Search for the decays $B_{s}^{0} \rightarrow \tau^{+}\tau^{-}$
and $B^{0} \rightarrow \tau^{+}\tau^{-}$

The LHCb collaboration†

Abstract

A search for the rare decays $B_{s}^{0} \rightarrow \tau^{+}\tau^{-}$ and $B^{0} \rightarrow \tau^{+}\tau^{-}$ is performed using proton–proton collision data collected with the LHCb detector. The data sample corresponds to an integrated luminosity of 3 fb$^{-1}$ collected in 2011 and 2012. The $\tau$ leptons are reconstructed through the decay $\tau^{-} \rightarrow \pi^{-}\pi^{+}\pi^{-}\nu_{\tau}$. Assuming no contribution from $B^{0} \rightarrow \tau^{+}\tau^{-}$ decays, an upper limit is set on the branching fraction $B(B_{s}^{0} \rightarrow \tau^{+}\tau^{-}) < 6.8 \times 10^{-3}$ at 95% confidence level. If instead no contribution from $B_{s}^{0} \rightarrow \tau^{+}\tau^{-}$ decays is assumed, the limit is $B(B^{0} \rightarrow \tau^{+}\tau^{-}) < 2.1 \times 10^{-3}$ at 95% confidence level. These results correspond to the first direct limit on $B(B_{s}^{0} \rightarrow \tau^{+}\tau^{-})$ and the world’s best limit on $B(B^{0} \rightarrow \tau^{+}\tau^{-})$.


© CERN on behalf of the LHCb collaboration, licence CC-BY-4.0

†Authors are listed at the end of this Letter.
Processes where a $B$ meson decays into a pair of oppositely charged leptons are powerful probes in the search for physics beyond the Standard Model (SM). Recently, the first observation of the $B^0 \rightarrow \mu^+\mu^-$ decay was made \cite{12} (the inclusion of charge-conjugate processes is implied throughout this Letter). Its measured branching fraction ($\mathcal{B}$) is compatible with the SM prediction \cite{3} and imposes stringent constraints on theories beyond the SM. Complementing this result with searches for the tauonic modes $B \rightarrow \tau^+\tau^-$, where $B$ can be either a $B^0$ or a $B^0_s$ meson, is of great interest in view of the recent hints of lepton flavour non-universality obtained by several experiments. In particular the measurements of $R(D^{(*)}) = \frac{B(B^0\rightarrow D^{(*)}\ell^+\nu)}{B(B^0\rightarrow D^{(*)}\ell^+\nu)}$, where $\ell^+$ represents either a muon, an electron or both, are found to be larger than the SM prediction by 3.9 standard deviations ($\sigma$) \cite{4}, and the measurement of $R_K = \frac{B(B^+\rightarrow K^+\mu^+\mu^-)}{B(B^+\rightarrow K^+e^+e^-)}$ is 2.6$\sigma$ lower than the SM prediction \cite{5}.

Possible explanations for these and other \cite{6} deviations from their SM expectations include leptoquarks, $W'/Z'$ bosons and two-Higgs-doublet models (see e.g. Refs. \cite{7,8}). In these models, the $B \rightarrow \tau^+\tau^-$ branching fractions could be enhanced with respect to the SM predictions, $\mathcal{B}(B^0 \rightarrow \tau^+\tau^-) = (2.22 \pm 0.19) \times 10^{-8}$ and $\mathcal{B}(B^0_s \rightarrow \tau^+\tau^-) = (7.73 \pm 0.49) \times 10^{-7}$ \cite{3}, by several orders of magnitude \cite{8,12}. All minimal-flavour-violating models predict the same enhancement of $\mathcal{B}(B^0_s \rightarrow \tau^+\tau^-)$ over $\mathcal{B}(B^0 \rightarrow \tau^+\tau^-)$ as in the SM.

The experimental search for $B \rightarrow \tau^+\tau^-$ decays is complicated by the presence of at least two undetected neutrinos, originating from the decay of the $\tau$ leptons. The BaBar collaboration has searched for the $B^0 \rightarrow \tau^+\tau^-$ mode \cite{13} and published an upper limit $\mathcal{B}(B^0 \rightarrow \tau^+\tau^-) < 4.10 \times 10^{-3}$ at 90\% confidence level (CL). There are currently no experimental results for the $B^0_s \rightarrow \tau^+\tau^-$ mode, though its branching fraction can be indirectly constrained to be less than 3\% at 90\% CL \cite{14,15}.

In this Letter, the first search for the rare decay $B^0_s \rightarrow \tau^+\tau^-$ is presented, along with a search for the $B^0 \rightarrow \tau^+\tau^-$ decay. The analysis is performed with proton–proton collision data corresponding to integrated luminosities of 1.0 fb$^{-1}$ and 2.0 fb$^{-1}$ recorded with the LHCb detector at centre-of-mass energies of 7 and 8 TeV, respectively. The $\tau$ leptons are reconstructed through the decay $\tau^- \rightarrow \pi^- \pi^+\pi^-\nu$, which proceeds predominantly through the decay chain $\tau^- \rightarrow a_1(1260)\pi^-\nu$, $a_1(1260)\rightarrow \rho(770)\pi^-\pi^-$ \cite{16}. The branching fraction $\mathcal{B}(\tau^- \rightarrow \pi^- \pi^+\pi^-\nu) = (9.31 \pm 0.05)\%$ \cite{17}. Due to the final-state neutrinos the $\tau^+\tau^-$ mass provides only a weak discrimination between signal and background, and cannot be used as a way to distinguish $B^0_s$ from $B^0$ decays. The number of signal candidates is obtained from a fit to the output of a multivariate classifier that uses a range of kinematic and topological variables as input. Data-driven methods are used to determine signal and background models. The observed signal yield is converted into a branching fraction using as a normalisation channel the decay $B^0 \rightarrow D^-D^{+*}_{s}$ \cite{18,19}, with $D^- \rightarrow K^+\pi^+\pi^-$ and $D^{+*}_{s} \rightarrow K^-K^+\pi^+$. The LHCb detector, described in detail in Refs. \cite{20,21}, is a single-arm forward spectrometer covering the pseudorapidity range $2 < \eta < 5$. The online event selection is performed by a trigger \cite{22}, which consists of a hardware stage, based on information from the calorimeter and muon systems, followed by a software stage, which applies a full event reconstruction. The hardware trigger stage requires events to have a muon with high transverse momentum ($p_T$) or a hadron, photon or electron with high transverse energy in the calorimeters. For hadrons, the transverse energy threshold is around 3.5 GeV, depending on the data-taking conditions. The software trigger requires a two-, three- or
four-track secondary vertex with a significant displacement from the primary \( pp \) interaction vertices (PVs). A multivariate classifier \[23\] is used for the identification of secondary vertices that are significantly displaced from the PVs, and are consistent with the decay of a \( b \) hadron. At least one charged particle must have \( p_T > 1.7 \text{ GeV/c} \) and be inconsistent with originating from any PV.

Simulated data are used to optimise the selection, obtain the signal model for the fit and determine the selection efficiencies. In the simulation, \( pp \) collisions are generated using PYTHIA \[24\] with a specific LHCb configuration \[25\]. Decays of hadrons are described by EvtGen \[26\], in which final-state radiation is generated using PHOTOS \[27\]. The interaction of the generated particles with the detector, and its response, are implemented using the GEANT4 toolkit \[28\] as described in Ref. \[29\]. The \( \tau^− \rightarrow \pi^−\pi^+\pi^−\nu_\tau \) decays are generated using the resonance chiral Lagrangian model \[30\] with a tuning based on the BaBar results for the \( \tau^− \rightarrow \pi^−\pi^+\pi^−\nu_\tau \) decays \[31\], implemented in the TAUOLA generator \[32\].

In the offline selection of the candidate signal and normalisation decays, requirements on the particle identification (PID) \[33\], track quality and the impact parameter with respect to any PV are imposed on all charged final-state particles. Three charged tracks, identified as pions for the \( B \rightarrow \tau^+\tau^− \) decays, and pions or kaons for the \( B^0 \rightarrow D^-D^+_s \) decays, forming a good-quality vertex are combined to make intermediate \( \tau, D^+ \) and \( D^+_s \) candidates. The kinematic properties of these candidates, like momenta and masses, are calculated from the three-track combinations. The flight directions of the \( \tau, D^+ \) and \( D^+_s \) candidates are estimated from their calculated momentum vectors. For the \( \tau \) candidates this is a biased estimate due to the missing neutrinos. In turn, \( B \)-meson candidates are reconstructed from two oppositely charged \( \tau \) or from \( D^- \) and \( D^+_s \) candidates with decay vertices well separated from the PVs. The \( B \)-meson candidates are required to have \( p_T > 2 \text{ GeV/c} \), at least one \( \tau \), \( D^+ \) and \( D^+_s \) candidate with \( p_T > 4 \text{ GeV/c} \) and at least one pion or kaon with \( p_T > 2 \text{ GeV/c} \). No further selection requirements are imposed on the normalisation mode.

For each \( \tau \) candidate, the plane spanned by the invariant masses \( m_{\pi^+\pi^-} \) of the two oppositely charged two-pion combinations is divided into nine sectors, as illustrated in Fig. \[1\]. Exploiting the intermediate \( \rho(770)^0 \) resonance of the \( \tau \) decays, these sectors are used to define three regions. The signal region consists of \( B \) candidates with both \( \tau \) candidates in sector 5, and is used to determine the signal yield. The signal-depleted region, composed of \( B \) candidates having at least one \( \tau \) candidate in sectors 1, 3, 7 or 9, provides a sample used when optimising the selection. The control region corresponds to \( B \) candidates with one \( \tau \) candidate in sectors 4, 5 or 8 and the other in sectors 4 or 8, and provides the background model.

For the \( B \rightarrow \tau^+\tau^- \) modes, further requirements are imposed on two types of isolation variables that are able to discriminate signal from background from partially reconstructed decays with additional charged or neutral particles. The first class of isolation variables, based on the decision of a multivariate classifier trained on simulated signal and other \( b \)-hadron decays, is sensitive to additional charged tracks. These tracks either make a good-quality vertex with any selected pion or \( \tau \) candidate, or belong to the same \( b \)-hadron decay as the selected pion candidates. The second class of isolation variables is based on calorimeter activity due to neutral particles in a cone, defined in terms of the pseudorapidity and polar angle, centred on the \( B \) candidate momentum.

In addition to the isolation variables, a method to perform an analytic reconstruction
of the $B \to \tau^+\tau^-$ decay chain, described in detail in Refs. \cite{34,35}, has been developed. It combines geometrical information about the decay and mass constraints on the particles $(B, \tau$ and $\nu)$ in the decay chain to calculate the $\tau$ momenta analytically. The possible solutions for the two $\tau$ momenta are found as solutions of a system of two coupled equations of second degree with two unknowns. The finite detector resolution and approximations made in the calculation prevent real solutions being found for a substantial fraction of the signal events. However, several intermediate quantities associated with the method are exploited to discriminate signal from background.

The selection is refined by adding a requirement on the output of a neural network \cite{36}, built using seven variables: the $\tau^\pm$ candidate masses and decay times, a charged track isolation variable for the pions, a neutral isolation variable for the $B$ candidate, and one variable from the analytic reconstruction method, introduced in Ref. \cite{35}. The classifier is trained on simulated $B \to \tau^+\tau^-$ decays, representing the signal, and data events from the signal-depleted region.

In order to determine the signal yield, a binned maximum likelihood fit is performed on the output of a second neural network (NN), built with 29 variables and using the same training samples. The NN inputs include the eight variables from the analytic reconstruction method listed in Ref. \cite{35}, further isolation variables, as well as kinematic and geometrical variables. The NN output is transformed to obtain a flat distribution for the signal over the range $[0.0, 1.0]$, while the background peaks towards zero.

Varying the two-pion invariant mass sector boundaries, the signal region is optimised for the $B^0_s \to \tau^+\tau^-$ branching fraction limit using pseudoexperiments. The boundaries are set to 615 and 935 MeV/$c^2$. The overall efficiency of the selection, determined using

Figure 1: Distribution of simulated $B^0_s \to \tau^+\tau^-$ candidates in the plane spanned by the invariant masses $m_{\pi^+\pi^-}$ of the two oppositely charged two-pion combinations. The distribution is symmetric by construction. The vertical and horizontal lines illustrate the sector boundaries.
simulated $B_s^0 \to \tau^+\tau^-$ decays, is approximately $2.2(2.4) \times 10^{-5}$, including the geometrical acceptance. Assuming the SM prediction, the number of $B_s^0 \to \tau^+\tau^-$ decays expected in the signal region is 0.02.

After the selection, the signal, signal-depleted and control regions contain, respectively, 16%, 13% and 58% of the simulated signal decays. The corresponding fractions of selected candidates in data are 7%, 37% and 47%. Most signal decays fall into the control region, but the signal region, which contains about 14 700 candidates in data after the full selection, is more sensitive due to its lower background contamination. For the fit, ten equally sized bins of NN output in the range [0.0, 1.0] are considered, where the high NN region [0.7, 1.0] was not investigated until the fit strategy was fixed. The signal model is taken from the $B_s^0 \to \tau^+\tau^-$ simulation, while the background model is taken from the data control region, correcting for the presence of expected signal events in this region. The fit model is given by

$$N_{\text{data}}^{\text{SR}} = s \times \hat{N}_{\text{sim}}^{\text{SR}} + f_b \times \left( N_{\text{data}}^{\text{CR}} \times s \times \frac{\varepsilon_{\text{SR}}}{\varepsilon_{\text{CR}}} \times \hat{N}_{\text{sim}}^{\text{CR}} \right),$$

where $N_{\text{sim/data}}^{\text{SR/(CR)}}$ is the NN output distribution in the signal (control) region from simulation/data, $s$ is the signal yield in the signal region, $f_b$ is a scaling factor for the background template, and $\varepsilon_{\text{SR}}$ ($\varepsilon_{\text{CR}}$) is the signal efficiency in the signal (control) region. The quantities $s$ and $f_b$ are left free in the fit. The corresponding normalised distributions $\hat{N}_{\text{sim}}^{\text{SR/(CR)}}$ and $\hat{N}_{\text{data}}^{\text{CR}}$ are shown in Fig. 2.

The agreement between the background NN output distributions in the control and signal regions has been tested in different samples: in the data for the background-dominated NN output bins [0.0, 0.7], in a generic $b\bar{b}$ simulated sample and in several specific simulated background modes (such as $B^0 \to D^-\pi^+\pi^-\pi^+$ with $D^- \to K^0\pi^-\pi^+\pi^-$, or $B_s^0 \to D_s^-\pi^+\pi^-\pi^+\pi^-$ with $D_s^- \to \tau^-\nu_\tau$). The distributions have been found to agree with each other in all cases. The background in the control region can therefore be used to characterise the background in the signal region.
Differences between the shapes of the background distribution in the signal and control regions of the data are the main sources of systematic uncertainties on the background model. These uncertainties are taken into account by allowing each bin in the $N_{\text{data}}^{\text{CR}}$ distribution to vary according to a Gaussian constraint. The width of this Gaussian function is determined by splitting the control region into two approximately equally populated samples and taking, for each bin, the maximum difference between the NN outputs of the two subregions and the unsplit sample. The splitting is constructed to have one region more signal-like and one region more background-like.

The signal can be mismodelled in the simulation. The $B^0_s \rightarrow \tau^+\tau^-$ decay is used to compare data and simulation for the variables used in the NN. Ten variables are found to be slightly mismodelled and their distributions are corrected by weighting. The difference in the shape of the NN output distribution compared to the original unweighted sample is used to derive the associated systematic uncertainty. The fit procedure is validated with pseudoexperiments and is found to be unbiased. Assuming no signal contribution, the expected statistical (systematic) uncertainty on the signal yield is 58 (33). The fit result on data is shown in Fig. 3 and gives a signal yield $s = -23 \pm 63 \text{ (stat)} \pm 31 \text{ (syst)}$.

The $B^0_s \rightarrow \tau^+\tau^-$ signal yield is converted into a branching fraction using $\mathcal{B}(B^0_s \rightarrow \tau^+\tau^-) = \alpha^s \times s$, with

$$\alpha^s \equiv \frac{\epsilon^{D^- D^+_s} \times \mathcal{B}(B^0 \rightarrow D^- D^+_s) \times \mathcal{B}(D^+ \rightarrow K^- \pi^+ \pi^+) \times \mathcal{B}(D^+_s \rightarrow K^+ K^- \pi^+)}{N_{\text{obs}}^{D^- D^+_s} \times \epsilon^{\tau^+ \tau^-} \times [\mathcal{B}(\tau^- \rightarrow \pi^- \pi^+ \pi^- \nu_\tau)]^2} \times \frac{f_d}{f_s}, \quad (2)$$
where $\epsilon^{\tau+\tau-}$ and $\epsilon^{D^+D^-}$ are the combined efficiencies of trigger, reconstruction and selection of the signal and normalisation channels. The branching fractions used are $\mathcal{B}(B^0 \rightarrow D^-D^+_s) = (7.5 \pm 1.1) \times 10^{-3}$\textsuperscript{18}, $\mathcal{B}(D^- \rightarrow K^+\pi^-\pi^-) = (9.46 \pm 0.24)\%$\textsuperscript{17} and $\mathcal{B}(D^+_s \rightarrow K^-K^+\pi^+) = (5.45 \pm 0.17)\%$\textsuperscript{17}, and $f_s/f_d = 0.259 \pm 0.015$\textsuperscript{37} is the ratio of $B^0_s$ to $B^0$ production fractions. The efficiencies are determined using simulation, applying correction factors derived from data. The $B^0 \rightarrow D^-D^+_s$ yield, $N^{\text{obs}}_{D^-D^+_s}$, is obtained from a fit to the mass distribution, which has four contributions: the $B^0 \rightarrow D^-D^+_s$ component, modelled by a Hypatia function\textsuperscript{38}, a combinatorial background component, described by an exponential function, and two partially reconstructed backgrounds, $B^0 \rightarrow D^+D^-_s$ and $B^0 \rightarrow D^-D^+_s$, modelled as in Ref.\textsuperscript{39}. The resulting fit is shown in Fig.\textsuperscript{4} and gives a yield of $N^{\text{obs}}_{D^-D^+_s} = 10\,629 \pm 114$, where the uncertainty is statistical.

Uncertainties on $\alpha^s$ arise from the $B^0 \rightarrow D^-D^+_s$ fit model, the finite size of the simulated samples, the uncertainty from the corrections to the simulation and external inputs, which include the branching fractions and hadronisation fractions in Eq.\textsuperscript{2}. The latter contribution is dominant, giving a relative uncertainty of 17\% on $\alpha^s$. The $B^0 \rightarrow D^-D^+_s$ fit model is varied using the sum of two Gaussian functions with a common mean and power-law tails instead of the Hypatia function for the signal, a second-order Chebychev polynomial instead of an exponential function for the combinatorial background, and adding two other background components from $B^0 \rightarrow D^-D^+_s$ and $B^0 \rightarrow a_1(1260)^-D^+_s$ decays. The change in signal yield compared to the nominal fit is taken as a systematic uncertainty, adding the contributions from the four variations in quadrature. The overall relative uncertainty on $\alpha^s$ due to $N^{\text{obs}}_{D^-D^+_s}$ (including the fit uncertainty) is 1.7\%. Corrections determined from $J/\psi \rightarrow \mu^+\mu^-$ and $B^0 \rightarrow K^-\pi^+$ data control samples are applied for the tracking, PID and the hadronic hardware trigger efficiencies. The relative uncertainty on $\alpha^s$ due to selection efficiencies is 2.9\%, taking into account both the limited size of the simulated samples and the systematic uncertainties. The normalisation factor is found to be $\alpha^s = (4.07 \pm 0.70) \times 10^{-5}$.

The efficiencies and the shapes of the NN output distributions depend on the parametrisation used in the simulation to model the $\tau^- \rightarrow \pi^-\pi^-\pi^-\nu_\tau$ decay. The result obtained with the TAUOLA BaBar-tune model is therefore compared to available alternatives\textsuperscript{40}, which are based on CLEO data for the $\tau^- \rightarrow \pi^-\pi^0\pi^0\nu_\tau$ decay\textsuperscript{41}. The selection efficiency for these alternative models can be up to 20\% higher, due to different structures in the two-pion invariant mass, resulting in lower limits. Dependence of the NN signal output distribution on the $\tau$-decay model is found to be negligible. Since the alternative models are based on a different $\tau$ decay, the BaBar-tune model is chosen as default and no systematic uncertainty is assigned.

The signal yield obtained from the likelihood fit is translated into an upper limit on the $B^0 \rightarrow \tau^+\tau^-$ branching fraction using the CL$_S$ method\textsuperscript{42,43}. Assuming no contribution from $B^0 \rightarrow \tau^+\tau^-$ decays, an upper limit is set on the $B^0_s \rightarrow \tau^+\tau^-$ branching fraction of $5.2 (6.8) \times 10^{-3}$ at 90 (95)\% CL. This is the first experimental limit on $\mathcal{B}(B^0_s \rightarrow \tau^+\tau^-)$. The analysis is repeated for the $B^0 \rightarrow \tau^+\tau^-$ decay. The fit is performed by replacing the signal model with that derived from simulated $B^0 \rightarrow \tau^+\tau^-$ decays, giving $s = -15 \pm 67$\textsuperscript{35}. The corresponding normalisation factor is $\alpha^d = (1.16 \pm 0.19) \times 10^{-5}$. The limit obtained is $\mathcal{B}(B^0 \rightarrow \tau^+\tau^-) < 1.6 (2.1) \times 10^{-3}$ at 90 (95)\% CL, which constitutes a factor 2.6 improvement with respect to the BaBar result\textsuperscript{13} and is the current best limit on $\mathcal{B}(B^0 \rightarrow \tau^+\tau^-)$.
Figure 4: Invariant mass distribution of the reconstructed $B^0 \to D^- D_s^+$ candidates in data (black points), together with the total fit result (blue line) used to determine the $B^0 \to D^- D_s^+$ yield. The individual components are described in the text.

Acknowledgements

We thank Jérôme Charles (CPT, Marseille, France) for fruitful discussions and help in developing the analytic reconstruction method. We express our gratitude to our colleagues in the CERN accelerator departments for the excellent performance of the LHC. We thank the technical and administrative staff at the LHCb institutes. We acknowledge support from CERN and from the national agencies: CAPES, CNPq, FAPERJ and FINEP (Brazil); MOST and NSFC (China); CNRS/IN2P3 (France); BMBF, DFG and MPG (Germany); INFN (Italy); FOM and NWO (The Netherlands); MNiSW and NCN (Poland); MEN/IFA (Romania); MinES and FASO (Russia); MinECo (Spain); SNSF and SER (Switzerland); NASU (Ukraine); STFC (United Kingdom); NSF (USA). We acknowledge the computing resources that are provided by CERN, IN2P3 (France), KIT and DESY (Germany), INFN (Italy), SURF (The Netherlands), PIC (Spain), GridPP (United Kingdom), RRCKI and Yandex LLC (Russia), CSCS (Switzerland), IFIN-HH (Romania), CBPF (Brazil), PL-GRID (Poland) and OSC (USA). We are indebted to the communities behind the multiple open source software packages on which we depend. Individual groups or members have received support from AvH Foundation (Germany), EPLANET, Marie Skłodowska-Curie Actions and ERC (European Union), Conseil Général de Haute-Savoie, Labex ENIGMASS and OCEVU, Région Auvergne (France), RFBR and Yandex LLC (Russia), GVA, XuntaGal and GENCAT (Spain), Herchel Smith Fund, The Royal Society, Royal Commission for the Exhibition of 1851 and the Leverhulme Trust (United Kingdom).
References

[1] CMS and LHCB collaborations, V. Khachatryan et al., Observation of the rare $B^0_s \rightarrow \mu^+ \mu^-$ decay from the combined analysis of CMS and LHCB data, Nature 522 (2015) 68 arXiv:1411.4413.

[2] LHCB collaboration, R. Aaij et al., Measurement of the $B^0 \rightarrow \mu^+ \mu^-$ branching fraction and effective lifetime and search for $B^0 \rightarrow \mu^+ \mu^-$ decays, LHCB-PAPER-2017-001, in preparation.


[35] See appendix A for further details.


[39] LHCb collaboration, R. Aaij et al., First observations of $\bar{B}^0_s \to D^+ D^-$, $D^+ D^-$ and $D^0\bar{D}^0$ decays, Phys. Rev. D87 (2013) 092007, arXiv:1302.5854.


[41] CLEO collaboration, D. M. Asner et al., Hadronic structure in the decay $\tau^- \to \nu_\tau \pi^- \pi^0 \pi^0$ and the sign of the tau neutrino helicity, Phys. Rev. D61 (2000) 012002, arXiv:hep-ex/9902022.


A Supplemental material

In Section A.1, details about the analytic reconstruction method are given. Sections A.2 and A.3 contain additional results related to the \( B_s^0 \rightarrow \tau^+\tau^- \) and \( B^0 \rightarrow \tau^+\tau^- \) decay channels.

A.1 Reconstruction method

A method to perform an analytic reconstruction of the \( B \rightarrow \tau^+\tau^- \) decay chain is described in the following and in detail in Ref. [34]. It combines geometrical information about the decay and sets mass constraints on the particles in the decay chain (\( B_s^0 \), \( \tau \) and \( \nu \)) to calculate the \( \tau \) momenta analytically. In these calculations, Lorentz invariance is kept manifest and the possible values for the two \( \tau \) momenta are found as analytic solutions of a system of two coupled equations of second degree in two unknowns. The only remaining degree of freedom is a single Lorentz scalar, introduced below as the angle \( \theta \), measuring the asymmetry of the decay triangle in the decay-time space of the two \( \tau \) leptons.

In the following, the unknown momenta of the \( \tau^\pm \) leptons, the primary parameters of interest, are labelled by the four-vectors \( p_\mu^\pm \); the \( B_s^0 \rightarrow \tau^+\tau^- \) decay plane is defined by the \( B \) production vertex, i.e. the PV, and the two \( \tau \) decay vertices; the three-vectors pointing from the PV to the \( \tau^\pm \) decay vertices are labelled \( \vec{w}_\pm \); the time intervals between the \( B \) production and the \( \tau^\pm \) decays, \( w_0^\pm \), which cannot be measured, act as the temporal counterpart to \( \vec{w}_\pm \). Together, they make the four-vectors \( w_\mu^\pm = (w_0^\pm, \vec{w}_\pm) \). Introducing the notation

\[
W \equiv (w_\mu^+, w_\mu^-), \quad P \equiv (p_\mu^+, p_\mu^-),
\]

the momenta \( p_\mu^\pm \) can now be obtained from the coupled set of equations

\[
W = H \cdot P
\]  

by imposing momentum conservation throughout the decay chain. Here

\[
H \equiv \left( \begin{array}{cc} \frac{t_B}{m_B} + \frac{t_\tau}{m_\tau} & \frac{t_B}{m_B} \\ \frac{t_B}{m_B} & \frac{t_\tau}{m_\tau} + \frac{t_\tau}{m_\tau} \end{array} \right)
\]

is given in terms of the \( B \) and \( \tau \) masses, \( m_i \), and decay times, \( t_i \). Using the on-shell and flight-direction constraints, it is possible to rewrite Eq. (4) in terms of a single unknown, chosen to be the rotation angle, \( \theta \), that diagonalises the matrix \( H \). Thanks to this transformation, solving Eq. (4) becomes equivalent to finding the roots of a fourth-order polynomial

\[
P^{(4)}(\xi) = \sum_{i=0}^{4} a_i(\theta)\xi^i = 0,
\]

where explicit expressions for the coefficients \( a_i(\theta) \) in terms of \( \theta \) and measurable quantities are given in Appendix D of Ref. [34].

Even though it is possible, in principle, to exactly determine the angle \( \theta \), a different approach has been used, because of the high complexity of the trigonometric equations involved. The value of the angle \( \theta \) is in fact approximated in the calculation of the complex solutions \( \xi_1, \ldots, \xi_4 \). Three approximations have been considered in Ref. [34]. They are
1. \( \theta = \langle \theta \rangle = \pi/4 \), representing the case where both \( \tau \) leptons have the same decay time.

2. \( \theta = \bar{\theta} \), where \( \bar{\theta} \) is the angle diagonalising the matrix

\[
\overline{H} \equiv \left( \begin{array}{cc}
1 + \frac{|\vec{w}_+|}{|\vec{w}_+ + \vec{w}_-|} & 1 + \frac{1}{|\vec{w}_+ + \vec{w}_-|} \\
1 + \frac{1}{|\vec{w}_+ + \vec{w}_-|} & 1 + \frac{|\vec{w}_-|}{|\vec{w}_+ + \vec{w}_-|}
\end{array} \right)
\]  

(7)

3. \( \theta = \theta^* \), where \( \theta^* \) is the result of applying iterative corrections to \( \theta = \bar{\theta} \).

These approximations, together with the finite detector resolution, prevent having real solutions for a substantial fraction of the signal events. Nonetheless, quantities appearing in intermediate calculations, though not having immediate physical meaning, have been found useful to discriminate between signal and background. The most powerful of these variables are exploited in the two neural networks that are used in the candidate selection. The variable

- \( \Re e[\vec{p}_+ (\theta = \bar{\theta}, \xi_3)] \)

is used by both the first and second NN, while the seven variables

- \( \Im m[\vec{p}_+ (\theta = \theta^*, \xi_1)] \),
- \( \Im m[\vec{p}_+(\theta = \theta^*, \xi_1) \vec{p}_-(\theta = \theta^*, \xi_1)] \),
- \( \Im m[\vec{p}_+(\theta = \theta^*, \xi_3) \vec{p}_-(\theta = \theta^*, \xi_3)] \),
- \( \Re e[\vec{p}_+ (\theta = \pi/4, \xi_1) \vec{p}_-(\theta = \pi/4, \xi_1)] \),
- \( \Re e[\xi_1(\theta = \theta^*)] \),
- \( \bar{\theta} \),
- \( p_+ \cdot p_- \),

are used by the second NN. Here \( \vec{p}_\pm^2 \equiv M_\pm^2 = m_\tau^2 \pm \sin(2\theta)s \), and \( s \equiv p_+ \cdot p_- \) (for signal events \( \hat{s} = \frac{M_\pm^2 - 2m_\tau^2}{2} \)).
A.2 Additional $B^{0}_{s}$ fit result

Figure 5 shows the fit result using only the background model. A likelihood-ratio test is performed comparing the nominal fit with the background-only alternative. The $p$-value of the likelihood-ratio test is 0.06, and the associated $z$-score is 1.60, showing that the data are consistent with the background-only hypothesis. Figure 6 shows the profile likelihood of the nominal fit. Figure 7 shows the expected and observed $C_{Ls}$ values as a function of the branching fraction. The expected limit for the $B^{0}_{s}$ mode is $\mathcal{B}(B^{0}_{s} \to \tau^{+}\tau^{-}) < 5.7 (7.4) \times 10^{-3}$ at 90 (95)% CL.

A.3 Additional $B^{0}$ fit result

The NN output distributions for simulated $B^{0} \to \tau^{+}\tau^{-}$ decays in the signal and control regions are shown in Fig. 8. The fit result, assuming no contribution from $B^{0}_{s} \to \tau^{+}\tau^{-}$ decays, is shown in Fig. 9 and Fig. 10 shows the expected and observed $C_{Ls}$ values as a function of the branching fraction. The expected limit for the $B^{0}$ mode is $\mathcal{B}(B^{0} \to \tau^{+}\tau^{-}) < 1.7 (2.1) \times 10^{-3}$ at 90 (95)% CL.
Figure 6: Profile likelihood scan of the $B^0_s \rightarrow \tau^+ \tau^-$ signal yield. The intersections of the likelihood curve with the horizontal lines define the 68.3%, 95.4% and 99.7% likelihood intervals.

Figure 7: The $p$-value derived with the CL$_s$ method as a function of $B(B^0_s \rightarrow \tau^+ \tau^-)$. Expected (observed) values are shown by a dashed (full) black line. The green (yellow) band covers the regions of 68% and 95% confidence for the expected limit. The red horizontal line corresponds to the limit at 95% CL.
Figure 8: Normalised NN output distribution in the signal ($\hat{N}_{\text{sim}}^{\text{SR}}$) and control ($\hat{N}_{\text{sim}}^{\text{CR}}$) region for $B^0 \rightarrow \tau^+ \tau^-$ simulated events.

Figure 9: Fit results for $B^0 \rightarrow \tau^+ \tau^-$. Distribution $N_{\text{data}}^{\text{SR}}$ (black points), overlaid with the total fit result (blue line), and background component (green line), assuming all signal events originate from $B^0 \rightarrow \tau^+ \tau^-$ decays. The $B^0 \rightarrow \tau^+ \tau^-$ signal component is also shown (red line), multiplied by $-1$. For each bin of the signal and background component the combined statistical and systematic uncertainty is shown as a light-coloured band. The difference between data and fit divided by its uncertainty (pull) is shown underneath.
Figure 10: The $p$-value derived with the CL$_s$ method as a function of $B(B^0 \rightarrow \tau^+ \tau^-)$. Expected (observed) values are shown by a dashed (full) black line. The green (yellow) band covers the regions of 68% and 95% confidence for the expected limit. The red horizontal line corresponds to the limit at 95% CL.
LHCb collaboration
Laboratori Nazionali dell'INFN di Frascati, Frascati, Italy
Sezione INFN di Genova, Genova, Italy
Sezione INFN di Milano Bicocca, Milano, Italy
Sezione INFN di Milano, Milano, Italy
Sezione INFN di Padova, Padova, Italy
Sezione INFN di Pisa, Pisa, Italy
Sezione INFN di Roma Tor Vergata, Roma, Italy
Sezione INFN di Roma La Sapienza, Roma, Italy
Henryk Niewodniczanski Institute of Nuclear Physics Polish Academy of Sciences, Kraków, Poland
AGH - University of Science and Technology, Faculty of Physics and Applied Computer Science, Kraków, Poland
National Center for Nuclear Research (NCBJ), Warsaw, Poland
Horia Hulubei National Institute of Physics and Nuclear Engineering, Bucharest-Mugurele, Romania
Petersburg Nuclear Physics Institute (PNPI), Gatchina, Russia
Institute of Theoretical and Experimental Physics (ITEP), Moscow, Russia
Institute of Nuclear Physics, Moscow State University (SINP MSU), Moscow, Russia
Institute for Nuclear Research of the Russian Academy of Sciences (INR RAN), Moscow, Russia
Yandex School of Data Analysis, Moscow, Russia
Budker Institute of Nuclear Physics (SB RAS), Novosibirsk, Russia
Institute for High Energy Physics (IHEP), Protvino, Russia
ICCB, Universitat de Barcelona, Barcelona, Spain
Universidad de Santiago de Compostela, Santiago de Compostela, Spain
European Organization for Nuclear Research (CERN), Geneva, Switzerland
Institute of Physics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
Physik-Institut, Universität Zürich, Zürich, Switzerland
Nikhef National Institute for Subatomic Physics, Amsterdam, The Netherlands
Nikhef National Institute for Subatomic Physics and VU University Amsterdam, Amsterdam, The Netherlands
NSC Kharkiv Institute of Physics and Technology (NSC KIPT), Kharkiv, Ukraine
Institute for Nuclear Research of the National Academy of Sciences (KINR), Kyiv, Ukraine
University of Birmingham, Birmingham, United Kingdom
H.H. Wills Physics Laboratory, University of Bristol, Bristol, United Kingdom
Cavendish Laboratory, University of Cambridge, Cambridge, United Kingdom
Department of Physics, University of Warwick, Coventry, United Kingdom
STFC Rutherford Appleton Laboratory, Didcot, United Kingdom
School of Physics and Astronomy, University of Edinburgh, Edinburgh, United Kingdom
School of Physics and Astronomy, University of Glasgow, Glasgow, United Kingdom
Oliver Lodge Laboratory, University of Liverpool, Liverpool, United Kingdom
Imperial College London, London, United Kingdom
School of Physics and Astronomy, University of Manchester, Manchester, United Kingdom
Department of Physics, University of Oxford, Oxford, United Kingdom
Massachusetts Institute of Technology, Cambridge, MA, United States
University of Cincinnati, Cincinnati, OH, United States
University of Maryland, College Park, MD, United States
Syracuse University, Syracuse, NY, United States
Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), Rio de Janeiro, Brazil, associated to 2
University of Chinese Academy of Sciences, Beijing, China, associated to 3
School of Physics and Technology, Wuhan University, Wuhan, China, associated to 3
Institute of Particle Physics, Central China Normal University, Wuhan, Hubei, China, associated to 3
Departamento de Fisica , Universidad Nacional de Colombia, Bogota, Colombia, associated to 8
Institut für Physik, Universität Rostock, Rostock, Germany, associated to 12
National Research Centre Kurchatov Institute, Moscow, Russia, associated to 32
Instituto de Fisica Corpuscular, Centro Mixto Universidad de Valencia - CSIC, Valencia, Spain, associated to 38
Van Swinderen Institute, University of Groningen, Groningen, The Netherlands, associated to 43

Universidade Federal do Triângulo Mineiro (UFTM), Uberaba-MG, Brazil
Laboratoire Leprince-Ringuet, Palaiseau, France
P.N. Lebedev Physical Institute, Russian Academy of Science (LPI RAS), Moscow, Russia
Università di Bari, Bari, Italy
Università di Bologna, Bologna, Italy
Università di Cagliari, Cagliari, Italy
Università di Ferrara, Ferrara, Italy
Università di Genova, Genova, Italy
Università di Milano Bicocca, Milano, Italy
Università di Roma Tor Vergata, Roma, Italy
Università di Roma La Sapienza, Roma, Italy
AGH - University of Science and Technology, Faculty of Computer Science, Electronics and Telecommunications, Kraków, Poland
LIFAELS, La Salle, Universitat Ramon Llull, Barcelona, Spain
Hanoi University of Science, Hanoi, Viet Nam
Università di Padova, Padova, Italy
Università di Pisa, Pisa, Italy
Università degli Studi di Milano, Milano, Italy
Università di Urbino, Urbino, Italy
Università della Basilicata, Potenza, Italy
Scuola Normale Superiore, Pisa, Italy
Università di Modena e Reggio Emilia, Modena, Italy
Iligan Institute of Technology (IIT), Iligan, Philippines
Novosibirsk State University, Novosibirsk, Russia

† Deceased