

Search for a $\mu^+\mu^-$ resonance in four-muon final states at Belle II

I. Adachi¹, K. Adamczyk², L. Aggarwal³, H. Ahmed⁴, H. Aihara⁵, N. Akopov⁶, A. Aloisio⁷, N. Anh Ky⁸, D. M. Asner⁹, H. Atmacan¹⁰, V. Aushev¹¹, M. Aversano¹², R. Ayad¹³, V. Babu¹⁴, H. Bae¹⁵, S. Bahinipati¹⁶, P. Bambade¹⁷, Sw. Banerjee¹⁸, S. Bansal¹⁹, M. Barrett²⁰, J. Baudot²¹, A. Baur²², A. Beaubien²³, F. Becherer²⁴, J. Becker²⁵, J. V. Bennett²⁶, F. U. Bernlochner²⁷, V. Bertacchi²⁸, M. Bertemes²⁹, E. Bertholet³⁰, M. Bessner³¹, S. Bettarini³², B. Bhuyan³³, F. Bianchi³⁴, T. Bilka³⁵, S. Bilokin³⁶, D. Biswas³⁷, A. Bobrov³⁸, D. Bodrov³⁹, A. Bolz⁴⁰, A. Bozek⁴¹, M. Bračko⁴², P. Branchini⁴³, T. E. Browder⁴⁴, A. Budano⁴⁵, S. Bussino⁴⁶, M. Campajola⁴⁷, L. Cao⁴⁸, G. Casarosa⁴⁹, C. Cecchi⁵⁰, J. Cerasoli⁵¹, M.-C. Chang⁵², P. Chang⁵³, R. Cheaib⁵⁴, P. Cheema⁵⁵, B. G. Cheon⁵⁶, K. Chilikin⁵⁷, K. Chirapatpimol⁵⁸, H.-E. Cho⁵⁹, K. Cho⁶⁰, S.-J. Cho⁶¹, S.-K. Choi⁶², S. Choudhury⁶³, L. Corona⁶⁴, L. M. Cremaldi⁶⁵, S. Das⁶⁶, F. Dattola⁶⁷, E. De La Cruz-Burelo⁶⁸, S. A. De La Motte⁶⁹, G. De Nardo⁷⁰, M. De Nuccio⁷¹, G. De Pietro⁷², R. de Sangro⁷³, M. Destefanis⁷⁴, R. Dhamija⁷⁵, A. Di Canto⁷⁶, F. Di Capua⁷⁷, J. Dingfelder⁷⁸, Z. Doležal⁷⁹, T. V. Dong⁸⁰, M. Dorigo⁸¹, K. Dort⁸², S. Dreyer⁸³, S. Dubey⁸⁴, G. Dujany⁸⁵, P. Ecker⁸⁶, M. Eliachevitch⁸⁷, D. Epifanov⁸⁸, P. Feichtinger⁸⁹, T. Ferber⁹⁰, D. Ferlewicz⁹¹, T. Fillinger⁹², C. Finck⁹³, G. Finocchiaro⁹⁴, A. Fodor⁹⁵, F. Forti⁹⁶, A. Frey⁹⁷, B. G. Fulsom⁹⁸, A. Gabrielli⁹⁹, E. Ganiev¹⁰⁰, M. Garcia-Hernandez¹⁰¹, R. Garg¹⁰², G. Gaudino¹⁰³, V. Gaur¹⁰⁴, A. Gaz¹⁰⁵, A. Gellrich¹⁰⁶, G. Ghevondyan¹⁰⁷, D. Ghosh¹⁰⁸, H. Ghumaryan¹⁰⁹, G. Giakoustidis¹¹⁰, R. Giordano¹¹¹, A. Giri¹¹², A. Glazov¹¹³, B. Gobbo¹¹⁴, R. Godang¹¹⁵, O. Gogota¹¹⁶, P. Goldenzweig¹¹⁷, W. Gradl¹¹⁸, T. Grammatico¹¹⁹, E. Graziani¹²⁰, D. Greenwald¹²¹, Z. Gruberová¹²², T. Gu¹²³, K. Gudkova¹²⁴, S. Halder¹²⁵, Y. Han¹²⁶, T. Hara¹²⁷, H. Hayashii¹²⁸, S. Hazra¹²⁹, C. Hearty¹³⁰, M. T. Hedges¹³¹, A. Heidelberg¹³², I. Heredia de la Cruz¹³³, M. Hernández Villanueva¹³⁴, T. Higuchi¹³⁵, M. Hoek¹³⁶, M. Hohmann¹³⁷, P. Horak¹³⁸, C.-L. Hsu¹³⁹, T. Humair¹⁴⁰, T. Iijima¹⁴¹, G. Inguglia¹⁴², N. Ipsita¹⁴³, A. Ishikawa¹⁴⁴, R. Itoh¹⁴⁵, M. Iwasaki¹⁴⁶, P. Jackson¹⁴⁷, W. W. Jacobs¹⁴⁸, E.-J. Jang¹⁴⁹, Q. P. Ji¹⁵⁰, S. Jia¹⁵¹, Y. Jin¹⁵², K. K. Joo¹⁵³, H. Junkerkalefeld¹⁵⁴, D. Kalita¹⁵⁵, J. Kandra¹⁵⁶, K. H. Kang¹⁵⁷, G. Karyan¹⁵⁸, T. Kawasaki¹⁵⁹, F. Keil¹⁶⁰, C. Kiesling¹⁶¹, C.-H. Kim¹⁶², D. Y. Kim¹⁶³, K.-H. Kim¹⁶⁴, Y.-K. Kim¹⁶⁵, H. Kindo¹⁶⁶, K. Kinoshita¹⁶⁷, P. Kodyš¹⁶⁸, T. Koga¹⁶⁹, S. Kohani¹⁷⁰, K. Kojima¹⁷¹, A. Korobov¹⁷², S. Korpar¹⁷³, E. Kovalenko¹⁷⁴, R. Kowalewski¹⁷⁵, T. M. G. Kraetzschmar¹⁷⁶, P. Križan¹⁷⁷, P. Krokovny¹⁷⁸, T. Kuhr¹⁷⁹, J. Kumar¹⁸⁰, M. Kumar¹⁸¹, R. Kumar¹⁸², K. Kumara¹⁸³, T. Kunigo¹⁸⁴, A. Kuzmin¹⁸⁵, Y.-J. Kwon¹⁸⁶, S. Lacaparra¹⁸⁷, Y.-T. Lai¹⁸⁸, T. Lam¹⁸⁹, L. Lanceri¹⁹⁰, J. S. Lange¹⁹¹, M. Laurenza¹⁹², K. Lautenbach¹⁹³, R. Lebourder¹⁹⁴, F. R. Le Diberder¹⁹⁵, M. J. Lee¹⁹⁶, D. Levit¹⁹⁷, P. M. Lewis¹⁹⁸, C. Li¹⁹⁹, L. K. Li²⁰⁰, Y. Li²⁰¹, Y. B. Li²⁰², J. Libby²⁰³, M. H. Liu²⁰⁴, Q. Y. Liu²⁰⁵, Z. Q. Liu²⁰⁶, D. Liventsev²⁰⁷, S. Longo²⁰⁸, T. Lueck²⁰⁹, C. Lyu²¹⁰, Y. Ma²¹¹, M. Maggiora²¹², S. P. Maharana²¹³, R. Maiti²¹⁴, S. Maity²¹⁵, G. Mancinelli²¹⁶, R. Manfredi²¹⁷, E. Manoni²¹⁸, M. Mantovano²¹⁹, D. Marcantonio²²⁰, S. Marcello²²¹, C. Marinas²²², L. Martel²²³, C. Martellini²²⁴, A. Martini²²⁵, T. Martinov²²⁶, L. Massaccesi²²⁷, M. Masuda²²⁸, K. Matsuoka²²⁹, D. Matvienko²³⁰, S. K. Maurya²³¹, J. A. McKenna²³², R. Mehta²³³, F. Meier²³⁴, M. Merola²³⁵, F. Metzner²³⁶, M. Milesi²³⁷, C. Miller²³⁸, M. Mirra²³⁹, K. Miyabayashi²⁴⁰, H. Miyake²⁴¹, R. Mizuk²⁴², G. B. Mohanty²⁴³, N. Molina-Gonzalez²⁴⁴, S. Mondal²⁴⁵, S. Moneta²⁴⁶, H.-G. Moser²⁴⁷, M. Mrvar²⁴⁸, R. Mussa²⁴⁹, I. Nakamura²⁵⁰, M. Nakao²⁵¹, Y. Nakazawa²⁵², A. Narimani Charan²⁵³, M. Naruki²⁵⁴, D. Narwal²⁵⁵, Z. Natkaniec²⁵⁶, A. Natchii²⁵⁷, L. Nayak²⁵⁸, M. Nayak²⁵⁹, G. Nazaryan²⁶⁰, C. Niebuhr²⁶¹, S. Nishida²⁶², S. Ogawa²⁶³, Y. Onishchuk²⁶⁴, H. Ono²⁶⁵, Y. Onuki²⁶⁶, P. Oskin²⁶⁷, F. Otani²⁶⁸, G. Pakhlova²⁶⁹, A. Panta²⁷⁰, S. Pardi²⁷¹, K. Parham²⁷², H. Park²⁷³, S.-H. Park²⁷⁴, B. Paschen²⁷⁵, A. Passeri²⁷⁶, S. Patra²⁷⁷, S. Paul²⁷⁸, T. K. Pedlar²⁷⁹, R. Peschke²⁸⁰, R. Pestotnik²⁸¹, M. Piccolo²⁸², L. E. Piilonen²⁸³, G. Pinna Angioni²⁸⁴, P. L. M. Podesta-Lerma²⁸⁵, T. Podobnik²⁸⁶, S. Pokharel²⁸⁷, C. Praz²⁸⁸, S. Prell²⁸⁹, E. Principe²⁹⁰, M. T. Prim²⁹¹, H. Purwar²⁹², P. Rados²⁹³, G. Raeuber²⁹⁴, S. Raiz²⁹⁵, N. Rauls²⁹⁶, M. Reif²⁹⁷, S. Reiter²⁹⁸, M. Remnev²⁹⁹, I. Ripp-Baudot³⁰⁰, G. Rizzo³⁰¹, S. H. Robertson³⁰², M. Roehrken³⁰³, J. M. Roney³⁰⁴, A. Rostomyan³⁰⁵, N. Rout³⁰⁶, G. Russo³⁰⁷, D. A. Sanders³⁰⁸, S. Sandilya³⁰⁹, L. Santelj³¹⁰, Y. Sato³¹¹, V. Savinov³¹², B. Scavino³¹³, C. Schmitt³¹⁴, C. Schwanda³¹⁵, M. Schwickardi³¹⁶, Y. Seino³¹⁷, A. Selce³¹⁸, K. Senyo³¹⁹, J. Serrano³²⁰, M. E. Sevier³²¹, C. Sfienti³²², W. Shan³²³, C. P. Shen³²⁴, X. D. Shi³²⁵, T. Shillington³²⁶, T. Shimasaki³²⁷, J.-G. Shiu³²⁸, D. Shtol³²⁹, A. Sibidanov³³⁰, F. Simon³³¹, J. B. Singh³³², J. Skorupa³³³, R. J. Sobie³³⁴, M. Sobotzik³³⁵, A. Soffer³³⁶, A. Sokolov³³⁷, E. Solovieva³³⁸, S. Spataro³³⁹, B. Spruck³⁴⁰, M. Starič³⁴¹, P. Stavroulakis³⁴², S. Stefkova³⁴³, R. Stroili³⁴⁴, M. Sumihama³⁴⁵, K. Sumisawa³⁴⁶, W. Sutcliffe³⁴⁷, H. Svidras³⁴⁸, M. Takizawa³⁴⁹, U. Tamponi³⁵⁰, S. Tanaka³⁵¹, K. Tanida³⁵², F. Tenchini³⁵³, O. Tittel³⁵⁴, R. Tiwary³⁵⁵, D. Tonelli³⁵⁶, E. Torassa³⁵⁷, K. Trabelsi³⁵⁸, I. Tsaklidis³⁵⁹, M. Uchida³⁶⁰, I. Ueda³⁶¹, K. Unger³⁶², Y. Unno³⁶³, K. Uno³⁶⁴, S. Uno³⁶⁵, P. Urquijo³⁶⁶, Y. Ushiroda³⁶⁷, S. E. Vahsen³⁶⁸, R. van Tonder³⁶⁹, K. E. Varvell³⁷⁰, M. Veronesi³⁷¹, A. Vinokurova³⁷², V. S. Vismaya³⁷³, L. Vitale³⁷⁴, V. Vobbiliseti³⁷⁵, R. Volpe³⁷⁶, B. Wach³⁷⁷, M. Wakai³⁷⁸, S. Wallner³⁷⁹, E. Wang³⁸⁰, M.-Z. Wang³⁸¹, X. L. Wang³⁸², Z. Wang³⁸³, A. Warburton³⁸⁴, S. Watanuki³⁸⁵, C. Wessel³⁸⁶, E. Won³⁸⁷, X. P. Xu³⁸⁸, B. D. Yabsley³⁸⁹, S. Yamada³⁹⁰, W. Yan³⁹¹, S. B. Yang³⁹², J. Yelton³⁹³, J. H. Yin³⁹⁴, K. Yoshihara³⁹⁵, C. Z. Yuan³⁹⁶, Y. Yusa³⁹⁷, L. Zani³⁹⁸, B. Zhang³⁹⁹, V. Zhilich⁴⁰⁰, Q. D. Zhou⁴⁰¹, X. Y. Zhou⁴⁰², and V. I. Zhukova⁴⁰³

(Belle II Collaboration)



(Received 7 March 2024; accepted 29 April 2024; published 14 June 2024)

We report on a search for a resonance X decaying to a pair of muons in $e^+e^- \rightarrow \mu^+\mu^-X$ events in the $0.212\text{--}9.000\text{ GeV}/c^2$ mass range, using 178 fb^{-1} of data collected by the Belle II experiment at the SuperKEKB collider at a center of mass energy of 10.58 GeV . The analysis probes two different models of X beyond the standard model: a Z' vector boson in the $L_\mu - L_\tau$ model and a muonphilic scalar. We observe no evidence for a signal and set exclusion limits at the 90% confidence level on the products of cross section and branching fraction for these processes, ranging from 0.046 fb to 0.97 fb for the $L_\mu - L_\tau$ model and from 0.055 fb to 1.3 fb for the muonphilic scalar model. For masses below $6\text{ GeV}/c^2$, the corresponding constraints on the couplings of these processes to the standard model range from 0.0008 to 0.039 for the $L_\mu - L_\tau$ model and from 0.0018 to 0.040 for the muonphilic scalar model. These are the first constraints on the muonphilic scalar from a dedicated search.

DOI: [10.1103/PhysRevD.109.112015](https://doi.org/10.1103/PhysRevD.109.112015)

I. INTRODUCTION

The standard model (SM) of particle physics is a highly predictive theoretical framework describing fundamental particles and their interactions. Despite its success, the SM is known to provide an incomplete description of nature. For example, it does not address the phenomenology related to dark matter, such as the observed relic density [1]. In addition, some experimental observations show inconsistencies with the SM. Prominent examples include the longstanding difference between the measured and the expected value of the muon anomalous magnetic-moment $(g-2)_\mu$ [2–4], possibly reduced by expectations based on lattice calculations [5], and the tensions in flavor observables reported by the BABAR, Belle, and LHCb experiments [6–8]. Some of these observations can be explained with the introduction of additional interactions, possibly lepton-universality-violating, mediated by non-SM neutral bosons [9–11]. Examples include the $L_\mu - L_\tau$ extension of the SM and a muonphilic scalar model.

The $L_\mu - L_\tau$ extension of the SM [12–14] gauges the difference between the muon and the τ -lepton numbers, giving rise to a new massive, neutral vector boson, the Z' . Among the SM particles, this particle couples only to μ , τ , ν_μ , and ν_τ , with a coupling constant g' . The Z' could also mediate interactions between SM and dark matter.

The muonphilic scalar S is primarily proposed as a solution for the $(g-2)_\mu$ anomaly [15–18]. This particle couples exclusively to muons through a Yukawa-like interaction, which is not gauge-invariant under the SM gauge symmetry and may arise from a high-dimension operator term at a mass scale beyond the SM. In contrast to

the $L_\mu - L_\tau$ model, the muonphilic scalar model needs a high-energy completion.

Searches for a Z' decaying to muons have been reported by the BABAR [19], Belle [20], and CMS [21] Collaborations. An invisibly decaying Z' has been searched for by the Belle II [22,23] and NA64- e [24] experiments. The Belle II experiment also searched recently for a Z' decaying to $\tau^+\tau^-$ [25]. Constraints on the existence of a muonphilic scalar have been obtained by reinterpretations of Z' searches into muons [18]. However, important experimental details may be unaccounted for in these reinterpretation studies, including the significantly different kinematic properties of the signal and the corresponding variation of the efficiency.

Here we report a search for the process $e^+e^- \rightarrow \mu^+\mu^-X$, with $X \rightarrow \mu^+\mu^-$, where X indicates Z' or S . The signal signature is a narrow enhancement in the mass distribution of oppositely charged muons $M(\mu\mu)$ in $e^+e^- \rightarrow \mu^+\mu^-\mu^+\mu^-$ events. We use data collected by the Belle II experiment at a center-of-mass (c.m.) energy \sqrt{s} corresponding to the mass of the $\Upsilon(4S)$ resonance. The $L_\mu - L_\tau$ model is used as a benchmark to develop the analysis; we then apply the same selections to the muonphilic scalar model and evaluate the performance. In both models, the X particle is at leading order emitted as final-state radiation (FSR) from one of the muons, as shown in Fig. 1. For the range of couplings explored in this study, the lifetime of X is negligible compared to the experimental resolution. The analysis techniques are optimized using simulated events prior to examining data.

We select events with exactly four charged particles with zero net charge, where at least three are identified as muons, with an invariant mass $M(4\mu)$ close to \sqrt{s}/c^2 , and with negligible detected energy in addition to that associated to the charged particles. The dominant, non-peaking background is the SM $e^+e^- \rightarrow \mu^+\mu^-\mu^+\mu^-$ process, whose main production diagrams are shown in Fig. 2. The analysis uses kinematic variables combined with a multivariate technique to enhance the signal-to-background ratio. A kinematic fit

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

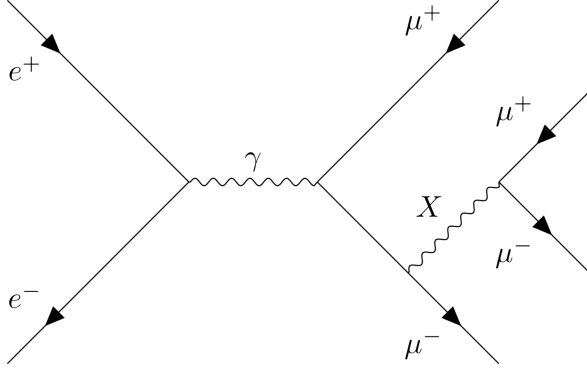


FIG. 1. Leading-order Feynman diagram for the process $e^+e^- \rightarrow \mu^+\mu^-X, X \rightarrow \mu^+\mu^-$.

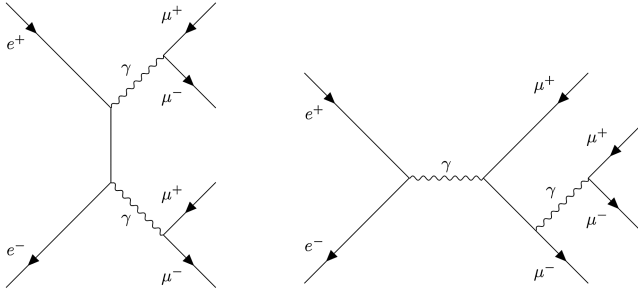


FIG. 2. Leading-order Feynman diagrams for the two main contributions to the $e^+e^- \rightarrow \mu^+\mu^-\mu^+\mu^-$ SM background: double photon conversion (left) and annihilation (right).

improves the dimuon mass resolution. The signal yield is extracted through a series of fits to the $M(\mu\mu)$ distribution, which allows an estimate of the background directly from data.

The paper is organized as follows. In Sec. II we briefly describe the Belle II experiment. In Sec. III we report the datasets and the simulation used. In Sec. IV we present the event selections. In Sec. V we describe the signal modeling and the fit technique to extract the signal. In Sec. VI we discuss the systematic uncertainties. In Sec. VII we describe and discuss the results. Section VIII summarizes our conclusions.

II. THE BELLE II EXPERIMENT

The Belle II detector [26,27] consists of several subdetectors arranged in a cylindrical structure around the e^+e^- interaction point. The longitudinal direction, the transverse plane, and the polar angle θ are defined with respect to the detector's cylindrical axis in the direction of the electron beam.

Subdetectors relevant for this analysis are briefly described here in order from innermost out; a full description of the detector is given in Refs. [26,27]. The innermost subdetector is the vertex detector, which consists of two inner layers of silicon pixels and four outer layers of silicon

strips. The second pixel layer was only partially installed for the data sample we analyze, covering one sixth of the azimuthal angle. The main tracking subdetector is a large helium-based small-cell drift chamber. The relative charged-particle transverse momentum resolution, $\frac{\Delta p_T}{p_T}$, is typically $0.1\% p_T \oplus 0.3\%$, with p_T expressed in GeV/ c . Outside of the drift chamber, time-of-propagation and aerogel ring-imaging Cherenkov detectors provide charged-particle identification in the barrel and forward end cap region, respectively. An electromagnetic calorimeter consists of a barrel and two end caps made of CsI(Tl) crystals: it reconstructs photons and identifies electrons. A superconducting solenoid, situated outside of the calorimeter, provides a 1.5 T magnetic field. A K_L^0 and muon subdetector (KLM) is made of iron plates, which serve as a magnetic flux-return yoke, alternated with resistive-plate chambers and plastic scintillators in the barrel and with plastic scintillators only in the end caps. In the following, quantities are defined in the laboratory frame unless specified otherwise.

III. DATA AND SIMULATION

We use a sample of e^+e^- collisions produced at c.m. energy $\sqrt{s} = 10.58$ GeV in 2020–2021 by the SuperKEKB asymmetric-energy collider [28] at KEK. The data, recorded by the Belle II detector, correspond to an integrated luminosity of 178 fb^{-1} [29].

Simulated signal $e^+e^- \rightarrow \mu^+\mu^-Z'$ with $Z' \rightarrow \mu^+\mu^-$ and $e^+e^- \rightarrow \mu^+\mu^-S$ with $S \rightarrow \mu^+\mu^-$ events are generated using MadGraph5_aMC@NLO [30] with initial-state radiation (ISR) included [31]. Two independent sets of Z' events are produced, with Z' masses, $m_{Z'}$, ranging from $0.212 \text{ GeV}/c^2$ to $10 \text{ GeV}/c^2$ in steps of $250 \text{ MeV}/c^2$, to estimate efficiencies, define selection requirements, and develop the fit strategy, and in steps of $5 \text{ MeV}/c^2$, exclusively dedicated to the training of the multivariate analysis. Samples of S events are generated in $40 \text{ MeV}/c^2$ steps for m_S masses between $0.212 \text{ GeV}/c^2$ and $1 \text{ GeV}/c^2$ and in $250 \text{ MeV}/c^2$ steps from $1 \text{ GeV}/c^2$ to $10 \text{ GeV}/c^2$.

Background processes are simulated using the following generators: $e^+e^- \rightarrow \mu^+\mu^-\mu^+\mu^-$, $e^+e^- \rightarrow e^+e^-\mu^+\mu^-$, $e^+e^- \rightarrow e^+e^-e^+e^-$, $e^+e^- \rightarrow \mu^+\mu^-\tau^+\tau^-$ and $e^+e^- \rightarrow e^+e^-\tau^+\tau^-$, with AAFH [32]; $e^+e^- \rightarrow \mu^+\mu^-(\gamma)$ with KKMC [33]; $e^+e^- \rightarrow \tau^+\tau^-(\gamma)$ with KKMC interfaced with TAUOLA [34]; $e^+e^- \rightarrow e^+e^-\pi^+\pi^-$ with TREPS [35]; $e^+e^- \rightarrow \pi^+\pi^-(\gamma)$ with PHOKHARA [36]; $e^+e^- \rightarrow e^+e^-(\gamma)$ with BabaYaga@NLO [37]; $e^+e^- \rightarrow u\bar{u}, d\bar{d}, s\bar{s}, c\bar{c}$ with KKMC interfaced with PYTHIA8 [38] and EvtGen [39] and $e^+e^- \rightarrow B^0\bar{B}^0$ and $e^+e^- \rightarrow B^+B^-$ with EvtGen interfaced with PYTHIA8. Electromagnetic FSR is simulated with PHOTOS [40,41] for processes generated with EvtGen. The AAFH generator, used for the four-lepton processes, including the dominant $e^+e^- \rightarrow \mu^+\mu^-\mu^+\mu^-$ background, does not simulate ISR effects. This is a source of disagreement

between data and simulation. Other sources of nonsimulated backgrounds include $e^+e^- \rightarrow \mu^+\mu^-\pi^+\pi^-$ and more generally $e^+e^- \rightarrow \mu^+\mu^-h$ and $e^+e^- \rightarrow \pi^+\pi^-h$, where h is typically a low-mass hadronic system; $e^+e^- \rightarrow J/\psi\pi^+\pi^-$ with $J/\psi \rightarrow \mu^+\mu^-$; $e^+e^- \rightarrow \psi(2S)\gamma$ with $\psi(2S) \rightarrow J/\psi\pi^+\pi^-$ and $J/\psi \rightarrow \mu^+\mu^-$; and $e^+e^- \rightarrow \Upsilon(nS)\pi^+\pi^-$ with $\Upsilon(nS) \rightarrow \mu^+\mu^-$ and $n = 1, 2, 3$.

The detector geometry and interactions of final-state particles with detector material are simulated using Geant4 [42] and the Belle II software [43,44].

IV. SELECTIONS

The selection requirements are divided into four categories: trigger, particle identification, candidate selections, and final background suppression.

A. Trigger selections

We filter events selected by the logical OR of a three-track trigger and a single-muon trigger. The efficiency of both triggers is measured using a reference calorimeter-only trigger, which requires a total energy deposit above 1 GeV in the polar angle region $22^\circ < \theta < 128^\circ$. We require a single electron of sufficient energy to activate the calorimeter trigger. The three-track trigger requires the presence of at least three tracks with $37^\circ < \theta < 120^\circ$. The efficiency of this trigger is measured in four-track events containing at least two pions and one electron and depends on the transverse momenta p_T of the two charged particles with lowest transverse momenta, reaching a plateau close to 100% for p_T above 0.5 GeV/ c . The single-muon trigger is based on the association of hits in the barrel KLM with geometrically matched tracks extrapolated from the inner tracker. The efficiency of this trigger is measured in a sample of two-track events with one electron and one muon, mostly from the $e^+e^- \rightarrow \tau^+\tau^-$ process, reaching a plateau of about 90% in the polar angle range $51^\circ < \theta < 117^\circ$. The efficiency for events with multiple muons is computed using the single-muon efficiency assuming no correlation. The overall trigger efficiency is 91% for $m_{Z'}$ close to the dimuon mass, increases smoothly to a plateau close to 99% in the mass range 2.5–8.5 GeV/ c^2 , and then drops to 89% at 10 GeV/ c^2 . It is slightly higher, 95%, for low masses in the S case, due to the harder spectrum of the muonphilic scalar (see Sec. IV E).

B. Particle identification

The identification of muons relies mostly on charged-particle penetration in the KLM for momenta larger than 0.7 GeV/ c and on information from the drift chamber and the calorimeter otherwise. The selection retains 93%–99% of the muons, and rejects 80%–97% of the pions, depending on their momenta. Electrons are identified mostly by comparing measured momenta to the energies of the associated calorimeter deposits. Photons are reconstructed from calorimeter

energy deposits greater than 100 MeV that are not associated with any track. Details of particle reconstruction and identification algorithms are given in Refs. [27,45].

C. Candidate selections

We require that events have exactly four charged particles with zero net charge and invariant mass $M(4\mu)$ between 10 GeV/ c^2 and 11 GeV/ c^2 . To suppress backgrounds from misreconstructed and single-beam induced tracks, the transverse and longitudinal projections of the distance of closest approach to the interaction point of the tracks must be smaller than 0.5 cm and 2.0 cm, respectively. At least three of the tracks must be identified as muons. This requirement provides better performance than requiring four identified muons or a pair of same-sign muons. It rejects almost all backgrounds other than $e^+e^- \rightarrow \mu^+\mu^-\mu^+\mu^-$, while retaining good efficiency for signal.

In the low dimuon-mass region below 1 GeV/ c^2 , there are residual backgrounds from $e^+e^- \rightarrow \mu^+\mu^-\gamma$, in which the photon converts to an electron-positron pair, and $e^+e^- \rightarrow e^+e^-\mu^+\mu^-$ events. Some of these electrons that are misidentified as muons have low momenta, and thus do not reach the KLM. The remaining electrons leave signals in the KLM at the gap between the barrel and end cap or in the gaps between adjacent modules. In this mass region, we therefore require that no track be identified as an electron.

To suppress radiative backgrounds and, in general, backgrounds with neutral particles, we require that the total energy of all photons be less than 0.4 GeV.

We add a further requirement when $M(4\mu) < 10.4$ GeV/ c^2 , exploiting the correlation between the invariant mass and the initial state radiation. This additional selection requires the total energy of all photons to be less than the expected energy of a single irradiated photon, which depends linearly on $M(4\mu)$.

In addition, we reject events in which the angle in the c.m. frame between the momentum of the four-muon system and that of the system composed of all the photons is larger than 160° .

At this level of the analysis, there is no *a priori* attempt to select a single $\mu^+\mu^-$ pair as a candidate X decay. Each event includes four possible $\mu^+\mu^-$ candidates, each with a different dimuon mass $M(\mu\mu)$, causing some combinatorial background. For each $\mu^+\mu^-$ candidate, the pair of the two remaining muons is labeled as the “recoil” pair. We consider independently all the $\mu^+\mu^-$ candidates, each with its recoil muons.

The resulting candidate $M(\mu\mu)$ distribution is shown in Fig. 3. The average data-to-simulation yield ratio is 0.76, due to the lack of ISR in the AAFH four-muon generator, in agreement with the values previously reported by BABAR [19] and Belle [20]. The excess of the simulation over data in the mass region below 2 GeV/ c^2 is also due to an overestimate of the three-track-trigger efficiency for very low transverse-momentum tracks. Specifically, the

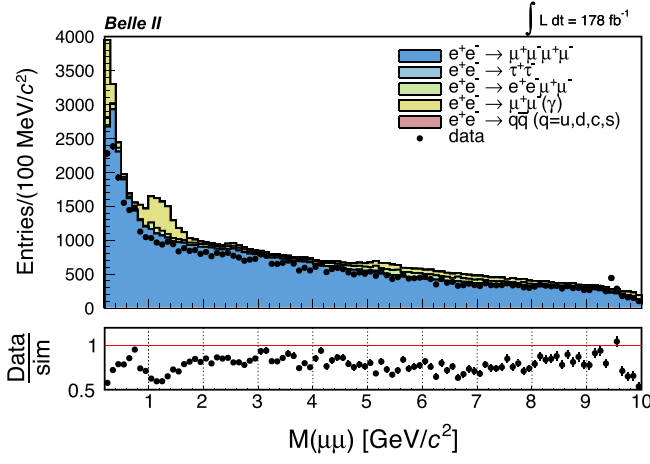


FIG. 3. Dimuon invariant-mass distribution in data and simulation for candidates passing all selections but the final background suppression. Contributions from the various simulated processes are stacked. The subpanel shows the data-to-simulation ratio.

enhancement in the range 1–2 GeV/c^2 originates from the process $e^+e^- \rightarrow \mu^+\mu^-\gamma$ with a near-beam-energy photon, followed by conversion of the photon into electron-positron pairs in detector material. These events are almost entirely removed by the final background suppression. Other visible features include the unsimulated contributions from the ρ , J/ψ , and $\Upsilon(1S)$ resonances.

D. Final background suppression

The final selection relies on a few distinctive features that allow the discrimination of signal from background: signal events include a $\mu^+\mu^-$ resonance, which can be seen both in the candidate muon pair and in the mass of the system recoiling against the two recoil muons; the signal is emitted through FSR from a muon (Fig. 1), while the dominant four-muon background proceeds through double-photon-conversion process (Fig. 2, left); and the double-photon-conversion process has a distinctive momentum

distribution. In the following, some of the relevant variables sensitive to these three classes of features are discussed: they are based both on the $\mu^+\mu^-$ candidate, where we search for signal, and on the recoil muons. For illustration, we show the case for a Z' signal with $m_{Z'} = 3 \text{ GeV}/c^2$ and for background, both with reconstructed candidate dimuon masses $2.75 < M(\mu\mu) < 3.25 \text{ GeV}/c^2$. The background in this mass region is dominated by the $e^+e^- \rightarrow \mu^+\mu^-\mu^+\mu^-$ process, see Fig. 3.

Magnitudes of the two candidate muon momenta, p_{μ^+} and p_{μ^-} , and their correlations are sensitive to the presence of a resonance (Fig. 4). Signal events cluster preferentially in the central part of the distribution, while background predominantly populates the extremes. A similar effect occurs for the momenta of the two recoil muons, $p_{\mu^+}^{\text{rec}}$ and $p_{\mu^-}^{\text{rec}}$ (Fig. 5), which provide instrumentally uncorrelated access to the same information, though with a different resolution. The cosine of the helicity angle of the candidate-muon pair $\cos \phi_{\text{hel}}$, defined as the angle between the momentum direction of the c.m. frame and the μ^- in the candidate-muon-pair frame, has a uniform distribution for a scalar or an unpolarized massive vector decaying to two fermions, but not for the background processes (Fig. 6). The slight departure from uniformity in the signal case is due to momentum resolution, which smears the determination of the boost to the muon-pair frame.

The double-photon-conversion process (Fig. 2, left) accounts for 80% of the four-muon background cross section. It also includes the case of off-shell photon emission (and subsequent dimuon production) from one of the initial-state electrons, ISR double-photon conversion, which contributes mainly in the low mass region. The annihilation process (Fig. 2, right) is very similar to the signal process and constitutes an nearly irreducible background: it accounts for 20% of the cross section for $M(\mu\mu) < 1 \text{ GeV}/c^2$ and for 10% above. Transverse projections of the candidate-muon-pair momentum $p_{\mu\mu}$ on the direction of the recoil muon with minimum momentum,

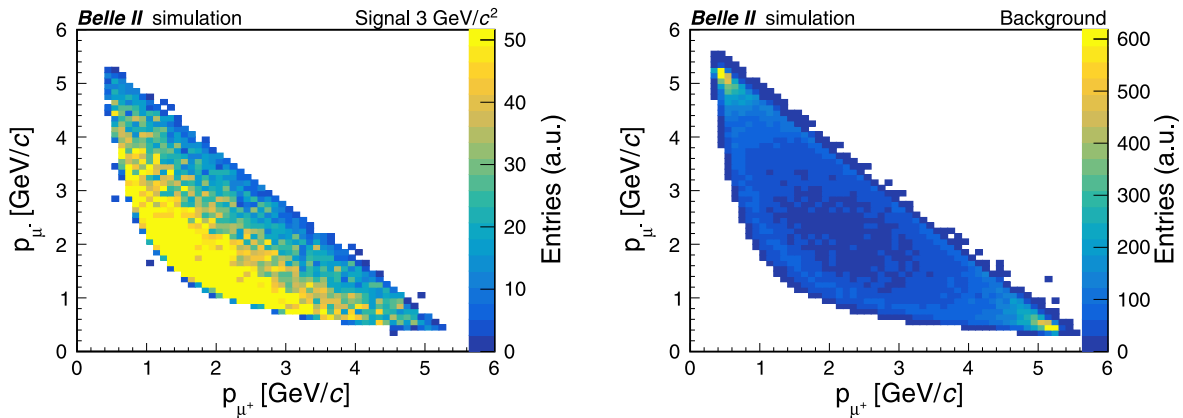


FIG. 4. Candidate- μ^+ momentum versus candidate- μ^- momentum for simulated signal (left) with $m_{Z'} = 3 \text{ GeV}/c^2$ and simulated background (right), for dimuon masses $2.75 < M(\mu\mu) < 3.25 \text{ GeV}/c^2$.

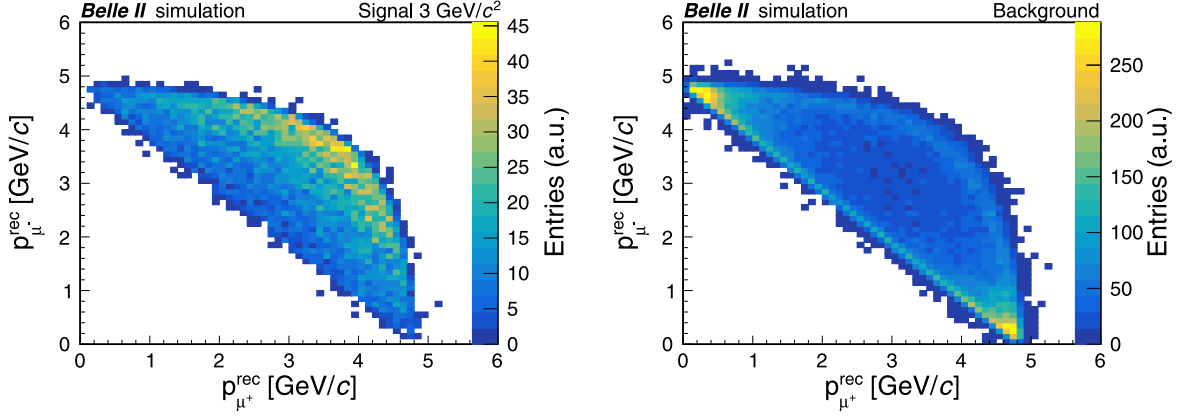


FIG. 5. Recoil- μ^+ momentum versus recoil- μ^- momentum for simulated signal (left) with $m_{Z'} = 3 \text{ GeV}/c^2$ and simulated background (right), for dimuon masses $2.75 < M(\mu\mu) < 3.25 \text{ GeV}/c^2$.

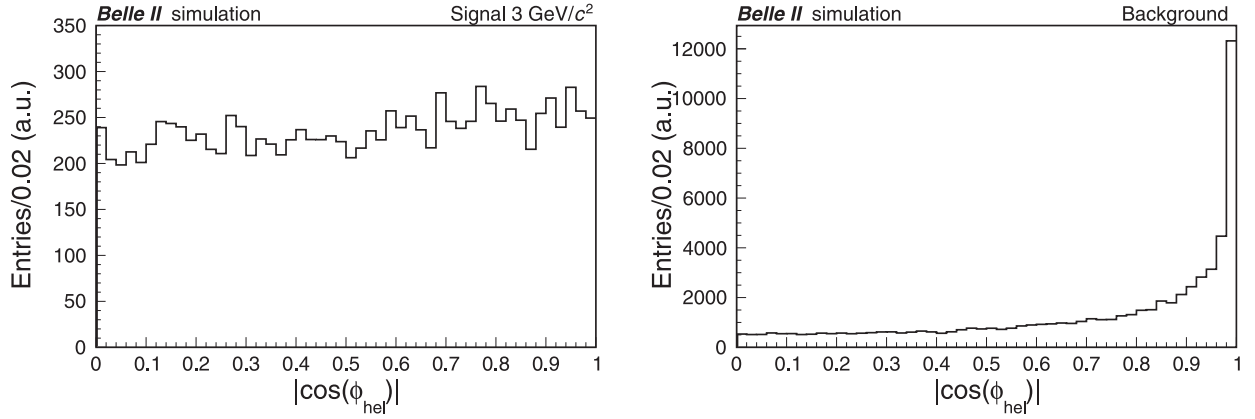


FIG. 6. Absolute value of the cosine of the helicity angle for simulated signal (left) with $m_{Z'} = 3 \text{ GeV}/c^2$ and simulated background (right), for dimuon masses $2.75 < M(\mu\mu) < 3.25 \text{ GeV}/c^2$.

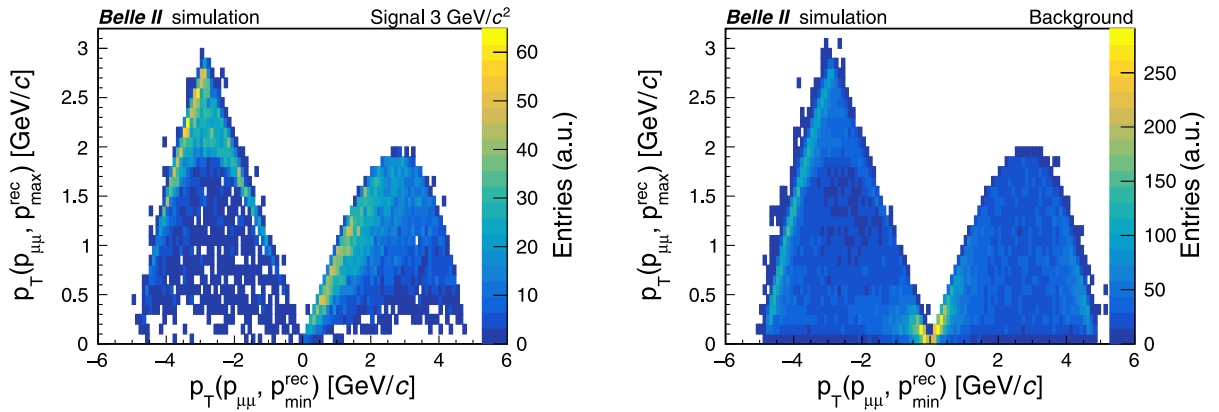


FIG. 7. Candidate-muon-pair transverse momentum with respect to the maximum momentum recoil-muon direction versus the candidate-muon-pair transverse momentum with respect to the minimum momentum recoil-muon direction (with the sign of the longitudinal projection) for simulated signal (left) with $m_{Z'} = 3 \text{ GeV}/c^2$ and simulated background (right), for dimuon masses $2.75 < M(\mu\mu) < 3.25 \text{ GeV}/c^2$.

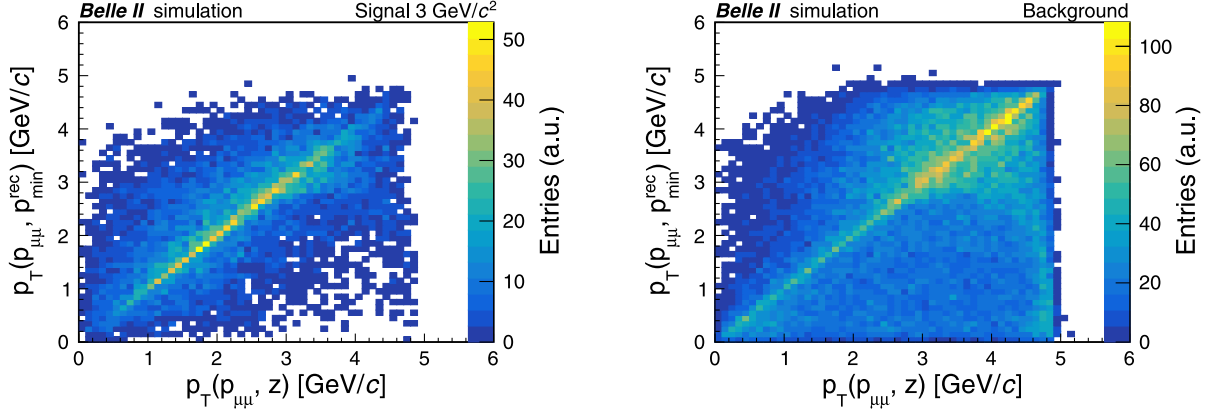


FIG. 8. Candidate-muon-pair transverse momentum with respect to the minimum-momentum recoil-muon direction versus candidate-muon-pair transverse momentum with respect to the beam direction, for simulated signal (left) with $m_{Z'} = 3 \text{ GeV}/c^2$ and simulated background (right), for dimuon masses $2.75 < M(\mu\mu) < 3.25 \text{ GeV}/c^2$.

$p_T(p_{\mu\mu}, p_{\min}^{\text{rec}})$, and on the direction of the recoil muon with maximum momentum, $p_T(p_{\mu\mu}, p_{\max}^{\text{rec}})$, are sensitive to FSR emission (Fig. 7). This is because, in case of signal, these are the transverse momenta of X with respect to the direction of the muon from which it was emitted, and with respect to the direction of the other muon. We assign to the transverse projection $p_T(p_{\mu\mu}, p_{\min}^{\text{rec}})$ the sign of the longitudinal projection, since this slightly increases the discriminating power. The transverse momentum of the candidate muon pair with respect to the z axis, $p_T(p_{\mu\mu}, z)$, which approximates the beam direction, is sensitive to the ISR double-photon conversion mechanism of emission because $p_T(p_{\mu\mu}, z)$ is the transverse momentum of the muon pair with respect to the initial-state-electron direction. This variable is shown in Fig. 8 in a two-dimensional distribution versus $p_T(p_{\mu\mu}, p_{\min}^{\text{rec}})$ to illustrate the correlation between variables sensitive to ISR and FSR, respectively.

The double photon conversion process produces two muon pairs from two off-shell photons. The dominant background at a mass m_0 is produced when one pair has

$M(\mu\mu)$ near m_0 and the other pair has a mass at the lowest possible value above $2m_\mu$. In these cases, the c.m. momentum p_0 of the two pairs can be analytically calculated. In $e^+e^- \rightarrow \mu^+\mu^-\mu^+\mu^-$ background events the dimuon c.m. momentum $p_{\mu\mu}$ peaks at p_0 , in contrast to the signal, at least for two of the dimuon candidates. This difference is visible in Fig. 9.

We select sixteen discriminating variables: the magnitude of the candidate-muon-pair momentum $p_{\mu\mu}$; the absolute value of the cosine of the helicity angle in the candidate-muon-pair rest frame; the magnitudes of the candidate-single-muon momenta; the candidate-single-muon transverse momenta; the magnitudes of the recoil-single-muon momenta; the recoil-single-muon transverse momenta; $p_T(p_{\mu\mu}, p_{\min}^{\text{rec}})$ and $p_T(p_{\mu\mu}, p_{\max}^{\text{rec}})$; the correlation of $p_T(p_{\mu\mu}, p_{\min}^{\text{rec}})$ with $p_T(p_{\mu\mu}, z)$; and the transverse projections of the recoil-muon-pair momentum on the directions of the momenta of the candidate muons with minimum and maximum momentum. All variables other than the helicity angle are defined in the c.m. frame.

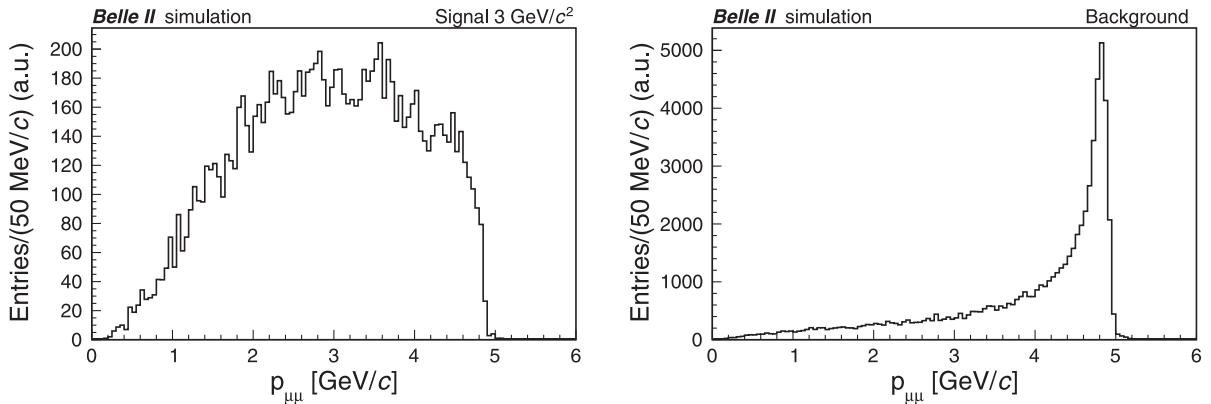


FIG. 9. Candidate-muon-pair momentum $p_{\mu\mu}$ for signal (left) with $m_{Z'} = 3 \text{ GeV}/c^2$ and background (right), for dimuon masses $2.75 < M(\mu\mu) < 3.25 \text{ GeV}/c^2$.

We use multilayer perceptron (MLP) artificial neural networks [46] with 16 input neurons, fed with the discriminant variables, and with one output neuron. The MLPs are developed using simulated Z' and simulated background events. To improve performance, we use five separate MLPs in different $M(\mu\mu)$ intervals, which we refer to as MLP ranges: 0.21–1.00 GeV/c^2 , 1.00–3.75 GeV/c^2 , 3.75–6.25 GeV/c^2 , 6.25–8.25 GeV/c^2 , and 8.25–10.00 GeV/c^2 . Within the MLP ranges, better performances are obtained if the dependence of the input variables on $m_{Z'}$ is reduced. This is achieved by scaling the momentum-dimensioned variables by p_0 , which is the maximum c.m. momentum of the two muon pairs. To ensure that MLPs are not biased to specific mass values, we use a training signal sample that has mass steps of 5 MeV/c^2 , so as to approximate a continuous distribution. For nearly all masses, the most discriminating variable is $p_{\mu\mu}$, followed by the correlation of p_{μ^+} and p_{μ^-} .

The selection applied on the MLP output is studied separately in each MLP range, by maximizing the figure of merit described in Ref. [47], and then expressed as a function of $M(\mu\mu)$ by interpolation. The background rejection factor achieved by the MLP selection varies from 2.5 to 14, with the best value around 5 GeV/c^2 . The resulting background is composed almost entirely of $e^+e^- \rightarrow \mu^+\mu^-\mu^+\mu^-$ events, with $e^+e^- \rightarrow \mu^+\mu^-(\gamma)$ and $e^+e^- \rightarrow e^+e^-\mu^+\mu^-$ processes contributing only below 1 GeV/c^2 . The MLP selection is applied separately to each of the four candidates per event, reducing the average candidate multiplicity per background event to 1.7. The average candidate multiplicity per signal event varies between 1.4 and 3, depending on the mass.

E. Efficiencies and dimuon spectrum

The efficiencies of the full selection for the $L_\mu - L_\tau$ and muonphilic scalar models are shown in Fig. 10. The S boson, due to angular momentum conservation, is produced through a p-wave process, and has a higher momentum spectrum than the Z' , which is produced via an s-wave process. For masses below 1 GeV/c^2 , this implies the presence of higher momentum muons in the case of the scalar, which are better identified and detected with higher efficiency. For masses above 1 GeV/c^2 , the muon identification efficiencies for S and Z' are similar, and the higher signal efficiencies for the Z' are due to the differences in the distributions of the momentum-dimensioned input variables, with the MLP optimized for the $L_\mu - L_\tau$ model. The signal efficiencies shown here are corrected for ISR. Although the signal generator includes ISR, it does not include the large-angle hard-radiation component that can produce photons in the acceptance, and thereby veto events. This effect is studied using $e^+e^- \rightarrow \mu^+\mu^-\gamma$ events, generated with KKMC that simulates ISR in a complete way. We require the dimuon mass in the range 10–11 GeV/c^2 , to

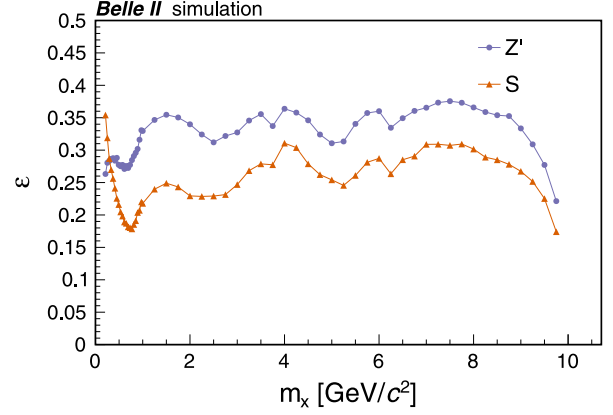


FIG. 10. Signal efficiency as a function of $m_{Z'}$ (purple dots) and m_S (orange triangles) masses after all selections are applied.

emulate the selection we apply on $M(4\mu)$, that intrinsically limits the maximum energy at which a photon can be radiated. Applying the selection on the photons (see Sec. IV) gives a relative reduction of 2.8% in efficiency. To improve the m_X resolution, a kinematic fit is applied requiring that the sum of the four-momenta of the muons be equal to the four-momentum of the c.m. system, thereby constraining the four-muon invariant mass to \sqrt{s}/c^2 . The resulting $M(\mu\mu)$ distribution is shown in Fig. 11. With the exception of the very low mass region, the data-to-simulation yield ratio is generally above one. This is because the MLPs perform worse on data, which naturally includes ISR, than on background simulation, which does not. This is not the case for the signal, which is simulated with the ISR contribution. Also visible in Fig. 11 are modulations originating from the five MLP ranges. Neither of these effects produce narrow peaking structures at the scale of the signal resolution, 2–5 MeV/c^2 (Sec. V). As in Fig. 3, contributions from the unsimulated ρ , J/ψ , and $\Upsilon(1S)$ resonances are visible.

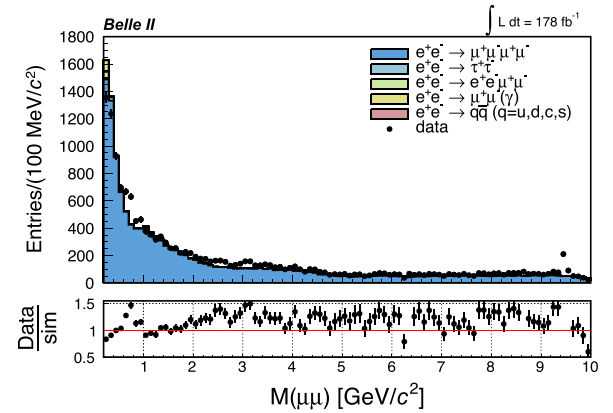


FIG. 11. Dimuon invariant-mass distribution in data and simulation for candidates passing all the selections. Contributions from the various simulated process are stacked. The subpanel shows the data-to-simulation ratio.

V. SIGNAL MODELING AND FIT

To search for the signal, we use the reduced dimuon mass $M_R \equiv \sqrt{M^2(\mu\mu) - 4m_\mu^2}$, which has smoother behavior than the dimuon mass near the kinematic threshold. The reduced-mass resolution is 2–2.5 MeV/ c^2 for $m_{Z'}$ below 1 GeV/ c^2 , increases smoothly to 5 MeV/ c^2 for $m_{Z'}$ around 5 GeV/ c^2 , then decreases to 2.5 MeV/ c^2 at 9 GeV/ c^2 .

The signal yields are obtained from a scan over the M_R spectrum through a series of unbinned maximum likelihood fits. The signal M_R distributions are parameterized from the simulation as sums of two Crystal Ball functions [48] sharing the same mean. The background is described with a quadratic function with coefficients as free parameters in the fit for masses below 1 GeV/ c^2 , and with a straight line above. Higher-order polynomials are investigated, but their corresponding fitted coefficients are compatible with zero over the full mass spectrum. The broad ρ contribution is accommodated by the quadratic fit.

The scan step-size is set equal to the mass resolution, which is sufficient to detect the presence of a X resonance regardless of its mass. The fit interval is 60 times the mass resolution, following an optimization study. A total of 2315 fits are performed, covering dimuon masses from 0.212 GeV/ c^2 to 9 GeV/ c^2 . If a fitting interval extends over two different MLP ranges, we use the MLP corresponding to the central mass. We exclude the dimuon mass interval 3.07–3.12 GeV/ c^2 , which corresponds to the J/ψ mass. The $\Upsilon(1S)$ peak is beyond the mass range of the search. The fit yields are scaled by 7% to account for a bias estimated in a study of the J/ψ in an $e^+e^-\mu^+\mu^-$ control sample, which obtains a width 25% larger than in simulated signals of that mass. Propagating this 25% degradation in resolution to all masses gives an average yield bias of 7%. This is also included as a systematic uncertainty (Sec. VI).

Signal yields from the fits are then converted into cross sections, after correcting for signal efficiency and luminosity.

VI. SYSTEMATIC UNCERTAINTIES

Several sources of systematic uncertainties affecting the cross-section determination are taken into account: these include signal efficiency, luminosity, and fit procedure.

Uncertainties due to the trigger efficiency in signal events are evaluated by propagating the uncertainties on the measured trigger efficiencies. They are 0.3% for most of the mass spectrum, increasing to 1.7% at low masses and 0.5% at high masses.

Uncertainties due to the tracking efficiency are estimated in $e^+e^- \rightarrow \tau^+\tau^-$ events in which one τ decays to a single charged particle and the other τ to three charged particles. The relative uncertainty on the signal efficiency is 3.6%.

Uncertainties due to the muon identification requirement are studied using $e^+e^- \rightarrow \mu^+\mu^-\gamma$, $e^+e^- \rightarrow e^+e^-\mu^+\mu^-$

events, and final states with a J/ψ . The relative uncertainty on the signal efficiency varies between 0.7% and 3%, depending on the X mass.

Beam backgrounds in the calorimeter can accidentally veto events due to the requirements on photons (Sec. IV C). The effect is studied by changing the level of beam backgrounds in the simulation and by varying the photon energy requirement (see Sec. IV) according to the calorimeter resolution. The relative uncertainty on the signal efficiency due to this source is estimated to be below 1%.

To evaluate uncertainties due to the data-to-simulation discrepancies in MLP selection efficiencies, we apply a tight selection on $M(4\mu)$ around \sqrt{s}/c^2 requiring it to be in the range 10.54–10.62 GeV/ c^2 . With this selection, data and background simulation are more directly comparable, because ISR and FSR effects are much less important. We compare MLP efficiencies, defined as the ratio of the number of events before and after the MLP selection, in data and simulation and assume that the uncertainties estimated in those signal-like conditions are representative of signal. We also assume that these uncertainties hold in the full $M(4\mu)$ interval 10–11 GeV/ c^2 for the signal, which is generated with ISR. The differences found in each MLP range vary between 1.1% and 8.1%, which are taken as estimates of the systematic uncertainties. To exclude potential bias from the presence of a signal, we check that these differences do not change if we exclude, in each MLP range and for each of the 2315 mass points, intervals ten times larger than the signal mass resolution around the test masses.

Uncertainties due to the interpolation of the signal efficiency between simulated points are estimated to be 3%, which is assigned as a relative uncertainty on the signal efficiency.

Uncertainties due to the fit procedure, in addition to that arising from mass resolution, are evaluated using a bootstrap technique [49]. A number of simulated signal events corresponding to the yield excluded at 90% confidence level are overlaid on simulated background and fitted for each Z' mass. The distribution of the difference between the overlaid and the fitted yields, divided by the fit uncertainty, shows a negligible average bias with a width that deviates from one by 4%, which is assigned as a relative uncertainty on the signal-yield determination. Additional uncertainties related to the fit procedure are those due to the mass resolution, discussed in Sec. V. An uncertainty of 7%, equal to the average yield bias, is included.

Systematic uncertainties from data-to-simulation differences in momentum resolution and beam-energy shift are found to be negligible, due to the kinematic fitting procedure. Finally, the integrated luminosity has a systematic uncertainty of 1% [29].

The uncertainties are summed in quadrature to give a total that ranges from 9.5% to 12.9% depending on the X

TABLE I. Systematic uncertainties affecting the cross-section determination.

Source	Uncertainty (%)
Trigger	0.3–1.7
Tracking	3.6
Particle identification	0.7–3
Beam background and calorimeter energy resolution	1
MLP selection	1.1–8.1
Efficiency interpolation	3
Fit bias	4
Mass resolution	7
Luminosity	1
Total	9.5–12.9

mass. The contributions to the systematic uncertainty are summarized in Table I. The systematic uncertainties are included as nuisance parameters with Gaussian constraints on the signal efficiency, with widths equal to the estimated systematic uncertainties.

VII. RESULTS

The significance of signal over background for each fit is evaluated as $\sqrt{2\log(\mathcal{L}/\mathcal{L}_0)}$, where \mathcal{L} and \mathcal{L}_0 are the likelihoods of the fits with and without signal. The largest local one-sided significance observed is 3.4σ at $M(\mu\mu) = 5.307 \text{ GeV}/c^2$, corresponding to a 1.6σ global significance after taking into account the look-elsewhere effect [50,51]. The corresponding fit is shown in Fig. 12. Three additional mass points have local significances that exceed 3σ . They are at $M(\mu\mu)$ masses of $1.939 \text{ GeV}/c^2$, $4.518 \text{ GeV}/c^2$, and $4.947 \text{ GeV}/c^2$, with global significances of 0.6σ , 1.2σ , and 1.1σ , respectively.

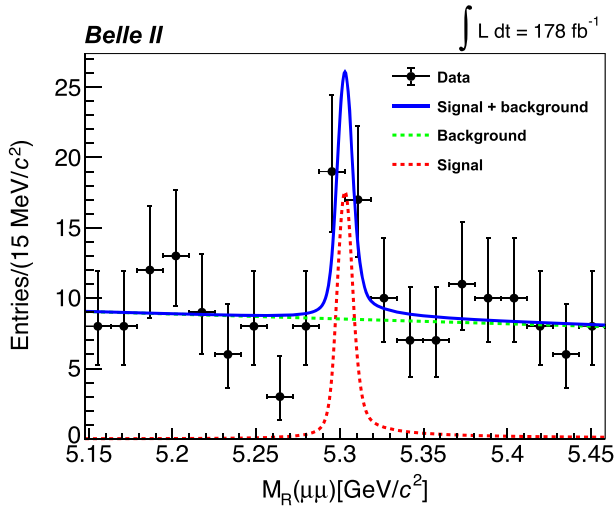


FIG. 12. Fit for a Z' mass hypothesis of $5.307 \text{ GeV}/c^2$, for which we obtain the maximum local significance of 3.4σ .

Since we do not observe any significant excess above the background, we derive 90% confidence level (CL) upper limits (UL) on the process cross sections $\sigma(e^+e^- \rightarrow \mu^+\mu^-X) \times \mathcal{B}(X \rightarrow \mu^+\mu^-)$ separately for Z' and S (Fig. 13), using the frequentist procedure CL_S [52]. The expected limits in Fig. 13 are the median limits from background-only simulated samples that use yields from fits to data. We obtain upper limits ranging from 0.046 fb to 0.97 fb for the $L_\mu - L_\tau$ model, and from 0.055 fb to 1.3 fb for the muonphilic scalar model. These upper limits are dominated by sample size, with systematic uncertainties worsening them on average by less than 1%.

The cross-section results are translated into upper limits on the coupling constant g' of the $L_\mu - L_\tau$ model and on the coupling constant g_S of the muonphilic scalar model (Fig. 14). For masses below $6 \text{ GeV}/c^2$, they range from 0.0008 to 0.039 for the $L_\mu - L_\tau$ model and from 0.0018 to 0.040 for the muonphilic-scalar model. These limits exclude the $L_\mu - L_\tau$ model and the muonphilic scalar model as explanations of the $(g-2)_\mu$ anomaly for $0.8 < m_{Z'} < 4.9 \text{ GeV}/c^2$ and $2.9 < m_S < 3.5 \text{ GeV}/c^2$, respectively.

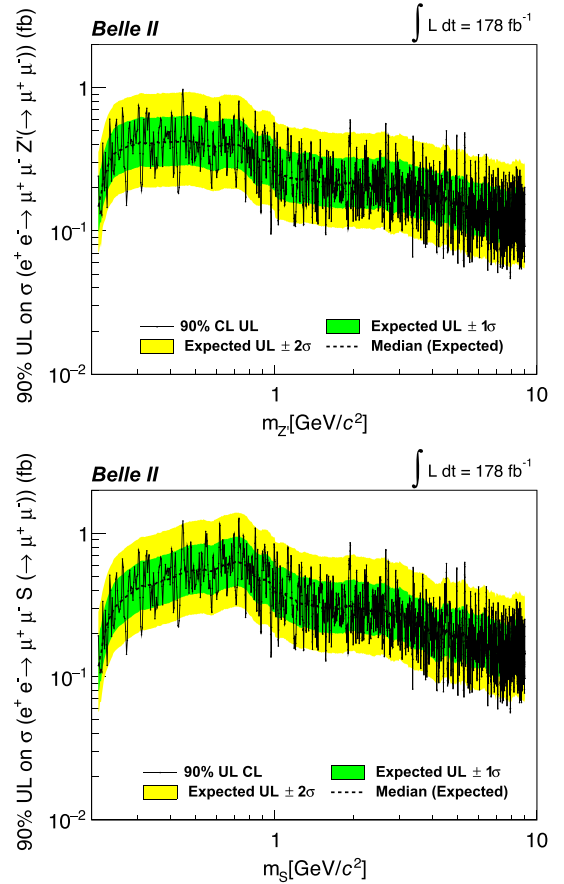


FIG. 13. Observed 90% confidence level upper limits and corresponding expected limits on the cross sections for the processes $e^+e^- \rightarrow \mu^+\mu^-X$ with $X \rightarrow \mu^+\mu^-$, $X = Z', S$, as functions of the Z' mass (top) and S mass (bottom).

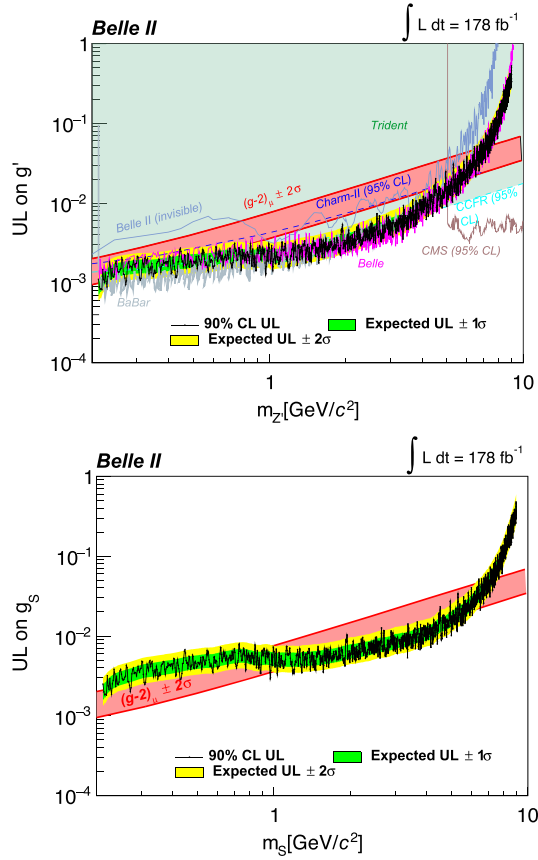


FIG. 14. Observed 90% CL upper limits and corresponding expected limits on (top) the $L_\mu - L_\tau$ model coupling g' and on (bottom) the muonphilic scalar model coupling g_S . Also shown in the top panel are constraints from Belle II [22,23] for invisible Z' decays, and from BABAR [19], Belle [20], and CMS [21] (95% CL) searches for Z' decays to muons, along with constraints (95% CL) derived from the trident production in neutrino experiments [53–55]. The red band in each panel shows the region that explains the muon anomalous magnetic moment $(g-2)_\mu \pm 2\sigma$.

Our constraints on g' are similar to those set by BABAR [19] for $m_{Z'}$ above 1 GeV/c^2 and to those set by Belle [20] on the full $m_{Z'}$ spectrum, both based on much larger integrated luminosities than ours. For the muonphilic scalar model, we do not show the constraints in Ref. [18], since they may not take into account all the experimental details affecting the signal efficiency, particularly those related to the higher momentum spectrum compared to the Z' . Numerical results of the cross-section and the coupling constant of the Z' and the S are available in the Supplemental Material [56].

VIII. CONCLUSION

We search for the process $e^+e^- \rightarrow \mu^+\mu^-X$ with $X \rightarrow \mu^+\mu^-$, $X = Z', S$ in a data sample of electron-positron collisions at 10.58 GeV collected by Belle II at SuperKEKB in 2020 and 2021, corresponding to an

integrated luminosity of 178 fb^{-1} . We find no significant excess above the background. We set upper limits on the cross sections for masses between 0.212 GeV/c^2 and 9 GeV/c^2 , ranging from 0.046 fb to 0.97 fb for the $L_\mu - L_\tau$ model, and from 0.055 fb to 1.3 fb for the muonphilic scalar model. We derive exclusion limits on the couplings for the two different models. For masses below 6 GeV/c^2 , they range from 0.0008 to 0.039 for the $L_\mu - L_\tau$ model and from 0.0018 to 0.040 for the muonphilic-scalar model. These limits exclude the $L_\mu - L_\tau$ model and the muonphilic scalar model as explanations of the $(g-2)_\mu$ anomaly for $0.8 < m_{Z'} < 4.9 \text{ GeV}/c^2$ and $2.9 < m_S < 3.5 \text{ GeV}/c^2$, respectively. These are the first results for the muonphilic scalar model based on a realistic evaluation of the signal efficiency that takes into account all the experimental details.

ACKNOWLEDGMENTS

This work, based on data collected using the Belle II detector, which was built and commissioned prior to March 2019, was supported by Higher Education and Science Committee of the Republic of Armenia Grant No. 23LCG-1C011; Australian Research Council and Research Grants No. DP200101792, No. DP210101900, No. DP210102831, No. DE220100462, No. LE210100098, and No. LE230100085; Austrian Federal Ministry of Education, Science and Research, Austrian Science Fund No. P 34529, No. J 4731, No. J 4625, and No. M 3153, and Horizon 2020 ERC Starting Grant No. 947006 “InterLeptons”; Natural Sciences and Engineering Research Council of Canada, Compute Canada and CANARIE; National Key R&D Program of China under Contract No. 2022YFA1601903, National Natural Science Foundation of China and Research Grants No. 11575017, No. 11761141009, No. 11705209, No. 11975076, No. 12135005, No. 12150004, No. 12161141008, and No. 12175041, and Shandong Provincial Natural Science Foundation Project ZR2022JQ02; the Czech Science Foundation Grant No. 22-18469S; European Research Council, Seventh Framework PIEF-GA-2013-622527, Horizon 2020 ERC-Advanced Grants No. 267104 and No. 884719, Horizon 2020 ERC-Consolidator Grant No. 819127, Horizon 2020 Marie Skłodowska-Curie Grant Agreement No. 700525 “NIOBE” and No. 101026516, and Horizon 2020 Marie Skłodowska-Curie RISE project JENNIFER2 Grant Agreement No. 822070 (European grants); L’Institut National de Physique Nucléaire et de Physique des Particules (IN2P3) du CNRS and L’Agence Nationale de la Recherche (ANR) under Grant No. ANR-21-CE31-0009 (France); BMBF, DFG, HGF, MPG, and AvH Foundation (Germany); Department of Atomic Energy under Project Identification No. RTI 4002, Department of Science and

Technology, and UPES SEED funding programs No. UPES/R&D-SEED-INFRA/17052023/01 and No. UPES/R&D-SOE/20062022/06 (India); Israel Science Foundation Grant No. 2476/17, U.S.-Israel Binational Science Foundation Grant No. 2016113, and Israel Ministry of Science Grant No. 3-16543; Istituto Nazionale di Fisica Nucleare and the Research Grants BELLE2; Japan Society for the Promotion of Science, Grant-in-Aid for Scientific Research Grants No. 16H03968, No. 16H03993, No. 16H06492, No. 16K05323, No. 17H01133, No. 17H05405, No. 18K03621, No. 18H03710, No. 18H05226, No. 19H00682, No. 20H05850, No. 20H05858, No. 22H00144, No. 22K14056, No. 22K21347, No. 23H05433, No. 26220706, and No. 26400255, and the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) of Japan; National Research Foundation (NRF) of Korea Grants No. 2016R1D1A1B-02012900, No. 2018R1A2B3003643, No. 2018R1A-6A1A06024970, No. 2019R1I1A3A01058933, No. 2021R1A6A1A03043957, No. 2021R1F1A1060423, No. 2021R1F1A1064008, No. 2022R1A2C1003993, and No. RS-2022-00197659, Radiation Science Research Institute, Foreign Large-Size Research Facility Application Supporting project, the Global Science Experimental Data Hub Center of the Korea Institute of Science and Technology Information and KREONET/GLORIAD; Universiti Malaya RU grant, Akademi Sains Malaysia, and Ministry of Education Malaysia; Frontiers of Science Program Contracts No. FOINS-296, No. CB-221329, No. CB-236394, No. CB-254409, and No. CB-180023, and SEP-CINVESTAV Research Grant No. 237 (Mexico); the Polish Ministry of Science and Higher Education and the National Science Center; the Ministry of Science and Higher Education of the Russian Federation

and the HSE University Basic Research Program, Moscow; University of Tabuk Research Grants No. S-0256-1438 and No. S-0280-1439 (Saudi Arabia); Slovenian Research Agency and Research Grants No. J1-9124 and No. P1-0135; Agencia Estatal de Investigacion, Spain Grant No. RYC2020-029875-I and Generalitat Valenciana, Spain Grant No. CIDEGENT/2018/020; The Knut and Alice Wallenberg Foundation (Sweden), Contracts No. 2021.0174 and No. 2021.0299; National Science and Technology Council, and Ministry of Education (Taiwan); Thailand Center of Excellence in Physics; TUBITAK ULAKBIM (Turkey); National Research Foundation of Ukraine, Project No. 2020.02/0257, and Ministry of Education and Science of Ukraine; the U.S. National Science Foundation and Research Grants No. PHY-1913789 and No. PHY-2111604, and the U.S. Department of Energy and Research Awards No. DE-AC06-76RLO1830, No. DE-SC0007983, No. DE-SC0009824, No. DE-SC0009973, No. DE-SC0010007, No. DE-SC0010073, No. DE-SC0010118, No. DE-SC0010504, No. DE-SC0011784, No. DE-SC0012704, No. DE-SC0019230, No. DE-SC0021274, No. DE-SC0021616, No. DE-SC0022350, No. DE-SC0023470; and the Vietnam Academy of Science and Technology (VAST) under Grants No. NVCC.05.12/22-23 and No. DL0000.02/24-25. These acknowledgements are not to be interpreted as an endorsement of any statement made by any of our institutes, funding agencies, governments, or their representatives. We thank the SuperKEKB team for delivering high-luminosity collisions; the KEK cryogenics group for the efficient operation of the detector solenoid magnet; the KEK Computer Research Center for on-site computing support; the NII for SINET6 network support; and the raw-data centers hosted by BNL, DESY, GridKa, IN2P3, INFN, and the University of Victoria.

-
- [1] G. Bertone, D. Hooper, and J. Silk, *Phys. Rep.* **405**, 279 (2005).
 - [2] G. W. Bennett *et al.* (Muon $g - 2$ Collaboration), *Phys. Rev. D* **73**, 072003 (2006).
 - [3] T. Aoyama *et al.*, *Phys. Rep.* **887**, 1 (2020).
 - [4] D. P. Aguillard *et al.* (Muon $g - 2$ Collaboration), *Phys. Rev. Lett.* **131**, 161802 (2023).
 - [5] S. Borsanyi *et al.*, *Nature (London)* **593** (2021).
 - [6] J. P. Lees *et al.* (BABAR Collaboration), *Phys. Rev. D* **88**, 072012 (2013).
 - [7] R. Aaij *et al.* (LHCb Collaboration), *Phys. Rev. D* **97**, 072013 (2018).
 - [8] G. Caria *et al.* (Belle Collaboration), *Phys. Rev. Lett.* **124**, 161803 (2020).
 - [9] F. Sala and D. M. Straub, *Phys. Lett. B* **774**, 205 (2017).
 - [10] C.-H. Chen and T. Nomura, *Phys. Lett. B* **777**, 420 (2018).
 - [11] A. Greljo, P. Stangl, and A. Eller Thomsen, *Phys. Lett. B* **820**, 136554 (2021).
 - [12] X. G. He, G. C. Joshi, H. Lew, and R. R. Volkas, *Phys. Rev. D* **43**, R22 (1991).
 - [13] B. Shuve and I. Yavin, *Phys. Rev. D* **89**, 113004 (2014).
 - [14] W. Altmannshofer, S. Gori, S. Profumo, and F. S. Queiroz, *J. High Energy Phys.* **12** (2016) 106.
 - [15] P. Harris, P. Schuster, and J. Zupan, *arXiv:2207.08990*.
 - [16] S. Gori, M. Williams, P. Ilten, N. Tran, G. Krnjaic, N. Toro, B. Batell, N. Blinov, C. Hearty, R. McGehee *et al.*, *arXiv:2209.04671*.
 - [17] D. Forbes, C. Herwig, Y. Kahn, G. Krnjaic, C. M. Suarez, N. Tran, and A. Whitbeck, *Phys. Rev. D* **107**, 116026 (2023).

- [18] R. Capdevilla, D. Curtin, Y. Kahn, and G. Krnjaic, *J. High Energy Phys.* **04** (2022) 129.
- [19] J. P. Lees *et al.* (BABAR Collaboration), *Phys. Rev. D* **94**, 011102 (2016).
- [20] T. Czakn *et al.* (Belle Collaboration), *Phys. Rev. D* **106**, 012003 (2022).
- [21] A. Sirunyan *et al.* (CMS Collaboration), *Phys. Lett. B* **792**, 345 (2019).
- [22] I. Adachi *et al.* (Belle II Collaboration), *Phys. Rev. Lett.* **124**, 141801 (2020).
- [23] I. Adachi *et al.* (Belle II Collaboration), *Phys. Rev. Lett.* **130**, 231801 (2023).
- [24] Y. M. Andreev *et al.* (NA64 Collaboration), *Phys. Rev. D* **106**, 032015 (2022).
- [25] Adachi *et al.* (Belle II Collaboration), *Phys. Rev. Lett.* **131**, 121802 (2023).
- [26] T. Abe *et al.*, [arXiv:1011.0352](https://arxiv.org/abs/1011.0352).
- [27] E. Kou *et al.*, *Prog. Theor. Exp. Phys.* **2019**, 123C01 (2019); **2020**, 029201(E) (2020).
- [28] K. Akai, K. Furukawa, and H. Koiso (SuperKEKB Accelerator Team), *Nucl. Instrum. Methods Phys. Res., Sect. A* **907**, 188 (2018).
- [29] F. Abudinén *et al.* (Belle II Collaboration), *Chin. Phys. C* **44**, 021001 (2020).
- [30] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, *J. High Energy Phys.* **07** (2014) 079.
- [31] Q. Li and Q.-S. Yan, [arXiv:1804.00125](https://arxiv.org/abs/1804.00125).
- [32] F. Berends, P. Daverveldt, and R. Kleiss, *Nucl. Phys.* **B253**, 441 (1985).
- [33] S. Jadach, B. F. L. Ward, and Z. Wąs, *Comput. Phys. Commun.* **130**, 260 (2000).
- [34] N. Davidson, G. Nanava, T. Przedzinski, E. Richter-Wąs, and Z. Wąs, *Comput. Phys. Commun.* **183**, 821 (2012).
- [35] S. Uehara, [arXiv:1310.0157](https://arxiv.org/abs/1310.0157).
- [36] H. Czyż, M. Gunia, and J. H. Kühn, *J. High Energy Phys.* **08** (2013) 110.
- [37] G. Balossini, C. Bignamini, C. M. C. Calame, G. Montagna, O. Nicosini, and F. Piccinini, *Phys. Lett. B* **663**, 209 (2008).
- [38] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, *Comput. Phys. Commun.* **191**, 159 (2015).
- [39] D. J. Lange, *Nucl. Instrum. Methods Phys. Res., Sect. A* **462**, 152 (2001).
- [40] E. Barberio, B. van Eijk, and Z. Wąs, *Comput. Phys. Commun.* **66**, 115 (1991).
- [41] E. Barberio and Z. Wąs, *Comput. Phys. Commun.* **79**, 291 (1994).
- [42] S. Agostinelli *et al.* (GEANT4 Collaboration), *Nucl. Instrum. Methods Phys. Res., Sect. A* **506**, 250 (2003).
- [43] T. Kuhr, C. Pulvermacher, M. Ritter, T. Hauth, and N. Braun (Belle II Framework Software Group), *Comput. Software Big Sci.* **3**, 1 (2019).
- [44] Belle II Analysis Software Framework (basf2), [10.5281/zenodo.5574115](https://zenodo.org/record/5574115).
- [45] V. Bertacchi *et al.* (Belle II Tracking Group), *Comput. Phys. Commun.* **259**, 107610 (2021).
- [46] A. A. Hoecker *et al.*, [arXiv:physics/0703039](https://arxiv.org/abs/physics/0703039).
- [47] G. Punzi, [arXiv:physics/0308063](https://arxiv.org/abs/physics/0308063).
- [48] T. Skwarnicki, Ph.D. thesis, INP, Cracow, 1986.
- [49] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap* (Chapman and Hall/CRC, New York, 1994).
- [50] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Eur. Phys. J. C* **71**, 1554 (2011); **73**, 2501(E) (2013).
- [51] E. Gross and O. Vitells, *Eur. Phys. J. C* **70**, 525 (2010).
- [52] A. L. Read, *J. Phys. G* **28**, 2693 (2002).
- [53] W. Altmannshofer, S. Gori, M. Pospelov, and I. Yavin, *Phys. Rev. Lett.* **113**, 091801 (2014).
- [54] G. Bellini *et al.* (Borexino Collaboration), *Phys. Rev. Lett.* **107**, 141302 (2011).
- [55] A. Kamada, K. Kaneta, K. Yanagi, and H.-B. Yu, *J. High Energy Phys.* **06** (2018) 117.
- [56] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevD.109.112015> for numerical results of the observed cross sections of $X \rightarrow \mu\mu$, as well as of the observed 90% CL upper limits on the cross section, g' , and g_S as functions of the mass.