Reply to Referee B

June 25, 2026

Thank you for your time reviewing this paper. We appreciate for your insightful comments to improve the paper draft. Plese find the following replies to your comments, and let us know if you have any further comments.

1. The $e^{\pm}\mu^{\mp}$ fit in Fig. 2 uses three bins [9.814, 9.862], [9.862, 9.886], [9.886, 9.920] GeV with all three $\chi_{bJ}(1P)$ yields floated simultaneously. These bin boundaries do not align with the PDG masses (9.859, 9.893, 9.912 GeV) to cover the corresponding $\chi_{bJ}(1P)$ states. What is the rationale for these bin bound- aries, and why are only three bins used here whereas Fig. 3 uses multiple bins with width 0.0025 GeV? How does the $M^{\rm recoil}_{\gamma_1\ell_1}$ distribution looks for the $e^{\pm}\mu^{\mp}$ final state? Finally, how stable are the upper limits under small boundary shifts, and was a cross-check performed by constraining the J = 0 and J = 2 yields via the production ratios in Table I?

   Ans: The label on the horizontal axis of Fig. 2 in paper draft was incorrect. We apologize for that. Figure 1 compares the $M^{\rm recoil}_{\gamma_1}$ distributions with incorrect vs correct horizontal axis labels. For all the decay processes, we extract the signal yields from $M^{\rm recoil}_{\gamma_1}$ within 9.81 to 9.95 (as mentioned on page 3 of the paper draft). Our fit uses the three bins with boundaries at 9.810, 9.880, 9.915, and 9.950 GeV. This method uses maximum likelihood estimation for three $\chi_{bJ}(1P)$ modes simultaneously using the histogram templates assuming Poisson errors for each of bin. Although the method itself does not prefer any bin-boundaries, they are chosen to avoid large Poisson uncertainties in the likelihood function because of empty bins in the background component. For $e^{\pm}\tau^{\mp}$ and $\mu^{\pm}\tau^{\mp}$ modes, the number of events is sufficient to a use unbinned maximum likelihood approach (as mentioned in the paper draft), and we use 0.0025 GeV bin for illustration purpose only.

   The distribution of $M^{\rm recoil}_{\gamma_1\ell_1}$ for $e^{\pm}\mu^{\mp}$ mode is a very narrow peak at the muon mass after 4-constraint kinametic fit, which is exclusively used in the $e^{\pm}\mu^{\mp}$ channel.

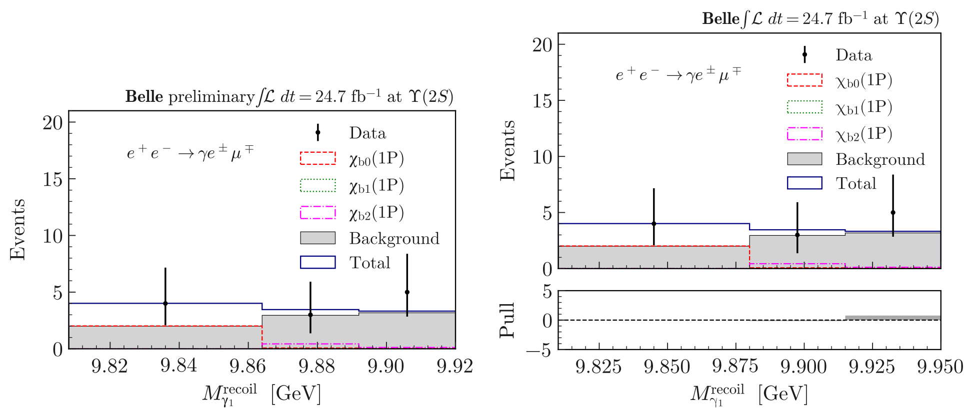

Figure 1: $M^{\rm recoil}_{\gamma_1}$ distribution with with incorrect label (left) and correct label (right).

   To check the boundary effect, we estimate the upper limits of branching fractions shifting the bin boundaries by ± 5 MeV. Obtained upper limit of branching fraction for $\chi_{b2}(1P) \to e^{\pm}\mu^{\mp}$ is 1.5×10^{-6} for +5 MeV shift of boundaries, while other results remain unchanged. The nominal result conservatively includes the augmented one. Also, obtained upper limit of branching fraction for $\chi_{b1}(1P) \to e^{\pm}\mu^{\mp}$ constraining the yields of $\chi_{b0}(1P)$ and $\chi_{b2}(1P)$ according the Table I in paper draft is 1.4×10^{-6}. The result slightly differs from the value in table III, with the value in the table being more conservative.

2. Upper limits for the $e^\pm\mu^\mp$ channels use pyhf, while the remaining six channels with $\ell\tau$ final states use a frequentist calculator from RooStats, both under the CLs formalism. What is the justification for using different implementations across channels that are jointly reported in Table III? Have the two calculators been validated against each other on a common test case?
Ans: For $\ell^\pm\tau^\mp$, we could obtain the shape of background due to sufficient number of events inside the signal search window. Although both methods use CLs formalism, we reach better sensivity for $\ell^\pm\tau^\mp$ cases using 2D probablity density function (PDF) with unbinned datasets. As the number of events for $e^\pm\mu^\mp$ case is not sufficient to obtain the reliable PDF for background shape, we use a method that does not require the PDF for background.

3. Table II shows that the "Control modes" systematic uncertainty (19.9% for $J = 0$, 10.0% for $J = 1$ and 2) dominates over all other sources combined by nearly an order of magnitude, yet no explanation of its origin is given in the text. What does this uncertainty represent? Additionally, what is the origin of the asymmetric PDF parameter uncertainty quoted as $^{+1.8}_{-1.2}\%$ for the τ channels?
Ans: The uncertainties stem from the $\Upsilon(2S) \to \gamma\chi_{bJ}(1P)$ branching fractions. The decay to $\chi_{b0}(1P)$ has a larger uncertainty than the others. We have changed the description of the uncertainty in the text to make this clearer. The asymmetric systematic is obtained from control modes by varying parameters of signal PDF by ±1 standard deviation. We have included this information in the new draft.

4. The best candidate selection in the τ modes retains the correct candidate only 75% of the time compared with 98% for $e^\pm\mu^\mp$. The paper mentions that this does not bias the results. Can you demonstrate that the 25% of mis-selected events do not sculpt a false signal peak or trough in $M^{\rm recoil}_{\gamma_1\ell_1}$?

   Ans: We rule out the potential bias by checking that $M^{\rm recoil}_{\gamma_1\ell_1}$, whose distribution we fit to, is not correlated with E_{γ_1}, which we use to select the best candidate. The following figure shows they have negligible correlation (-3.9%). Also, $M^{\rm recoil}_{\gamma_1\ell_1}$ distributions of events with one reconstructed candidate and accepted events from multiple reconstructed candidates are shown in Figure 2 using data for $e^\pm\tau^\mp$ mode, which shows no false peaking in the entire signal region.

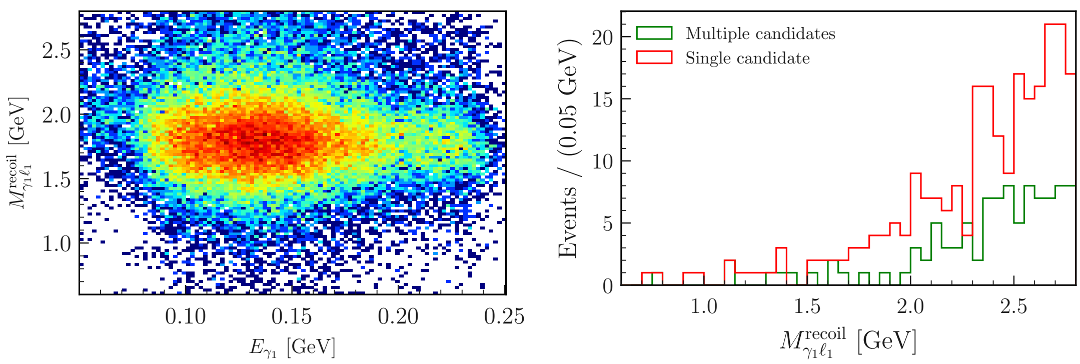

Figure 2: Distribution of E_{γ_1} vs $M^{\rm recoil}_{\gamma_1\ell_1}$ for simulated $e^\pm\tau^\mp$ events (left) and events with vs without best candidate selection (right) using data for $e^\pm\tau^\mp$ mode.

5. The Wilson coefficient bounds in Fig. 4 are derived from Eq. (4) using several theory inputs. The construction of the exclusion circles is not well described. Is it constructed via a simultaneous two-parameter scan or a fixed ratio of C_{SL} and C_{SR}? Can the authors clarify how Fig. 4 is constructed, so that the bounds can be reproduced independently?
Ans: We obtain the following equation from equation (16) and (17) of Ref [1]. Substituting all the values of parameters including the upper limits of branching fractions, we have obtained the fixed circles, which have been plotted on Figure 4 of the paper draft. We have modified the text to include such information.

$$\left(\frac{C^{\rm q\ell_1\ell_2}_{\rm SL}}{\Lambda^2}\right)^2 + \left(\frac{C^{\rm q\ell_1\ell_2}_{\rm SR}}{\Lambda^2}\right)^2 = \frac{16\pi m_\chi \Gamma_\chi \mathcal{B}(\chi \to \ell_1\ell_2)}{G_{\rm F}^2 f_\chi^2 m_{\rm b}^2 m_{\ell_2}^2 (m_\chi^2 - m_{\ell_2}^2)^2}. \quad (1)$$

6. The requirement $M^{\rm recoil}_{\gamma_1\ell_1} \in [0.6, 2.8]$ GeV for the τ modes is stated without motivation. What drives this choice? Background suppression, signal efficiency or something else? Has the sensitivity been checked for small variations of this boundary?
Ans: The number of backgrounds events increases exponentially on higher $M^{\rm recoil}_{\gamma_1\ell_1}$ region, and it potentially decreases to 0 on the lower $M^{\rm recoil}_{\gamma_1\ell_1}$ region. Therefore, we choose the region $M^{\rm recoil}_{\gamma_1\ell_1} \in [0.6, 2.8]$ GeV, which corresponds to around 2.5 times of its resolution. Text is modified to include such information. We have varified that the upper limits remain unchanged within ± 0.1 GeV of the $M^{\rm recoil}_{\gamma_1\ell_1}$ window.

7. Table III states that the efficiencies $\epsilon_{J,\ell_1,\ell_2}$ include "the relevant τ branching fractions", each of which is approximately 17%. Could the authors confirm explicitly that the tabulated values already incorporate this factor, so that the quoted upper limits on $\mathcal{B}(\chi_{bJ}(1P) \to \ell_1\ell_2)$ can be used directly without further rescaling?
Ans: Yes, $\epsilon_{J,\ell_1,\ell_2}$ includes branching fraction of relevant τ decays, $\mathcal{B}(\chi_{bJ}(1P) \to \ell_1\ell_2)$ does not require any further scaling. We have updated the text.

8. The EFT interpretation presented in the manuscript focuses primarily on scalar operators, motivated by the sensitivity of the $\chi_{b0}(1P)$ state to scalar currents, while contributions from vector and tensor operators are only briefly mentioned. However, the broader BSM implications of the results are not discussed. In particular, can the authors provide a more explicit mapping of the reported limits onto representative BSM scenarios, such as leptoquark or Z' models? A more complete discussion along these lines would significantly improve the physical interpretation and impact of the results.
Ans: The main focus of this work is to test lepton flavor violation in direct searches of $\chi_{bJ}(1P)$ meson decays. We provide the sensitivity of EFT scalar operators using the results for $\chi_{b0}(1P)$ decays, which is the most interesting part of this analysis. Providing more explicit mapping of BSM scenarios like leptoquarks or Z' models is beyond the scope of this work.

9. The correlation coefficients Cmn between the $J = 0, 1, 2$ signal yields are listed in Table III where the value $C_{20} = -33.0\%$ for $e^\pm\tau^\mp$ is unexpectedly large. Can the authors comment on its physical origin and confirm that it is stable across pseudo-experiments?
Ans: High value of C_{20} arises from the fluctuation of the experimental data, that is now mentioned in the modified draft. We have studied the correlation coefficients using toy Monte Carlo samples, where the mean values of C_{02} sample is around $+10\%$ with a standard deviation of 11%.

10. The CM energy of 10.023 GeV is mentioned with a citation to Ref. [22], which appears in the reference list as the statement that "natural units are used throughout the Letter". Could an appropriate reference be cited for this instead?
Ans: We wanted to add a footnote, which ended up being reflected as a reference. We have corrected it.

11. The author list includes both the Belle and Belle II collaboration names, but the analysis uses exclusively data recorded by the Belle detector running till 2010. A brief clarifying statement in the text would prevent confusion for readers less familiar with the two experiments.
Ans: Belle II is the successor of Belle experiment. Although the data was taken before 2010, we share software frameworks, computing facilities, collaborators for this work.

References

[1] Derek E. Hazard and Alexey A. Petrov. Lepton flavor violating quarkonium decays. Phys. Rev. D, 94(7):074023, 2016.