LHCb-CONF-2024-002
June 6, 2024

Observation of the $\Sigma^+ \to p\mu^+\mu^-$ rare decay at LHCb

LHCb collaboration[†]

Abstract

The first observation of the $\Sigma^+ \to p\mu^+\mu^-$ decay is reported with high significance using proton-proton collision data, corresponding to an integrated luminosity of 5.4 fb^{-1}, collected with the LHCb detector at a centre-of-mass energy of 13 TeV. A yield of 279 ± 19 $\Sigma^+ \to p\mu^+\mu^-$ decays is observed. No significant structure is observed in the dimuon invariant-mass distribution, which is compatible with expectations from the Standard Model.

[†]Conference report prepared for the 15th International Conference on Beauty, Charm, Hyperons in Hadronic Interactions (BEACH 2024), Charleston, South Carolina, United States, 3–7 June 2024. Contact authors: Francesco Dettori, francesco.dettori@cern.ch and Gabriele Martelli, gabriele.martelli@cern.ch.

1 Introduction

The $\Sigma^+ \to p\mu^+\mu^-$ decay is a flavour-changing neutral-current process that in the Standard Model (SM) is allowed only at loop level. The decay rate could be significantly modified by the presence of so-called new physics (NP) effects. The SM short-distance contributions arise from Z-penguin, box and electromagnetic-penguin processes, whose combined branching fraction is predicted to be $\mathcal{O}(10^{-12})$ [1]. While this is significantly smaller than the predicted long-distance contribution and can thus be ignored in the branching fraction calculation, it is worth noting that NP contributions would be expected to manifest at short distance [2]. The long-distance SM contribution is calculated from weak nonleptonic decays $\Sigma^+ \to (N\pi)^+$ and the subsequent reactions, $(N\pi)^+ \to p\gamma^{(*)}$, where N represents either a proton, p, or a neutron, n, and γ^* is a virtual photon. This branching fraction prediction produces an eight-fold ambiguity due to the description of the form factors that needs experimental inputs in order to be resolved. Four different complex form factors model this contribution, corresponding to P-even and P-odd processes that are mediated by a real and virtual intermediate photon, respectively. They are studied in chiral perturbation theory (χPT) and are computed to the lowest order with both the relativistic- and heavy-baryon approaches, which do not agree well. A unitarity argument determines the imaginary parts [1, 2], while the real components of these form factors are experimentally determined from the $\Sigma^+ \to p\gamma$ decay rate. The latter is responsible for a four-fold degeneracy in each χPT approach. The theoretical branching fraction is predicted to be in the range $\mathcal{B}(\Sigma^+ \to p\mu^+\mu^-) = [1.6, 9.1] \times 10^{-8}$ [2]. Recently, a new measurement of the branching fraction and polarisation asymmetry in $\Sigma^+ \to p\gamma$ decays has been released by the BESIII collaboration [3]. The former is significantly more precise than the world average [4], from which it also deviates considerably. Relying on this recent value, a new prediction of the branching fraction of $\mathcal{B}(\Sigma^+ \to p\mu^+\mu^-) = [1.2, 7.8] \times 10^{-8}$ has been obtained [5]. An accurate determination of the $\Sigma^+ \to p\mu^+\mu^-$ branching fraction is awaited to clarify the situation and pinpoint the correct theoretical approach.

The interest in this channel increased significantly after evidence was published by the HyperCP experiment [6] with a measured branching fraction of $\mathcal{B}(\Sigma^+ \to p\mu^+\mu^-) = (8.6^{+6.6}_{-5.4} \pm 5.5) \times 10^{-8}$, compatible with the SM, and an unexpected hint of a structure in the dimuon invariant mass distribution. Indeed, the three observed decays have almost the same dimuon invariant mass of $m_{X^0} = 214.3 \pm 0.5\,\mathrm{MeV}/c^2$, which lies close to the kinematic limit. If confirmed, this would point towards a decay with an intermediate particle decaying into two muons, *i.e.* a $\Sigma^+ \to pX^0(\to \mu^+\mu^-)$ decay, with mass $m_{X^0} = 214.3 \pm 0.5$ MeV/c^2. Albeit limited in statistical significance, this result attracted significant theoretical attention attempting to explain the origin of this supposed state [7–17]. In general, a pseudoscalar particle is favoured over a scalar particle and a lifetime of the order of 10^{-14} s is estimated for the former. Considerable experimental efforts have been made in order to search for this particle in other experiments and other decay modes [18–29]. However, no other search was performed on the $\Sigma^+ \to p\mu^+\mu^-$ decay prior to LHCb using Run 1 data [30], corresponding to an integrated luminosity of 3 fb^{-1}. At the Large Hadron Collider (LHC) hyperons are produced copiously in high energy proton-proton (pp) collisions. An excess of $10.2^{+3.9}_{-3.5}$ events on top of the background was observed with a significance of 4.1σ, corresponding to a branching fraction of $\mathcal{B}(\Sigma^+ \to p\mu^+\mu^-) = (2.2^{+1.8}_{-1.3}) \times 10^{-8}$, compatible with the SM. The background-subtracted dimuon invariant-mass distribution was consistent with a phase-space simulation, placing

an upper limit for a hypothetical X^0 particle, $\mathcal{B}(\Sigma^+ \to pX^0(\to \mu\mu)) < 1.4 \times 10^{-8}$ at 90% confidence level (CL), which disfavours the central value determined by the HyperCP collaboration.

In this conference report, a search for the $\Sigma^+ \to p\mu^+\mu^-$ decay[1] is presented as performed using pp collision data recorded by the LHCb experiment in 2016–2018 at a centre-of-mass energy of $\sqrt{s} = 13$ TeV, corresponding to an integrated luminosity of $5.4\,\text{fb}^{-1}$. Moreover, the dimuon invariant-mass distribution is also reported. A measurement of the branching fraction, using the $\Sigma^+ \to p\pi^0$ decay as a control channel, is left for future publication. This search follows a similar strategy to that performed using Run 1 data [30], namely a loose preselection is applied followed by a multivariate selection to reject most of the remaining background. A boosted decision tree algorithm (BDT) [31] is used together with particle identification (PID) variables. The signal yield is obtained from a fit to the $p\mu^+\mu^-$ invariant-mass distribution, used also to subtract the background in the dimuon invariant-mass distribution. To avoid experimenter bias, the dimuon invariant-mass distribution is not examined until the full analysis is finalised. With respect to the Run 1 analysis, several improvements are implemented with Run 2 data, namely a more efficient PID discrimination, larger simulated samples and, most importantly, new trigger lines that increase the signal efficiency by a factor ~ 10, in addition to the yield increase due to cross-section and luminosity.

2 The LHCb detector

The LHCb detector [32,33] is a single-arm forward spectrometer covering the pseudorapidity range $2 < \eta < 5$, designed for the study of particles containing b or c quarks. The detector elements that are particularly relevant to this analysis are: a silicon-strip vertex detector surrounding the pp interaction region that allows a precise measurement of the primary pp interaction vertex (PV) and of the Σ^+ baryon decay vertex (DV); a tracking system that provides a measurement of the momentum of charged particles thanks to a dipole magnet with a bending power of $4\,\text{T}\,\text{m}$; a set of ring-imaging Cherenkov detectors that provide charged particle identification; a calorimeter system which identifies photons and provides a measurement of their energy; and a muon system composed of alternating layers of iron and multiwire proportional chambers. Note that the Σ^+ baryon, due to its non-negligible lifetime of $\tau_{\Sigma^+} = (8.018 \pm 0.026) \times 10^{-11}\,\text{s}$ [4], can decay inside the vertex detector such that all decay products are reconstructed using the full tracking system (long tracks), or downstream of the vertex detector (downstream tracks). In this analysis, only long tracks are used. The short lifetime estimated for the X^0 particle would result in a prompt signal in this search; hence, no attempt is made to distinguish the dimuon origin vertex from the decay vertex of the Σ^+ baryon.

The online event selection is performed by a trigger consisting of a hardware stage, using information from the calorimeter and muon systems, followed by two software stages. In this analysis, no explicit requirements are imposed at the hardware stage, since all hardware selections are found to contribute to the sample. At the software stage, a full event reconstruction is performed. Since 2016, two inclusive dimuon trigger selections have been added at the two software trigger stages specifically designed to retain low transverse-momentum p_{T} combinations whilst remaining within the strict time constraints

[1]The inclusion of charge conjugated processes is implied throughout this report.

imposed for the software trigger. In these selections, muon tracks are required to be inconsistent with originating from any PV and to have PID information consistent with their mass hypothesis. More details can be found in Refs. [34,35]. In addition, an exclusive trigger selection has been introduced for the $\Sigma^+ \to p\mu^+\mu^-$ decay channel. Combinations of a pair of oppositely-charged muons and one proton candidate that are not consistent with originating from any PV, with a good track-fit quality and which form a good-quality vertex are combined to form a Σ^+ candidate. The Σ^+ candidate is required to have $p_{\rm T} > 500$ MeV/c, to be consistent with originating from a PV and from which it is required to be displaced. Given the large production rate of Σ^+ baryons in pp collisions, the present search is conducted including also data selected at one or more trigger stages by other particles in the event. In the offline processing, trigger decisions are associated with reconstructed candidates. A trigger decision can thus be ascribed to the reconstructed candidate, the rest of the event or a combination of both. To retain the highest possible efficiency, all the candidates passing the trigger selections above are used in the search for $\Sigma^+ \to p\mu^+\mu^-$ decays.

Simulation is used to devise and optimise the analysis strategy, as well as to estimate reconstruction and selection efficiencies. In the simulation, pp collisions are generated using Pythia [36] with a specific LHCb configuration [37]. Decays of hadronic particles are described by EvtGen [38], in which final-state radiation is generated using Photos [39]. The interaction of the generated particles with the detector, and its response, are implemented using the Geant4 toolkit [40] as described in Ref. [41]. The signal $\Sigma^+ \to p\mu^+\mu^-$ decay is generated according to a phase-space model. Per-event weights are also computed to reproduce the SM prediction and its uncertainty [2,5].

3 Selection

Offline, the only remaining sources of background are of two kinds: one is the combinatorial background, composed of random associations of tracks present in a pp event; the second component, referred to in the following as Λ background, is composed of genuine $\Lambda \to p\pi^-$ decays where the pion is misidentified as a muon, combined with a third unrelated track identified as a muon. No other kind of background contributes, mainly owing to the very small available energy in the $\Sigma^+ \to p\mu^+\mu^-$ reaction, only 39.8 MeV/c^2. This implies that, for example, $\Sigma^+ \to p\pi^+\pi^-$ and $\Sigma^+ \to \Lambda\mu^+\nu_\mu$ decays are kinematically forbidden. Misidentified decays from other hadrons also do not contribute. For example, the $K^+ \to \pi^+\pi^-\pi^+$ and $K^+ \to \pi^+\mu^-\mu^+$ decays could be misidentified as $\Sigma^+ \to p\mu^+\mu^-$ but the peak of the invariant mass of the final state particles $m_{p\mu^+\mu^-}$ is shifted considerably higher in mass. Higher mass mesons would be shifted even more. As far as background from baryons is concerned, where the final state proton is correctly identified, the Σ^+ triplet is the lightest state with a significant lifetime, hence all other possible background sources will have a mass considerably larger. Background from four or more body final states with unreconstructed particles could contribute but would not peak in the $m_{p\mu^+\mu^-}$ and is included in the combinatorial background. Finally background from clones, *i.e.* duplicated tracks from the same particle hits, is rejected offline with a requirement on the minimum angle between each pair of tracks.

To further reduce the background a multivariate operator is devised based on a Boosted Decision Tree (BDT) algorithm [42,43] implemented in the TMVA toolkit [31,44]. This

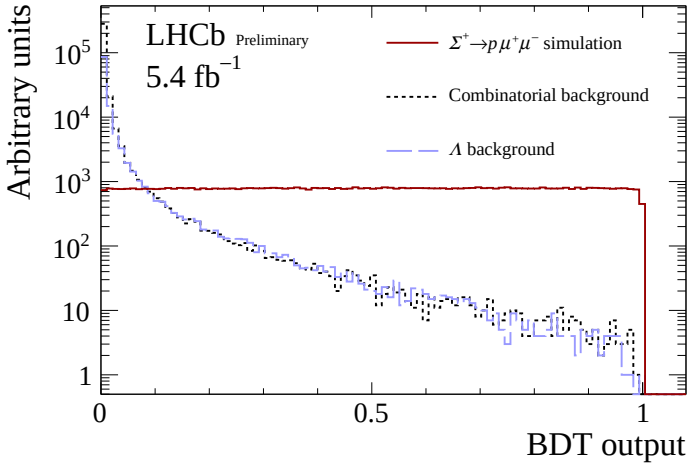

Figure 1: Distribution of the BDT output variable for simulated signal (red-solid line) and data divided into combinatorial background (black-dashed line) and Λ background (blue long-dashed line).

BDT algorithm combines the following variables: the $\chi^2_{\rm IP}$ of the Σ^+ and of the final state tracks with respect to the best PV, where the $\chi^2_{\rm IP}$ is defined as the difference in the vertex-fit χ^2 of a given PV reconstructed with and without the particle being considered; the Σ^+ DOCA, *i.e.* the maximum distance of closest approach between any pair of the three daughter tracks; the Σ^+ FD significance, *i.e.* the flight distance of the Σ^+ from the primary vertex divided by its uncertainty; the pointing angle, *i.e.* the angle between the Σ^+ momentum and the lines joining the primary and the decay vertex; the χ^2 of the Σ^+ vertex fit; the transverse momentum of the final state tracks; an isolation variable counting the transverse momentum of charged particles around the signal ones inside a cone of $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 1$. The BDT is trained using the signal simulation sample and the data sidebands as a sample of combinatorial background ($m_{p\mu^+\mu^-} < 1173\,{\rm MeV}/c^2$ or $m_{p\mu^+\mu^-} > 1205\,{\rm MeV}/c^2$). In both samples, a veto on the $p\mu^-$ mass under the $p\pi^-$ hypothesis, of $\pm 10\,{\rm MeV}/c^2$ around the known Λ mass value (Λ veto) is applied to enforce training against combinatorial background only. To avoid overtraining, the k-folding technique, with $k = 9$, is applied [45]. The BDT output ranges from zero, background-like candidates, to one, signal-like candidates. The distribution of the BDT output for signal simulation and data is shown in Fig. 1. The data is divided into a sample with the Λ veto and its complementary sample, containing mostly the Λ background. It can be seen that the BDT distribution is very similar for the two background sources.

The same BDT classifier was applied to a sample of "same-sign" $\Sigma^+ \to \bar{p}\mu^+\mu^+$ candidates in data, where a signal would violate lepton number conservation, to verify that it would not create fake structures in the background both in the $m_{p\mu^+\mu^-}$ and the $m_{\mu^+\mu^-}$ distributions.

The final selection is based on the BDT output, the muon and proton particle-identification variables [46], and the width of the Λ veto window. Criteria on these four variables are optimised, on a four-dimensional grid, to have the largest significance, defined as $s = \frac{N_{\rm S}}{\sqrt{N_{\rm S}+N_{\rm B}}}$, where $N_{\rm S}$ is the expected signal and $N_{\rm B}$ the expected background yield. The $N_{\rm S}$ estimate is based on a preliminary fit to the data after tight selection criteria applied along with the signal efficiency obtained in simulation. The N_B estimate is the

sum of two contributions: one based on a fit to the $m_{p\mu^+\mu^-}$ sidebands for the combinatorial background, and one based on a fit to the $m_{p\pi^-}$ distribution without the Λ veto to estimate the residual Λ background.

4 Fit to the $m_{p\mu^+\mu^-}$ distribution

The $m_{p\mu^+\mu^-}$ distribution for candidates satisfying the final selection criteria is shown in Fig. 2. A clear peak at the Σ^+ mass is observed, with a small residual background. An extended unbinned maximum-likelihood fit [47] is performed to the selected candidates. The signal component is described by a Hypatia function [48], with the z parameter fixed to zero. The remaining parameters are obtained from a fit to the simulated sample with the exception of the mean and resolution parameters, which are left free to vary in the fit to data. The background is described by a modified Argus function [49] where the threshold parameter is fixed to the sum of the daughter masses and the remaining parameters are left free to float. The result of the fit is superimposed onto the data in Fig. 2, both in linear and logarithmic scale. The fit results in a signal yield of

$$N_{\Sigma^+\to p\mu^+\mu^-} = 279 \pm 19\,,$$

where the uncertainty is statistical only. This result represents the first observation of the $\Sigma^+ \to p\mu^+\mu^-$ decay with overwhelming significance. The fit to the data is repeated leaving all the parameters of the signal function free to float instead of fixed to the simulation results. The variation of the signal yield is negligible, hence no systematic uncertainty is assigned. A similar amount of Σ^+ and $\overline{\Sigma}^-$ are seen in the sample.

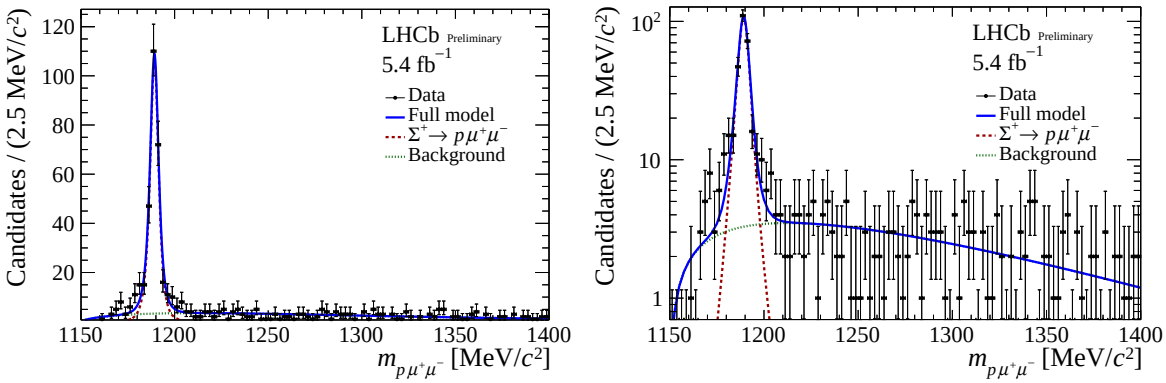

Figure 2: Distribution of the invariant mass of $\Sigma^+ \to p\mu^+\mu^-$ candidates with the result of an extended unbinned maximum-likelihood fit superimposed (blue-solid line) in (left) linear scale and (right) logarithmic scale. The signal (red-dashed line) and background (green-dotted line) components are also shown.

5 Dimuon invariant-mass distribution

The distribution of the dimuon invariant mass is shown for data in Fig. 3 after background subtraction. The background is subtracted using per-event signal weights derived with

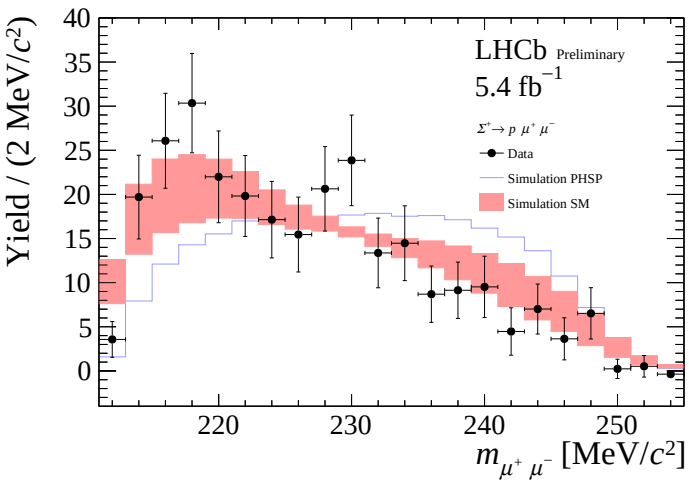

Figure 3: Distribution of dimuon invariant mass for $\Sigma^+ \to p\mu^+\mu^-$ candidates in data compared with simulation. LHCb phase space is shown as is (blue line), and weighted according to the SM amplitude [2, 5] (red band).

the *sPlot* method [50] using $m_{p\mu^+\mu^-}$ as the discriminant variable. The $m_{p\mu^+\mu^-}$ and $m_{\mu^+\mu^-}$ variables are found to be uncorrelated aside from the higher border of the kinematics space. It is checked that the same distribution is obtained performing the unbinned maximum-likelihood fit described in Sec. 4 in bins of the dimuon invariant mass. From a comparison of the data distribution with the phase-space simulated sample, it is clear that no significant peaking structures are visible. The simulation is weighted with a per-event weight obtained from the ratio of the SM prediction [2, 5] and the generated phase-space distribution. This is shown by the red band in Fig. 3, whose range accounts for the maximum spread in central values of the four-fold degeneracies present in each χPT approach as the weight uncertainty. When comparing the data to this reweighted simulated sample a rather good agreement is retrieved in the full dimuon invariant-mass range.

A scan for a possible resonant structure in the dimuon invariant mass is performed selecting candidates in a region within twice the signal resolution in the $p\mu^+\mu^-$ invariant mass around the known Σ^+ mass, using the same method as the Run 1 analysis [30]. The distribution of these candidates as a function of the dimuon invariant mass is shown in Fig. 4. Steps of half the resolution on the dimuon invariant mass, $\sigma(m_{\mu^+\mu^-})$, are considered in this scan, following the method outlined in Ref. [51]. The value of $\sigma(m_{\mu^+\mu^-})$ varies in the range [0.5, 2] MeV/c^2 depending on the dimuon invariant mass as shown in Fig. 7. For each step the putative signal is estimated in a window of $\pm 1.5 \times \sigma(m_{\mu^+\mu^-})$ around the considered particle mass, while the background is estimated from the lower and upper mass sidebands contained in the range $[1.5 - 4.0] \times \sigma(m_{\mu^+\mu^-})$ from the same mass. Only one of the two sidebands is considered when the second is outside the allowed kinematic range. The local p-value of the background-only hypothesis is shown in Fig. 4 as a function of the dimuon invariant mass, and no significant signal is found: the most significant point is at 229.37 MeV/c^2, with a p-value of 3%. When considering a putative candidate with a mass $m_{X^0} = 214.3\,\text{MeV}/c^2$ the fractional contribution to all the candidates in the mass window is 3.7% and the difference with respect to the expected background from the $m_{\mu^+\mu^-}$ sidebands (i.e. nonresonant) is -4 events.

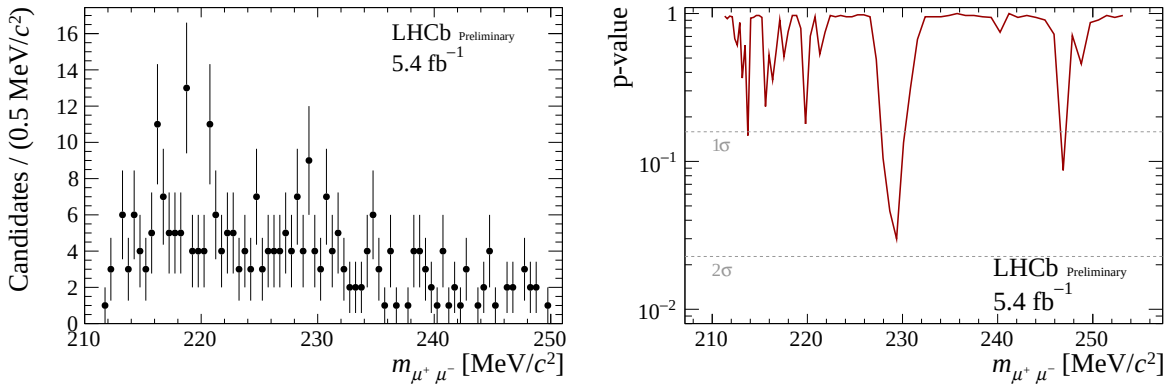

Figure 4: (left) Distribution of the dimuon invariant mass of $\Sigma^+ \to p\mu^+\mu^-$ decays in a signal region of ± 2 times the resolution on $m_{p\mu^+\mu^-}$. No background subtraction is applied. (right) The local p-value in each dimuon invariant mass window, obtained as described in the text. The horizontal dashed lines correspond to the p-value of one and two standard deviations.

6 Conclusions and outlook

The $\Sigma^+ \to p\mu^+\mu^-$ decay is observed with very large significance in data collected in Run 2 by the LHCb experiment in pp collisions, with a yield of $N_{\Sigma^+\to p\mu^+\mu^-} = 279 \pm 19$. No structure is seen in the dimuon invariant mass distribution which is compatible with expectations from the Standard Model. A precise determination of the branching fraction is left for a future publication. With the collected signal yield, in addition to a measurement of the integrated and differential branching fraction, other measurements are possible, such as charge-parity symmetry violation and forward-backward asymmetries.

7 Supplementary material for LHCb-CONF-2024-002

Additional figures are presented in this section.

In Fig. 5 an example of Feynman diagrams for (a) and (b) the SM and (c) a possible NP contribution to the $\Sigma^+ \to p\mu^+\mu^-$ decay is shown.

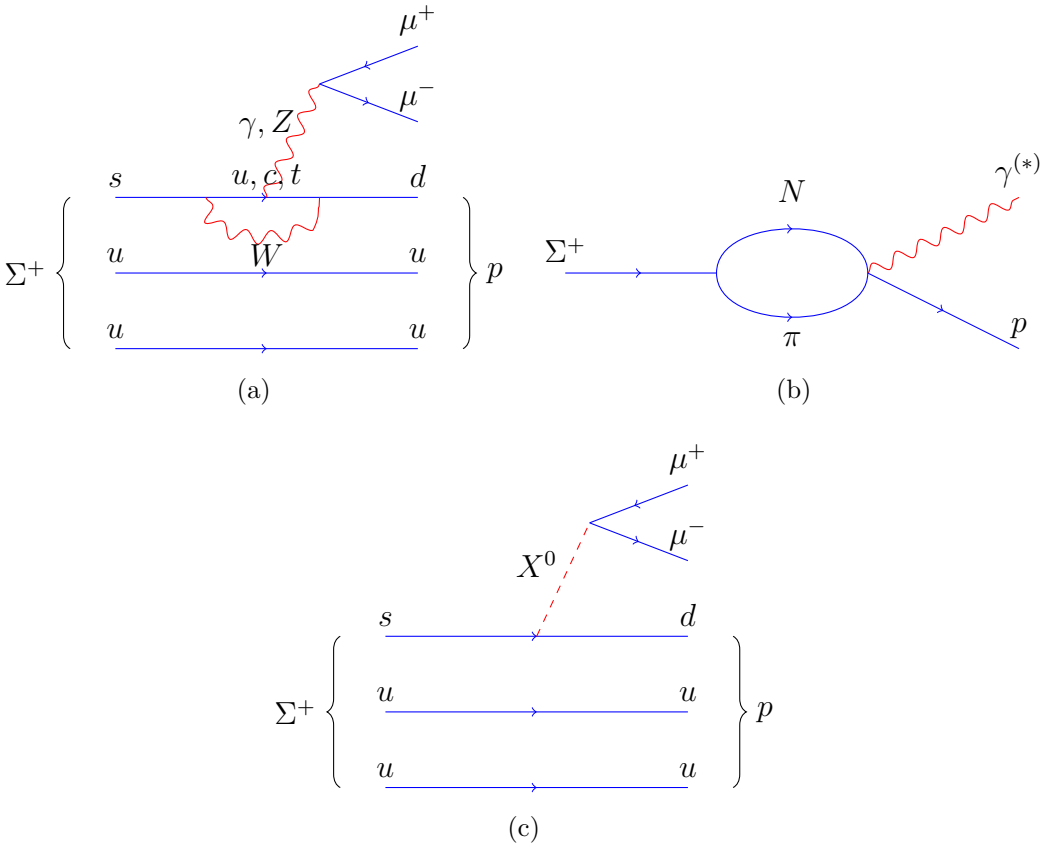

Figure 5: Example of Feynman diagrams for (a) the SM short-distance contribution, (b) the SM long-distance contribution, and (c) a possible NP contribution to the $\Sigma^+ \to p\mu^+\mu^-$ decay.

The distribution of the dimuon invariant mass is shown for data in Fig. 6 using a finer binning.

The resolution of the dimuon invariant mass as determined from data is shown in Fig. 7 varying from about $0.5\,\text{MeV}/c^2$ at low dimuon mass to about 2 MeV/c^2 at higher masses. This shows that any intermediate particle contribution would be present as a very narrow peak in the dimuon mass distribution.

In Fig. 8 the distribution of the $p\mu^-$ invariant mass in the $p\pi^-$ hypothesis is shown for $\Sigma^+ \to p\mu^+\mu^-$ candidates in data within a $\pm 6\,\text{MeV}/c^2$ window from the Σ^+ mass peak, without the $\Lambda\Lambda$ veto. The veto is indicated by the two vertical red lines. The residual background coming from Λ decays outside of the veto represents only a small tail distributed along the $m_{p\mu^+\mu^-}$ mass, taken into account by the combinatorial background.

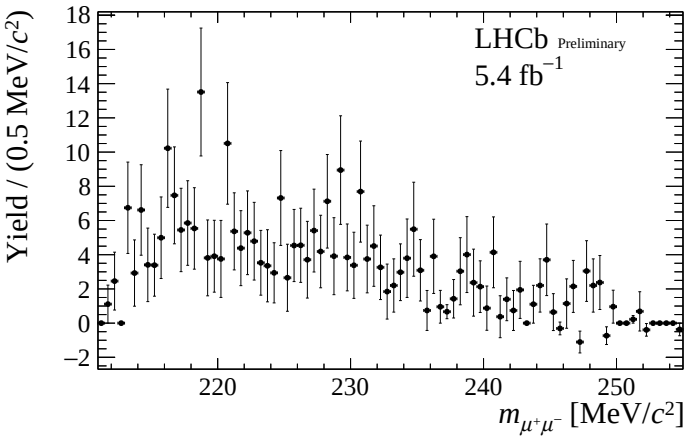

Figure 6: Distribution of the dimuon invariant mass for $\Sigma^+ \to p\mu^+\mu^-$ candidates in data using a finer binning.

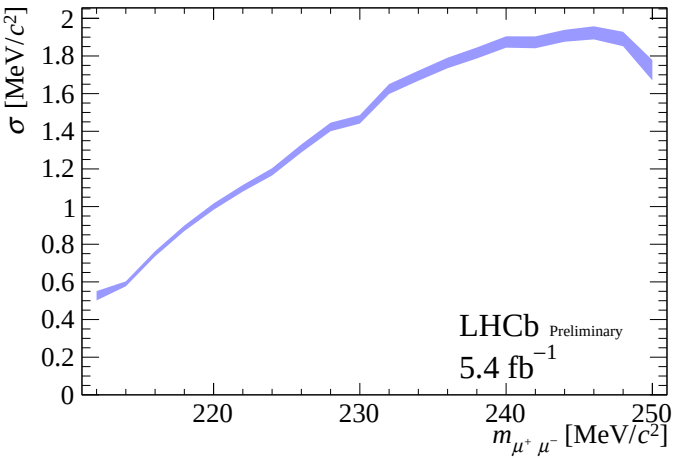

Figure 7: Dimuon invariant-mass resolution σ for $\Sigma^+ \to p\mu^+\mu^-$ candidates versus the mass itself.

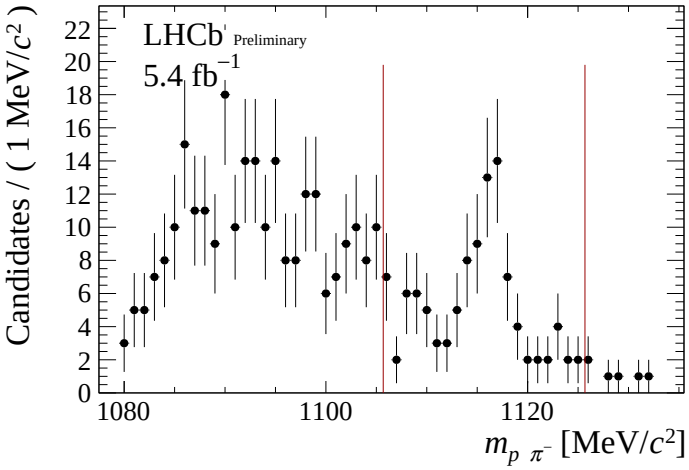

Figure 8: Distribution of the $p\mu^-$ invariant mass in the $p\pi^-$ hypothesis for $\Sigma^+ \to p\mu^+\mu^-$ candidates in data within a $\pm 6\,\mathrm{MeV}/c^2$ window from the Σ^+ mass peak, without the Λ veto (indicated between red lines).

References

[1] X.-G. He, J. Tandean, and G. Valencia, *The Decay $\Sigma^+ \to p\ell^+\ell^-$ within the standard model*, Phys. Rev. **D72** (2005) 074003, arXiv:hep-ph/0506067.

[2] X.-G. He, J. Tandean, and G. Valencia, *Decay rate and asymmetries of $\Sigma^+ \to p\mu^+\mu^-$*, JHEP **10** (2018) 040, arXiv:1806.08350.

[3] BESIII collaboration, M. Ablikim *et al.*, *Precision Measurement of the Decay $\Sigma^+ \to p\gamma$ in the Process $J/\psi \to \Sigma^+\overline{\Sigma}^-$*, Phys. Rev. Lett. **130** (2023) 211901, arXiv:2302.13568.

[4] Particle Data Group, R. L. Workman *et al.*, *Review of Particle Physics*, PTEP **2022** (2022) 083C01.

[5] A. Roy, J. Tandean, and G. Valencia, *The decays $\Sigma^+ \to p\ell^+\ell^-$ within the standard model and beyond*, arXiv:2404.15268.

[6] HyperCP Collaboration, H. Park *et al.*, *Evidence for the decay $\Sigma^+ \to p\mu^+\mu^-$*, Phys. Rev. Lett. **94** (2005) 021801, arXiv:hep-ex/0501014.

[7] X.-G. He, J. Tandean, and G. Valencia, *Does the HyperCP Evidence for the Decay $\Sigma^+ \to p\mu^+\mu^-$ Indicate a Light Pseudoscalar Higgs Boson?*, Phys. Rev. Lett. **98** (2007) 081802, arXiv:hep-ph/0610362.

[8] X.-G. He, J. Tandean, and G. Valencia, *Light Higgs production in hyperon decay*, Phys. Rev. **D74** (2006) 115015, arXiv:hep-ph/0610274.

[9] D. S. Gorbunov and V. A. Rubakov, *On sgoldstino interpretation of HyperCP events*, Phys. Rev. **D73** (2006) 035002, arXiv:hep-ph/0509147.

[10] S. V. Demidov and D. S. Gorbunov, *More about the sgoldstino interpretation of HyperCP events*, JETP Lett. **84** (2007) 479, arXiv:hep-ph/0610066.

[11] X.-G. He, J. Tandean, and G. Valencia, *Implications of a new particle from the HyperCP data on $\Sigma^+ \to p\mu^+\mu^-$*, Phys. Lett. **B631** (2005) 100, arXiv:hep-ph/0509041.

[12] C. Q. Geng and Y. K. Hsiao, *Constraints on the new particle in $\Sigma^+ \to p\mu^+\mu^-$*, Phys. Lett. **B632** (2006) 215, arXiv:hep-ph/0509175.

[13] N. G. Deshpande, G. Eilam, and J. Jiang, *On the possibility of a new boson X0 (214-MeV) in $\Sigma^+ \to p\mu^+\mu^-$*, Phys. Lett. **B632** (2006) 212, arXiv:hep-ph/0509081.

[14] C.-H. Chen, C.-Q. Geng, and C.-W. Kao, *U-boson and the HyperCP exotic events*, Phys. Lett. **B663** (2008) 400, arXiv:0708.0937.

[15] G. Xiangdong, C. S. Li, Z. Li, and H. Zhang, *Contributions from SUSY-FCNC couplings to the interpretation of the HyperCP events for the decay $\Sigma^+ \to p\mu^+\mu^-$*, Eur. Phys. J. **C55** (2008) 317, arXiv:0712.0257.

[16] M. L. Mangano and P. Nason, *Radiative quarkonium decays and the NMSSM Higgs interpretation of the hyperCP $\Sigma^+ \to p\mu^+\mu^-$ events*, Mod. Phys. Lett. **A22** (2007) 1373, arXiv:0704.1719.

[17] M. Pospelov, *Secluded U(1) below the weak scale*, Phys. Rev. **D80** (2009) 095002, arXiv:0811.1030.

[18] CLEO collaboration, W. Love *et al.*, *Search for Very Light CP-Odd Higgs Boson in Radiative Decays of Upsilon(S-1)*, Phys. Rev. Lett. **101** (2008) 151802, arXiv:0807.1427.

[19] E391a collaboration, Y. C. Tung *et al.*, *Search for a light pseudoscalar particle in the decay* $K^0_{(L)} \to \pi^0\pi^0 X$, Phys. Rev. Lett. **102** (2009) 051802, arXiv:0810.4222.

[20] D0 collaboration, V. M. Abazov *et al.*, *Search for NMSSM Higgs bosons in the* $h \to aa \to \mu\mu\mu\mu, \mu\mu\tau\tau$ *channels using p anti-p collisions at s**(1/2) = 1.96-TeV*, Phys. Rev. Lett. **103** (2009) 061801, arXiv:0905.3381.

[21] BaBar collaboration, B. Aubert *et al.*, *Search for Dimuon Decays of a Light Scalar Boson in Radiative Transitions* $\Upsilon \to \gamma A^0$, Phys. Rev. Lett. **103** (2009) 081803, arXiv:0905.4539.

[22] Belle collaboration, H. J. Hyun *et al.*, *Search for a Low Mass Particle Decaying into* $\mu^+\mu^-$ *in* $B^0 \to K^{*0}X$ *and* $B^0 \to \rho^0 X$ *at Belle*, Phys. Rev. Lett. **105** (2010) 091801, arXiv:1005.1450.

[23] KTeV collaboration, E. Abouzaid *et al.*, *Search for the Rare Decays* $K_L \to \pi^0\pi^0\mu^+\mu^-$ *and* $K_L \to \pi^0\pi^0 X^0 \to \pi^0\pi^0\mu^+\mu^-$, Phys. Rev. Lett. **107** (2011) 201803, arXiv:1105.4800.

[24] BESIII collaboration, M. Ablikim *et al.*, *Search for a light Higgs-like boson* A^0 *in* J/ψ *radiative decays*, Phys. Rev. **D85** (2012) 092012, arXiv:1111.2112.

[25] BaBar collaboration, J. P. Lees *et al.*, *Search for a Dark Photon in* e^+e^- *Collisions at BaBar*, Phys. Rev. Lett. **113** (2014) 201801, arXiv:1406.2980.

[26] LHCb collaboration, R. Aaij *et al.*, *Search for rare* $\mathcal{B}^0_{(s)} \to \mu^+\mu^-\mu^+\mu^-$ *decays*, Phys. Rev. Lett. **110** (2013) 211801, arXiv:1303.1092.

[27] LHCb collaboration, R. Aaij *et al.*, *Search for hidden-sector bosons in* $\mathcal{B}^0 \to K^{*0}\mu^+\mu^-$ *decays*, Phys. Rev. Lett. **115** (2015) 161802, arXiv:1508.04094.

[28] LHCb collaboration, R. Aaij *et al.*, *Search for long-lived scalar particles in* $\mathcal{B}^+ \to K^+\chi(\mu^+\mu^-)$ *decays*, Phys. Rev. **D95** (2017) 071101, arXiv:1612.07818.

[29] LHCb collaboration, R. Aaij *et al.*, *Search for dark photons produced in 13 TeV pp collisions*, Phys. Rev. Lett. **120** (2018) 061801, arXiv:1710.02867.

[30] LHCb collaboration, R. Aaij *et al.*, *Evidence for the rare decay* $\Sigma^+ \to p\mu^+\mu^-$, Phys. Rev. Lett. **120** (2018) 221803, arXiv:1712.08606.

[31] H. Voss, A. Hoecker, J. Stelzer, and F. Tegenfeldt, *TMVA - Toolkit for Multivariate Data Analysis with ROOT*, PoS **ACAT** (2007) 040.

[32] LHCb collaboration, A. A. Alves Jr. *et al.*, *The LHCb detector at the LHC*, JINST **3** (2008) S08005.

[33] LHCb collaboration, R. Aaij *et al.*, *LHCb detector performance*, Int. J. Mod. Phys. **A30** (2015) 1530022, arXiv:1412.6352.

[34] F. Dettori, D. Martinez Santos, and J. Prisciandaro, *Low-p_T dimuon triggers at LHCb in Run 2*, LHCb-PUB-2017-023. CERN-LHCb-PUB-2017-023, CERN, Geneva, 2017.

[35] LHCb collaboration, R. Aaij *et al.*, *Design and performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC*, JINST **14** (2019) P04013, arXiv:1812.10790.

[36] T. Sjöstrand, S. Mrenna, and P. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, arXiv:0710.3820; T. Sjöstrand, S. Mrenna, and P. Skands, *PYTHIA 6.4 physics and manual*, JHEP **05** (2006) 026, arXiv:hep-ph/0603175.

[37] I. Belyaev *et al.*, *Handling of the generation of primary events in Gauss, the LHCb simulation framework*, J. Phys. Conf. Ser. **331** (2011) 032047.

[38] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. **A462** (2001) 152.

[39] N. Davidson, T. Przedzinski, and Z. Was, *PHOTOS interface in C++: Technical and physics documentation*, Comp. Phys. Comm. **199** (2016) 86, arXiv:1011.0937.

[40] Geant4 collaboration, J. Allison *et al.*, *Geant4 developments and applications*, IEEE Trans. Nucl. Sci. **53** (2006) 270; Geant4 collaboration, S. Agostinelli *et al.*, *Geant4: A simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250.

[41] M. Clemencic *et al.*, *The LHCb simulation application, Gauss: Design, evolution and experience*, J. Phys. Conf. Ser. **331** (2011) 032023.

[42] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and regression trees*, Wadsworth international group, Belmont, California, USA, 1984.

[43] Y. Freund and R. E. Schapire, *A decision-theoretic generalization of on-line learning and an application to boosting*, J. Comput. Syst. Sci. **55** (1997) 119.

[44] A. Hoecker *et al.*, *TMVA 4 — Toolkit for Multivariate Data Analysis with ROOT. Users Guide.*, arXiv:physics/0703039.

[45] A. Blum, A. Kalai, and J. Langford, *Beating the hold-out: bounds for k-fold and progressive cross-validation*, in *Proceedings of the twelfth annual conference on computational learning theory*, COLT '99, (New York, NY, USA), 203, ACM, 1999.

[46] R. Aaij *et al.*, *Selection and processing of calibration samples to measure the particle identification performance of the LHCb experiment in Run 2*, Eur. Phys. J. Tech. Instr. **6** (2019) 1, arXiv:1803.00824.

[47] R. J. Barlow, *Extended maximum likelihood*, Nucl. Instrum. Meth. A **297** (1990) 496.

[48] D. Martinez Santos and F. Dupertuis, *Mass distributions marginalized over per-event errors*, Nucl. Instrum. Meth. A **764** (2014) 150, arXiv:1312.5000.

[49] ARGUS collaboration, H. Albrecht *et al.*, *Search for Hadronic $b \to u$ Decays*, Phys. Lett. B **241** (1990) 278.

[50] M. Pivk and F. R. Le Diberder, *sPlot: A statistical tool to unfold data distributions*, Nucl. Instrum. Meth. **A555** (2005) 356, arXiv:physics/0402083.

[51] M. Williams, *Searching for a particle of unknown mass and lifetime in the presence of an unknown non-monotonic background*, JINST **10** (2015) P06002, arXiv:1503.04767.