Search for dark photons produced in 13 TeV $pp$ collisions

LHCb collaboration

Abstract

Searches are performed for both prompt-like and long-lived dark photons, $A'$, produced in proton-proton collisions at a center-of-mass energy of 13 TeV, using $A' \rightarrow \mu^+ \mu^-$ decays and a data sample corresponding to an integrated luminosity of 1.6 fb$^{-1}$ collected with the LHCb detector. The prompt-like $A'$ search covers the mass range from near the dimuon threshold up to 70 GeV, while the long-lived $A'$ search is restricted to the low-mass region $214 < m(A') < 350$ MeV. No evidence for a signal is found, and 90% confidence level exclusion limits are placed on the $\gamma-A'$ kinetic-mixing strength. The constraints placed on prompt-like dark photons are the most stringent to date for the mass range $10.6 < m(A') < 70$ GeV, and are comparable to the best existing limits for $m(A') < 0.5$ GeV. The search for long-lived dark photons is the first to achieve sensitivity using a displaced-vertex signature.
The possibility that dark matter particles may interact via unknown forces, felt only feebly by Standard Model (SM) particles, has motivated substantial effort to search for dark-sector forces (see Ref. 1 for a review). A compelling dark-force scenario involves a massive dark photon, $A'$, whose coupling to the electromagnetic current is suppressed relative to that of the ordinary photon, $\gamma$, by a factor of $\varepsilon$. In the minimal model, the dark photon does not couple directly to charged SM particles; however, a coupling may arise via kinetic mixing between the SM hypercharge and $A'$ field strength tensors \cite{2,7}. This mixing provides a potential portal through which dark photons may be produced if kinematically allowed. If the kinetic mixing arises due to processes whose amplitudes involve one or two loops containing high-mass particles, perhaps even at the Planck scale, then $10^{-12} \lesssim \varepsilon^2 \lesssim 10^{-4}$ is expected \cite{1}. Fully exploring this few-loop range of kinetic-mixing strength is an important goal of dark-sector physics.

Constraints have been placed on visible $A'$ decays by previous beam-dump \cite{7, 21}, fixed-target \cite{22, 24}, collider \cite{25, 25}, and rare-meson-decay \cite{29, 38} experiments. The few-loop region is ruled out for dark photon masses $m(A') \lesssim 10$ MeV ($c = 1$ throughout this Letter). Additionally, the region $\varepsilon^2 \gtrsim 5 \times 10^{-7}$ is excluded for $m(A') < 10.2$ GeV, along with about half of the remaining few-loop region below the dimuon threshold. Many ideas have been proposed to further explore the \cite{39, 51} parameter space, including an inclusive search for $A' \rightarrow \mu^+\mu^-$ decays with the LHCb experiment, which is predicted to provide sensitivity to large regions of otherwise inaccessible parameter space using data to be collected during Run 3 of the LHC (2021–2023) \cite{52}.

A dark photon produced in proton-proton, $pp$, collisions via $\gamma-A'$ mixing inherits the production mechanisms of an off-shell photon with $m(\gamma^*) = m(A')$; therefore, both the production and decay kinematics of the $A' \rightarrow \mu^+\mu^-$ and $\gamma^* \rightarrow \mu^+\mu^-$ processes are identical. Furthermore, the expected $A' \rightarrow \mu^+\mu^-$ signal yield is given by \cite{52}

$$n_{ex}^{A'}[m(A'), \varepsilon^2] = \varepsilon^2 \left[ \frac{n_{ob}^{\gamma^*}[m(A')]}{2\Delta m} \right] \mathcal{F}[m(A')][\varepsilon_\gamma^*][m(A'), \tau(A')],$$

(1)

where $n_{ob}^{\gamma^*}[m(A')]$ is the observed prompt $\gamma^* \rightarrow \mu^+\mu^-$ yield in a small $\pm \Delta m$ window around $m(A')$, the function $\mathcal{F}[m(A')]$ includes phase-space and other known factors, and $\varepsilon_\gamma^*[m(A'), \tau(A')]$ is the ratio of the $A' \rightarrow \mu^+\mu^-$ and $\gamma^* \rightarrow \mu^+\mu^-$ detection efficiencies, which depends on the $A'$ lifetime, $\tau(A')$. If $A'$ decays to invisible final states are negligible, then $\tau(A') \propto [m(A')\varepsilon^2]^{-1}$ and $A' \rightarrow \mu^+\mu^-$ decays can potentially be reconstructed as displaced from the primary $pp$ vertex (PV) when the product $m(A')\varepsilon^2$ is small. When $\tau(A')$ is small compared to the experimental resolution, $A' \rightarrow \mu^+\mu^-$ decays are reconstructed as prompt-like and are experimentally indistinguishable from prompt $\gamma^* \rightarrow \mu^+\mu^-$ production, resulting in $\varepsilon_\gamma^*[m(A'), \tau(A')] \approx 1$. This facilitates a fully data-driven search and the cancelation of most experimental systematic effects, since the observed $A' \rightarrow \mu^+\mu^-$ yields, $n_{ob}^{A'}[m(A')]$, can be normalized to $n_{ex}^{A'}[m(A'), \varepsilon^2]$ to obtain constraints on $\varepsilon^2$.

This Letter presents searches for both prompt-like and long-lived dark photons produced in $pp$ collisions at a center-of-mass energy of 13 TeV, using $A' \rightarrow \mu^+\mu^-$ decays and a data sample corresponding to an integrated luminosity of 1.6 fb$^{-1}$ collected with the LHCb detector in 2016. The prompt-like $A'$ search is performed from near the dimuon threshold up to 70 GeV, above which the $m(\mu^+\mu^-)$ spectrum is dominated by the $Z$ boson. The long-lived $A'$ search is restricted to the mass range $214 < m(A') < 350$ MeV, where the data sample potentially provides sensitivity.
The LHCb detector is a single-arm forward spectrometer covering the pseudorapidity range \( 2 < \eta < 5 \), described in detail in Refs. \([53,54]\). Simulated data samples, which are used to validate the analysis, are produced using the software described in Refs. \([55,57]\). The online event selection is performed by a trigger \([58]\), which consists of a hardware stage using information from the calorimeter and muon systems, followed by a software stage, which performs a full event reconstruction. At the hardware stage, events are required to have a muon with \( p_T > 1.8 \text{ GeV} \), where \( p_T \) is the momentum transverse to the beam direction, or a dimuon in which the product of the \( p_T \) of each muon is in excess of \((\approx 1.5 \text{ GeV})^2\). The long-lived \( A' \) search also uses events selected at the hardware stage independently of the \( A' \rightarrow \mu^+\mu^- \) candidate. In the software stage, \( A' \rightarrow \mu^+\mu^- \) candidates are built from two oppositely charged tracks that form a good quality vertex and satisfy stringent muon-identification criteria. The muons are required to have \( 2 < \eta < 4.5 \), \( p_T > 0.5 \) \((1.0) \text{ GeV} \), momentum \( p > 10 \) \((20) \text{ GeV} \), and be inconsistent \((\text{consistent})\) with originating from the PV in the long-lived \((\text{prompt-like})\) \( A' \) search. Finally, the \( A' \) candidates are required to satisfy \( p_T > 1 \text{ GeV} \), \( 2 < \eta < 4.5 \), and have a decay topology consistent with originating from the PV.

The prompt-like \( A' \) search is based on a data sample where all online-reconstructed particles are stored, but most lower-level information is discarded, greatly reducing the event size. This data-storage strategy, made possible by advances in the LHCb data-taking scheme introduced in 2015 \([59,60]\), permits the recording of all events that contain a prompt-like dimuon candidate without placing any requirements on \( m(\mu^+\mu^-) \). The \( m(\mu^+\mu^-) \) spectrum recorded by the trigger is provided in the Supplemental Material to this Letter \([61]\).

Three main types of background contribute to the prompt-like \( A' \) search: prompt \( \gamma^* \rightarrow \mu^+\mu^- \) production, which is irreducible; resonant decays to \( \mu^+\mu^- \), whose mass-peak regions are avoided in the search; and various types of misreconstruction. The misreconstruction background consists of three dominant contributions: double misidentification of prompt hadrons as muons, \( hh \); a misidentified prompt hadron combined with a muon produced in a decay of a hadron containing a heavy-flavor quark, \( Q \), where the muon is misreconstructed as prompt-like, \( h\mu_Q \); and the misreconstruction of two muons produced in \( Q \)-hadron decays, \( \mu_Q\mu_Q \). These backgrounds are highly suppressed by the stringent muon-identification and prompt-like requirements applied in the trigger; however, in the region \([m(\phi),m(T)]\), the misreconstructed backgrounds overwhelm the signal-like \( \gamma^* \rightarrow \mu^+\mu^- \) contribution.

For masses below \((\text{above})\) the \( \phi \) meson mass, dark photons (expected to be predominantly produced in meson-decay \((\text{Drell-Yan})\) processes in \( pp \) collisions at LHCb). A well-known signature of Drell-Yan production is dimuons that are largely isolated, and a high-mass dark photon would inherit this property. The signal sensitivity is enhanced by applying a jet-based isolation requirement for \( m(A') > m(\phi) \), which improves the sensitivity by up to a factor of two at low masses and by \( \mathcal{O}(10\%) \) for \( m(A') > 10 \text{ GeV} \). Jet reconstruction is performed by clustering charged and neutral particle-flow candidates \([62]\) using the anti-\(k_T\) clustering algorithm \([63]\) with \( R = 0.5 \) as implemented in \textsc{FastJet} \([64]\). Muons with \( p_T(\mu)/p_T(\text{jet}) < 0.7 \) are rejected, where the contribution to \( p_T(\text{jet}) \) from the other muon is excluded if both muons are clustered in the same jet, as this is found to provide nearly optimal sensitivity for all \( m(A') > m(\phi) \). Figure 1 shows the resulting prompt-like \( m(\mu^+\mu^-) \) spectrum using \( \Delta m \) bins that are \( \sigma[m(\mu^+\mu^-)]/2 \) wide, where \( \sigma[m(\mu^+\mu^-)] \) is the mass resolution which varies from about 0.7 \text{ MeV} \) near threshold.
The prompt-like $A'$ search strategy involves determining the observed $A' \rightarrow \mu^+\mu^-$ yields from fits to the $m(\mu^+\mu^-)$ spectrum, and normalizing them using Eq. [1] to obtain constraints on $\varepsilon^2$. To determine $n_{\text{ob}}^*\{m(A')\}$ for use in Eq. [1], binned extended maximum likelihood fits are performed using the dimuon vertex-fit quality, $\chi^2_{\text{VF}}(\mu^+\mu^-)$, and $\min[\chi^2_{\text{IP}}(\mu^\pm)]$ distributions, where $\chi^2_{\text{IP}}(\mu)$ is defined as the difference in $\chi^2_{\text{VF}}(\text{PV})$ when the PV is reconstructed with and without the muon track. The $\chi^2_{\text{VF}}(\mu^+\mu^-)$ and $\min[\chi^2_{\text{IP}}(\mu^\pm)]$ fits are performed independently at each mass, with the mean of the $n_{\text{ob}}^*\{m(A')\}$ results used as the nominal value and half the difference assigned as a systematic uncertainty.

Both fit quantities are built from features that approximately follow $\chi^2$ probability density functions (PDFs) with minimal mass dependence. The prompt-dimuon PDFs are taken directly from data at $m(J/\psi)$ and $m(Z)$, where prompt resonances are dominant (see Fig. [1]). Small $p_T$-dependent corrections are applied to obtain the PDFs at all other masses. These PDFs are validated near threshold, at $m(\phi)$, and at $m(T(1S))$, where the data predominantly consist of prompt dimuons. The sum of the $h\bar{h}$ and $h\mu_Q$ contributions, which each involve misidentified prompt hadrons, is determined using same-sign $\mu^+\mu^+$ contributions, as the nominal value and half the difference assigned as a systematic uncertainty.

The prompt-like $A'$ mass spectrum is scanned in steps of $\sigma[m(\mu^+\mu^-)]/2$ searching for $A' \rightarrow \mu^+\mu^-$ contributions. At each mass, a binned extended maximum likelihood fit is performed using all prompt-like candidates in a $\pm12.5\sigma[m(\mu^+\mu^-)]$ window around $m(A')$. The profile likelihood is used to determine the $p$-value and the confidence interval for
Figure 2: Regions of the \([m(A'), \varepsilon^2]\) parameter space excluded at 90% CL by the prompt-like \(A'\) search compared to the best existing limits \cite{27,38}.

The most significant excess is \(3.3\sigma\) at \(m(A') \approx 5.8\) GeV, corresponding to a \(p\)-value of 38% after accounting for the trials factor due to the number of prompt-like signal hypotheses.

Regions of the \([m(A'), \varepsilon^2]\) parameter space where the upper limit on \(n_{\text{ex}}^{A'}[m(A')]\) is less than \(n_{\text{ob}}^{A'}[m(A'), \varepsilon^2]\) are excluded at 90% CL. Figure 2 shows that the constraints placed on prompt-like dark photons are comparable to the best existing limits below 0.5 GeV, and are the most stringent for \(10.6 < m(A') < 70\) GeV. In the latter mass range, a nonnegligible model-dependent mixing with the \(Z\) boson introduces additional kinetic-mixing parameters altering Eq. 1; however, the expanded \(A'\) model space is highly constrained by precision electroweak measurements. This search adopts the parameter values suggested in Refs. \cite{67,68}. The LHCb detector response is found to be independent of which quark-annihilation process produces the dark photon above 10 GeV, making it easy to recast the results in Fig. 2 for other models.

For the long-lived dark photon search, the stringent criteria applied in the trigger make contamination from prompt muon candidates negligible. The dominant background contributions to the long-lived \(A'\) search are as follows: photon conversions to \(\mu^+\mu^-\) in the silicon-strip vertex detector (the VELO) that surrounds the \(pp\) interaction region \cite{69}; \(b\)-hadron decays where two muons are produced in the decay chain; and the low-mass tail from \(K^0_s \to \pi^+\pi^-\) decays, where both pions are misidentified as muons. Additional sources of background are negligible, \(e.g.\) kaon and hyperon decays, and \(Q\)-hadron decays producing a muon and a hadron that is misidentified as a muon.

Photon conversions in the VELO dominate the long-lived data sample at low masses. A
new method, which is described in detail in Ref. [70], was recently developed for identifying particles created in secondary interactions with the VELO material. A high-precision three-dimensional material map was produced from a data sample of secondary hadronic interactions. Using this material map, along with properties of the \( A' \rightarrow \mu^+\mu^- \) decay vertex and muon tracks, a \( p \)-value is assigned to the photon-conversion hypothesis for each long-lived \( A' \rightarrow \mu^+\mu^- \) candidate. A mass-dependent requirement is applied to these \( p \)-values that reduces the expected photon-conversion yields to a negligible level.

A characteristic signature of muons produced in \( b \)-hadron decays is the presence of additional displaced tracks. Events are rejected if they are selected by the inclusive \( Q \)-hadron software trigger [71] independently of the presence of the \( A' \rightarrow \mu^+\mu^- \) candidate. Furthermore, two boosted decision tree (BDT) classifiers, originally developed for studying \( B^{0(s)} \rightarrow \mu^+\mu^- \) decays [72], are used to identify other tracks in the event that are consistent with having originated from the same \( b \)-hadron decay as the signal muon candidates. The requirements placed on the BDT responses, which are optimized using a data sample of \( K^0_S \) decays as a signal proxy, reject 70% of the \( b \)-hadron background at a cost of about 10% loss in signal efficiency.

As in the prompt-like \( A' \) search, the normalization is based on Eq. [1] however, in the long-lived \( A' \) search, \( \epsilon_{\gamma}\gamma'[m(A'), \tau(A')] \) is not unity, in part because the efficiency depends on the decay time, \( t \). Furthermore, the looser kinematic, muon-identification, and hardware-trigger requirements applied to long-lived \( A' \rightarrow \mu^+\mu^- \) candidates, cf. prompt-like candidates, increase the efficiency by a factor of 7 to 10, ignoring \( t \)-dependent effects. These \( m(A') \)-dependent factors are determined using a small control data sample of dimuon candidates consistent with originating from the PV, but otherwise satisfying the long-lived criteria. A relative 10% systematic uncertainty is assigned to the long-lived \( A' \rightarrow \mu^+\mu^- \) normalization due to background contamination in the control sample.

The fact that the kinematics are identical for \( A' \rightarrow \mu^+\mu^- \) and prompt \( \gamma^* \rightarrow \mu^+\mu^- \) decays for \( m(A') = m(\gamma^*) \) enables the \( t \) dependence of the signal efficiency to be determined using a data-driven approach. For each value of \( [m(A'), \tau(A')] \), prompt \( \gamma^* \rightarrow \mu^+\mu^- \) candidates in the control data sample near \( m(A') \) are resampled many times as long-lived \( A' \rightarrow \mu^+\mu^- \) decays, and all \( t \)-dependent properties, e.g. \( \min(\chi^2_{DF}(\mu^\pm)) \), are recalculated based on the resampled decay-vertex locations. This approach is validated in simulation by using prompt \( A' \rightarrow \mu^+\mu^- \) decays to predict the properties of long-lived \( A' \rightarrow \mu^+\mu^- \) decays, and based on these studies a 2% systematic uncertainty is assigned to the signal efficiencies. The \( \epsilon_{\gamma}\gamma'[m(A'), \tau(A')] \) values integrated over \( t \) are provided in Ref. [61].

A scan is again performed in discrete steps of \( \sigma[m(\mu^+\mu^-)]/2 \) looking for \( A' \rightarrow \mu^+\mu^- \) contributions; however, in this case, discrete steps in \( \tau(A') \) are also considered. Binned extended maximum likelihood fits are performed using all long-lived candidates and the three-dimensional feature space of \( m(\mu^+\mu^-) \), \( t \), and the consistency of the decay topology as quantified in the decay-fit \( \chi^2_{DF} \), which has three degrees of freedom (the data distribution is provided in Ref. [61]). The expected conversion contribution is derived in each bin from the number of candidates rejected by the conversion criterion. Two large control data samples are used to develop and validate the modeling of the \( b \)-hadron and \( K^0_S \) contributions: candidates that fail the \( b \)-hadron suppression requirements, and candidates that fail but nearly satisfy the muon-identification requirements. The profile likelihood is used to obtain the \( p \)-values and confidence intervals on \( n_{\text{obs}}[m(A'), \tau(A')] \). The most significant excess occurs at \( m(A') = 239 \text{ MeV} \) and \( \tau(A') = 0.86 \text{ ps} \), where the \( p \)-value corresponds to 3.0\( \sigma \). Considering only the long-lived-search trials factor reduces this to
Figure 3: Ratio of the observed upper limit on $n_{\text{ob}}^\gamma[m(A'),\varepsilon^2]$ to its expected value, where regions less than unity are excluded. There are no constraints from previous experiments in this region.

2.0$\sigma$. More details about these fits are provided in Ref. [61].

Under the assumption that $A'$ decays to invisible final states are negligible, there is a fixed (and known) relationship between $\tau(A')$ and $\varepsilon^2$ at each mass [52]; therefore, the upper limits on $n_{\text{ob}}^\gamma[m(A'),\tau(A')]$ can be translated into limits on $n_{\text{ob}}^\gamma[m(A'),\varepsilon^2]$. Regions of the $[m(A'),\varepsilon^2]$ parameter space where the upper limit on $n_{\text{ob}}^\gamma[m(A'),\varepsilon^2]$ is less than $n_{\text{ex}}^\gamma[m(A'),\varepsilon^2]$ are excluded at 90% CL (see Fig. [3]). While only small regions of $[m(A'),\varepsilon^2]$ space are excluded, a sizable portion of this parameter space will soon become accessible as more data are collected.

In summary, searches are performed for both prompt-like and long-lived dark photons produced in $pp$ collisions at a center-of-mass energy of 13 TeV, using $A'\to\mu^+\mu^-$ decays and a data sample corresponding to an integrated luminosity of 1.6 fb$^{-1}$ collected with the LHCb detector during 2016. The prompt-like $A'$ search covers the mass range from near the dimuon threshold up to 70 GeV, while the long-lived $A'$ search is restricted to the low-mass region $214 < m(A') < 350$ MeV. No evidence for a signal is found, and 90% CL exclusion regions are set on the $\gamma-A'$ kinetic-mixing strength. The constraints placed on prompt-like dark photons are the most stringent to date for the mass range $10^{.6} < m(A') < 70$ GeV, and are comparable to the best existing limits for $m(A') < 0.5$ GeV. The search for long-lived dark photons is the first to achieve sensitivity using a displaced-vertex signature.

These results demonstrate the unique sensitivity of the LHCb experiment to dark photons, even using a data sample collected with a trigger that is inefficient for low-mass $A'\to\mu^+\mu^-$ decays. Using knowledge gained from this analysis, the software-trigger efficiency for low-mass dark photons has been significantly improved for 2017 data taking. Looking forward to Run 3, the planned increase in luminosity and removal of the hardware-trigger stage should increase the number of expected $A'\to\mu^+\mu^-$ decays in the low-mass region by a factor of $\mathcal{O}(100-1000)$ compared to the 2016 data sample.
Acknowledgements

We express our gratitude to our colleagues in the CERN accelerator departments for the excellent performance of the LHC. We thank the technical and administrative staff at the LHCb institutes. We acknowledge support from CERN and from the national agencies: CAPES, CNPq, FAPERJ and FINEP (Brazil); MOST and NSFC (China); CNRS/IN2P3 (France); BMBF, DFG and MPG (Germany); INFN (Italy); NWO (The Netherlands); MNiSW and NCN (Poland); MEN/IFA (Romania); MinES and FASO (Russia); MinECo (Spain); SNSF and SER (Switzerland); NASU (Ukraine); STFC (United Kingdom); NSF (USA). We acknowledge the computing resources that are provided by CERN, IN2P3 (France), KIT and DESY (Germany), INFN (Italy), SURF (The Netherlands), PIC (Spain), GridPP (United Kingdom), RRCKI and Yandex LLC (Russia), CSCS (Switzerland), IFIN-HH (Romania), CBPF (Brazil), PL-GRID (Poland) and OSC (USA). We are indebted to the communities behind the multiple open-source software packages on which we depend. Individual groups or members have received support from AvH Foundation (Germany), EPLANET, Marie Sklodowska-Curie Actions and ERC (European Union), ANR, Labex P2IO, ENIGMASS and OCEVU, and Région Auvergne-Rhône-Alpes (France), RFBR and Yandex LLC (Russia), GVA, XuntaGal and GENCAT (Spain), Herchel Smith Fund, the Royal Society, the English-Speaking Union and the Leverhulme Trust (United Kingdom).

References


[61] See Supplemental Material to this Letter for additional fit details and figures.


[70] Mapping the material in the LHCb vertex locator using secondary hadronic interactions, in preparation.


Supplemental Material

Prompt-Like Fits

The fit strategy denoted by aic-o and described in detail in Ref. [65] is used in the prompt-like $A'$ search. The $m(\mu^+\mu^-)$ spectrum is scanned in steps of $\sigma[m(\mu^+\mu^-)]/2$ searching for $A' \to \mu^+\mu^-$ contributions. At each mass, a binned extended maximum likelihood fit is performed, and the profile likelihood is used to determine the $p$-value and the confidence interval on $n_{\text{ob}}^A[m(A')]$. The prompt-like-search trials factor is obtained using pseudoexperiments. As in Ref. [65], each fit is performed in a $\pm 12.5\sigma[m(\mu^+\mu^-)]$ window around the scan-mass value using bins with widths of $\sigma[m(\mu^+\mu^-)]/20$. Near threshold, the quantity $g(\mu^+\mu^-) \equiv \sqrt{m(\mu^+\mu^-)^2 - 4m(\mu)^2}$ is used instead of the mass since it is easier to model. The confidence intervals are defined using the bounded likelihood approach, which involves taking $\Delta \log \mathcal{L}$ relative to zero signal, rather than the best-fit value, if the best-fit signal value is negative. This approach enforces that only physical (nonnegative) upper limits are placed on $n_{\text{ob}}^A[m(A')]$, and prevents defining exclusion regions that are much better than the experimental sensitivity in cases where a large deficit in the background yield is observed.

The signal models are determined at each $m(A')$ using a combination of simulated $A' \to \mu^+\mu^-$ decays and the widths of the large resonance peaks that are clearly visible in the data. The background models are chosen following the method of Ref. [65]. This method takes as input a large set of potential background components, then performs a data-driven model-selection process whose uncertainty is included in the profile likelihood following Ref. [66]. In this analysis, the set of possible background components includes all Legendre modes with $\ell \leq 10$ at every $m(A')$. Additionally, dedicated background components are included to model the near-threshold turn-on behavior and all sizable known resonance contributions.

The use of 11 Legendre modes adequately describes every double-misidentified peaking background that contributes at a significant level, e.g., $\phi \to K^+K^-$ and $D \to K^+\pi^+$ double misidentified as dimuons, and in the $D$ case misreconstructed as prompt-like, do not require dedicated background components. In mass regions where such complexity is not required, the data-driven model-selection procedure reduces the complexity which increases the sensitivity to a potential signal contribution. As in Ref. [65], all fit regions are transformed onto the interval $[-1, 1]$, where the scan $m(A')$ value maps to zero. After such a transformation, the signal model is (approximately) an even function; therefore, odd Legendre modes are orthogonal to the signal component, which means that the presence of odd modes has minimal impact on the variance of $n_{\text{ob}}^A[m(A')]$. In the prompt-like fits, all odd Legendre modes up to ninth order are included in every background model, while only a subset of the even modes is selected for inclusion in each fit.

Regions in the mass spectrum where large known resonance contributions are observed are vetoed in the prompt-like $A'$ search. Furthermore, the regions near the $\eta'$ meson and the excited $\Upsilon$ states (beyond the $\Upsilon(4S)$ meson) are treated specially. For example, since it is not possible to distinguish between $A' \to \mu^+\mu^-$ and $\eta' \to \mu^+\mu^-$ contributions at $m(\eta')$, the $p$-values near this mass are ignored. Any excess at $m(\eta')$ is treated as signal when setting the limits on $n_{\text{ob}}^A[m(A')]$, which is conservative in that a $\eta' \to \mu^+\mu^-$ contribution will weaken the constraints on $A' \to \mu^+\mu^-$ decays. The same strategy is used near the excited $\Upsilon$ masses. The treatment of all mass regions is summarized in Table 1.
The long-lived signal yields are determined from binned extended maximum likelihood fits performed on all long-lived $A' \rightarrow \mu^+\mu^-$ candidates using the three-dimensional feature space of the dimuon invariant mass, $m(\mu^+\mu^-)$, the $A'$ decay time, $t$, and the decay-fit quality, $\chi^2_{DF}$. As in the prompt-like $A'$ search, a scan is performed in discrete steps of $\sigma[m(\mu^+\mu^-)]/2$; however, in this case, discrete steps in $\tau(A')$ are also considered. The profile likelihood is again used to obtain the $p$-values and the confidence intervals on $n_{\text{ob}}(m(A'), \tau(A'))$. The binning scheme involves four bins in $\chi^2_{DF}$: [0,2], [2,4], [4,6], and [6,8]. Eight bins in $t$ are used: [0,2,0.6], [0.6,1.1], [1.1,1.6], [1.6,2.2], [2.2,3.0], [3.5], [5,10], and $>10$ ps. The binning scheme used for $m(\mu^+\mu^-)$ depends on the scan $m(A')$ value, and is chosen such that the majority of the signal falls into a single bin. Signal decays mostly have small $\chi^2_{DF}$ values, with about 50% (80%) of $A' \rightarrow \mu^+\mu^-$ decays satisfying $\chi^2_{DF} < 2$ (4). Background from $b$-hadron decays populates the small $t$ region and is roughly uniformly distributed in $\chi^2_{DF}$, whereas background from $K_s^0$ decays is signal-like in $\chi^2_{DF}$ and roughly uniformly distributed in $t$. Figure 4 shows the three-dimensional distribution of all long-lived $A' \rightarrow \mu^+\mu^-$ candidates.

The expected contribution in each bin from photon conversions is derived from the number of candidates rejected by the conversion criterion. As discussed in the Letter, two large control data samples are used to develop and validate the modeling of the $b$-hadron and $K_s^0$ contributions. Both contributions are well modeled by the function
\[
\Theta[g(\mu^+\mu^-) - q_0] \times \{a[g(\mu^+\mu^-) - q_0] + b[g(\mu^+\mu^-) - q_0]^2\},
\]
where $q_0$, $a$, and $b$ are fitted to the data, and $\Theta$ denotes the Heaviside step function. While no evidence for $t$ or $\chi^2_{DF}$ dependence is observed for these parameters in either the $b$-hadron or $K_s^0$ control sample, all parameters are allowed to vary independently in each $[t, \chi^2_{DF}]$ region in the fits used in the long-lived $A'$ search.

Figure 3 shows the long-lived $A' \rightarrow \mu^+\mu^-$ candidates, along with the pull values obtained from fits performed to the data where no signal contributions are included. All of the pulls are in the range $[-2,2]$. N.b., due to the fact that the background threshold parameters are free to vary in each $[t, \chi^2_{DF}]$ region, the lowest-mass nonempty bin for each $[t, \chi^2_{DF}]$ is biased towards a positive pull in the absence of a signal contribution.
Figure 4: Three-dimensional distribution of $\chi^2_{DF}$ versus $t$ versus $m(\mu^+\mu^-)$, which is fit to determine the long-lived signal yields. The data are consistent with being predominantly due to $b$-hadron decays at small $t$, and due to $K^0_S$ decays for large $t$ and $m(\mu^+\mu^-) \gtrsim 280$ MeV. The largest signal-like excess occurs at $m(A') = 239$ MeV and $\tau(A') = 0.86$ ps.

Figure 5: Efficiency ratio $\epsilon_{A'}^2[m(A'), \tau(A')]$ for long-lived dark photons, integrated over decay time. The sharp decrease at larger values of $\epsilon^2$ is due to the stringent $\min[\chi^2_{IP}(\mu^\pm)]$ criterion applied in the 2016 trigger.
Figure 6: Long-lived $A' \rightarrow \mu^+\mu^-$ candidates (black points) showing $t$ versus $m(\mu^+\mu^-)$ in bins of $\chi^2_{DF}$, compared to the pulls from binned fits performed without a signal contribution (color axis). Positive pulls denote an excess of data candidates.
Additional Figures

Figure 7: Mass spectrum selected by the prompt-like $A' \rightarrow \mu^+ \mu^-$ trigger.

Figure 8: Example $\chi^2_{IP}(\mu_{\pm})^{1/2}$ distributions with fit results overlaid for prompt-like candidates near (left) $m(A') = 0.5$, (middle) 5, and (right) 50 GeV. The square root of $\chi^2_{IP}(\mu_{\pm})$ is used in the fits to increase the bin occupancies at large $\chi^2_{IP}(\mu_{\pm})$ values.

Figure 9: Comparison of the results presented in this Letter to existing constraints from previous experiments (see Ref. [1] for details about previous experiments).
Sezione INFN di Bologna, Bologna, Italy
Sezione INFN di Cagliari, Cagliari, Italy
Università e INFN, Ferrara, Ferrara, Italy
Sezione INFN di Firenze, Firenze, Italy
Laboratori Nazionali dell’INFN di Frascati, Frascati, Italy
Sezione INFN di Genova, Genova, Italy
Sezione INFN di Padova, Padova, Italy
Sezione INFN di Pisa, Pisa, Italy
Sezione INFN di Roma Tor Vergata, Roma, Italy
Sezione INFN di Roma La Sapienza, Roma, Italy
Henryk Niewodniczanski Institute of Nuclear Physics Polish Academy of Sciences, Kraków, Poland
AGH - University of Science and Technology, Faculty of Physics and Applied Computer Science, Kraków, Poland
National Center for Nuclear Research (NCBJ), Warsaw, Poland
Horia Hulubei National Institute of Physics and Nuclear Engineering, Bucharest-Magurele, Romania
Petersburg Nuclear Physics Institute (PNPI), Gatchina, Russia
Institute of Theoretical and Experimental Physics (ITEP), Moscow, Russia
Institute of Nuclear Physics, Moscow State University (SINP MSU), Moscow, Russia
Institute for Nuclear Research of the Russian Academy of Sciences (INR RAN), Moscow, Russia
Yandex School of Data Analysis, Moscow, Russia
Budker Institute of Nuclear Physics (SB RAS), Novosibirsk, Russia
Institute for High Energy Physics (IHEP), Protvino, Russia
ICCB, Universitat de Barcelona, Barcelona, Spain
Universidad de Santiago de Compostela, Santiago de Compostela, Spain
European Organization for Nuclear Research (CERN), Geneva, Switzerland
Institute of Physics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
Physik-Institut, Universität Zürich, Zürich, Switzerland
Nikhef National Institute for Subatomic Physics, Amsterdam, The Netherlands
Nikhef National Institute for Subatomic Physics and VU University Amsterdam, Amsterdam, The Netherlands
NSC Kharkiv Institute of Physics and Technology (NSC KIPT), Kharkiv, Ukraine
Institute for Nuclear Research of the National Academy of Sciences (KINR), Kyiv, Ukraine
University of Birmingham, Birmingham, United Kingdom
H.H. Wills Physics Laboratory, University of Bristol, Bristol, United Kingdom
Cavendish Laboratory, University of Cambridge, Cambridge, United Kingdom
Department of Physics, University of Cambridge, Cambridge, United Kingdom
STFC Rutherford Appleton Laboratory, Didcot, United Kingdom
School of Physics and Astronomy, University of Edinburgh, Edinburgh, United Kingdom
School of Physics and Astronomy, University of Glasgow, Glasgow, United Kingdom
Oliver Lodge Laboratory, University of Liverpool, Liverpool, United Kingdom
Imperial College London, London, United Kingdom
School of Physics and Astronomy, University of Manchester, Manchester, United Kingdom
Department of Physics, University of Oxford, Oxford, United Kingdom
Massachusetts Institute of Technology, Cambridge, MA, United States
University of Cincinnati, Cincinnati, OH, United States
University of Maryland, College Park, MD, United States
Syracuse University, Syracuse, NY, United States
Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), Rio de Janeiro, Brazil, associated to
University of Chinese Academy of Sciences, Beijing, China, associated to
School of Physics and Technology, Wuhan University, Wuhan, China, associated to
Institute of Particle Physics, Central China Normal University, Wuhan, Hubei, China, associated to
Departamento de Física, Universidad Nacional de Colombia, Bogota, Colombia, associated to
Institut für Physik, Universität Rostock, Rostock, Germany, associated to
National Research Centre Kurchatov Institute, Moscow, Russia, associated to
National Research Tomsk Polytechnic University, Tomsk, Russia, associated to
Instituto de Física Corpuscular, Centro Mixto Universidad de Valencia - CSIC, Valencia, Spain, associated to
Van Swinderen Institute, University of Groningen, Groningen, The Netherlands, associated to

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

†Deceased