delta\lambda_1(R_\lambda, \mu) = R_\lambda^2 \frac{\alpha_s^2(\mu)}{\pi^2} \left[\delta\lambda_1^{(2)} + \frac{\alpha_s(\mu)}{\pi} \left(\delta\lambda_1^{(3)} + \delta\lambda_1^{(2)} \beta_0 \ln \frac{\mu}{R_\lambda} \right) \right] + \mathcal{O}(\alpha_s^4),$$

$$\delta\rho_1(R_\rho, \mu) = R_\rho^3 \frac{\alpha_s^3(\mu)}{\pi^3} \delta\rho_1^{(3)} + \mathcal{O}(\alpha_s^4), \tag{3.7}$$

where we set $R_\lambda = R_\rho = 1\,\mathrm{GeV}$ by default. Note that λ_1 and ρ_1 appear in the second and third moments of the shape function and contain an $\mathcal{O}(\Lambda_{\mathrm{QCD}}^2)$ and $\mathcal{O}(\Lambda_{\mathrm{QCD}}^3)$ renormalon ambiguities, respectively. Therefore by dimensional analysis they must scale like $\delta\lambda_1 \propto R_\lambda^2$ and $\delta\rho_1 \propto R_\rho^3$. Furthermore, we take $\delta\rho_1 \propto \alpha_s^3$ to impose a plausible "invisibility" of our scheme choice and to avoid diluting the lower-order corrections arising from δm_b and $\delta\lambda_1$. For $\delta\lambda_1^{(2)}$ we use the value given in ref. [32] for the invisible scheme, $\delta\lambda_1^{(2)} = \pi^2/3 - 1$. Values of $\delta\lambda_1^{(3)}$ and $\delta\rho_1^{(3)}$ are not available so far.

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, here we take a pragmatic approach and simply define our "invisible" scheme for $\widehat{\lambda}_1$ and $\widehat{\rho}_1$ at $\mathcal{O}(\alpha_s^3)$ 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, we find a satisfactory convergence for the soft function, see figure 11, by taking

$$\delta\lambda_1^{(3)} = 16, \qquad \delta\rho_1^{(3)} = -3. \tag{3.8}$$

In principle, we could also consider the perturbative convergence of the final 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, we use the convergence of the soft function to define the short-distance scheme. This is also the most natural, since the soft function is the object containing the renormalons to be subtracted. This is somewhat similar to the "shape-function" scheme in ref. [53], 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 that approach is that it yields $\delta\lambda, \delta\rho_1 \sim \alpha_s$, which leads to massive oversubtraction similar to the kinetic scheme.

3.4 Subleading δm_b corrections

As discussed in section 3.1, 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 in 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. [15] the resulting corrections from changing to the $\widehat{m}_b$ scheme could effectively be absorbed into the nonsingular corrections. As we will see, at N^{3}LL$'$ this is no longer viable, so instead we will explicitly switch both hard and jet functions to a short-distance mass scheme.

To derive the scheme change, we consider the partonic function $W(k)$ appearing in eq. (2.3), 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

$$W(k) = \frac{1}{m_b} w\left[\frac{k}{m_b}, \alpha_s(\mu), \ln\frac{\mu}{m_b}\right] = \frac{1}{m_b} w\left[\frac{k}{m_b}, \alpha_s(m_b)\right], \tag{3.9}$$

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 is only the internal μ dependence from $\alpha_s(\mu)$ which cancels order by order. Therefore, we can pick $\mu = m_b$, which eliminates all logarithms and allows us to track the associated m_b dependence via the dependence on $\alpha_s(m_b)$. After switching the scheme, we can easily reintroduce the μ dependence by reexpanding $\alpha_s(m_b)$ in terms of $\alpha_s(\mu)$.

The partonic rate in the pole scheme is given by eq. (3.9) evaluated at $k = m_b^{\rm pole} - 2E_\gamma$ and $m_b = m_b^{\rm pole}$. To switch to a short-distance scheme, we thus have to replace $k \to k + \delta m_b$ and $m_b \to m_b^{\rm pole} = \widehat{m}_b + \delta m_b$ in eq. (3.9) and expand in δm_b. This gives

$$\begin{aligned}
\widehat{m}_b W(\widehat{m}_b x) &= \frac{1}{1 + \delta m_b/\widehat{m}_b} w\left(\frac{x + \delta m_b/\widehat{m}_b}{1 + \delta m_b/\widehat{m}_b}, \alpha_s(\widehat{m}_b + \delta m_b)\right) \\
&= \left\{1 + \frac{\delta m_b}{\widehat{m}_b}\frac{\mathrm{d}}{\mathrm{d}x}(1-x) + \frac{1}{2}\frac{\delta m_b^2}{\widehat{m}_b^2}\frac{\mathrm{d}^2}{\mathrm{d}x^2}(1-x)^2 \right. \\
&\left. \quad + \frac{1}{3!}\frac{\delta m_b^3}{\widehat{m}_b^3}\frac{\mathrm{d}^3}{\mathrm{d}x^3}(1-x)^3 + \frac{\delta m_b}{\widehat{m}_b}\beta[\alpha_s(\widehat{m}_b)]\frac{\mathrm{d}}{\mathrm{d}\alpha_s} + \mathcal{O}(\alpha_s^4)\right\} w[x, \alpha_s(\widehat{m}_b)], \tag{3.10}
\end{aligned}$$

where for convenience we multiplied by $\widehat{m}_b$ and switched variables, $k = \widehat{m}_b x$, such that $\widehat{m}_b W(\widehat{m}_b x)$ is a dimensionless function of x. In the second step, we 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 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 $\Delta w_{77}^{{\rm s}(n)}$ and $\Delta w_{77}^{{\rm ns}(n)}$ appearing in eqs. (2.8)

and (2.10). Writing the α_s expansion of $\delta m_b / \widehat{m}_b$ as

$$\frac{\delta m_b}{\widehat{m}_b} = \frac{\alpha_s(\mu)}{4\pi} \delta_m^{(1)}(\mu) + \left[\frac{\alpha_s(\mu)}{4\pi} \right]^2 \delta_m^{(2)}(\mu) + \left[\frac{\alpha_s(\mu)}{4\pi} \right]^3 \delta_m^{(3)}(\mu) + \mathcal{O}(\alpha_s^4), \tag{3.11}$$

we find for the singular

$$\Delta w_{77}^{\mathrm{s}(1)}(\mu, x) = \frac{1}{4} \delta_m^{(1)}(\mu) \frac{\mathrm{d}}{\mathrm{d}x} w_{77}^{\mathrm{s}(0)}(x),$$

$$\Delta w_{77}^{\mathrm{s}(2)}(\mu, x) = \frac{1}{16} \left\{ \delta_m^{(2)}(\mu) \frac{\mathrm{d}}{\mathrm{d}x} + \frac{1}{2} [\delta_m^{(1)}(\mu)]^2 \frac{\mathrm{d}^2}{\mathrm{d}x^2} \right\} w_{77}^{\mathrm{s}(0)}(x)$$

$$+ \frac{1}{4} \delta_m^{(1)}(\mu) \frac{\mathrm{d}}{\mathrm{d}x} [(1-x) w_{77}^{\mathrm{s}(1)}(x)],$$

$$\Delta w_{77}^{\mathrm{s}(3)}(\mu, x) = \frac{1}{64} \left\{ \delta_m^{(3)}(\mu) \frac{\mathrm{d}}{\mathrm{d}x} + \delta_m^{(1)}(\mu) \delta_m^{(2)}(\mu) \frac{\mathrm{d}^2}{\mathrm{d}x^2} + [\delta_m^{(1)}(\mu)]^3 \frac{1}{3!} \frac{\mathrm{d}^3}{\mathrm{d}x^3} \right\} w_{77}^{\mathrm{s}(0)}(x)$$

$$+ \frac{1}{16} \left\{ -2\beta_0 \delta_m^{(1)}(\mu) + \left[\delta_m^{(2)}(\mu) + 2\beta_0 \delta_m^{(1)}(\mu) \ln \frac{\mu}{\widehat{m}_b} \right] \frac{\mathrm{d}}{\mathrm{d}x} (1-x) \right.$$

$$+ \left. \frac{1}{2} [\delta_m^{(1)}(\mu)]^2 \frac{\mathrm{d}^2}{\mathrm{d}x^2} (1-x)^2 \right\} w_{77}^{\mathrm{s}(1)}(x)$$

$$+ \frac{1}{4} \delta_m^{(1)}(\mu) \frac{\mathrm{d}}{\mathrm{d}x} (1-x) w_{77}^{\mathrm{s}(2)}(x), \tag{3.12}$$

and the nonsingular

$$\frac{\Delta w_{77}^{\mathrm{ns}(2)}(\mu_{\mathrm{ns}}, x)}{(1-x)^3} = \frac{1}{4} \delta_m^{(1)}(\mu_{\mathrm{ns}}) \frac{\mathrm{d}}{\mathrm{d}x} \left[\frac{w_{77}^{\mathrm{ns}(1)}(x)}{(1-x)^2} \right],$$

$$\frac{\Delta w_{77}^{\mathrm{ns}(3)}(\mu_{\mathrm{ns}}, x)}{(1-x)^3} = \frac{1}{16} \left\{ -2\beta_0 \delta_m^{(1)}(\mu_{\mathrm{ns}}) + \left[\delta_m^{(2)}(\mu_{\mathrm{ns}}) + 2\beta_0 \delta_m^{(1)}(\mu_{\mathrm{ns}}) \ln \frac{\mu_{\mathrm{ns}}}{\widehat{m}_b} \right] \frac{\mathrm{d}}{\mathrm{d}x} (1-x) \right.$$

$$+ \left. \frac{1}{2} [\delta_m^{(1)}(\mu_{\mathrm{ns}})]^2 \frac{\mathrm{d}^2}{\mathrm{d}x^2} (1-x)^2 \right\} \frac{w_{77}^{\mathrm{ns}(1)}(x)}{(1-x)^3}$$

$$+ \frac{1}{4} \delta_m^{(1)}(\mu_{\mathrm{ns}}) \frac{\mathrm{d}}{\mathrm{d}x} \frac{w_{77}^{\mathrm{ns}(2)}(x)}{(1-x)^2}. \tag{3.13}$$

As we have seen above, 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. (3.10),

$$\widehat{m}_b W^{\mathrm{s}}(\widehat{m}_b x) = \left[1 + \frac{\delta m_b}{\widehat{m}_b} \frac{\mathrm{d}}{\mathrm{d}x} + \frac{1}{2} \frac{\delta m_b^2}{\widehat{m}_b^2} \frac{\mathrm{d}^2}{\mathrm{d}x^2} + \frac{1}{3!} \frac{\delta m_b^3}{\widehat{m}_b^3} \frac{\mathrm{d}^3}{\mathrm{d}x^3} \right] w^{\mathrm{s}}[x, \alpha_s(\widehat{m}_b)]. \tag{3.14}$$

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 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 $\delta m_b \sim \Lambda_{\mathrm{QCD}}$, we see again that all terms in eq. (3.14) 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, we see that they are explicitly power-suppressed by x and hence nonsingular. For this reason, they could be included as part of the nonsingular correction terms $\Delta w^{\mathrm{ns}}_{77}$, as was done in ref. [15]. 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. (3.10),

$$\frac{\delta m_b}{\widehat{m}_b}\beta(\alpha_s)\frac{\mathrm{d}}{\mathrm{d}\alpha_s}[w(x, \alpha_s(\widehat{m}_b)] = -\frac{\alpha_s^3 C_F}{8\pi^3}\,\delta_m^{(1)}\,\beta_0\,w_{77}^{\mathrm{s}(1)}(x) + \mathcal{O}(\alpha_s^4)\,. \qquad (3.15)$$

Since it starts at $\mathcal{O}(\alpha_s^3)$ it first appears at N^{3}LL$'$. Formally, this term is also subleading power, because $\delta m_b/\widehat{m}_b \sim \lambda$. 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, we consistently switch the hard and jet functions to the short-distance mass $\widehat{m}_b$. The details of this procedure are discussed in the following two subsections.

3.5 Hard function

The hard function in a short-distance scheme is obtained by writing the b-quark mass in the pole-scheme hard function in eq. (A.5) in terms of a short-distance mass and reexpanding the result strictly in powers of $\alpha_s(\mu)$. At N^{3}LO we obtain

$$\widehat{h}_s(\widehat{m}_b, \mu) = h_s(\widehat{m}_b, \mu) - \sum_{n=2}^{\infty}\sum_{m=0}^{2n-3}\Delta H_m^{(n)}(\mu)\left[\frac{\alpha_s(\mu)}{4\pi}\right]^n \ln^m \frac{\mu}{\widehat{m}_b}\,, \qquad (3.16)$$

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

$$\Delta H_0^{(2)}(\mu) = \delta_m^{(1)}(\mu) H_1^{(1)} \,,$$
$$\Delta H_1^{(2)}(\mu) = 2\delta_m^{(1)}(\mu) H_2^{(1)} \,,$$
$$\Delta H_0^{(3)}(\mu) = \delta_m^{(2)}(\mu) H_1^{(1)} + \delta_m^{(1)}(\mu) H_1^{(2)} - \frac{1}{2}(\delta_m^{(1)}(\mu))^2 \Big(H_1^{(1)} + 2H_2^{(1)} \Big) \,,$$
$$\Delta H_1^{(3)}(\mu) = 2\delta_m^{(2)}(\mu) H_2^{(1)} + 2\delta_m^{(1)}(\mu) H_2^{(2)} - (\delta_m^{(1)}(\mu))^2 H_2^{(1)} \,,$$
$$\Delta H_2^{(3)}(\mu) = 3\delta_m^{(1)}(\mu) H_3^{(2)} \,,$$
$$\Delta H_3^{(3)}(\mu) = 4\delta_m^{(1)}(\mu) H_4^{(2)} \,. \tag{3.17}$$

Here $H_m^{(n)}$ are the coefficients of the pole-scheme hard function defined in eq. (A.5), and $\delta_m^{(n)}(\mu)$ are defined in eq. (3.11).

Similarly, the RGE for the hard function in a short-distance mass scheme is derived by rewriting the pole mass in eq. (A.6) in terms of a short-distance mass,

$$\frac{\mathrm{d}\widehat{h}_s(\widehat{m}_b,\mu)}{\mathrm{d}\ln\mu} = \left\{ \Gamma^H[\alpha_s(\mu)] \ln\frac{\mu}{\widehat{m}_b} + \gamma^H[\alpha_s(\mu)] + \Delta\gamma^H(\mu) \right\} \widehat{h}_s(\widehat{m}_b,\mu) \,, \tag{3.18}$$

where

$$\Delta\gamma^H(\mu) = \Gamma^H[\alpha_s(\mu)] \ln\frac{\widehat{m}_b}{m_b^{\mathrm{pole}}} = -\Gamma^H[\alpha_s(\mu)] \ln\left(1 + \frac{\delta m_b}{\widehat{m}_b} \right)$$
$$= -\left[\frac{\alpha_s(\mu)}{4\pi} \right]^2 \Gamma_0^H \delta_m^{(1)}(\mu)$$
$$- \left[\frac{\alpha_s(\mu)}{4\pi} \right]^3 \left\{ \Gamma_1^H \delta_m^{(1)}(\mu) + \Gamma_0^H \left[\delta_m^{(2)}(\mu) - \frac{1}{2}(\delta_m^{(1)}(\mu))^2 \right] \right\} + \mathcal{O}(\alpha_s^4) \,, \tag{3.19}$$

and the hard cusp anomalous dimension coefficients $\Gamma_n^H = -2\Gamma_n$ are defined in eq. (A.1). Note that the reexpansion in eqs. (3.16) and (3.18) must be performed in terms of the same $\alpha_s(\mu)$, such that $\widehat{h}_s(\widehat{m}_b,\mu)$ in eq. (3.16) indeed satisfies the RGE in eq. (3.18).

We expect that the renormalon associated with the pole mass cancels 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 know about 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 (3.18) can be written as

$$\widehat{U}_H(\widehat{m}_b,\mu_H,\mu) = U_H(\widehat{m}_b,\mu_H,\mu) \times \Delta U_H(\mu_H,\mu) \,, \tag{3.20}$$

where U_H is the usual hard evolution factor given in eq. (A.10), and the correction factor ΔU_H is given by

$$\Delta U_H(\mu_H,\mu) = \exp\left[\int_{\mu_H}^{\mu} \mathrm{d}\ln\mu\, \Delta\gamma^H(\mu) \right] \,. \tag{3.21}$$

– 21 –

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 N³LL, we can still perform the integral by employing an analytic approximation analogous to the one used for the K and η integrals in eq. (A.3). We find up to N³LL

$$
\begin{aligned}
\ln \Delta U_H(\mu_H, \mu) = {} & \frac{\alpha_s(\mu_H)}{4\pi} \frac{\Gamma_0^H}{2\beta_0} \delta_{\mathrm{MSR}}^{(1)}(r-1) \\
& + \frac{\alpha_s(\mu_H)}{4\pi} \frac{\Gamma_0^H}{4\beta_0} \bigg\{ \frac{\alpha_s(\mu_H)}{4\pi} \bigg[\Big(\delta_{\mathrm{MSR}}^{(2)} - \frac{1}{2} (\delta_{\mathrm{MSR}}^{(1)})^2 + 2\delta_{\mathrm{MSR}}^{(1)} \beta_0 \ln \frac{\mu_H}{R} \Big) \\
& \qquad + \delta_{\mathrm{MSR}}^{(1)} \bigg(\frac{\Gamma_1^H}{\Gamma_0^H} - \frac{\beta_1}{\beta_0} \bigg) \bigg] (r^2 - 1) - \delta_{\mathrm{MSR}}^{(1)}(r-1)^2 \bigg\} , \quad (3.22)
\end{aligned}
$$

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

$$
\delta_{\mathrm{MSR}}^{(n)} \equiv \frac{R}{m_b^{\mathrm{MSR}}(R)} \, a_n^{\mathrm{MSR}} . \tag{3.23}
$$

At NNLL only the first line on the right-hand side of eq. (3.22) is kept.

3.6 Jet function

Similar to the hard function in the previous section, we define the jet function in a short-distance scheme by expressing m_b^{pole} in terms of a short-distance mass $\widehat{m}_b$. To this end, we start with the jet function in the pole scheme given in eq. (A.11) and set $s = m_b^{\mathrm{pole}} \omega$. Then we write m_b^{pole} in terms of $\widehat{m}_b$ and reexpand the result strictly in powers of $\alpha_s(\mu)$, such that $m_b^{\mathrm{pole}} J(m_b^{\mathrm{pole}} \omega, \mu_J) = \widehat{m}_b \widehat{J}(\widehat{m}_b \omega, \mu_J)$ order by order. This yields up to N³LO

$$
\widehat{J}(\widehat{m}_b \omega, \mu) = J(\widehat{m}_b \omega, \mu) + \sum_{n=2}^{3} \sum_{m=-1}^{2n-4} \Delta J_m^{(n)} \bigg[\frac{\alpha_s(\mu)}{4\pi} \bigg]^n \frac{1}{\mu^2} \mathcal{L}_m \bigg(\frac{\widehat{m}_b \omega}{\mu^2} \bigg) . \tag{3.24}
$$

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

$$
\begin{aligned}
\Delta J_{-1}^{(2)}(\mu) &= \delta_m^{(1)}(\mu) J_0^{(1)} , \\
\Delta J_0^{(2)}(\mu) &= \delta_m^{(1)}(\mu) J_1^{(1)} , \\
\Delta J_{-1}^{(3)}(\mu) &= \delta_m^{(2)}(\mu) J_0^{(1)} + \delta_m^{(1)}(\mu) J_0^{(2)} - \frac{1}{2} (\delta_m^{(1)}(\mu))^2 \big(J_0^{(1)} - J_1^{(1)} \big) , \\
\Delta J_0^{(3)}(\mu) &= \delta_m^{(2)}(\mu) J_1^{(1)} + \delta_m^{(1)}(\mu) J_1^{(2)} - \frac{1}{2} (\delta_m^{(1)}(\mu))^2 J_1^{(1)} , \\
\Delta J_1^{(3)}(\mu) &= 2\delta_m^{(1)}(\mu) J_2^{(2)} , \\
\Delta J_2^{(3)}(\mu) &= 3\delta_m^{(1)}(\mu) J_3^{(2)} , \tag{3.25}
\end{aligned}
$$

where the $\delta_m^{(n)}$ are defined in eq. (3.11) and the $J_m^{(n)}$ coefficients are those of the original pole-scheme jet function as defined in eq. (A.11).

4 Perturbative uncertainties

For our predictions we can distinguish perturbative and parametric uncertainties. Parametric uncertainties arise from the uncertainty in input parameters, such as C_7^{incl}, $|V_{\mathrm{tb}} V_{ts}^*|^2$, m_b, $\mathcal{F}(k)$. These are not considered in the following. We normalize our numerical results by dividing out the overall prefactor $\Gamma_0 |C_7^{\mathrm{incl}}|^2$, so the associated uncertainties drop out. We also ignore the parametric uncertainties due to the shape function $\mathcal{F}(k)$ and m_b, which do affect the shape of the spectrum, because we do not compare with experimental $B \to X_s \gamma$ measurements, in which case they would be determined by the fit to the data.

Our primary focus is on the perturbative results and their perturbative uncertainties arising from missing higher-order corrections. For these there are various different sources, which fall into two categories:

- **Profile scale variations:** we identify three sources of perturbative uncertainties that are estimated by a suitable set of variations of the profile scales discussed in section 2.4. The resummation uncertainty Δ_{resum} is obtained by taking the maximum envelope of all 27 simultaneous 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 e_{ns} parameter, 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 considered independent and are thus added in quadrature,

$$\Delta_{\mathrm{profile}} = \Delta_{\mathrm{resum}} \oplus \Delta_{\mathrm{ns}} \oplus \Delta_{\mathrm{match}} \,. \tag{4.1}$$

  For notational convenience, we use the symbol $\oplus$ to denote addition in quadrature, $x \oplus y \equiv \sqrt{x^2 + y^2}$.

- **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 2.2 and 2.3, where we abbreviate $\Delta_{c^{\mathrm{ns}}} = (\sum_{k=0}^5 \Delta_{c_k^{\mathrm{ns}}}^2)^{1/2}$. 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 profile scale variations estimate the uncertainty due to the missing next order $\mathrm{N^4LL'}+\mathrm{N^4LO}$, 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,

$$\Delta_{\mathrm{total}} = \Delta_{\mathrm{profile}} \oplus \Delta_{h_3} \oplus \Delta_{c^{\mathrm{ns}}} \,. \tag{4.2}$$

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

Table 1. Numerical values of required input parameters.

The nuisance parameter uncertainties only contribute at the highest order $\text{N}^3\text{LL}'+\text{N}^3\text{LO}(c_k)$, while at lower orders we simply have $\Delta_{\text{total}} = \Delta_{\text{profile}}$.

Note that in ref. [15] the perturbative uncertainty was estimated from profile scale variations by taking the maximum envelope of $3^5 = 243$ profile scale variations corresponding to simultaneous variations of the above five profile scale parameters. Here we have refined the estimation procedure 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 $\text{N}^3\text{LL}'$. In part, this becomes possible because we are now able to reexpand the fixed-order hard function against the product of the fixed-order jet and soft functions. We have also checked that the total Δ_{profile} estimated as described above is comparable in size to what we obtain in our setup by taking the maximum envelope of all $3^5 = 243$ variations excluding a small set of obvious outliers.

5 Results

In this section, we present our numerical results for the photon energy spectrum. All numerical results are implemented and obtained with the `SCETlib` [54] library.

Our default numerical setup is as follows. The values used for input parameters are summarized in table 1. The spectrum is always divided by the overall normalization factor $\Gamma_0|C_7^{\text{incl}}|^2$, so its numerical value is not needed. For the shape-function model in eq. (2.26) we use $\lambda = 0.6\,\text{GeV}$ and $p = 4$ as the default settings. We illustrate the impact of changing p and λ later in this section. We neglect finite-charm-mass corrections and work in QCD with $n_f = 4$ massless quarks. We always use the 4-loop running of α_s, which is sufficient for resummation at N^3LL. We use the MSR scheme for the b-quark mass and adopt short-distance schemes for λ_1 and ρ_1 as discussed in section 3.3. The impact of the short-distance mass scheme is discussed in section 5.3. Throughout this section, the colored bands always show the perturbative uncertainties obtained from profile scale variations Δ_{profile}.

5.1 Main results

To begin, in figure 3, we show the contribution of the resummed singular corrections to the full result at NNLL' (left panel) and $\text{N}^3\text{LL}'$ (right panel). The resummed contribution is indeed dominant across the peak of the spectrum, while it decreases rapidly in the tail and eventually changes sign at $E_\gamma \lesssim 1.8\,\text{GeV}$, where the resummation is getting turned
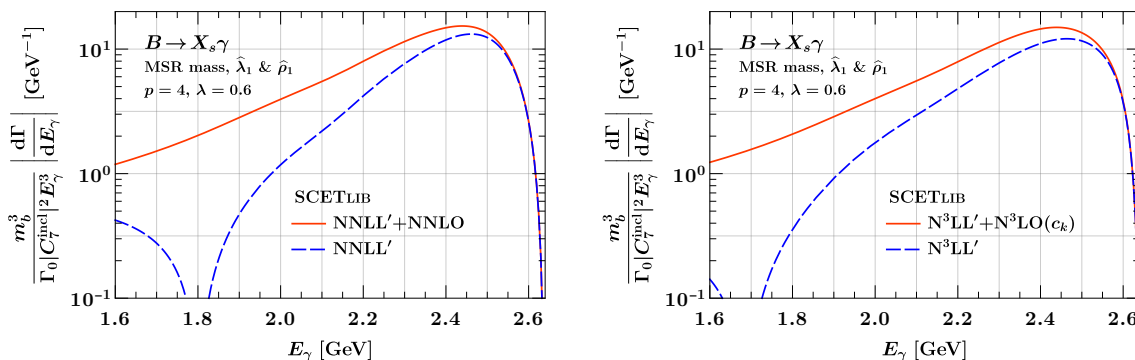
Figure 3. Comparison of the absolute value of resummed contribution to the full $B \to X_s\gamma$ photon energy spectrum in the MSR mass scheme. The left panel shows the 2-loop predictions, the right panel shows the 3-loop predictions. The overall factor $\Gamma_0|C_7^{\text{incl}}|^2(E_\gamma/\widehat{m}_b)^3$ is divided out (see eq. (2.1)).

off. Here, only the full matched result is meaningful, which remains positive and slowly approaches zero in the far tail.

Our main predictions for the $B \to X_s\gamma$ photon energy spectrum at different perturbative orders are presented in figure 4. In addition to the colored bands showing Δ_{profile}, the gray dashed line shows $\Delta_{\text{profile}} \oplus \Delta_{h_3}$, and the black solid line shows Δ_{total} defined in eq. (4.2). The first column shows the results for the spectrum, and the second column shows the relative difference to the central value at the highest order, i.e. to the red solid line on the left. In the first row, we use the same value for the shape-function parameter $p = 4$ for all orders, whereas in the second row, we use different values for p. In figure 5 we show the breakdown of the relative perturbative uncertainties into the individual contributions at NNLL'+NNLO and N³LL'+N³LO(c_k).

We remind the reader that the choice $p = 4$ was needed to ensure that the spectrum at the highest order N³LL' 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 $\mathcal{F}$. By using the same fixed model for $\mathcal{F}$ at each order, we can directly assess the perturbative convergence in the spectrum. However, by using a common p, the spectrum at lower orders vanishes correspondingly faster, i.e., quadratically at NNLL' and cubically at NLL', which also affects to some extent the shape of the spectrum into the peak. Therefore, in the lower panels of figure 4 we also show an alternative order comparison, where we do change the model 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 4 and 5 manifest a good perturbative convergence, especially from NNLL'+NNLO to N³LL'+N³LO(c_k). The relative uncertainties are under good control over the entire E_γ range, except at the very endpoint where the spectrum vanishes so the relative uncertainties necessarily blow up. Apart from the very endpoint, the total uncertainty at N³LL'+N³LO(c_k) is at most 15% and in most of the E_γ range well below that.
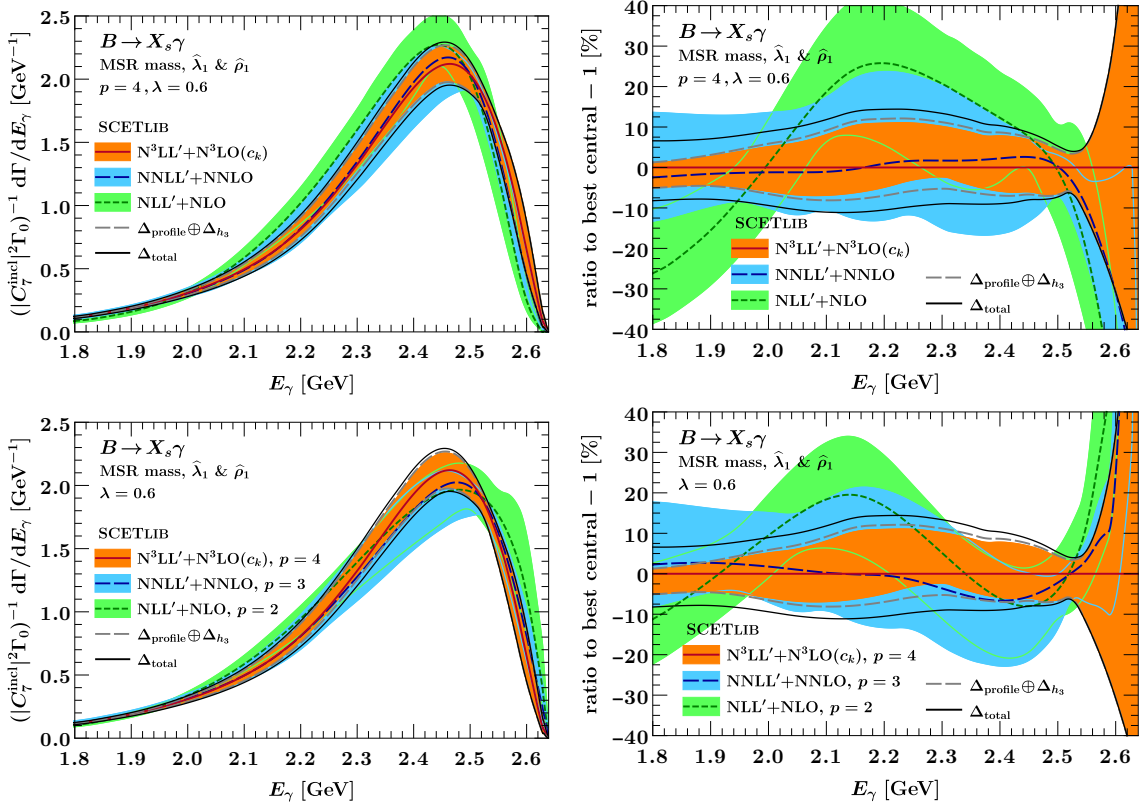
Figure 4. The $B \to X_s \gamma$ spectrum at different perturbative orders. The E_γ spectrum itself is shown on the left, while the relative differences to the highest-order central value are shown on the right. In the top row, we use the same value $p = 4$ at each order, while in the bottom row we use successive values at each order. The colored bands show the perturbative uncertainty estimated by just profile scale variations, Δ_{profile}. The dashed gray line includes in addition the uncertainty due to h_3. The solid black line further includes in addition the nonsingular nuisance parameters, corresponding to the total perturbative uncertainty at $N^3LL'+N^3LO(c_k)$.
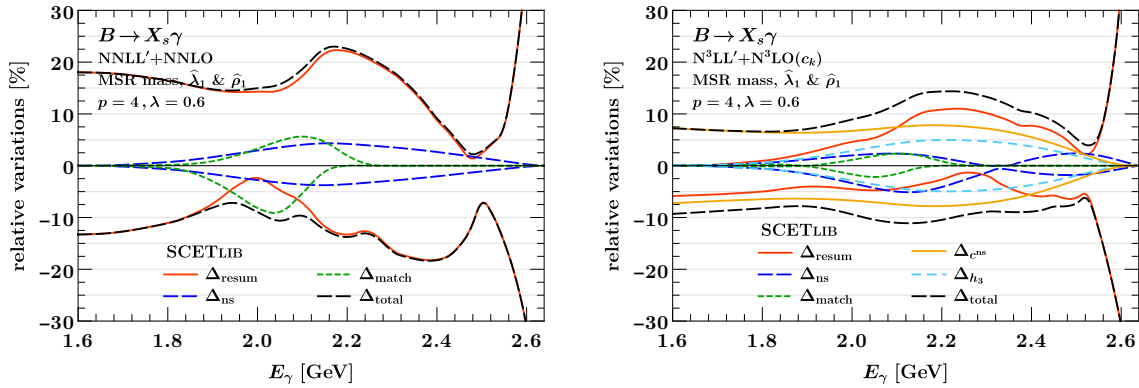
Figure 5. Breakdown of the perturbative uncertainty for the $B \to X_s \gamma$ spectrum into its components at $NNLL'+NNLO$ (left panel) and $N^3LL'+N^3LO(c_k)$ (right panel).
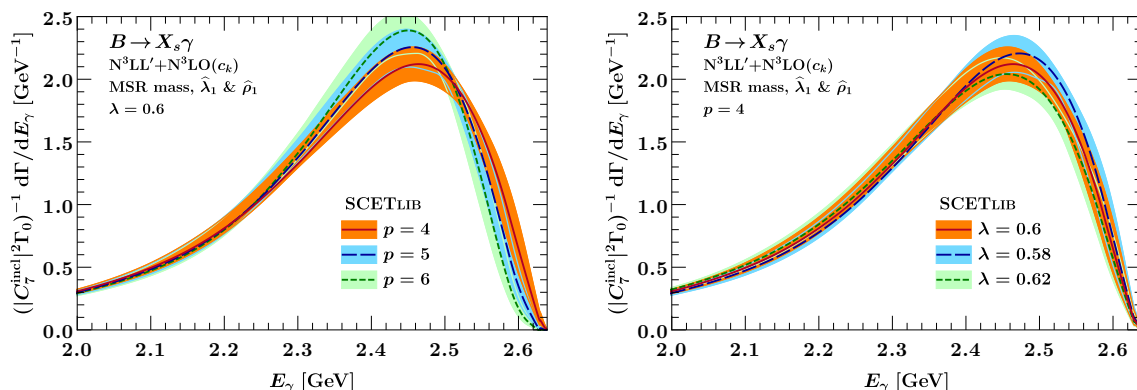
**Figure 6**. The $B \to X_s\gamma$ spectrum using different values for the shape-function parameters p (left panel) and λ (right panel) at $\mathrm{N}^3\mathrm{LL}'+\mathrm{N}^3\mathrm{LO}(c_k)$. The remaining theory parameters are set to their default values.

The uncertainties at $\mathrm{N}^3\mathrm{LL}'+\mathrm{N}^3\mathrm{LO}(c_k)$ are substantially reduced compared to $\mathrm{NNLL}'+\mathrm{NNLO}$, even accounting for the fact that not all 3-loop perturbative ingredients are known. As expected, the higher-order uncertainties estimated from profile scale variations (Δ_{resum}, Δ_{ns}, Δ_{match}) are much reduced. Also recall that in the fixed-order tail Δ_{resum} turns into the usual fixed-order scale variation. 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 below $E_\gamma \lesssim E_1 = 2.1\,\mathrm{GeV}$ it starts to take over and becomes the dominant uncertainty. This demonstrates that in our approach we are able to substantially benefit from the increased precision of the $\mathrm{N}^3\mathrm{LL}'$ resummation even in the absence of the full $\mathrm{N}^3\mathrm{LO}$ result. Furthermore, in the fixed-order tail the total uncertainty dominated by $\Delta_{c^{\mathrm{ns}}}$ is still reduced compared to the scale-variation based estimate at $\mathrm{NNLL}'+\mathrm{NNLO}$. As discussed at the end of section 2.3, this is anticipated and justified because of the additional nontrivial perturbative information included at $\mathrm{N}^3\mathrm{LO}(c_k)$.

The effect of varying the shape function model parameters p and λ is illustrated in figure 6. As expected, the position and height of the peak depend on the model parameters. We can clearly see how the value of p controls how fast the spectrum vanishes toward 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. We stress that these results are just meant as an additional illustration. In particular, the differences seen in figure 6 are not to be interpreted as an additional theoretical uncertainty in the predictions. As mentioned before, the actual shape of $\mathcal{F}$ will be determined by fitting to the experimental data.

5.2 Different treatments of higher-order singular cross terms

When evaluating the singular correction using the factorization theorem in eq. (2.4) by default we expand the fixed-order boundary conditions of the hard, jet, and partonic soft functions against each other. This is usually done, because it ensures that when switching off the resummation in the far-tail region and adding the fixed-order nonsingular
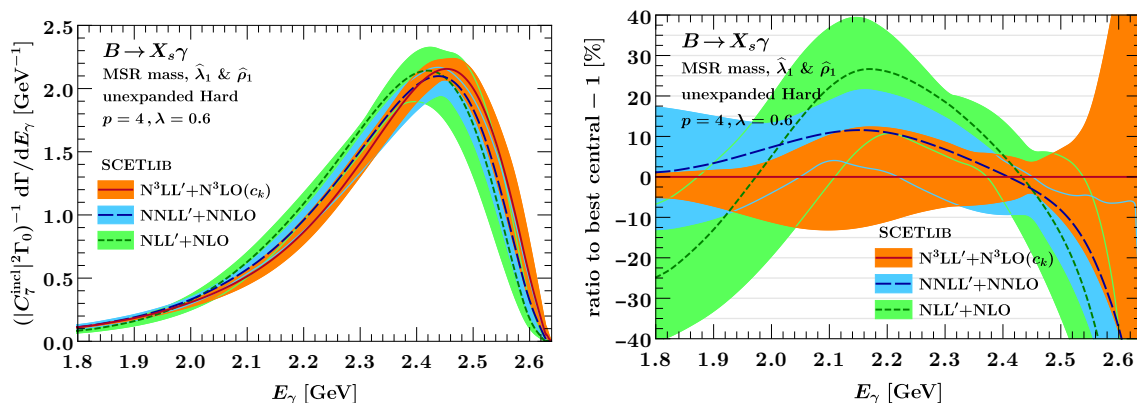
Figure 7. $B \to X_s\gamma$ spectrum with unexpanded hard function.

correction, the fixed-order results are exactly reproduced. In ref. [15] it was observed that in the $1S$ scheme and up to NNLL$'$+NNLO the perturbative convergence in the resummation region is substantially improved by keeping the hard function unexpanded as an overall factor. The main reason is that this ensures that the hard function does not affect the shape of the spectrum, which receives relatively large corrections from the jet and soft functions. The disadvantage is that this has the danger of generating unphysically large higher-order corrections in the fixed-order limit. In ref. [15] it was checked that this does not happen in the region of interest.

In this section, we study the differences in our perturbative setup between expanding the hard function against soft and jet functions (our default) vs. 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. To fix this problem, we compensate for these higher-order terms by adding a constant term that cancels them in the fixed-order region but is only a power-suppressed correction in the peak region,

$$\left[W_{77}^{\mathrm{s}}(k) + W_{77}^{\mathrm{ns}}(k)\right] \to \left[W_{77}^{\mathrm{s,unexp}}(k) - W_{77,\mathrm{FO}}^{\mathrm{s,unexp}}(\widehat{m}_b)\right] + \left[W_{77}^{\mathrm{ns}}(k) + W_{77,\mathrm{FO}}^{\mathrm{s,exp}}(\widehat{m}_b)\right], \quad (5.1)$$

where we subtract the singular with unexpanded hard function, $W_{77,\mathrm{FO}}^{\mathrm{s,unexp}}(\widehat{m}_b)$ and add it back with expanded out hard function, $W_{77,\mathrm{FO}}^{\mathrm{s,exp}}(\widehat{m}_b)$. 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 us 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 7, where we adopt the MSR scheme with short-distance schemes for $\widehat{\lambda}_1$ and $\widehat{\rho}_1$. We compare these results to our default setup using the expanded hard function in figure 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
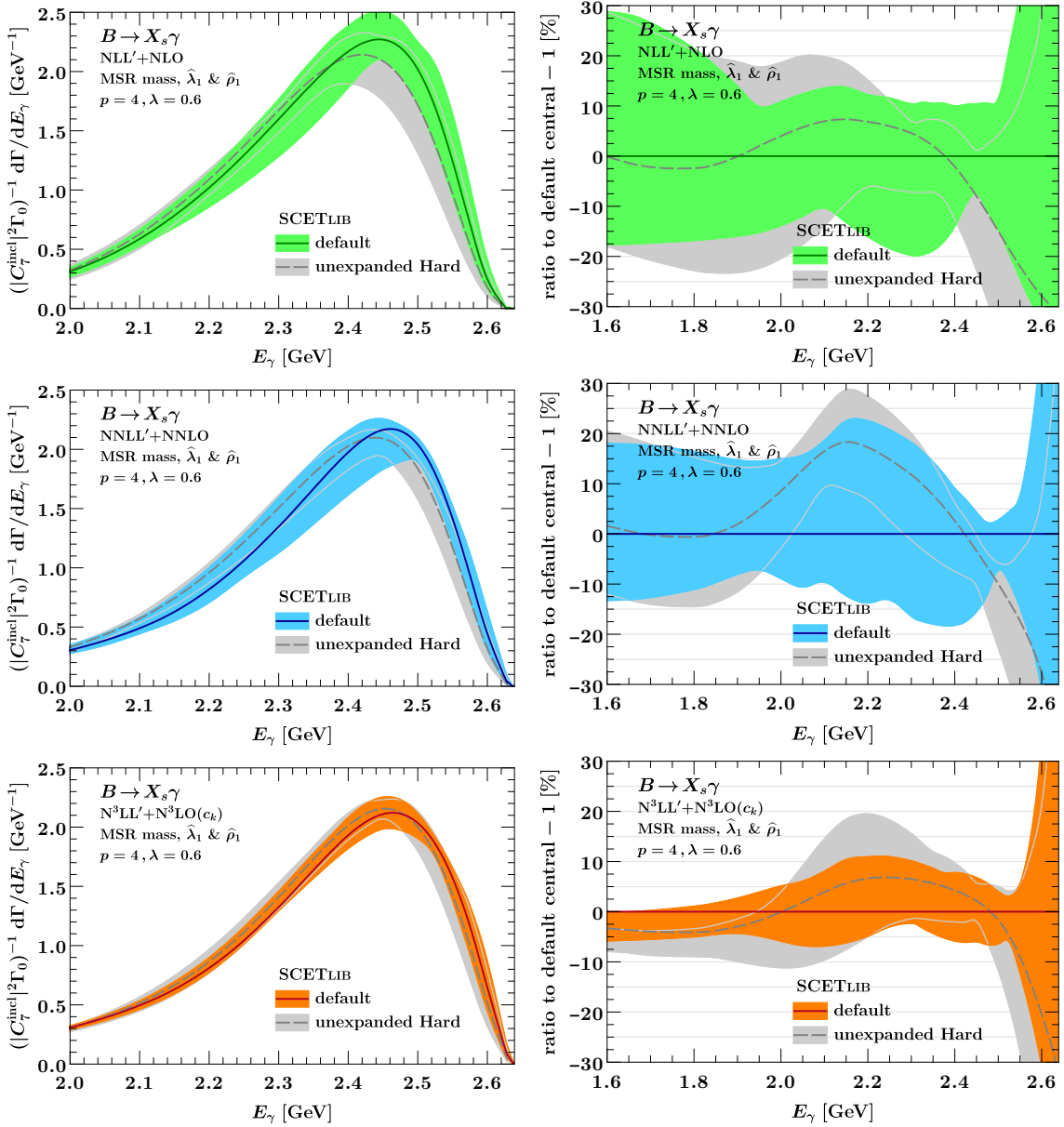
Figure 8. Comparison of $B \to X_s\gamma$ spectrum with expanded and unexpanded hard function. The colored and gray bands at each order display the spectrum with expanded and unexpanded hard function, respectively.

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 N³LL'+N³LO(c_k). This is also visible in figure 7, where the blue band suspiciously narrows down in the peak region of the spectrum and
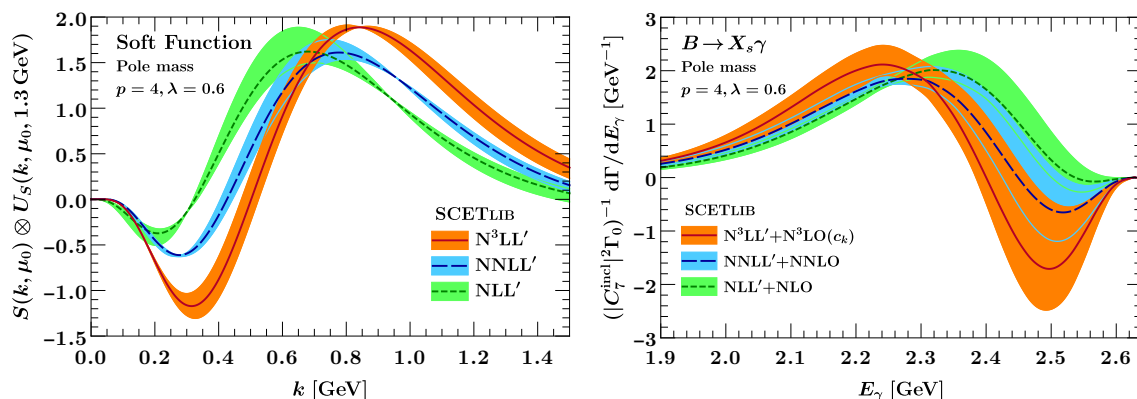
Figure 9. The hadronic soft function (left panel) and the $B \to X_s\gamma$ spectrum (right panel) at different orders in the pole mass scheme. Small values of k in the soft function correspond to large E_γ in the spectrum. The pole-mass renormalon leads to the large negative dip at small k and large E_γ and an unstable position of the peak.

competes with the orange band. Based on this set of observations we conclude that in our setup keeping the hard function unexpanded is disfavored. Hence, for our final numerical results we opt for the conventional treatment of fully expanding the fixed-order boundary conditions of the hard, jet, and partonic soft functions against each other, since it yields an overall more consistent picture of perturbative uncertainties and convergence.

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

As already explained in section 3 a suitable 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 we illustrate the numerical impact of expressing m_b, λ_1, and ρ_1 in short-distance schemes. Moreover, we show that the $1S$ mass scheme, which was successfully used in previous works [15, 32] up to NNLL′+NNLO, starts to break down at N³LL′, while the MSR mass scheme yields convergent, stable results.

In the following figures we show our numerical predictions for the hadronic soft function (left column) and photon energy spectrum (right column) at different orders using the pole (figure 9), $1S$ (figure 10 and figure 12), and MSR (figure 11) schemes. In these plots, we always use our default setup modulo the different short-distance schemes as indicated. For the soft function, we always show the combination $S(k,\mu_0) \otimes U_S(k,\mu_0,1.3\,\text{GeV})$, i.e., we use the soft evolution kernel to evolve the soft function to a fixed scale $\mu = 1.3\,\text{GeV}$. In this way, the μ_0 dependence cancels up to higher order, so we can use our default μ_0 variations to estimate the perturbative uncertainties for the soft function.

In figure 9 we clearly see 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. All these features are also mirrored in the corresponding results for the spectrum. This behaviour of
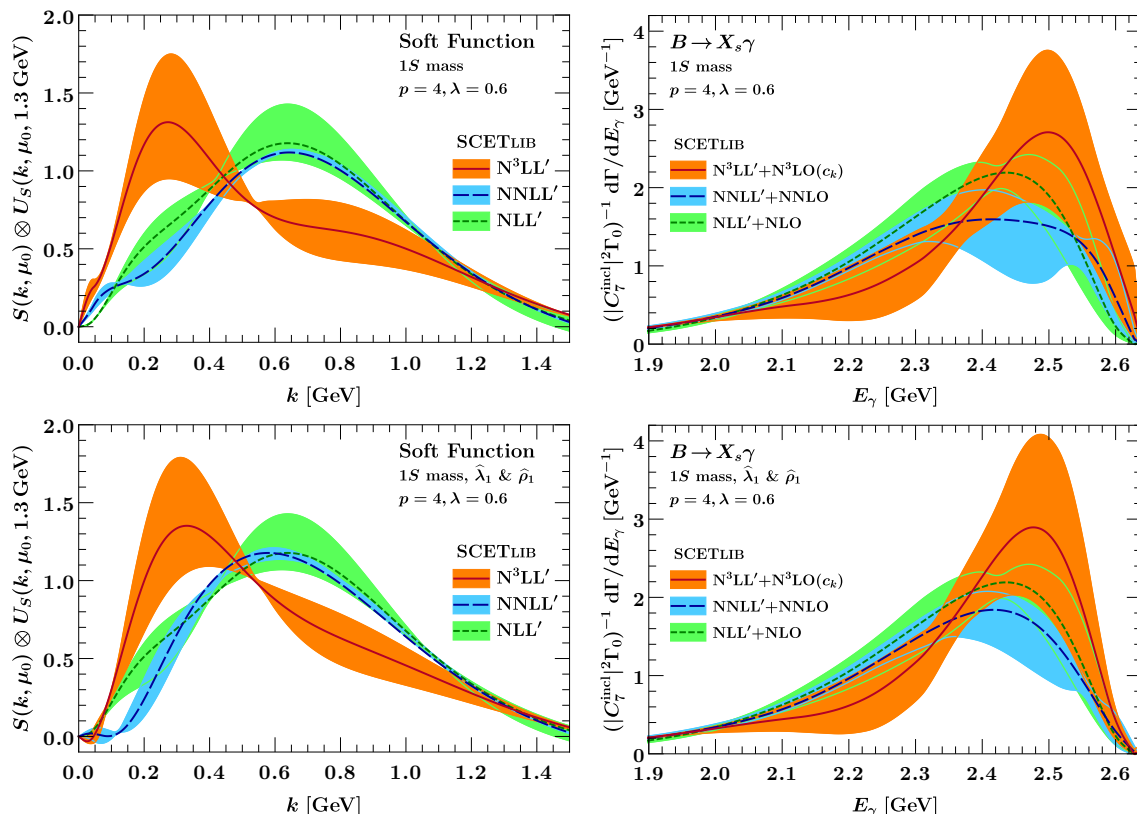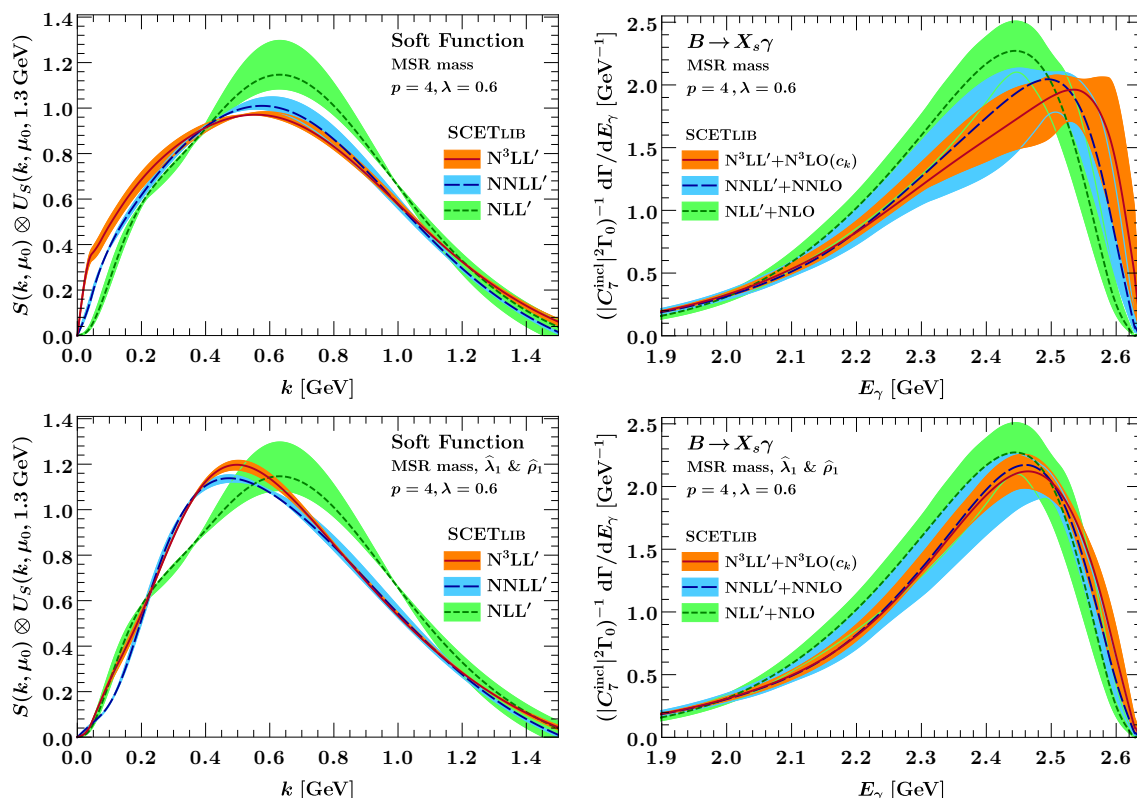
Figure 10. The hadronic soft function (left panel) and the $B \to X_s \gamma$ spectrum (right panel) in the $1S$ mass scheme. The top panels use the pole scheme for λ_1 and ρ_1, while the bottom panels use a short-distance scheme. In both cases the spectrum, and in particular the position of its peak, are stable at NLL$'$+NLO and NNLL$'$+NNLO, but at N^{3}LL$'$+N^{3}LO(c_k) the predictions start breaking down.

the pole scheme was already observed in ref. [32] up to NNLL$'$, and we see that it continues to get worse at N^{3}LL$'$.

The predictions in the $1S$ mass scheme are presented in figure 10. Although the differential decay rate up to NNLL$'$ is somewhat stable, the prediction fails dramatically at N^{3}LL$'$. 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 N^{3}LL$'$+N^{3}LO(c_k) does not change. Neither does it seem to be possible to substantially improve the convergence by adjusting the values of the residual short-distance 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 does not help at all with the bad behaviour at N^{3}LL$'$+N^{3}LO(c_k).

The results in the MSR scheme are presented in figure 11. In this case we observe much more stable results in comparison to the $1S$ scheme. Here in the first row, where we only switch the b-quark mass to the MSR scheme, we still observe that the spectrum is very sensitive to the behavior of the soft function at small k. This sensitivity is reflected in

Figure 11. The hadronic soft function (left panel) and the $B \to X_s \gamma$ spectrum (right panel) in the MSR mass scheme. 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.

the uncertainty estimates from scale variation. By subtracting the subleading renormalons present in λ_1 and ρ_1, we find a substantial improvement in the peak region of the spectrum, and the result exhibits excellent convergence between all orders (second row). From these results we can conclude that the MSR scheme is indeed a much more suitable mass scheme for the $B \to X_s \gamma$ spectrum when going beyond NNLL$'$.

To understand the reason for the breakdown of the $1S$ scheme at N^{3}LL$'$ we recall the relation between the pole and $1S$ schemes up to the 3-loop order,

$$m_b^{\text{pole}} = m_b^{1S} + R^{1S}(\mu) \sum_{n=1}^{3} \sum_{m=0}^{n-1} c_{nm}^{1S} \left[\frac{\alpha_s(\mu)}{4\pi} \right]^n \left[\ln \frac{\mu}{R^{1S}(\mu)} \right]^m. \tag{5.2}$$

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

numerical values for the c_{nm} coefficients are

$$
\begin{aligned}
c_{10}^{1S} &= 2.0944 \,, \\
c_{20}^{1S} &= 135.438 - 10.2393 \, n_f \,, \\
c_{21}^{1S} &= 92.1534 - 5.58505 \, n_f \,, \\
c_{30}^{1S} &= 11398.2 - 1372.75 \, n_f + 38.9677 \, n_f^2 \,, \\
c_{31}^{1S} &= 7766.02 - 1077.92 \, n_f + 33.5103 \, n_f^2 \,, \\
c_{32}^{1S} &= 3041.06 - 368.614 \, n_f + 11.1701 \, n_f^2 \,.
\end{aligned}
\tag{5.3}
$$

By contrast to the MSR scheme where the R scale is a parameter we can choose, in the $1S$ scheme $R^{1S}(\mu)$ 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 we find $R^{1S}(4.75 \, \text{GeV}) = 1.36 \, \text{GeV}$, whereas at the soft scale we obtain $R^{1S}(\mu_S = 1.3 \, \text{GeV}) = 2.40 \, \text{GeV}$, which is almost twice the size of the soft scale itself. Such a large infrared scale violates the power counting of HQET that is used to describe the heavy quarks in the B meson with the residual soft momenta $k \sim \Lambda_{\text{QCD}}$ 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. To see this explicitly, we consider the conversion between the $1S$ and MSR schemes $\Delta m_b(R) \equiv m_b^{\text{MSR}}(R) - m_b^{1S}$, which formally does not involve a renormalon. We obtain the following perturbative series for $\Delta m_b(R = \mu)$ at various scales,

$$
\begin{aligned}
\Delta m_b(R = \mu)\Big|_{\mu=4.2\,\text{GeV}} &= -0.35 \, \epsilon - 0.12 \, \epsilon^2 - 0.04 \, \epsilon^3 \, [\text{GeV}] \,, \\
\Delta m_b(R = \mu)\Big|_{\mu=1.93\,\text{GeV}} &= -0.15 \, \epsilon - 0.06 \, \epsilon^2 + 0.02 \, \epsilon^3 \, [\text{GeV}] \,, \\
\Delta m_b(R = \mu)\Big|_{\mu=1.3\,\text{GeV}} &= -0.06 \, \epsilon - 0.06 \, \epsilon^2 + 0.10 \, \epsilon^3 \, [\text{GeV}] \,,
\end{aligned}
\tag{5.4}
$$

where $\epsilon \equiv 1$ is an auxiliary parameter denoting the perturbative order of the corrections. The perturbative series in the first line shows a rather good convergence for Δm_b at the hard scale. In this regime the perturbative expansion for the $1S$ scheme contains logarithms of the form $\ln(\mu/R^{1S}(\mu))|_{\mu=4.7\,\text{GeV}} \sim \ln(4.7/1.36)$. These logarithms are suppressed by $R^{1S}(\mu = m_b^{1S})/m_b^{1S} = 1.36/4.7$ 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 bottom mass.

We can also define a natural scale for the $1S$ scheme, which we denote as μ_{1S}, at which all logarithms of the form $\ln(\mu_{1S}/R^{1S}(\mu_{1S}))$ are resummed. It corresponds to the fixed point of the R^{1S} scale where $R^{1S}(\mu_{1S}) = \mu_{1S}$, which yields $\mu_{1S} = 1.93 \, \text{GeV}$. The perturbative series for $\Delta m_b(\mu_{1S})$ is shown in the second line of eq. (5.4). It shows the same good convergence as the first line, but with overall smaller corrections and a change of sign in the $\mathcal{O}(\epsilon^3)$ coefficient.

Finally, the last line in eq. (5.4) shows the perturbative series for $\Delta m_b(\mu_S)$ at the soft scale $\mu_S = 1.3 \, \text{GeV}$. Contrary to the first two lines, the resulting series exhibits no conver-
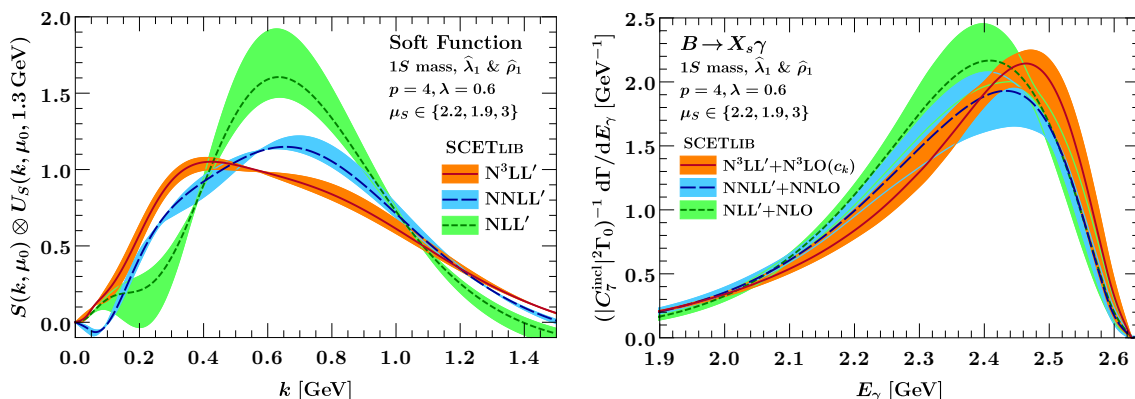
Figure 12. The hadronic soft function (left panel) and the $B \to X_s\gamma$ spectrum (right panel) in the $1S$ mass scheme. A short-distance scheme is used for the hadronic parameters λ_1 and ρ_1. In contrast to the other plots, larger values are used for the soft scale: $\mu_0 \in \{2.2, 1.9, 3\}$ GeV. The larger soft scale amends the breakdown of $1S$ scheme at $N^3LL'+N^3LO(c_k)$, but is not compatible with the SCET scale hierarchy.

gence 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 of $R^{1S} = \mu_{1S} \sim 1.93$ GeV. To examine this hypothesis, in figure 12 we show 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, $\mu_S \in \{2.2, 1.9, 3\}$ GeV. Indeed the resulting spectrum exhibits a significant improvement at all orders compared to figure 10, and in particular the $N^3LL'+N^3LO(c_k)$ 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 our default soft scale, which leads to larger unresummed logarithms in the soft function. Consequently, the results in figure 12 do not reach the same level of stability as those in the MSR scheme in figure 11.

6 Conclusions

In this paper, we presented the photon energy spectrum in $B \to X_s\gamma$ at $N^3LL'+N^3LO(c_k)$, mediated by the electromagnetic operator in the weak Hamiltonian. For the singular contributions we used the SCET factorization theorem to resum large logarithms, which arise in the description of the spectrum close to the kinematic endpoint. We accounted for the complete soft and jet functions at N^3LO and treated the unknown nonlogarithmic constant of the 3-loop hard function as a nuisance parameter. In addition, the RG evolution is performed at complete N^3LL, taking advantage of the fully-known 3-loop anomalous dimensions. We matched the resummed predictions to fixed order at the formal N^3LO order, developing a method that allows us to consistently include the known NNLO results, while parametrizing the missing nonsingular corrections at 3-loop order in terms of suitable

theory nuisance parameters c_k. The variation of these nuisance parameters provides an estimate of the uncertainty that arises from our ignorance of these missing terms.

We incorporated nonperturbative effects by convolving the partonic spectrum with a universal shape function. The first moments of the shape function depend on the b-quark mass m_b and the HQET parameters λ_1 and ρ_1. It is crucial to define these parameters in a suitable short-distance scheme to avoid renormalon ambiguities that would spoil the convergence of perturbative series. Another important aspect of our analysis was the implementation and choice of an appropriate short-distance scheme at N^{3}LL$'$. By using the MSR mass scheme for m_b and adopting an analogous scheme to the "invisible" scheme for λ_1 and ρ_1 we are able to obtain stable predictions. We also find, quite unexpectedly, that the $1S$ mass scheme, which has been successfully used in the past for this process up to NNLL$'$, starts to badly break down at N^{3}LL$'$+N^{3}LO(c_k). We demonstrated that the reason for this sudden breakdown is that the $1S$ scheme is ultimately not designed for use at very low scales. This is because its built-in infrared cutoff scales with $\alpha_s(\mu)$ and thus increases at lower scales and quickly becomes too large when the soft scale is lower than $\mu_{1S} = 1.93\,\mathrm{GeV}$, such that it effectively breaks the power counting of the underlying HQET.

Our main results are presented in section 5.1. Our final predictions for the photon energy spectrum in the MSR scheme and invisible schemes for λ_1 and ρ_1 exhibit excellent perturbative stability, in particular considering the rather low scales involved in the problem. In particular, we observe a substantial improvement of the perturbative theory uncertainties from NNLL$'$+NNLO to N^{3}LL$'$+N^{3}LO(c_k). Importantly, even without the complete fixed $\mathcal{O}(\alpha_s^3)$ information available, with our method we are able to benefit from the increased perturbative precision at N^{3}LL$'$, allowing us to improve the precision of the theory predictions across the entire phenomenologically important peak region. Indeed, the uncertainties due to the missing fixed-order results at $\mathcal{O}(\alpha_s^3)$ only become dominant in the fixed-order tail of the spectrum, which is phenomenologically less relevant. Nevertheless, their calculation is still encouraged and needed to further reduce the theory uncertainties.

In the future, we look forward to confronting our improved predictions with both existing and future experimental measurements, enabling more precise determinations of the shape function $\mathcal{F}(k)$, the b-quark mass, and the normalization of the $B \to X_s\gamma$ rate.

Acknowledgments

A Resummation ingredients

A.1 General definitions

We write the perturbative series for the cusp and noncusp anomalous dimensions as

$$\Gamma_{\mathrm{cusp}}(\alpha_s) \equiv \Gamma^q_{\mathrm{cusp}}(\alpha_s) = \sum_{n=0}^{\infty} \Gamma_n \left(\frac{\alpha_s}{4\pi}\right)^{n+1}, \qquad \gamma^F(\alpha_s) = \sum_{n=0}^{\infty} \gamma_n^F \left(\frac{\alpha_s}{4\pi}\right)^{n+1}. \qquad (\mathrm{A.1})$$

At N³LL′, we need the quark cusp anomalous coefficients up to four loops [55–58]. The coefficients β_n of the QCD β function in $\overline{\text{MS}}$ are defined as

$$\frac{\mathrm{d}\alpha_s(\mu)}{\mathrm{d}\ln\mu} = \beta[\alpha_s(\mu)] = -2\alpha_s(\mu)\sum_{n=0}^{\infty}\beta_n\left[\frac{\alpha_s(\mu)}{4\,\pi}\right]^{n+1}. \tag{A.2}$$

At N³LL′ they are also needed up to four loops [59–62].

The RGE solutions are written in terms of the following standard integrals

$$K_\Gamma(\mu_0,\mu) = \int_{\alpha_s(\mu_0)}^{\alpha(\mu)}\frac{\mathrm{d}\alpha_s}{\beta(\alpha_s)}\,\Gamma_{\text{cusp}}(\alpha_s)\int_{\alpha_s(\mu_0)}^{\alpha_s}\frac{\mathrm{d}\alpha_s'}{\beta(\alpha_s')}\,,$$

$$\eta_\Gamma(\mu_0,\mu) = \int_{\alpha_s(\mu_0)}^{\alpha_s(\mu)}\frac{\mathrm{d}\alpha_s}{\beta(\alpha_s)}\,\Gamma_{\text{cusp}}(\alpha_s)\,,$$

$$K_\gamma(\mu_0,\mu) = \int_{\alpha_s(\mu_0)}^{\alpha_s(\mu)}\frac{\mathrm{d}\alpha_s}{\beta(\alpha_s)}\,\gamma(\alpha_s)\,. \tag{A.3}$$

In principle, they can be performed analytically [63]. For simplicity, in our numerical results we employ the standard approximate analytic solutions [64], which are obtained by performing the integrals after expanding the numerators in α_s. For our purposes here, the numerical accuracy of the approximate analytic solutions is sufficient [64].

Following ref. [32], we define the plus distributions

$$\mathcal{L}_{-1}(x) = \delta(x)\,,$$

$$\mathcal{L}_n(x) = \left[\frac{\theta(x)\ln^n x}{x}\right]_+ = \lim_{\epsilon\to 0}\frac{\mathrm{d}}{\mathrm{d}x}\left[\theta(x-\epsilon)\frac{\ln^{n+1}x}{n+1}\right] \qquad (n\geq 0)\,,$$

$$\mathcal{L}^a(x) = \left[\frac{\theta(x)}{x^{1-a}}\right]_+ = \lim_{\epsilon\to 0}\frac{\mathrm{d}}{\mathrm{d}x}\left[\theta(x-\epsilon)\frac{x^a-1}{a}\right]\,. \tag{A.4}$$

A.2 Hard function

We write the perturbative series for the hard function in the pole-mass scheme as

$$h_s(m_b^{\text{pole}},\mu) = \sum_{n=0}^{\infty}\sum_{m=0}^{2n}H_m^{(n)}\left[\frac{\alpha_s(\mu)}{4\,\pi}\right]^n\ln^m\left(\frac{\mu}{m_b^{\text{pole}}}\right). \tag{A.5}$$

It satisfies the following RGE,

$$\frac{\mathrm{d}h_s(m_b^{\text{pole}},\mu)}{\mathrm{d}\ln\mu} = \left\{\Gamma^H[\alpha_s(\mu)]\ln\frac{\mu}{m_b^{\text{pole}}} + \gamma^H[\alpha_s(\mu)]\right\}h_s(m_b^{\text{pole}},\mu)\,, \tag{A.6}$$

where $\Gamma^H(\alpha_s) \equiv -2\Gamma_{\text{cusp}}(\alpha_s)$ and $\gamma^H(\alpha_s) = 2\gamma^q(\alpha_s) + 2\gamma^Q(\alpha_s)$ are the hard cusp and noncusp anomalous dimensions. The noncusp anomalous coefficients $\gamma_n^H = 2\gamma_n^q + 2\gamma_n^Q$ are known to three loops [23].

The nonlogarithmic coefficients of the hard function in eq. (A.5) are known to two loops [32, 39],

$$H_0^{(0)} = 1\,,$$

$$H_0^{(1)} = -C_F\left(12 + \frac{\pi^2}{6}\right),$$

$$H_0^{(2)} = 16\,C_F\left[3.88611\,C_F + 5.89413\,C_A - \left(\frac{7859}{3456} + \frac{109\pi^2}{576} + \frac{13\zeta_3}{48}\right)\beta_0 \right.$$
$$\left. + \frac{3563}{1296} - \frac{29\pi^2}{108} - \frac{\zeta_3}{6}\right], \tag{A.7}$$

The 3-loop coefficient $H_0^{(3)}$ is currently unknown and treated as a nuisance parameter as discussed in section 2.2, where $h_n \equiv H_0^{(n)}/4^n$.

The coefficients of the logarithmic terms $H_{m\geq 1}^{(n)}$ are determined by iteratively solving the RGE in eq. (A.6) order by order. Substituting eq. (A.5) into eq. (A.6), we obtain a recurrence relation that expresses them in terms of the anomalous dimensions and lower-order nonlogarithmic coefficients,

$$H_m^{(n)} = \frac{1}{m}\left\{\sum_{j=0}^{t_1}(\gamma_j^H + 2\,(n-j-1)\,\beta_j)\,H_{m-1}^{(n-j-1)} + \theta(m \geq 2)\sum_{j=0}^{t_2}\Gamma_j^H\,H_{m-2}^{(n-j-1)}\right\}, \tag{A.8}$$

where $m \geq 1$, the summation limits are $t_1 = \lfloor n - (m+1)/2 \rfloor$ and $t_2 = \lfloor n - m/2 \rfloor$, and the symbol $\lfloor\rfloor$ denotes the floor function. The condition $\theta(m \geq 2)$ indicates that the second sum is present only if $m \geq 2$. The explicit expressions up to three loops read

$$H_1^{(1)} = \gamma_0^H\,,$$

$$H_2^{(1)} = \frac{1}{2}\,\Gamma_0^H\,,$$

$$H_1^{(2)} = H_0^{(1)}\,(2\,\beta_0 + \gamma_0^H) + \gamma_1^H\,,$$

$$H_2^{(2)} = \frac{1}{2}\left\{2\,\beta_0\,\gamma_0^H + (\gamma_0^H)^2 + H_0^{(1)}\Gamma_0^H + \Gamma_1^H\right\},$$

$$H_3^{(2)} = \frac{1}{6}\,\Gamma_0^H\,(2\,\beta_0 + 3\,\gamma_0^H)\,,$$

$$H_4^{(2)} = \frac{1}{8}\,(\Gamma_0^H)^2\,,$$

$$H_1^{(3)} = H_0^{(2)}\,(4\,\beta_0 + \gamma_0^H) + H_0^{(1)}\,(2\,\beta_1 + \gamma_1^H) + \gamma_2^H\,,$$

$$H_2^{(3)} = \frac{1}{2}\left\{2\,\beta_1\,\gamma_0^H + 4\,\beta_0\,\gamma_1^H + 2\,\gamma_0^H\,\gamma_1^H + H_0^{(2)}\Gamma_0^H \right.$$
$$\left. + H_0^{(1)}\left(8\,\beta_0^2 + 6\,\beta_0\,\gamma_0^H + (\gamma_0^H)^2 + \Gamma_1^H\right) + \Gamma_2^H\right\},$$

$$H_3^{(3)} = \frac{1}{6}\left\{8\,\beta_0^2\,\gamma_0^H + (\gamma_0^H)^3 + (2\,\beta_1 + 3\,\gamma_1^H)\,\Gamma_0^H + 3\,\gamma_0^H\,(H_0^{(1)}\,\Gamma_0^H + \Gamma_1^H) \right.$$
$$\left. + \beta_0\left(6\,(\gamma_0^H)^2 + 8\,H_0^{(1)}\,\Gamma_0^H + 4\,\Gamma_1^H\right)\right\},$$

$$H_4^{(3)} = \frac{1}{24}\,\Gamma_0^H\,(8\,\beta_0^2 + 20\,\beta_0\,\gamma_0^H + 6\,(\gamma_0^H)^2 + 3\,H_0^{(1)}\Gamma_0^H + 6\,\Gamma_1^H)\,,$$

$$H_5^{(3)} = \frac{1}{24} \left(\Gamma_0^H \right)^2 \left(4\beta_0 + 3\gamma_0^H \right),$$

$$H_6^{(3)} = \frac{1}{48} (\Gamma_0^H)^3 \,. \tag{A.9}$$

The all-order solution of the RGE in eq. (A.6) is given by

$$h_s(m_b^{\text{pole}}, \mu) = h_s(m_b^{\text{pole}}, \mu_H) \, U_H(m_b^{\text{pole}}, \mu_H, \mu) \,, \tag{A.10}$$

$$U_H(m_b, \mu_H, \mu) = \exp\left[-2K_\Gamma(\mu_H, \mu) - 2\eta_\Gamma(\mu_H, \mu) \ln \frac{\mu_H}{m_b} + K_{\gamma^H}(\mu_H, \mu) \right] \,.$$

A.3 Jet function

The perturbative series for the renormalized jet function reads

$$J(s, \mu) = \sum_{n=0}^{\infty} \sum_{m=-1}^{2n-1} J_m^{(n)} \left[\frac{\alpha_s(\mu)}{4\pi} \right]^n \frac{1}{\mu^2} \mathcal{L}_m \left(\frac{s}{\mu^2} \right) \,. \tag{A.11}$$

The jet function is normalized such that $J_{-1}^{(n)} = 1$. Explicit expressions for the coefficients $J_m^{(n)}$ up to 3-loop order in our notation are given in refs. [22, 65].

For completeness, the jet function obeys the RGE

$$\frac{\mathrm{d}J(s, \mu)}{\mathrm{d}\ln\mu} = \left\{ \Gamma^J[\alpha_s(\mu)] \frac{1}{\mu^2} \mathcal{L}_0\left(\frac{s}{\mu^2} \right) + \gamma^J[\alpha_s(\mu)] \, \delta(s) \right\} \otimes_s J(s, \mu) \tag{A.12}$$

where $\Gamma^J(\alpha_s) = -2\Gamma_{\text{cusp}}(\alpha_s)$ and the symbol $\otimes_s$ denotes the convolution of the form

$$f(s) \otimes_s g(s) = \int \mathrm{d}s' \, f(s - s') \, g(s') \,. \tag{A.13}$$

In our numerical implementation, we do not need to explicitly solve the jet-function RGE, because we always evolve the hard and soft functions to the jet scale.

A.4 Partonic soft function

In the pole scheme the partonic soft function $C_0(\omega, \mu)$ is given by the b-quark matrix element

$$C_0(\omega, \mu) = \langle b_v | \bar{b}_v \delta(iD_+ + \omega) b_v | b_v \rangle \,. \tag{A.14}$$

Its perturbative expansion is written as

$$C_0(\omega, \mu) = \sum_{n=0}^{\infty} \sum_{m=-1}^{2n-1} S_m^{(n)} \left[\frac{\alpha_s(\mu)}{4\pi} \right]^n \frac{1}{\mu} \mathcal{L}_m \left(\frac{\omega}{\mu} \right) \,. \tag{A.15}$$

The expansion coefficients $S_m^{(n)}$ can be found in ref. [23].

The RGE for the partonic soft function reads

$$\frac{\mathrm{d}C_0(\omega, \mu)}{\mathrm{d}\ln\mu} = \left\{ \Gamma^S[\alpha_s(\mu)] \frac{1}{\mu} \mathcal{L}_0\left(\frac{\omega}{\mu} \right) + \gamma^S[\alpha_s(\mu)] \, \delta(\omega) \right\} \otimes_\omega C_0(\omega, \mu) \,, \tag{A.16}$$

where $\Gamma^S(\alpha_s) = 2\Gamma_{\mathrm{cusp}}(\alpha_s)$ and the noncusp anomalous coefficients are known to three loops [23]. By solving it iteratively, we can obtain a recurrence relation for all logarithmic coefficients $S_{m \geq 0}^{(n)}$,

$$S_m^{(n)} = -\frac{1}{m + \delta_{m0}} \left\{ \sum_{j=0}^{t_1} \left[\gamma_j^S + 2(n-j-1)\beta_j \right] S_{m-1}^{(n-j-1)} + \sum_{j=0}^{t_2} \sum_{i=t_m-2}^{2(n-j)-3} \Gamma_j^S V_{m-1}^{0i} S_i^{(n-j-1)} \right\}, \tag{A.17}$$

where the summation limits are $t_1 = \lfloor n - 1 - m/2 \rfloor$, $t_2 = \lfloor n - (t_m + 1)/2 \rfloor$ and $t_m = \max(m, 1)$. The coefficients V_k^{mn} appear in the convolution algebra of plus distributions,

$$\mathcal{L}_n(x) \otimes_x \mathcal{L}_m(x) = \sum_{k=-1}^{n+m+1} V_k^{nm} \mathcal{L}_k(x), \tag{A.18}$$

and are given in ref. [32]. It is easy to check that eq. (A.17) reproduces the explicit results to three loops in ref. [23].

The all-order solution of the soft RGE is [32, 66–68]

$$C_0(\omega, \mu) = C_0(\omega, \mu_S) \otimes_\omega U_S(\omega, \mu_S, \mu),$$

$$U_S(\omega, \mu_S, \mu) = \exp\left[-2K_\Gamma(\mu_S, \mu) + K_{\gamma^S}(\mu_S, \mu) \right] \mathcal{V}[2\eta_\Gamma(\mu_S, \mu), \mu_S, \omega],$$

$$\mathcal{V}(\eta, \mu, \omega) = \frac{e^{-\gamma_E \eta}}{\Gamma(1+\eta)} \left[\frac{\eta}{\mu} \mathcal{L}^\eta\left(\frac{\omega}{\mu}\right) + \delta(\omega) \right] \tag{A.19}$$

where K_Γ, η_Γ, K_γ are defined in eq. (A.3), γ_E is the Euler-Mascheroni constant, and the $\mathcal{L}^\eta(x)$ plus distribution is defined in eq. (A.4).

Note that in the short-distance schemes we consider, the residual terms δm_b, $\delta\lambda_1$, $\delta\rho_1$ are formally μ independent, such that the evolution for the partonic soft function in a short-distance scheme $\widehat{C}_0(\omega, \mu)$ is the same as in the pole scheme.

Open Access. This article is distributed under the terms of the Creative Commons Attribution License ([CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

References

[1] S. Bertolini, F. Borzumati and A. Masiero, *QCD enhancement of radiative B decays*, *Phys. Rev. Lett.* **59** (1987) 180 [INSPIRE].

[2] B. Grinstein, R.P. Springer and M.B. Wise, *Effective Hamiltonian for weak radiative B meson decay*, *Phys. Lett. B* **202** (1988) 138 [INSPIRE].

[3] M. Misiak et al., *Estimate of $\mathcal{B}(\bar{B} \to X_s \gamma)$ at $O(\alpha_s^2)$*, *Phys. Rev. Lett.* **98** (2007) 022002 [hep-ph/0609232] [INSPIRE].

[4] M. Misiak et al., *Updated NNLO QCD predictions for the weak radiative B-meson decays*, *Phys. Rev. Lett.* **114** (2015) 221801 [arXiv:1503.01789] [INSPIRE].

[5] B. Grinstein and M.B. Wise, *Weak radiative B meson decay as a probe of the Higgs sector*, *Phys. Lett. B* **201** (1988) 274 [INSPIRE].

[6] W.-S. Hou and R.S. Willey, *Effects of charged Higgs bosons on the processes $b \to s\gamma$, $b \to sg*$ and $b \to s\ell^+\ell^-$*, *Phys. Lett. B* **202** (1988) 591 [INSPIRE].

[7] M. Misiak and M. Steinhauser, *Weak radiative decays of the B meson and bounds on $M_{H^\pm}$ in the two-Higgs-doublet model*, *Eur. Phys. J. C* **77** (2017) 201 [arXiv:1702.04571] [INSPIRE].

[8] M. Neubert, *Analysis of the photon spectrum in inclusive $B \to X_s\gamma$ decays*, *Phys. Rev. D* **49** (1994) 4623 [hep-ph/9312311] [INSPIRE].

[9] I.I.Y. Bigi, M.A. Shifman, N.G. Uraltsev and A.I. Vainshtein, *On the motion of heavy quarks inside hadrons: universal distributions and inclusive decays*, *Int. J. Mod. Phys. A* **9** (1994) 2467 [hep-ph/9312359] [INSPIRE].

[10] M. Neubert, *QCD based interpretation of the lepton spectrum in inclusive $\bar{B} \to X_u\ell\bar{\nu}$ decays*, *Phys. Rev. D* **49** (1994) 3392 [hep-ph/9311325] [INSPIRE].

[11] BaBar collaboration, *Measurement of the $B \to X_s\gamma$ branching fraction and photon energy spectrum using the recoil method*, *Phys. Rev. D* **77** (2008) 051103 [arXiv:0711.4889] [INSPIRE].

[12] Belle collaboration, *Measurement of inclusive radiative B-meson decays with a photon energy threshold of 1.7 GeV*, *Phys. Rev. Lett.* **103** (2009) 241801 [arXiv:0907.1384] [INSPIRE].

[13] BaBar collaboration, *Measurement of $B(B \to X_s\gamma)$, the $B \to X_s\gamma$ photon energy spectrum, and the direct CP asymmetry in $B \to X_{s+d}\gamma$ decays*, *Phys. Rev. D* **86** (2012) 112008 [arXiv:1207.5772] [INSPIRE].

[14] BaBar collaboration, *Exclusive measurements of $b \to s\gamma$ transition rate and photon energy spectrum*, *Phys. Rev. D* **86** (2012) 052012 [arXiv:1207.2520] [INSPIRE].

[15] SIMBA collaboration, *Precision global determination of the $B \to X_s\gamma$ decay rate*, *Phys. Rev. Lett.* **127** (2021) 102001 [arXiv:2007.04320] [INSPIRE].

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

[17] K.S.M. Lee and I.W. Stewart, *Shape-function effects and split matching in $B \to X_s\ell^+\ell^-$*, *Phys. Rev. D* **74** (2006) 014005 [hep-ph/0511334] [INSPIRE].

[18] K.S.M. Lee, Z. Ligeti, I.W. Stewart and F.J. Tackmann, *Universality and m_X cut effects in $B \to X_s\ell^+\ell^-$*, *Phys. Rev. D* **74** (2006) 011501 [hep-ph/0512191] [INSPIRE].

[19] Particle Data Group collaboration, *Review of particle physics*, *PTEP* **2022** (2022) 083C01 [INSPIRE].

[20] Belle-II collaboration, *Measurement of the photon-energy spectrum in inclusive $B \to X_s\gamma$ decays identified using hadronic decays of the recoil B meson in* 2019–2021 *Belle II data*, arXiv:2210.10220 [INSPIRE].

[21] Belle-II collaboration, *Snowmass white paper: Belle II physics reach and plans for the next decade and beyond*, arXiv:2207.06307 [INSPIRE].

[22] R. Brüser, Z.L. Liu and M. Stahlhofen, *Three-loop quark jet function*, *Phys. Rev. Lett.* **121** (2018) 072003 [arXiv:1804.09722] [INSPIRE].

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

[24] F.J. Tackmann, *Theory uncertainties from nuisance parameters*, talk at *SCET* 2019 *workshop*, https://indico.physics.lbl.gov/event/694/contributions/3344/, March 2019.

[25] F.J. Tackmann, *Beyond scale variations: perturbative theory uncertainties from nuisance parameters*, DESY-19-021, DESY, Hamburg, Germany (2019).

[26] I.R. Blokland et al., *The electromagnetic dipole operator effect on* $\bar{B} \to X_s \gamma$ *at* $O(\alpha_s^2)$, *Phys. Rev. D* **72** (2005) 033014 [hep-ph/0506055] [INSPIRE].

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

[28] H.M. Asatrian et al., *Magnetic dipole operator contributions to the photon energy spectrum in* $\bar{B} \to X_s \gamma$ *at* $O(\alpha_s^2)$, *Nucl. Phys. B* **762** (2007) 212 [hep-ph/0607316] [INSPIRE].

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

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

[31] 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] [INSPIRE].

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

[33] 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 [arXiv:0803.4214] [INSPIRE].

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

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

[36] 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] [INSPIRE].

[37] 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] [INSPIRE].

[38] 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] [INSPIRE].

[39] 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] [INSPIRE].

[40] J.R. Andersen and E. Gardi, *Radiative B decay spectrum: DGE at NNLO*, *JHEP* **01** (2007) 029 [hep-ph/0609250] [INSPIRE].

[41] M. Misiak, *QCD calculations of radiative B decays*, in the proceedings of the *Heavy Quarks and Leptons* 2008 *(HQ&L08)*, (2008) [arXiv:0808.3134] [INSPIRE].

[42] R. Abbate et al., *Thrust at* N^3LL *with power corrections and a precision global fit for* $\alpha_s(mZ)$, *Phys. Rev. D* **83** (2011) 074021 [arXiv:1006.3080] [INSPIRE].

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

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

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

[46] M. Gremm and A. Kapustin, *Order $1/m_b^3$ corrections to $B \to X_c \ell \bar{\nu}$ decay and their implication for the measurement of $\bar{\Lambda}$ and λ_1*, *Phys. Rev. D* **55** (1997) 6924 [hep-ph/9603448] [INSPIRE].

[47] 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] [INSPIRE].

[48] A.H. Hoang, *The top mass: interpretation and theoretical uncertainties*, in the proceedings of the 7$^{\text{th}}$ *International Workshop on Top Quark Physics*, (2014) [arXiv:1412.3649] [INSPIRE].

[49] A.H. Hoang et al., *The MSR mass and the $O(\Lambda_{\text{QCD}})$ renormalon sum rule*, *JHEP* **04** (2018) 003 [arXiv:1704.01580] [INSPIRE].

[50] M. Butenschoen et al., *Top quark mass calibration for Monte Carlo event generators*, *Phys. Rev. Lett.* **117** (2016) 232001 [arXiv:1608.01318] [INSPIRE].

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

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

[53] 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] [INSPIRE].

[54] M.A. Ebert et al., *SCETlib: a C++ package for numerical calculations in QCD and soft-collinear effective theory*, DESY-17-099, DESY, Hamburg, Germany (2018).

[55] G.P. Korchemsky and A.V. Radyushkin, *Renormalization of the Wilson loops beyond the leading order*, *Nucl. Phys. B* **283** (1987) 342 [INSPIRE].

[56] 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] [INSPIRE].

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

[58] 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 [arXiv:2002.04617] [INSPIRE].

[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 [INSPIRE].

[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] [INSPIRE].

[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] [INSPIRE].

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

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

[64] G. Billis, F.J. Tackmann and J. Talbert, *Higher-order Sudakov resummation in coupled gauge theories*, *JHEP* **03** (2020) 182 [arXiv:1907.02971] [INSPIRE].

[65] J. Gaunt, M. Stahlhofen, F.J. Tackmann and J.R. Walsh, *N-jettiness subtractions for NNLO QCD calculations*, *JHEP* **09** (2015) 058 [arXiv:1505.04794] [INSPIRE].

[66] 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] [INSPIRE].

[67] M. Neubert, *Renormalization-group improved calculation of the $B \to X_s\gamma$ branching ratio*, *Eur. Phys. J. C* **40** (2005) 165 [hep-ph/0408179] [INSPIRE].

[68] 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 [arXiv:0711.2079] [INSPIRE].