Strong constraints on the $K_S^0 \rightarrow \mu^+ \mu^-$ branching fraction

LHCb collaboration

Abstract

A search for the decay $K_S^0 \rightarrow \mu^+ \mu^-$ is performed using proton-proton collision data, corresponding to an integrated luminosity of 5.6 fb$^{-1}$, collected with the LHCb experiment during 2016, 2017 and 2018 at a center-of-mass energy of 13 TeV. The observed number of signal decays is consistent with zero, yielding an upper limit of $\mathcal{B}(K_S^0 \rightarrow \mu^+ \mu^-) < 2.2 (2.6) \times 10^{-10}$ at 90 (95)% CL. The limit reduces to $\mathcal{B}(K_S^0 \rightarrow \mu^+ \mu^-) < 2.1 (2.4) \times 10^{-10}$ at 90 (95)% CL once combined with the full Run I dataset.
The decay $K_S^0 \rightarrow \mu^+\mu^-$ is a Flavor Changing Neutral Current (FCNC) process which has not been observed yet. In the Standard Model (SM), this decay is highly suppressed, with an expected branching fraction $\mathcal{B}(K_S^0 \rightarrow \mu^+\mu^-)_{\text{SM}} = (5.18 \pm 1.50_{\text{LD}} \pm 0.02_{\text{SD}}) \times 10^{-12}$, where the uncertainty with subscript LD (SD) relates to long- (short-) distance effects. The main contributions to the $K_S^0 \rightarrow \mu^+\mu^-$ decay amplitude are illustrated in Fig. 1. The related channel $K_L^0 \rightarrow \mu^+\mu^-$, is predicted in the SM to occur with a branching fraction of $\mathcal{B}(K_L^0 \rightarrow \mu^+\mu^-)_{\text{SM}} = (6.85 \pm 0.80_{\text{LD}} \pm 0.06_{\text{SD}}) \times 10^{-9}$ or $\mathcal{B}(K_L^0 \rightarrow \mu^+\mu^-)_{\text{SM}} = (8.11 \pm 1.49_{\text{LD}} \pm 0.13_{\text{SD}}) \times 10^{-9}$ for an (unknown) positive or negative relative sign of the $K_L^0 \rightarrow \gamma\gamma$ amplitude, respectively. Both $\mathcal{B}(K_L^0 \rightarrow \mu^+\mu^-)$ theory predictions are in a good agreement with the experimental world average, based on Refs. [6–8], $\mathcal{B}(K_L^0 \rightarrow \mu^+\mu^-) = (6.84 \pm 0.11) \times 10^{-9}$.

![Figure 1: Diagrams representing SM contributions to the $K_S^0 \rightarrow \mu^+\mu^-$ decay amplitude. (top) Long-distance contribution, generated by two intermediate photons, and (bottom) short-distance contributions.](image-url)

Both the $K_S^0$ and the $K_L^0$ decay amplitudes are dominated by LD contributions in the SM. The reason for the large difference between the two branching fractions is that the $S$-wave component is charge-parity (CP) violating and CP conserving for the $K_S^0$ and $K_L^0$ modes, respectively. In the $K_S^0$ case, the CP-conserving long-distance contribution can only proceed through $P$-wave, and the CP-violating short distance component in the SM is even more suppressed.

Due to the strong suppression of the SM decay amplitude, dynamics beyond the Standard Model (BSM) can induce large deviations of $\mathcal{B}(K_S^0 \rightarrow \mu^+\mu^-)$ with respect to the SM prediction. This has been shown to be the case in SUSY scenarios as well as in leptoquark models [10, 11]. The current best limit, $\mathcal{B}(K_S^0 \rightarrow \mu^+\mu^-) < 0.8 (1.0) \times 10^{-9}$ at 90 (95)% CL, was set by LHCb with the data collected during Run 1 (2010–2012).

In this note, a significantly improved limit is presented. Results are based on proton-proton ($pp$) collision data collected with the LHCb detector at a center-of-mass energy of 13 TeV during 2016, 2017, and 2018, corresponding to an integrated luminosity of 5.6 fb$^{-1}$. A major improvement with respect to the previous analysis is achieved by employing dedicated software triggers that were not present in Run 1. While the analysis strategy closely follows what was done for Run 1, the event reconstruction and selection have been improved, as discussed below. This measurement benefits from the huge $K_S^0$ production cross section at the LHC, of approximately 0.6 barn at a center-of-mass energy
of 13 TeV [13], and from the forward geometry of the vertex detector of LHCb.

The LHCb detector [14,15] 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 includes a high-precision tracking system consisting of a silicon-strip vertex detector (VELO) surrounding the $pp$ interaction region, a large-area silicon-strip detector located upstream of a dipole magnet with a bending power of about 4 Tm, and three stations of silicon-strip detectors and straw drift tubes placed downstream of the magnet. The tracking system provides a measurement of the momentum, $p$, of charged particles with a relative uncertainty that varies from 0.5% at low momentum to 1.0% at 200 GeV/c. The minimum distance of a track to a primary vertex (PV), the impact parameter (IP), is measured with a resolution of $(15 + 29/p_T) \mu m$, where $p_T$ is the component of the momentum transverse to the beam, in GeV/c. Different types of charged hadrons are distinguished using information from two ring-imaging Cherenkov (RICH) detectors. Photons, electrons and hadrons are identified by a calorimeter system consisting of scintillating-pad and preshower detectors, an electromagnetic and a hadronic calorimeter. Muons are identified by a system composed of alternating layers of iron and multiwire proportional chambers.

Information from the tracking system, the calorimeter system, the muon system, and the RICH detectors is used for the precise muon identification and momentum measurement.

Events are first required to pass a hardware-trigger selection [16], based on information from the calorimeter and muon system. Subsequently, a full event reconstruction is applied in a two-step software selection. In the previous analysis, the search was limited by the muon $p_T$ threshold of approximately 1.8 GeV/c. In Run 2, a new tracking method was applied and a dedicated software trigger selection was developed, which allowed to lower the $p_T$ muon thresholds to 80 MeV/c. This translates into an increase of the trigger efficiency for $K_S^0 \rightarrow \mu^+\mu^-$ of about an order of magnitude with respect to Run 1 [17]. While the trigger efficiency at the software stage is approximately 60%, it is only 11% at the hardware stage, as it relies on high-$p_T$ event signatures.

The $K_S^0 \rightarrow \pi^+\pi^-$ decay is used as a normalization mode given its abundance, its similar topology to $K_S^0 \rightarrow \mu^+\mu^-$ and its well-known branching fraction [5]. Common offline preselection criteria are applied to $K_S^0 \rightarrow \mu^+\mu^-$ and $K_S^0 \rightarrow \pi^+\pi^-$ candidates in order to cancel many systematic effects in the efficiency ratio. Candidates for the $K_S^0 \rightarrow \mu^+\mu^-$ ($K_S^0 \rightarrow \pi^+\pi^-$) decay are obtained from two tracks with opposite charge identified as muons (pions), forming a well detached secondary vertex with an invariant mass in the vicinity of the known $K_S^0$ mass value [5]. Kaon candidates are required to decay inside the VELO, where the best $K_S^0$ mass resolution is achieved. Approximately 22% of $K_S^0$ produced at the $pp$ interaction point decay within the acceptance of the VELO. The secondary vertex (SV) must be well detached from the PV by requiring the $K_S^0$ candidate decay time to be larger than 6% of the known $K_S^0$ lifetime [5]. The $K_S^0$ candidate origin must be compatible with a PV, while its decay products should be inconsistent with originating from any PV. Decays of $\Lambda$ baryons to $p\pi^-$ [1] are suppressed by removing candidates close to the expected elliptical kinematic regions in the Armenteros-Podolanski plane [18]. The loss in signal efficiency is negligible. Muon tracks are required to have associated hits in the muon system [19]. Pions from $K_S^0 \rightarrow \pi^+\pi^-$ decays are required to be within the muon system acceptance. The main background sources are combinatorial and $K_S^0 \rightarrow \pi^+\pi^-$ decays. In $K_S^0 \rightarrow \pi^+\pi^-$, the pions are misidentified as muons and in which the invariant

---

1 The inclusion of charge-conjugate processes is implied throughout this paper, unless otherwise noted.
mass of the kaon candidate is underestimated on average by 40 MeV/c² corresponding to ten times the $K^0_S \to \mu^+\mu^-$ resolution.

Combinatorial background is suppressed using an Adaptive Boosted Decision Tree (BDT) \cite{20,21} algorithm, taken from the XGBoost library \cite{22}. Simulated $K^0_S \to \mu^+\mu^-$ decays is used as a proxy for signal, and $K^0_S \to \mu^+\mu^-$ candidates from data in the dimuon invariant mass region above 520 MeV/c² as a proxy for combinatorial background. Before the BDT training, the simulated $K^0_S \to \mu^+\mu^-$ candidates are weighted using a Gradient Boost algorithm trained with simulated and real data $K^0_S \to \pi^+\pi^-$ candidates, to take into account small differences between data and simulation. Since the background candidates used in the training are part of the fitted sample, the $k$-folding approach \cite{23} is applied to avoid overtraining.

The BDT is trained independently in two mutually exclusive candidate categories. In the first category, referred to as exclusively trigger-on-signal (xTOS), the events are triggered at the hardware stage exclusively by the signal candidate decay products. The second category, referred to as trigger-independent-of-signal (TIS), contains the rest of the events \cite{24}. Both categories are required to fulfi the same software trigger requirements.

The BDT variables are: the kaon candidate decay time and IP significance, the IP significance and the track-fit $\chi^2$ of each of the two tracks, the distance of closest approach between the two tracks, the cosine of the helicity angle, the $\chi^2$ of the SV fit, two SV isolation variables, and a VELO material veto variable \cite{25}. The SV isolation variables are defined as the difference in the $\chi^2$ of the vertex fit with only the two final-state tracks and when adding the one or two nearest tracks. The VELO material veto variable suppresses efficiently background originating from inelastic interactions with the VELO stations and RF foil which separates the VELO modules from the beam vacuum. A selection requirement is placed on the BDT, rejecting 99% of the combinatorial background with a signal efficiency of approximately 63% for both trigger categories.

Another significant background source is $K^0_L \to \mu^+\mu^-$, which at LHCb is suppressed by a factor of approximately $2.3 \times 10^{-3}$ relative to $K^0_S \to \mu^+\mu^-$ decays due to its longer lifetime. Interference between $K^0_S$ and $K^0_L$ vanishes due to $K^0$ and $K^0$ being produced in equal amounts \cite{3} at the LHC. Backgrounds such as $K^0 \to \mu^+\mu^-\gamma(\gamma)$, $\Sigma^+ \to p\mu^+\mu^-$, $K^{0,+} \to \pi^0+\mu^+\mu^-$, $K^0 \to \pi^+\mu^+\nu\mu$, $\Lambda \to p\pi^-$, $\omega \to \pi^0\mu^+\mu^-$, $\eta \to \mu^+\mu^-\gamma$ decays are found to be negligible.

The data sample is divided into twenty categories: two trigger decisions times ten bins of the BDT response. The BDT bins are chosen to have the same fraction of simulated signal candidates in each bin. A dedicated muon identification Boosted Decision Tree ($\mu$BDT) is trained to suppress $K^0_S \to \pi^+\pi^-$ decays using the algorithm presented in Ref. \cite{12}. The selection criterion on the $\mu$BDT is optimized independently for each of the twenty categories.

The $K^0_S \to \mu^+\mu^-$ branching fraction is determined in an unbinned maximum-likelihood fit to the kaon candidate invariant mass in the range 480–595 MeV/c². Taking into account the ratio of detection efficiencies, the signal yield is normalized to $K^0_S \to \pi^+\pi^-$ decays to cancel uncertainties due to the $K^0_S$ cross-section, luminosity, reconstruction, and selection cuts including the BDT binning. The fit is performed simultaneously in the twenty data categories. The contributions considered are: $K^0_S \to \mu^+\mu^-$ signal, modelled with a Hypathia distribution \cite{24}; the combinatorial background, described by an exponential function; $K^0_S \to \pi^+\pi^-$ background, modelled with a power law distribution, and $K^0_L \to \mu^+\mu^-$, described with the same probability density function as the $K^0_S \to \mu^+\mu^-$. The $\chi^2$ defined as the difference in the isolation variables, and a VELO material veto variable \cite{25}. The SV isolation variables are...
decay. All yields are free to vary in the fit. A Gaussian constraint is applied to the yield of the $K^0_S \to \mu^+ \mu^-$ component, based on its known branching fraction [5] and on the efficiency ratio between $K^0_L \to \mu^+ \mu^-$ and $K^0_S \to \mu^+ \mu^-$. Additional Gaussian constraints are applied to the efficiency ratio between $K^0_S \to \mu^+ \mu^-$ and $K^0_S \to \pi^+ \pi^-$, based on the systematic uncertainties. An independent sample of $K^0_S \to \pi^+ \pi^-$ decays obtained from a trigger-unbiased sample is used to calibrate the $K^0_S$ candidate mass peak position and resolution parameters (see Fig. 2). It is also used to correct the simulation to obtain the efficiencies of the signal and the normalization channels.

The response of the muon system is calibrated using $J/\psi \to \mu^+ \mu^-$ decays complemented with the use of $K^0_{S,L} \to \pi^+ \mu^+ \nu$ decays at low transverse momentum. The yield of $K^0_{S,L} \to \pi^+ \mu^+ \nu$ decays as a function of the data taking period is also used to evaluate the variation of the total efficiency with time, mostly caused by changes in the thresholds of the hardware trigger. The obtained single event sensitivity is $(3.01 \pm 0.56) \times 10^{-12}$, meaning that approximately two $K^0_S \to \mu^+ \mu^-$ and five $K^0_L \to \mu^+ \mu^-$ signal decays originating from SM processes are expected to be present in the dataset.

Various sources of systematic uncertainty have been taken into account. The main source is due to the knowledge of the trigger efficiency, yielding 11% for the hardware trigger and 13% for the software trigger. Other sources include data-simulation differences in the muon identification, the correction applied on simulation, the efficiency ratio between the signal and normalization modes, the BDT response due to changes of the experimental conditions, and the knowledge of the $K^0_S \to \pi^+ \pi^-$ branching fraction. The total systematic uncertainty is between 19% and 23%, depending on the trigger category and the BDT bin. It tends to be lower in the TIS trigger category and higher in lower BDT bins, which have lower signal to background ratio. The systematic uncertainties are taken into account as Gaussian constraints in the fit to the data.

The fit shows no evidence for $K^0_S \to \mu^+ \mu^-$ decays (see Fig. 3), yielding a signal significance of 1.5$\sigma$ (1.4$\sigma$ when combined with Run 1 data). An upper limit is obtained by integrating the profile likelihood multiplied by a flat prior in the positive branching fraction domain, yielding $2.2 (2.6) \times 10^{-10}$ at 90 (95)% confidence level (CL). The likelihood is combined with Run 1, obtaining a limit of $2.1 (2.4) \times 10^{-10}$ at 90 (95)% CL. A log-likelihood interval of one standard deviation ($-2\Delta \log \mathcal{L} = 1$) from
the Run 2 dataset yields $B(K_S^0 \rightarrow \mu^+\mu^-) = 1.03^{+0.76}_{-0.68} \times 10^{-10}$. Combined with Run 1 it yields $B(K_S^0 \rightarrow \mu^+\mu^-) = 0.94^{+0.72}_{-0.64} \times 10^{-10}$. This corresponds to a signal yield of $34 \pm 23$ candidates. The profile likelihoods are shown in Fig. 4.

In summary, a search for the rare decay $K_S^0 \rightarrow \mu^+\mu^-$ has been performed on a LHCb dataset of about 8.6 fb$^{-1}$. These results supersede those of our previous publications \[12,27\]. The data are consistent both with the background-only hypothesis and the combined background and SM signal expectation at the 1.4\(\sigma\) and 1.3\(\sigma\) level respectively. The most stringent upper limit on the $K_S^0 \rightarrow \mu^+\mu^-$ branching fraction to date of $2.1 (2.4) \times 10^{-10}$ at 90 (95)% confidence level is set, improving by a factor of four the previous best limit.

We would like to thank M. Moulson, J.M. Camalich, and G. D’Ambrosio for fruitful discussions.
Figure 4: $-2\Delta \log \mathcal{L}$, where $\mathcal{L}$ is the profile likelihood vs $B(K_0^0 \to \mu^+\mu^-)$. The blue line corresponds to the Run 2 result, the orange line to the Run 1 result and the green line shows the combination. The vertical line shows the location of the upper limit of the combined result at 90% confidence level.

References

[7] E137 collaboration, T. Akagi et al., Experimental study of the rare decays $K_L^0 \to \mu e$, $K_L^0 \to ee$, $K_L^0 \to \mu \mu$ and $K_L^0 \to eee$, Phys. Rev. D51 (1995) 2061.


