Review

Challenges in semileptonic $B$ decays

P. Gambino$^{1,4}$, A. S. Kronfeld$^2$, M. Rotondo$^3$, C. Schwanda$^4$, F. Bernlochner$^5$, A. Bharucha$^6$, C. Bozzi$^7$, M. Calvi$^8$, L. Cao$^9$, G. Ciezarek$^9$, C. T. H. Davies$^{10}$, A. X. El-Khadra$^{11}$, S. Hashimoto$^2$, M. Jung$^1$, A. Khodjamirian$^{13}$, Z. Ligi$^{14}$, E. Lunghi$^5$, V. Lüth$^6$, T. Mannel$^{13}$, S. Meinel$^{17}$, G. Paz$^{18}$, S. Schacht$^{1,19}$, S. Simula$^{20}$, W. Sutcliffe$^5$, A. Vaquero Avilés-Casco$^{21}$

$^1$ Physics Department, University of Turin and INFN Sezione di Torino, 10125 Turin, Italy
$^2$ Theoretical Physics Department, Fermilab, Batavia, IL 60510-5011, USA
$^3$ Laboratori Nazionali di Frascati, INFN, 00044 Frascati, RM, Italy
$^4$ Institut für Hochenergiephysik, Österreichische Akademie der Wissenschaften, 1050 Wien, Austria
$^5$ Physikalisches Institut der Rheinischen Friedrich-Wilhelms-Universität Bonn, 53115 Bonn, Germany
$^6$ Aix Marseille Univ, Université de Toulon, CNRS, CPT, Marseille, France
$^7$ INFN, Sezione di Ferrara, Ferrara, Italy
$^8$ University of Milano Bicocca and INFN, Milano Bicocca, Milan, Italy
$^9$ European Organization for Nuclear Research (CERN), Geneva, Switzerland
$^{10}$ SUPA, School of Physics and Astronomy, University of Glasgow, Glasgow G12 8QQ, UK
$^{11}$ Department of Physics, University of Illinois, Urbana, IL 61801, USA
$^{12}$ Theory Center, High Energy Accelerator Research Organization (KEK), School of High Energy Accelerator Science (SOKENDAI), Tsukuba 305-0801, Japan
$^{13}$ Theoretische Physik 1, Naturwissenschaftlich-Technische Fakultät, Universität Siegen, 57068 Siegen, Germany
$^{14}$ Lawrence Berkeley National Laboratory, University of California, Berkeley, CA 94720, USA
$^{15}$ Physics Department, Indiana University, Bloomington, IN 47405, USA
$^{16}$ SLAC National Accelerator Laboratory, Stanford, CA 94309, USA
$^{17}$ Department of Physics, University of Arizona, Tucson, AZ 85721, USA
$^{18}$ Department of Physics and Astronomy, Wayne State University, Detroit, MI 48201, USA
$^{19}$ Department of Physics, LEPP, Cornell University, Ithaca, NY 14853, USA
$^{20}$ INFN, Sezione di Roma Tre, Rome, Italy
$^{21}$ Department of Physics and Astronomy, University of Utah, Salt Lake City, UT 84112, USA

Received: 11 June 2020 / Accepted: 16 September 2020 / Published online: 19 October 2020
© The Author(s) 2020

Abstract Two of the elements of the Cabibbo–Kobayashi–Maskawa quark mixing matrix, $|V_{ub}|$ and $|V_{cb}|$, are extracted from semileptonic $B$ decays. The results of the $B$ factories, analysed in the light of the most recent theoretical calculations, remain puzzling, because for both $|V_{ub}|$ and $|V_{cb}|$ the exclusive and inclusive determinations are in clear tension. Further, measurements in the $\tau$ channels at Belle, Babar, and LHCb show discrepancies with the Standard Model predictions, pointing to a possible violation of lepton flavor universality. LHCb and Belle II have the potential to resolve these issues in the next few years. This article summarizes the discussions and results obtained at the MITP workshop held on April 9–13, 2018, in Mainz, Germany, with the goal to develop a medium-term strategy of analyses and calculations aimed at solving the puzzles. Lattice and continuum theorists working together with experimentalists have discussed how to reshape the semileptonic analyses in view of the much higher luminosity expected at Belle II, searching for ways to systematically validate the theoretical predictions in both exclusive and inclusive $B$ decays, and to exploit the rich possibilities at LHCb.

Contents

1 Executive summary ........................................ 2
2 Heavy-to-heavy exclusive decays .......................... 3
3 Heavy-to-light exclusive .................................... 18
4 Quark masses and leptonic decays ......................... 25
5 Heavy-to-heavy inclusive .................................. 28
6 Heavy-to-light inclusive ................................... 33
7 Outlook .................................................. 38
8 References ................................................ 39

$^4$ e-mail: gambino@to.infn.it (corresponding author)
1 Executive summary

The magnitudes of two of the elements of the Cabibbo–Kobayashi–Maskawa (CKM) quark mixing matrix \([1, 2]\), \(|V_{ub}|\) and \(|V_{cb}|\), are extracted from semileptonic \(B\)-meson decays. The results of the \(B\) factories, analysed in the light of the most recent theoretical calculations, remain puzzling, because – for both \(|V_{ub}|\) and \(|V_{cb}|\) – the determinations from exclusive and inclusive decays are in tension by about \(3\sigma\). Recent experimental and theoretical results reduce the tension, but the situation remains unclear. Meanwhile, measurements in the semitauonic channels at Belle, Babar, and LHCb show discrepancies with the Standard Model (SM) predictions, pointing to a possible violation of lepton-flavor universality. LHCb and the upcoming experiment Belle II have the potential to resolve these issues in the next few years.

Thirty-five participants met at the Mainz Institute for Theoretical Physics to develop a medium-term strategy of analyses and calculations aimed at the resolution of these issues. Lattice and continuum theorists discussed with experimentalists how to reshape the semileptonic analyses in view of the much larger luminosity expected at Belle II and how to best exploit the new possibilities at LHCb, searching for ways to systematically validate the theoretical predictions, to confirm new physics indications in semitauonic decays, and to identify the kind of new physics responsible for the deviations.

Format of the workshop

The program took place during a period of five days, allowing for ample discussion time among the participants. Each of the five workshop days was devoted to specific topics: the inclusive and exclusive determinations of \(|V_{cb}|\) and \(|V_{ub}|\), semitauonic \(B\) decays and how they can be affected by new physics, as well as related subjects such as purely leptonic \(B\) decays and heavy quark masses. In the mornings, we had overview talks from the experimental and theoretical sides, reviewing the main aspects and summarizing the state of the art. In the late afternoon, we organized discussion sessions led by experts of the various topics, addressing questions that have been brought up before or during the morning talks.

Exclusive heavy-to-light decays

The \(B \to D^{(*)}\ell\nu\) decays have received significant attention in the last few years. New Belle results for the \(q^2\) and angular distributions have allowed studies of the role played by the parametrization of the form factors in the extraction of \(|V_{cb}|\). It turns out that the extrapolation to zero-recoil is very sensitive to the parametrization employed, a problem that can be solved only by precise calculations of the form factors at non-zero recoil. Until these are completed, the situation remains unclear, with repercussions on the calculation of \(R(D^*)\) as well, with diverging views on the theoretical uncertainty of present estimates based on heavy quark effective theory (HQET) expressions.

Beside a critical reexamination of these recent developments, we discussed several incremental and qualitative improvements in lattice QCD, also in baryonic decays. Though unlikely to carry much weight in determining \(|V_{cb}|\), the latter offer great opportunities to test lepton-flavor universality violation (LFUV) and lattice QCD. The discussions also addressed the fact that QCD errors are now almost as small as effects from QED. Thus, further improvement must be theoretically made by properly studying the effect of QED radiation, especially the treatment of soft photons and photons that are neither soft nor hard and their sensitivity to the meson wave functions.

Concerning studies of LFUV, we discussed the role played by higher excited charmed states in establishing new physics and the challenges that the present \(R(D^{(*)})\) measurements represent for model building.

Exclusive heavy-to-light decays

This determination of \(|V_{ub}|\) relies on nonperturbative calculations of the form factor of \(B \to \pi\ell\nu\), which is the most precise channel. We discussed the status of the light-cone sum rule (LCSR) calculations and several recent improvements in lattice QCD, in particular the most recent results from the Fermilab Lattice & MILC Collaborations and from the RBC & UKQCD Collaborations, as well as future prospects. The Fermilab/MILC calculation alone leads to a remarkably small total error on \(|V_{ub}|\), about 4%. While at present the most precise extraction of \(|V_{ub}|\) comes from \(B \to \pi\ell\nu\), it is worth considering the channel \(B_s \to K\ell\nu\) as well, because here the lattice-QCD calculations are affected by somewhat smaller uncertainties. \(B_s \to K\ell\nu\) can be accessible at Belle II in a run at the \(T(5S)\) and a precision of about 5–10% could be achieved with 1 fb\(^{-1}\). On the other hand, LHCb has an ongoing analysis of the ratio \(B(B_s \to K\ell\nu)/B(B_s \to D_s\ell\nu)\), which will provide a new determination of \(|V_{ub}/V_{cb}|\). This approach follows the success that LHCb demonstrated for semileptonic baryon decays via the precise measurement of the ratio \(B(A_b \to p\mu\nu)/(A_b \to \Lambda_c\mu\nu)\) in the high-\(q^2\) region. This measurement, combined with precise lattice-QCD calculations of the form factors, allowed the extraction of ratio \(|V_{ub}/V_{cb}|\) with an uncertainty of 7%. We discussed also other channels, in particular how to study \(B \to \pi\pi\ell\nu\) including the resonant structures. Careful studies of other heavy-to-light channels will also be crucial to improve the signal model for the inclusive \(|V_{ub}|\) measurements.
Inclusive heavy-to-heavy decays

The theoretical predictions in this case are based on an operator product expansion. Theoretical uncertainties already dominate current determinations, and better control of all higher-order corrections is needed to reduce them. In this respect, it would be important to have the perturbative-QCD corrections to the complete coefficient of the Darwin operator and to check the treatment of QED radiation in the experimental analyses. A full $O(a_s^2)$ calculation of the total width may be within reach with recently developed techniques. From the experimental point of view, new and more accurate measurements will be most welcome, in particular to better understand the correlations between different moments and moments with different cuts. A better determination of the higher hadronic mass moments and a first measurement of the forward–backward asymmetry would benefit the global fit, as would a better understanding of higher power corrections. The importance of having global fits to the moments in different schemes and by different groups has also been stressed. This calls for an update of the 1S scheme fit and could lead to a cross-check of the present theoretical uncertainties. Lattice QCD already provides inputs to the fit with the calculation of the heavy quark masses, which have been reviewed. New developments discussed at the workshop may soon be able to provide additional information that can be fed into the fits, such as constraints on the heavy-quark quantities $\mu_G^2$ and $\mu_G^2$. The two main approaches are (i) computing inclusive rates directly with lattice QCD and (ii) using the heavy quark expansion for meson masses, precisely computed at different quark mass values. The state of theoretical calculations for inclusive semitauonic decays has also been discussed, as they represent an important cross-check of the LFUV signals.

Inclusive heavy-to-light decays

This determination is based on various well-founded theoretical methods, most of which agree well. The 2017 endpoint analysis by BaBar seems to challenge this consolidated picture, suggesting discrepancies between some of the methods and a lower value of $|V_{ub}|$. For the future, the complete NNLO corrections in the full phase space should be implemented and the various methods should be upgraded in order to make the best use of the Belle II differential data based on much higher statistics. These data will make it possible to test the various methods and to calibrate them, as they will contain information on the shape functions. The SIMBA and NNVub methods seem to have the potential to fully exploit the $B \to X_u\ell\nu$ (and possibly radiative) measurements through combined fits to the shape function(s) and $|V_{ub}|$. The separation of $B^+ \to \tau\nu$ and $B^0$ in the experimental analyses will certainly help to constrain weak annihilation, but the real added value of Belle II could be precise measurements of kinematic distributions in $M_X$, $q^2$, $E_\ell$, etc. A detailed measurement of the high $q^2$ tail might be very useful, also in view of attempts to check quark-hadron duality. Experimentally, better hybrid (inclusive + exclusive) Monte Carlos are badly needed; $s$-$\bar{s}$ popping should be investigated to develop a better understanding of kaon vetos. The $b \to c$ background will be measured better, which will benefit these analyses.

Leptonic decays

The measurement of $B \to \tau\nu$ is not yet competitive with semileptonic decays for measuring $|V_{ub}|$, because of a 20% error on the rate. Belle II will improve on this. The corresponding lattice-QCD calculation is however very precise, with an error below 1%, according to the 2019 report from FLAG [3] and based mainly on a result from Fermilab/MILC that was presented at the workshop. That said, the mode is useful today to model builders trying to understand new physics explanations of the tension between inclusive and exclusive determinations of $|V_{ub}|$. Belle II will also access $B \to \mu\nu(\gamma)$ with the possibility to reach an uncertainty on the branching fraction of about 5% with $50 ab^{-1}$, allowing for a new determination of $|V_{ub}|$ in the long term. We discussed also the LHCb contribution to leptonic decays with the process $B \to \mu\mu\nu\nu$, where two of the muons come from virtual $\gamma$ or light vector meson decays. A study of this channel has been published in [4] and a very stringent upper limit obtained, inconsistent with the existing branching fraction predictions, indicating the need for reliable theoretical calculations.

2 Heavy-to-heavy exclusive decays

The aim of this section is to present an overview of $b \to c$ exclusive decays. After an introduction to the parametrization of the relevant form factors between hadronic states we describe the status of current lattice QCD calculations with particular focus on $B \to D^*$ and $A_b \to \Lambda_c$. Next, we discuss experimental measurements of $B \to D^{(*)}$ semileptonic decays with special focus on the ratios $R(D^{(*)})$, and several phenomenological aspects of these decays: the extraction of $V_{cb}$, theoretical predictions for $R(D^{(*)})$, the role of $B \to D^{(*)}$ transitions and constraints on new physics. We also briefly discuss the information that is required to reproduce results presented in experimental analyses and to incorporate older measurements into approaches based on modern form factor parametrizations. We conclude with the description of HAMMER, a tool designed to more easily calculate the change in signal acceptancies, efficiencies and signal yields in the presence of new physics.
2.1 Parametrization of the form factors

In this section, we introduce the form factors for the hadronic matrix elements that arise in semileptonic decays. Several different notations appear in the literature, often using different conventions depending on whether the final-state meson is heavy (e.g., $D$) or light (e.g., $\pi$). A general decomposition relies, however, only on Lorentz covariance and other symmetry properties of the matrix elements. As discussed below, it is advantageous to choose the Lorentz structure so that the form factors have definite parity and spin.

In this spirit, let us consider the matrix elements for a meson decay $B\rightarrow X^\ell \ell\bar{\nu}$, where the quark content of the $B$ is $bl$ with $l$ a light quark ($u$, $d$, or $s$), and the quark content of the $X$ is $\bar{q}l$ where $q$ can be either a light quark or the $c$ quark. The desired decomposition can be written as

\[
\langle X(p')|S|B(p)\rangle = \frac{M^2-m^2}{m_b-m_q} f_0(q^2),
\]

\[
\langle X(p')|V|B(p)\rangle = \left[ (p+p')^\mu - \frac{M^2-m^2}{q^2} q^\mu \right] f_+(q^2) + \frac{M^2-m^2}{q^2} q^\mu f_0(q^2),
\]

\[
\langle X(p')|T|B(p)\rangle = \frac{2 m}{M+m} (e^\ast \cdot q) A_0(q^2),
\]

\[
\langle X(p')|A|B(p)\rangle = \frac{2i}{M+m} \epsilon^{\mu\nu\alpha\beta} e_\alpha q_\beta p_\mu p_\nu V(q^2).
\]

(2.6) with $q^\mu$ and using the appropriate Ward identities shows that the scalar form factor, $f_0$, and pseudoscalar form factor, $A_0$, appear in the vector and axial vector transitions. The $J^P$ quantum numbers of the form factors are given in Table 1. The tensor form factors in Eqs. (2.3) and (2.7) appear in extensions of the Standard Model.

One can impose bounds on the shape of these form factors by using QCD dispersion relations for a generic decay $H_0 \rightarrow H_q \ell \ell\bar{\nu}$. Since the amplitude for production of $H_bH_q$ from a virtual $W$ boson is determined by the analytic continuation of the form factors from the semileptonic region of momentum transfer $m_q^2 < q^2 < M^2-m^2$ to the pair production region $q^2 \geq M^2+m^2$, one can find constraints in the pair-production region, amenable to perturbative QCD calculations, and then propagate the constraint to the semileptonic region by using analyticity. The result of this process applied to the form factors is the model-independent Boyd–Griessmann–Lebed (BGL) parametrization [5,6], which expands a form factor $F(z)$ in the dimensionless variable $z$ as

\[
F(z) = \frac{1}{B_F(z)\phi_F(z)} \sum_{j=0}^\infty a_j z^j,
\]

where $t_{\pm} = (M \pm m)^2$, $B_F(z)$ are known as the Blaschke factors, which incorporate the below- or near-threshold [7] poles in the s-channel process $\ell\nu \rightarrow \bar{B}X$, and $\phi_F(z)$ is called the outer function. The poles, and hence the Blaschke factor, depend on the spin and the parity of the intermediate state, which is why it is useful to use fixed $J^P$ for the form factors. See Sect. 3.5.5 for more details. Of course, in practical applications the series (2.8) is truncated at some power $z^n$. By taking certain linear combinations of form factors with the same spin and parity one obtains the BGL notation for the helicity amplitudes,

\[
f_+^{\text{BGL}} = f_+,
\]

\[
f_0^{\text{BGL}} = (M^2-m^2) f_0.
\]
\[ g = \frac{2}{M + m} V, \]  
\[ f = (M + m) A_1, \]  
\[ \mathcal{F}_1 = M(M + m)(w - r) A_1 - \frac{2Mm(w^2 - 1)}{1 + r} A_2, \]  
\[ \mathcal{F}_2 = 2A_0, \]

leaving aside the (BSM) tensor form factors. Here the velocity transfer
\[ w = v_M \cdot v_m = \frac{M^2 + m^2 - q^2}{2Mm}, \]

with \( v_M = p/M \) and \( v_m = p'/m, \) is often used in heavy-to-heavy decays. For heavy-to-light decays it can be helpful to work with the energy of the daughter meson in the rest frame of the parent, i.e.,
\[ E = p' \cdot v_M = \frac{M^2 + m^2 - q^2}{2M}. \]

These form factors are subject to three kinematic constraints, namely
\[ (M^2 - m^2) f^\text{BGL}_B(q^2 = 0) = f^\text{BGL}_B(q^2 = 0), \]  
\[ (M - m) f(q^2 = q^2_{\text{max}}) = \mathcal{F}_1(q^2 = q^2_{\text{max}}), \]  
\[ \frac{2}{M^2 - m^2} \mathcal{F}_1(q^2 = 0) = \mathcal{F}_2(q^2 = 0), \]

where \( q^2_{\text{max}} = (M - m)^2, \) corresponding to \( w = 1 \) and \( E = m. \)

The variable \( z \) can also be expressed via \( w, \)
\[ z = \frac{\sqrt{w + 1} - \sqrt{2N}}{\sqrt{w + 1} + \sqrt{2N}}, \]

where \( N = (t_+ - t_0)/(t_+ - t_-), \) is real for \( q^2 \leq (M + m)^2, \) and it becomes a pure phase beyond that limit. The constant \( t_0 \) defines the point at which \( z = 0. \) Often \( t_0 = t_-, \) one end of the kinematic range, so \( z \) ranges from 0 at maximum \( q^2 \) to \( z_{\text{max}} = (1 - \sqrt{r})/(1 + \sqrt{r})^2 \) when \( m_1 \approx 0. \) Alternatively, the choice \( t_0 = (M + m)(\sqrt{M} - \sqrt{m})^2 \) sets \( z = 0 \) exactly in the middle of the kinematic range. Even for \( B \to \pi \ell \nu, \) \( z \) is always a small quantity, which ensures a fast convergence of the power series defined in (2.8).

Unitarity constraints from the QCD dispersion relations are translated into constraints for the coefficients of the BGL expansion. In general,
\[ \sum_{j=0}^{\infty} (a_j^F)^2 \leq 1, \]
for each form factor \( F, \) but in the particular case of \( B \to D^* \ell \bar{\nu} \) the bound becomes
\[ \sum_{j=0}^{\infty} \left( a_j^f + (a_j^{\mathcal{F}_1})^2 \right) \leq 1, \]
for the \( f \) and \( \mathcal{F}_1 \) form factors, because they have the same quantum numbers. These bounds are known as the weak unitarity constraints.

A modification of the BGL parametrization by Bourrely, Lellouch and Caprini (BCL) [8] is often chosen in analyses of heavy-to-light decays. The BCL parametrization improves BGL by fixing two artifacts of the truncated BGL series. In particular, it removes an unphysical singularity at the pair production threshold and corrects the large \( q^2 \) behavior (see [9,10]) in the functional form. These two modifications improve the convergence of the expansion. However, the kinematic range is much more constrained in the heavy-to-heavy case, and lies farther from both the production threshold and the large \( q^2 \) region. Therefore, the presence of far singularities or an incorrect asymptotic behavior are not expected to spoil the \( z \)-expansion in that case.

In the heavy-to-heavy case, one can sharpen the weak unitarity constraints on the BGL coefficients using heavy quark symmetry (HQS) which relates the different \( B^{(*)} \to D^{(*)} \ell \bar{\nu} \) channels and their form factors: each form factor is either proportional to the Isgur–Wise function \( \xi(w) \) or zero. Using heavy quark effective theory (HQET) one can improve the precision by introducing radiative and power (i.e. in inverse powers of the heavy masses) corrections. Then we can define any form factor in such a way that it admits the expansion in both \( \alpha \) and the heavy quark masses
\[ F(w) = \xi(w) \left( 1 + c_0 \frac{\alpha}{\pi} + c_1 \frac{A_{\text{QCD}}}{m_b} + c_2 \frac{A_{\text{QCD}}}{m_c} + \cdots \right). \]

These expansions can be used to link the \( z \) expansion coefficients of different form factors, leading to the so-called strong unitarity constraints [11,12]. The power corrections depend on subleading Isgur–Wise functions that have been estimated with QCD sum rules [13–15].

Previous analyses of \( B \to D^* \ell \nu \) have used the Caprini–Lellouch–Neubert (CLN) parametrization [11]. CLN employ a notation for the form factors that satisfies (2.24),\(^2\)
\[ S_1^{\text{CLN}} = \frac{f_0^{\text{BGL}}}{M^2(1 - r) \sqrt{r}(1 + w)}, \quad P_1^{\text{CLN}} = \frac{\sqrt{r}}{1 + r} F_2, \]
\[ V_1^{\text{CLN}} = \frac{2\sqrt{r}}{1 + r} f^{\text{BGL}}_+ , \quad V_4^{\text{CLN}} = M \sqrt{r} g, \]

\(^2\) See [12] for a comprehensive table including other decays.
\[ A_{1}^{\text{CLN}} = \frac{f}{M\sqrt{1 + w}}, \quad A_{3}^{\text{CLN}} = \frac{F_{l}}{M^{2}(1 - r)\sqrt{1 + w}}, \]  
\[ R_{1}^{\text{CLN}} = \frac{V_{4}^{\text{CLN}}}{A_{1}^{\text{CLN}}}, \quad R_{2}^{\text{CLN}} = \frac{w - r}{w - 1} - \frac{(1 - r)A_{3}^{\text{CLN}}}{A_{1}^{\text{CLN}}}. \]  
\[ (2.27) \]

where the letter naming the form factor (S, P, V, and A) encodes its quantum numbers (scalar, pseudoscalar, vector, and axial vector), and \( R_{1,2}^{\text{CLN}} \) are two convenient ratios of form factors. Sometimes the ratio \( R_{0}^{\text{CLN}} = P_{1}^{\text{CLN}}/A_{1}^{\text{CLN}} \) is considered.

In the CLN parametrization the strong unitarity constraints obtained with HQET at NLO are used to remove some of the coefficients of the \( z \) expansion. Further, specific numerical coefficients are introduced in a polynomial in \( w \) for \( R_{1,2}^{\text{CLN}} \). The numerical values were determined using information available in 1997, which has been partly superseded but not updated. The numerical values also omit error estimates (which were discussed in the original CLN paper [11], although in an optimistic manner) because at the time the experimental statistical errors dominated, which is no longer the case. A consensus of the workshop recommends that CLN no longer be used, certainly not unless the numerical coefficients have been updated and the ensuing theoretical uncertainties are accounted for. It is better to use a general form of the \( z \) expansion.

HQET naturally presents another basis for the form factors of the \( B \rightarrow D^{(*)}\bar{\ell}\ell \) processes. Using velocities instead of momenta and otherwise mimicking the Lorentz structure of Eqs. (2.2), (2.5), and (2.6), the notation is \( h_{+} \) and \( h_{-} \) for \( B \rightarrow D^{(*)}\bar{\ell}\ell \), and \( h_{Y}^{+} \) and \( h_{Y}^{-} \) for \( B \rightarrow D^{(*)}\bar{\epsilon}^{\pm} \). In the heavy quark limit, these form factors tend to

\[ h_{X}(w) = \eta(\alpha_{s})\xi(\ell) + O\left(\frac{A_{QCD}}{m_{b,c}}\right), \]  
\[ (2.29) \]

for \( X = +, A_{1}, A_{3}, V \), and

\[ h_{Y}(w) = \beta(\alpha_{s})\xi(\ell) + O\left(\frac{A_{QCD}}{m_{b,c}}\right), \]  
\[ (2.30) \]

with \( Y = -, A_{2} \). Here \( \eta(\alpha_{s}) = 1 + O(\alpha_{s}) \), while \( \beta(\alpha_{s}) = O(\alpha_{s}) \). In this representation, the identities expressed in Eqs. (2.18)–(2.20) become evident.

Finally, for the case of a baryonic decay \( A_{b} \rightarrow Y_{(q)}\ell\bar{\ell} \), with \( Y = p, A_{c} \), we define

\[ \langle Y(p')|V^{\mu}|A_{b}(p)\rangle = \bar{u}_{q}(p') \frac{M - m_{b}}{m_{b} - m_{q}}f_{0}(q^{2})u_{b}(p), \]  
\[ (2.31) \]

\[ \langle Y(p')|P|A_{b}(p)\rangle = \bar{u}_{q}(p')\gamma_{5} \frac{M + m_{b} + m_{q}}{m_{b} + m_{q}}g_{0}(q^{2})u_{b}(p), \]  
\[ (2.32) \]

where \( M \) is the mass of the \( A_{b}, m \) is the mass of the daughter baryon and \( s_{\pm} = (M \pm m)^{2} - q^{2} \). The \( z \) expansions for the baryonic form factors employed in Ref. [16] use trivial outer functions and do not impose unitarity bounds on the coefficients of the expansion. As a result, the coefficients are unconstrained and reach values as high as \( \sim 10 \). See also Sect. 3.5.5.

2.2 Heavy-to-heavy form factors from lattice QCD

The lattice QCD calculation of the form factors for the semileptonic decay of a hadron uses two- and three-point correlation functions, which are constructed from valence quark propagators obtained by solving the Dirac equation on a set of gluon field configurations. Averaging the correlation functions over the gluon field configurations then yields the appropriate Feynman path integral. The two-point correlation functions give the amplitude for a hadron to be created at the time origin and then destroyed at a time \( T \). The three-point correlation functions include the insertion of a current \( J \) at time \( t \) on the active quark line, changing the active quark from one flavor to another. Usually calculations are performed with the initial hadron at rest. Momentum is inserted at the current so that a range of momentum transfer, \( q \), from initial to final hadron can be mapped out.

The three-point correlation functions (for multiple \( q \) values) and the two-point correlation functions (with multiple momenta in the case of the final-state hadron) are fit as functions of \( t \) and \( T \) to determine the matrix elements of the currents between initial and final hadrons that yield the required form factors. An important point here is that the initial and final hadrons that we focus on are the ground-state particles.
in their respective channels. However, terms corresponding to excited states must be included in the fits in order to make sure that systematic effects from excited-state contamination are taken into account in the fit parameters that yield the ground-state to ground-state matrix element of $J$ and hence the form factors.

Statistical uncertainties in the form factors obtained obviously depend on the numbers of samples of gluon-field configurations on which correlation functions are calculated. To improve statistical accuracy further, calculations usually include multiple positions of the time origin for the correlation functions on each configuration. The numerical cost of the calculation of quark propagators falls as the quark mass increases and so heavy ($b$ and $c$) quark propagators are typically numerically inexpensive. The accompanying light quark propagators for heavy-light hadrons are much more expensive, especially if $u/d$ quarks with physically light masses are required. It is this cost that limits the statistical accuracy that can be obtained, especially since the statistical uncertainty for a heavy-light hadron correlation function (on a given number of gluon field configurations) also grows as the separation in mass between the heavy and light quarks increases.

A key issue for heavy-to-heavy ($b$ to $c$) form factor calculations is how to handle heavy quarks on the lattice. Discretization of the Dirac equation on a space-time lattice gives systematic discretization effects that depend on powers of the quark mass in lattice units. The size of these effects depends on the value of the lattice spacing and the power with which the effects appear (i.e., the level of improvement used in the lattice Lagrangian).

Since the $b$ quark is so heavy, its mass in lattice units will be larger than 1 on all but the finest lattices ($a < 0.05$ fm) currently in use. Highly-improved discretizations of the Dirac equation are needed to control the discretization effects. A good example of such a lattice quark formalism is the highly improved staggered quark (HISQ) action developed by HPQCD [17] for both light and heavy quarks with discretization errors appearing at $O(\alpha_s (am)^2)$ and $O((am)^4)$. An alternative approach is to make use of the fact that $b$ quarks are nonrelativistic inside their bound states. This means that a discretization of a nonrelativistic action (NRQCD) can be used, expanding the action to some specified order in the $b$ quark velocity. Discretization effects then depend on the scales associated with the internal dynamics and these scales are all much smaller than the $b$ quark mass. Relativistic effects can be included and discretization effects corrected at the cost of complicating the action with additional operators. A third possibility is to start from the Wilson quark action and improved versions of it but to tune the parameters (such as the quark mass) using a nonrelativistic dispersion relation for the meson, which is known as the Fermilab method [18]. This removes the leading source of mass-dependent discretization effects, while retaining a discretization that connects smoothly to the continuum limit. Again, improved versions of this approach (such as the Oktay–Kronfeld action [19]) include additional operators.

The $c$ quark has a mass larger than $\Lambda_{QCD}$ but within lattice QCD it can be treated successfully as a light quark because its mass in lattice units is less than 1 on lattices in current use (with $a < 0.15$ fm). This means that, although discretization effects are visible in lattice QCD calculations with $c$ quarks, they are not large and can easily be extrapolated away accurately for a continuum result. For example, discretization effects are less than 10% at $a = 0.15$ fm in calculations of the decay constant of the $D_s$ using the HISQ action [20]. Purely nonrelativistic approaches to the $c$ quark are therefore not useful on the lattice. There can be some advantage for $b$-to-$c$ form factor calculations in using the same action for $b$ and $c$, however, as we discuss below.

Because lattice and continuum QCD regularize the theory in a different way, the lattice current $J$ needs a finite renormalization factor to match its continuum counterpart so that matrix elements of $J$, and form factors derived from them, can be used in continuum phenomenology. For NRQCD and Wilson/Fermilab quarks the current $J$ must be normalized using lattice QCD perturbation theory. Since this is technically rather challenging it has only been done through $O(\alpha_s)$ and this leaves a sizeable (possibly several percent) systematic error from missing higher-order terms in the perturbation theory. If Wilson/Fermilab quarks are used for both $b$ and $c$ quarks, then arguments can be made about the approach to the heavy-quark limit that can reduce, but not eliminate, this uncertainty [21].

Relativistic treatments of the $b$ and $c$ quarks have a big advantage here, because $J$ can generally be normalized in a fully nonperturbative way within the lattice QCD calculation and without additional systematic errors. The advantages of this approach were first demonstrated by the HPQCD collaboration using the HISQ action to determine the decay constant of the $B_s$ [22]. The HISQ PCAC relation normalizes the axial-vector current in this case. Calculations for multiple quark masses on lattices with multiple values of the lattice spacing allow both the physical dependence of the decay constant on quark mass and the dependence of the discretization effects to be mapped out so that the physical result at the $b$ quark mass can be determined. This calculation has now been updated and extended to the $B$ meson by the Fermilab Lattice and MILC collaborations [23], achieving better than 1% uncertainty. HPQCD is now carrying out a similar approach to $b$-to-$c$ form factor calculations [24], and the JLQCD collaboration is also working in that direction [25] with Möbius domain-wall quarks.

An equivalent approach, using ratios of hadronic quantities at different quark masses where normalization factors cancel, has been developed by the European Twisted Mass
collaboration using the twisted-mass action [26,27] for Wilson fermions.

2.2.1 \( B \rightarrow D^{(*)} \) form factors from lattice QCD

Early lattice QCD calculations of \( B \rightarrow D \) form factors were limited to the determination of \( \mathcal{G}^{B \rightarrow D}(w) = 4r f_+(q^2)/(1 + r) \) (with notation defined near (2.16)) at the zero-recoil point \( w = 1 \). Results include the \( N_f = 2 + 1 \) calculation of Fermilab/MILC [28,29] and the \( N_f = 2 \) calculation of Atoui et al. [30]. More recently Fermilab/MILC [31] and HPQCD [32,33] have presented \( N_f = 2 + 1 \) calculations of the \( B \rightarrow D \) form factor at non-zero recoil based on partially overlapping subsets of the same MILC asqtad (a\(^2\) tadpole improved) ensembles.

The Fermilab/MILC calculation [31] uses configurations with four different lattice spacings and with pion masses in the range \([260, 670]\) MeV. The bottom and charm quarks are implemented in the Fermilab approach. The form factors \( f_{B \rightarrow D}^{B \rightarrow D}(w) \) are extracted from double ratios of three point functions up to a matching factor which is calculated at 1-loop in lattice perturbation theory. The results are presented in terms of three synthetic data points which can be subsequently fitted using any form factor parametrization. The systematic uncertainty due to the joint continuum-chiral extrapolation is about 1.2% and dominates the error budget.

The HPQCD calculations [32,33] rely on ensembles with two different lattice spacings and two/three light-quark masses values, respectively. The treatment of heavy quarks is different from that used in the Fermilab/MILC papers: the bottom and charm quarks are described in NRQCD and the charm quark using HISQ. The form factors are extracted from appropriate three-point functions and the results are presented in terms of the parameters of a modified BCL \( z \) expansion that incorporates dependence on lattice spacing and light-quark masses into the expansion coefficients.

In order to combine the Fermilab/MILC and HPQCD results [3], it is necessary to generate a set of synthetic data which is (almost exactly) equivalent to the HPQCD calculation. The two sets of synthetic data can then be combined while taking into account the correlation due to the fact the Fermilab/MILC and HPQCD share MILC asqtad configurations. As mentioned above, dominant uncertainties are of systematic nature, implying that this correlation (whose estimate is rather uncertain) is a subdominant effect. A simultaneous fit of Fermilab/MILC and HPQCD synthetic data together with the available Belle and Babar data yields a determination of \( |V_{cb}| \) with an overall 2.5% uncertainty (dominated by the experimental error which contributes about 2% to the total error).

Finally, both collaborations present values for both the \( f_+ \) and \( f_0 \) form factors, which allow for a lattice only calculation of the SM prediction for \( R(D) \). The uncertainty on the Fermilab/MILC and HPQCD combined determination of \( R(D) \), without experimental input, is about 2.5% and is negligible compared to current experimental errors.

The advantage of an approach in which currents can be nonperturbatively normalized has been demonstrated by HPQCD for \( B_s \rightarrow D_s \) form factors in [34]. They use the HISQ action for all quarks, extending the method developed for decay constants. The range of heavy quark masses can be increased on successively finer lattices (keeping the value in lattice units below 1) until the full range from \( c \) to \( b \) is reached. The full \( q^2 \) range of the decay can also be covered by this method since the spatial momentum of the final state meson (which should also be less than 1 in lattice units) grows in step with the heavy meson/quark mass. Results from [34] improve on the uncertainties obtained in [33] with NRQCD \( b \) quarks and this promising all-HISQ approach is now being extended to other processes. It is interesting to observe that the \( B_s \rightarrow D_s \) form factors are very close to the \( B \rightarrow D \) form factors over the entire kinematic range, see also [35,36].

Calculations of \( B \rightarrow D^* \) form factors at non-zero recoil are considerably more involved due to difficulties in describing the resonant \( D^* \rightarrow D \pi \) decay. Up to now, lattice QCD simulations have focused on the single \( B \rightarrow D^* \) form factor that contributes to the rate at zero recoil, \( A_1(q_{\text{max}}^2) \). The quantity generally quoted is \( h_{A_1}(1) \) where

\[
h_{A_1}(1) = \frac{M_B + M_{D^*}}{2\sqrt{M_B M_{D^*}}} A_1(q_{\text{max}}^2)
\]  

The combination of the lattice QCD result and the experimental rate, extrapolated to zero recoil, yields a value for \( V_{cb} \).

The Fermilab Lattice/MILC Collaborations have achieved the highest precision for this result so far [37]. They use improved Wilson quarks within the Fermilab approach for both \( b \) and \( c \) quarks and work on gluon field configurations that include \( u/d \) (with equal mass) and \( s \) quarks in the sea \((n_f = 2 + 1)\) using the asqtad action. By taking a ratio of three-point correlation functions they are able simultaneously able to improve their statistical accuracy and reduce part of the systematic uncertainty from the normalization of their current operator. Their result is \( h_{A_1}(1) = 0.906(4)(12) \) where the uncertainties are statistical and systematic respectively. Their systematic error is dominated by discretization effects. They take the systematic uncertainty from missing higher-order terms in the perturbative current matching [38] to be 0.1\( \alpha_s^2 \).

The HPQCD collaboration have calculated \( h_{A_1}(1) \) on gluon field configurations that include \( n_f = 2 + 1 + 1 \) HISQ sea quarks using NRQCD \( b \) quarks and HISQ \( c \) quarks [39]. Their result, \( h_{A_1}(1) = 0.895(10)(24) \) has a larger uncertainty, dominated by the systematic uncertainty of 0.5\( \alpha_s^2 \) allowed for in the current matching. They were also able to calculate the equivalent result for \( B_s \rightarrow D^*_s \), obtaining \( h_{A_1}^s(1) = 0.879(12)(26) \) and demonstrating that the depen-
Fig. 1 Plot taken from Ref. [24] showing the comparison of lattice QCD results for $h_{A_1}(1)$ (left side) and $h_{A_1}^s(1)$ (right side). Raw results for $h_{A_1}(1)$ are from [39] and [37] and are plotted as a function of valence (= sea) light quark mass, given by the square of $M_\pi$. On the right are points for $h_{A_1}^s(1)$ from [39] plotted at the appropriate valence mass for the $s$ quark, but obtained at physical sea light quark masses. The final result for $h_{A_1}(1)$ from [37], with its full error bar, is given by the inverted blue triangle. The inverted red triangles give the final results for $h_{A_1}(1)$ and $h_{A_1}^s(1)$ from [39]. The HPCQCD results of [24] are given by the black stars.

The importance of being able to compare lattice QCD and experiment away from the zero recoil point is now clear and several lattice QCD calculations are underway, attempting to cover the full $q^2$ range of the decay and all 4 form factors. This includes calculations for $B \rightarrow D^*$ from JILQCD [25] with Möbius domain-wall quarks, Fermilab/MILC [40] (see also talk at Lattice 2019) with improved Wilson/Fermilab quarks and LANL/SWME with an improved version of this formalism known as the Oktay–Kronfeld action [41]. Calculations for other $b \rightarrow c$ pseudoscalar-to-vector form factors, $B_s \rightarrow D_{s1}^*$ [42] and $B_c \rightarrow (J/\psi, \eta_c)$ are also underway from HPQCD [43,44] using the all-HISQ approach. At the same time further $B \rightarrow D$ and $B_s \rightarrow D_s$ form factor calculations are in progress, including those using a variant of the Fermilab approach known as relativistic heavy quarks on RBC/UKQCD configurations [45]. In future we should be able to compare results from multiple actions with experiment for improved accuracy in determining $|V_{cb}|$.

2.2.2 $A_b \rightarrow A_c^{(*)}$ form factors from lattice QCD

The $A_b \rightarrow A_c$ form factors have been calculated with $2 + 1$ dynamical quark flavors; the vector and axial vector form factors can be found in Refs. [16], while the tensor form factors (which contribute to the decay rates in many new-physics scenarios) where added in Ref. [46]. This calculation used two different lattice spacings of approximately 0.11 fm and 0.08 fm, sea quark masses corresponding to pion masses in the range from 360 down to 300 MeV, and valence quark masses corresponding to pion masses in the range from 360 down to 230 MeV. The lattice data for the form factors, which cover the kinematic range from near $q_{max}^2 \approx 11 \text{GeV}^2$ down to $q^2 \approx 7 \text{GeV}^2$, were fitted with a modified version of the BCL $z$ expansion [8] discussed in Sect. 2.1, where simultaneously to the expansion in $z$, an expansion in powers of the lattice spacing and quark masses is performed. No dispersive bounds were used in the $z$ expansion here (this is something that can perhaps be improved in the future, see also Sect. 3.5.5). The form factors extrapolated to the continuum limit and physical pion mass yield the following Standard Model predictions:

$$\frac{1}{|V_{cb}|^2} \Gamma(A_b \rightarrow A_c \mu^- \bar{\nu}_\mu) = (21.5 \pm 0.8_{\text{stat}} \pm 1.1_{\text{syst}}) \text{ ps}^{-1}$$  \hspace{1cm} (2.37)

for the fully integrated decay rate, which has a total uncertainty of 6.3% (corresponding to a 3.2% theory uncertainty in a possible $|V_{cb}|$ determination from this decay rate),

$$\frac{1}{|V_{cb}|^2} \int_{7 \text{GeV}^2}^{q_{max}^2} \frac{d\Gamma(A_b \rightarrow A_c \mu^- \bar{\nu}_\mu)}{dq^2} \text{ d}q^2 = (8.37 \pm 0.16_{\text{stat}} \pm 0.34_{\text{syst}}) \text{ ps}^{-1}$$ \hspace{1cm} (2.38)

for the partially integrated decay rate, which has a total uncertainty of 4.5% (corresponding to 2.3% for $|V_{cb}|$), and

$$R(A_c) = \frac{\Gamma(A_b \rightarrow A_c \tau^- \bar{\nu}_\tau)}{\Gamma(A_b \rightarrow A_c \mu^- \bar{\nu}_\mu)} = 0.3328 \pm 0.0074_{\text{stat}} \pm 0.0070_{\text{syst}}$$ \hspace{1cm} (2.39)

for the lepton-flavor-universality ratio, which has a total uncertainty of 3.1%. The systematic uncertainties of the vector and axial vector form factors are dominated by finite-volume effects and the chiral extrapolation. Both of these can be reduced substantially in the future by adding a new lattice gauge field ensemble with physical light-quark masses and a large volume, and dropping the “partially quenched” data sets that have $m_\pi^{(\text{val})} < m_\pi^{(\text{sea})}$. Adding another ensemble at a third, finer lattice spacing will also be beneficial to better control the continuum extrapolation.

At this workshop, there was some discussion about the validity of the modified $z$ expansion; it has been argued that it would be safer to first perform chiral/continuum extrapo-
lations and then perform a secondary z expansion fit. This is expected to make a difference mainly if nonanalytic quark-mass dependence from chiral perturbation theory is included. However, the fits used in Ref. [16] for the $A_\ell$ form factors were analytic in the lattice spacing and light-quark mass. Note that the shape of the $A_b \to A_\ell \mu^+ \bar{\nu}_\mu$ differential decay rate was later measured by LHCb, and found to be in good agreement with the lattice QCD prediction all the way down to $q^2 = 0$ [47].

Motivated by the prospect of an LHCb measurement of $R(A^*_\ell)$, work is now also underway to compute the $A_b \to A^*_\ell$ form factors in lattice QCD, for the $A^*_\ell(2595)$ and $A^*_\ell(2625)$, which have $J^P = \frac{1}{2}^-$ and $J^P = \frac{3}{2}^-$, respectively. Preliminary results were shown at the workshop. For these form factors, the challenge is that, to project the $A^*_\ell$ interpolating field exactly to negative parity and avoid contamination from the lower-mass positive parity states, one needs to perform the lattice calculation in the $A^*_\ell$ rest frame. With the $b$-quark action currently in use, discretization errors growing with the $\Lambda_b$ momentum then limit the accessible kinematic range to a small region near $q^2_{\text{max}}$. To predict $R(A^*_\ell)$, it will be necessary to combine the lattice QCD results for the form factors in the high-$q^2$ region with heavy-quark effective theory and LHCb data for the shapes of the $A_b \to A^*_\ell \mu^+ \bar{\nu}_\mu$ differential decay rates [48].

2.3 Measurements of $B \to D^{(*)} \ell \nu$ and related processes

2.3.1 Measurements with light leptons

The decays $B \to D^{(*)} \ell \nu$ and $B \to D \ell \nu$ have been measured at Belle and BaBar as well as at older experiments (CLEO, LEP). Unfortunately, most of these measurements assume the Caprini–Lellouch–Neubert parametrization of the form factors (see Sect. 2.1) and report results in terms of $|V_{cb}|$ times the only form factors relevant at the zero-recoil point $w = 1$, namely $F(1) \equiv h_{A_1}(1)$ for $B \to D^{(*)} \ell \nu$ and $G(1) \equiv 2\sqrt{M_DM_B/(M_D + M_B)} f^+ (1)$ for $B \to D \ell \nu$, and of the other CLN parameters, instead of a general form of the $z$ expansion or the raw spectra. The heavy flavor averaging group (HFLAV) has performed an average of these CLN measurements [49] and reports

\begin{equation}
\eta_{\text{EW}}F(1)|V_{cb}| = (35.27 \pm 0.11) \pm 0.36(\text{sys}) \times 10^{-3},
\end{equation}

\begin{equation}
\eta_{\text{EW}}G(1)|V_{cb}| = (42.00 \pm 0.45) \pm 0.89(\text{sys}) \times 10^{-3},
\end{equation}

Notice that Eq. (2.40) together with $h_{A_1}(1) = 0.904(12)$ [3] leads to the low value $|V_{cb}| = 38.76(69) \times 10^{-3}$. Eq. (2.41) together with $G(1) = 1.054(1) 10^{-3}$. Eq. (2.41) together with $G(1) = 1.0541(83)$ [31] leads to a consistent result $|V_{cb}| = 39.58(99) \times 10^{-3}$. In the case of $B \to D \ell \nu$ one can also use the existing lattice calculations at non-zero recoil [31,32] to guide the extrapolation to zero recoil, together with the $w$ spectrum measured by Belle [50]. In the BGL parametrization, this leads to a higher value, $|V_{cb}| = 40.83(1.13) \times 10^{-3}$, a more reliable determination than (2.41).

In the following we will have a closer look at the most recent measurements by the various experiments.

Belle has recently updated the untagged measurement of the $B^0 \to D^{*+} \ell^+ \nu$ mode [51]. While the new analysis is based on the same 711 fb$^{-1}$ Belle data set, the reanalysis takes advantage of a major improvement of the track reconstruction software, which was implemented in 2011, leading to a substantially higher slow pion tracking efficiency and hence to much larger signal yields than in the previous publication [52]. Again $D^{*+}$ mesons are reconstructed in the cleanest mode, $D^{*+} \to D^0 \pi^+$ followed by $D^0 \to K^- \pi^+$, combined with a charged, light lepton (electron or muon) and yields are extracted in 10 bins for each of the 4 kinematic variables describing the $B^0 \to D^{*+} \ell^+ \nu$ decay. These yields are published along with their full error matrix. The updated publication also contains an analysis of these yields using both the CLN and the BGL form factors (where BGL has only 5 free parameters). The CLN analysis results in $\eta_{\text{EW}}F(1)|V_{cb}| = (35.06 \pm 0.15(\text{stat}) \pm 0.56(\text{sys})) \times 10^{-3}$, while the BGL fit gives $\eta_{\text{EW}}F(1)|V_{cb}| = (34.93 \pm 0.23(\text{stat}) \pm 0.59(\text{sys})) \times 10^{-3}$.

Both results are thus well consistent. This contrasts with a tagged measurement of $B^0 \to D^{*+} \ell^+ \nu$ first shown by Belle in November 2016 [53]. Analyzing the raw data of this measurement in terms of the CLN and BGL form-factors gives a difference of almost two standard deviations in $|V_{cb}|$ [54,55]. However, this result has remained preliminary and will not be published. A new tagged analysis, using an improved version of the hadronic tag is now underway and should clarify the experimental situation.

Babar has presented a full four-dimensional angular analysis of $B^0 \to D^{*0} \ell^+ \nu$ decays, using both CLN and BGL parametrizations [56]. This analysis is based on the full data set of 450 fb$^{-1}$, and exploits the hadronic B-tagging approach. The full decay chain $e^+ e^- \to \gamma(4S) \to B_{\text{tag}} B_{\text{sig}} (\to D^* \ell \nu \nu)$ is considered in a kinematic fit that includes constraints on the beam properties, the secondary vertices, the masses of $B_{\text{tag}}$, $B_{\text{sig}}$, $D^*$ and the missing neutrino. After applying requirements on the probability of the $\chi^2$ of this constrained fit, which is the main discriminating variable, the remaining background is only about 2% of the sample. The resolution on the kinematic variables is about a factor five better than the one possible with untagged measurements. The shape of the form factors is extracted using an unbinned maximum likelihood fit where the signal events are described by the four dimensional differential decay rate. The extraction of $|V_{cb}|$ is performed indirectly by adding to the likelihood the constraint that the integrated rate $\Gamma = B/\tau_B$, where $B$ is the $B \to D^{*+} \ell^+ \nu$ branching fraction and $\tau_B$ is the B-
The values of these external inputs are taken from HFLAV [49]. The final result, using $h_{A(1)}$ from [37], is $|V_{cb}| = (38.36 \pm 0.90) \times 10^{-3}$ with a 5-parameter BGL version and $|V_{cb}| = (38.40 \pm 0.84) \times 10^{-3}$ in the CLN case, both compatible with the above HFLAV average. Nevertheless, the individual form factors show significant deviations from the world average CLN determination by HFLAV.

LHCb has extracted $V_{cb}$ from semileptonic $B^0_s$ decays for the first time [57]. The measurement uses both $B^0 \to D_s^+ \mu^- \nu_\mu$ and $B^0 \to D_s^- \mu^+ \nu_\mu$ decays using 3 fb$^{-1}$ collected in 2011 and 2012. The value of $|V_{cb}|$ is determined from the observed yields of $B^0_s$ decays normalized to those of $B^0$ decays after correcting for the relative reconstruction and selection efficiencies. The normalization channels are $B^0 \to D^- \mu^+ \nu_\mu$ and $B^0 \to D^\ast^- \mu^+ \nu_\mu$ with the $D^- \to [K^+ K^-]_\nu \pi^-$, to minimize the systematic uncertainties. The shapes of the form factors are extracted as well, exploiting the kinematic variable $p_{\perp} (D_s)$ which is the component of the $D_s^\ast$ momentum perpendicular to the $B^0_s$ flight direction. This variable is correlated with $q^2$. In this analysis both the CLN parametrization and a 5-parameter version of BGL have been used. The results for $V_{cb}$ are

\[
V_{cb}\text{[CLN]} = (41.4 \pm 0.6\text{(stat)} \pm 0.9\text{(syst)} \pm 1.2\text{(ext)}) \times 10^{-3}, \\
V_{cb}\text{[BGL]} = (42.3 \pm 0.8\text{(stat)} \pm 0.9\text{(syst)} \pm 1.2\text{(ext)}) \times 10^{-3},
\]

where the first uncertainty are statistical, the second systematic and the third due to the limited knowledge of the external input, in particular the $B^0_s$ to $B^0_s$ production ratio $f_{s}/f_d$, which is known with an uncertainty of about 5%. The results are compatible with both the inclusive and exclusive decays. Although not competitive with the results obtained at the $B$ factories, the novel approach used can be extended to the semileptonic $B^0$ decays.

### 2.3.2 Past measurements of $R(D)$ and $R(D^*)$

$R_D$ and $R_{D^*}$ are defined as the ratios of the semileptonic decay width of $B_d$ and $B_u$ meson to a $\tau$ lepton and its associated neutrino $\nu_\tau$ over the $B$ decay width to a light lepton. A summary of the currently available measurements of $R_D$ and $R_{D^*}$ is presented in Table 2, showing the yield of $B$ signal and $B$ normalization decays and the stated uncertainties.

The data were collected by the BaBar and Belle experiments at $e^+e^-$ colliders operating at the $\Upsilon(4S)$ resonance, which decays exclusively to pairs of $B^+B^-$ or $B^0\bar{B}^0$ mesons. The LHCb experiment operates at the high energy $p\bar{p}$ collider at CERN at total energies of 7 and 8 TeV, where pairs of $b$-hadrons (mesons or baryons) along with a large number of other charged and neutral particles are produced. While the maximum production rate of the $\Upsilon(4S) \to B\bar{B}$ events has been 20 Hz, the rates observed at LHCb exceed 100 kHz.

Currently we have only two measurements [58–60] of the ratios $R_D$ and $R_{D^*}$ based on two distinct samples of hadronic tagged $B\bar{B}$ events with signal $B \to D\tau\nu_\tau$ and $B \to D^\ast\tau\nu_\tau$ decays and purely leptonic tau decays, $\tau^- \to e^-\bar{\nu}_e\nu_\tau$ or $\tau^- \to \mu^-\bar{\nu}_\mu\nu_\tau$. In addition, there is a measurement from Belle [62,63] of $R_{D^*}$ with hadronic tags and a semileptonic one-prong $\tau$ decay ($\tau^- \to \pi^-\nu_\tau$ or $\tau^- \to \rho^-\nu_\tau$). A Belle measurement [61] of $R_D$ and $R_{D^*}$ with semi-leptonic tags and purely leptonic $\tau$ decays appeared recently, superseding a previous measurement [67] of $R_{D^*}$ obtained with the same technique.

At LHCb only decays of neutral $B$ mesons producing a charged $D^\ast$ meson and a muon of opposite charge are selected, with a single decay chain $D^{\ast+} \to D^0(\to K^+\pi^-)\pi^+$. LHCb published two measurements [64–66] of $R_{D^*}$, the first relying on purely leptonic $\tau$ decays and normalized to the $B^0 \to D^{\ast+}\mu^-\bar{\nu}_\mu$ decay rate, and the more recent one using 3-prong semileptonic $\tau$ decays, $\tau^- \to \pi^+\pi^-\pi^-\nu_\tau$ and normalization to the decay $B^0 \to D^{\ast+}\pi^+\pi^-\pi^-$. This LHCb measurement extracts directly the ratio of branching fractions $K(D^*) = B(B^0 \to D^{\ast+}\tau^-\bar{\nu}_\tau)/B(B^0 \to D^{\ast+}\pi^+\pi^-\pi^-)$. The ratio $K(D^*)$ is then converted to $R(D^*)$ by using the known branching fractions of $B^0 \to D^{\ast+}\pi^+\pi^-\pi^-$ and $B^0 \to D^{\ast+}\mu^-\bar{\nu}_\mu$.

BaBar and Belle analyses rely on the large detector acceptance to detect and reconstruct all final state particles from the decays of the two $B$ mesons, except for the neutrinos. They exploit the kinematics of the two-body $\Upsilon(4S)$ decay and known quantum numbers to suppress non-$BB$ and combinatorial backgrounds. They differentiate the signal decays involving two or three missing neutrinos from decays involving a low mass charged lepton, an electron or muon, plus an associated neutrino.

LHCb isolates the signal decays from very large backgrounds by exploiting the relatively long $B$ decay lengths which allows for a separation of the charged particles from the $B$ and charm decay vertex from many others originating from the $pp$ collision point. There are insufficient kinematic constraints and therefore the total $B$ meson momentum is estimated from its transverse momentum, degrading the resolution of kinematic quantities like the missing mass and the momentum transfer squared $q^2$. Also, the production of $D^{\ast+}D_s^-$ pairs with the decay $D_s^- \to \tau^-\bar{\nu}_\tau$ leads to sizable background in the signal sample.

The summary in Table 2 indicates that the results are not inconsistent. For BaBar and Belle the systematic uncertainties are comparable for $R_{D^*}$, while Belle systematic uncertainties are smaller for $R_D$. However the differences in the signal yield and the background suppression lead to smaller statistical errors for BaBar. The Belle measurements based on semileptonic tagged samples result in a 50% smaller signal yield than for the hadronic tag samples. For the two LHCb measurements, the event yields exceed the BaBar yields by
Table 2 Summary of $R_D$ and $R_{D^*}$ measurements and theoretical predictions. The number of observed signal and normalization events is also reported. The normalization channel is $B \to D^{(*)} \ell \nu \ell$ for all measurements but the LHCb one with three-prong $\tau$ decays, where the normalization channel is $B \to D^\star \pi \pi \pi$. The latter LHCb measurement has been updated using the latest HFLAV average for $B(B \to D^\star \ell \nu \ell)$. The quoted theory predictions are arithmetic averages of the values reported in Table 3 below; they are given for illustration only and do not imply consent from the authors of the calculations.

| Experiment | tag | $\tau$ decay | $N(D \tau \nu \ell)$ | $N_{norm}$ | $R(D)$  |
|------------|-----|---------------|---------------------|------------|---------|
| Babar [58,59] | Had. | $\ell \nu \ell$ | 489 ± 63            | 2891 ± 65  | 0.440 ± 0.058 ± 0.042 |
| Belle [60] | Had. | $\ell \nu \ell$ | 320 ± 55            | 3147 ± 72  | 0.375 ± 0.064 ± 0.026 |
| Belle [61] | SL   | $\ell \nu \ell$ | 1778 ± 204          | 22896 ± 471| 0.307 ± 0.037 ± 0.016 |
| HFLAV      |      |               |                     |            | 0.340 ± 0.027 ± 0.013 |
| Theory     |      |               |                     |            | 0.299 ± 0.003      |

| Experiment | tag | $\tau$ decay | $N(D^\star \tau \nu \ell)$ | $N_{norm}$ | $R(D^\star)$  |
|------------|-----|---------------|-----------------------------|------------|--------------|
| Babar [58,59] | Had. | $\ell \nu \ell$ | 888 ± 63                   | 11953 ± 122| 0.332 ± 0.024 ± 0.018 |
| Belle [60] | Had. | $\ell \nu \ell$ | 503 ± 65                   | 3797 ± 74  | 0.293 ± 0.038 ± 0.015 |
| Belle [62,63] | Had. | $\pi \nu \ell$, $\rho \nu \ell$ | 298 ± 29                   | 7213 ± 96  | 0.270 ± 0.035 ± 0.028 |
| LHCb [64] | $\nu \mu \nu \ell$ | 16480                 | 363000                       | 0.336 ± 0.027 ± 0.030 |
| LHCb [65,66] | $\pi \pi \pi \nu \ell$ | 1273                | 17660                       | 0.280 ± 0.018 ± 0.029 |
| Belle [61] | SL   | $\ell \nu \ell$ | 651 ± 46                   | 16942 ± 148| 0.283 ± 0.018 ± 0.014 |
| HFLAV      |      |               |                             |            | 0.295 ± 0.011 ± 0.008 |
| Theory     |      |               |                             |            | 0.253 ± 0.005      |

Close to a factor of 20, but the relative statistical errors on $R_{D^*}$ are comparable to BaBar, and the systematic uncertainties are larger by a factor of 2.

2.3.3 Lessons learned

All currently available measurements are limited by the difficulty of separating the signal from large backgrounds from many sources, leading to sizable statistical and systematic uncertainties. The measurement of ratios of two $B$ decay rates with the very similar – if not identical – final state particles, significantly reduces the systematic uncertainties due to detector effects, tagging efficiencies, and also from uncertainties in the kinematics due to form factors and branching fractions. For all three experiments the largest systematic uncertainties are attributed to the limited size of the MC samples, the fraction and shapes of various backgrounds, especially from decays involving higher mass charm states, and uncertainties in the relative efficiency of signal and normalization, the efficiency of other backgrounds, as well as lepton mis-identification. Though the total number of $B \bar{B}$ events of the full Belle data set exceeds the one for BaBar by 65%, the signal BaBar signal yield for $B \to D^{(*)} \tau \nu \ell$ exceeds Belle by 67% due to differences in event selection and fit procedures.

While the use by Belle of semileptonic $B$ decays as tags for $B \bar{B}$ events benefits from the fewer decay modes with higher BFs, the presence of a neutrino in the tag decays results in the loss of stringent kinematic constraints. The resulting signal yields are lower by 50% compared to hadronic tags, and the backgrounds are much larger. The use of the ECL, namely the sum of the energies of the excess photons in a tagged event, in the fit to extract the signal yield is somewhat problematic, since it includes not only the photons left over from incorrectly reconstructed $B \bar{B}$ events, but also photons emitted from the high intensity beams. As a result the signal contributions are difficult to separate from the very sizable backgrounds.

2.3.4 Outlook for $R(D)$ and $R(D^*)$

Belle II and the upgraded LHCb are expected to collect large data samples with considerably improved detector performances. This should lead to much reduced detector related uncertainties, higher signal fractions, and opportunities to measure many related processes. The goal is to push the sensitivity of many measurements of critical variables and distributions beyond theory uncertainties and thereby increase the sensitivity to non-Standard Model processes.

Currently there are only two measurements of the ratio $R_D$, one each by BaBar and Belle, based on two distinct samples of hadronic tagged $B \bar{B}$ events for the signal $B \to D \tau \nu \ell$ and $B \to D^\star \tau \nu \ell$ decays. The decay $B \to D \tau \nu \ell$ is dominated by a P-wave, whereas in the $B \to D^\star \tau \nu \ell$ S, P, and D waves contribute and the impact for contributions from new physics processes is expected to be smaller. A contribution of a hypothetical charged Higgs would result in an S-wave for $B \to D \tau \nu \ell$, and a P-wave for $B \to D^\star \tau \nu \ell$, thus measurements of the angular distributions and the polarization...
of the $\tau$ lepton or $D$ and $D^*$ mesons will be important. Such measurements would of course also serve as tests of other hypotheses, for instance contributions from leptoquarks. The studies for many decay modes, the detailed kinematics of the signal events, the four-momentum transfer $q^2$, the lepton momentum, the angles and momenta of $D$ and $D^*$ and the $\tau$ spin should be extended to perform tests for potential new physics contributions.

Belle II will benefit from major upgrades to all detector components, except for the barrel sections of the calorimeter and the muon detector. In addition, a new data acquisition and analysis software are being developed to benefit from the very high data rates and improved detector performance. Upgrades to the precision tracking and lepton identification, especially at lower momenta, are expected to significantly improve the mass resolution and purity of the signal samples. This should also improve the detector modeling of efficiencies for signal and backgrounds and fake rates that are the major contributions to the current systematic uncertainties. The much larger data rates should allow choice of cleaner and more efficient $B\bar{B}$ tagging algorithms.

Major improvements to the MC simulation signal and backgrounds will be needed. They require much better understanding of all semileptonic $B$ decays, contributing to signal and backgrounds, i.e., updated measurements of branching fractions and form factors and theoretical predictions, especially for backgrounds involving higher mass charm mesons, either resonances or states resulting from charm quark fragmentation. The fit to extract the signal yields could be improved by reducing the backgrounds and making use of fully 2D or 3D distributions of kinematic variables, and by avoiding simplistic parametrizations. The suppression of fake photons and $\pi^0$s needs to be scrutinized to avoid unnecessary signal loss and very large backgrounds for $D^{*0}$ decays. Shapes of distributions entering multi-variable methods to reduce the backgrounds should be scrutinized by comparisons with data or MC control samples, and any significant differences should be addressed. The use of ECL, the sum of the energies of all unassigned photon in an event, may be questionable, given the expected high rate of beam generated background.

The first study by Belle of the $\tau$ spin in $B \to D^* \tau \nu_\tau$ decays with $\tau^- \to \rho^- \nu_\tau$ or $\tau^- \to \pi^- \nu_\tau$ is very promising, it indicates that much larger and cleaner data samples will be needed. The systematic uncertainty on the $R_{D^*}$ measurement of 11% is dominated by the hadronic $B$ decay composition of 7% and the size of the MC sample [63]. The measured transverse $\tau$ polarization of $P_\tau = -0.38 \pm 0.51_{-0.16}^{+0.21}$ is totally statistics dominated, and implies $P_\tau < 0.5$ at 90% C.L.

Among the many other measurements Belle II is planning, ratios $R$ for both inclusive and inclusive semileptonic $B$ decays are of interest, for instance in addition to $R_D, R_{D^*}$, and $R_{D^{*+}}$ also $R_{X_s}$, as well as $R_\pi$ and $R_{X_s}$, which rely on unique capabilities of Belle II.

The LHCb detector is currently undergoing a major upgrade with the goal to switch to an all software trigger and to be able to select and record data up to rates of 100 kHz. Replacements of all tracking devices are planned, ranging from radiation hard pixel detector near interaction region to scintillation fibers downstream. Improvements to electron and muon detection and reduction in pion misidentification will be critical for the suppression of backgrounds, and should also allow rate comparison for decays involving electron or muons. LHCb relies on large data samples rather than MC simulation to assess signal efficiencies and most importantly the many sources of backgrounds and their suppression.

Several analyses are underway based on Run 1 and Run 2 data samples, and are benefiting from improved trigger capabilities. The first analysis based on 3-prong $\tau$ decays showed a clear separation of the $\tau$ decay vertex from both the $D$ and the proton interaction point, improving the signal purity to about 11%, compared to 4.4% for the purely lepton 1-prong $\tau$ decay. This may therefore be the favored $\tau$ decay mode, and should also be tried for $B^+ \to D^0 \tau^+ \nu_\tau$. Improved measurements of the branching fractions for normalization and the $\tau$ decays will be essential.

As a follow-up on the first LHCb measurement of $R_{D^*}$, a simultaneous fit to two disjoint $D^0 \mu^- \bar{D}^0 \mu^+$ samples is in preparation, taking into account the large feed-down from $D^*$ decay present in the $D^0 \mu^-$ sample. As pointed out above, the decay $B^+ \to D^0 \tau^+ \nu_\tau$ is more sensitive to new physics processes than $B^0 \to D^{*+} \tau^- \nu_\tau$ and thus this analysis is expected to be very important to establish the excess in these decay modes and its interpretation. This analysis will benefit from the addition of dedicated triggers sensitive to $D^0 \mu^-, D^{*+} \mu^-, \Lambda^+_c \mu$ and $D^{*+} \mu^+$ final states.

LHCb is considering a series of other ratios measurements, among several $b \to c$ transitions ($\bar{B}_b^0 \to D^- \tau^+ \nu_\tau$, $B \to D^{*+} \tau^- \nu_\tau$ and $\Lambda_b^+ \to \Lambda_c^{*+} (\tau^+ \nu_\tau)$) and certain $b \to u$ transitions ($B^+ \to \rho^0 \tau^+ \nu_\tau$, $B^+ \to p \tau^+ \nu_\tau$ and $\Lambda_b^0 \to p \tau^- \nu_\tau$), most of which will be challenging to observe and not trivial to normalize. The decay $\Lambda_b^+ \to \Lambda_c^{*+} \tau^+ \nu_\tau$ probes a different spin structure, and a precise measurement of $R_{\Lambda_c}$ would be of great interest for the interpretation of the excess of events in $R_D$. The observation of the decay $B_c^- \to J/\psi (\to \mu^+ \mu^-) \tau^- (\to \mu^- \nu_\mu \nu_\tau \nu_\tau) \tau^- \nu_\tau$ has recently been reported. It is a very rare process which is only observable at LHCb. The final state of 3 muons is a unique signature, though impacted especially at lower momenta, are expected to significantly improve the mass resolution and purity of the signal samples. This should also improve the detector modeling of efficiencies for signal and backgrounds and fake rates that are the major contributions to the current systematic uncertainties. The much larger data rates should allow choice of cleaner and more efficient $B\bar{B}$ tagging algorithms.

Major improvements to the MC simulation signal and backgrounds will be needed. They require much better understanding of all semileptonic $B$ decays, contributing to signal and backgrounds, i.e., updated measurements of branching fractions and form factors and theoretical predictions, especially for backgrounds involving higher mass charm mesons, either resonances or states resulting from charm quark fragmentation. The fit to extract the signal yields could be improved by reducing the backgrounds and making use of fully 2D or 3D distributions of kinematic variables, and by avoiding simplistic parametrizations. The suppression of fake photons and $\pi^0$s needs to be scrutinized to avoid unnecessary signal loss and very large backgrounds for $D^{*0}$ decays. Shapes of distributions entering multi-variable methods to reduce the backgrounds should be scrutinized by comparisons with data or MC control samples, and any significant differences should be addressed. The use of ECL, the sum of the energies of all unassigned photon in an event, may be questionable, given the expected high rate of beam generated background.

The first study by Belle of the $\tau$ spin in $B \to D^* \tau \nu_\tau$ decays with $\tau^- \to \rho^- \nu_\tau$ or $\tau^- \to \pi^- \nu_\tau$ is very promising, it indicates that much larger and cleaner data samples will be needed. The systematic uncertainty on the $R_{D^*}$ measurement of 11% is dominated by the hadronic $B$ decay composition of 7% and the size of the MC sample [63]. The measured transverse $\tau$ polarization of $P_\tau = -0.38 \pm 0.51_{-0.16}^{+0.21}$ is totally statistics dominated, and implies $P_\tau < 0.5$ at 90% C.L.

Among the many other measurements Belle II is planning, ratios $R$ for both inclusive and inclusive semileptonic $B$ decays are of interest, for instance in addition to $R_D, R_{D^*}$, and $R_{D^{*+}}$ also $R_{X_s}$, as well as $R_\pi$ and $R_{X_s}$, which rely on unique capabilities of Belle II.
2.4 Extraction of $V_{cb}$ and predictions for $R_{D^{(*)}}$

The values of $V_{cb}$ extracted from inclusive and exclusive decays have been in tension for a long time [68]. In order to extract $V_{cb}$ from $B \rightarrow D^{(*)} l \nu$ data we need information on the form factors, which is mostly provided by lattice QCD. For the $B \rightarrow D$ form factors $f_{+0}$ there are lattice results at $w \geq 1$ [3, 31, 32]. A fit to all the available experimental and lattice data of $B \rightarrow D l \nu$ leads to [69]

$$V_{cb} \cdot 10^3 = 40.49(97),$$

(2.42)

with $\chi^2$/dof = 19.0/22. Similar results have been obtained in [3]. For $B \rightarrow D^*$ at the moment there is only information on one of the four form factors at zero-recoil, $A_1(w = 1)$ [37, 39], however further developments look promising [70–72]. At the other end of the $w$ or $q^2$ spectrum there are results available from LCSR [73, 74]. In view of the advanced experimental precision, a key question for the precise extraction of $V_{cb}$ and a robust prediction of $R(D^{(*)}) \equiv B(B \rightarrow D^{(*)} \tau \nu)/B(B \rightarrow D^{(*)} l \nu)$ is how large the theoretical uncertainties are. For example, whenever relations such as (2.24) are used, how large are HQET corrections beyond NLO, i.e. of $O \left( \alpha_s^2, \Lambda_{QCD}^2/m_{c,b}^2, \alpha_s \Lambda_{QCD}/m_{c,b} \right)$ and how accurate are the QCDR results that are used at NLO? A guideline for an answer to these questions can be provided by studying the size of NLO corrections in the HQET expansion and by a comparison with corresponding available lattice results [12]. A definite answer, especially for the pseudoscalar form factor $P_1$, which is needed for the prediction of $R(D^*)$, will be given only by future lattice results [70–72].

In all experimental analyses prior to 2017, HQET relations have been employed in terms of a form of the CLN parametrization [11] where theoretical uncertainties noted in Ref. [11] were set to zero by fixing coefficients to definite numbers. Moreover, the slope and curvature of $R_{1,2}(w)$ depend on the same underlying theoretical quantities as $R_{1,2}(1)$, which makes the variation of the latter and fixing of the former inconsistent. In future experimental analyses this has to be taken into account.

Recent preliminary Belle data [53] allowed for a reappraisal of fits to $B \rightarrow D^* l \nu$ by several groups [12, 39, 54, 55, 75–77]. For the first time, Ref. [53] reported deconvoluted $w$ and angular distributions which are independent of the parametrization. This allowed to test the possible influence of different parametrizations on the extracted value of $V_{cb}$. Indeed, based on that data set the central values for $|V_{cb}|$ varied by up to 6% between CLN and BGL fits [54, 55, 77]. By floating some additional parameters of the less flexible CLN parametrization, the agreement between BGL and CLN could be restored [54, 76]. Furthermore, in the literature one could observe a correlation of smaller central values for $V_{cb}$ with stronger HQET+QCDSR input [12, 39, 53–55, 75–77].

Recently, on top of the tagged analysis Ref. [53] a new untagged Belle analysis of $B \rightarrow D^* l \nu$ appeared [78]. The new, more precise data brought the $|V_{cb}|$ central values of the CLN and BGL fits closer together. However, in order to obtain a reliable error, it is necessary to employ the BGL parametrization with a sufficient number of coefficients rather than the CLN parametrization. Including the new data, Ref. [79] obtains

$$V_{cb} \cdot 10^3 = 39.6 \left( ^{+1.1}_{-1.0} \right),$$

(2.43)

with $\chi^2$/dof = 80.1/72. The inclusion of LCSR or strong unitarity constraints, where input from HQET is used in a conservative way, basically does not change the fit result [79]. The $V_{cb}$ value in Eq. (2.43) differs by 1.9σ from the inclusive result.

The shortcomings of the CLN parametrization have been addressed in several recent articles [36, 75, 77, 80, 86]: varying the coefficients of the HQE consistently allows for a simultaneous description of the available experimental and lattice data in $B \rightarrow D$, while the parametrization dependence in the extraction of $V_{cb}$ from Ref. [53] remains [75]. Including additionally contributions at $O(1/m_c^2)$ and higher orders in the $\varepsilon$ expansion, the extracted values for $V_{cb}$ using the BGL parametrization and the HQE become compatible [80].

For the above reasons, older HFLAV averages, which are based on the CLN parametrization, should not be employed in future analyses, with the exception of the total branching ratios, whose parametrization dependence is expected to be negligible. The two most recent experimental analyses of $B_{(s)} \rightarrow D_{(s)} \ell \nu$ [56, 57] present results obtained in both CLN and a simplified version of the BGL parametrization. They did not observe sizeable parametrization dependence, but found very different values of $V_{cb}$, however, they did not provide data in a format that allows for independent reanalyses.

For the lepton flavor nonuniversality observables $R(D^{(*)})$ we list a few recent theoretical predictions in Table 3. Predictions for further lepton flavor non-universality observables of underlying $b \rightarrow c \ell \nu$ transitions can be found in Refs. [87, 88]. Compared to predictions from before 2016, the predictions in Table 3 make use of new lattice results and new experimental data. The results are based on different methodologies and a different treatment of the uncertainties of HQET + QCDSR. We have a very good consensus for $R(D)$ predictions because in this case the predictions are dominated by the recent comprehensive lattice results from Refs. [31, 32, 89]. QED corrections to $R(D)$ remain a topic which deserves further study [90, 91]. In the case of $R(D^*)$, as we do not have
yet lattice information on the form factor $P_1$, we can use the
exact endpoint relation $P_1(w_{\text{max}}) = A_s(w_{\text{max}})$ and results
from HQET and QCDSR. Depending on the estimate of the
corresponding theory uncertainty one obtains different the-oretical
effects for the prediction of $R(D^*)$. As soon as we have lattice results for $P_1$ [70], the different fits will stabilize
and we expect a similar consensus as for $R(D)$. Despite the
most recent experimental results being closer to the SM pre-
dictions, the $R(D^{(*)})$ anomaly persists and remains a tough challenge for model builders.

2.5 Semileptonic $B \to D^{**} \ell \bar{\nu}$ decays

Semileptonic $B$ decays to the four lightest excited charm
mesons, $D^{**} = \{D^0_0, D^+_1, D_1, D_2^*\}$, are important both
because they are complementary signals of possible new
physics contributions to $b \to c \tau \nu$, and because they are sub-
stantial backgrounds to the $R(D^{(*)})$ measurements (as well
as to some $|V_{cb}|$ and $|V_{ub}|$ measurements). Thus, the correct
interpretation of future $B \to D^{(*)} \ell \bar{\nu}$ measurements requires
consistent treatment of the $D^{**}$ modes.

The spectroscopy of the $D^{**}$ states is important, because
in addition to the impact on the kinematics, it also affects the
expansion of the form factors [92,93] in HQET [94,95].

The isospin averaged masses and widths of the six lightest charm
mesons, rounded to 1 MeV [99] (from Ref. [100])

| Particle | $m$ (MeV) | $\Gamma$ (MeV) |
|----------|-----------|---------------|
| $D^0_0$  | 2349      | 236           |
| $D^+_1$  | 2427      | 384           |
| $D_1$    | 2421      | 31            |
| $D_2^*$  | 2461      | 47            |
| $D^*$    | 2099      | 0             |
| $D$      | 1866      | 0             |

The deviations $\Gamma_{\text{exp}} - \Gamma_{\text{theory}} = \sigma$ are
measured to be $\sigma = 0.5$ MeV. The corresponding results in the heavy quark limit were obtained in Ref. [102].

A simplifying assumption used in Refs. [92,93] to
reduce the number of subleading Isgur–Wise functions
was to neglect certain $O(\Lambda_{\text{QCD}}/m_c)$ contributions involving
the chromomagnetic operator in the subleading HQET
Lagrangian, motivated by the fact that the mass splittings in
both the $s^{i\pi}_1 = \frac{1}{2}^+$ and $s^{j\pi}_1 = \frac{3}{2}^+$ doublets were measured to
be much smaller than $m_{D^*} - m_{D}$. This is not supported by
the more recent data (see Table 4), so Ref. [101] extended the predictions of Refs. [92,93] accordingly, including deriving
the HQET expansions of the form factors which do not contribute in the $m_{\ell} = 0$ limit. The impact of arbitrary new
physics operators was analyzed in Ref. [100], including the
$O(\Lambda_{\text{QCD}}/m_c)$ and $(\alpha_s)$ corrections in HQET. The corre-
sponding results in the heavy quark limit were obtained in
Ref. [102].

The large impact of the $O(\Lambda_{\text{QCD}}/m_c)$ contributions to
the form factors can be understood qualitatively by considering
how heavy quark symmetry constrains the structure of the expansions near zero recoil. It is useful to think of a simul-
taneous expansion in powers of $(w - 1)$ and $(\Lambda_{\text{QCD}}/m_c)$. (The kinematical ranges are $0 < w - 1 < 0.2$ for $\tau$ final
states, and $0 < w - 1 < 0.3$ for $e$ and $\mu$.) The decay rates to
the spin-1 $D^{**}$ states, which are not helicity suppressed near
$w = 1$, are of the form

$$d \Gamma_{D^*_1, D^*_2} / dw \sim \sqrt{w^2 - 1} \left[ (0_{\text{HQS}} + 0_{\text{HQS}}) e + e^2 + \cdots \right]$$

Here $e$ is a power-counting parameter of order
$\Lambda_{\text{QCD}}/m_c$, and the $0$s are consequences of heavy quark symmetry. The $e^2$ term in the first parenthesis is fully determined by the leading order Isgur–Wise function and hadron mass splittings [92,93,100,101]. The same also holds for those new physics contributions to $B \to D^*_0 \ell \bar{\nu}$, which are not helicity sup-

---

**Table 3** Recent theory predictions for $R(D^{(*)})$. The deviations are calculated from the HFLAV spring 2019 updates $R(D)^{\exp} = 0.340(27)(13)$ [49, 58–61] and $R(D)^{\exp} = 0.295(11)(8)$ [49, 58–67], respectively. For older predictions see Refs. [82–84]. Table adapted and extended from Ref. [85]

| References | $R(D)$ | Exp. deviation |
|------------|--------|---------------|
| [69]       | 0.299(3) | 1.4σ          |
| [75]       | 0.299(3) | 1.4σ          |
| [77]       | 0.302(3) | 1.3σ          |
| [80]       | 0.297(3) | 1.4σ          |

| References | $R(D^*)$ | Exp. deviation |
|------------|----------|---------------|
| [75]       | 0.257(3) | 2.7σ          |
| [79]       | 0.254(3) | 2.7σ          |
| [81]       | 0.251(3) | 3.1σ          |
| [80]       | 0.250(3) | 3.2σ          |
pressed. This explains why the $O(\Lambda_{QCD}/m_{c,b})$ corrections to the form factors are very important, and can make $O(1)$ differences in physical predictions, without being a sign of a breakdown of the heavy quark expansion. The sensitivity of the $D^{**}$ modes to new physics is complementary and sometimes greater than those of the $D$ and $D^*$ modes [100,102]. Thus, using HQET, the predictions for $B \to D^{**}\tau\bar{\nu}$ are systematically improvable by better data on the $e$ and $\mu$ modes, just like they are for $B \to D^{(*)}\tau\bar{\nu}$ [75], and are being implemented in HAMMER [103–105].

2.6 New physics in $B \to D^{(*)}\tau\nu$

Independently of the recent discussion on form factor parametrizations and their influence on the extraction of $V_{cb}$ (covered in Sect. 2.1) it is clear from Table 3 that the SM cannot accommodate the present experimental data on $R(D^{(*)})$. Even after the inclusion of the most recent Belle measurement [106], the significance of the anomaly remains $3.1\sigma$. This leaves, apart from an underestimation of systematic uncertainties on the experimental side, NP as an exciting potential explanation. The required size of such a contribution comes as a surprise, however: defining $\hat{R}(X) \equiv R(X)/R(X)_{SM}$, the new average corresponds to $\hat{R}(D) = 1.14 \pm 0.10$ and $\hat{R}(D^{*}) = 1.14 \pm 0.06$; for NP to accommodate these data, a contribution of 5–10% relative to a SM tree-level amplitude is required for NP interfering with the SM, and $O(40\%)$ for NP without interference. An effect of this size can be clearly identified with upcoming measurements by LHCb and Belle II [107,108]. It would also immediately imply large effects in other observables.

The potential of $R(D^{(*)})$ as discovery modes does not diminish the importance of additional measurements with $b$-hadrons. Specifically, even with a potential discovery, model discrimination will require measurements beyond these ratios. These additional measurements fall in four categories:

- Additional $R(X)$ measurements such as $R(D^{**})$, $R(\Lambda_c)$, $R(X_c)$, $R(J/\psi)$ and $R(B^{0}_{s}^{(*)})$, are important cross-checks to establish $R(D^{(*)})$ as NP with independent systematics and provide independent NP sensitivity (especially $R(X_c)$ and $R(\Lambda_c)$), as discussed in Sects. 2.3 and 2.5. Note, however, the existence of an approximate sum rule relating the NP contributions to $R(\Lambda_c)$, $R(D)$, and $R(D^{*})$ [109].

- Integrated angular and polarization asymmetries and polarization fractions are excellent model discriminators. In many models they are completely determined once the measurements of $R(D^{(*)})$ are taken into account. For instance, the recent measurement of the longitudinal polarization fraction of the $D^*$ in $B \to D^*\tau\bar{\nu}$, $F_L(D^*)$, was able to rule out solutions that remained compatible with the whole set of the remaining $b \to c\tau\nu$ data [109–114]. The model-discriminating potential of both $R(D^{(*)})$ and selected angular quantities is visualized in Fig. 2, where fit results for pairs of $B \to D^{(*)}\tau\nu$ observables within all phenomenologically viable single-mediator scenarios with left-handed neutrinos to the state-of-the-art data are shown.

- Differential distributions in $q^2$ and the different angles are extremely powerful in distinguishing between NP models, as can be seen for instance from a recent analysis of data with light leptons in the final state [86]. They require, however, large amounts of data and the insufficient information on the decay kinematics can pose difficulties for the interpretation of the data, as discussed in Sect. 2.7. However, already the rather rough available information on the differential rates $d\Gamma/dq^2(B \to D^{(*)}\tau\nu)$ [58,60] is excluding relevant parts of the parameter space [114–118].

- An analysis of the flavor structure of the observed effect, e.g. in $b \to c(e,\mu)\nu$, $b \to u\tau\nu$ and $t \to b\tau\nu$ transitions.

In addition to the above observables, the leptonic decay $B_c \to \tau\nu$ plays a special role. Although it is not expected to be measured in the foreseeable future, it provides nevertheless a strong constraint on NP, since the relative influence of scalar NP is enhanced in this mode. A limit can then be obtained even from the total width of the $B_c$ meson [119]. Theoretical estimates for the partial width assumed unaffected by NP can be used to strengthen these bounds [117,120,121], and also data from LEP [122]. Both approaches rely on additional assumptions, however, see Refs. [109,112] for recent extensive discussions.

The constraints discussed so far are relevant in any scenario trying to address the existing anomalies. An interesting subclass of such models is that where the existence of a single mediator coupling to only the known SM degrees of freedom is assumed, classified in [116], creating only a subset of the possible operators at the $b$ scale. Among those, only five scenarios remain that can reasonably well accommodate the data described above, see also Refs. [84,109,116,118,123–127] for comparisons (additional constraints in specific scenarios are commented on below): Scenario I yields only a left-handed vector operator, created by either a heavy color-less vector particle [128–131] (phenomenologically highly disfavored) or a leptoquark, see Refs. [119,132–156] for this and other leptoquark variants. Scenario II includes Scenario I, but yields also a right-handed scalar operator, realized for example by a vector leptoquark. Scenario III involves both left- and right-handed scalar operators, generated for instance by a charged Higgs [83,117,157–162] (with a limited capability to accommodate $R(D^{*})$ due to the $B_c$ constraint discussed above). Scenarios IV and V involve the left-handed scalar and tensor operator which are generated proportionally to each

@ Springer
Fig. 2: State-of-the-art fit results in single-mediator models for selected pairs of observables in $B \rightarrow D^{(*)}\tau\nu$ decays (following Ref. [114] for form factor and input treatment). All outer ellipses correspond to 95% confidence level, inner (where present) to 68%. We show the SM prediction in grey, the experimental measurement/average in yellow (where applicable) and scenarios I, II, III IV and V in dark green, green, dark blue, dark red and red, respectively, see text. Contours outside the experimental ellipse imply that the measured central values cannot be accommodated within that scenario. The limit $BR(B_c \rightarrow \tau\nu) \leq 30\%$ has been applied throughout, but affects only the fits with scalar coefficients. Dark green contours are missing in the two graphs on the right, because the predictions of scenario I are identical to the SM ones.

other ($C_{S_L} = \pm 4C_T$ at the NP scale $\Lambda$), in the latter case with the addition of the left-handed vector operator, again realized in leptoquark models. It is also possible to analyze the available data in more general contexts. For example, within SMEFT the right-handed vector current is expected to be universal [163–165], see [114] for a global analysis in this framework, while this does not hold when the electroweak symmetry breaking is realized non-linearly [165]. Allowing for additional light degrees of freedom beyond the SM opens the possibility of contributions with right-handed neutrinos, see Refs. [145,166–172].

Once specific models are considered, typically additional constraints apply. Important ones include high-$p_T$ searches, looking for collider signatures of the mediators related to the anomaly [173–176], RGE-induced flavor-non-universal effects in $\tau$ decays [177], lepton-flavor violating decays [177], precision universality tests in quarkonia decays [178], charged-lepton magnetic moments [174] and electric dipole moments in models with non-vanishing imaginary parts [179].

2.7 Interpretation of experimental results

The reconstructed kinematic distributions used in measurements are sensitive to both the modeling of required non-perturbative inputs (e.g., form factors, light-cone meson wave functions), and to assumptions about the underlying fundamental theory (e.g., possible presence of operators with chiral structures different from those found in the SM). Current measurements assume the SM operator structure, and include the non-perturbative uncertainties as they are known at the time of publication. While this is a valid strategy for testing the SM, if in future the presence of a non-SM contribution with a different chiral structure is established then past measurements will require reinterpretation.

In order to present experimental results in such a way to allow a-posteriori analyses to have maximum flexibility in the description of non-perturbative inputs and BSM content, the following strategies might be considered. The techniques to allow for reinterpretation of results overlap with those used to make differential measurements designed to be sensitive to the chiral structure and non-perturbative quantities.

A first possibility, is the publication of unfolded distributions (see, for instance, the $B \rightarrow D^{(*)}\ell\nu$ spectrum presented in Ref. [53]). This method offers the possibility to fit with ease the experimental results to arbitrary parametrizations of the form factors [54,55,75]; its downside is that it requires relatively high statistics and that the unfolded distributions do not contain the whole experimental information.

A second option, which has been employed in the untagged Belle analysis of Ref. [51], is to provide folded distributions in which detector effects are not removed and no extrapolation is performed, together with experimental efficiencies and the detector response matrix (which reproduces detector effects to a given accuracy). This allows the use of any parametrization of SM and BSM effects in comparing with the experimental result. This approach, while requiring slightly more involved a posteriori fitting strategies, avoids the statistical problems associated with unfolding and can be extended more easily to higher dimensions.

Finally, the most complete information is contained in the Likelihood function which depends on a set of SM parameters (e.g., for $B \rightarrow \pi\ell\nu$ they could be the coefficients of the $z$-expansion of the form factors and $V_{ub}$) and on the Wil-
son coefficients of BSM operators). This method has not been currently pursued in any \( B \) decay measurement, in part because of difficulties related to the extremely large amount of information that would need to be presented. Two differing approaches are to publish the full experimental Likelihood in the full parameter space of BSM Wilson coefficients and SM non-perturbative coefficients, or to publish the tools for external readers to be able to repeat the full experimental fit with the signal model varied. For representing the experimental Likelihood in a high-dimensional space, possible approaches include the use of Markov chain sampling, or MVA surface modelling. These are the only strategies which would allow the entirety of the experimental information to be available in a posteriori theoretical investigations. It is essential to this approach for the experimental measurement to cover the full parameter space in a sufficiently general way, including alternative Likelihoods with different parametrizations for non-perturbative effects.

2.8 HAMMER

Future new physics searches in \( b \to c \tau \nu_\tau \) decays are a challenging endeavor: most experimental results make use of kinematic properties of the process to discriminate between the signal of interests and backgrounds. For instance recent kinematic properties of the process to discriminate between the non-resonant \( \gamma \ell \nu_\ell \), the non-resonant \( B \to \pi \pi \ell \nu_\ell \) decays, and some subtlety of the \( z \)-expansion.

3.1 Form factors for semileptonic \( b \)-hadron decays into light hadrons from lattice QCD

3.1.1 Form factor parametrizations

The matrix elements that describe the hadronic part of the semileptonic transitions \( \rightarrow X \ell \nu_\ell \) or \( B \rightarrow X \ell \ell \) are parametrized in terms of the form factors in Eqs. (2.1)–(2.3), where \( X \) now denotes a pion or kaon. The transitions \( B \rightarrow X^* \ell \nu_\ell \) or \( B \rightarrow X^* \ell \ell \) are parametrized in terms of the form factors in Eqs. (2.4)–(2.7), where \( X^* \) now denotes a \( \rho \), \( K^* \), or \( \phi \) meson. As discussed in Sect. 2.1, modern theoretical calculations of the form factors employ \( z \)-parametrizations to describe their shapes, which can be implemented in a model-independent way, being based on analyticity and unitarity constraints. For the case at hand, an often used choice for the \( z \)-parameter defined in Eq. (2.9) is \( t_0 = (M + m)/\sqrt{M^2 - m^2} \), which results in a range \( |z| < 0.3 \), centered around \( z = 0 \). In general, the small range of \( z \) coupled with unitarity constraints on the coefficients ensure that the polynomial expansions converge quickly. As discussed already in Sect. 2.1, for \( B \)-meson decays to light hadrons with their larger \( q^2 \) range, the BCL parametrization [8] is the standard choice, as the resulting forms satisfy the expected asymptotic \( q^2 \) and near threshold scaling behaviors [9, 10]:

\[
f_+(q^2) = \frac{1}{1 - q^2/M_{B^*(1-)}^2} \sum_{n=0}^{N_z-1} b_n^+(t_0) \left[ z^n - (-1)^n \frac{n}{N_z} z^{N_z} \right], \tag{3.1}
\]

\[
f_0(q^2) = \frac{1}{1 - q^2/M_{B^*(0^+)}^2} \sum_{n=0}^{N_z} b_n^0(t_0) z^n. \tag{3.2}
\]

3.1.2 Lattice QCD results for \( B \)-meson decay form factors to light pseudoscalars

Lattice-QCD calculations of the form factors for semileptonic \( B_{(s)} \)-meson decays to light hadrons proceed along the same lines as discussed in Sect. 2.2. In particular, there are a number of different, well-developed strategies for dealing with the heavy \( b \)-quark in lattice QCD, see Ref. [3] for a
review. The same two- and three-point functions as for the heavy-to-heavy case are needed here, albeit with the appropriate valence quark propagators, to describe the heavy-to-light decay process. While this affects the statistical errors in the next step, the fits to the spectral representations of the correlation functions to obtain the desired matrix elements on each gauge ensemble and each recoil momentum, the procedure is essentially the same. The resulting “lattice data” are then used in combined chiral-continuum fits coupled with a systematic errors analysis to obtain the form factors in the continuum over the range of recoil energies that are included in the simulations. Here, a well known challenge is that the recoil energies that are accessible in lattice-QCD calculations cover only a fraction of the entire kinematic region. A related challenge is that the validity of Chiral Perturbation Theory (used to extrapolate or interpolate to the physical pion mass) is limited to pion energies of \( \approx 1 \) GeV. The final step is the \( z \)-expansion fit, from which the form factors are obtained over the entire kinematic range, albeit with larger errors in the region not directly covered in the lattice calculation.

Lattice-QCD calculations of the \( B \rightarrow \pi \) vector current form factors \( f_+ \) and \( f_0 \) can be used to determine \(|V_{ub}|\) from experimental measurements of the \( B \rightarrow \pi \ell \nu\) decay rate. There are currently two independent, published lattice-QCD computations that employ the modern methods outlined above, including the model-independent \( z \)-expansion [180,181]. The RBC/UKQCD collaboration [180] uses ensembles with \( N_f = 2 + 1 \) flavors of Domain Wall fermions at two lattice spacings with sea-pion masses in the range \([300,400]\) MeV. The Fermilab/MILC collaboration [181] uses ensembles with \( N_f = 2 + 1 \) flavors of asqtad (improved staggered) fermion at four lattice spacings covering the range \( a \approx 0.045–0.12 \) fm and a range of sea-pion masses down to 177 MeV. Earlier work [182] used a subset of these ensembles. The treatment of the \( b \)-quark is similar in the two works; Ref. [180] uses a variant of the Fermilab approach, called the relativistic heavy quark (RHQ) action, while Ref. [181] employs the original Fermilab formalism. Both groups also use the mostly nonperturbative renormalization method to compute the renormalization factors. The form factors obtained by the two lattice groups are in good agreement with each other, and can be combined in joint fits together with experimental data for an improved \(|V_{ub}|\) determination [3].

Ongoing work by RBC/UKQCD is extending the calculation to include more ensembles [45]. Ongoing work by the Fermilab/MILC collaboration employs the HISQ \( N_f = 2 + 1 + 1 \) ensembles with sea-pion masses at (or near) the physical point, and the Fermilab formalism for the \( b \)-quark [183]. The HPQCD collaboration has published a calculation of the scalar form factor for the \( B \rightarrow \pi \) transition at zero recoil \( f_0(q^2_{max}) \) on a subset of the \( N_f = 2 + 1 + 1 \) HISQ ensembles and treating the \( b \)-quark in NRQCD [184], which provides a nice test of the soft-pion theorem, but cannot be used in \(|V_{ub}|\) determinations. Ongoing work includes a calculation of the \( B \rightarrow \pi \) form factors over a range of \( q^2 \) on a subset of the asqtad ensembles using NRQCD \( b \)-quarks and HISQ light-valence quarks [185]. The JLQCD collaboration has an ongoing project to calculate the \( B \rightarrow \pi \) form factors on \( N_f = 2 + 1 \) Domain Wall ensembles using also Domain Wall fermions for the heavy and light valence quarks [186]. They focus their calculation on small lattice spacings \((a \approx 0.044–0.080 \) fm\) and include a series of heavy-quark masses to extrapolate to the physical \( b \)-quark mass.

The vector current form factors \( f_+ \) and \( f_0 \) needed for rare \( B \rightarrow \pi \ell \nu \) decay are the same as for \( B \rightarrow \pi \ell \nu \) decay (up to small isospin corrections), but the tensor form factor \( f_T \) is also needed to describe the rare process in the SM, while it can contribute to \( B \rightarrow \pi \ell \nu \) decay only in BSM theories. So far, \( f_T \) has been calculated only by the Fermilab/MILC collaboration [187] using the same ensembles and methods as for the vector current form factors. However, most (if not all) of the ongoing projects described above, now include the complete set of form factors in their analyses, and new results for this form factor will therefore also be forthcoming.

The \( B_s \) to \( K \ell \nu \) process can be used for an alternate determination of \(|V_{ub}|\), and there currently are three independent, published lattice-QCD computations of the vector-current form factors [180,188,189]. In Ref. [188] the HPQCD collaboration used NRQCD \( b \)-quarks and HISQ light-valence quarks to calculate the form factors on a subset of asqtad ensembles. The RBC/UKQCD [180] work is already described above, since they calculated the \( B_s \) to \( K \) and \( B \rightarrow \pi \) transition form factors together. The Fermilab/MILC collaboration [189] used the same methods and set-up as for their \( B \rightarrow \pi \) project [181] but on a subset of asqtad ensembles. Both Fermilab/MILC [189] and, in a follow-up paper, HPQCD [190] also computed ratios of \( B_s \) to \( K \) and \( B_s \rightarrow D_s \) observables, which can be used in combination with LHCb measurements to determine \(|V_{ub}|/|V_{cb}|\).

**3.1.3 Challenges of vector mesons**

Lattice calculations of \( B_{(s)} \) decay form factors with vector mesons (\( \rho, K^*, \phi \)) in the final state are substantially more challenging, as these vector mesons are unstable resonances for sufficiently light quark masses. The asymptotic final state in the continuum then contains (at least) two hadrons, and the relation with the finite-volume matrix elements computed on the lattice becomes nontrivial. The formalism that allows a mapping of finite-volume to infinite-volume \( 1 \rightarrow 2 \) hadron matrix elements has been developed [191–197] and will be discussed in more detail below. First numerical applications to a form factor with nonzero momentum transfer have been published for the electromagnetic process \( \pi \gamma^* \rightarrow \pi \pi \),...
where the $\pi\pi$ final state in a $P$ wave couples to the $\rho$ resonance [198–200].

The lattice QCD calculations of $B(s) \to V$ form factors published to date did not implement this $1 \to 2$ formalism. For the $B \to \rho$ form factors, there is only an early study by the UKQCD collaboration [201], performed in the quenched approximation and with heavy up and down quark masses for which the $\rho$ is stable. For the $B \to K^*, B_s \to K^*, B_s \to \phi$ form factors, an unquenched lattice QCD calculation is available [202]. This work used three different ensembles of lattice gauge field configurations with pion masses of approximately 310, 340, and 520 MeV. For the lower two pion masses, the $K^*$ is expected to be unstable, but the analysis was performed as if the $K^*$ were stable. This entails using only a quark-antiquark interpolating field for the $K^*$, and assuming that the information extracted from exponential fits to the two-point and three-point correlation functions corresponds to the “$K^*$” contribution. The systematic errors introduced by this treatment are difficult to quantify. For unstable $K^*$, none of the actual discrete finite-volume energy levels directly corresponds to the resonance, and the actual ground state may be far from the resonance location (for typical lattice volumes, this problem is more severe at nonzero momentum). However, a quark-antiquark interpolating field couples more strongly to energy levels in the vicinity of the resonance, and ground-state saturation is typically not seen in the correlation functions before the statistical noise becomes overwhelming. In these cases, exponential fits are still dominated by one or multiple energy levels in the vicinity of the resonance.

In the following, we will denote the vector meson resonance as $V$, and the two pseudoscalar mesons whose scattering shows the resonance as $P_1$ and $P_2$. The finite-volume energy levels for a given total momentum and irreducible representation of the appropriate symmetry group are determined by the Lüscher quantization condition [203] and its generalizations, as reviewed in Ref. [204]. In the absence of interactions, they would consist of $P_1 P_2$ scattering states with energies equal to the sums of the $P_1$ and $P_2$ energies, where the $P_1$ and $P_2$ momenta take on the discrete values allowed by the periodic boundary conditions. Through the $P_1 P_2$ interactions, these energy levels are shifted away from their non-interacting values in a volume-dependent way. In the simplest case (considering only elastic scattering and neglecting the partial-wave mixing induced by the finite volume), each interacting finite-volume energy level can be mapped to a corresponding value of the infinite-volume $P_1 P_2$ scattering phase shift, or, equivalently, scattering amplitude; more complicated cases with coupled channels and partial-wave mixing can also be treated. The dependence of the scattering amplitude on the $P_1 P_2$ invariant-mass-squared, $s$, can be described by a Breit-Wigner-type function. By analytically continuing the scattering amplitude to complex $s$, one finds poles on the second Riemann sheet at $s = (m_V \pm i \Gamma_V/2)^2$, where $\Gamma_V$ is the width of the resonance. This procedure has been applied successfully to the $\rho$, $K^*$, and other resonances (see Ref. [204] for a review).

The $B(s) \to V$ form factors correspond to the residues at the pole at $s = (m_V \pm i \Gamma_V/2)^2$ in the $B(s) \to P_1 P_2$ form factors, where the $P_1 P_2$ final state is projected to the $\ell = 1$ partial wave. These $B(s) \to P_1 P_2$ finite-volume matrix elements both for the ground states and multiple excited states. (ii) Obtain the infinite-volume $P_1 P_2$ scattering amplitude from the finite-volume energy spectrum using the Lüscher method, and fit a suitable parametrisation of the $s$-dependence to the data. (iii) Map the finite-volume $B(s) \to P_1 P_2$ matrix elements to infinite-volume $B(s) \to P_1 P_2$ matrix elements using the Lellouch–Lüscher factor, which depends on the energy-derivative of the scattering phase shift and a known finite-volume function.

The finite-volume formalism requires the center-of-mass energy $\sqrt{s}$ to be small enough so that no more than two particles can be produced by the scattering through the strong interaction (however, the total momentum of the $P_1 P_2$ system can in principle be arbitrarily large). For example, in the case of the $B \to \pi\pi$ form factors, the formalism requires $\sqrt{s} \lesssim 4 m_\pi$, which becomes more restrictive when performing the calculation at lighter quark masses. However, it is likely that the coupling to four pions has negligible effects even at somewhat higher values of $\sqrt{s}$, as needed to map out the $\rho$ resonance region when using physical quark masses.

### 3.1.4 $\Lambda_b \to p$ and $\Lambda_b \to \Lambda^{(*)}$ form factors from lattice QCD

The $\Lambda_b \to p$ form factors relevant for the decay $\Lambda_b \to p \mu^- \bar{\nu}_\mu$ have been computed in lattice QCD together with the $\Lambda_b \to \Lambda_c$ form factors [16]; some aspects of this work were already discussed in Sect. 2.2.2. The lattice data for $\Lambda_b \to p$ cover the kinematic range from $q^2 \approx 15$ GeV$^2$ to near $q^2_{\text{max}} \approx 22$ GeV$^2$, and consequently the predicted $\Lambda_b \to p \mu^- \bar{\nu}_\mu$ differential decay rate is most precise in this range. The integrated decay rates in the Standard Model were found to be

$$
\frac{1}{|V_{ub}|^2} \Gamma(\Lambda_b \to p \mu^- \bar{\nu}_\mu) = (25.7 \pm 2.6_{\text{stat}} \pm 4.6_{\text{syst}}) \text{ ps}^{-1}
$$

and

$$
\frac{1}{|V_{ub}|^2} \int_{15 \text{ GeV}^2}^{q^2_{\text{max}}} \frac{d\Gamma(\Lambda_b \to p \mu^- \bar{\nu}_\mu)}{dq^2} dq^2 = (12.31 \pm 0.76_{\text{stat}} \pm 0.77_{\text{syst}}) \text{ ps}^{-1}.
$$
The latter has a total uncertainty of 8.8% (corresponding to a 4.4% theory uncertainty in a $|V_{ub}|$ determination from this rate), and the ratio to the partially integrated $A_b \to A_c \mu^- \bar{v}$ decay rate (2.38) has a total uncertainty of 9.8%, corresponding to a 4.9% theory uncertainty in the determination of $|V_{ub}/V_{cb}|$ performed by LHCb [205], commensurate with the experimental uncertainty. The $A_b \to p$ form factors from Ref. [16] can also be used to predict the Standard-Model value of the baryonic $b \to u \ell \bar{v}$ lepton-flavor-universality ratio,

$$\frac{\Gamma(A_b \to p \tau^- \bar{v}_\tau)}{\Gamma(A_b \to p \mu^- \bar{v}_\mu)} = 0.689 \pm 0.058_{\text{stat}} \pm 0.064_{\text{syst}}.$$  (3.5)

By increasing statistics, removing the partially quenched data sets (cf. Sect. 2.2.2), adding one ensemble with physical light-quark masses, and another ensemble with a third, finer lattice spacing, it will likely be possible to reduce the uncertainties in both the $A_b \to p$ and $A_b \to A_c$ form factors by a factor of 2 in the near future.

The same methods have also been used to compute the $A_b \to A$ [206], $A_c \to p$ [207], and $A_c \to A$ [208] form factors with lattice QCD. The latter calculation already includes an ensemble with the physical pion mass, and gave results for the $A_c \to A \pi^+ \nu_e$ and $A_c \to A \mu^+ \nu_\mu$ branching fractions consistent with, and two times more precise than, the measurements performed recently by the BESIII Collaboration [209,210]. This is a valuable test of the lattice methods used to determine the heavy-baryon decay form factors.

A lattice-QCD calculation is also in progress for the $A_b \to A^*(1520)$ form factors (in the narrow-width approximation) [211], which are relevant for the rare decay $A_b \to A^* \to p K \mu^+ \mu^-$. As with $A_b \to A^*$, discussed in Sect. 2.2.2, this initial calculation only reaches $q^2$ in the vicinity of $q^2_{\text{max}}$.

3.2 Light-cone sum rules calculations of heavy-to-light form factors

QCD sum rules on the light cone (LCSR) is a non-perturbative method for calculating hadronic quantities [212–214]. It has been applied to obtain the form factors for $B$ decays (see the definitions in Sect. 2.1). The first LCSR calculations relevant for $V_{ub}$ were performed in 1997 when the next-to-leading order (NLO) twist-2 corrections to $f_+(q^2)$ were calculated [215,216]. The leading order (LO) corrections up to twist-4 were calculated in Ref. [217]. Since the LO twist-3 contribution was found to be large, further improvements were made by calculating the smaller NLO corrections [218]. A more recent update where the $MS$ mass is used in place of the pole mass for $m_b$ can be found in Refs. [219,220] for the $B \to \pi$ case and in Ref. [221] for the $B_s \to K$ case.

Here we will discuss a selection of the more recent LCSR calculations.

For $B \to \pi$, a NNLO $O(\alpha_s^2 \beta_0)$ calculation of $f_+(0)$ was performed, with the result $f_+(0) = 0.262^{+0.020}_{-0.023}$ with uncertainties $\lesssim 9\%$ [222]. This calculation tested the argument that radiative corrections to $f_+, f_-$ and $f_B$ should cancel when both calculated in sum rules (the 2-loop contribution to $f_B$ in QCDSR is sizeable). It was found that despite $\sim 9\% O(\alpha_s^2 \beta_0)$ change to $f_B$, the effect on $f_+(0)$ was only $\sim 2\%$.

More recently unitarity bounds and extrapolation were used to perform a Bayesian analysis of the form factor $f_+(q^2)$ for $B \to \pi$ [223]. Prior distributions were taken for inputs, a likelihood function was constructed based on fulfilling the sum rule for $m_B$ to 1%, and posterior distributions were obtained using Bayes’ theorem. The posterior distributions of the inputs differed only for $s_0$, which was pushed to higher values $s_0 = 41 \pm 4$ GeV (mainly due to the choice of $m_b$). Finally the results were fit to the BCL parametrisation, finding a central value of $f_+(0) = 0.31 \pm 0.02$. Obtaining $f_+(q^2)$ and the first two derivatives at 0 and 10 GeV$^2$ has allowed the extrapolation to higher $q^2$ using improved unitarity bounds.

$V_{ub}$ can also be obtained from the channels $B \to \rho/\omega$, and updated LCSR results were made available in 2015 [224]. The improvements in these results include: the computation of full twist-4 (+partial twist-5) 2-particle DA contribution to FFs, plus the determination of certain so-far unknown twist-5 DAs in the asymptotic limit; a discussion of the non-resonant background for vector meson final states; the determination and usage of updated hadronic matrix elements, specifically the decay constants; fits with full error correlation matrix for the $\varepsilon$ expansion coefficients, as well as an interpolation to the most recent lattice computation. The result for $|V_{ub}|$ from $B \to \rho \ell \nu$ has comparable errors to the $B \to \pi$ determination. In general the $B \to V$ results agree with previous exclusive determinations and global fits within errors.

Future prospects for exclusive $V_{ub}$ from LCSR include extending the subset of NNLO corrections calculated both in $q^2$ and to include all NNLO twist 2 and 3 contributions. It would also be beneficial to perform a Bayesian uncertainty analysis of all $B \to P, D \to P$ LCSRs (along the lines of the aforementioned analysis for $B \to \pi$ [223]). Finally the measurement of $B_s \to K \ell \nu$ at LHCb/Belle II will allow an important complementary determination of $V_{ub}$ using results from Ref. [225].

3.3 Measuring $|V_{ub}|$ exclusively and the prospects for Belle II

The most precise exclusive determinations of $|V_{ub}|$ will ultimately come from the most theoretically clean $b \to u \ell^- \bar{v}_\ell$ modes: $B^0 \to \pi^+ \ell^- \bar{v}_\ell$, $B^+ \to K^+ \ell^- \bar{v}_\ell$ and $A_{1p}^0 \to p \ell^- \bar{v}_\ell$, which involve ground state hadrons in the final state. The main challenge facing measurements of $|V_{ub}|$ from these
modes is the large background from \( b \to c\ell^-\bar{\nu}_\ell \) decays, which is \( \text{O}(|V_{cb}|^2)/|V_{ub}|^2 \) \( \approx 100 \) more likely to occur. This background is difficult to separate from signal given the need to partially reconstruct the missing signal neutrino.

Several measurements of exclusive \( B^0 \to \pi^+\ell^-\bar{\nu}_\ell \) decays were made at the \( B \) factories CLEO, Babar and Belle. These measurements fall in to two categories of tagged and untagged measurements, which exploit the unique \( e^+e^- \to T(4S) \to B\bar{B} \) topology and fully hermetic detector design of the \( B \) factories. In tagged measurements [226] the non-signal \( B \) meson in the event is first reconstructed in a number of hadronic modes before selecting the signal pion and lepton. Exploiting the known energies and momenta of the interacting \( e^+e^- \) beams allows for neutrino 4-momentum, \( p_\nu \), to be reconstructed and the signal to be extracted using the missing mass squared of the neutrino, \( M^2 = p_\nu^2 \). In untagged measurements [227,228] the signal pion and lepton are first selected with a tight selection to reduce background from \( b \to c\ell^-\bar{\nu}_\ell \) decays. The neutrino is then reconstructed by inclusively reconstructing the other \( B \) in the event as a sum of remaining tracks and photons. The beam constrained mass, \( M_{bc} \), and beam energy difference\(^3\) are used as fit variables to simultaneously extract the signal. While tagged measurements give a high purity and better \( q^2 \) resolution they suffer from a much lower efficiency resulting from the branching fractions and reconstruction efficiencies for tagged modes.

In both tagged and un-tagged measurements the exclusive \( B^0 \to \pi^+\ell^-\bar{\nu}_\ell \) signal is fitted in bins of \( q^2 \) to determine the partial branching fraction in each bin. These measurements together with LQCD and LCSR predictions can be used as constraints to simultaneously fit the form factors of decays and determine the parameter \( |V_{ub}| \). HFLAV performed a fit for \( |V_{ub}| \) the \( B^0 \to \pi^+\ell^-\bar{\nu}_\ell \) form factor, \( f_+(q^2) \), under a BCL parametrisation utilising Babar and Belle tagged and untagged datasets and state of the art theory predictions [68]. This resulted in the most precise determination of \( |V_{ub}| \) to date, \( |V_{ub}| = 3.67 \pm 0.09(\text{exp}) \pm 0.12(\text{theo}) \), which has a total uncertainty of 4%.

Untagged and tagged measurements of \( |V_{ub}| \) from \( B^0 \to \pi^+\ell^-\bar{\nu}_\ell \) decays at Belle II will significantly improve the precision on \( |V_{ub}| \). In order to project the reduction in uncertainty both tagged and untagged analyses were performed on simulated Belle II Monte Carlo. The expected uncertainty on \( |V_{ub}| \) was determined for a given luminosity by extracting the partial branching fractions from pseudo-datasets generated from Monte Carlo expectations and fitting these together with LQCD predictions. With 50 ab\(^{-1}\) and future expected improvements in LQCD predictions the projected uncertainties on \( |V_{ub}| \) from \( B^0 \to \pi^+\ell^-\bar{\nu}_\ell \) decays were 1.7% (tagged) and 1.3% (untagged). The dominant systematic for the tagged analysis is the calibration of the tagging efficiency which is assumed irreducible at 1% on \( |V_{ub}| \). For the untagged analysis the dominant systematic uncertainty results from the uncertainty on the number of \( B\bar{B} \) pairs which is assumed irreducible at 0.5%. Several systematics relating to the branch fractions and form factors of \( b \to c\ell^-\bar{\nu}_\ell \) and \( b \to u\ell^-\bar{\nu}_\ell \) decays are also considered irreducible in the untagged analysis given the lower purity than the tagged analysis.

### 3.4 Measuring \( |V_{ub}|/|V_{cb}| \) at LHCb

All \( b \)-hadron species are accessible at hadron colliders thus opening to LHCb a wide possibility of \( |V_{ub}| \) measurements from exclusive \( b \to u \) transitions, while inclusive \( |V_{ub}| \) measurements do not seem feasible at the moment. In proton-proton collision at high energy \( b\bar{b} \) quark pairs are produced mainly from gluon splitting and hadronize independently, as a consequence \( b \)-hadrons have a wide continuum momentum spectrum and the reconstruction of semileptonic decays can not profit of the beam-energy constraints used at B-factories. However, thanks to the large boost acquired by the \( b \)-hadrons, the direction of the momentum can be well determined by the vector connecting the primary vertex of proton-proton interactions and the \( b \)-hadron decay vertex. By imposing the \( b \)-hadron mass constraint, the missing neutrino momentum can be calculated with a two-fold ambiguity. A small fraction of unphysical solutions arises from the imperfect reconstruction of vertices positions. The best way to choose between the two solutions depends on the specific decay mode under study. The choice can be optimized considering additional variables related to the decay kinematics by using linear regression algorithms [229].

The precise determination of an absolute branching fraction requires the precise knowledge of the total \( b \)-hadron production rate and of the experimental detection efficiency, which includes reconstruction, trigger and final states selection. To minimize the experimental uncertainty it is preferred to determine ratios of branching fractions, normalizing the \( b \)-hadron decay mode under study to a well-known \( b \)-hadron decay mode, that has as similar as possible decay topology. Choosing a decay of the same \( b \)-hadron removes the dependence on the production fraction of the specific \( b \)-hadron.

The first determination of \( |V_{ub}| \) at LHCb was done with baryons, measuring the branching fractions for \( A_0 \to \mu^+\nu \) and \( A_0 \to \Lambda^+_c \mu^-\bar{\nu}_\mu \) decays [205]. What is directly determined is the ratio of the CKM matrix elements

\[
\frac{|V_{ub}|^2}{|V_{cb}|^2} = \frac{\mathcal{B}(A_0^0 \to p\mu^-\bar{\nu})}{\mathcal{B}(A_0^0 \to \Lambda^+_c \mu^-\bar{\nu})} \times R_{FF}
\]

where \( R_{FF} \) is the ratio of the relevant form factors, calculated using LQCD. The ratio represents a band in the \( |V_{ub}| \) versus \( |V_{cb}| \) plane and can be converted into a measurement of \( |V_{ub}| \) using existing measurements of \( |V_{cb}| \). Approximately 10%
of $b$-hadrons produced at LHC are $A_{b}$ and a clean signal identification is possible imposing stringent proton identification requirements. The large background from $b$-hadron decays with additional charged tracks in the decay products is strongly reduced employing isolation criteria by means of multivariate machine-learning algorithms. The signal yields are determined from a $\chi^2$ fit to the B corrected mass distributions of $A_{b}^0 \rightarrow p \mu^- \bar{\nu}$ and $A_{b}^0 \rightarrow A_{1}^0 \mu^- \bar{\nu}$ candidates. The corrected mass is defined as $m_{corr} = \sqrt{m_{b\mu}^2 + p_{T\perp}^2 + p_{L}}$ where $p_{L}$ is the momentum of the hadron-$\mu$ pair transverse to the $A_{b}$ flight direction.

The LQCD form-factors that are used in the calculation of $|V_{ub}|$ [16] are most precise in the kinematic region where $q^2$, the invariant mass squared of the leptonic system, is high. When the branching fractions of the $b \rightarrow u \ (b \rightarrow c)$ decays are integrated in the region $q^2 > 15(7)$ GeV$^2$ the theory uncertainty on $|V_{ub}|/|V_{cb}|$ is 4.9%. This measurements, performed with Run 1 data, gives $|V_{ub}|/|V_{cb}| = 0.83 \pm 0.004$ (stat) $\pm 0.004$ (syst), consistent with previous exclusive measurements of the two CKM matrix elements.

A new measurement of this type is currently under study at LHCb. It uses $B_s^0 \rightarrow K^{+}\mu^-\bar{\nu}$ decays whose branching fraction is predicted to be of the same order of magnitude of the $A_{b}^0 \rightarrow p \mu^- \bar{\nu}$ one.

The signal selection is challenging due to the large background from partially reconstructed decays of all species of $b$-hadrons, but it can exploit the good efficiency and purity of kaon and muon identification provided by the LHCb detector, the separation of the $K\mu$ vertex from primary vertex and the already mentioned isolation tools. The chosen normalization mode $B_s^0 \rightarrow D_{s}^{\pm}\mu^-\bar{\nu}$, $D_{s}^{\pm} \rightarrow K^{-}K^{+}\pi^{\pm}$ benefits of small uncertainty in the $D_{s}^{+}$ branching fraction. The good identification of this decay mode, despite the large feed-down from $B_s^0$ decays to excited $D_s$ mesons with un-reconstructed neutral particles, has been proven to be possible at LHCb with the measurement of $B_s^0$ lifetime [230].

Form factors for the $B_s^0$ mesons decays to $K$ and $D_s$ have been calculated with LQCD by several groups [180,188]. The calculation are performed in the high $q^2$ region and extrapolated to the full region with BGL or BCL $z$-expansions. Different calculations agree at high $q^2$, but there is currently a disagreement in the $q^2 \approx 0$ extrapolated value. For $B_s^0 \rightarrow K^{+}\mu^-\bar{\nu}$ in the low $q^2$ region (up to 12 GeV$^2$) form factors calculated with LCSF are also available [225].

The uncertainties on the experimental measurement of the $B_s^0 \rightarrow K^{+}\mu^-\bar{\nu}$ yield increase at high $q^2$ (low kaon momentum) due to the reduced efficiency and the larger background contamination. It is foreseen to perform the measurement in few $q^2$ bins so that the use of different calculations of form factors will be possible. Larger data samples, accumulated during the LHCb Upgrade period will allow a differential measurement in finer $q^2$ bins.

Purely leptonic $B^- \rightarrow \mu^-\bar{\nu}$ decays are not accessible at LHCb. An alternate way has been tested, searching for the decay $B^- \rightarrow \mu^-\bar{\nu}\mu^+\bar{\mu}^- \rightarrow$ an hard photon is irradiated from the initial state and materializes into two muons. This decay has the experimental advantage of the presence of additional particles in the final state and of a larger branching fraction, due to the removal of the helicity suppression. An upper limit on the branching fraction of $1.6 \times 10^{-8}$ has been determined with $4.7 \text{ fb}^{-1}$ of integrated luminosity [4], making it a possible candidate for a $|V_{ub}|$ measurement in the LHCb Upgrade period [231].

3.5 Related issues

3.5.1 $R_\pi$

The experimental signature of $B \rightarrow \pi \tau \nu_\tau$ is challenging: low in rate due to CKM suppression, this final state can only be isolated from backgrounds using multivariate analysis techniques. Due to the pseudoscalar nature of the pion in the final state, an increased sensitivity to certain new physics models involving scalar exchange particles is expected and measurements of this branching fraction offer an orthogonal path to probe the anomalies observed in $R(D)$ and $R(D^*)$. The first limit on the branching fraction using lepton and one-prong $\tau$ decay modes was reported by Ref. [232]. They reported

$$B(B^0 \rightarrow \pi^- \tau^+ \nu_\tau) < 2.8 \times 10^{-4} \text{ at 95% CL,} \quad (3.6)$$

using a frequentist method. This result can be converted into a value of $R_\pi = \Gamma(B^0 \rightarrow \pi^- \tau^+ \nu_\tau)/\Gamma(B^0 \rightarrow \pi^- \ell^+ \nu_\ell)$ with $\ell = e, \mu$ of

$$R_\pi = 1.05 \pm 0.51, \quad (3.7)$$

which can in turn be compared to the SM prediction of Refs. [181,233] of

$$R_\pi = 0.641 \pm 0.016. \quad (3.8)$$

Although the current precision is very limited, this result can already exclude the model parameter space of new physics models e.g. charged Higgs bosons, cf. Ref. [233]. Albeit a challenging signature, the final state with a charged pion has excellent prospects to be discovered in the large future Belle II data set. A naive extrapolation of Eq. 3.7 using assuming SM couplings results in evidence with $4 \text{ ab}^{-1}$ and discovery with $11 \text{ ab}^{-1}$ of integrated luminosity. The theoretical precision in $R_\pi$ will further increase with progress in lattice and with combined light lepton and lattice fits (the measured spectra can constrain the low $q^2$ region, which the lattice has difficulties in predicting reliably).
3.5.2 Experimental status and prospects of $B \to \ell \nu \ell \gamma$

The experimental study of $B \to \ell \nu \ell \gamma$ with $\ell = e, \mu$ is challenging and requires the clean laboratory of an $e^+ e^-$ machinery: in such a setting the known initial state and the full reconstruction of the second $B$-meson produced in the collision provide the necessary constraint to successfully identify this signature. In addition, to not be overwhelmed with background, only photons at high energies ($\approx 1$ GeV or larger) can be studied this way. The difficulties lie in the low efficiency of the reconstruction performance of the second $B$-meson, which have to happen in low branching fraction hadronic modes, and the still sizeable cross-feed from $B \to \pi^0 \ell \nu \ell$ and $B \to \eta \ell \nu \ell$ decays. These two semileptonic processes produce very similar final states, namely $B \to \ell \nu \ell \gamma$, but can be reduced by looking for an unassigned second high-energetic photon in the collision event under study. To separate $B \to \ell \nu \ell \gamma$ from such decays successfully a fit to

$$m_{\ell
\\nu}^2 \simeq m_{\text{miss}}^2 = (p_{\text{miss}} - p_{\ell} - p_{\gamma})^2$$

(3.9)
can be carried out. Here $p_{\ell}$ and $p_{\gamma}$ denote the reconstructed four-vectors of the visible final states of $B \to \ell \nu \ell \gamma$. The four-vector of the decaying signal $B$-meson, $p_{\text{miss}}$, can be reconstructed using the information from the reconstructed tag-side $B$-meson. Correctly reconstructed signal decays peak at $m_{\ell \nu}^2 \approx 0$ GeV$^2$, whereas the dominant semileptonic decays are shifted to higher values due to the absence of the additional photon in the four-vector sum. The sensitivity can be further increased by explicitly reconstructing the semileptonic backgrounds and combine this information into a global analysis. This was the strategy pursued by Ref. [234], which constrained the $\pi^0$ semileptonic background this way. The current experimental limit with a lower photon energy cut of 1 GeV is

$$\Delta B(B \to \ell \nu \ell \gamma) < 3.0 \times 10^{-6} \text{ at 95% CL.}$$

(3.10)
The above limit was determined using a flat Bayesian prior.

The discovery prospects for this decay at Belle II are excellent: the improved tracking capabilities, better calorimeter electronics, and the continuous development of modern tagging algorithms such as Ref. [235] will help improving the sensitivity. Extrapolating from the central value and uncertainty of the currently most precise limit of Eq. 3.10 of $\Delta B(B \to \ell \nu \ell \gamma) = (1.4 \pm 1.1) \times 10^{-6}$, evidence should be possible with 5 ab$^{-1}$ and a discovery is possible with 50 ab$^{-1}$ [236]. In principle, after discovery the value of $|V_{ub}|$ could be extracted from this decay as well, along with the first inverse momentum of the light-cone distribution amplitude, $\lambda_B$. An extrapolation from the current sensitivity is shown in Fig. 3, based on the numbers from Ref. [236]. The sensitivity for $|V_{ub}|$ will not be competitive with other methods (leptonic and semileptonic), but the achievable precision on $\lambda_B$ will help measurements and interpretations, which rely on our understanding of the light-cone distribution amplitude properties.

3.5.3 Theoretical progress for $B \to \gamma \ell \nu \ell$

The photoleptonic decay $B \to \gamma \ell \nu \ell$ determined by two independent form factors is the simplest probe of the $B$-meson light-cone distribution amplitudes (LCDAs), which represent one of the most important inputs in the theory of semileptonic and nonleptonic $B$-decays based on QCD factorization and LCSRs. The calculation of the form factors in HQET and at large photon recoil in the leading power is well developed and can be found in Ref. [237]. The $1/m_B$ and $1/E_\gamma$ power suppressed effects, expressed in a form of the soft overlap part of the form factors, were quantified using a technique [238] based on dispersion relations and quark-hadron duality (see also Ref. [239]). The most advanced calculation of the $B \to \gamma \ell \nu \ell$ form factors, including power suppressed terms, was done recently [240] resulting in the prediction of the decay branching fraction at $E_\gamma > 1.0$ GeV as a function of the key unknown theoretical quantity: the inverse moment $\lambda_B$ of the B-meson LCDA. An alternative approach [241] calculates the power-suppressed corrections due to photon emission at long distances in terms of the photon LCDAs in the LCSR framework. The proof of concept for a lattice QCD calculation of radiative leptonic decays was recently done in [242], see also [243].

3.5.4 $B \to \pi \pi \ell \nu \ell$ decay beyond $\rho$

Calculations of $B \to \rho$ form factors both in lattice QCD and from LCSRs usually adopt a narrow $\rho$ approximation and by default ignore the influence of nonresonant effects (radially excited $\rho$’s) in the mass interval around $\rho$. The role
of these effects has to be assessed at a quantitative level. In Refs. [244,245] the first attempt to calculate more general $B \to \pi\pi$ form factors from LCSRs, using two-pion LCDAs at low mass of dipion system and at large recoil, was undertaken. The currently limited knowledge of these non-perturbative inputs calls for their further development and also for alternative methods. In Ref. [246] a different version of LCSRs with $B$ meson LCDAs was obtained which predicts the convolutions of the $B^0 \to \pi^+\pi^0$ form factors in $P$ wave with the timelike pion form factor. In the narrow $\rho$-meson limit these sum rules reproduce analytically the known LCSRs for $B \to \rho$ form factors. Using data for the pion vector form factor from $\tau$ decay, the finite-width effects and the contribution of excited $\rho$-resonances to the $B \to \pi\pi$ form factors were found to amount up to $\approx 20\%$ in the small dipion mass region where they can be interpreted as a nonresonant ($P$-wave) background to the $B \to \rho$ transition. For a more general analysis of $B \to \pi\pi\ell\nu_\ell$ decays see e.g. Refs. [247,248].

3.5.5 Remarks on the $z$ expansion

The use of the so-called $z$ expansion for form factors has become a standard practice for semileptonic decays, see Refs. [249,250] for a pedagogical discussion. In the workshop several issues concerning it were discussed, in particular its application to baryon form factors.

Form factors which parametrize matrix elements of the form $\langle L|J|H\rangle$ have known analytic structure. In particular, they are analytic in the complex $t = q^2$ plane outside a cut on the real axis. The cut starts at some positive $t_{\text{cut}}$ equal to the invariant mass squared of the lightest state the current $J$ can produce. The domain of analyticity can be mapped onto the unit circle via $z = (\sqrt{t_{\text{cut}} - t} - \sqrt{t_{\text{cut}} - t_0}) / (\sqrt{t_{\text{cut}} - t} + \sqrt{t_{\text{cut}} - t_0})$, where $t_0$ is a free parameter denoting the point that is mapped to $z = 0$. The form factor can be expanded as a Taylor series in $z$ which is a model-independent parametrization. For heavy-to-light form factors the maximum value of $|z|$ is related to the distance between $(m_H - m_L)^2$ and $t_{\text{cut}}$. As a result, increasing $t_{\text{cut}}$ decreases the maximum value of $|z|$ leading to a faster convergence of the series.

Naively one might assume that the lightest state is the two-particle state $\bar{H}H$. This would imply that $t_{\text{cut}} = (m_H + m_L)^2$, but this is not the case in general. For example, for the proton electric and magnetic form factors ($H = L = p$) the cut starts at the two-pion threshold and not at the $p\bar{p}$ threshold. As another example, for one of the $B \to \pi$ form factors ($f_+$) the cut starts at $m_{B^*}^2$. Since this is a simple pole, it can be easily “removed” by considering $(t - m_{B^*})f_+$ as a Taylor series in $z$. For $(t - m_{B^*})f_+$ the cut starts at $(m_B + m_{\pi})^2$. If one uses a higher value of $t_{\text{cut}}$ than the physical one, one faces the danger of trying to expand the form factor in a region where it is not analytic. One of the immediate results of the workshop was the identification of such a problem in the literature. For baryon form factors, e.g. $A_B \to p$, analyses have used the wrong value of $t_{\text{cut}} = (m_{A_B} + m_p)^2$, see Ref. [251] and arXiv.org version 2 of Ref. [16]. In fact, $t_{\text{cut}}$ for the baryon form factors is the same as for the meson form factors of analogous decays.

Another issue discussed in the workshop is the use (or lack of use) of bounds on the coefficients of the $z$ expansion. Although the form factor is expressed as an infinite series, in practice the series is truncated after a few terms. One would like to ensure that the value of a physical parameter such as $|V_{ub}|$ is independent of the number of parameters used, by bounding the coefficients. For example, one can use a unitarity bound [252] or a bound from the heavy quark expansion [253]. It seems that currently there is no consistent use of bounds in the extraction of $|V_{ub}|$. As the analysis [254] shows, this can be a problem as the data improve and the number of necessary parameters increases. This can be especially problematic if one needs to use the $z$-expansion for extrapolation. The community needs to be aware of this issue and at least test that results do not change if bounds are applied to the coefficients.

The unitarity bounds for meson decays such as $B \to \pi$ rely on the fact that for $(t - m_{B^*})f_+$ the cut starts at the $(m_B + m_{\pi})^2$. For baryon decays such as $A_B \to p$, unitarity can only constrain the region above $(m_{A_B} + m_p)^2$. The region between $(m_B + m_{\pi})^2$ and $(m_{A_B} + m_p)^2$ is left unconstrained. Following the analysis of Ref. [254], one might worry that the contribution of the latter region is the dominant one. While considering together mesons and baryons contributions to the dispersive bounds might overcome the problem [88], further study is warranted.

4 Quark masses and leptonic decays

4.1 Quark masses

In the Standard Model (and many extensions), quark masses and the CKM matrix all stem from Higgs-Yukawa couplings between the quark fields and the Higgs doublet. It is therefore natural to consider the bottom-quark mass, $m_b$, in this report. As discussed in Sect. 5, $m_b$ can be extracted from the inclusive semileptonic $B$ decay distributions, along with $|V_{cb}|$. In the theory of inclusive decays, the charm-quark mass, $m_c$, is also needed to control an infrared sensitivity; see Sect. 5.

Figure 4 compares results from lattice QCD with realistic sea content of $n_f = 2 + 1 + 1$ or $2 + 1$ sea quarks with the FLAG 2019 [3] average for the $2 + 1 + 1$ sea. The average for $m_b$ is dominated by the very precise result from the Fermilab Lattice, MILC, and TUMQCD Collaborations, while that for $m_c$ is dominated by the corresponding
what smaller uncertainties based on Refs. [27,256–259] and [255,256,259–261], respec-
2
[3] stretches the error bar by

Fermilab/MILC/TUMQCD result together with two separate results from the HPQCD Collaboration. The FLAG 2019 [3] averages (for 2 + 1 + 1 sea flavors) are

\[ \bar{m}_b = m_{b,\overline{MS}}(m_{b,\overline{MS}}) = 4.198(12) \text{ GeV}, \]

(4.1)

\[ \bar{m}_c = m_{c,\overline{MS}}(m_{c,\overline{MS}}) = 1.280(13) \text{ GeV}, \]

(4.2)

based on Refs. [27,256–259] and [255,256,259–261], respectively. Another recent review [280] finds averages with somewhat smaller uncertainties

\[ \tilde{m}_b = m_{b,\overline{MS}}(m_{b,\overline{MS}}) = 4.188(10) \text{ GeV}, \]

(4.3)

\[ \tilde{m}_c = m_{c,\overline{MS}}(m_{c,\overline{MS}}) = 1.2735(35) \text{ GeV}, \]

(4.4)

based on the same original sources. In the case of \( \tilde{m}_c \), two results [260,261] agree poorly with the others, increasing \( \chi^2/\text{dof} \) of the average by a factor of around 5. FLAG 2019 [3] stretches the error bar by \( \sqrt{\chi^2/\text{dof}} \), while Ref. [280] discards them: the resulting error bar is smaller because of the compatibility of the inputs as well as the lack of stretching. Other differences in averaging methodology are quantitatively unimportant. In either case, the quoted averages are much more precise than those in the PDG.

A last remark is that the most precise results [255,256,259] all use the very high statistics MILC HISQ ensembles with staggered fermions for the sea quarks [23,281]. In the future, other groups [282–285] will have to collect similar statistics to enable a complete cross check.

Four distinct methods are used in the results shown in Fig. 4: (1) converting the bare lattice mass to the \( \overline{\text{MS}} \) scheme, (2) fitting to a formula for the heavy-light hadron mass in the heavy-quark expansion [286,287], and (3) computing moments of quarkonium correlation functions [288,289]. The first two require an intermediate renormalization scheme that can be defined for any ultraviolet regulator: quark masses defined this way can be computed with lattice gauge theory or dimensional regularization. For example, HPQCD 13 (\( \Upsilon \) decays) [265] uses two-loop lattice perturbation theory to convert the bare NRQCD mass to the pole mass [290,291], and dimensional regularization to convert the pole mass into the \( \overline{\text{MS}} \) mass.

Instead of the pole mass, one can use a regularization-independent momentum-subtracted mass [292]. Like the \( \overline{\text{MS}} \) scheme these RI-MOM schemes are mass-independent renormalization schemes, but they depend on the gauge. In lattice gauge theory, Landau gauge is easily obtained on each gauge-field configuration via a minimization procedure [293]. The mass renormalization factor, \( Z_m \), can be computed from the three point function for the scalar or pseudoscalar density, because \( Z_m^{-1} = Z_S = Z_P \) (up to technical details for Wilson fermions). For example, the matrix element \( \langle p'|\bar{P}|p \rangle \), between gauge-fixed quark states, can be used to define \( Z_P \) using the same formulas for lattice gauge

\[ e^+ e^- \rightarrow b \bar{b} \]

\[ e^+ e^- \rightarrow c \bar{c} \]

2 + 1 flavors of sea quark [262–266]; circles denote results extracted from \( e^+ e^- \) collisions near \( Q \bar{Q} \) threshold [267–279]. The vertical band shows the FLAG 2019 average for 2 + 1 + 1 sea flavors. Note that 2 + 1-flavor calculations are in rough (good) agreement for bottom (charm)

\[ 1.20 \quad 1.30 \quad 1.40 \]

(4.3)

(4.4)

(4.5)

(4.6)

(4.7)

(4.8)

(4.9)

(4.10)

(4.11)

(4.12)

(4.13)

(4.14)

(4.15)

(4.16)

(4.17)

(4.18)

(4.19)

(4.20)

(4.21)

(4.22)

(4.23)

(4.24)

(4.25)

(4.26)

(4.27)

(4.28)

(4.29)

(4.30)

(4.31)

(4.32)

(4.33)

(4.34)

(4.35)

(4.36)

(4.37)

(4.38)

(4.39)

(4.40)

(4.41)

(4.42)

(4.43)

(4.44)
theory as for continuum gauge theory (with dimensional regularization) \[292\]. The schemes labeled RI-MOM and RI-MOM use \( p' = p \) and slightly different definitions of the quark-field normalization \( Z_2 \); for a review see Ref. \[294\]. The momentum transfer \( q \equiv p' - p = 0 \) here, namely it is “exceptional” in the sense of Weinberg’s theorem \[295\]. On the other hand, the RI-sMOM scheme \[296\] chooses \( p' \) and \( p \) such that \( p^2 = q^2 = p'^2 \equiv \mu^2 \). Without the exceptional momentum, the extraction of \( Z_P \) is more robust. It would be interesting to see whether RI-sMOM on the ETM 2+1+1 ensembles yields \( \bar{m}_b \) favoring the RI’-MOM results or the RI-sMOM results on MILC’s ensembles.

The HQE method starts with the HQE formula for a heavy-light hadron mass \[297,298\],
\[
M = m + \bar{\Lambda} + \frac{\mu^2}{2m} - \frac{dJ}{2m} G_G(m) + \cdots ,
\]
(4.5)
where \( M \) is the hadron mass, which is computed in lattice QCD as a function of the quark mass, \( m \), and \( dJ \) depends on the spin of the hadron. The quantities can be identified with the energy of gluons and light quarks, \( \bar{\Lambda} \), the Fermi motion of the heavy quark, \( \mu^2 \), and the hyperfine splitting, \( \mu^2_G \). \( \mu^2_G \) depends logarithmically on \( m \). Although this idea is not new \[286,287\], to be precise one has to confront the definition of \( m \). Although the pole mass is natural in the context of the HQE, it is not suitable in practice, because of its infrared sensitivity. The \( \bar{M} \) scheme, on the other hand, breaks the power counting: \( m_{\text{pole}} - \bar{m}_{\text{MS}} \propto \alpha_s m_{\text{pole}} \). Instead, one chooses mass definitions that, in some sense, lie in between these two choices. Gambino et al. \[257\] choose the kinetic mass \[299\], while Fermilab/MILC/TUMQCD \[256\] choose the minimal renormalon subtracted (MRS) mass \[300\]. After extracting \( m_{\text{kin}} \) or \( m_{\text{MRS}} \) from fitting Eq. (4.5), the result can be converted to the \( \bar{M} \) scheme with three- and four-loop perturbation theory, respectively. In addition to the different matching, the error bar from Fermilab/MILC/TUMQCD is so small because it is based on the largest data set of all calculations in Fig. 4. See Sect. 5.3 for further discussion and results for \( \bar{\Lambda}, \mu^2, \mu^2_G, \) and higher-dimension corrections to the HQE.

One can avoid an intermediate scheme by computing a short-distance quantity in lattice QCD, taking the continuum limit, and analyzing the result with \( \bar{M} \) perturbation theory. For example, one can compute moments of quarkonium correlation functions \[288,289\],
\[
G^{(n)}_t = \sum_t t^n G(t),
\]
(4.6)
\[
G(t) = c_t \sum_x (\bar{Q} \Gamma Q(x, t) \bar{Q} \Gamma Q(0, 0)),
\]
(4.7)
for some Dirac matrix \( \Gamma \). In lattice gauge theory, the pseudoscalar density needs no renormalization if \( \Gamma = \gamma^5 \) and \( c_{\gamma^5} = m^2_Q \). The moments \( G^{(n)}_t \) are physical observables with a good continuum limit, which is proportional to \( m_Q \) to the appropriate power, multiplied by a dimensionless function of \( \alpha_s(m_Q) \). Thus, these moments also yield determinations of the strong coupling as well as quark masses. In Fig. 4, results obtained in this way are labeled “moments”.

The same moments \( G^{(n)}_t \) can be obtained from the cross section for \( e^+e^- \) annihilation into \( Q\bar{Q} \) hadrons via a suitably subtracted dispersion relation. In this case, \( \Gamma = \gamma^\nu \) for the electromagnetic current, and \( c_{\gamma^\nu} = 1 \) because the electromagnetic current is conserved. Thus, the same perturbative calculations (only changing \( \Gamma \)) can be used to extract the bottom- and charm-quark masses and \( \alpha_s \) from experimental measurements. The dispersion relation, related sum rules, and the perturbative series for the moments are the basis of the result labeled \( e^+e^- \to b\bar{b} \) and \( e^+e^- \to c\bar{c} \) in Fig. 4. The order \( \alpha_s^p, p = 1, 2, 3 \), became available in 1993 \[301\], 1997 \[302\], and 2006 \[278,303\], respectively.

4.2 Leptonic decays

Instead of semileptonic decays, CKM matrix elements can also be determined from purely leptonic decays. For example, a goal of Belle II is to improve the determination of \( V_{ub} \) from \( B^+ \to \tau^+\nu \), as well as \( V_{cd} \) from \( D^+ \to \ell^+\nu \) and \( V_{cs} \) from \( D_s^+ \to \ell^+\nu \), and a goal of LHCb to observe \( B_c \to \tau\nu \). The rates for leptonic decays suffer a helicity suppression, making tauonic and muonic decays preferred experimentally. Leptonic decays are mediated by the axial-vector part of the electroweak current, as well as possible pseudoscalar currents, so they complement semileptonic decays in this way.

The hadronic quantity describing the decay is known as the decay constant, defined by
\[
\langle 0| \bar{b} \gamma^\mu \gamma^5 u| B^+(p) \rangle = i p^\mu f_{B^+},
\]
(4.8)
where \( p^\mu \) is the four-momentum of the \( B \) meson and \( f_{B^+} \) is the decay constant. For other mesons, the axial currents and notation change in obvious ways. From the partial conservation of the flavor-nonsinglet axial current, the pseudoscalar density can also be used to compute the decay constant:
\[
(m_b + m_u) \langle 0| \bar{b} \gamma^5 u| B^+(p) \rangle = M_{B^+}^2 f_{B^+},
\]
(4.9)
where \( m_b \) and \( m_u \) are bare quark masses.

Equations (4.8) and (4.9) are the basis of lattice-QCD calculations. In general, the axial current used is not a Noether current, so it is not absolutely normalized. Fermion formulations with good chiral symmetry (staggered, overlap, domain wall) provide an absolutely normalized pseudoscalar density. Until recently, however, lattice spacings have not been small enough to use these approaches for the \( b \) quark. Methods developed especially for heavy quarks have therefore been used, and they do not provide any absolutely normalized \( \bar{b}\Gamma u \) bilinears.
calculation omitting QED. The results are arguably precise enough for the foreseeable future.

The results in Eqs. (4.10)–(4.15) again use the very high statistics MILC HISQ ensembles with staggered fermions for the sea quarks. Here the lattice spacing is, for some ensembles, small enough to reach the $b$ quark, so the calculation uses the HISQ action for all $b$ and light quarks alike. Thus, an absolutely normalized pseudoscalar density is available, so the uncertainty is essentially statistical, as propagated through a fit to the continuum limit with physical quark mass. Again, other groups will have to collect similar statistics in the future to enable a complete cross check.

To go beyond the precision quoted here, analyses of leptonic decays will have to include QED radiative corrections to the measured rates. The issues and an elegant solution for light mesons (pion and kaon) can be found in Refs. [312–314]. Radiative corrections for heavy-light mesons will be more difficult to incorporate, because of the hierarchy of soft scales $\Lambda_{QCD}$, $A_{QCD}/m_Q$, $A_{QCD}/m^2_Q$, etc.

5 Heavy-to-heavy inclusive

5.1 Heavy quark expansion for $b \rightarrow c$

5.1.1 Review of the current status

The heavy quark expansion (HQE) for the inclusive semileptonic $b \rightarrow c$ transitions starts form a correlation function for the $b \rightarrow c$ current

$$d\Gamma \propto \sum_x (2\pi)^4 \delta^4(P_B - P_X - q)(B(v)\bar{b}\gamma_\mu(1 - \gamma_5)c\gamma_\nu(1 - \gamma_5)b)(B(v))$$

$$= \int d^4x e^{iq\cdot x} \langle B(v)|\bar{b}(x)\gamma_\mu(1 - \gamma_5)c(x)\gamma_\nu(1 - \gamma_5)b|B(v)\rangle$$

$$= 2 \text{Im} \int d^4x e^{iq\cdot x} \langle B(v)|T[\bar{b}(x)\gamma_\mu(1 - \gamma_5)c(x)\gamma_\nu(1 - \gamma_5)b]|B(v)\rangle$$

$$\times (1 - \gamma_5)b|B(v)\rangle = 2 \text{Im} \int d^4x e^{i(m_b - q\cdot v)\cdot x}$$

$$\langle B(v)|T[\bar{b}_c(x)\gamma_\mu(1 - \gamma_5)c(x)\gamma_\nu(1 - \gamma_5)b_c]|B(v)\rangle$$

(5.1)

with

$$b(x) = e^{-im_b v\cdot x} b_c(x).$$

The time ordered product in the last line can be expanded in an operator product expansion which for large $m_b$ and $m_c$ yields an expansion in terms of local hadronic matrix elements which parametrize the hadronic input. Within this approach, the differential rate can be expressed as a series in $1/m$

$$d\Gamma = d\Gamma_0 + \left(\frac{\Lambda_{QCD}}{m_b}\right)^2 d\Gamma_2 + \left(\frac{\Lambda_{QCD}}{m_b}\right)^3 d\Gamma_3 + \left(\frac{\Lambda_{QCD}}{m_b}\right)^4 d\Gamma_4$$
\[ + d \Gamma_5 \left( a_0 \left( \frac{\Lambda_{QCD}}{m_b} \right)^5 + a_2 \left( \frac{\Lambda_{QCD}}{m_b} \right)^3 \left( \frac{\Lambda_{QCD}}{m_c} \right)^2 \right) \\
+ \cdots + d \Gamma_5 \left( \frac{\Lambda_{QCD}}{m_b} \right)^3 \left( \frac{\Lambda_{QCD}}{m_c} \right)^4 \right) \tag{5.2} \]

The coefficients \( d \Gamma_i \) are given by

\[ d \Gamma_i = \sum_k C_i^{(k)} \langle B(v) | O_i^{(k)} | B(v) \rangle \tag{5.3} \]

where the \( O_i^{(k)} \) are operators of mass-dimension \( i + 3 \) and the sum over \( k \) runs over all elements of the operator basis, \( C_i^{(k)} \) are coefficients that can be calculated in QCD perturbation theory as a series in \( \alpha_s(m_b) \). Note that starting at order \( 1/m_b^2 \) the \( b \rightarrow c \) HQE exhibits an infrared sensitivity to the charm quark mass; for the total rate, \( \Gamma_3 \) contains a \( \log(m_c^2) \) while \( \Gamma_5 \) contains inverse powers of \( m_c^2 \) which are explicitly shown in Eq. (5.2).

The leading term \( d \Gamma_0 \) is the partonic result which turns out to be independent of any unknown hadronic matrix element. This term is fully known (triple differential rate) at tree level, at order \( \alpha_s \) [315,316] and order \( \alpha_s^2 \) [316–320].

Due to heavy quark symmetry, there is no term \( d \Gamma_1 \) and the leading power corrections appear at order \( 1/m^2 \). These are given in terms of two non-perturbative matrix elements

\[ 2M_B \mu_\pi^2 = -\langle B(v) | \bar{b}_\nu (iD)^2 b_\nu | B(v) \rangle \tag{5.4} \]

\[ 2M_B \mu_\pi^2 \gamma = -i \langle B(v) | \bar{b}_\nu \sigma_{\mu\nu} (iD^\mu) (iD^\nu) b_\nu | B(v) \rangle \tag{5.5} \]

The coefficients of these two matrix elements are known to order \( \alpha_s \) [321–325]. At order \( 1/m_b^3 \) there are again only two matrix elements which are given by

\[ 2M_H \rho^3_D = -\langle B(v) | \bar{b}_\nu (iD) \nu (iD^\mu) b_\nu H(v) \rangle \tag{5.6} \]

\[ 2M_H \rho^3_{LS} = -i \langle B(v) | \bar{b}_\nu \sigma_{\mu\nu} (iD^\mu) (iD^\nu) b_\nu | B(v) \rangle \tag{5.7} \]

For these matrix elements only the tree level coefficients are known. Furthermore, if the matrix elements are defined as above,\(^5\) the coefficient of \( \rho^3_{LS} \) vanishes for the total rate, which is related to reparametrization invariance of the HQE [326].

The HQE predictions of the inclusive semileptonic rates depend on \( m_b \) and \( m_c \), and the size of the perturbative QCD corrections depends on the choice of the quark-mass scheme. The quark masses are discussed in detail in a different section of this paper, and we refer the reader to this section.

5.1.2 Higher power corrections

At order \( 1/m^4_b \) and higher the number of independent non-perturbative parameters starts to proliferate. In addition, due to the dependence on powers of \( 1/m_c \) the power counting needs to be re-defined: since we have parametrically \( m_c^2 \sim \Lambda_{QCD} m_b \) one has to count the term \( d \Gamma_3 \delta_2 \) as a part of \( d \Gamma_4 \), see (5.2). Thus the full complexity of the dim-8 operators already enters an analysis of the \( 1/m_b^5 \) contribution.

We shall not list the independent matrix elements appearing at order \( 1/m_b^4 \) and \( 1/m_b^5 \), rather we refer the reader to the list given in Refs. [327,328]. However, the proper counting of the number of independent operators has been settled only recently [329], using the method of Hilbert series. It turns out that at tree level there are 9 dimension 7 operators [327] while QCD corrections will increase this number to 11 [329].

The reason is very simple. At order \( 1/m^4_b \) we have operators with four covariant derivatives, which can be written as \( \langle E^2 \rangle \) (chromoelectric field squared) and \( \langle B^2 \rangle \) (chromomagnetic field squared) where \( E \) and \( B \) are both color-octets. Thus the combination appearing at tree level is

\[ E^2 = E^a \cdot E^b T^a T^b \quad \text{and likewise for } B^2. \tag{5.8} \]

However, the symmetric product of \( T^a \) and \( T^b \) contains a singlet and an octet component

\[ \frac{1}{2} (T^a T^b + T^b T^a) = \delta^{ab} + d^{abc} T^c. \tag{5.9} \]

The two terms on the right-hand side acquire different coefficients once QCD corrections are taken into account, and thus become independent operators. Although this observation [329] is correct, it has no impact unless QCD corrections are considered at order \( 1/m^5_b \). The same argument explains the different counting at order \( 1/m^3_b \) where we have 18 parameters at tree level [327], while the general case involves 25 matrix elements [329].

Clearly the number of independent parameters appearing at order \( 1/m^4_b \) is too large to extract them from experiment, even if data will become very precise in the future. To this end, one has to rely on some additional theoretical input, which should better be model dependent. A systematic approach has been proposed in Ref. [327] and refined in Ref. [328]: it is based on the “lowest-lying state saturation Ansatz” (LLSA) and corresponds to a naive factorization of the matrix elements. The LLSA allows us to write all matrix elements appearing in \( 1/m^4_b \) and \( 1/m^5_b \) in terms of four parameters, which are \( \mu^2_D \) and \( \mu^2_{LS} \) (see Eqs. (5.4) and (5.5)) and \( \epsilon_1/2 \) and \( \epsilon_3/2 \), where \( \epsilon \) are the excitation energies of the lowest orbitally excited spin symmetry doublets with \( j \) the spin of the light degrees of freedom. Note that in this setup also \( \rho_D \) and \( \rho_{LS} \) can be computed which may serve as a check, since these parameters can also be extracted from experiment.

The LLSA has been used to study the impact of the \( 1/m^4_b \) terms on the extraction of \( V_{cb} \) in Ref. [330]. It turns out that, even if a generous margin is allowed for the uncertainties, the...
shift in the extracted $V_{cb}$ remains well below 1%, and with the default choices of Ref. [330] a shift of $-0.25\%$ is found.

Recently the impact of the reparametrization invariance on the HQE has been re-investigated. In Refs. [326,331] it has been shown that the number of independent parameters in higher orders can be reduced by reparametrization invariance, for the total rate and the $q^2$ moments. While the number of HQE parameters up to order $1/m_b^2$ is still two, there is only one parameter at $1/m_b^3$, since the spin-orbit term can be absorbed into $\mu_\tau^2$. At order $1/m_b^4$ there will be only four parameters, which opens up the possibility of constraining the higher dimensional matrix elements directly with experimental data, at least if Belle II will be able to measure several moments of the $q^2$ distribution.

### 5.1.3 Heavy quark expansion for $B \to X_c \tau \bar{\nu}$

The recent data on the exclusive decays $B \to D^{(*)} \tau \bar{\nu}$ indicate that the branching ratios of these channels lie above the prediction of the SM. This issue is discussed in detail in Sect. 2.4, but we may also consider the inclusive decay $B \to X_c \tau \bar{\nu}$ for which the HQE provides us with a precise prediction.

While a new measurement of $B \to X_c \tau \bar{\nu}$ has to wait until Belle II has collected a sufficient data sample, we may compare with a measurement performed at LEP resulting in [99]

$$\text{Br}(b\text{-admix} \to X \tau \bar{\nu}) = (2.41 \pm 0.23)\%$$

where $b$-admix refers to the $b$-hadron admixture produced in a $Z$ decay. Since to leading order the inclusive semitauonic branching fraction of all $b$-hadrons are the same, we may take this as an estimate of $B \to X_c \tau \bar{\nu}$.

This has to be compared with the measured sum of $B \to D \tau \bar{\nu}$ and $B \to D^{(*)} \tau \bar{\nu}$

$$\text{Br}(B \to [D + D^{(*)}] \tau \bar{\nu}) = (2.68 \pm 0.16)\%,$$

indicating that the two ground states tend to oversaturate the inclusive decay.

The decay $B \to X_c \tau \bar{\nu}$ has been studied in the HQE [332] up to $1/m_b^2$ and $a_\tau^2$ in the $1S$ scheme, resulting in

$$\text{Br}(B^- \to X_c \tau \bar{\nu}) = (2.42 \pm 0.05)\%.$$

More recently, sizable effects of order $1/m_b^3$ have been found [333], which using the kinetic scheme, but without $O(a_\tau^2)$ contributions, found

$$\text{Br}(B^- \to X_c \tau \bar{\nu}) = (2.26 \pm 0.05)\%.$$

The additional inclusion of $O(a_\tau^2)$ effects in the kinetic scheme appears to lead to a very similar value [334]. These HQE calculations are compatible with the LEP measurement.

However, the LEP measurement is not very precise and thus leaves room for new physics contributions. In the context of $R(D^{(*)})$ many new physics scenarios have been discussed, and we will not repeat any of this here. Instead we use a very simple ansatz to explore qualitatively the effect of new physics. To this end, we add an additional interaction of the form

$$H_{\text{NP}} = \frac{G_F V_{cb}}{\sqrt{2}} (\alpha \, O_{V+A} + \beta \, O_{S-P})$$

with

$$O_{V+A} = (\bar{c} \gamma_\mu (1 + \gamma_5) b) \, (\bar{\tau} \gamma^\mu (1 - \gamma_5) \nu),$$

$$O_{S-P} = (\bar{c} (1 - \gamma_5) b) \, (\bar{\tau} (1 - \gamma_5) \nu).$$

We may fit the two parameters $\alpha$ and $\beta$ to the data on $B \to D^{(*)} \tau \bar{\nu}$ and find $\alpha = -0.15 \pm 0.04$ and $\beta = 0.35 \pm 0.08$ [333]. This may be inserted back into the calculation of the total rate for $B \to X_c \tau \bar{\nu}$ for which we find

$$\text{Br}(B^- \to X_c \tau \bar{\nu}) = (3.15 \pm 0.19)\%$$

indicating a significant shift of the inclusive rate. This result is graphically presented in Fig. 6 and indicates that generically the exclusive and inclusive data are in tension, unless the new physics is such that it almost cancels in the inclusive rate.

### 5.2 Inclusive processes in lattice QCD

Until recently, the application of lattice QCD has been limited to the calculation of form factors of exclusive processes such as $B \to D^{(*)} \ell \nu$ or $B \to \pi \ell \nu$, for which initial and final states contain a single hadron. A first proposal to evaluate the structure functions relevant to the inclusive decays $B \to X_{u,c} \ell \nu$ in lattice QCD was put forward in [335]. As

![Fig. 6](https://example.com/fig6.png)

**Fig. 6** Fit of the data to the parameters $\alpha$ and $\beta$. The green ellipse represents the fit result to the exclusive channels, the green band represents the LEP measurement, the red band the SM result obtained from the HQE.

---

\( \odot \) Springer
mentioned above, the differential decay rate for the inclusive decay \( B(p_B) \to X_c(p_X) \ell(\nu)\bar{\nu}(p_\nu) \) may be written in terms of the structure functions of \( W_{\mu\nu}(p_B, q) \), which contains the sum over all possible final states:

\[
W_{\mu\nu}(p_B, q) = \sum_x (2\pi)^3 \delta^4(p_B - q - p_X) \frac{1}{2M_B} \times \langle B(p_B)|J_\mu^+(x)|X(p_X)\rangle \langle X(p_X)|J_\nu|B(p_B)\rangle,
\]

where \( J_\mu \) stands for the \( b \to c \) weak current and \( q^\mu = (p_\ell + p_\nu)^\mu \) is the momentum transfer. The optical theorem relates this to the forward scattering matrix element \( T_{\mu\nu}(p_B, q) \),

\[
T_{\mu\nu}(p_B, q) = \int d^4x \ e^{-iqx} \frac{1}{2M_B} \langle B(p_B)|T\{J_\mu^+(x)J_\nu(0)\}|B(p_B)\rangle,
\]

as \(-1/\pi\text{Im}T_{\mu\nu} = W_{\mu\nu}\), see for instance [336,337].

One can calculate these forward matrix elements on the lattice as long as the momenta \( p_B \) and \( q \) are in the region where no singularity develops. It means that the lattice calculation is possible in an unphysical kinematical region where no real decay is allowed. This kinematical region corresponds to the situation where the energy given to the final charm system \( p_X^0 \) is too small to create real states such as the \( D \) and \( D^* \) mesons or the \( D^0 \) continuum states. The connection to the physical region can be established by using Cauchy’s integral on the complex plane of \( p_X^0 \). An alternative method is to reconstruct the spectral density (of the states \( X \) appearing in the sum) directly from the lattice correlation function [338].

An explorative lattice calculation has been performed at relatively light \( b \) quark masses [335]. The numerical results suggest that the matrix element is nearly saturated by the ground state \( D^{(*)} \) meson contribution at the zero-recoil limit.

Since the non-perturbative lattice calculation may be obtained at the kinematical point away from the resonance region, it may also be used to validate the heavy quark expansion (HQE) method. So far, the HQE calculation is available in the unphysical region only at the tree-level, \( O(a_s^0) \). The one-loop and two-loop corrections have been calculated for the differential decay rate. They have to be transformed to the unphysical kinematical point by applying the Cauchy integral. Such work is in progress.

As already mentioned, the lattice calculation can be made only in the unphysical kinematical region and its comparison with the experimentally observed \( B \) decay distribution is not straightforward. One should first perform an integral of the experimental data with an appropriate weight to reproduce Cauchy’s integral in the complex plane of \( p_X^0 \), which requires the experimental data obtained as a function of two kinematical variables \( q^2 \) and \( p_B \cdot q \). It still does not cover the whole complex plane, and one need to supplement by a perturbative QCD calculation for the region of \( p_X^0 > p_B^0 \). The perturbative expansion in this unphysical region should be well-behaved, but the details should be investigated further.

More recently, a different approach that in principle allows to calculate the total decay rate has been proposed [339]. In the new method, the integral corresponding to the phase space of the \( B \to X_c \ell\nu \) is directly performed rather than the Cauchy’s integral. As a result, information about the unphysical kinematical region is no longer necessary. A first comparison with the HQE with a small \( m_b \sim 2.7 \) GeV shows good agreement with the lattice calculation, despite large uncertainties. This method may open an opportunity to compute the inclusive decay rate fully non-perturbatively using lattice QCD, and can also be applied to calculate various moments of the \( B \to X_c \ell\nu \) decays, as well as the more challenging \( B \to X_u \ell\nu \) decays.

### 5.3 HQE matrix elements from lattice QCD

The same hadronic parameters appearing in the OPE analysis of inclusive semileptonic \( B \)-meson decays appear also in the HQE of the pseudoscalar (PS) and vector (V) heavy-light meson masses. Therefore, one can try to determine them from a lattice calculation of the latter at different values of the heavy quark mass. After the pioneering work of Ref. [287], new unquenched results have been presented recently [256, 257]. These papers are mentioned in Sect. 4.1 for their results on quark masses.

In Ref. [257] a precise lattice computation of PS and V heavy-light meson masses has been performed for heavy-quark masses ranging from the physical charm mass up to \( \approx 4 \) times the physical \( b \)-quark mass, adopting the gauge configurations generated by the European Twisted Mass Collaboration (ETMC) with \( N_f = 2 + 1 + 1 \) dynamical quarks at three values of the lattice spacing \((a \approx 0.062, 0.082, 0.089 \text{ fm})\) with pion masses in the range \( M_\pi \approx 210-450 \text{ MeV} \). The heavy-quark mass is simulated directly on the lattice up to \( \approx 3 \) times the physical charm mass. The interpolation to the physical \( b \)-quark mass is obtained with the ETMC \textit{ratio method} [26, 27], based on ratios of the spin-averaged meson masses computed at nearby heavy-quark masses, and the kinetic scheme is adopted. The extrapolation to the physical pion mass and to the continuum limit yields \( m_b^{\text{kin}}(1 \text{ GeV}) = 4.61(20) \text{ GeV} \), corresponding to \( \text{m}_b(\text{m}_b) = 4.26(18) \text{ GeV} \) in the \( \overline{\text{MS}} \) scheme, in agreement with other \( m_b \) determinations; see Sect. 4.1. The ratio method is applied above the physical \( b \)-quark mass to provide heavy-light meson masses towards the static point. The lattice data are analyzed in terms of the HQE and the matrix elements of dimension-4 and dimension-5 operators are determined with good precision, namely:

\[
\overline{\Lambda} = 0.552 (26) \text{ GeV},
\]
The size of two combinations of the matrix elements of dimension-6 operators is also determined:

\[
\rho_D^3 - \rho_{\pi\pi}^3 - \rho_S^3 = 0.153 \text{ (34) GeV}^3,
\]

(5.17)

\[
\rho_{\pi\pi}^3 + \rho_A^3 - \rho_{LS}^3 = -0.158 \text{ (84) GeV}^3,
\]

(5.18)

with the full covariance matrix provided in Ref. [257]. Although all the above results refer to the asymptotic limit, namely to infinitely heavy quarks, and differ from the matrix elements extracted in the inclusive fits described above by higher power corrections, they are found to be mutually consistent. In the future lattice results could be used as additional constraints in the semileptonic fits. Another interesting future application concerns the heavy-quark sum rules for the form factor entering the semileptonic decay \( B \to D^* \ell \nu \) at zero-recoil; here the non-local correlators \( \rho_{A,S,\pi\pi,\pi G} \) play an important role; see Ref. [340].

The analysis by the Fermilab, MILC and TUMQCD Collaborations [256], based on [300], employs only PS mesons and the minimal renormalon subtracted (MRS) heavy quark mass. The results are obtained using MILC ensembles with five values of lattice spacing ranging from approximately 0.12–0.03 fm, enabling good control over the continuum extrapolation, and both physical and unphysical values of the two light and the strange sea-quark masses. This leads to

\[
\Lambda_{\text{MRS}} = 0.555 \text{ (31) GeV}
\]

(5.19)

while power corrections are controlled by the difference \( \mu_{\pi}^2 - \mu_G^2(m_H) \). Assuming \( \mu_G^2(m_H) = 0.35(7) \text{ GeV}^2 \) as a prior, the authors find \( \mu_{\pi}^2 = 0.05(21) \text{ GeV}^2 \). Notice that the definition of \( \mu_{\pi}^2 \) used here still has a renormalon ambiguity of order \( \Lambda_{\text{QCD}}^2 \).

5.4 Experimental status

### 5.4.1 Measurements of inclusive observables in \( B \to X_c \ell \nu \)

Several experiments have measured the partial branching fraction of the inclusive decay \( B \to X_c \ell \nu \) (\( \ell = e, \mu \)) as a function of the lower threshold on the lepton momentum \( (E_{\ell}) \), or other inclusive observables in this decay such as the moments of the lepton energy and of the \( X_c \) mass distribution. Available measurements are listed in Table 5, where it should be noted that the most recent experimental result is from the year 2010.

The Belle collaboration has measured spectra of the lepton energy \( E_\ell \) and the hadronic mass \( M(X_c) \) in \( B \to X_c \ell \nu \) using 152 million \( \Upsilon(4S) \to BB \) events [343,344]. These analyses proceed as follows: first, the decay of one \( B \) meson in the event is fully reconstructed in a hadronic mode \( (B_{\text{tag}}) \). Next, the semileptonic decay of the second \( B \) meson in the event \( (B_{\text{sig}}) \) is identified by searching for a charged lepton amongst the remaining particles in the event. In Ref. [343], the electron momentum spectrum in the \( B \) meson rest frame is measured down to 0.4 GeV. In [344], all remaining particles in the event, excluding the charged lepton (electron or muon), are combined to reconstruct the hadronic \( X \) system. The \( M(X_c) \) spectrum is measured for different lepton energy thresholds in the \( B \) meson rest frame. The observed spectra are distorted by resolution and acceptance effects and cannot be used directly to obtain the moments. In the Belle analyses, acceptance and finite resolution effects are corrected by unfolding the observed spectra using the singular value decomposition (SVD) algorithm [348].

Table 5 List of available measurements of inclusive moments in \( B \to X_c \ell \nu \)

| Experiment | Lepton spectrum \( (E_\ell^n) \) | Hadron spectrum \( (M_{2n}^{X_c}) \) |
|------------|-------------------------------|----------------------------------|
| BaBar      | \( n = 0, 1, 2, 3 \) [341,342] | \( n = 1, 2, 3 \) [341]          |
| Belle      | \( n = 0, 1, 2, 3 \) [343]    | \( n = 1, 2 \) [344]             |
| CDF        | \( n = 1, 2 \) [345]          |                                  |
| CLEO       | \( n = 1, 2 \) [346]          |                                  |
| DELPHI     | \( n = 1, 2, 3 \) [347]       | \( n = 1, 2 \) [347]             |

The resolution in \( M(X_c) \) is improved by a kinematic fit to the whole event, taking into account 4-momentum conservation and constraining the missing mass to zero. To derive the true moments from the reconstructed ones, BaBar applies
a set of linear corrections. These corrections depend on the charged particle multiplicity of the $X$ system, the normalized missing mass, $E_{\text{miss}} - p_{\text{miss}}$, and the lepton momentum. In this way, BaBar measures the moments of the hadronic mass spectrum up to $(M_X^2)$ for minimum lepton energies ranging from 0.8 to 1.9 GeV.

5.4.2 Determination of $|V_{cb}|$ from inclusive decays

The heavy flavor averaging group (HFLAV) has used the measurements discussed in the previous section to determine $|V_{cb}|$ from a fit to HQEs of inclusive observables [68]. Using expressions in the so-called kinetic scheme [320,323,349–351] and a precise determination of the $c$-quark mass, $m_c^{\text{MS}}(3 \text{ GeV}) = 0.986 \pm 0.013 \text{ GeV}$ [277], as external input, HFLAV obtains

$$|V_{cb}| = (42.19 \pm 0.78) \times 10^{-3},$$  

$$m_b^{\text{kin}} = 4.554 \pm 0.018 \text{ GeV},$$  

$$\mu_\pi^2 = 0.464 \pm 0.076 \text{ GeV}^2.$$  

The $\chi^2$ of the fit is 15.6 for 43 degrees of freedom. Using expressions in the so-called 1S scheme [352,353] the same set of measurements results in

$$|V_{cb}| = (41.98 \pm 0.45) \times 10^{-3},$$  

$$m_b^{1S} = 4.691 \pm 0.037 \text{ GeV},$$  

$$\lambda_1 = -0.362 \pm 0.067 \text{ GeV}^2,$$

with a $\chi^2$ of the fit of 23.0 for 59 degrees of freedom. This analysis uses measurements of the photon energy moments in $B \to X_s \gamma$ [354–357] to constrain the $b$-quark mass and does not include higher order corrections of $O(\alpha_s^2)$ and $O(\alpha_s/m_b^2)$.

As mentioned above, the semileptonic moments have been analysed also including higher order power corrections estimated using the LSSA [330]. In this case a kinetic scheme fit to the experimental data that additionally includes a constraint $m_b^{\text{kin}} = 4.550(42) \text{ GeV}$ from PDG (after scheme con-

version) leads to a slightly more precise value,

$$|V_{cb}| = (42.00 \pm 0.64) \times 10^{-3}. \tag{5.26}$$

6 Heavy-to-light inclusive

6.1 Introduction and theoretical background

Inclusive semileptonic heavy to light decays can in principle be analyzed similarly to $B \to X_c \ell \nu$ by using a local OPE. In practice, due to the large charm background, experimental cuts are generally imposed and reduce the “inclusivity” of the theoretical prediction. In particular, the local OPE does not converge well when the invariant mass of the hadronic system is $M_X \lesssim M_D$. In such a case the decay spectra are described using a “non-local” OPE [358–360], where perturbative coefficients are convoluted with non-perturbative “shape functions” (SFs), the $B$ meson analogs of parton distribution functions. In this SF region, the perturbative coefficients themselves can be factorized into “hard” and “jet” pieces, where the former has a typical scale of $m_b$ and the latter has a typical scale of $\sqrt{m_b\Lambda_{\text{QCD}}}$. In the infinite mass limit $m_b \to \infty$ there is a single non-perturbative SF. Power corrections start at $1/m_b$ and include multiple “subleading” SFs [361–365].

One can classify the terms based on their suppression by $1/m_b$ and $\alpha_s$. The perturbative components of the leading power term are known at $O(\alpha_s^2)$ [366–371]. The $1/m_b$ power corrections include terms convoluted with the leading power SF whose perturbative parts are known at $O(\alpha_s)$ [372] and terms convoluted with subleading SFs whose perturbative parts are known at $O(\alpha_s^0)$ [361–363]. At this order one can still use subleading functions of one light-cone variable. The inclusion of $O(\alpha_s)$ contributions of subleading SFs requires functions of multiple light-cone momenta in analogy to higher twist effects in Deep Inelastic Scattering [373]. Schematically, in the SF region we have the factorization for-
mula
\[ d\Gamma \sim H \cdot J \otimes S + \frac{1}{m_b} \sum_i h \cdot J_0 \otimes s_i \]
\[ + \frac{1}{m_b} \sum_k h \cdot j_k \otimes S + O \left( \frac{1}{m_b^2} \right), \]

(6.1)

where \( H \) is the leading power hard function, \( J \) is the leading power jet function, both known at \( O(\alpha_s^2) \), \( J_0 \) is the \( O(\alpha_s) \) part of \( J \), \( h = 1 + O(\alpha_s) \), \( s_i \) are given in Refs. [361–363], and \( j_k \) in Ref. [372]. The symbol \( \otimes \) denotes an integral over the light-cone momentum.

The moments of the leading and subleading SFs are related to the HQE parameters measured in the inclusive semileptonic decays to charm. The relations are known for the leading SF up to at least the fifth moment [374], although the current large uncertainty of higher HQE parameters [327,330] might limit the use of higher moments relations. The formalism in Ref. [374] allows to construct such relations for the subleading SFs too, but at present only the first three moments are known [364,375]. A detailed knowledge of the SFs is necessary only in a portion of the phase space where \( p_+ = E_X - p_X \sim \Lambda_{\text{QCD}} \), elsewhere only the first few moments of the SFs are relevant and one recovers the local OPE description.

The present \(|V_{ub}|\) determination by HFLAV [68] is based on various approaches which are all rooted in (6.1) and differ in the inclusion and treatment of perturbative and nonperturbative contributions, see Ref. [376] for a detailed discussion.

The approach known as BLNP (Bosch–Lange–Neubert–Paz) [377] aimed at a precision extraction of \(|V_{ub}|\) from \( B \to X_s \ell \nu \) and \( B \to X_s \gamma \), based on the knowledge in 2005. It used the first two terms in (6.1), in particular the \( O(\alpha_s) \) expression for \( H \cdot J \otimes S \) and the \( O(\alpha_s^0) \) expression for the \( h \cdot J_0 \otimes s_i \) terms. Kinematic corrections that scale as \( \alpha_s/m_b \) and \( \alpha_s/m_b^2 \) [378], as well as \( 1/m_b^2 \) corrections [336,337], for which factorization formulas were not known, were also included by convolution with the leading power shape function. Using Renormalisation Group methods \( H \) is evolved from “hard” to the “jet” scale to resum Sudakov double logs. As for the non-perturbative inputs, the leading order SF was to be taken from \( B \to X_s \gamma \) and subleading SFs \( s_i \) to be modeled using \( \sim 700 \) models. In practice, the current treatment of \( S \) by experiments is to use an exponential or Gaussian model constrained by the first two moments of \( S \) obtained from the global fit of HQE parameters in the kinetic scheme [68].

Since Ref. [377] appeared, there have been many theoretical advances. Two-loop calculations of \( H \) [367–370] and \( J \) [379] as well as one-loop calculation of \( j_k \) [372] became available. The free quark differential decay rate were calculated at \( O(\alpha_s^2) \) [366,380–382] and at complete \( O(\alpha_s^2) \) [371]. Running effects from the “hard” to the “jet” at \( O(\alpha_s^2) \) were studied [383]. It was found there that the factorization of the perturbative coefficient into jet and hard functions is not strictly necessary. More recently, three loop calculations of \( J \) [384] and the partonic \( S \) [385] were performed. Implementing these within the BLNP framework would probably require also the calculation of \( H \) at three-loops, which is not available yet. There were also theoretical advances in the description of non-perturbative effects in \( B \to X_s \gamma \) [386–388]. In particular, new subleading shape functions unique to \( B \to X_s \gamma \) were identified [387], making it more difficult to use data from radiative \( B \) decays as input for the extraction of \(|V_{ub}|\). These new features are not yet implemented in the BLNP approach. An alternative implementation of the same conceptual framework has been presented in Ref. [389], together with a systematic procedure to account for the uncertainties in the modelling of the leading SF, to be discussed below.

The GGOU (Gambino–Giordano–Ossola–Uraltsev) approach [390] avoids the expansion in \( 1/m_b \) and the introduction of subleading SFs. The perturbative coefficients are computed at fixed order to \( O(\alpha_s^2 \beta_0) \) in the kinetic scheme. The effect of RGE evolution in the SF region and all subleading SFs are absorbed into three \( q^2 \)-dependent SF \( F_i(k,q^2) \), whose first moments are fixed by present semileptonic fits. The uncertainty due to the functional form is estimated comparing \( \sim 100 \) models.

The emergence of the SF can also be seen in perturbation theory: soft-gluon resummation together with an infrared prescription gives rise to a hard quark SF. In the DGE (Dressed-Gluon Exponentiation) approach [391,392] this is achieved by an internal resummation of running coupling corrections in the Sudakov exponent, thus providing a perturbative model for the leading SF. A somewhat similar line of action is followed in Ref. [393] where the infrared prescription is provided by the so-called analytic QCD coupling.

The so-called Weak Annihilation (WA) contributions are a source of theoretical uncertainty common to all approaches. In the local OPE they emerge at \( O(1/m_b^3) \) but are enhanced by a large Wilson coefficient [394] and may give rise to a difference between \( B^+ \) and \( B^0 \) decays. As they are expected to be much more important in charm decays, the latter constrain them most effectively at present. In particular, the \( D^0, D^+ \) and \( D_s \) total semileptonic rates and the electron spectra measured by the CLEO Collaboration [395] have been employed [396–398]. From the absence of clear indications for WA effects in semileptonic charm decays, one can conclude that the WA correction to the total rate of \( B \to X_s \ell \nu \) must be smaller than about 2% [398]. However, WA is localized in the high \( q^2 \) region and therefore the related uncertainty on \(|V_{ub}|\) depends on the kinematical cuts, and this is taken into account in the current HFLAV averages. Because the high \( q^2 \) tail is particularly sensitive to higher power corrections (and not to the SFs), see for instance Refs. [381,390,399].
one might eventually expect the cleanest determinations of $|V_{ub}|$ to come from the low $q^2$ region only. An upper cut on $q^2$ might therefore be beneficial [377,390].

A few recent experimental analyses [400,401] have relaxed the kinematic cuts, making use of experimental information to subtract the background. As a result, most of the $B \rightarrow X_u \ell \nu$ phase space is taken into account and the sensitivity to the SFs is substantially reduced, while a description based on the local OPE sets in. In these cases the quoted theoretical uncertainties are smaller, but one should keep in mind that these analyses still depend on the SFs treatment and modelling for the determination of the reconstruction efficiencies, whose uncertainty contribute to the final experimental systematic error. As will be discussed later on, a realistic signal simulation requires the implementation of so-called hybrid models that transform the inclusive predictions of the approaches mentioned above into individual final hadronic states. The uncertainties related to such hybrid models remain a major issue for the inclusive determination of $|V_{ub}|$.

6.2 Status of the experimental results

The most difficult task of the inclusive measurements is the discrimination between the $B \rightarrow X_u \ell \nu$ signal and the much more abundant decays involving Cabibbo-favoured $B \rightarrow X_c \ell \nu$ decays. The signal events are studied in restricted regions of the phase space to improve the signal-to-background ratio. Compared to $B \rightarrow X_c \ell \nu$ events, the signal tends to have higher lepton momenta $p_\ell$, lower invariant mass of the $X_u$ state $M_X$, higher $q^2$, and smaller values of the light-cone momentum $P_+ = E_X - |p_X|$, where $E_X$ and $p_X$ are energy and momentum of the hadronic system $X_u$ in the $B$ meson rest frame. As explained above, these restrictions introduce difficulties in the calculation of the expected partial branching fraction, enhancing perturbative and non-perturbative QCD corrections which lead to large theoretical uncertainties in the measurement of $|V_{ub}|$.

The measurement of the partial branching fraction $\Delta B$ can be obtained with tagged or untagged analyses.

6.2.1 Tagged analyses

In tagged analyses, the $\Upsilon(4S) \rightarrow B \overline{B}$ events are identified by reconstructing one of the $B$ mesons, $B_{\text{reco}}$, via fully hadronic decays. The signal decay of the second $B$ meson ($B_{\text{signal}}$) is identified just by the presence of an electron or a muon. The tracks and neutral objects not associated with the $B_{\text{reco}}$ can be uniquely assigned to the signal side, so that the inclusive $X_u$ state can be clearly reconstructed. The neutrino four-momentum $p_\nu$ can be estimated from the missing momentum $p_{\text{miss}} = p_{e^+e^-} - p_{B_{\text{reco}}} - p_{X_u} - p_\nu$, where $p_{e^+e^-}$ is the initial state four-momentum. From this, all the kinematic variables of the signal state can be easily computed.

Because the momentum of the signal $B$ meson is determined from of the $B_{\text{reco}}$, the signal decay products can be computed directly in the $B$-meson rest frame, resulting in an improved resolution of the accessible observables. Moreover, the constrained kinematics allow for a better separation of the signal from the background.

The downside of the tagged analysis is the low signal efficiency (about 0.3–0.5%) which implies that for kinematic variables like the lepton momentum $p_\ell$, the untagged analyses at the $B$-factories can give competitive or better results. Undetected and poorly reconstructed tracks or photons lead to irreducible background from the dominant $B \rightarrow X_c \ell$ decays even in regions of the phase space potentially free of such background, and this can affect the final resolution on the signal kinematics.

The hadronic $B$-tagging approach was used for the first time by BaBar to extract the $B \rightarrow X_u \ell \nu$ signal in the phase space region $M_X < 1.55$ GeV, with the further requirement that $p_\ell > 1$ GeV [402]. Using the same sample, BaBar removed the constraint on $M_X$ and obtained the $B \rightarrow X_u \ell \nu$ partial branching ratio requiring only $p_\ell > 1$ GeV, which covers about 90% of the phase space [403]. This challenging analysis was affected by the large statistical uncertainties and limited by the knowledge on the $B \rightarrow X_c$ background components and the signal composition available when it was published.

Exploiting the full datasets collected, both Belle [400] and BaBar [401] published measurements of $B \rightarrow X_u \ell \nu$ partial branching fraction, performing a fit in $M_X$ and $q^2$, and requiring only $p_\ell > 1$ GeV. BaBar determined also the partial branching fractions in other several restricted regions of the phase space.

6.2.2 U ntagged analyses

The untagged measurements allow to collect large samples but are affected by considerable backgrounds. The untagged measurements have access only to a few kinematic variables, namely the lepton momentum $p_\ell$, and the $q^2$ spectra,

- lepton spectrum: this can be studied inclusively without requirements on the rest of the event. In this case the momentum spectrum can only be given in the $\Upsilon(4S)$ rest frame.
- $q^2$ distribution: this requires the reconstruction of the neutrino 4-momentum, which exploits the high hermeticity of the $B$ factories’ detectors. The neutrino 4-momentum is given by the event missing 4-momentum, $p_{\text{miss}} = p_{e^+e^-} - p_{\text{vis}}$, where $p_{e^+e^-}$ is the initial state 4-momentum, and $p_{\text{vis}}$ is the total visible 4-momentum determined by all the charged tracks from the collision point, identified pairs of charged tracks from $K_s$, $\Lambda$ and
\[ \gamma \rightarrow e^+e^-, \text{ and energy deposits in the electromagnetic calorimeter.} \]

The lepton momentum spectrum is affected by large backgrounds from \( B \rightarrow X_c \ell \nu \) via the \( D\ell\nu, D^*\ell\nu, D^{**}\ell\nu \) (where \( D^{**} \) is a mixture of charm excited state and non resonant \( D^{(*)} - n\pi \text{ transitions} \), \( D_s K \ell\nu X \) and also secondary leptons from \( D \) mesons decays, and a background from \( e^+e^- \rightarrow q\bar{q} \) events, where the main contribution comes from \( c\bar{c} \), which is assessed from control data samples recorded below the \( \Upsilon(4S) \) resonance. Because of the large background, usually the signal is extracted only for regions with high momentum lepton, typically \( p_\ell > 1.9-2.1 \text{ GeV} \). Old analyses of the lepton endpoints are from CLEO \[404\], Belle \[405\] and BaBar \[406\].

Recently, BaBar published a study \[407\] of the lepton spectrum using the full data set, and exploiting all the knowledge about the rate and the form factors of the various \( B \rightarrow X_c \ell \nu \) exclusive decays which are the major source of backgrounds. The signal is extracted from a fit to the electron momentum spectrum, which is described as the sum of predicted signal (model dependent shape) and various specific backgrounds yields with shapes fixed by MC. The fit covers lepton momentum in the \( \Upsilon(4S) \) rest frame from 0.8 to 2.7 GeV, in 50 MeV bins, except for the data in the interval 2.1–2.7 GeV which are combined in a single bin to avoid effects from differences in the shape of the theoretically predicted signal spectrum. In a given momentum interval, the excess of events above the sum of the fitted background contributions is taken as the number of signal events.

An important difference of this analysis with respect to the other ones is that different theoretical models are considered in the extraction of the partial branching fractions. Instead, all other measurements determine the partial branching fraction by using a single model, and its partial rate is then converted in a measurement of \(|V_{ub}|\) by taking the corresponding partial rate predicted by the theory calculations.

The extracted inclusive signal branching fractions and the values of \(|V_{ub}|\) agree well for GGOU and DGE, although they are about 13% smaller than the average of the other measurements. This difference can be attributed to the shape of the predicted signal spectrum and/or the shapes of some of the large background contributions above 2 GeV where the signal fraction is largest. On the other hand, the value of \(|V_{ub}|\) based on BLNP agrees well with other measurements.

A subset of all the measurements of the inclusive \(|V_{ub}|\) are reported in Fig. 8 for the various frameworks considered, see \[49\] for more details.

6.2.3 Lessons learned from the past

The measurements based on tagged samples have considerably larger statistical uncertainties. The sample size allows for only a few bins in the 2D fit, but there are regions of the phase space (e.g. low \( M_X \)) where the background fractions are modest. The current sensitivity to the details of the shapes of the signal and background distributions is however limited.

For untagged measurements only the high end of the spectrum is sensitive to the signal and also to the background near their kinematic endpoints. Both approaches have their pros and cons, given the size of the currently available data. The latest BaBar measurement of the lepton spectrum, shows a high dependence of the result from the signal model. The same effect, even if not directly evident, was observed also in tagged measurements from the sensitivity of the signal yield extraction on the shape function parameters in the analyses that cover larger portion of the phase space.

Semileptonic \( B \rightarrow X_u \ell\nu \) decays are simulated as a combination of resonant decays with \( X_u = \pi, \eta, \eta', \rho, \omega \) and decays to nonresonant hadronic final states \( X_u \). The latter is simulated with a continuous invariant mass spectrum following the theory predictions by De Fazio and Neubert \[378\], which depend on the SF parameters and \( m_B \). The nonresonant and the resonant part are combined such that the sum of their branching fractions is equal to the measured one for the inclusive \( B \rightarrow X_u \ell\nu \). The events generated with this model, are reweighted to obtain predictions for different SF parameters and different branching fraction of the resonant states. This model is usually called “hybrid model”. Belle in \[400\], corrects the hybrid model to match the moments of the \( M_X \) and \( q^2 \) distributions predicted by the the GGOU model. A picture of the model of the invariant mass \( M_X \) shape used to describe the \( B \rightarrow X_u \ell\nu \) is reported in Fig. 9.

Another effect not considered so far, is the impact of the fragmentation of the generated \( u \) quark into final state hadrons, which is performed using JETSET. The modeling of the final state multiplicity could affect both the signal effi-
ciency and the signal templates used to separate signal from background.

The measurement of the partial branching fraction separately for neutral and charged $B$ mesons has been used to constrain the WA contribution. Both tagged approach, in various regions of the phase space [401], and untagged approach, in the high lepton region [408], have been used, but these have given weak upper limits mainly because of the large statistical uncertainties. More stringent upper limit on WA has been obtained by CLEO which used a model dependent approach studying the high $q^2$ region in $B \to X_u \ell \nu$ decays [409]. Both these bounds are milder than those estimated from the SIMBA framework with two projected single-differential measurements for charged and neutral $B$-meson decays are required to allow for an improved extraction of $|V_{ub}|$ from these decays. Additional external constraints, such from $B \to X_s \ell \nu$, can also be employed.

In the tagged measurements the suppression of the $b \to c$ background is performed by vetoing events where a $K^+$ or a $K^0_s$ is detected in the hadronic $X$ system. This causes a loss in the signal contribution where a $s \bar{s}$ pair is produced (usually called $s \bar{s}$-popping). The fraction of these events is about 12% of the non-resonant component and it is fixed in the fragmentation parameters of JETSET/PYTHIA. The uncertainty on this fraction is assumed to be about 30%, so for analyses that aim to cover larger regions of the phase space, with higher statistics this could be an irreducible source of systematic uncertainty. This is another point that should be improved in future analyses at Belle II.

6.3 Fitting distributions: SIMBA and NNVub

As we discussed above, SFs modelling is an important source of theoretical uncertainty in the study of $B \to X_u \ell \nu$ and particularly in the extraction of $|V_{ub}|$ from these decays. While the first few moments of the SFs must satisfy OPE constraints, direct experimental information on the SFs is somewhat limited. Indeed, the measured photon spectrum in $B \to X_s \gamma$ is sensitive to a different set of subleading SFs. However differential distributions in $B \to X_u \ell \nu$ such as the lepton energy and the invariant mass distributions depend directly on all the SFs and can therefore be used to constrain them. Conversely, they can be used to validate SFs models and approaches where the SFs are calculated, such as DGE. The high luminosity expected makes the measurement of differential distributions possible at Belle II.

The extraction of $|V_{ub}|$ performed by HFLAV in the BLNP and GGOU frameworks assumes a set of two-parameter functional forms, and it is unclear to what extent the chosen set is representative of the available functional space, and whether the estimated uncertainty really reflects the limited knowledge of the SFs. This point was first emphasized in Ref. [389], where a different strategy was proposed, based on the expansion of the leading SF in a basis of orthogonal functions, whose coefficients are fitted to the $B \to X_s \gamma$ spectrum, and on the modeling of the subleading SFs. The SIMBA project [410] aims at performing a global fit to $B \to X_s \gamma$ and $B \to X_u \ell \nu$ spectra, to simultaneously determine $|V_{ub}|$, $m_b$, the leading SF, as well as the Wilson coefficient of radiative $b$ decays. Additional external constraints, such from $B \to X_s \ell \nu$, can also be employed.

Another strategy, called NNVub and explored in [411] for the GGOU approach, employs artificial neural networks as unbiased interpolants for the SFs, in a way similar to what the NNPDF Collaboration do in fitting for Parton Distribution Functions [412]. This method allows for unbiased estimates of the SFs functional form uncertainty, and for a straightforward implementation of new experimental data, including $B \to X_s \gamma$ and $B \to X_u \ell \nu$ spectra and other inputs on quark masses and OPE matrix elements. Both SIMBA and NNVub appear well posed to analyse the Belle II data in a model independent and efficient way.

6.4 Prospect for the future: Belle II outlook

The measurements of fully differential spectra on the kinematic variables, e.g. $q^2$, $M_X^2$, $p_T^X$, $E_l$, and separate measurements for charged and neutral $B$-meson decays are required to allow for an improved extraction of $|V_{ub}|$ in the long term. Therefore, the future measurements should provide these unfolded spectra independent of theoretical assumptions.

Combining both $B \to X_u \ell \nu$ and $B \to X_s \gamma$ as well as constraints on the SF moments from $B \to X_s \ell \nu$ in a global fit can simultaneously provide the inclusive $|V_{ub}|$ and the leading SF functional form with its uncertainties as they follow from the uncertainties in the included experimental measurements. Figure 10 shows the projections for a global fit in the SIMBA framework with two projected single-differential spectra of $M_X$ and $E_l$ for $B \to X_u \ell \nu$ and a $E_\gamma$ spectrum for $B \to X_s \gamma$ from 1 ab$^{-1}$ and 5 ab$^{-1}$ Belle II data set [108].
The new tagging algorithm developed for Belle II can perform better than the old neural network method used in the previous Belle publications with about 3 times higher efficiency [235]. With a larger data set, the systematic uncertainties counted for reconstruction efficiencies, fake leptons and continuum background knowledge are expected to improve for this measurement. The projections for inclusive $|V_{cb}|$ are summarized in Table 6.

7 Outlook

We have summarized our main results in Sect. 1. In this final Section, we would like to look at the prospects of our field over the next five years. What can we expect for semileptonic $b$ decays at the two main experiments? What kind of progress can we reasonably anticipate in lattice QCD and continuum calculations?

Belle II has started data taking with a complete detector in March 2019 and recorded about 10/fb in its first year of operation. By introducing the crab waist scheme at the collision point, SuperKEKB achieved the world’s highest instantaneous luminosity of $2.4 \times 10^{34}$/cm$^2$/s in June 2020 with acceptable background conditions for Belle II to take data. In total 64/fb of $\Upsilon(4S)$ data were recorded in the spring 2020 run, bringing the total to 74/fb. Data taking will resume in October 2020 with the goal to reach a total integrated luminosity of more than 100/fb before the end of the year break. Belle II plans to accumulate a data set equivalent to the Belle luminosity of about 1/ab by the end of 2021. In 2022 the experiment will enter a long shutdown to install the second pixel detector layer and replace the silicon photomultipliers in the barrel particle identification device. Data taking will resume in 2023 and by 2025 Belle II expects to have recorded a data sample exceeding 10/ab.

Given these luminosity prospects, competitive Belle II results for semileptonic $B$ decays can be expected in the years to follow. In addition, a three times more efficient hadronic tag and better low momentum tracking of the slow pion from the $D^*$ decay will further benefit semileptonic analyses in particular. This will allow to take a fresh look at the CKM matrix element magnitudes $|V_{cb}|$ and $|V_{ub}|$ and to improve measurements which are still statistically limited, such as $R(D)$ and $R(D^*)$.

The LHCb experiment has shown great capabilities with the results on $R(D^*)$, $|V_{ub}|/|V_{cb}|$ with $\Lambda_b$ decays, and $|V_{cb}|$ with $B_s$ decays. These measurements are based on the data collected in 2011 and 2012 (Run 1), corresponding at 3/fb of integrated luminosity. The data collected in 2015–2018 (Run 2) at $pp$ collision energy of $\sqrt{s} = 13$ TeV, correspond to about 6/fb of integrated luminosity. There are various ongoing analyses on the full dataset. Most of the measurements are limited by systematic uncertainties, among which the largest ones are generally due to external inputs from other experiments and to the limited available samples of Monte Carlo simulations. Nevertheless the large dataset available is going to be fully exploited.

The LHCb experiment is at present undergoing a major upgrade of the detector. The construction and commissioning should end in 2021, when LHC will resume the activity. The upgrade will allow to collect data at higher instantaneous luminosity, so about five $pp$ collisions per bunch crossing are foreseen, to be compared with about one-two $pp$ collisions in Run 1 and Run 2. To handle the higher occupancy expected in the detector, besides the improvements in the various subdetectors, a full software L0 trigger will be employed. The software L0 trigger will add flexibility to the data taking, allowing to reduce the thresholds for muon and hadron trigger decisions, enlarging in this way the physics capabilities. The analyses of semileptonic decays with taus and electrons will benefit from the lower trigger thresholds in terms of signal efficiencies. With this upgraded detector, LHCb is planning to integrate a luminosity of 23/fb by 2024, and to collect a total sample of 50/fb by 2028–2029, after LHC will have switched to higher luminosity.

By now, lattice QCD is the tool of choice for the form factors describing semileptonic decays of $b$-hadrons. At present, the most urgent need is the $q^2$ (or, equivalently, $w$) dependence of the form factors of $B \rightarrow D^*l\nu$, both to see how the form-factor slopes affect the $|V_{cb}|$ determination and to solidify the SM prediction of $R(D^*)$. A few such calculations are underway. Given the success of LHCb with $\Lambda_b$ semileptonic decays, updates of the baryon form factors are desirable, and we encourage other lattice-QCD practitioners to turn their attention to these decays. Another topic for future research
are rigorous calculations with a $\rho$ or $\phi$ vector meson in the final state.

The leptonic decay constants are now at the subpercent level of uncertainty, and the prospects for extending these methods to semileptonic form factors are underway. In general, near-term lattice-QCD calculations of this precision will be based on the MILC collaboration’s HISQ ensembles, which, among all lattice data sets, span the largest range of lattice spacing at physical light-quark masses and with high statistics. We consider it important that other ensemble sets be extended to a similar range, to enable further (sub)percent-level calculations with different systematics from the fermion discretization.

The inclusive determination of $|V_{ub}|$ will benefit from the calculation of new higher order effects, such as the $O(\alpha_s^3)$ contributions to the total width, and from a reassessment of QED effects. However, the next frontier is represented by the integration with lattice QCD calculations to improve the determination of HQE matrix elements, and eventually by the calculation of the inclusive rates directly on the lattice. For what concerns inclusive charmless decays, the general theoretical framework appears solid but needs to be updated in the light of recent higher order calculations and should be extensively validated by experimental data which will become available at Belle II. In particular, the measurement of the lepton energy and hadronic invariant mass distributions will provide important information on the Shape Functions, while the $q^2$ distribution will allow us to constrain and possibly avoid the effect of Weak Annihilation. The wealth of data expected at Belle II, a close cooperation between theorists and experimentalists, and hopefully new lattice data should help resolve various open issues, so that we might eventually expect the uncertainty on inclusive $|V_{ub}|$ to become lower than 3%.

Acknowledgements This work was supported by the Mainz Institute for Theoretical Physics (MITP) of the Cluster of Excellence PRISMA+ (Project ID 39083149) which hosted the workshop. We acknowledge the friendly effectiveness of its staff and thank the scientific coordinator Tobias Hurth and director Matthias Neubert for their encouragement. We are grateful to G. Caria, B. Dey, A. Grejlo, B. Grinstein, N. Gubernari, G. Heroiza, Y. Kwon, H. Meyer, R. Laha, W. Lee, P. Owen, S. Stefkova, P. Urquijo, who also participated in the workshop, F. Bernlochner and L. Cao were supported by the DFG Emmy-Noether Grant No. BE 6075/1-1. C. Davies was supported by the UK Science and Technology Facilities Council. A. El-Khadra was supported by the U.S. Department of Energy, Office of Science, Office of High Energy Physics under Award Number DE-SC0015655 and by the Fermilab Distinguished Scholars Program. P. Gambino and M. Jung were supported by the Italian Ministry of Research (MIUR) under grant PRIN 20172LNEEZ. S. Hashimoto was supported by JSPS KAK-ENHI Grant Number JP26247043 and by the Post-K and Fugaku supercomputer project through the Joint Institute for Computational Fundamental Science (JICFuS). The work of A. Khodjamirian and T. Mannel was supported by the DFG (German Research Foundation) under Grant 396021762–TRR 257 “Particle Physics Phenomenology after the Higgs Discovery”. Z. Ligeti was supported in part by the Office of High Energy Physics of the U.S. Department of Energy under contract DE-AC02-05CH11231. S. Meinel was supported by the U.S. Department of Energy, Office of Science, Office of High Energy Physics under Award Number DE-SC0009913. G. Paz was supported by the U.S. Department of Energy grant DE-SC0007983 and by a Career Development Chair award from Wayne State University. S. Schacht was supported by a DFG Forschungsstipendium under contract No. SCHA 2125/1-1. A. Vaquero was supported by the U.S. National Science Foundation under grants PHY14–14614 and PHY17–19626. This manuscript has been authored by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics.

Data Availability Statement This manuscript has no associated data or the data will not be deposited. [Authors’ comment: Data sharing to applicable to this article as no datasets were generated or analysed during the current study.]

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. Funded by SCOAP3.

References
1. N. Cabibbo, Phys. Rev. Lett. 10, 531 (1963). https://doi.org/10.1103/PhysRevLett.10.531
2. M. Kobayashi, T. Maskawa, Prog. Theor. Phys. 49, 652 (1973). https://doi.org/10.1143/PTP.49.652
3. S. Aoki, et al., (2019). arXiv:1902.08191 [hep-lat]
4. R. Aaij et al., Eur. Phys. J. C 79(8), 675 (2019). https://doi.org/10.1140/epjc/s10052-019-7112-x
