

**DEUTSCHES ELEKTRONEN-SYNCHROTRON**  
Ein Forschungszentrum der Helmholtz-Gemeinschaft



DESY 20-115  
MIT-CTP 5220  
arXiv:2007.04320  
July 2020

**Precision Global Determination of the  
 $B \rightarrow X_s \gamma$  Decay Rate**

SIMBA Collaboration

ISSN 0418-9833

**NOTKESTRASSE 85 - 22607 HAMBURG**

DESY behält sich alle Rechte für den Fall der Schutzrechtserteilung und für die wirtschaftliche Verwertung der in diesem Bericht enthaltenen Informationen vor.

DESY reserves all rights for commercial use of information included in this report, especially in case of filing application for or grant of patents.

To be sure that your reports and preprints are promptly included in the  
HEP literature database  
send them to (if possible by air mail):

DESY Zentralbibliothek Notkestraße 85 22607 Hamburg Germany	DESY Bibliothek Platanenallee 6 15738 Zeuthen Germany
---	---

# Precision Global Determination of the $B \rightarrow X_s \gamma$ Decay Rate

Florian U. Bernlochner,<sup>1</sup> Heiko Lacker,<sup>2</sup> Zoltan Ligeti,<sup>3</sup>  
Iain W. Stewart,<sup>4</sup> Frank J. Tackmann,<sup>5</sup> and Kerstin Tackmann<sup>5</sup>  
(SIMBA Collaboration)

<sup>1</sup>*Physikalisches Institut, Rheinische Friedrich-Wilhelms-Universität Bonn, D-53113 Bonn, Germany*

<sup>2</sup>*Humboldt University of Berlin, D-12489 Berlin, Germany*

<sup>3</sup>*Lawrence Berkeley National Laboratory, University of California, Berkeley, California 94720, USA*

<sup>4</sup>*Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*

<sup>5</sup>*Deutsches Elektronen-Synchrotron (DESY), D-22607 Hamburg, Germany*

(Dated: July 8, 2020)

We perform the first global fit to inclusive  $B \rightarrow X_s \gamma$  measurements using a model-independent treatment of the nonperturbative  $b$ -quark distribution function, with next-to-next-to-leading logarithmic resummation and  $\mathcal{O}(\alpha_s^2)$  fixed-order contributions. The normalization of the  $B \rightarrow X_s \gamma$  decay rate, given by  $|C_7^{\text{incl}} V_{tb} V_{ts}^*|^2$ , is sensitive to physics beyond the Standard Model (SM). We determine  $|C_7^{\text{incl}} V_{tb} V_{ts}^*| = (14.77 \pm 0.51_{\text{fit}} \pm 0.59_{\text{theory}} \pm 0.08_{\text{param}}) \times 10^{-3}$ , in good agreement with the SM prediction, and the  $b$ -quark mass  $m_b^{1S} = (4.750 \pm 0.027_{\text{fit}} \pm 0.033_{\text{theory}} \pm 0.003_{\text{param}}) \text{ GeV}$ . Our results suggest that the uncertainties in the extracted  $B \rightarrow X_s \gamma$  rate have been underestimated by up to a factor of two, leaving more room for beyond-SM contributions.

*Introduction* The flavor-changing neutral-current  $b \rightarrow s \gamma$  transition is well known for its high sensitivity to contributions beyond the Standard Model (SM). The main goal of our global analysis of the  $B \rightarrow X_s \gamma$  decay rate is to obtain a precise constraint on the short-distance physics it probes, which can then be compared to predictions in the SM [1–4] or beyond [5–7]. In our approach, this amounts to extracting a precise value of the Wilson coefficient  $|C_7^{\text{incl}}|$  from the measurements.

Since  $b \rightarrow s \gamma$  is a two-body decay at tree level, the photon energy spectrum,  $d\Gamma/dE_\gamma$ , peaks only a few hundred MeV below the kinematic limit  $E_\gamma \lesssim m_B/2$ . In this peak region, the measurements are most precise, but the theory predictions depend on a nonperturbative function,  $\mathcal{F}(k)$ , often called the shape function, which encodes the distribution of the residual momentum  $k$  of the  $b$ -quark in a  $B$  meson [8, 9]. A key aspect of our analysis is a model-independent treatment of  $\mathcal{F}(k)$  based on expanding it in a suitable basis [10]. This approach can incorporate any given shape function model, by using it as the generating function for the basis expansion, and thus goes beyond existing approaches that use specific models [11–15].

While  $\mathcal{F}(k)$  primarily affects the shape of the decay spectrum, its normalization is determined by  $|C_7^{\text{incl}}|^2$ , up to small corrections. Thus, with our treatment of  $\mathcal{F}(k)$ , we can perform a global fit to the measurements of  $d\Gamma/dE_\gamma$ , including the precisely measured peak region, to simultaneously determine  $\mathcal{F}(k)$  and a precise value of  $|C_7^{\text{incl}}|$ . Our global fit is the first to exploit the full available experimental information on the spectrum [16–19], together with the most precise theoretical knowledge of its perturbative contributions. This provides a more robust approach than the current method of using theoretical predictions for the  $B \rightarrow X_s \gamma$  rate with a fixed cut at  $E_\gamma > 1.6 \text{ GeV}$  [4] and corresponding extrapolated measurements [20].

*The  $B \rightarrow X_s \gamma$  Spectrum* Using SCET [21–24], we can write the photon energy spectrum in a factorized form,

$$\frac{d\Gamma}{dE_\gamma} = 2\Gamma_0 \frac{(2E_\gamma)^3}{\hat{m}_b^3} \left[ \int dk \hat{P}(k) \mathcal{F}(m_B - 2E_\gamma - k) + \frac{1}{\hat{m}_b} \sum_a (\hat{P}_a \otimes g_a)(m_B - 2E_\gamma) \right], \quad (1)$$

where

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

and  $\hat{m}_b$  denotes a short-distance  $b$ -quark mass, for which we use the  $1S$  scheme [25–27].

The first term in Eq. (1) is the dominant contribution, where  $\mathcal{F}(k)$  contains the leading nonperturbative shape function plus a combination of subleading shape functions specific for  $B \rightarrow X_s \gamma$ . The function  $\hat{P}(k)$  encodes the perturbatively calculable  $b \rightarrow s \gamma$  spectrum, with  $k \sim m_b - 2E_\gamma$ . It receives contributions from different operators in the effective electroweak Hamiltonian,

$$\hat{P}(k) = |C_7^{\text{incl}}|^2 \left[ W_{77}^s(k) + W_{77}^{\text{ns}}(k) \right] + 2 \text{Re}(C_7^{\text{incl}}) \sum_{i \neq 7} C_i W_{7i}^{\text{ns}}(k) + \sum_{i,j \neq 7} C_i C_j W_{ij}^{\text{ns}}(k). \quad (3)$$

Here,  $W_{77}^s(k)$  contains the universal “singular” contributions proportional to  $\alpha_s^i \ln^j(k/m_b)/k$  and  $\alpha_s^i \delta(k)$ , which dominate in the peak region where  $k$  is small [28]. It is included following Ref. [10] to NNLL’ order, which includes next-to-next-to-leading-logarithmic (NNLL) resummation and all singular terms at  $\mathcal{O}(\alpha_s^2)$  [22, 29–37].

The coefficient  $C_7^{\text{incl}}$  is dominated by the Wilson coefficient  $\bar{C}_7(\mu)$  in the electroweak Hamiltonian,

$$C_7^{\text{incl}} = \bar{C}_7(\mu) + \sum_{i \neq 7} \bar{C}_i(\mu) [s_i(\mu, \hat{m}_b) + r_i(\mu, \hat{m}_b, \hat{m}_c)]. \quad (4)$$

The  $s_i$  terms are defined to cancel the  $\mu$  dependence of  $\overline{C}_7(\mu)$  and to satisfy  $s_i(\widehat{m}_b, \widehat{m}_b) = 0$ . The  $\overline{C}_i r_i$  terms contain all virtual corrections proportional to  $\overline{C}_{i \neq 7}$  that give rise to singular contributions. In particular, they contain the sizable corrections from virtual  $c\bar{c}$  loops, and the resulting sensitivity to the charm quark mass,  $\widehat{m}_c$ , which are a dominant theory uncertainty in the decay rate. Since in our approach these contributions are included in  $C_7^{\text{incl}}$ , they only affect its SM prediction, but not its determination from the experimental data. The results of Refs. [3, 4, 38, 39] yield the NNLO SM prediction [28],

$$|C_7^{\text{incl}}|_{\text{SM}} = 0.3624 \pm 0.0128_{c\bar{c}} \pm 0.0080_{\text{scale}}. \quad (5)$$

The remaining  $W_{ij}^{\text{ns}}(k)$  terms in Eq. (3) are “nonsingular” contributions with  $\mathcal{C}_i = \overline{C}_i(\widehat{m}_b)$  [28]. They start at  $\mathcal{O}(\alpha_s)$  and are suppressed by at least  $k/m_b$  relative to  $W_{77}^s(k)$ , and are therefore subleading in the peak region. They are included to full  $\mathcal{O}(\alpha_s^2)$  for  $ij = 77, 78$  [40–42], while the remaining ones are known and included to  $\mathcal{O}(\alpha_s^2\beta_0)$  [43–45]. Since  $W_{77}^s(k)$  dominates in the peak region, the normalization of the spectrum is determined by  $|C_7^{\text{incl}}|$ , enabling its precise extraction.

The second term in Eq. (1) is subdominant, and describes so-called resolved and unresolved contributions, where  $\widehat{P}_a$  are perturbative coefficients starting at  $\mathcal{O}(\alpha_s)$ , and the  $g_a$  are additional subleading shape functions [46]. The uncertainties from resolved contributions are much smaller than suggested by earlier estimates [47], and are not relevant at the current level of accuracy [28] (see also Ref. [48]). The only marginally relevant contribution is related to the known  $\mathcal{O}(1/\widehat{m}_c^2)$  correction to the total rate [49–51], and is included in our analysis via a subleading  $\mathcal{O}(\Lambda_{\text{QCD}}^2)$  shape function  $g_{27}(k)$ .

The nonperturbative shape function  $\mathcal{F}(k)$  is dominated by the leading-order shape function, so we assume it is positive. We introduce a dimension-1 parameter  $\lambda$ , and expand  $\mathcal{F}(k)$  as [10],

$$\mathcal{F}(k) = \frac{1}{\lambda} \left[ \sum_{n=0}^{\infty} \tilde{c}_n f_n\left(\frac{k}{\lambda}\right) \right]^2, \quad (6)$$

where  $f_n(x)$  are a suitably chosen complete set of orthonormal functions on  $[0, \infty)$ . The normalization condition  $\int_0^{\infty} dk \mathcal{F}(k) = 1$  implies

$$\sum_{n=0}^{\infty} \tilde{c}_n^2 = 1. \quad (7)$$

In practice, the expansion for  $\mathcal{F}(k)$  must be truncated at a finite order  $N$ . Therefore, the form of  $\mathcal{F}(k)$  used for the fit is given by the following approximation

$$\mathcal{F}(k) = \sum_{m,n=0}^N c_m c_n F_{mn}(k), \quad (8)$$

where

$$F_{mn}(k) = \frac{1}{\lambda} f_m\left(\frac{k}{\lambda}\right) f_n\left(\frac{k}{\lambda}\right). \quad (9)$$

The effect of the truncation in Eq. (8) is approximated by the modified coefficients  $c_n$ , which differ from the  $\tilde{c}_n$  in Eq. (8). In particular, we always keep the normalization of  $\mathcal{F}(k)$  exact by enforcing

$$\sum_{n=0}^N c_n^2 = 1. \quad (10)$$

Using the expansion for  $\mathcal{F}(k)$  in Eq. (8) we get

$$\begin{aligned} \frac{d\Gamma}{dE_\gamma} &= 16\Gamma_0 \frac{E_\gamma^3}{\widehat{m}_b^3} \sum_{m,n=0}^N c_m c_n \int dk \widehat{P}(k) F_{mn}(m_B - 2E_\gamma - k) \\ &\quad + 16\Gamma_0 \frac{E_\gamma^3}{\widehat{m}_b^3} \frac{1}{\widehat{m}_b^2} \int dk \widehat{P}_{27}(k) g_{27}(m_B - 2E_\gamma - k) \\ &\equiv N_s \sum_{m,n=0}^N c_m c_n \frac{d\Gamma_{77,mn}}{dE_\gamma} + \dots \end{aligned} \quad (11)$$

Here,  $N_s = |C_7^{\text{incl}} V_{tb} V_{ts}^*|^2 \widehat{m}_b^2$ , and Eq. (11) defines  $d\Gamma_{77,mn}/dE_\gamma$ , which we precompute from Eq. (3). The ellipses denote subleading terms not proportional to  $|C_7^{\text{incl}}|^2$ , which are also written in terms of  $N_s$  and  $c_n$  as explained in [28]. Then,  $N_s$  and the  $c_n$  are fitted from the measured spectra, with the uncertainties and correlations in the measurements captured in the uncertainties and correlations of the fit parameters. Using the moment relations for  $\mathcal{F}(k)$  [28], we obtain  $C_7^{\text{incl}}$  and  $\widehat{m}_b$ , as well as the heavy-quark parameters  $\widehat{\lambda}_1$  and  $\widehat{\rho}_1$  from the fitted  $N_s$  and  $c_n$ . The other coefficients  $\mathcal{C}_{i \neq 7}$  are fixed to their SM values [28]. Of these, only  $\mathcal{C}_1$  and  $\mathcal{C}_2$  are numerically relevant, which are known to be SM dominated, while  $\mathcal{C}_8$ , which is sensitive to new physics, gives only a small contribution. We use input values for  $\widehat{\lambda}_2$  and  $\widehat{\rho}_2$ , which are obtained from the  $B$  and  $D$  meson mass splittings [28].

*Fit procedure* We implement a binned  $\chi^2$  fit, with

$$\chi^2 = \sum_{i,j} (\Gamma_i^{\text{meas}} - \Gamma_i) (V^{-1})_{ij} (\Gamma_j^{\text{meas}} - \Gamma_j). \quad (12)$$

Here  $\Gamma_i^{\text{meas}}$  is the measured  $B \rightarrow X_s \gamma$  rate in bin  $i$ ,  $\Gamma_i$  is the integral of Eq. (11) over bin  $i$ ,  $V$  is the full experimental covariance matrix, and the sum runs over all bins of all measurements included in the fit.

The orthonormal basis  $\{f_n\}$  is constructed [10] such that the first  $F_{00}(k)$  term in the expansion of  $\mathcal{F}(k)$  can have any (nonnegative) functional form, while the higher  $F_{mn}(k)$  terms provide a complete expansion generated from it. If  $F_{00}(k)$  provides a good approximation to  $\mathcal{F}(k)$ , the expansion converges very quickly due to the constraint in Eq. (7), and consequently a good fit can be obtained with small  $N$ , making the best use of the data

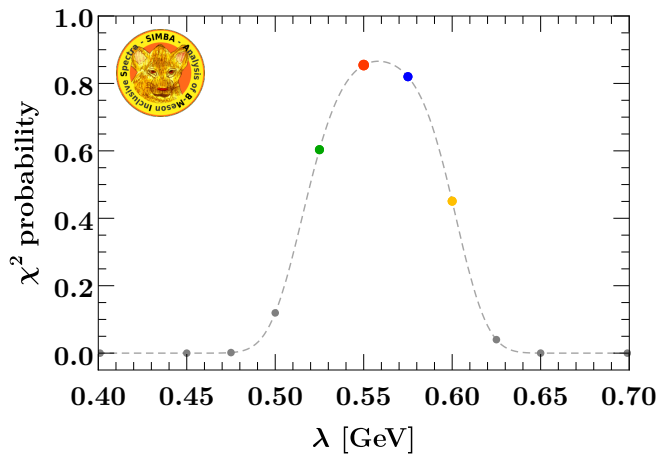


FIG. 1. The pre-fit  $\chi^2$  probability for different  $\lambda$  corresponding to different bases. See text for details.

to constrain  $\mathcal{F}(k)$ . Hence,  $F_{00}(k)$  should already provide a reasonable description of the data. To find such  $F_{00}(k)$ , we perform a pre-fit to the data using three different functional forms for  $F_{00}(k)$ , given in [28], over a wide range of  $\lambda$ . We choose the form that provides the best pre-fits. Its  $\chi^2$  probability is shown in Fig. 1 for sufficiently different values of  $\lambda$  such that each can be considered as a different basis. We choose the best  $\lambda = 0.55$  GeV (orange) as our default basis, and use  $\lambda = 0.525, 0.575, 0.6$  GeV (green, blue, yellow), which also have good pre-fits, as alternative bases to test the basis independence.

The truncation in Eq. (8) induces a residual dependence on the functional form of the basis. To ensure that the corresponding uncertainty is small compared to others, the truncation order  $N$  is chosen based on the available data, by increasing  $N$  until there is no significant improvement in fit quality. This is done by constructing nested hypothesis tests using the difference in  $\chi^2$  between fits of increasing number of coefficients. If the  $\chi^2$  improves by more than 1 from the inclusion of an additional coefficient, the higher number of coefficients is retained. To account for the truncation uncertainty, we include one additional coefficient in the fit. It is in this sense that our analysis is model independent within the quoted uncertainties. The final truncation order is found to be  $N = 3$  for each considered basis. To ensure that the entire fit procedure including the choice of the basis and truncation order is unbiased, it is validated using pseudo-experiments generated around the best fit values, using the full experimental covariance matrices.

*Results* We include four differential  $B \rightarrow X_s \gamma$  measurements [16–19] in the fit. The measurements in Ref. [16–18] include  $B \rightarrow X_d \gamma$  contributions, which are subtracted assuming identical shapes for  $B \rightarrow X_s \gamma$  and  $B \rightarrow X_d \gamma$  and that the ratio of branching ratios is  $|V_{td}/V_{ts}|^2 = 0.0470$  [52]. For Ref. [19], we combine the highest six  $E_\gamma$  bins to stay insensitive to possible quark-

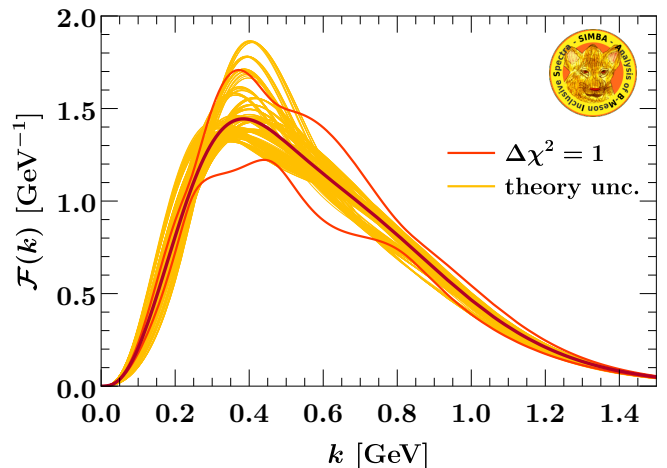


FIG. 2. The fitted shape function  $\mathcal{F}(k)$  with central result (dark red) and fit uncertainties (dark orange lines). The yellow curves show the variation of the fitted shape when varying the perturbative inputs as discussed in the text.

hadron duality violation and resonances with masses near  $m_{K^*}$ . We use the measurements of Refs. [17, 18] in the  $\Upsilon(4S)$  rest frame and boost the predictions accordingly. We use the uncorrected measurement from Ref. [17] and apply the experimental resolution matrix [53] to the predictions.

The fit results for  $N_s$  and  $c_{0-3}$  including their correlations are given in [28]. The resulting shape function is shown in Fig. 2, and the results for  $|C_7^{\text{incl}}|$  and  $\hat{m}_b \equiv m_b^{1S}$  are shown in Fig. 3. We also determine the kinetic energy parameter  $\hat{\lambda}_1$  in the invisible scheme [10], with plots analogous to Fig. 3 given in Fig. S2 in [28]. We find the following results:

$$\begin{aligned}
 |C_7^{\text{incl}} V_{tb} V_{ts}^*| &= (14.77 \pm 0.51_{\text{fit}} \pm 0.59_{\text{theory}} \\
 &\quad \pm 0.08_{\text{param}}) \times 10^{-3}, \\
 m_b^{1S} &= (4.750 \pm 0.027_{\text{fit}} \pm 0.033_{\text{theory}} \\
 &\quad \pm 0.003_{\text{param}}) \text{ GeV}, \\
 \hat{\lambda}_1 &= (-0.210 \pm 0.046_{\text{fit}} \pm 0.040_{\text{theory}} \\
 &\quad \pm 0.056_{\text{param}}) \text{ GeV}^2. \quad (13)
 \end{aligned}$$

The first uncertainty with subscript “fit” is evaluated from the  $\Delta\chi^2 = 1$  variation around the best fit point. It incorporates the experimental uncertainties as well as the uncertainty due to the unknown shape function, which is simultaneously constrained in the fit. The theory and parametric uncertainties are evaluated by repeating the fit with different theory inputs [28]. The theory uncertainties are due to unknown higher-order perturbative corrections to the shape of the spectrum in the peak region, which are evaluated by a large set of resummation profile scale variations. The results for all variations are shown by the yellow lines in Fig. 2 and scatter points in Fig. 3. To be conservative, the theory uncertainty quoted

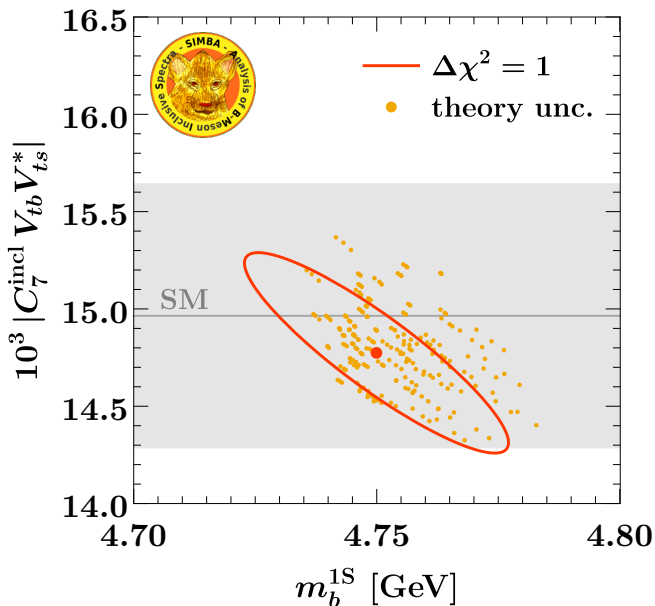


FIG. 3. Results for  $|C_7^{\text{incl}} V_{tb} V_{ts}^*|$  and  $m_b^{1S}$ . The central fit result is shown by the dark orange point and ellipse. The yellow scattered points show the variation of the fit results when varying the perturbative inputs as discussed in the text.

in Eq. (13) is obtained from the largest absolute deviation for a given quantity (ignoring the apparent asymmetry in the variations). The parametric uncertainty is only relevant for  $\hat{\lambda}_1$ , for which it comes entirely from  $\hat{\rho}_2$ .

Varying the residual  $c\bar{c}$ -loop contributions in the theory inputs for the fit, equivalent to the  $c\bar{c}$  uncertainty in Eq. (5), changes the extracted  $|C_7^{\text{incl}}|$  by  $\pm 0.2\%$  and  $m_b^{1S}$  by  $\pm 1$  MeV, showing that by far the dominant dependence on and uncertainty from these contributions is factorized into  $C_7^{\text{incl}}$ . The uncertainty due to the numerical value of  $\hat{m}_c^2/\hat{m}_b^2$  contributes most of the parametric uncertainty of  $|C_7^{\text{incl}}|$  in Eq. (13).

From Eq. (5) and  $|V_{tb} V_{ts}^*| = (41.29 \pm 0.74) \times 10^{-3}$  [52], we find the SM value  $|C_7^{\text{incl}} V_{tb} V_{ts}^*| = (14.96 \pm 0.68) \times 10^{-3}$ , with the uncertainty dominated by  $|C_7^{\text{incl}}|$  in Eq. (5). This is shown by the gray band in Fig. 3, and is in excellent agreement with our extracted value.

Converting our result for  $m_b^{1S}$  to the  $\overline{\text{MS}}$  scheme at three loops including charm-mass effects [54], we find

$$\overline{m}_b(\overline{m}_b) = (4.224 \pm 0.040 \pm 0.013) \text{ GeV}, \quad (14)$$

where the first uncertainty comes from the total uncertainty in  $m_b^{1S}$  in Eq. (13), and the second one is the conversion uncertainty. This result agrees with the world average of  $\overline{m}_b(\overline{m}_b) = (4.18_{-0.02}^{+0.03}) \text{ GeV}$  [52].

In Fig. 4, we demonstrate the basis independence by comparing the results for  $|C_7^{\text{incl}}|$  and  $m_b^{1S}$  for the four basis choices in Fig. 1. The results using these bases are consistent within a fraction of the fit uncertainties. This would not be the case without including an additional

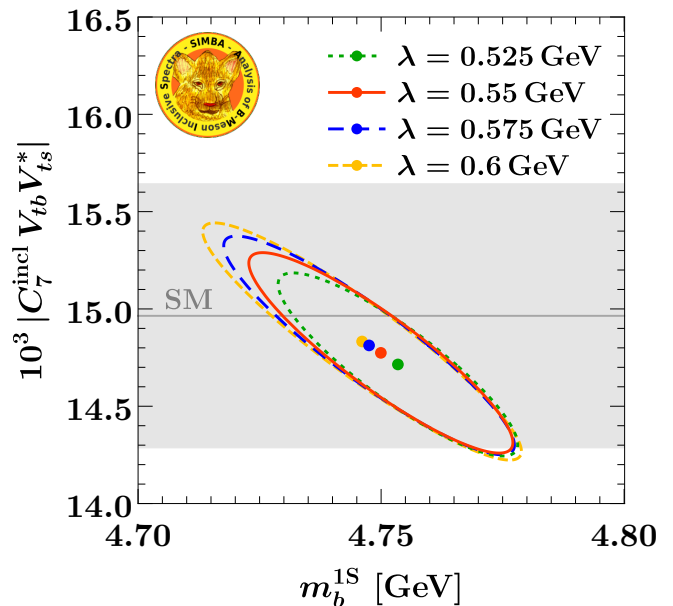


FIG. 4. Comparison of the fit results for  $|C_7^{\text{incl}} V_{tb} V_{ts}^*|$  and  $m_b^{1S}$  for four different bases. The results are consistent within a fraction of the fit uncertainties.

coefficient ( $c_3$ ) to account for the truncation uncertainty.

*Conclusions* We presented the first global analysis of inclusive  $B \rightarrow X_s \gamma$  measurements to determine  $|C_7^{\text{incl}}|$  within a framework that allows a model-independent and data-driven treatment of the nonperturbative  $b$ -quark distribution function  $\mathcal{F}(k)$ . The value extracted from Eq. (13),  $|C_7^{\text{incl}}| = 0.3578 \pm 0.0199$ , is consistent with the SM prediction in Eq. (5).

In comparison, in the past, the SM prediction for the rate in the  $E_\gamma > 1.6$  GeV region,  $\mathcal{B}(B \rightarrow X_s \gamma) = (3.36 \pm 0.23) \times 10^{-4}$  [4], was compared with its measurement,  $\mathcal{B}(B \rightarrow X_s \gamma) = (3.32 \pm 0.15) \times 10^{-4}$  [20], which have 6.8% and 4.5% uncertainties, respectively. The latter relies on an extrapolation to the 1.6 GeV cut and on corresponding uncertainty estimates, which entail insufficient variations of the nonperturbative shape-function models and perturbative uncertainties that affect the spectrum. In addition, correlations in these uncertainties in calculating and measuring the rate for  $E_\gamma > 1.6$  GeV cannot be fully assessed. In contrast, in our approach,  $C_7^{\text{incl}}$  is reliably calculable in the SM or in models beyond it, and the relevant hadronic physics and its uncertainties are determined from the data, together with the extraction of  $|C_7^{\text{incl}}|$ . Hence, our approach is more reliable, as it makes optimal use of the data, uncertainties from nonperturbative parameters and perturbative inputs are clearly traceable, and no double counting can occur.

The uncertainty in our extracted  $|C_7^{\text{incl}} V_{tb} V_{ts}^*|^2$  from Eq. (13) is 10.6%, about twice the uncertainty in HFLAV's result for the  $E_\gamma > 1.6$  GeV rate. If we neglect the theory uncertainties as well as the truncation uncer-

tainty (by repeating the fit only including up to  $c_2$ ), we would obtain a smaller uncertainty of 5.5%, close to that of HFLAV's result. This suggests that HFLAV's uncertainty is underestimated by about a factor of two, which leaves more room for new physics. More importantly, the precision of testing the SM is currently limited by the extraction of  $|C_7^{\text{incl}}|$  from data, and can be improved significantly with high-precision Belle II measurements.

*Acknowledgments* We thank Antonio Limosani for information about the detector resolution in the Belle measurement, and Francesca di Lodovico for information about the correlations in the *BABAR* inclusive measurement. We thank Anna Sophia Lacker for the artwork for the SIMBA logo. We thank the DESY and LBL theory groups, KIT, and the Aspen Center for Physics (supported by the NSF Grant PHY-1607611) for hospitality while portions of this work were carried out. This work was also supported in part by the Offices of High Energy and Nuclear Physics of the U.S. Department of Energy under DE-AC02-05CH11231 and DE-SC0011090, the Simons Foundation through grant 327942, the DFG Emmy-Noether Grants TA 867/1-1 and BE 6075/1-1, and the Helmholtz Association Grant W2/W3-116.

- 
- [1] S. Bertolini, F. Borzumati, and A. Masiero, *Phys. Rev. Lett.* **59**, 180 (1987).
- [2] B. Grinstein, R. P. Springer, and M. B. Wise, *Phys. Lett. B* **202**, 138 (1988).
- [3] M. Misiak, H. Asatrian, K. Bieri, M. Czakon, A. Czarnecki, *et al.*, *Phys. Rev. Lett.* **98**, 022002 (2007), hep-ph/0609232.
- [4] M. Misiak *et al.*, *Phys. Rev. Lett.* **114**, 221801 (2015), arXiv:1503.01789 [hep-ph].
- [5] B. Grinstein and M. B. Wise, *Phys. Lett. B* **201**, 274 (1988).
- [6] W.-S. Hou and R. S. Willey, *Phys. Lett. B* **202**, 591 (1988).
- [7] M. Misiak and M. Steinhauser, *Eur. Phys. J. C* **77**, 201 (2017), arXiv:1702.04571 [hep-ph].
- [8] M. Neubert, *Phys. Rev. D* **49**, 4623 (1994), hep-ph/9312311.
- [9] I. I. Y. Bigi, M. A. Shifman, N. G. Uraltsev, and A. I. Vainshtein, *Int. J. Mod. Phys. A* **9**, 2467 (1994), hep-ph/9312359.
- [10] Z. Ligeti, I. W. Stewart, and F. J. Tackmann, *Phys. Rev. D* **78**, 114014 (2008), arXiv:0807.1926 [hep-ph].
- [11] D. Benson, I. I. Bigi, and N. Uraltsev, *Nucl. Phys. B* **710**, 371 (2005), hep-ph/0410080.
- [12] B. O. Lange, M. Neubert, and G. Paz, *Phys. Rev. D* **72**, 073006 (2005), hep-ph/0504071.
- [13] J. R. Andersen and E. Gardi, *JHEP* **01**, 097 (2006), hep-ph/0509360.
- [14] P. Gambino, P. Giordano, G. Ossola, and N. Uraltsev, *JHEP* **10**, 058 (2007), arXiv:0707.2493 [hep-ph].
- [15] U. Aglietti, F. Di Lodovico, G. Ferrera, and G. Ricciardi, *Eur. Phys. J. C* **59**, 831 (2009), arXiv:0711.0860 [hep-ph].
- [16] B. Aubert *et al.* (BaBar), *Phys. Rev. D* **77**, 051103 (2008), arXiv:0711.4889 [hep-ex].
- [17] A. Limosani *et al.* (Belle), *Phys. Rev. Lett.* **103**, 241801 (2009), arXiv:0907.1384 [hep-ex].
- [18] J. P. Lees *et al.* (BaBar), *Phys. Rev. D* **86**, 112008 (2012), arXiv:1207.5772 [hep-ex].
- [19] J. Lees *et al.* (BaBar Collaboration), *Phys. Rev. D* **86**, 052012 (2012), arXiv:1207.2520 [hep-ex].
- [20] Y. S. Amhis *et al.* (HFLAV), (2019), arXiv:1909.12524 [hep-ex].
- [21] C. W. Bauer, S. Fleming, and M. E. Luke, *Phys. Rev. D* **63**, 014006 (2000), hep-ph/0005275.
- [22] C. W. Bauer, S. Fleming, D. Pirjol, and I. W. Stewart, *Phys. Rev. D* **63**, 114020 (2001), hep-ph/0011336.
- [23] C. W. Bauer and I. W. Stewart, *Phys. Lett. B* **516**, 134 (2001), hep-ph/0107001.
- [24] C. W. Bauer, D. Pirjol, and I. W. Stewart, *Phys. Rev. D* **65**, 054022 (2002), hep-ph/0109045.
- [25] A. H. Hoang, Z. Ligeti, and A. V. Manohar, *Phys. Rev. Lett.* **82**, 277 (1999), hep-ph/9809423.
- [26] A. H. Hoang, Z. Ligeti, and A. V. Manohar, *Phys. Rev. D* **59**, 074017 (1999), hep-ph/9811239.
- [27] A. Hoang and T. Teubner, *Phys. Rev. D* **60**, 114027 (1999), hep-ph/9904468.
- [28] See the supplemental material at the end of the paper.
- [29] G. P. Korchemsky and G. Marchesini, *Nucl. Phys. B* **406**, 225 (1993), hep-ph/9210281.
- [30] E. Gardi, *JHEP* **02**, 053 (2005), hep-ph/0501257.
- [31] I. R. Blokland, A. Czarnecki, M. Misiak, M. Slusarczyk, and F. Tkachov, *Phys. Rev. D* **72**, 033014 (2005), hep-ph/0506055.
- [32] C. W. Bauer and A. V. Manohar, *Phys. Rev. D* **70**, 034024 (2004), hep-ph/0312109.
- [33] T. Becher and M. Neubert, *Phys. Lett. B* **633**, 739 (2006), hep-ph/0512208.
- [34] T. Becher and M. Neubert, *Phys. Lett. B* **637**, 251 (2006), hep-ph/0603140.
- [35] C. Balzereit, T. Mannel, and W. Kilian, *Phys. Rev. D* **58**, 114029 (1998), hep-ph/9805297.
- [36] M. Neubert, *Eur. Phys. J. C* **40**, 165 (2005), hep-ph/0408179.
- [37] S. Fleming, A. H. Hoang, S. Mantry, and I. W. Stewart, *Phys. Rev. D* **77**, 114003 (2008), arXiv:0711.2079 [hep-ph].
- [38] M. Misiak and M. Steinhauser, *Nucl. Phys. B* **764**, 62 (2007), hep-ph/0609241.
- [39] M. Czakon, P. Fiedler, T. Huber, M. Misiak, T. Schutzmeier, and M. Steinhauser, *JHEP* **04**, 168 (2015), arXiv:1503.01791 [hep-ph].
- [40] K. Melnikov and A. Mitov, *Phys. Lett. B* **620**, 69 (2005), hep-ph/0505097.
- [41] T. Ewerth, *Phys. Lett. B* **669**, 167 (2008), arXiv:0805.3911 [hep-ph].
- [42] H. Asatrian, T. Ewerth, A. Ferroglia, C. Greub, and G. Ossola, *Phys. Rev. D* **82**, 074006 (2010), arXiv:1005.5587 [hep-ph].
- [43] Z. Ligeti, M. E. Luke, A. V. Manohar, and M. B. Wise, *Phys. Rev. D* **60**, 034019 (1999), hep-ph/9903305.
- [44] A. Ferroglia and U. Haisch, *Phys. Rev. D* **82**, 094012 (2010), arXiv:1009.2144 [hep-ph].
- [45] M. Misiak and M. Poradzinski, *Phys. Rev. D* **83**, 014024 (2011), arXiv:1009.5685 [hep-ph].
- [46] K. S. M. Lee and I. W. Stewart, *Nucl. Phys. B* **721**, 325 (2005), hep-ph/0409045.

- [47] M. Benzke, S. J. Lee, M. Neubert, and G. Paz, JHEP **08**, 099 (2010), arXiv:1003.5012 [hep-ph].
- [48] A. Gunawardana and G. Paz, JHEP **11**, 141 (2019), arXiv:1908.02812 [hep-ph].
- [49] M. Voloshin, Phys. Lett. B **397**, 275 (1997), hep-ph/9612483.
- [50] Z. Ligeti, L. Randall, and M. B. Wise, Phys. Lett. B **402**, 178 (1997), hep-ph/9702322.
- [51] A. K. Grant, A. Morgan, S. Nussinov, and R. Peccei, Phys. Rev. D **56**, 3151 (1997), hep-ph/9702380.
- [52] M. Tanabashi *et al.* (Particle Data Group), Phys. Rev. D **98**, 030001 (2018), and 2019 update.
- [53] A. Limosani (Belle Collaboration), private communication.
- [54] A. Hoang, (2000), hep-ph/0008102.
- [55] K. S. Lee and I. W. Stewart, Phys. Rev. D **74**, 014005 (2006), hep-ph/0511334.
- [56] K. S. Lee, Z. Ligeti, I. W. Stewart, and F. J. Tackmann, Phys. Rev. D **75**, 034016 (2007), hep-ph/0612156.
- [57] A. Kapustin and Z. Ligeti, Phys. Lett. B **355**, 318 (1995), hep-ph/9506201.
- [58] G. Buchalla, A. J. Buras, and M. E. Lautenbacher, Rev. Mod. Phys. **68**, 1125 (1996), hep-ph/9512380.
- [59] M. Czakon, U. Haisch, and M. Misiak, JHEP **03**, 008 (2007), hep-ph/0612329.
- [60] C. Greub, T. Hurth, and D. Wyler, Phys. Rev. D **54**, 3350 (1996), hep-ph/9603404.
- [61] A. J. Buras, A. Czarnecki, M. Misiak, and J. Urban, Nucl. Phys. **B611**, 488 (2001), hep-ph/0105160.
- [62] A. J. Buras, A. Czarnecki, M. Misiak, and J. Urban, Nucl. Phys. **B631**, 219 (2002), hep-ph/0203135.
- [63] K. Bieri, C. Greub, and M. Steinhauser, Phys. Rev. D **67**, 114019 (2003), hep-ph/0302051.
- [64] M. Misiak and M. Steinhauser, Nucl. Phys. **B840**, 271 (2010), arXiv:1005.1173 [hep-ph].
- [65] G. P. Korchemsky and G. F. Sterman, Phys. Lett. B **340**, 96 (1994), hep-ph/9407344.
- [66] H. Asatrian, T. Ewerth, H. Gabrielyan, and C. Greub, Phys. Lett. B **647**, 173 (2007), hep-ph/0611123.
- [67] H. M. Asatrian, T. Ewerth, A. Ferroglia, P. Gambino, and C. Greub, Nucl. Phys. **B762**, 212 (2007), hep-ph/0607316.
- [68] M. Misiak, in *Heavy Quarks and Leptons 2008 (HQ&L08)* (2008) arXiv:0808.3134 [hep-ph].
- [69] A. Ali and C. Greub, Z. Phys. C **49**, 431 (1991).
- [70] A. Ali and C. Greub, Phys. Lett. B **259**, 182 (1991).
- [71] A. Ali and C. Greub, Phys. Lett. **B361**, 146 (1995), hep-ph/9506374.
- [72] N. Pott, Phys. Rev. **D54**, 938 (1996), hep-ph/9512252.
- [73] R. Abbate, M. Fickinger, A. H. Hoang, V. Mateu, and I. W. Stewart, Phys. Rev. **D83**, 074021 (2011), arXiv:1006.3080 [hep-ph].
- [74] F. J. Tackmann, Phys. Rev. D **72**, 034036 (2005), hep-ph/0503095.
- [75] M. Gremm and A. Kapustin, Phys. Rev. D **55**, 6924 (1997), hep-ph/9603448.
- [76] C. W. Bauer, Phys. Rev. **D57**, 5611 (1998), [Erratum: Phys. Rev. **D60**, 099907 (1999)], hep-ph/9710513.
- [77] A. Kapustin, Z. Ligeti, and H. D. Politzer, Phys. Lett. B **357**, 653 (1995), hep-ph/9507248.
- [78] S. J. Lee, M. Neubert, and G. Paz, Phys. Rev. D **75**, 114005 (2007), hep-ph/0609224.
- [79] M. Misiak, Acta Phys. Polon. **B40**, 2987 (2009), arXiv:0911.1651 [hep-ph].
- [80] S. Watanuki *et al.* (Belle), Phys. Rev. **D99**, 032012 (2019), arXiv:1807.04236 [hep-ex].
- [81] C. Bobeth, M. Misiak, and J. Urban, Nucl. Phys. **B574**, 291 (2000), hep-ph/9910220.
- [82] M. Misiak and M. Steinhauser, Nucl. Phys. **B683**, 277 (2004), hep-ph/0401041.
- [83] A. J. Buras, M. Jamin, M. E. Lautenbacher, and P. H. Weisz, Nucl. Phys. **B400**, 37 (1993), hep-ph/9211304.
- [84] M. Ciuchini, E. Franco, G. Martinelli, and L. Reina, Nucl. Phys. **B415**, 403 (1994), hep-ph/9304257.
- [85] K. G. Chetyrkin, M. Misiak, and M. Münz, Phys. Lett. B **400**, 206 (1997), [Erratum: Phys. Lett. **B425**, 414 (1998)], hep-ph/9612313.
- [86] P. Gambino, M. Gorbahn, and U. Haisch, Nucl. Phys. **B673**, 238 (2003), hep-ph/0306079.
- [87] M. Gorbahn and U. Haisch, Nucl. Phys. **B713**, 291 (2005), hep-ph/0411071.
- [88] M. Gorbahn, U. Haisch, and M. Misiak, Phys. Rev. Lett. **95**, 102004 (2005), hep-ph/0504194.
- [89] K. G. Chetyrkin, J. H. Kuhn, and M. Steinhauser, Comput. Phys. Commun. **133**, 43 (2000), hep-ph/0004189.
- [90] C. W. Bauer, Z. Ligeti, M. Luke, and A. V. Manohar, Phys. Rev. D **67**, 054012 (2003), hep-ph/0210027.
- [91] P. Urquijo, private communication.
- [92] A. H. Hoang, A. Jain, I. Scimemi, and I. W. Stewart, Phys. Rev. **D82**, 011501 (2010), arXiv:0908.3189 [hep-ph].
- [93] A. G. Grozin, P. Marquard, J. H. Piclum, and M. Steinhauser, Nucl. Phys. **B789**, 277 (2008), arXiv:0707.1388 [hep-ph].
- [94] A. H. Hoang, A. Jain, I. Scimemi, and I. W. Stewart, Phys. Rev. Lett. **101**, 151602 (2008), arXiv:0803.4214 [hep-ph].



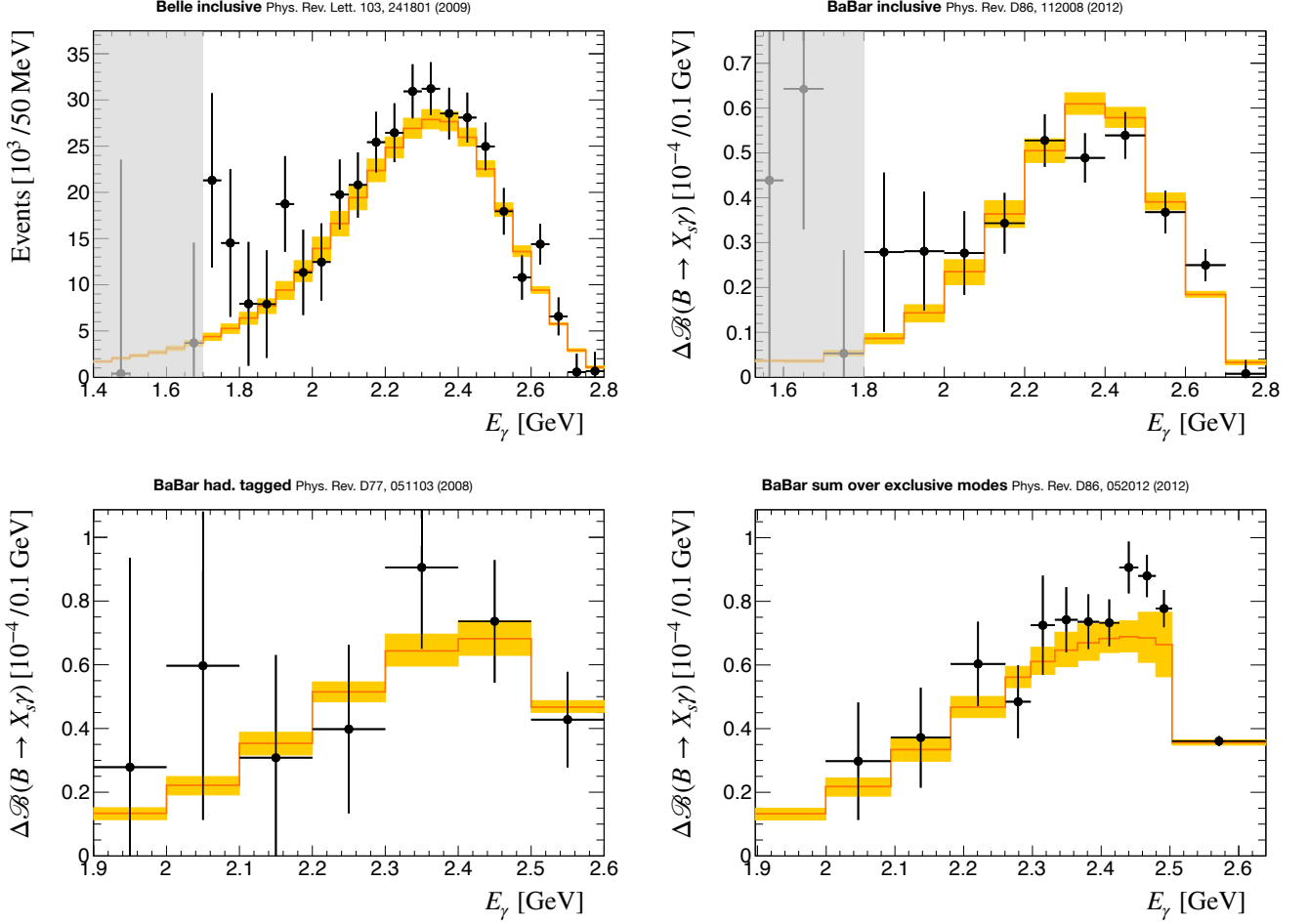


FIG. S1. Fit results to the measured photon energy spectra [16–19]. The orange lines are the fitted central values, and the yellow bands correspond to the  $\Delta\chi^2 = 1$  variation. We omit the first 6 bins of the Belle inclusive spectrum and the first 3 bins of the *BaBar* inclusive spectrum, as these have very large uncertainties and provide no additional information.

## SUPPLEMENTAL MATERIAL

### A. Additional fit results

The full expression of Eq. (11) used in the fit including the non-77 terms is given by

$$\begin{aligned}
 \frac{d\Gamma}{dE_\gamma} &= N_s \sum_{m,n=0}^N c_m c_n \frac{d\Gamma_{77,mn}}{dE_\gamma} + \sqrt{N_s} \sum_{ij=27,78} N_{ij} \sum_{m,n=0}^N c_m c_n \frac{d\Gamma_{ij,mn}}{dE_\gamma} + \sum_{ij=22,28,88} N_{ij} \sum_{m,n=0}^N c_m c_n \frac{d\Gamma_{ij,mn}}{dE_\gamma} \\
 &\quad + \sqrt{N_s} N_{27} \frac{\hat{\lambda}_2}{\hat{m}_b^2} \sum_{n=0}^2 d_n \frac{d\Gamma_{g27,n}}{dE_\gamma}, \\
 N_{27} &= -2 \left( c_2 - \frac{c_1}{6} \right) |V_{tb} V_{ts}^*| \hat{m}_b, \quad N_{78} = -2 c_8 |V_{tb} V_{ts}^*| \hat{m}_b, \\
 N_{22} &= \left( c_2 - \frac{c_1}{6} \right)^2 |V_{tb} V_{ts}^*|^2 \hat{m}_b^2, \quad N_{28} = 2 \left( c_2 - \frac{c_1}{6} \right) c_8 |V_{tb} V_{ts}^*|^2 \hat{m}_b^2, \quad N_{88} = c_8^2 |V_{tb} V_{ts}^*|^2 \hat{m}_b^2, \quad (S1)
 \end{aligned}$$

where the normalization  $N_s$  is defined by

$$N_s = |C_7^{\text{incl}} V_{tb} V_{ts}^*|^2 \hat{m}_b^2. \quad (S2)$$

Parameter	Fit result		$N_s$	$c_0$	$c_1$	$c_2$	$c_3$
$10^3 N_s$	$4.925 \pm 0.294$	$N_s$	1	-0.804332	+0.809278	+0.703457	+0.579938
$c_0$	$0.9956 \pm 0.0063$	$c_0$	-0.804332	1	-0.860744	-0.980474	-0.738548
$c_1$	$0.0641 \pm 0.0361$	$c_1$	+0.809278	-0.860744	1	+0.844262	+0.325844
$c_2$	$0.0624 \pm 0.0458$	$c_2$	+0.703457	-0.980474	+0.844262	1	+0.666741
$c_3$	$0.0267 \pm 0.0727$	$c_3$	+0.579938	-0.738548	+0.325844	+0.666741	1

TABLE S1. Fit results (left) and correlations (right) for the normalization parameter  $N_s$  and the fitted shape-function coefficients  $c_{0,1,2,3}$  for the default fit with  $\lambda = 0.55$  GeV and  $N = 3$ .

Parameter	Fit result		$ C_7^{\text{incl}} V_{tb} V_{ts}^* $	$m_b^{1S}$	$\hat{\lambda}_1$	$\hat{\rho}_1$
$10^3  C_7^{\text{incl}} V_{tb} V_{ts}^* $	$14.77 \pm 0.51$	$ C_7^{\text{incl}} V_{tb} V_{ts}^* $	1	-0.895754	-0.788116	+0.685843
$m_b^{1S}/\text{GeV}$	$4.750 \pm 0.027$	$m_b^{1S}$	-0.895754	1	+0.917563	-0.770155
$\hat{\lambda}_1/\text{GeV}^2$	$-0.210 \pm 0.046$	$\hat{\lambda}_1$	-0.788116	+0.917563	1	-0.953347
$\hat{\rho}_1/\text{GeV}^3$	$0.134 \pm 0.036$	$\hat{\rho}_1$	+0.685843	-0.770155	-0.953347	1

TABLE S2. Results (left) and correlations (right) for  $|C_7^{\text{incl}} V_{tb} V_{ts}^*|$ ,  $m_b^{1S}$ ,  $\hat{\lambda}_1$ ,  $\hat{\rho}_1$  obtained from the default fit results in Table S1 for  $\lambda = 0.55$  GeV and  $N = 3$  by inverting the moment relations for  $\mathcal{F}(k)$ . Only the fit uncertainties are included.

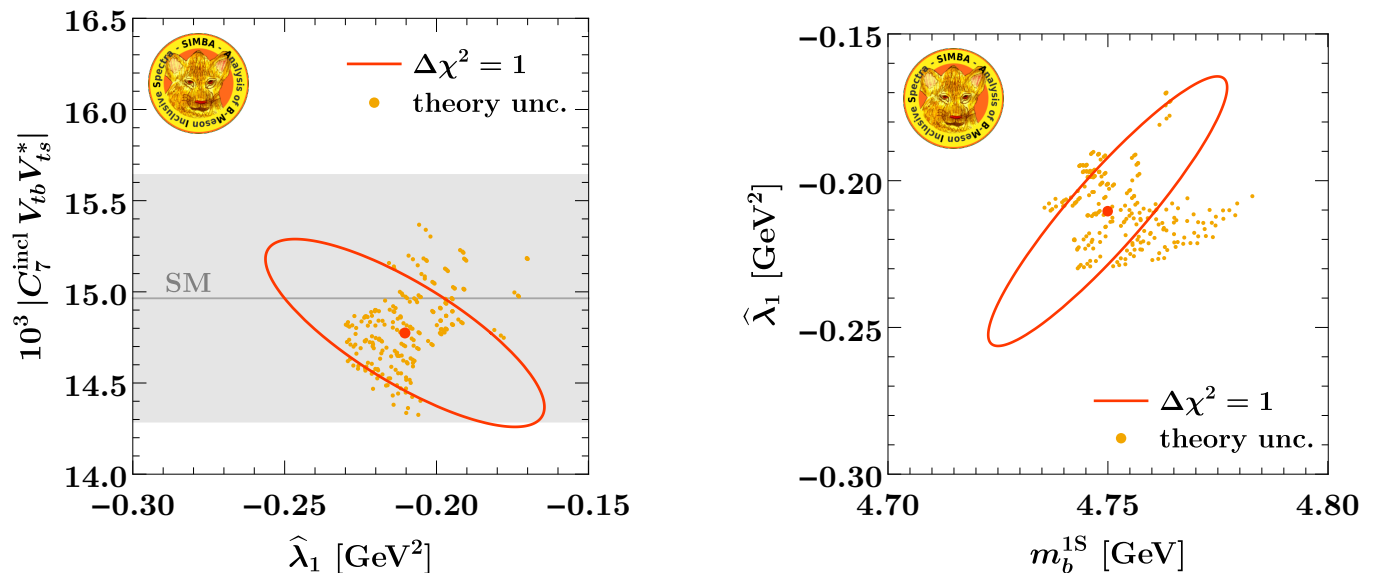


FIG. S2. Fit results in the projection of  $|C_7^{\text{incl}} V_{tb} V_{ts}^*|$  vs.  $\hat{\lambda}_1$  (left) and  $\hat{\lambda}_1$  vs.  $m_b^{1S}$  (right), analogous to Fig. 3 in the main text. The orange ellipse shows the  $\Delta\chi^2 = 1$  contour. The yellow points show fit results from varying the perturbative inputs.

The  $d\Gamma_{ij,mn}$  and  $d\Gamma_{g27,n}$  are precomputed from the basis expansion of the shape function. The  $c_n$  and the normalization  $N_s$  are determined from the fit. For the normalization prefactors of the remaining non-77 nonsingular terms in Eq. (S1) we use the SM input values collected in Sec. E. The overall minus sign in  $N_{27}$  and  $N_{78}$  arises from assuming the SM negative sign for  $\text{Re}(C_7^{\text{incl}}) = -|C_7^{\text{incl}}|$ , and assuming the SM imaginary part of  $C_7^{\text{incl}}$ , which is negligible. The value for  $\hat{m}_b$  in the prefactors is obtained during the fit from the  $c_n$  as discussed in Sec. D 2. The coefficients  $d_n$  parametrize the  $g_{27}$  subleading shape function that cannot be absorbed into the leading shape function, cf. Sec. D 3.

The fitted experimental spectra with the fit results overlaid are shown in Fig. S1. The central value is shown by the orange line, and the yellow band corresponds to the  $\Delta\chi^2 = 1$  uncertainties of the fit. The fit results for  $N_s$  and  $c_n$  and their correlation matrix are given in Table S1. The final results for  $|C_7^{\text{incl}}|$ ,  $\hat{m}_b \equiv m_b^{1S}$ ,  $\hat{\lambda}_1$ , and  $\hat{\rho}_1$  together with their correlation matrix are given in Table S2. They are obtained from the fitted  $N_s$  and  $c_n$  by using Eq. (S2) and the moment relations discussed in Sec. D 2. In Fig. S2 these results and the corresponding theory uncertainties are shown as well, analogous to Fig. 3 in the main text.

In Fig. S3 the convergence of the fit results for our default basis with  $\lambda = 0.55$  GeV for an increasing number of

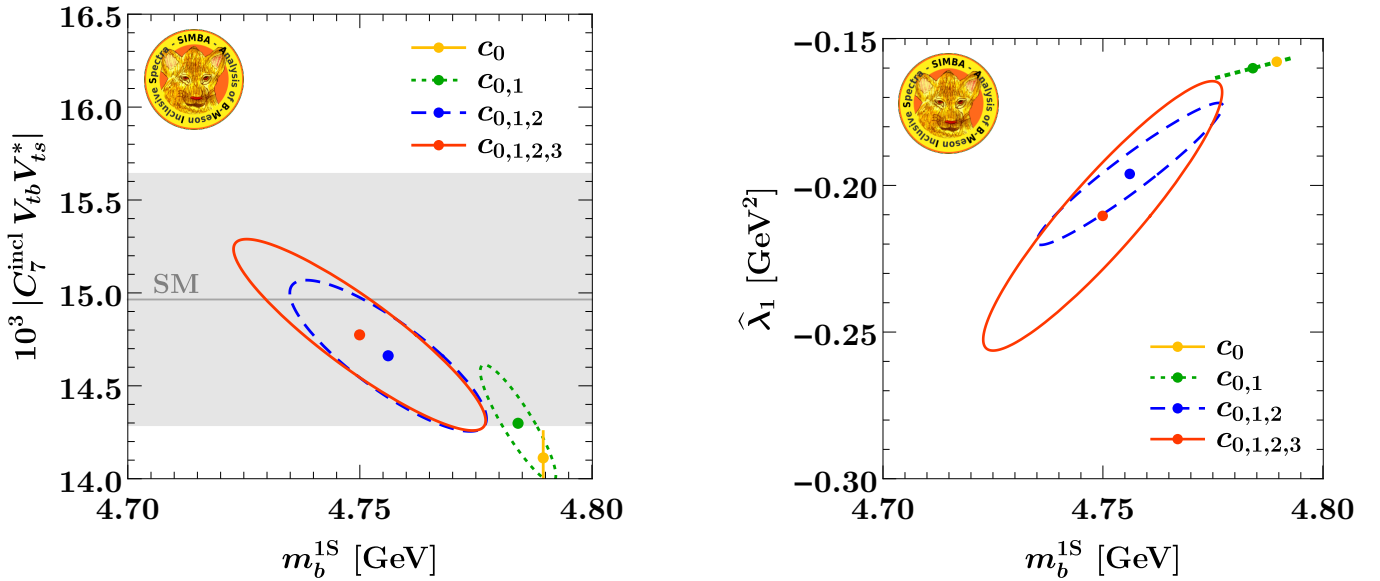


FIG. S3. Fit results as a function of the number of fitted basis coefficients in the projection of  $|C_7^{\text{incl}} V_{tb} V_{ts}^*|$  vs.  $m_b^{1S}$  (left) and  $\hat{\lambda}_1$  vs.  $m_b^{1S}$  (right). For more details see text.

basis coefficients is shown. As discussed in the main text, the truncation order is determined using a nested hypothesis test to determine the appropriate number of coefficients given the available data sets. For the nominal fit we use 4 coefficients ( $c_{0,1,2,3}$ ). Note that all fits with fewer coefficients also have acceptable  $\chi^2$ , so the fit quality alone is not a sufficient criterion for choosing the number of coefficients. On the other hand, the results with only  $c_0$  and  $c_{0,1}$ , which effectively correspond to using a fixed model for the shape function, clearly show a model bias and underestimated uncertainties. The central values change only moderately by the inclusion of the fourth coefficient  $c_3$ . The resulting increase in the fit uncertainties illustrates the effect of accounting for the truncation uncertainty by including this additional coefficient. Including  $c_3$  is essential for the results with different basis choices to be consistent as in Fig. 4. Without including  $c_3$ , the results still show a clear bias between different bases.

## B. Wilson coefficients and $C_7^{\text{incl}}$

### 1. Split matching

The effective Hamiltonian for  $B \rightarrow X_s \gamma$  is

$$\mathcal{H}_{\text{eff}} = -\frac{4G_F}{\sqrt{2}} V_{tb} V_{ts}^* \sum_{i=1}^8 C_i O_i. \quad (\text{S3})$$

The dominant contributions are from

$$\begin{aligned} O_1 &= (\bar{s} \gamma_\mu T^a P_L c) (\bar{c} \gamma^\mu T^a P_L b), & O_2 &= (\bar{s} \gamma_\mu P_L c) (\bar{c} \gamma^\mu P_L b), \\ O_7 &= \frac{e}{16\pi^2} \bar{m}_b \bar{s} \sigma_{\mu\nu} F^{\mu\nu} P_R b, & O_8 &= \frac{g}{16\pi^2} \bar{m}_b \bar{s} \sigma_{\mu\nu} G^{\mu\nu} P_R b, \end{aligned} \quad (\text{S4})$$

where  $P_{R,L} = (1 \pm \gamma_5)/2$ , and we neglected the mass of the strange quark (giving  $m_s^2/m_b^2$  suppressed corrections). The  $O_{3-6}$  in Eq. (S3) are four-quark operators generated at one-loop level in the SM. The renormalized Wilson coefficients  $C_i(\mu)$  and operators  $O_i(\mu)$  are defined in the  $\overline{\text{MS}}$  scheme, and  $\bar{m}_b(\mu)$  is the  $\overline{\text{MS}}$   $b$ -quark mass.

The Wilson coefficient  $C_7^{\text{incl}}$  arises when we carry out a split matching procedure to separate the scale dependence above and below the scale  $\mu_0 \sim m_b$  [55, 56]. Above  $\mu_0$ , we have the matching onto  $\mathcal{H}_{\text{eff}}$  at the weak scale  $\mu_{\text{weak}} \sim m_W$  and its renormalization group evolution down to  $\mu_0 \sim m_b$ . At  $\mu_0$ , we have virtual matrix element corrections from all operators  $O_{i \neq 7}$  that are proportional to the tree-level matrix element of the chromomagnetic operator  $O_7$ . Together,

these effects can be combined to define the effective Wilson coefficient

$$C_7^{\text{incl}} = \bar{C}_7(\mu_0) + \sum_i \bar{C}_i(\mu_0) [s_i(\mu_0, \hat{m}_b) + r_i(\mu_0, \hat{m}_b, \hat{m}_c)] = \mathcal{C}_7 + \sum_{i \neq 7} \bar{C}_i(\mu_0) r_i(\mu_0, \hat{m}_b, \hat{m}_c), \quad (\text{S5})$$

which is the main short-distance perturbative coefficient that is constrained by the  $B \rightarrow X_s \gamma$  measurements. Its value is sensitive to beyond Standard-Model physics, while the shape of the photon spectrum is not [57]. The two terms in the last equality in Eq. (S5) are separately  $\mu_0$  independent order by order in  $\alpha_s$ . The barred coefficients  $\bar{C}_i(\mu)$  are defined as

$$\bar{C}_i(\mu_0) = \begin{cases} C_i(\mu_0), & i = 1, \dots, 6, \\ C_i^{\text{eff}}(\mu_0) \frac{\bar{m}_b(\mu_0)}{\hat{m}_b}, & i = 7, 8, \end{cases} \quad (\text{S6})$$

where  $C_{7,8}^{\text{eff}}(\mu_0)$  are the standard scheme-independent effective Wilson coefficients [58]. The additional factors  $\bar{m}_b(\mu_0)/\hat{m}_b$  are included in  $\bar{C}_{7,8}$  to convert to a short-distance  $b$ -quark mass scheme,  $\hat{m}_b$ , which improves the convergence of perturbation theory.

The  $r_i(\mu_0, \hat{m}_b, \hat{m}_c)$  in Eq. (S5) encode the finite virtual corrections from operators other than  $O_7$  that give rise to singular contributions to the photon energy spectrum, and are responsible for the difference between  $\mathcal{C}_7$  and  $C_7^{\text{incl}}$ . Hence, the split matching procedure essentially amounts to matching  $\mathcal{H}_{\text{eff}}$  at  $\mu_0$  onto a single  $O_7$  chromomagnetic operator, which is subsequently matched onto its corresponding operator in SCET. In doing so, we treat the charm quark as a heavy quark and integrate out both charm and bottom quarks at the scale  $\mu_0$ . As a result, (most of) the sizable contributions from  $c\bar{c}$ -loops proportional to  $\bar{C}_{1,2}\bar{C}_7$  and  $\bar{C}_{1,2}^2$  are contained within  $|C_7^{\text{incl}}|^2$ , including their full  $\hat{m}_c$  dependence, namely in the terms  $r_{1,2}(\mu_0, \hat{m}_b, \hat{m}_c)$ . As already mentioned in the main body, this organization of the perturbative contributions has the advantage that the associated theory uncertainty due to the  $\hat{m}_c$  dependence only enters in the SM prediction for  $C_7^{\text{incl}}$ , but does not limit the accuracy with which  $C_7^{\text{incl}}$  can be extracted from the experimental data. This treatment is furthermore motivated by the fact that in the experimental measurements of  $B \rightarrow X_s \gamma$ , charmed final states are not included in the signal and are treated as background.

## 2. Perturbative results

The coefficient  $\mathcal{C}_7$  in Eq. (S5) is defined to be  $\mu_0$  independent and to satisfy  $\mathcal{C}_7 = \bar{C}_7(\hat{m}_b)$ . Thus it is equal to  $\bar{C}_7(\mu_0)$  plus the additional terms from the renormalization group that cancel the  $\mu_0$  dependence of  $\bar{C}_7(\mu_0)$  and vanish when  $\mu_0 = \hat{m}_b$ . Explicitly, up to  $\mathcal{O}(\alpha_s^2)$  with  $\hat{m}_b$  in the  $1S$  mass scheme, we have

$$\begin{aligned} \mathcal{C}_7 &= \bar{C}_7(\mu_0) + \sum_i \bar{C}_i(\mu_0) s_i(\mu_0, \hat{m}_b) \\ &= \bar{C}_7(\mu_0) + \frac{\alpha_s(\mu_0)}{4\pi} \ln \frac{\hat{m}_b}{\mu_0} \left[ \gamma_m^{(0)} \bar{C}_7(\mu_0) + \sum_i \gamma_{i7}^{(0)} \bar{C}_i(\mu_0) \right] \\ &\quad + \frac{\alpha_s^2(\mu_0)}{(4\pi)^2} \ln \frac{\hat{m}_b}{\mu_0} \left\{ \gamma_m^{(1)} \bar{C}_7(\mu_0) + \sum_i \gamma_{i7}^{(1)} \bar{C}_i(\mu_0) + \frac{1}{2} \ln \frac{\hat{m}_b}{\mu_0} \left[ (\gamma_m^{(0)} + \gamma_{77}^{(0)})^2 \bar{C}_7(\mu_0) + (\gamma_m^{(0)} + \gamma_{77}^{(0)}) \sum_{j \neq 7} \gamma_{j7}^{(0)} \bar{C}_j(\mu_0) \right. \right. \\ &\quad \left. \left. + \gamma_{87}^{(0)} \gamma_m^{(0)} \bar{C}_8(\mu_0) + \gamma_{87}^{(0)} \sum_{j \neq 7} \gamma_{j8}^{(0)} \bar{C}_j(\mu_0) + \sum_{j \neq 7, 8} \sum_k \gamma_{j7}^{(0)} \gamma_{kj}^{(0)} \bar{C}_k(\mu_0) \right] \right. \\ &\quad \left. - \sum_{j \neq 7, 8} 4C_F \gamma_{j7}^{(0)} \bar{C}_j(\mu_0) \left[ 1 - \frac{C_F \pi \alpha_s(\mu_0)}{8} - \frac{3}{4} \ln \frac{\hat{m}_b}{\mu_0} \right] - \beta_0 \ln \frac{\hat{m}_b}{\mu_0} \left[ \gamma_m^{(0)} \bar{C}_7(\mu_0) + \sum_i \gamma_{i7}^{(0)} \bar{C}_i(\mu_0) \right] \right\}, \quad (\text{S7}) \end{aligned}$$

where  $\beta_0 = (11C_A - 4T_F n_f)/3$  and  $n_f = 5$  is the number of active flavors above the scale  $m_b$ , and  $C_A = 3$ ,  $C_F = 4/3$ ,  $T_F = 1/2$ . The  $\gamma_{ij}^{(k)}$  and  $\gamma_m^{(k)}$  are anomalous dimension coefficients defined via

$$\begin{aligned} \mu \frac{d}{d\mu} C_j(\mu) &= \sum_i C_i(\mu) \gamma_{ij}(\mu), & \gamma_{ij}(\mu) &= \frac{\alpha_s(\mu)}{4\pi} \gamma_{ij}^{(0)} + \frac{\alpha_s^2(\mu)}{(4\pi)^2} \gamma_{ij}^{(1)} + \mathcal{O}(\alpha_s^3), \\ \mu \frac{d}{d\mu} \bar{m}_b(\mu) &= \bar{m}_b(\mu) \gamma_m(\mu), & \gamma_m(\mu) &= \frac{\alpha_s(\mu)}{4\pi} \gamma_m^{(0)} + \frac{\alpha_s^2(\mu)}{(4\pi)^2} \gamma_m^{(1)} + \mathcal{O}(\alpha_s^3). \end{aligned} \quad (\text{S8})$$

For example,  $\gamma_m^{(0)} = -8$  and  $\gamma_{77}^{(0)} = 32/3$ . The full set of required anomalous dimension coefficients can be found in Ref. [59].

To fully implement the split matching procedure it is convenient to also define scale-independent coefficients  $\mathcal{C}_{i \neq 7}$ , which appear in the nonsingular terms in Eq. (3). Analogous to  $\mathcal{C}_7$  above, they are defined to be  $\mu_0$  independent and to satisfy  $\mathcal{C}_i = \bar{\mathcal{C}}_i(\mu_0 = \hat{m}_b)$ . To one-loop order they are given by

$$\mathcal{C}_8 = \bar{\mathcal{C}}_8(\mu_0) + \frac{\alpha_s(\mu_0)}{4\pi} \ln \frac{\hat{m}_b}{\mu_0} \left[ \gamma_m^{(0)} \bar{\mathcal{C}}_8(\mu_0) + \sum_i \gamma_{i8}^{(0)} \bar{\mathcal{C}}_i(\mu_0) \right], \quad \mathcal{C}_{j \neq 7,8} = \bar{\mathcal{C}}_j(\mu_0) + \frac{\alpha_s(\mu_0)}{4\pi} \ln \frac{\hat{m}_b}{\mu_0} \sum_i \gamma_{ij}^{(0)} \bar{\mathcal{C}}_i(\mu_0). \quad (\text{S9})$$

The  $\mu_0$  dependence of  $r_i(\mu_0, \hat{m}_b, \hat{m}_c)$  in Eq. (S5) is defined such that it cancels that of the coefficients  $\bar{\mathcal{C}}_i(\mu_0)$  in Eq. (S5), while for  $\mu_0 = \hat{m}_b$ , the  $r_i(\hat{m}_b, \hat{m}_b, \hat{m}_c)$  agree with their usual definitions in the literature. Denoting their  $\alpha_s$  expansions at  $\mu_0 = \hat{m}_b$  as

$$r_i(\hat{m}_b, \hat{m}_b, \hat{m}_c) = \frac{\alpha_s(\hat{m}_b)}{4\pi} r_i^{(1)} + \frac{\alpha_s^2(\hat{m}_b)}{(4\pi)^2} r_i^{(2)} + \mathcal{O}(\alpha_s^3), \quad (\text{S10})$$

we have at NNLO

$$r_8(\mu_0, \hat{m}_b, \hat{m}_c) = \frac{\alpha_s(\mu_0)}{4\pi} r_8^{(1)} + \frac{\alpha_s^2(\mu_0)}{(4\pi)^2} \left[ r_8^{(2)} + \ln \frac{\hat{m}_b}{\mu_0} r_8^{(1)} (\gamma_m^{(0)} + \gamma_{88}^{(0)} - 2\beta_0) \right] + \mathcal{O}(\alpha_s^3), \quad (\text{S11})$$

while for  $k = 1, \dots, 6$ ,

$$r_k(\mu_0, \hat{m}_b, \hat{m}_c) = \frac{\alpha_s(\mu_0)}{4\pi} r_k^{(1)} + \frac{\alpha_s^2(\mu_0)}{(4\pi)^2} \left[ r_k^{(2)} + \ln \frac{\hat{m}_b}{\mu_0} \left( \sum_{i \neq 7} \gamma_{ki}^{(0)} r_i^{(1)} - 2\beta_0 r_k^{(1)} \right) \right] + \mathcal{O}(\alpha_s^3). \quad (\text{S12})$$

The results of Refs. [41, 42] give

$$\begin{aligned} r_8^{(1)} &= \frac{C_F}{3} \left( 11 - \frac{2\pi^2}{3} + 2i\pi \right), \\ \text{Re } r_8^{(2)} &= C_F \left\{ \left( \frac{c_{78,0}^a}{2} + \frac{55}{3} + \frac{34\pi^2}{9} - \frac{8\pi^4}{27} \right) C_F + \left( \frac{c_{78,0}^{na}}{2} - \frac{3454}{81} + \frac{176\pi^2}{81} + \frac{88\zeta_3}{9} \right) C_A \right. \\ &\quad \left. + \left( \frac{314}{27} - \frac{16\pi^2}{27} - \frac{8\zeta_3}{3} \right) \beta_0(n_l) + \left[ \frac{976}{81} - \frac{4\pi}{\sqrt{3}} - \frac{244\pi^2}{81} + 16\sqrt{3} \text{Cl}_2\left(\frac{\pi}{3}\right) - \frac{32\zeta_3}{27} \right] T_F n_h \right\}. \end{aligned} \quad (\text{S13})$$

A value for  $\text{Im } r_8^{(2)}$  is not yet known, but it only contributes to the spectrum at  $\mathcal{O}(\alpha_s^3)$ . In Eq. (S13),  $n_h = 1$  is the number of flavors with mass  $\hat{m}_b$ , and  $n_l = 4$  is the number of massless flavors, since we neglected for simplicity the  $\hat{m}_c$  dependence in  $r_8^{(2)}$ . The full  $\hat{m}_c$  dependence of  $\text{Re } r_8^{(2)}$ , arising from  $c\bar{c}$  loops inserted into gluon propagators, is known [41, 42], but the massless approximation is sufficiently accurate for our purposes.

For  $r_{1-6}(\mu_0)$  we have the NLO coefficients [60–62]

$$\begin{aligned} r_1^{(1)} &= -\frac{1}{6} r_2^{(1)}, & r_2^{(1)} &= -\frac{1666}{243} + 2a(\rho) + 2b(\rho) - \frac{80}{81} i\pi, \\ r_3^{(1)} &= \frac{2392}{243} + \frac{8\pi}{3\sqrt{3}} - a(1) + 2b(1) + \frac{32}{9} X_b + \frac{56}{81} i\pi, & r_4^{(1)} &= \frac{145}{243} - \frac{1}{6} r_3^{(1)} + 2b(\rho) + 2b(1) - \frac{40}{81} i\pi, \\ r_5^{(1)} &= \frac{6136}{81} - \frac{32\pi}{\sqrt{3}} + 16r_3^{(1)} - \frac{128}{3} X_b, & r_6^{(1)} &= -\frac{310}{27} + 6r_2^{(1)} - \frac{4}{3} r_3^{(1)} + 4r_4^{(1)} + \frac{1}{3} r_5^{(1)} - \frac{104}{27} i\pi. \end{aligned} \quad (\text{S14})$$

Here,  $\rho = \hat{m}_c^2/\hat{m}_b^2$  and  $a(\rho)$ ,  $b(\rho)$ , and  $X_b$  are given in Ref. [62]. Since the Wilson coefficients  $C_{3-6}$  are small, the  $r_{3-6}$  terms only have very small impacts, and their NNLO contributions  $r_{3-6}^{(2)}$  can be safely neglected.

The NNLO contributions  $r_{1,2}^{(2)}(\mu_0)$  are only fully known in the large  $\beta_0$  approximation, where they are obtained as an expansion in  $m_c/m_b$  [63]. They are given by

$$r_1^{(2)} = -r_2^{(2)}/6, \quad r_2^{(2)} = -\frac{3}{2}\beta_0 \left[ \left( \text{Re } r_2^{(2)} \right)_{\text{Ref. [63]}} + i \left( \text{Im } r_2^{(2)} \right)_{\text{Ref. [63]}} \right] + \dots, \quad (\text{S15})$$

where the terms in the square brackets are given in Eqs. (26) and (27) of Ref. [63], and the ellipses denote other independent color structures. The full NNLO contributions to  $r_{1,2}^{(2)}$  are required to cancel the  $\widehat{m}_c$ -scheme dependence and have been computed in the limits  $m_c \gg m_b/2$  and  $m_c = 0$  [38, 39, 64].

The SM prediction for  $C_7^{\text{incl}}$  in Eq. (5) is obtained using the above results together with the input parameters and numerical values for the Wilson coefficients given below in Sec. E. Although the two terms in Eq. (S5) are formally  $\mu_0$  independent, there is still residual  $\mu_0$  dependence from the truncation of perturbation theory. We vary  $\mu_0$  between  $\widehat{m}_b/2$  and  $2\widehat{m}_b$  to obtain an estimate of the associated perturbative uncertainty, quoted in Eq. (5) with the subscript “scale”. In addition, to estimate the uncertainty from missing  $\mathcal{O}(\alpha_s^2)$  charm-loop contributions we use the  $\alpha_s^2/\beta_0$  result in Eq. (S15) with a multiplicative prefactor of  $1.0 \pm 0.5$ , yielding the uncertainty quoted in Eq. (5) with a subscript “ $c\bar{c}$ ”. The parametric uncertainties from input parameters, including the numerical value of  $\widehat{m}_c$  itself, are much smaller than these two sources of uncertainties and can be safely neglected.

### C. Perturbative ingredients for the photon energy spectrum

The perturbative components of the photon energy spectrum are described by Eq. (3), which we repeat here for convenience

$$\widehat{P}(k) = |C_7^{\text{incl}}|^2 \left[ W_{77}^s(k) + W_{77}^{\text{ns}}(k) \right] + 2 \text{Re}(C_7^{\text{incl}}) \sum_{i \neq 7} C_i W_{7i}^{\text{ns}}(k) + \sum_{i,j \neq 7} C_i C_j W_{ij}^{\text{ns}}(k). \quad (\text{S16})$$

Definitions for the Wilson coefficients  $C_7^{\text{incl}}$  and  $C_i$  are given above in Sec. B,  $W_{77}^s(k)$  contains the dominant singular contributions and the  $W_{ij}^{\text{ns}}(k)$  are the various nonsingular terms. In our formula for  $d\Gamma/dE_\gamma$  in Eq. (1) we have kept an overall  $E_\gamma^3$  kinematic prefactor. Here one power of  $E_\gamma$  arises from the photon phase-space integration, and for the dominant 77-like contributions two more factors of  $E_\gamma$  arise from the derivative that acts on the photon field in each  $F^{\mu\nu}$ . Since these factors are universal we do not expand them about the singular limit. This improves the behavior of the decomposition into singular and nonsingular terms in the tail region where these components become comparable.

#### 1. Singular contributions

The all-order factorization theorem for the singular contributions  $W_{77}^s(k)$  is well known [24, 65]. For our treatment we follow Ref. [10] and make use of the SCET-based factorization theorem, expressing the perturbative ingredients in a short-distance scheme. (In the notation of Ref. [10],  $\mu_i = \mu_J$  and  $\mu_\Lambda = \mu_S$ .)

$$W_{77}^s(k) = h_s(\widehat{m}_b, \widehat{m}_c, \mu_b) U_H(\widehat{m}_b, \mu_b, \mu_J) \int d\omega d\omega' \widehat{m}_b J[\widehat{m}_b(k - \omega), \mu_J] U_S(\omega - \omega', \mu_J, \mu_S) \widehat{C}_0(\omega', \mu_S). \quad (\text{S17})$$

Here  $h_s$ ,  $J$ , and  $\widehat{C}_0$  are the fixed-order hard, jet, and soft functions, which we include up to NNLO. The evolution kernels  $U_H$  and  $U_S$  sum large logarithms of  $k/\widehat{m}_b \sim 1 - 2E_\gamma/\widehat{m}_b$  to all orders in perturbation theory, and are included at NNLL order. The perturbative expressions for  $h_s$ ,  $J$ ,  $\widehat{C}_0$  as well as  $U_H$  and  $U_S$  together with the required anomalous dimensions can be found in Ref. [10], where they were obtained using results from Refs. [22, 29–37].

In the appropriate region the logarithmic summation is achieved by choosing  $\mu_b \sim \widehat{m}_b$ ,  $\mu_S \sim \widehat{m}_b - 2E_\gamma \gtrsim \Lambda_{\text{QCD}}$ ,  $\mu_J^2 \sim \mu_b \mu_S$ . The dependence of  $W_{77}^s(k)$  on  $\mu_b$ ,  $\mu_J$ , and  $\mu_S$  cancels between the fixed-order functions and evolution kernels order by order in resummed perturbation theory, and will be used to estimate higher-order perturbative uncertainties. The precise procedure we use to estimate these uncertainty and to transition into and out of this resummation region is described in more detail in Sec. C3 below.

The hard function  $h_s$  arises from matching the QCD chromomagnetic operator  $O_7$  onto a corresponding SCET operator at the scale  $\mu_b$ , which is the second step in the split matching procedure described in the previous section. To be consistent with the first step of the split matching, also here we integrate out bottom and charm quarks at the hard matching scale  $\mu_b$ . As a result, the hard function includes all effects of virtual massive charm loops inserted into

gluon propagators, which starting at two loops gives rise to its  $\widehat{m}_c$  dependence given by

$$\begin{aligned}
h_s(\widehat{m}_b, \widehat{m}_c, \mu_b) &= h_s(\widehat{m}_b) + \frac{\alpha_s^2(\mu_b)}{(4\pi)^2} C_F T_F f_{2s}\left(\frac{\widehat{m}_c^2}{\widehat{m}_b^2}\right) \\
f_{2s}(\rho) &= \frac{4}{9} \ln^3 \rho + \frac{50}{9} \ln^2 \rho + \left[ -\frac{124}{9} \ln(1-\rho) + \frac{8\pi^2}{9} + \frac{794}{27} \right] \ln \rho - \frac{16}{3} \text{Li}_3(\rho) + \left( \frac{8 \ln \rho}{3} - \frac{124}{9} \right) \text{Li}_2(\rho) \\
&\quad + \frac{124\pi^2}{27} + \frac{5578}{81} + \rho \left( \frac{32}{9} \ln \rho + \frac{172}{9} \right) + 12\rho^2 \left[ -\text{Li}_2(\rho) + \frac{1}{2} \ln^2 \rho - \log(1-\rho) \log(\rho) + \frac{\pi^2}{3} \right] \\
&\quad - \frac{2}{9} \sqrt{\rho} (35\rho + 81) \left[ -4\text{Li}_2(\sqrt{\rho}) + \text{Li}_2(\rho) + 2\text{arctanh}(\sqrt{\rho}) \ln \rho + \pi^2 \right], \tag{S18}
\end{aligned}$$

where  $h_s(\widehat{m}_b)$  is the two-loop result for massless quarks given in Ref. [10] and  $f_{2s}(\rho)$  is extracted from the results of Ref. [66]. At the same time, all SCET ingredients are defined for  $n_f = 3$  massless flavors.

## 2. Nonsingular contributions

The remaining nonsingular terms  $W_{ij}^{\text{ns}}(k)$  in Eq. (S16) are included using fixed-order perturbation theory. These terms are power-suppressed by  $k/\widehat{m}_b$  in the  $B \rightarrow X_s \gamma$  peak region, but lose this suppression in the tail of the spectrum where  $k \sim \widehat{m}_b$ . By using  $C_7^{\text{incl}}$  and  $\mathcal{C}_i$  in Eq. (S16), the  $W_{ij}^{\text{ns}}$  are also  $\mu$  independent order by order in  $\alpha_s$ . We use the notation  $\mu = \mu_{ns}$  for the residual scale dependence in all nonsingular terms, and will vary this scale as part of our perturbative uncertainty estimate. Up to  $\mathcal{O}(\alpha_s^2)$  we write

$$\widehat{m}_b W_{ij}^{\text{ns}}(\widehat{m}_b x) = \frac{1}{(1-x)^3} \left\{ \frac{\alpha_s(\mu_{ns})}{\pi} C_F w_{ij}^{\text{ns}(1)}(x) + \frac{\alpha_s^2(\mu_{ns})}{\pi^2} C_F \left[ w_{ij}^{\text{ns}(2)}(x) + \frac{1}{2} \beta_0 \ln \frac{\mu_{ns}}{m_b} w_{ij}^{\text{ns}(1)}(x) + \Delta w_{ij}^{\text{ns}}(x) \right] \right\}. \tag{S19}$$

The NLO and NNLO coefficient functions,  $w_{ij}^{\text{ns}(1)}(x)$  and  $w_{ij}^{\text{ns}(2)}(x)$ , are determined by taking the full fixed-order results for  $d\Gamma/dE_\gamma$  calculated in the literature, reorganizing the Wilson coefficients as in Eq. (S16), and then using Eq. (1) with  $\mathcal{F}(m_B - 2E_\gamma - k) = \delta(\widehat{m}_b - 2E_\gamma - k)$  and subtracting the fixed-order singular terms predicted by  $W_{77}^s(k)$  at each order. When this construction is carried out with the results consistently expressed in a short-distance mass scheme, there is an additional  $\mathcal{O}(\alpha_s^2)$  correction induced, which is denoted as  $\Delta w_{ij}^{\text{ns}}(x)$  in Eq. (S19).

Note that the extraction of the nonsingular corrections is somewhat nontrivial. For example, if we take the full theory result in a short-distance mass scheme and extract the coefficient of the terms proportional to  $\text{Re}[\overline{C}_7 \overline{C}_8^*]$  (setting  $\mu = \widehat{m}_b$ ), we find

$$\begin{aligned}
\left. \frac{\widehat{m}_b}{2\Gamma_0} \frac{d\Gamma}{dE_\gamma} \right|_{\overline{C}_7 \overline{C}_8} &= \frac{\alpha_s(\widehat{m}_b)}{4\pi} \left[ \text{Re}(r_8^{(1)}) \delta(x) + 4C_F w_{78}^{\text{ns}(1)}(x) \right] \\
&\quad + \frac{\alpha_s^2(\widehat{m}_b)}{(4\pi)^2} \left\{ \text{Re}(r_8^{(2)}) \delta(x) + \text{Re}(r_8^{(1)}) 4C_F [w_{77}^{\text{s}(1)}(x) + w_{77}^{\text{ns}(1)}(x)] + 16C_F [w_{78}^{\text{ns}(2)}(x) + \Delta w_{78}^{\text{ns}}(x)] \right\} \\
&\quad + \mathcal{O}(\alpha_s^3). \tag{S20}
\end{aligned}$$

Here the  $\delta(x)$  and  $w_{77}^{\text{s}(1)}(x)$  terms are both reproduced by the singular  $E_\gamma^3 |C_7^{\text{incl}}|^2 W_{77}^s$  term. Furthermore, the  $\text{Re}(r_8^{(1)}) w_{77}^{\text{ns}(1)}(x)$  term is reproduced by  $E_\gamma^3 |C_7^{\text{incl}}|^2 W_{77}^{\text{ns}}$ . Only the remaining terms contribute to  $W_{78}^{\text{ns}}$ , as indicated by their superscripts.

By far the numerically dominant nonsingular corrections come from  $W_{77}^{\text{ns}}$ . Using as input the results from Refs. [31, 40, 67], we find that the one-loop and two-loop nonsingular coefficient functions are

$$\begin{aligned}
w_{77}^{\text{ns}(1)}(x) &= -\frac{8-7x+2x^2}{2} \ln x - \frac{3}{4} (1-x)(5-3x), \\
w_{77}^{\text{ns}(2)}(x) &= \frac{1}{2} \left( C_F - \frac{1}{2} C_A \right) \left\{ (2+x-x^2-x^3) \left[ L_1(x) + \frac{1}{x} \left( \ln x - \frac{2}{3} \right) L_3(x) \right] \right. \\
&\quad \left. + (2-x+x^2) [2\text{Li}_3(x) - \text{Li}_2(x) \ln x] - x^2 [\text{Li}_3(x^2) - \text{Li}_2(x^2) \ln x] + (-15+11x-3x^2) \frac{\zeta_3}{2} \right\}
\end{aligned}$$

$$\begin{aligned}
& + C_F \left\{ (2 + 10x - x^2) \frac{1}{4x} \left[ \text{Li}_3(1-x) + \text{Li}_3(x) - \text{Li}_2(x) \ln x - \frac{1}{2} \ln(1-x) \ln^2 x - \zeta_3 \right] \right. \\
& + \frac{8 - 32x + 8x^2 + 47x^3 - 46x^4 - 2x^5 + 8x^6}{12(1-x)} L_2(x) + \frac{-6 + 28x - 9x^2 - 9x^3 - 2x^4 + x^5}{12(1-x)} L_3(x) \\
& + \frac{8 - 13x + 9x^2 - 3x^3}{4(1-x)} \ln^3 x + \frac{195 - 405x + 266x^2 - 66x^3 + 2x^4 - x^5}{24(1-x)} \ln^2 x \\
& + \left[ \frac{725 - 572x + 29x^2 + 28x^3 + 32x^4}{48} + (7 - 7x + x^2) \frac{\pi^2}{12} \right] \ln x \\
& + \left. \frac{451 - 1531x + 1553x^2 - 525x^3 - 16x^4 + 32x^5}{96(1-x)} + \frac{59 - 149x + 150x^2 - 53x^3 - 2x^4 + x^5}{12(1-x)} \frac{\pi^2}{6} \right\} \\
& + C_A \left\{ \frac{-4 + 12x + 8x^2 - 11x^3 + 3x^4 + x^5}{12} L_2(x) + \frac{-3 - 10x + 15x^2 + 3x^3 + 2x^4 - x^5}{24(1-x)} L_3(x) \right. \\
& - \frac{x^2}{8} \ln^3 x + \frac{24 + 9x + 4x^2 + x^3 - x^4}{48} \ln^2 x \\
& + \left[ \frac{-110 + 188x - 5x^2 - 45x^3 + 10x^4 + 4x^5}{48(1-x)} + (9 - 7x + 3x^2) \frac{\pi^2}{24} \right] \ln x \\
& + \left. \frac{-10 + 129x - 65x^2 + 4x^3 + 2x^4}{48} + \frac{22 - 67x + 54x^2 - 7x^3 + 2x^4 - x^5}{24(1-x)} \frac{\pi^2}{6} \right\} \\
& + \beta_0 \left\{ \frac{2 + 2x - x^2}{8x} \left[ \text{Li}_2(1-x) - \frac{\pi^2}{6} \right] + \frac{3(8 - 7x + 2x^2)}{16} \ln^2 x + \frac{1 - 45x + 45x^2 - 19x^3}{48(1-x)} \ln x \right. \\
& + \left. \frac{-63 + 92x - 41x^2}{32} + (8 - 7x + 2x^2) \frac{\pi^2}{48} \right\}. \tag{S21}
\end{aligned}$$

To write  $w_{77}^{\text{ns}(2)}(x)$  we defined the following functions of  $x$ , which diverge at most logarithmically for  $x \rightarrow 0$

$$\begin{aligned}
L_1(x) &= \frac{1}{x} \left\{ \text{Li}_3(1-x) + 2\text{Li}_3\left(\frac{1}{1+x}\right) - 2\text{Li}_3\left(\frac{1-x}{1+x}\right) + \frac{1}{4}\text{Li}_3\left[\left(\frac{1-x}{1+x}\right)^2\right] - \frac{1}{6}[2\ln(1+x)^2 - \pi^2] \ln(1+x) - \frac{5\zeta_3}{4} \right\}, \\
L_2(x) &= \frac{1}{2x^3} \left\{ \text{Li}_2(x^2) + 2[x^2 + \ln(1-x^2)] \ln x - x^2 \right\}, \\
L_3(x) &= \frac{1}{2}\text{Li}_2(x^2) - \text{Li}_2(x) + \ln(1+x) \ln x. \tag{S22}
\end{aligned}$$

Above we mentioned that a factor of  $E_\gamma^3 \propto (1-x)^3$  was universal for the 77 contributions. It follows that  $(1-x)^3 W_{77}^{\text{ns}}$  should also vanish as  $(1-x)^3$  as  $x \rightarrow 1$ , and hence that the  $w_{77}^{\text{ns}}(x)$  coefficients should vanish like  $(1-x)^3$  for  $x \rightarrow 1$  to cancel the overall factor  $1/(1-x)^3$  in Eq. (S19). Expanding the above results in the limit  $x \rightarrow 1$ , we find

$$\begin{aligned}
w_{77}^{\text{ns}(1)}(x) &= \frac{9}{4}(1-x)^3 + \mathcal{O}[(1-x)^4], \\
w_{77}^{\text{ns}(2)}(x) &= \left[ C_F \left( -\frac{4933}{1728} - \frac{3\pi^2}{8} + \frac{\zeta_3}{2} \right) + C_A \left( \frac{599}{576} - \frac{3\pi^2}{16} - \frac{\zeta_3}{4} \right) + \beta_0 \left( \frac{5}{4} - \frac{\pi^2}{24} \right) \right] (1-x)^3 + \mathcal{O}[(1-x)^4]. \tag{S23}
\end{aligned}$$

If we would expand the  $E_\gamma^3$  in the singular SCET contribution  $W_{77}^s$ , then this would modify the nonsingular contribution, such that it would not vanish like  $E_\gamma^3$  either. In this situation, as was also noted in Ref. [68], the proper  $E_\gamma^3$  behavior of the spectrum would be obtained only by nontrivial cancellations between the singular and nonsingular contributions. Although formally the difference between these approaches corresponds to a different treatment of nonsingular corrections, this difference can be numerically important even to rather low values of  $x$  because of the third power, and the fact that the resummation in the singular terms can potentially spoil the cancellation at small  $x$ . For this reason, our approach of keeping the  $E_\gamma^3$  prefactor unexpanded is preferred.

For the remaining nonsingular coefficient functions, the fixed-order  $\overline{C}_7 \overline{C}_8$  result from Refs. [41, 42] allows us to extract  $w_{78}^{\text{ns}(1)}$  and  $w_{78}^{\text{ns}(2)}$ . Although both of these coefficients are used in our analysis, for brevity of the presentation we only quote here the first-order term

$$w_{78}^{\text{ns}(1)}(x) = \frac{x}{3(1-x)} \ln x + \frac{5 - 2x + x^2}{12}. \tag{S24}$$



Finally, for the remaining nonsingular structures, the full theory results at one loop are well known [69–72]. They enable us to determine the following  $\mathcal{O}(\alpha_s)$  nonsingular coefficient functions

$$\begin{aligned} w_{88}^{\text{ns}(1)}(x) &= \frac{1}{36(1-x)} \left[ 2(1+x^2) \left( \ln x - 2 \ln \frac{\mu}{m_b} \right) - 3 - 7x^2 + 2x^3 \right], \\ w_{72}^{\text{ns}(1)}(x) &= -\frac{8}{3} \rho^2 \mathcal{G}_1 \left( \frac{1-x}{4\rho} \right), \quad w_{71}^{\text{ns}(1)}(x) = -\frac{1}{6} w_{72}^{\text{ns}(1)}(x), \quad w_{18}^{\text{ns}(1)}(x) = -\frac{1}{6} w_{82}^{\text{ns}(1)}(x) = \frac{1}{18} w_{72}^{\text{ns}(1)}(x), \\ w_{22}^{\text{ns}(1)}(x) &= \frac{4}{9} \rho \mathcal{G}_2 \left( \frac{1-x}{4\rho} \right), \quad w_{11}^{\text{ns}(1)}(x) = -\frac{1}{6} w_{12}^{\text{ns}(1)}(x) = -\frac{1}{6} w_{21}^{\text{ns}(1)}(x) = \frac{1}{36} w_{22}^{\text{ns}(1)}(x), \end{aligned} \quad (\text{S25})$$

where  $\rho = \widehat{m}_c^2 / \widehat{m}_b^2$  and the charm-loop functions  $\mathcal{G}_{1,2}(u)$  are given by

$$\begin{aligned} \mathcal{G}_1(u) &= \int_0^u du' \text{Re}[\mathcal{G}(u') + u'], & \mathcal{G}_2(u) &= \int_0^u du' (1 - 4\rho u') \left| \frac{\mathcal{G}(u')}{u'} + 1 \right|^2, \\ \mathcal{G}(u) &= \begin{cases} -[\arctan \sqrt{u/(1-u)}]^2, & u \leq 1, \\ [\ln(\sqrt{u} + \sqrt{u-1}) - i\pi/2]^2, & u > 1. \end{cases} \end{aligned} \quad (\text{S26})$$

The corresponding  $\mathcal{O}(\alpha_s^2)$  nonsingular coefficient functions  $w_{ij}^{\text{ns}(2)}$  are not yet fully known. However, we stress that analogous to the  $\overline{C}_7 \overline{C}_8$  contribution in Eq. (S20), all singular contributions as well as a subset of the nonsingular contributions appearing at two loops that behave  $O_7$ -like are already accounted for via the  $|C_7^{\text{incl}}|^2 (W_{77}^{\text{s}} + W_{77}^{\text{ns}})$  term. For the remaining two-loop contributions  $w_{ij}^{\text{ns}(2)}$  we use the known results for the  $\alpha_s^2 \beta_0$  terms obtained from the full theory results of Refs. [43–45], and thus leave out contributions with the color structure  $C_A$ . Again for brevity, we do not list here the results for these coefficient functions. The nonsingular corrections for  $i, j = 3, 4, 5, 6$  are known to be very small [72] and are neglected.

### 3. Scale choices and estimation of perturbative uncertainties

We now discuss our treatment of the central scales  $\mu_i$  and their variations used to estimate perturbative uncertainties. The soft and jet scales take different forms in the three parametrically distinct regions of the spectrum:

$$\begin{aligned} 1) \text{ SCET shape function region: } & \Lambda_{\text{QCD}} \sim (m_B - 2E_\gamma) \ll \widehat{m}_b, \\ 2) \text{ Shape function OPE: } & \Lambda_{\text{QCD}} \ll (m_B - 2E_\gamma) \ll \widehat{m}_b, \\ 3) \text{ Local OPE: } & \Lambda_{\text{QCD}} \ll (m_B - 2E_\gamma) \sim \widehat{m}_b. \end{aligned} \quad (\text{S27})$$

This can be properly accounted for by using profile functions,  $\mu_S = \mu_S(E_\gamma)$  and  $\mu_J = \mu_J(E_\gamma)$ , as discussed in Ref. [10] (see also Ref. [73]). The hard scale  $\mu_b \sim \widehat{m}_b$  is independent of  $E_\gamma$ . For the remaining scales we use

$$\begin{aligned} \mu_S(E_\gamma) &= \begin{cases} \mu_0 & E_1 \leq E_\gamma \\ \mu_0 + (\mu_b - \mu_0) \frac{2(E_\gamma - E_1)^2}{(E_2 - E_1)^2} & \frac{1}{2}(E_1 + E_2) \leq E_\gamma < E_1 \\ \mu_b - (\mu_b - \mu_0) \frac{2(E_\gamma - E_2)^2}{(E_2 - E_1)^2} & E_2 \leq E_\gamma < \frac{1}{2}(E_1 + E_2) \\ \mu_b & E_\gamma < E_2, \end{cases} \\ \mu_J(E_\gamma) &= [\mu_S(E_\gamma)]^{(1-e_J)/2} \mu_b^{(1+e_J)/2}, \\ \mu_{ns}(E_\gamma) &= [\mu_J(E_\gamma)]^{(1-e_{ns})/2} \mu_b^{(1+e_{ns})/2}. \end{aligned} \quad (\text{S28})$$

The constant parameters  $\mu_0$ ,  $\mu_b$ ,  $E_1$ ,  $E_2$ ,  $e_J$ , and  $e_{ns}$  can be varied to assess perturbative uncertainties. In Eq. (S28) the soft scale  $\mu_S(E_\gamma)$  takes the value  $\mu_S = \mu_0 \sim 1 \text{ GeV} \gtrsim \Lambda_{\text{QCD}}$  in the SCET region given by  $E_1 \leq E_\gamma$ . In the local OPE region,  $E_\gamma < E_2$ , all the scales become equal,  $\mu_S = \mu_J = \mu_{ns} = \mu_b$ , which turns off the resummation and is crucial for the singular and nonsingular contributions to properly recombine to reproduce the local OPE prediction for the spectrum. In between these two we have a transition region where we join the soft scales in a smooth manner, as given by the quadratic functions of  $E_\gamma$  shown in Eq. (S28). Since the transition scales  $E_1$  and  $E_2$  are not very widely separated, there is no need to separately implement a shape function OPE scaling region for the soft scale,

noting that it is anyway well captured by the form of  $\mu_S(E_\gamma)$  used in the transition. The parameters  $e_J$  and  $e_{ns}$  provide a means to independently vary the jet and nonsingular scales when assessing perturbative uncertainties. By default we have  $e_J = e_{ns} = 0$ . For  $\mu_J$  this gives the geometric mean of the soft and hard scales as required. For  $\mu_{ns}$  we choose our default as the geometric mean between the hard and jet scales, and we will vary this choice up to the hard scale and down to the jet scale. This allows us to capture the fact that the nonsingular perturbative series are sensitive to lower scales than the hard scale (as would be made explicit in subleading power factorization theorems for these terms).

Taken together we consider a total of  $3^5 = 243$  different variations for the profile parameters to assess the perturbative uncertainty, given by the choices

$$\begin{aligned} \mu_b &= \{4.7, 2.35, 9.4\} \text{ GeV}, & \mu_0 &= \{1.3, 1.1, 1.8\} \text{ GeV}, & E_1 &= \{2.2, 2.1, 2.3\} \text{ GeV}, & E_2 &= 1.6 \text{ GeV}, \\ e_J &= \{0, -1/3, +1/3\}, & e_{ns} &= \{0, -1/2, +1/2\}. \end{aligned} \quad (\text{S29})$$

For each parameter, the first case in the list is the default central value, and the next two are the variations. We do not vary  $E_2$  since our fit analysis is not sensitive to the uncertainty in the spectrum in the region  $E_\gamma \lesssim 1.6 \text{ GeV}$ . To assess the theoretical uncertainty we separately carry out the fit for each of these 243 cases and then consider the spread of the results as giving the range of possibilities for the central values. The results for these 243 fits are shown by the dark yellow shape function curves in Fig. 2 and as the yellow scatter points in Figs. 3 and S2. The theoretical uncertainty for a given quantity is then obtained by using the largest absolute deviation of these results from the default central value.

## D. Shape Functions

### 1. Shape-function basis

We briefly summarize the functional basis used for expanding the shape function in Eq. (6). For more details we refer to Ref. [10]. The orthonormal basis functions  $f_n(x)$  are given by

$$f_n(x) = \sqrt{y'(x)} \phi_n[y(x)], \quad \phi_n(y) = \sqrt{\frac{2n+1}{2}} \frac{1}{2^n n!} \frac{d^n}{dy^n} (y^2 - 1)^n, \quad (\text{S30})$$

where  $\phi_n(y)$  is an orthonormal basis on  $y \in [-1, 1]$ , given by the normalized Legendre polynomials. The function  $y(x)$  can be any variable transformation that maps  $x \in [0, \infty)$  to  $y \in [-1, 1]$ , i.e., it has to satisfy  $y(0) = -1$ ,  $y(\infty) = +1$ , and  $y'(x) > 0$ . Given any positive and normalized function  $Y(x)$  on  $x \in [0, \infty)$ , we can construct  $y(x)$  from its integral

$$y(x) = -1 + 2 \int_0^x dx' Y(x'), \quad y'(x) = 2Y(x). \quad (\text{S31})$$

With this construction we have

$$f_0^2(x) = y'(x) \phi_0^2[y(x)] = Y(x), \quad F_{00}(k) = \frac{1}{\lambda} Y\left(\frac{k}{\lambda}\right). \quad (\text{S32})$$

Hence,  $Y(x)$  or equivalently  $F_{00}(k)$  acts as the generating function for the basis, for which we can use any suitable model function.

We consider the following functional forms

$$\begin{aligned} Y_{\text{exp}}(x, p) &= \frac{(p+1)^{p+1}}{\Gamma(p+1)} x^p e^{-(p+1)x}, \\ Y_{\text{gauss}}(x, p) &= \frac{2 a^{p+1}}{\Gamma[(1+p)/2]} x^p e^{-a^2 x^2}, \quad a = \frac{\Gamma(1+p/2)}{\Gamma[(1+p)/2]}. \end{aligned} \quad (\text{S33})$$

where the parameter  $p$  determines the behavior of  $F_{00}(k) \sim k^p$  for  $k \rightarrow 0$ . As explained in Ref. [10], for integer  $p$  we need at least  $p \geq 3$  to ensure that after short-distance subtractions, which involve taking two derivatives of  $\hat{F}(k)$ , the spectrum vanishes at the kinematic endpoint. We have tested the three functional forms  $Y_{\text{exp}}(x, 3)$ ,  $Y_{\text{exp}}(x, 4)$ ,  $Y_{\text{gauss}}(x, 3)$  in the pre-fit. Of these,  $Y_{\text{exp}}(x, 3)$  provides the best pre-fits and is thus used as the default functional form.

## 2. Treatment of leading and subleading contributions to $\mathcal{F}(k)$

Our definition of the shape function  $\mathcal{F}(k)$ , appearing in the leading power contributions to the cross section, absorbs the non-resolved subleading power shape functions appearing in  $B \rightarrow X_s \gamma$ . This induces corrections in the formulas for the moments of  $\mathcal{F}(k)$  which are used in our analysis.

Taking a set of values  $\{c_n\}$  as input, from a fit or otherwise, the  $i$ th moment of  $\mathcal{F}(k)$  is given by

$$M^i[\mathcal{F}] = \int dk k^i \mathcal{F}(k) = \sum_{m,n} c_m c_n \int dk k^i F_{mn}(k) \equiv \sum_{m,n} c_m c_n M_{mn}^i. \quad (\text{S34})$$

Here in the second step we inserted the basis expansion for  $\mathcal{F}(k)$ , and in the last relation we defined the moment matrices  $M_{mn}^i \equiv M^i[F_{mn}]$  as the moments of the  $F_{mn}(k)$  basis functions defined in Eq. (9).

Theoretically moments of  $\mathcal{F}(k)$  up to  $\mathcal{O}(\Lambda_{\text{QCD}}^3)$  are given in terms of HQET hadronic parameters by [10, 74]

$$\begin{aligned} M^0[\mathcal{F}] &= \sum_n |c_n|^2 = 1 + \mathcal{O}(\alpha_s \Lambda_{\text{QCD}}^2 / \hat{m}_b^2), \\ M^1[\mathcal{F}] &= \sum_{m,n} c_m c_n M_{mn}^1 = m_B - \hat{m}_b + \frac{-\hat{\lambda}_1 + 3\hat{\lambda}_2}{2\hat{m}_b} + \frac{5\hat{\rho}_1 - 3\hat{\rho}_2}{6\hat{m}_b^2} + \mathcal{O}(\alpha_s \Lambda_{\text{QCD}}^2 / \hat{m}_b), \\ M^2[\mathcal{F}] &= \sum_{m,n} c_m c_n M_{mn}^2 = -\frac{\hat{\lambda}_1}{3} + \frac{\hat{\rho}_1 + 3\hat{\rho}_2}{3\hat{m}_b} - (m_B - \hat{m}_b)^2 + 2(m_B - \hat{m}_b)M^1[\mathcal{F}] + \mathcal{O}(\alpha_s \Lambda_{\text{QCD}}^3 / \hat{m}_b), \\ M^3[\mathcal{F}] &= \sum_{m,n} c_m c_n M_{mn}^3 = \frac{\hat{\rho}_1}{3} + (m_B - \hat{m}_b)^3 - 3(m_B - \hat{m}_b)^2 M^1[\mathcal{F}] + 3(m_B - \hat{m}_b) M^2[\mathcal{F}] + \mathcal{O}(\alpha_s \Lambda_{\text{QCD}}^4 / \hat{m}_b), \end{aligned} \quad (\text{S35})$$

where  $\hat{m}_b$  is defined in the 1S scheme, and

$$\hat{\lambda}_1 = \lambda_1^i(R) + \frac{\mathcal{T}_1 + 3\mathcal{T}_2}{\hat{m}_b}, \quad \hat{\lambda}_2 = \lambda_2(\mu) + \frac{\mathcal{T}_3 + 3\mathcal{T}_4}{3\hat{m}_b}. \quad (\text{S36})$$

Here,  $\lambda_1^i(R)$  is defined in the invisible scheme [10] with  $R = 1 \text{ GeV}$ , and  $\lambda_2(\mu)$  is the usual chromomagnetic matrix element (defined in the  $\overline{\text{MS}}$  scheme). The  $\hat{\rho}_i$  are matrix elements of local dimension-6 operators in HQET in a suitable short-distance scheme, and the  $\mathcal{T}_i$  are matrix elements of time-ordered products [75].

The  $1/\hat{m}_b$  corrections in Eq. (S35) arise from absorbing the subleading shape functions into  $\mathcal{F}(k)$ . By doing so, the moment expansion of  $\mathcal{F}(k)$  in Eq. (S35) reproduces the complete  $\mathcal{O}(\Lambda_{\text{QCD}}^3 / \hat{m}_b^3)$  local OPE corrections for  $B \rightarrow X_s \gamma$  [74, 76]. Note that the normalization of  $\mathcal{F}(k)$  does not receive  $\mathcal{O}(\Lambda_{\text{QCD}}^2 / \hat{m}_b^2)$  and  $\mathcal{O}(\Lambda_{\text{QCD}}^3 / \hat{m}_b^3)$  corrections. At  $\mathcal{O}(\alpha_s)$  and beyond, the subleading shape functions will in general involve different perturbative prefactors than the leading shape function. Since  $\mathcal{O}(\alpha_s \Lambda_{\text{QCD}} / \hat{m}_b)$  corrections are beyond the order we are working, they are also effectively absorbed into  $\mathcal{F}(k)$ , which means the moments receive relative corrections of  $\mathcal{O}(\alpha_s \Lambda_{\text{QCD}} / \hat{m}_b)$  as indicated in Eq. (S35), which we neglect. The exception is the normalization of  $\mathcal{F}(k)$ , which only receives relative  $\mathcal{O}(\alpha_s \Lambda_{\text{QCD}}^2 / \hat{m}_b^2)$  corrections. (The  $\mathcal{O}(\Lambda_{\text{QCD}}^4)$  corrections to the moments are not included and not explicitly indicated.)

It turns out that the numerical effect of the included subleading shape functions on the first moment is significant. For typical values of the  $\hat{\lambda}_i$  and  $\hat{\rho}_i$  parameters, the  $1/\hat{m}_b$  corrections to the first moment contribute about 70–80 MeV causing a corresponding 70–80 MeV shift in the extracted value of  $\hat{m}_b$ . In other words, without including these effects we would obtain a value of  $\hat{m}_b$  that is 70–80 MeV too small.

Values for  $\hat{\lambda}_2$  and  $\hat{\rho}_2$  are obtained from meson mass relations as discussed below in Sec. E3, whereas values of  $\hat{m}_b$ ,  $\hat{\lambda}_1$ , and  $\hat{\rho}_1$  are obtained whenever necessary from  $M^1[\mathcal{F}]$ ,  $M^2[\mathcal{F}]$ , and  $M^3[\mathcal{F}]$  by inverting the moment relations in Eq. (S35). In particular, the moment relations are inverted when the current value of  $\hat{m}_b$  is needed inside the fit.

## 3. Resolved-photon contributions

Considerable attention has been paid to the so-called resolved photon contributions, as they were estimated to yield a 5% theoretical uncertainty in the total rate, not reducible below 4% [47]. More recently Ref. [48] estimated their impact to be substantially smaller. Using somewhat different considerations, we also find that these contributions are not as large as estimated in Ref. [47]. From our analysis we find that the only marginally relevant contributions are

those related to the calculable  $\mathcal{O}(\lambda_2/\widehat{m}_c^2)$  corrections to the total rate [49–51], which enter via the subleading shape function  $g_{27}(k)$  as discussed below.

The resolved-photon contributions coming from  $O_8O_7$  [77, 78] and  $O_2O_7$  are expected to be most significant [47], while contributions from  $O_2O_2$ ,  $O_2O_8$ , and  $O_8O_8$  can be neglected.

As pointed out in Ref. [79], the potentially relevant  $O_8O_7$  contribution can be constrained using the measured isospin asymmetry in  $B \rightarrow X_s\gamma$ , defined by

$$\Delta_{0-} = \frac{\Gamma(\bar{B}^0 \rightarrow X_s\gamma) - \Gamma(B^- \rightarrow X_s\gamma)}{\Gamma(\bar{B}^0 \rightarrow X_s\gamma) + \Gamma(B^- \rightarrow X_s\gamma)} \equiv \frac{\Gamma^0 - \Gamma^-}{\Gamma^0 + \Gamma^-}. \quad (\text{S37})$$

To see this, we decompose these contributions to  $\Gamma^0$  and  $\Gamma^-$ , denoted as  $\delta\Gamma^-$  and  $\delta\Gamma^0$ , according to the quark to which the photon couples (besides the  $O_7$  operator),

$$\begin{aligned} \delta\Gamma^- &= Q_u\delta\Gamma^a + (Q_d + Q_s)\delta\Gamma^b = Q_u(\delta\Gamma^a - \delta\Gamma^b), \\ \delta\Gamma^0 &= Q_d\delta\Gamma^a + (Q_u + Q_s)\delta\Gamma^b = Q_d(\delta\Gamma^a - \delta\Gamma^b), \end{aligned} \quad (\text{S38})$$

where for  $\delta\Gamma^a$  the photon couples to the valence quark flavor, and for  $\delta\Gamma^b$  to any non-valence flavors. For the non-valence contribution we used that  $SU(3)$  flavor symmetry implies that  $\delta\Gamma^b$  is universal at leading order. Since  $Q_u + Q_d + Q_s = 0$ , both contributions are proportional to  $\delta\Gamma^a - \delta\Gamma^b$ . The isospin asymmetry is given by [79]

$$\Delta_{0-} = \frac{\delta\Gamma^0 - \delta\Gamma^-}{\Gamma^0 + \Gamma^-} = -\frac{\delta\Gamma^a - \delta\Gamma^b}{\Gamma^0 + \Gamma^-}. \quad (\text{S39})$$

Hence, the relative impact of these contributions to the isospin-averaged rate is given by

$$\frac{\delta\Gamma^- + \delta\Gamma^0}{\Gamma^0 + \Gamma^-} = \frac{1}{3} \frac{\delta\Gamma^a - \delta\Gamma^b}{\Gamma^0 + \Gamma^-} = -\frac{\Delta_{0-}}{3} = (0.16 \pm 0.71)\%. \quad (\text{S40})$$

where we used the latest Belle measurement  $\Delta_{0-} = -(0.48 \pm 2.12)\%$  [80] (for  $m_{X_s} < 2.8 \text{ GeV}$  or equivalently  $E_\gamma > 1.9 \text{ GeV}$ ), which is nearly a factor of three more precise than earlier results. Hence, the  $O_7O_8$  contribution is experimentally constrained to be much smaller than the current sensitivity, and can be neglected.

Concerning the  $O_2O_7$  contribution, unlike Ref. [47], we treat the charm quark as heavy in our analysis, which amounts to expanding the charm loop in  $\Lambda_{\text{QCD}}/m_c$ . The resulting contribution to the spectrum is then given in terms of an unknown  $\mathcal{O}(\Lambda_{\text{QCD}}^2)$  subleading shape function  $g_{27}(k)$  as

$$\begin{aligned} \frac{d\Gamma_{g_{27}}}{dE_\gamma} &= 2\Gamma_0 \frac{(2E_\gamma)^3}{\widehat{m}_b^3} \frac{1}{\widehat{m}_b^2} \int dk \widehat{P}_{27}(k) g_{27}(m_B - 2E_\gamma - k), \\ \widehat{P}_{27}(k) &= 2\text{Re}(C_7^{\text{incl}}) \left( C_2 - \frac{C_1}{6} \right) \left( -\frac{\widehat{m}_b^2}{18\widehat{m}_c^2} \right) U_{\text{NLL}}(k). \end{aligned} \quad (\text{S41})$$

In a complete factorization analysis, this contribution would involve some evolution between hard, jet, and soft contributions, which is currently not known. To provide some reasonable Sudakov suppression in the peak region, which is important to avoid artificially enhancing this contribution relative to the leading, resummed  $W_{77}^s$  in Eq. (S17), we include in it the NLL evolution factor of the leading contribution given by the product  $U_{\text{NLL}} = [U_H(\mu_b, \mu_J) U_S(k, \mu_J, \mu_S)]_{\text{NLL}}$  with  $\mu_{b,J,S}$  fixed to their central scales.

The subleading shape function  $g_{27}(k)$  is not known, but its moments can be calculated in terms of local matrix elements. To parametrize it, we expand it as

$$g_{27}(k) = \widehat{\lambda}_2 \sum_{n=0}^2 d_n F_n(k), \quad F_n(k) = \frac{1}{\lambda} f_n\left(\frac{k}{\lambda}\right), \quad (\text{S42})$$

where we use our default  $\lambda = 0.55 \text{ GeV}$  and the functional basis  $f_n(x)$  is generated from  $Y_{\text{exp}}(x, p)$ . For our central results we use  $p = 2$ , corresponding to linear scaling  $F_n(k) \sim k$  for  $k \rightarrow 0$ , and for the uncertainties we also use  $p = 4$ . The  $d\Gamma_{g_{27,n}}/dE_\gamma$  in Eq. (S1) are obtained by inserting the basis expansion in Eq. (S42) into Eq. (S41).

At present, we have no sensitivity to determine the basis coefficients  $d_i$  from the data. Instead, we determine  $d_0$  and  $d_1$  for a given value of  $d_2$  from the norm and first moment of  $g_{27}(k)$ , which are given by

$$M^0[g_{27}] = \widehat{\lambda}_2, \quad M^1[g_{27}] = \frac{\widehat{\rho}_2}{2}. \quad (\text{S43})$$

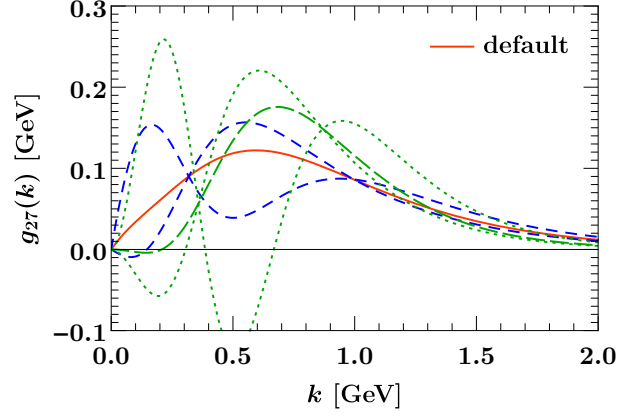


FIG. S4. Variations for  $g_{27}(k)$  for fixed norm and first moment. The solid orange line shows the default choice ( $p = 2$ ,  $d_2 = 0$ ), the blue dashed lines the  $d_2$  variations, and the green lines the different basis ( $p = 4$ ) for  $d_2 = 0$  (long-dashed) and with  $d_2$  varied (dotted).

For our central results we set  $d_2 = 0$ , and to estimate the uncertainties we vary  $d_2$  by an  $\mathcal{O}(1)$  amount to provide a reasonably large variation in the shape of  $g_{27}(k)$ . The variations for  $g_{27}(k)$  for fixed norm and first moment are illustrated in Fig. S4, with the solid orange line showing the default central choice.

The main impact of this contribution is due to the norm of  $g_{27}(k) \sim \hat{\lambda}_2$ , which reproduces the well-known  $\mathcal{O}(\lambda_2/m_c^2)$  correction to the total rate [49–51]. The central values and uncertainties used for  $\hat{\lambda}_2$  and  $\hat{\rho}_2$  are discussed in Sec. E3. The uncertainties due to the unknown shape of  $g_{27}(k)$  beyond its norm and first moment are much smaller than the fit uncertainties. They change the extracted  $|C_7^{\text{incl}}|$  by at most 0.25% and  $m_b^{1S}$  by 4 MeV, and are thus irrelevant at the present level of accuracy and can be neglected. With more available data in the future, the  $d_n$  coefficients could also be included in the fit and constrained by the data.

### E. Numerical inputs

Here, we collect all numerical input values entering in our analysis. The following values are taken from Ref. [52]:

$$\begin{aligned}
 \alpha_s^{(5)}(m_Z) &= 0.1181, & \alpha_{\text{em}}(0) &= 1/137.036, & G_F &= 1.1663787 \times 10^{-5}, \\
 m_Z &= 91.1876 \text{ GeV}, & m_W &= 83.379 \text{ GeV}, & m_t &= 173 \text{ GeV}, \\
 \overline{m}_c(\overline{m}_c) &= (1.27 \pm 0.02) \text{ GeV}, & \overline{m}_b(\overline{m}_b) &= (4.18_{-0.02}^{+0.03}) \text{ GeV}, \\
 |V_{td}| &= 0.00896_{-0.00023}^{+0.00024}, & |V_{ts}| &= 0.04133 \pm 0.00074, & |V_{tb}| &= 0.999105 \pm 0.000032, \\
 m_B &= 5.279 \text{ GeV}, & \tau_B &= 1.581 \text{ ps}, \\
 \Delta m_B &= 45.22 \text{ MeV}, & \Delta m_D &= 141.315 \text{ MeV},
 \end{aligned} \tag{S44}$$

where  $m_B$ ,  $\Delta m_B = m_{B^*} - m_B$ ,  $\Delta m_D = m_{D^*} - m_D$  are averaged over charged and neutral mesons.

#### 1. Wilson coefficients

At and above the split-matching scale  $\mu_0 = 4.7 \text{ GeV}$ , we always use the exact 4-loop running of  $\alpha_s(\mu)$  with  $n_f = 5$  flavors. To obtain the SM values of the Wilson coefficients at  $\mu_0$ , we start from the full NNLO  $\mathcal{O}(\alpha_s^2)$  boundary conditions [81, 82] at the weak scale  $\mu_{\text{weak}} = 160 \text{ GeV}$  and evolve them down to  $\mu_0$  with the anomalous dimensions up to  $\mathcal{O}(\alpha_s^3)$  [59, 83–88]. To perform the evolution, we use the exact numerical solution of the coupled RGE system. For the boundary conditions we convert the above top-quark pole mass to  $\overline{m}_b(160 \text{ GeV}) = 163.3 \text{ GeV}$ . For  $\overline{C}_{7,8}(\mu_0)$  we evolve  $\overline{m}_b(\overline{m}_b)$  to  $\mu_0$  with 4-loop running and for  $\hat{m}_b$  we use  $\hat{m}_b^{\text{fix}} = 4.7 \text{ GeV}$  (see Sec. E2 below).

The resulting NNLO Wilson coefficients  $\mathcal{C}_i$  are given by

$$\begin{aligned}
 \mathcal{C}_1 &= -0.27189, & \mathcal{C}_2 &= 1.00967, & \mathcal{C}_3 &= -5.195 \times 10^{-3}, & \mathcal{C}_4 &= -8.060 \times 10^{-2}, \\
 \mathcal{C}_5 &= 3.611 \times 10^{-4}, & \mathcal{C}_6 &= 9.375 \times 10^{-4}, & \mathcal{C}_7 &= -0.25594, & \mathcal{C}_8 &= -0.13801.
 \end{aligned} \tag{S45}$$

For theory predictions at  $\mu_0$ ,  $\mu_b$ , and below, we treat the coefficients as fixed input values, i.e., we do not expand them in  $\alpha_s$  against the perturbative corrections they multiply. Varying  $\mu_{\text{weak}} = 160 \text{ GeV}$  by a factor of two has very little impact on the  $C_i(\mu_0)$ . In particular, the combinations of  $C_{i \neq 7}$  that enter the theory predictions for the photon energy spectrum via Eq. (3) only vary below the percent level when varying  $\mu_{\text{weak}}$ , so we can safely neglect their uncertainties and keep their values fixed in the fit. Similarly, their uncertainties are irrelevant for the SM prediction of  $C_7^{\text{incl}}$ .

## 2. $\alpha_s$ , $\hat{m}_b$ , $\hat{m}_c$

As discussed in Secs. B and C, we integrate out both bottom and charm quarks at the scale  $\mu_0 = \mu_b = 4.7 \text{ GeV}$ . At  $\mu_b$  and below we then always use the 3-loop running for  $\alpha_s(\mu)$ , consistent with the NNLL resummation, with  $n_f = 3$  flavors. As the starting value we use  $\alpha_s^{(3)}(\mu_b) = 0.207$ , which is obtained as follows. We first use  $\alpha_s^{(5)}$  to evolve  $\overline{m}_b(\overline{m}_b)$  to  $\overline{m}_b(\mu_b)$  and use it to decouple the  $b$  quark at  $\mu_b$  to obtain  $\alpha_s^{(4)}(\mu_b)$ . Then, we use  $\alpha_s^{(4)}$  to evolve  $\overline{m}_c(\overline{m}_c)$  to  $\overline{m}_c(\mu_b)$  and use it to decouple the  $c$  quark at  $\mu_b$  to obtain  $\alpha_s^{(3)}(\mu_b)$ . The decoupling and running of the  $\overline{\text{MS}}$  masses is performed at 4 loops using the RunDec package [89]. The uncertainties on  $\overline{m}_b(\overline{m}_b)$  and  $\overline{m}_c(\overline{m}_c)$  are negligible for this purpose, only affecting the result in the 4th digit.

Several perturbative ingredients, such as the hard-matching coefficient  $h_s$  and the nonsingular corrections  $W_{ij}^{\text{ns}}$ , depend on the value of  $\hat{m}_b$ . As a result, the perturbative fit inputs  $d\Gamma_{ij, mn}/dE_\gamma$  in Eq. (S1) have a mild dependence on  $\hat{m}_b$ , which is subleading compared to the dominant dependence entering through the shape function. To be able to precompute the perturbative inputs, for simplicity we use a fixed value  $\hat{m}_b = 4.7 \text{ GeV}$  obtained from  $\overline{m}_b(\overline{m}_b) = 4.18 \text{ GeV}$  for their computation. We have checked that changing this value by  $\pm 50 \text{ MeV}$ , which also covers our final fit result for  $m_b^{1S}$ , has a negligible impact on the fit. (For the  $d\Gamma_{77}$  terms it changes the fit results for  $|C_7^{\text{incl}}|$  by 0.15% and for  $m_b^{1S}$  by less than 1 MeV.)

The corrections  $\sim C_{1,2}$  also require a value for the charm-quark mass  $\hat{m}_c$ . In fact, the main dependence in the perturbative inputs on both  $\hat{m}_b$  and  $\hat{m}_c$  comes from the dependence on  $\rho = \hat{m}_c^2/\hat{m}_b^2$  in  $w_{12}^{\text{ns}}$  and  $w_{22}^{\text{ns}}$ . (The sensitivity of  $h_s$  on the precise value of  $\rho$  is negligible.) While  $\overline{m}_c(\overline{m}_c)$  is known precisely, the perturbative scheme to use for  $m_c$  is also relevant, and this scheme dependence is only canceled by the still unknown non- $\beta_0$   $\mathcal{O}(\alpha_s^2)$  corrections. Since the difference between the bottom and charm pole masses,  $\delta_{bc} = m_b^{\text{pole}} - m_c^{\text{pole}}$ , is free of renormalons, we use  $\hat{m}_c \equiv \hat{m}_b - \delta_{bc}$  as a suitable charm-mass definition consistent with our treatment of the charm quark. We obtain a value for  $\delta_{bc}$  by converting  $\overline{m}_b(\mu_b)$  and  $\overline{m}_c(\mu_b)$  obtained above to the pole scheme and taking their difference. As expected, while the individual values for  $m_{b,c}^{\text{pole}}$  strongly depend on the order at which the conversion is performed, the resulting  $\delta_{bc} = 3.4 \text{ GeV}$  only changes by about 15 MeV when the conversion is performed at two vs. three loops and when using  $\alpha_s^{(4)}(\mu_b)$  vs.  $\alpha_s^{(5)}(\mu_b)$ . Accounting also for the uncertainties in  $\overline{m}_c(\overline{m}_c)$  and  $\overline{m}_b(\overline{m}_b)$ , we assign a conservative uncertainty of 50 MeV for  $\delta_{bc}$ .

To summarize, to compute all perturbative inputs for the fit, we use

$$\alpha_s^{(3)}(\mu = 4.7 \text{ GeV}) = 0.207, \quad \hat{m}_b = (4.70 \pm 0.05) \text{ GeV}, \quad \delta_{bc} = (3.40 \pm 0.05) \text{ GeV}, \quad \hat{m}_c = \hat{m}_b - \delta_{bc}. \quad (\text{S46})$$

## 3. $\hat{\lambda}_2$ and $\hat{\rho}_2$

The HQET parameters  $\lambda_2$  and  $\rho_2$  are needed in the moment relations for  $\mathcal{F}(k)$  in Eq. (S35), and also in the moment constraints for  $g_{27}(k)$  in Eq. (S43). Here, we discuss how we extract  $\lambda_2$  and  $\rho_2$  from the measured heavy meson masses using relations that are free of leading renormalon ambiguities.

The  $B$  and  $D$  meson masses can be expanded in  $1/m_Q$ , following the notation of Ref. [75], as

$$m_M = m_Q + \bar{\Lambda} - \frac{\lambda_1 + d_M C_G(m_Q, \mu) \lambda_2(\mu)}{2m_Q} + \frac{\rho_1 + d_M \rho_2}{4m_Q^2} - \frac{\mathcal{T}_1 + \mathcal{T}_3 + d_M(\mathcal{T}_2 + \mathcal{T}_4)}{4m_Q^2} + \mathcal{O}\left(\frac{\Lambda_{\text{QCD}}^4}{m_Q^3}, \frac{\alpha_s \Lambda_{\text{QCD}}^3}{m_Q^2}\right), \quad (\text{S47})$$

where  $m_M$  with  $M = B, B^*, D, D^*$  are the masses of the lightest pseudoscalar and vector mesons containing the heavy quark  $Q$  with  $d_{B,D} = 3$  and  $d_{B^*, D^*} = -1$  for the pseudoscalar and vector mesons. Here,  $m_Q$  is the pole mass of the heavy quark  $Q$ . We have included the  $\overline{\text{MS}}$  Wilson coefficient  $C_G(m_Q, \mu)$  for the scale-dependent chromomagnetic  $\overline{\text{MS}}$  matrix element  $\lambda_2(\mu)$ , but neglect Wilson coefficients for the terms of higher order in  $1/m_Q$ .

Only three linear combinations of  $\mathcal{T}_i$  appear in expressions for inclusive  $B$  decays [90], which are

$$\begin{aligned}\tau_1 &= \mathcal{T}_1 - 3\mathcal{T}_4 = (0.161 \pm 0.122) \text{ GeV}^3, \\ \tau_2 &= \mathcal{T}_2 + \mathcal{T}_4 = (-0.017 \pm 0.062) \text{ GeV}^3, \\ \tau_3 &= \mathcal{T}_3 + 3\mathcal{T}_4 = (0.213 \pm 0.102) \text{ GeV}^3.\end{aligned}\tag{S48}$$

The numerical values are taken from a global fit in the  $1S$  scheme to semileptonic  $B \rightarrow X_c \ell \nu$  and radiative  $B \rightarrow X_s \gamma$  moments [20]. The parameters are only weakly correlated and are insensitive to whether or not the radiative moments are included in the fit [91].

Denoting  $\Delta m_B = m_{B^*} - m_B$  and  $\Delta m_D = m_{D^*} - m_D$ , we have

$$\begin{aligned}\Delta m_B &= 2C_G(m_b, \mu) \frac{\lambda_2(\mu)}{m_b} - \frac{\rho_2 - \tau_2}{m_b^2} = 2C_G^R(\widehat{m}_b, \mu, R) \frac{\lambda_2(\mu)}{\widehat{m}_b} - \frac{\widehat{\rho}_2(R) - \tau_2}{\widehat{m}_b^2}, \\ \Delta m_D &= 2C_G(m_c, \mu) \frac{\lambda_2(\mu)}{m_c} - \frac{\rho_2 - \tau_2}{m_c^2} = 2C_G^R(\widehat{m}_c, \mu, R) \frac{\lambda_2(\mu)}{\widehat{m}_c} - \frac{\widehat{\rho}_2(R) - \tau_2}{\widehat{m}_c^2}.\end{aligned}\tag{S49}$$

The first equality in each of these expressions involves the  $\overline{\text{MS}}$  Wilson coefficient  $C_G(m_Q, \mu)$ , which has a  $\mathcal{O}(\Lambda_{\text{QCD}}/m_Q)$  renormalon ambiguity that is canceled by a corresponding ambiguity in  $\rho_2$ . In the second equalities, we switched to renormalon-free quantities, where the Wilson coefficient  $C_G^R(\widehat{m}_Q, \mu, R)$  and matrix element  $\widehat{\rho}_2(R)$  are defined in the renormalon-free MSR scheme [92]. To evaluate the Wilson coefficient  $C_G$  or  $C_G^R$  we use the fixed-order results evaluated at  $\mu = R = \sqrt{\widehat{m}_b \widehat{m}_c}$  (which are known to 3-loops [93]) and evolve down to  $\mu = R = 1 \text{ GeV}$ , using the  $\overline{\text{MS}}$  RGE or RRGE [92, 94] respectively. To highlight the improvement obtained in the renormalon-free scheme, we note that in  $\overline{\text{MS}}$  we have  $C_G(\widehat{m}_c, \sqrt{\widehat{m}_b \widehat{m}_c}) = \{1.26, 1.46, 1.69\}$  at 1, 2, and 3-loops respectively, whereas in MSR the results exhibit convergence with  $C_G^R(\widehat{m}_c, \sqrt{\widehat{m}_b \widehat{m}_c}, \sqrt{\widehat{m}_b \widehat{m}_c}) = \{0.991, 1.034, 1.045\}$  at 1, 2, and 3-loops. Inverting the MSR results in Eq. (S49), we obtain

$$\begin{aligned}\lambda_2(\mu) &= \frac{1}{2} \frac{\widehat{m}_b^2 \Delta m_B - \widehat{m}_c^2 \Delta m_D}{\widehat{m}_b C_G^R(\widehat{m}_b, \mu, R) - \widehat{m}_c C_G^R(\widehat{m}_c, \mu, R)}, \\ \widehat{\rho}_2(R) - \tau_2 &= \frac{\widehat{m}_c C_G^R(\widehat{m}_c, \mu, R) \widehat{m}_b^2 \Delta m_B - \widehat{m}_b C_G^R(\widehat{m}_b, \mu, R) \widehat{m}_c^2 \Delta m_D}{\widehat{m}_b C_G^R(\widehat{m}_b, \mu, R) - \widehat{m}_c C_G^R(\widehat{m}_c, \mu, R)}.\end{aligned}\tag{S50}$$

With the input values for the meson and quark masses from above we then find

$$\lambda_2(\mu = 1 \text{ GeV}) = (0.128 \pm 0.005) \text{ GeV}^2, \quad \widehat{\rho}_2(R = 1 \text{ GeV}) - \tau_2 = (0.110 \pm 0.052) \text{ GeV}^3.\tag{S51}$$

Here the uncertainty in  $\widehat{\rho}_2$  comes from varying the low scale down to 0.8 GeV and up to 1.3 GeV, which in MSR provides an estimate of the size of neglected  $\mathcal{O}(\Lambda_{\text{QCD}}/\widehat{m}_{b,c}^3)$  corrections in Eq. (S49). We then combine this in quadrature with an estimate of  $\mathcal{O}(\alpha_s \Lambda_{\text{QCD}}^3/\widehat{m}_c^2)$  corrections to Eq. (S49). (The additional variation of the starting scale by a factor of two has a very small effect.) These uncertainty estimates for higher-order terms are then propagated to obtain the uncertainty quoted for  $\lambda_2(\mu = 1 \text{ GeV})$  in Eq. (S51).

Finally, for the parameters appearing in the moment relations we take  $\mu = \sqrt{\widehat{m}_b \widehat{m}_c} \simeq 2.5 \text{ GeV}$  for  $\lambda_2(\mu)$ , still taking  $R = 1$  for  $\widehat{\rho}_2(R)$ , and hence will use

$$\widehat{\lambda}_2 = \lambda_2(2.5 \text{ GeV}) + \frac{\tau_3}{3\widehat{m}_b} = (0.135 \pm 0.009) \text{ GeV}^2, \quad \widehat{\rho}_2(1 \text{ GeV}) = (0.093 \pm 0.081) \text{ GeV}^3,\tag{S52}$$

where we added the uncertainties from Eqs. (S48) and (S51) in quadrature.