ATLAS CONF Note

ATLAS-CONF-2023-041

29th August 2023

Search for narrow low-mass resonances in the four-muon final state with the ATLAS detector at the LHC

The ATLAS Collaboration

A search for new low-mass resonances in the four-muon final state in the four-muon invariant mass range from 10 − 50 GeV is performed using pp collision data recorded with the ATLAS detector at the Large Hadron Collider. The data correspond to an integrated luminosity of 20.3 fb^{-1} at a centre-of-mass energy ($\sqrt{s}$) of 8 TeV collected in 2012, and 51.5 fb^{-1} and 58.5 fb^{-1} collected at $\sqrt{s}$ = 13 TeV in 2015–2017 and 2018, respectively. An excess at 18 GeV in the invariant mass spectrum of the four-muon system of the selected $\Upsilon(1S)+\mu^+\mu^- \rightarrow \mu^+\mu^-\,\mu^+\mu^-$ final state is observed in the 8 TeV data. Although not blinded, typical selection criteria are used and variations studied to explore the stability of the excess, resulting in local significances of the excess of between be 3.6 σ and 6.3 σ corresponding to global significances of 1.9 σ to 5.4 σ. A baseline choice of selection cuts defined for the analysis at 8 TeV was defined in the 8 TeV data to be further probed at Run 2. Subsequent analysis of 13 TeV data finds a smaller excess at the same four-muon invariant mass with a significance of 1.9 σ in early 13 TeV data runs, while data collected with more restrictive trigger selections in the 2018 dataset finds no evidence of a signal. LHCb and CMS have observed no excess at this mass. Signal MC and di-$\Upsilon(1S)$ production studies in data are used to quantify performance across the three data-taking periods, finding reductions in expected sensitivity in the 13 TeV dataset compared to that at 8 TeV. The lack of observation of a signal in the 13 TeV is assessed to be in tension with the structure observed at an invariant mass of 18 GeV in the 8 TeV dataset, up to a maximum of 2.7σ. We subsequently set and report limits on the total production cross section times branching fraction to a four-muon final state for a new state at a mass of 18 GeV for a variety of models.

ATLAS-CONF-2023-041
04 September 2023

1 Introduction

New resonances decaying to four-lepton final states are predicted in theories of exotic mesons comprised of four valence quarks (tetraquark) and in a variety of extensions to the standard model which predict the existence of new fundamental particles. Tetraquark states of four heavy quarks [1–4] can decay to four muons, providing a clean signature for the search even in a busy hadron collider environment. In models with a hidden or dark sector [5–7] or in the next-to-minimal-supersymmetric standard model [8], a non-SM Higgs boson can decay to four leptons through intermediate SM or exotic gauge bosons [9], or may decay through vector quarkonium final states [10–13].

Many searches in the four-lepton (electron or muon) final states have been performed at the Large Hadron Collider (LHC), but these studies have largely focused on searches at high invariant mass. Recent studies from the LHCb [14] Collaboration (at 8 and 13 TeV centre-of-mass energy) and the CMS Collaboration [15] (at 13 TeV with 35.9 fb^{-1} of proton-proton collisions) in the $\Upsilon(1S) + \mu^+\mu^-$ final state have set limits on the production of potential resonances at low mass (< 27 GeV), while new results from the LHCb collaboration [16] observed the existence of a new state at ~ 6.9 GeV in the di-J/ψ final state, which is also confirmed by ATLAS [17]. The state can be interpreted as a fully-charmed tetraquark resonance.

This paper reports on a search for new low-mass resonances, X, decaying to four-muon (4μ) final states, via $\Upsilon(1S) + \mu^+\mu^-$, in pp collisions, making use of the full Run-1 and Run-2 datasets at centre-of-mass energies of 8 and 13 TeV, respectively. There are significant differences in the triggers available through the different periods of data taking, complicating the analysis and the interpretation of results. As a cross-check, a search is performed for the simultaneous production of $\Upsilon(1S)$ pairs, acting as a standard candle for performance, using the same datasets and applying the same selection criteria with the exception an additional requirement of a $\Upsilon(1S)$ mass window applied on the second muon pair.

2 Detector and triggers

The ATLAS detector [18, 19] is composed of an inner tracking system, calorimeters, and a muon spectrometer. The inner detector (ID) consists of silicon pixel and microstrip detectors, and a transition radiation tracker, all immersed in a 2 T axial magnetic field. The ID is enclosed by a system of electromagnetic and hadronic calorimeters. Surrounding the calorimeters is the muon spectrometer (MS) consisting of three large air-core superconducting magnets, a system of precision tracking chambers, and fast detectors for triggering. The most important detector systems for this search are the ID and MS. The ID provides precise track reconstruction within the pseudorapidity range $|\eta| < 2.5$[1]. The MS provides muon identification and precision measurements in the bending plane of muons within the pseudorapidity range $|\eta| < 2.7$.

An extensive software suite [20] is used in data simulation, in the reconstruction and analysis of real and simulated data, in detector operations, and in the trigger and data acquisition systems of the experiment.

[1] ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the centre of the detector and the z-axis along the beam pipe. The x-axis points from the IP to the centre of the LHC ring, and the y-axis points upward. Cylindrical coordinates (r, ϕ) are used in the transverse plane, ϕ being the azimuthal angle around the beam pipe. The pseudorapidity is defined in terms of the polar angle θ as $\eta = -\ln\tan(\theta/2)$. Rapidity is defined as $y = 0.5\ln[(E+p_z)/(E-p_z)]$ where E denotes the energy and p_z is the component of the momentum along the beam direction.

A multi-stage trigger system is employed to reduce the tens of megahertz of the bunch collision rate to about 400 hertz and 1 kilohertz for recording in 8 and 13 TeV data-taking, respectively. This analysis is based on level-1 muon triggers that search for hit coincidences between muon trigger detector layers of muon candidates above predetermined transverse momentum (p_{T}) thresholds. Subsequent software-based trigger selections with increased precision and information from both the MS and the ID approach the position and momentum resolution close to that of the offline event reconstruction.

Muons from the decays of low-mass resonances are expected to have a soft p_{T} spectrum. Triggering these events in a busy collision environment is challenging. For this search, events of interest are primarily recorded using triggers requiring two or three muon candidates with a minimum muon p_{T} of 4 GeV. The dimuon triggers are based on the identification of $J/\psi \to \mu^+\mu^-$, $B \to \mu^+\mu^-$ and $\Upsilon \to \mu^+\mu^-$ decays. The two muon candidates are required to have opposite charges and have their invariant masses $m_{\mu\mu}$ in the range of 2.5 – 12 GeV. There is no dimuon mass or muon charge requirement for the three-muon trigger in 8 TeV data or in 13 TeV data collected between 2015–17. For 2018 data-taking, due to data processing limitations, three-muon trigger selections additionally required that, of the three muon candidates identified by the trigger, a pair of opposite charge should have an invariant mass in the range 8 – 12 GeV.

3 Data and Monte Carlo samples

The search is performed using the proton-proton collision data collected with the ATLAS detector during LHC runs in 2012 at a centre-of-mass energy ($\sqrt{s}$) of 8 TeV and runs in 2015–18 at an energy of $\sqrt{s} = 13$ TeV. At $\sqrt{s} = 8$ TeV, the data used were recorded with a combination of the un-prescaled dimuon and three-muon triggers corresponding to a full integrated luminosity of 20.3 fb^{-1}. At $\sqrt{s} = 13$ TeV, there were no dimuon triggers with muon p_{T} thresholds of 4 GeV due to their prohibitively high rates, while the three-muon trigger was prescaled in 2015–17 and was switched off during 2018 data-taking, where this trigger selection was replaced with a more restrictive three-muon trigger. For the search at 13 TeV, the full Run-2 dataset covering the 2015–17 and 2018 datasets is used, corresponding to integrated luminosities of 51.5 fb^{-1} and 58.5 fb^{-1} after trigger prescales, respectively. All datasets have been reconstructed under the same conditions with the same software release.

Monte Carlo (MC) samples are used to aid the characterisation of the detector performance, to determine the detector resolution as a function of the four-muon mass, the development of the event selection and to help the understanding of background contributions. Beyond-the-standard-model simulations of the production of low-mass narrow-width ($\lesssim 1$ MeV) scalar or pseudoscalar tetraquark states and Higgs-like scalar states decaying to $\Upsilon(1\mathrm{S}) + \mu^+\mu^-$ are used solely for interpretation of the experimental results. Signal simulations for tetraquark models [21] are produced using MadGraph5_aMC@NLO v2.3 [22] at leading-order interfaced to Pythia 8.186 [23]. Signal simulations for the Higgs-like scalar are produced using the leading-order gluon-fusion process simulated in Pythia 8.186 interfaced to EvtGen [24]. Both sets of signal models use the A14 tune [25] and the `NNPDF2.3lo` [26] PDF set.

Backgrounds to the $X \to \Upsilon(1\mathrm{S}) + \mu^+\mu^-$ search are complex. Muons can be produced from decays of heavy quarks (b, c), tau leptons and vector bosons (W, Z/γ^*). Muon candidates can also originate from charged pion and kaon decays in flight, or from misidentification of other charged particles. Various MC generators [23, 27–31] were used to create simulations of $b\bar{b}$, $c\bar{c}$, $Z/\gamma^* + b\bar{b}$, $Z/\gamma^* + Z/\gamma^*$ production and double parton scattering interactions in which a $b\bar{b}$ pair and an $\Upsilon(1\mathrm{S})$ are created together from a single pp collision.

Generated events are passed through a GEANT 4 [32] simulation of the ATLAS detector [33]. The EVTGEN program [24] is used for simulating the properties of bottom and charm hadron decays. The simulated events are reconstructed with the same reconstruction software as used for the data including the effect of additional coincident pp interactions (pile-up) in the same or neighbouring bunch crossings. The MC samples used are reweighted to match the pile-up conditions observed in the data. These background samples are used to assess whether data selection criteria could introduce structures into the observed four-muon mass distributions and to validate data-driven approaches to the background modelling.

4 Muon identification and performance

Muon candidates are reconstructed from ID tracks combined with independent MS tracks or track segments [34]. A muon candidate identified using a combination of MS and ID tracks is referred to as a combined muon. A muon candidate formed from an MS track segment that is matched to an ID track extrapolated to the MS is referred to as a segment-tagged muon. The segment-tagged muons are intended to reconstruct low p_{T} muons that have insufficient energy to traverse all three MS layers and also to recover part of the inefficiency due to the limited coverage of the muon spectrometer.

Both combined and segment-tagged muons are used for this analysis. They are required to have transverse momenta $p_{\mathrm{T}} > 3$ GeV and to be in the pseudorapidity range of $|\eta| < 2.5$. The muon track parameters are taken from the ID measurement alone, since the precision of the measured track parameters for muons in the p_{T} range of interest is dominated by the ID tracks. The muon track must also be consistent with the hypothesis that they originate from the primary vertex, satisfying the requirements of $|z_0 \sin\theta| < 1$ mm and $|d_0/\sigma_{d_0}| < 6$. Here $z_0 \sin\theta$ is the longitudinal impact parameter of the ID track, while d_0 and σ_{d_0} are the transverse impact parameter and its uncertainty. All these vertex quantities are measured with respect to the associated primary vertex.

Combined muon candidates defined using the "LowPt" selection criteria [34–36] are required to have one MS station hit if within $|\eta| < 1.3$, two station hits if between $1.3 \leq |\eta| < 1.55$, and to pass "Medium" quality criteria if $|\eta| \geq 1.55$. Combined or segment-tagged muons failing the LowPt selection criteria provide additional acceptance for low invariant mass parent systems.

5 Event selection

Events are retained for analysis if they are recorded with all detector subsystems operating normally and have at least one reconstructed primary vertex (PV) formed from two or more associated ID tracks with $p_{\mathrm{T}} > 400$ MeV.

The $X \to \Upsilon(1\mathrm{S}) + \mu^+\mu^- \to \mu^+\mu^-\,\mu^+\mu^-$ selection begins with the identification of four muon candidates: the muon quadruplet. At least three of these muon candidates must satisfy the LowPt muon quality requirements. The sum of charges of the four muon candidates must be zero and the charge measurement in the ID and MS must be consistent for each combined muon. The four muon candidate trajectories are fit to a common vertex and must pass a $\chi^2/N_{\mathrm{d.o.f}} \leq 10$, $N_{\mathrm{d.o.f}} = 5$, criterion to be retained for further analysis. This loose criterion is imposed because of the presence of the long χ^2 tails observed in signal MC and reference quarkonia datasets at low muon p_T. The four-muon invariant mass $m_{4\mu}$, calculated using the track parameters from the four-muon vertex fit, must satisfy 10 GeV $\leq m_{4\mu} \leq$ 50 GeV.

Any two muons of the quadruplet are then fit to a common vertex. If the di-muon vertex fit satisfies a quality criterion of $\chi^2 < 3$ then the two muons are said to form a muon doublet. A muon can form multiple valid doublets with different muons. Doublets formed by muons of opposite charge (q_μ) are called OS (opposite sign) doublets, whereas those formed by muons of the same charge are called SS (same sign) doublets. OS doublets with 9.2 GeV $\leq m_{\mu^+\mu^-} \leq$ 9.7 GeV are identified as $\Upsilon(1S) \rightarrow \mu^+\mu^-$ candidates. Here the mass window corresponds to approximately ±2 times the detector resolution centred around the $\Upsilon(1S)$ mass [37].

The $\Upsilon(1S) + \mu^+\mu^-$ events are selected by requiring one $\Upsilon(1S) \rightarrow \mu^+\mu^-$ candidate and another OS doublet. The momenta of both doublets are required to point to a common PV, which is then defined as the primary vertex of the event. This vertex is the PV with the highest $\sum p_{\mathrm{T}}^2$ over all associated ID tracks in 98.3% and 96.7% of selected events in 8 and 13 TeV, respectively. The two muon candidates associated with the $\Upsilon(1S)$ decay are further required to have $p_{\mathrm{T}} > 4$ GeV, to satisfy online trigger requirements. No $\Upsilon(1S)$ mass constraint is applied in the $m_{4\mu}$ calculation. The second di-muon pair is required to have invariant mass $m_{\mu^+\mu^-} > 1$ GeV to reduce backgrounds from photon conversions. If two or more doublets satisfy this minimum di-muon invariant requirement (which occurs in 2 – 3% of events), the one with the smallest χ^2 is chosen. If more than one muon quadruplet satisfying the above criteria is present in an event, only the quadruplet with the best four-muon vertex χ^2 value is retained. This occurs in 6% of events in 8 TeV data, in 10% of events in 2015–17 13 TeV data, and in 8% of events in 2018 13 TeV data.

This defines the "baseline" selection of the $X \rightarrow \Upsilon(1S) + \mu^+\mu^-$ search and the key selection criteria are summarised in Table 1. Since the search covers a wide variety of potential signal models our selections are intentionally loose and guided by previous ATLAS publications studying low p_{T} phenomena and associated production of quarkonia [17, 38, 39].

Table 1: Criteria for the baseline event selection in the $X \rightarrow \Upsilon(1S) + \mu^+\mu^-$ search. See text for full details.

Candidate object	Requirements
Muons	$p_{\mathrm{T}}(\mu) > 3$ GeV and $\lvert\eta\rvert < 2.5$, $\lvert z_0 \sin\theta\rvert < 1$ mm and $\lvert d_0/\sigma_{d_0}\rvert < 6$
Muon quadruplet	≥ 3 muons passing LowPt selection criteria, $\sum q_\mu = 0$, four-muon vertex fit $\chi^2/N_{\mathrm{d.o.f}} \leq 10$, 10 GeV $\leq m_{4\mu} \leq$ 50 GeV
Muon doublet	di-muon vertex fit $\chi^2 < 3$
$\Upsilon(1S)$ candidate	OS muon doublet with $p_{\mathrm{T}}(\mu_{1,2}) > 4$ GeV, 9.2 GeV $\leq m_{\mu^+\mu^-} \leq$ 9.7 GeV
$\Upsilon(1S) + \mu^+\mu^-$ candidate events	$\Upsilon(1S)$ candidate plus OS muon doublet with $m_{\mu^+\mu^-} > 1$ GeV, both muon doublets point to a common PV

5.1 Analysis of 8 TeV data

The $m_{4\mu}$ distribution of the selected events in the range of 10 – 50 GeV in the 8 TeV dataset following the above selection is shown in Figure 1. The selection is mostly sensitive to this mass range due to the doublet mass requirements. Although the data analysis at 8 TeV was not blinded, typical selection criteria are applied and the stability and characteristics of the observed excess is confirmed by an extensive variation of the cuts.

A same-sign control sample, $\Upsilon(1S) + \mu^{\pm}\mu^{\pm}$, is selected by requiring one $\Upsilon(1S) \to \mu^{+}\mu^{-}$ candidate and another SS doublet with the baseline selection. The shape of the $m_{4\mu}$ distribution of these $\Upsilon(1S) + \mu^{\pm}\mu^{\pm}$ events is compared to that of the $\Upsilon(1S) + \mu^{+}\mu^{-}$ events in Figure 1 (top). Overall, the shape of the $m_{4\mu}$ distribution of the $\Upsilon(1S) + \mu^{+}\mu^{-}$ events is well reproduced by that of the $\Upsilon(1S) + \mu^{\pm}\mu^{\pm}$ events.

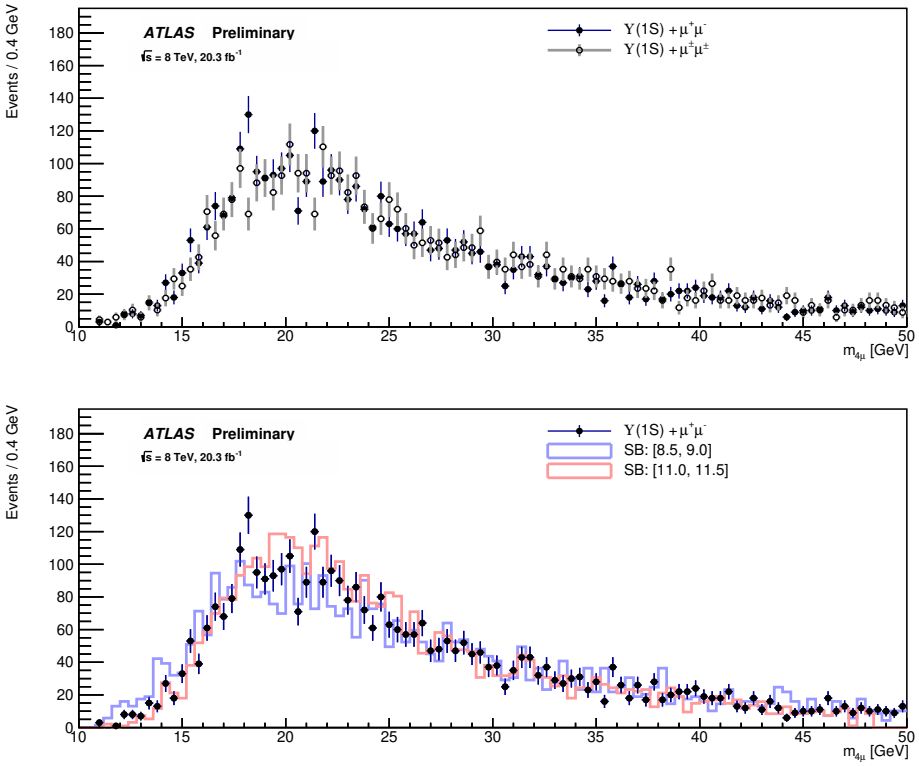

Figure 1: Comparisons of the $m_{4\mu}$ distribution of the selected $\Upsilon(1S) + \mu^{+}\mu^{-}$ events with those of data control samples at $\sqrt{s} = 8$ TeV: (top) the $\Upsilon(1S) + \mu^{\pm}\mu^{\pm}$ events and (bottom) events having an OS doublet with its mass in the $\Upsilon(n\mathrm{S})$ sidebands of 8.5 – 9.0 GeV and 11.0 – 11.5 GeV. In all cases, the numbers of events of the control sample are normalised to that of the $\Upsilon(1S) + \mu^{+}\mu^{-}$ events. The mass values of the muon quadruplet in the sidebands are shifted to match the range of the $\Upsilon(1S) \to \mu^{+}\mu^{-}$ events for comparison, see text.

To investigate the nature of the apparent excesses at 18 and 21 GeV, additional data control and background MC samples are examined to explore the possibility of whether or not artificial structures in the $m_{4\mu}$ distributions could be manufactured due to the event selection. This includes analysis of $\Upsilon(1S)$+di-track data and single-muon plus three-track data, progressively removing muon identification requirements in the muon quadruplet object reconstruction in Table 1. No indications of a source of artifical structures

were found in either data or MC. No attempt is made to understand precisely the relative background composition in data given its complexity.

Studies are performed for events in the $\Upsilon(1S) \to \mu^+\mu^-$ sidebands. Instead of requiring the mass of the $\Upsilon(1S) \to \mu^+\mu^-$ doublet in the window of 9.2 – 9.7 GeV, two sidebands are defined: a lower sideband of 8.5 – 9.0 GeV and an upper sideband of 11.0 – 11.5 GeV. These mass ranges are chosen to be away from the $\Upsilon(n\mathrm{S})$ resonances to minimise their contamination and to be away from the invariant mass thresholds of the dimuon triggers. All other selections are kept the same. There are inherent kinematic differences between the sideband and $\Upsilon(1S) + \mu^+\mu^-$ events, with the lower and upper sidebands being less and more energetic on average, respectively, than the search region. To better compare the overall shapes, the mass values of muon quadruplets are shifted[2] upwards by +0.7 GeV for events in the lower sideband and downwards by −1.8 GeV for events in the upper sideband. Figure 1 (bottom) compares the $m_{4\mu}$ distributions of the $\Upsilon(1S) + \mu^+\mu^-$ events and the sideband events. The numbers of events in the sidebands are normalised to that of the $\Upsilon(1S) + \mu^+\mu^-$ events in the comparison. The three distributions are characteristically similar and no statistically significant feature is visible in the sideband distributions.

The numbers of events in the $\Upsilon(1S) + \mu^\pm\mu^\pm$ and sideband control samples are limited. Additional studies are performed using events with two or three combined muons together with two or one ID tracks, respectively, without requirement that these tracks are associated with identified muons to increase the available statistics and assess if selection criteria can introduce structures in the four-muon invariant mass spectrum. In these events, ID tracks are used as muon replacements to form "muon" quadruplets. The tracks are subjected to the same selection as muons other than the muon identification requirements. Additional studies are performed using an event mixing technique. In this approach, muon doublets from independent proton-proton collision events passing the baseline selection cuts are combined to form "muon" quadruplets. The "four-muon" mass distributions of these high statistics data-driven control samples are examined and no statistically significant excesses are observed in these distributions.

The $\Upsilon(1S) + \mu^+\mu^-$ selection is also applied to the signal and background MC samples. Besides the gradual turn-on in the $m_{4\mu}$ distribution resulting from the muon p_T and $\Upsilon(1S)$ mass requirement, sharp turn-ons at twice the $\Upsilon(1S)$ mass caused by its pair production are seen in some of these samples. However, no clear indication of a manufactured peak is evident in any $m_{4\mu}$ distribution of these samples.

5.2 Analysis of 13 TeV data

The analysis of the 13 TeV data is used to provide an independent cross-check of the excess observed in the 8 TeV data. Thus, it follows closely the analysis at 8 TeV, with analysis selections fixed before unblinding of the 13 TeV data. The dimuon triggers with a p_T threshold of 4 GeV were unavailable at 13 TeV, while triggers with higher thresholds are found to be inefficient for events contributing to the four-muon invariant mass region near where the excess was observed in the 8 TeV data. Nevertheless a p_T threshold of 4 GeV was maintained for the three-muon trigger for part of the 13 TeV dataset, collected in 2015–17. Hence the search at 13 TeV is based on the data recorded with this trigger, corresponding to an integrated luminosity of 51.5 fb^{-1}. In 2018, a more restrictive trigger selection was applied, requiring two of the three muon candidates identified by the trigger to have opposite charge and have an invariant mass in the range 8 − 12 GeV. This trigger is hereafter referred to as "3μ_bUpsi". This 2018 dataset corresponds to an integrated luminosity of 58.5 fb^{-1} and is analysed separately.

[2] Shifts correspond to the average mass correction from sidebands to the $\Upsilon(1S)$ target mass.

Much of the study of the 13 TeV data prior to unblinding relates to the understanding of low p_{T} muon reconstruction and identification. Muon quadruplets are chosen as for the 8 TeV dataset. The numbers of events selected at a few key steps of the selection are compared to those at 8 TeV in Table 2 without and with the requirement of the three-muon trigger. Conditions at 13 TeV lead to enhanced combinatorial backgrounds as indicated in Table 2 which illustrates that there are more than twice as many quadruplets passing selection criteria per unit integrated luminosity in 13 TeV data than in 8 TeV data collected with equivalent triggers. The presence of larger combinatorial backgrounds not only increases background yields but also reduces the efficiency of selecting any true four muon signal.

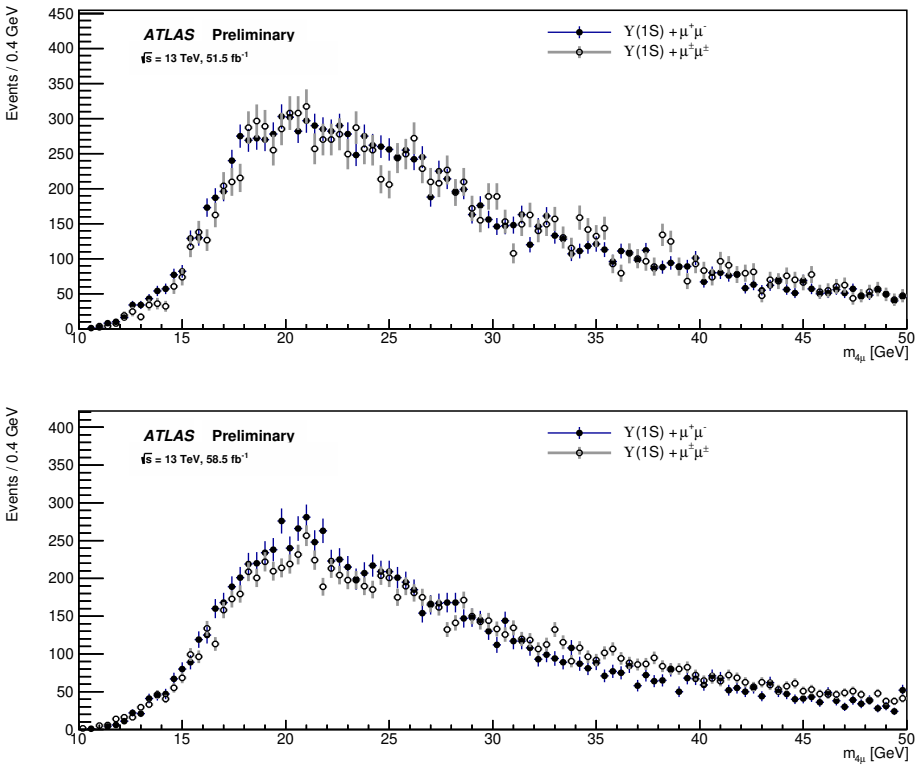

Figure 2: Comparisons of the $m_{4\mu}$ distribution of the selected $\Upsilon(1S) + \mu^+\mu^-$ events at $\sqrt{s}$ = 13 TeV in 2015–17 data (top) and 2018 data (bottom) compared to $\Upsilon(1S) + \mu^\pm\mu^\pm$ events from the data control sample. The numbers of events of the control sample are normalised to that of the $\Upsilon(1S) + \mu^+\mu^-$ events.

The $m_{4\mu}$ distributions of $\Upsilon(1S) + \mu^+\mu^-$ and $\Upsilon(1S) + \mu^\pm\mu^\pm$ events passing the final selections are compared in Figure 2. The lack of low p_{T} dimuon triggers in 13 TeV data means that Figure 2 and Figure 1 cannot be directly compared. The distributions of $\Upsilon(1S) + \mu^+\mu^-$ events are well described by those of the $\Upsilon(1S) + \mu^\pm\mu^\pm$ events in 2015–17 data. For the 2018 dataset, there is a broad enhancement of $\Upsilon(1S) + \mu^+\mu^-$ yields in the 16 to 24 GeV range compared to $\Upsilon(1S) + \mu^\pm\mu^\pm$ yields due to the opposite charge and invariant mass requirements of triggers in this dataset. The kinematic spectra of muon candidates in $\Upsilon(1S) + \mu^+\mu^-$ events across all datasets was studied in detail but found to be broadly comparable (with the exception of lower p_{T} acceptance in datasets where a three muon trigger was required).

Table 2: The numbers of selected events at major steps of the event selection at $\sqrt{s}$ = 8 and 13 TeV. The numbers are shown for the cases without and with the requirement of the three-muon trigger for 8 TeV data where di-muon triggers are also available. The numbers in parentheses are numbers of events per fb^{-1}.

Dataset	8 TeV		13 TeV	
Luminosity (fb^{-1})	20.3		51.5	58.5
Trigger	All triggers	3μ only	3μ only	3μ_bUpsi only
Four muons, $\geq$ 3 LowPt, $p_{\mathrm{T}} > (4, 4, 3, 3)$ GeV	261,893	170,467	1,152,307	231,318
One $\Upsilon(1S)$ and $10 < m_{4\mu} < 50$ GeV	6,467	3,641 (179)	20,887 (406)	19,125 (327)
$\Upsilon(1S) + \mu^+\mu^-$	3,849	2,218 (109)	13,657 (265)	10,862 (186)
$\Upsilon(1S) + \mu^\pm\mu^\pm$	2,618	1,423 (70)	7,230 (140)	8,263 (141)

6 Signal/background models and statistical fits

Unbinned maximum-likelihood fits based on the framework described in Refs. [40–42] are used to assess the compatibility of the $m_{4\mu}$ distribution of the selected $\Upsilon(1S) + \mu^+\mu^-$ events in each of the datasets with the background-only hypothesis and to extract information of a potential resonance in the signal-plus-background hypothesis. Denoting the background contribution by a function $f_B(m_{4\mu}; \vec{\theta}_{\mathrm{B}})$, and the signal contribution by a function $f_S(m_{4\mu}; m_X, \sigma_X, \vec{\theta}_{\mathrm{S}})$, the likelihood for the signal-plus-background fit of the observed $m_{4\mu}$ distribution is constructed as

$$L(N_S, m_X, \sigma_X, \vec{\theta}) = \prod_{n \text{ events}} \left[N_B \cdot f_B(m_{4\mu}; \vec{\theta}_{\mathrm{B}}) + N_S \cdot f_S(m_{4\mu}; m_X, \sigma_X, \vec{\theta}_{\mathrm{S}}) \right] \cdot \frac{e^{-(N_B+N_S)}(N_B + N_S)^n}{n!}.$$

Here m_X is the mass of the resonance and σ_X is the width of its $m_{4\mu}$ distribution, vectors $\vec{\theta}_{\mathrm{B}}$ and $\vec{\theta}_{\mathrm{S}}$ denote other free parameters of their respective functions and N_S and N_B are the numbers of signal and background events respectively. The vector $\vec{\theta}$ represents all non-explicitly listed free parameters of the likelihood function, namely N_B, $\vec{\theta}_{\mathrm{B}}$ and $\vec{\theta}_{\mathrm{S}}$. Both $f_B(m_{4\mu}; \vec{\theta}_{\mathrm{B}})$ and $f_S(m_{4\mu}; m_X, \sigma_X, \vec{\theta}_{\mathrm{S}})$ functions are normalised to unity over the considered $m_{4\mu}$ range.

The local p-value for the compatibility with the background-only hypothesis when testing a hypothesised resonance at m_X is based on the profile likelihood ratio test statistic:

$$q_0(m_X, \sigma_X) = -2 \ln \left(\frac{L(0, m_X, \sigma_X, \hat{\hat{\vec{\theta}}})}{L(\hat{N}_S, m_X, \sigma_X, \hat{\hat{\vec{\theta}}})} \right).$$

The single circumflex denotes the unconditional maximum-likelihood estimates of the parameters, i.e., all parameters are varied to maximise the likelihood function. The double circumflex are conditional maximum-likelihood estimates of the parameters assuming there is no signal. The p-value is calculated using the asymptotic approximation [42]. Global significance values are calculated accounting for the trial factors or the look-elsewhere effect of the search range, based on the method described in Ref. [43].

The $m_{4\mu}$ distribution from the simulation at $\sqrt{s} = 8$ TeV of a narrow resonance at $m_X = 18.0$ GeV is tested to be adequately described by a Gaussian distribution and also by a double-sided crystal-ball (DSCB) function [44, 45]. The Gaussian/DSCB parameters, including the width parameter σ_X, of the signal model are obtained from simulation for each m_X mass point with width increasing from 0.15 GeV at $m_X = 15$ GeV to 0.57 GeV at $m_X = 50$ GeV. In both parameterisations, the core of the distribution has a width of approximately 0.20 GeV at $m_X = 18.0$ GeV. These widths are significantly larger than the intrinsic width Γ_X assumed for the resonances in the simulation and are therefore taken to be the detector resolutions.

The characteristic features of the background $m_{4\mu}$ distribution make it difficult to find functions with a small number of free parameters to model it well over the entire mass search range of $10 - 50$ GeV. However for smaller mass windows, fourth-order Chebyshev polynomials are found to provide adequate descriptions of the $m_{4\mu}$ distribution of the same-sign $\Upsilon(1S) + \mu^{\pm}\mu^{\pm}$ events. Therefore, they are used to model the $m_{4\mu}$ distributions of background events. Variations of the fit ranges, signal models, and changes in background parameterisations to third- or to a fifth-order Chebyshev polynomials are used to determine systematic uncertainties on the modelling of the $m_{4\mu}$ distribution. The fits are performed in sliding $m_{4\mu}$ windows of 10 GeV around the test mass m_X between $15 - 45$ GeV[3] at a step size of 0.05 GeV.

6.1 Four-muon invariant mass fits in 8 TeV data

For the $m_{4\mu}$ distribution of the 8 TeV data as shown in Figure 1, the three smallest p-values are found at $m_X \sim 18.05$ GeV, 21.4 GeV, and 31.7 GeV with local significances of 5.5, 2.4, and 2.6 standard deviations (σ) respectively with corresponding global significances of $4.6\,\sigma$, $0.1\,\sigma$, and $0.5\,\sigma$ for observing these excesses anywhere in the m_X search range of $15 - 45$ GeV. Fits with a free mass resolution parameter σ_X result in 0.18 ± 0.05 GeV, 0.12 ± 0.04 GeV, and 0.08 ± 0.04 GeV, compared with the expected detector resolutions of 0.20 GeV, 0.24 GeV, and 0.36 GeV for the excesses at 18 GeV, 21 GeV, and 32 GeV, respectively. The observed $m_{4\mu}$ width is consistent with the expected detector resolution for the excess at 18 GeV whereas it is narrower than the resolution for both the 21 GeV and 32 GeV excesses, indicating these latter two are incompatible with resulting from a physical resonance.

Figure 3 (left) shows the local p-value as a function of the test mass m_X over the range of $15 - 45$ GeV. The dependence of the excess on different background models of Chebyshev polynomials from third to fifth order are tested according to the procedure of Ref. [46]. Impacts of different $m_{4\mu}$ fit ranges are also studied. While the changes in the fitted m_X and N_S values are found to be small, the local significance of the excess depends somewhat on the $m_{4\mu}$ fit range. Small fit ranges generally lead to reduced significances as the background contribution becomes less constrained. The fit ranges are varied from 6 GeV to 12 GeV centred around the excess, resulting in significance changes from $4.5\,\sigma$ to $5.6\,\sigma$. The 6 GeV fit range corresponds to approximately ± 30 times the detector resolution. The variation of the local p-value around the 18 GeV excess from the fit range variation as well as from changes to the background and signal models described above is illustrated in Figure 3 (right) by the shaded band.

Figure 4 (left) compares the data with the signal-plus-background and background-only fits in the $m_{4\mu}$ range of $15 - 25$ GeV. The fit gives a mass of $m_X = 18.05 \pm 0.05\,(\text{stat.})$ GeV and a signal yield of $N_S = 83 \pm 17\,(\text{stat.})$ events. No attempt is made to estimate systematic uncertainties. The $m_{4\mu}$ distribution of the selected events passing only the three-muon trigger is shown in Figure 4 (right) along with the signal-plus-background fit.

[3] The test range is chosen to be narrower than the m_X range in the event selection to allow background constraint from both sides of the test mass.

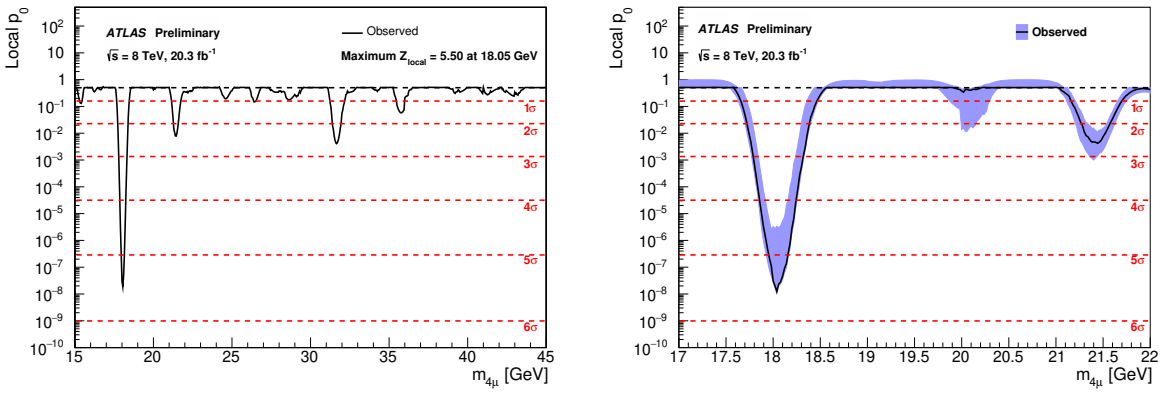

Figure 3: The compatibility with the background-only hypothesis of the data $m_{4\mu}$ distribution of the selected $\Upsilon(1S) + \mu^+\mu^-$ events at 8 TeV expressed using the local p-value as a function of the test mass in the range of (left) 15 – 45 GeV and (right) 17 – 22 GeV. The shaded band in the right-hand plot represents the variation from different $m_{4\mu}$ fit ranges, and signal and background fit models.

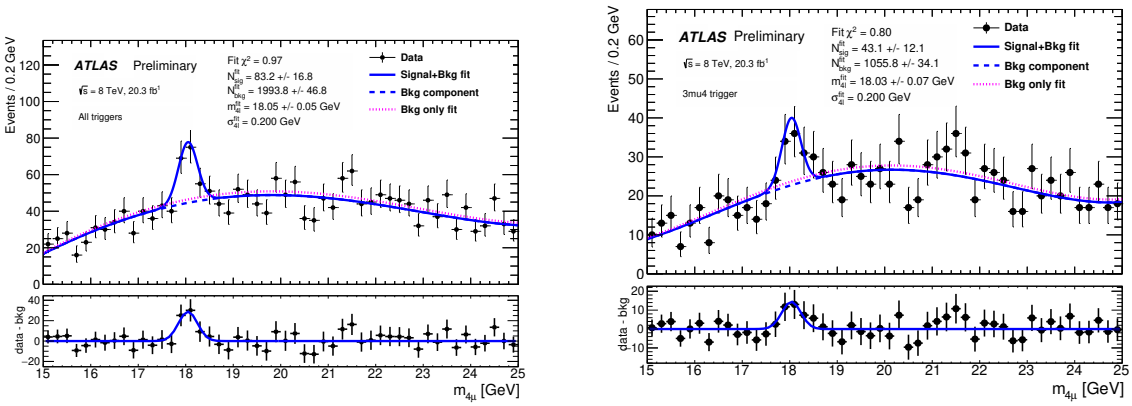

Figure 4: Distributions of $m_{4\mu}$ in the range of 15 – 25 GeV of the selected $\Upsilon(1S) + \mu^+\mu^-$ events at $\sqrt{s} = 8$ TeV for cases (left) without the explicit trigger requirement and (right) requiring events to fire the three-muon trigger. The curves show both the signal-plus-background fits and background-only fits to the data. In the fits, the backgrounds are modelled by fourth-order Chebyshev polynomials and the signal by a Gaussian function with its width fixed to the expected detector resolution. The differences between the data and the fitted background are shown in the bottom panes.

Extensive investigations of the excess at 18 GeV have been performed. The data sample is split into pairs of mutually exclusive subsets by pile-up conditions, 4μ and $\Upsilon(1S)$ rapidities, transverse momentum of the four-muon and $\Upsilon(1S)$ systems, the mass $m^{\text{non-res}}_{\mu^+\mu^-}$ of the non-resonant muon doublet, as well as trigger dependence. There is no indication of the excess being due to any particular detector-dependent subset of events studied.

Due to the unblinded nature of the 8 TeV analysis, local significances should be interpreted with caution. Variations with respect to the baseline event selection are studied to test the stability of the apparent excess. Table 3 summarises the results of these modified selections. The numbers of signal and background events vary with the modifications, but the excess remains across analysis variations, including those known to reduce potential signal efficiency, resulting in significances between 3.6 and 6.3 σ (local corresponding to

Table 3: Comparison of the results from alternative event selections at 8 TeV (see text for details) obtained from the fits to the $m_{4\mu}$ mass distributions in the range of 15 − 25 GeV of the selected $\Upsilon(1S) + \mu^+\mu^-$ events. The background is modelled by a fourth-order Chebyshev polynomial and the signal by a Gaussian distribution with its width fixed to the expected detector resolution. N_B and N_S are the numbers of background and signal events over the fit range. σ_m refers to the per-event mass resolution, with $m_{\Upsilon(1S)} \pm 2\sigma_m$ corresponding to an invariant mass range of approximately 8.94 to 9.91 GeV. Local significances are shown. The uncertainties are statistical only.

Selection criteria	N_B	Mass (GeV)	N_S	Significance (σ)
Baseline	1994 ± 47	18.05 ± 0.05	83 ± 17	5.5
Selection variations from the baseline				
≥ 2 LowPt muons	3124 ± 59	18.09 ± 0.06	94 ± 20	5.0
= 4 LowPt muons	689 ± 28	18.03 ± 0.07	37 ± 10	4.1
$m^{\text{non-res}}_{\mu^+\mu^-} > 0$ GeV	2515 ± 53	18.00 ± 0.06	81 ± 19	4.7
$m^{\text{non-res}}_{\mu^+\mu^-} > 0.5$ GeV	2306 ± 51	18.00 ± 0.05	87 ± 18	5.3
$m^{\text{non-res}}_{\mu^+\mu^-} > 2$ GeV	1696 ± 43	18.05 ± 0.07	58 ± 15	4.3
Vertex fit $\chi^2/N_{\text{d.o.f}} \leq 4$	1705 ± 43	18.03 ± 0.05	69 ± 15	5.0
Vertex fit $\chi^2/N_{\text{d.o.f}} \leq 20$	2077 ± 48	18.04 ± 0.05	81 ± 17	5.0
$m_{\Upsilon(1S)} \pm 2\sigma_m$ window	3705 ± 64	18.09 ± 0.06	90 ± 22	4.5
$\Upsilon(1S)$ mass correction	1998 ± 47	18.02 ± 0.08	64 ± 17	4.1
$m^{\text{non-res}}_{\mu^+\mu^-} < m_{\Upsilon(1S)}$	1418 ± 40	18.06 ± 0.05	94 ± 17	6.3
$p_T > 2.5$ GeV non-res. muons	2741 ± 55	18.05 ± 0.05	70 ± 19	4.1
$p_T > 4$ GeV non-res. muons	982 ± 33	18.06 ± 0.08	35 ± 11	3.6
Tight IP cuts	1469 ± 40	18.01 ± 0.05	71 ± 15	5.5
Lifetime $\lvert\tau/\sigma_\tau\rvert < 3$	1873 ± 45	18.04 ± 0.05	86 ± 17	5.6
MBS < 3	1749 ± 44	18.05 ± 0.04	83 ± 16	5.8

a global significance of between 1.9 σ and 5.4 σ, although the list of investigated variations should not be considered exhaustive. Here, the global significance takes into account only the mass range explored and the mass resolution in quantifying the look-elsewhere effect. The modifications include study of the dependence of the result on the number of muons in the quadruplet which are required to pass the LowPt selection criteria (from the baseline where three LowPt muons are required in order to match implicit trigger conditions). The modifications also include removal or tightening of the minimum invariant mass requirement on the non-resonant doublet, loosening and tightening of the four-muon vertex fit quality requirement, and replacement of the 9.2 − 9.7 GeV mass window with a $\pm 2\sigma_{m_{\mu\mu}}$ window around $m_{\Upsilon(1S)}$ for the $\Upsilon(1S) \to \mu^+\mu^-$ candidates. Some of these variations (in particular the requirement for four LowPt muons and the increased p_T thresholds) would be expected to reduce the selection efficiency of any real signal present at an invariant mass of 18 GeV.

For $\Upsilon(1S) + \mu^+\mu^-$ events, the $m_{4\mu}$ resolution can in principle be improved by recalculating the four-muon mass as $m_{4\mu} - m_{2\mu} + m_{\Upsilon(1S)}$, where $m_{2\mu}$ is the measured $\Upsilon(1S) \to \mu^+\mu^-$ mass. The fit with this mass correction results in a smaller signal yield of $N_S = 64 \pm 17$ events, leading to a reduced significance of 4.1 σ for the excess with this alternative selection. In MC simulation a reduction in signal efficiency is observed due to combinatoric ambiguity in selecting $\Upsilon(1S) \to \mu^+\mu^-$ candidates in the presence of large

backgrounds. A further variation imposes an upper threshold on the invariant mass of the non-resonant muon pair: for a true resonance at 18 GeV, decays via on-shell di-$\Upsilon(1S)$ production are kinematically forbidden and so to potentially reduce background contributions the requirement of $m^{\text{non-res}}_{\mu^+\mu^-} < m_{\Upsilon(1S)}$ is imposed, which leads to a 30% background reduction over the baseline analysis and an increased significance for the excess.

Also tested is the impact of modifying the lower p_{T} threshold on the non-resonant muon candidates, the tightening of the muon impact parameter requirements to $|d_0/\sigma_{d_0}| < 3$ and $|z_0 \sin\theta| < 0.5$ mm, the imposition of a pseudo-proper decay time significance requirement of $|\tau/\sigma_\tau| < 3$ and improving the consistency between the ID and MS momentum measurements for muon candidates failing the LowPt quality requirement by imposing a cut on the momentum bias significance (MBS) < 3. The pseudo-proper decay time τ is calculated as $\tau = L_{xy} \cdot m_{4\mu}/p_T^{4\mu}$, where L_{xy} is the projection of the transverse decay distance between the four-muon vertex and the PV along the transverse momentum of the 4μ system $p_T^{4\mu}$. Consistent results from these different selections indicate any potential resonance has no significant lifetime.

6.2 Four-muon invariant mass fits in 13 TeV data

The 13 TeV dataset provides an independent and blinded sample to check the excess observed at $m_X = 18$ GeV in the 8 TeV data. Selection cuts were restricted to those used for study of the 8 TeV data before analysis of the 13 TeV datasets, and the 13 TeV data serves as an independent cross-check of the former. The $m_{4\mu}$ distributions of the selected $\Upsilon(1S) + \mu^+\mu^-$ events, shown in Figure 2, are fit to the signal-plus-background hypothesis in the range of $15 - 25$ GeV following the same procedure described above for the 8 TeV data. In the fit, the signal is modelled using a Gaussian function with its width fixed to 0.20 GeV.

Figure 5 (left) compares the fit results with the observed distribution with the requirement of the three-muon trigger (2015–17 data). The fit to the data with $m_{4\mu}$ constrained to the value obtained from the 8 TeV fit, $m_X = 18.05$ GeV, results in a signal yield of 44 ± 25 and a significance of $1.9\,\sigma$. With a free-floating mass fit the preferred invariant mass is 17.95 ± 0.12 GeV with a signal yield of 48 ± 25.

Figure 5 (right) shows the results of the analysis of the 2018 dataset which has additional restrictions on charge and invariant mass imposed in the three-muon trigger selection. In this dataset, no significant deviation from the background expectation is observed and the fitted signal yield (with mass fixed to $m_X = 18.05$ GeV) is -4 ± 22.

Consistent with selection yields in Section 5.2, combinatorial backgrounds in Figure 5 are observed to be approximately 2.5 times larger per unit integrated luminosity in the 13 TeV dataset than at 8 TeV.

7 Di-Upsilon production as a validation tool

The searches in $\Upsilon(1S) + \mu^+\mu^-$ final states span datasets with different centre-of-mass energies, beam conditions, and triggering strategies. In addition to the use of signal MC simulations to characterise the impact on expected sensitivity from these changing data-taking conditions, we can use the associated production of a pair of $\Upsilon(1S)$ mesons decaying to a four-muon final state acts as a valuable cross-check to quantify how the yields of a known and experimentally-distinctive signature vary across the datasets. This study also serves a secondary purpose to allow us to assess whether di-$\Upsilon(1S)$ production may be

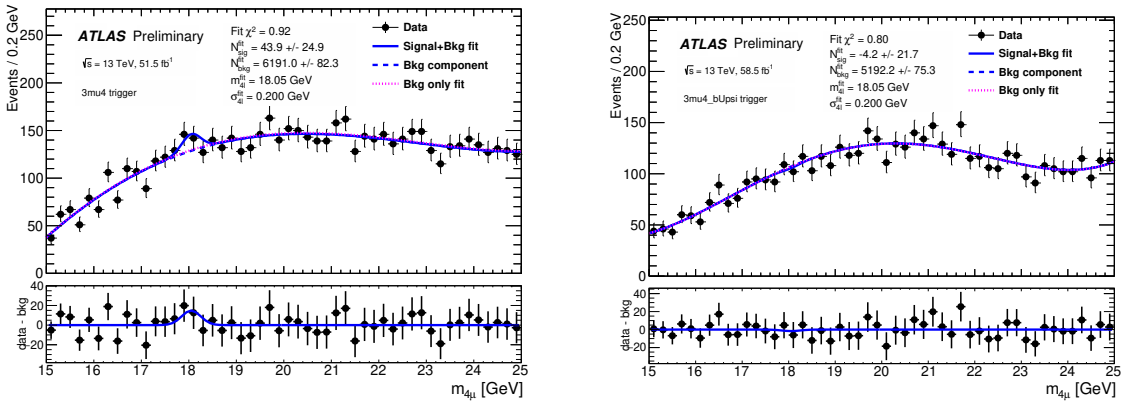

Figure 5: Distribution of $m_{4\mu}$ in the range of 15−25 GeV of the selected $\Upsilon(1S)+\mu^+\mu^-$ events at $\sqrt{s} = 13$ TeV requiring the events to pass the three-muon trigger in the 2015–17 dataset (left) or the three-muon trigger with additional charge and invariant mass requirements in the 2018 dataset (right). The curves show both the signal-plus-background fits and background-only fits to the data. The signal is modelled by a Gaussian function with its width fixed to the expected detector resolution, whereas the backgrounds are modelled by fourth-order Chebyshev polynomials. The differences between the data and the fitted background are shown in the bottom pane.

the cause of any observed resonant structures in the $m_{4\mu}$ spectrum. To understand and contextualise the results, di-$\Upsilon(1S)$ production is studied in each of the three data periods, with identical event selections as described in Table 1.

Figure 6 shows the dimuon invariant mass distribution of the second muon pair in the vicinity of the $\Upsilon(1S)$ invariant mass for $\Upsilon(1S) + \mu^+\mu^-$ events in the 8 TeV dataset (for all triggers and for the three-muon triggered subset) and in the 2015–17 and 2018 datasets at 13 TeV.

There is evidence for a new resonance in the 8 TeV dataset at a dimuon invariant mass of 9.43 ± 0.04 GeV, corresponding to a signal of di-$\Upsilon(1S)$ production with a signal yield of 40 ± 13. The $\Upsilon(1S)$ signal is modelled with a single Gaussian function. The width of the Gaussian is constrained to 100 MeV and used across all data periods to allow stable comparison of yields across the three datasets. The value is determined from a fit to the 8 TeV data. Requiring selected events in the 8 TeV dataset to fire the three-muon trigger decreases the di-$\Upsilon(1S)$ signal yield by approximately half to 25 ± 9.

In the 13 TeV dataset, for assumed equal reconstruction and trigger performance and a benchmark production cross section ratio $\sigma_{13\,\mathrm{TeV}}/\sigma_{8\,\mathrm{TeV}} = 1.4$, 89 di-$\Upsilon(1S)$ events would be expected in the 2015-17 dataset (compared with 51 ± 22 observed) and 101 events would be expected in the 2018 dataset (compared with 42 ± 21 observed). Despite increased integrated luminosity the significance of a di-$\Upsilon(1S)$ signal is substantially reduced and background levels are 2.5 times higher per unit integrated luminosity compared to the 8 TeV data.

The observed di-$\Upsilon(1S)$ yield in the 2015–2017 dataset is approximately 60% of what would be expected extrapolating from 8 TeV observations with equivalent detector performance.

These observations show similar behaviour to those of the excess at 18 GeV in the four-muon invariant mass and are consistent with the expected sensitivity reductions from signal MC. The ratios of observed yields of di-$\Upsilon(1S)$ production to fitted signal at 18 GeV in $m_{4\mu}$ are found to be statistically compatible at the 1σ level across the datasets considered. The 3μ_bUpsi trigger, active in 2018 data-taking, requires the two opposite sign muons out of the three reconstructed online to be compatible with the $\Upsilon(n\mathrm{S})$ mass.

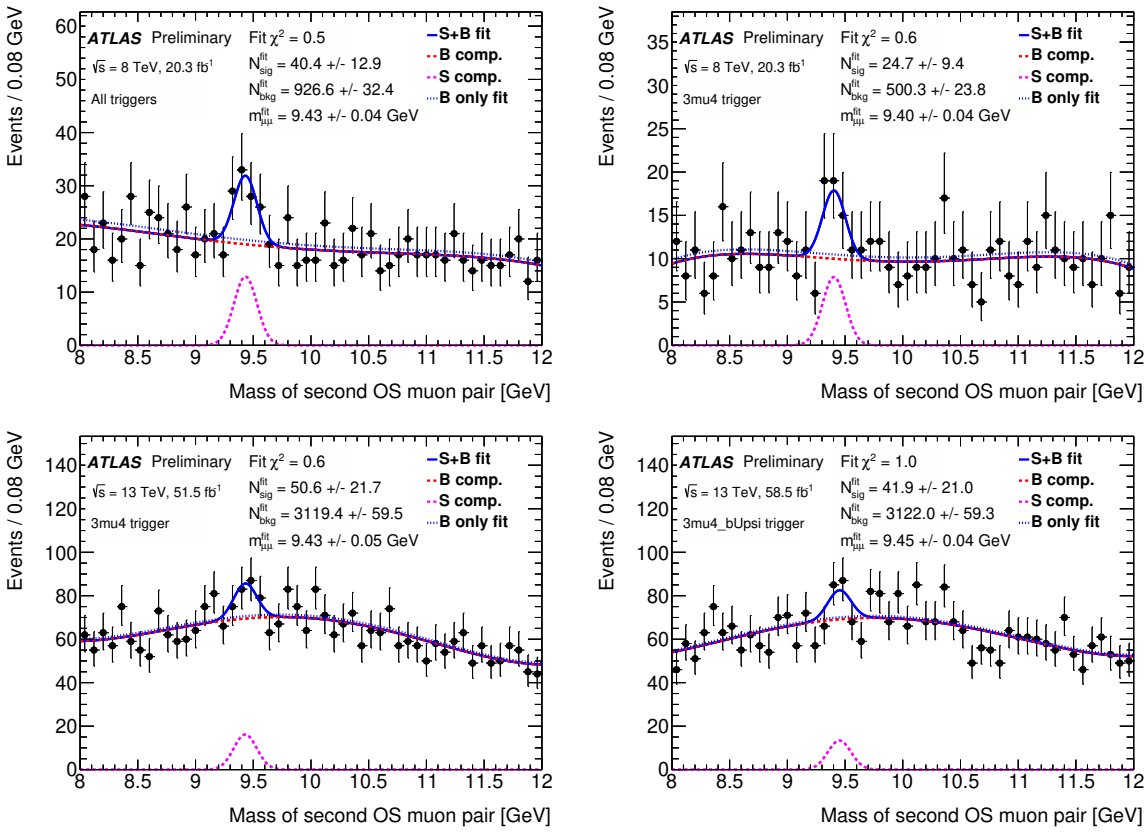

Figure 6: Dimuon invariant mass distributions of the second muon pair in $\Upsilon(1S) + \mu^+\mu^-$events in $\sqrt{s} = 8$ TeV collisions (top left) without explicit trigger requirement, (top right) requiring events to pass the three-muon trigger, and in $\sqrt{s} = 13$ TeV collisions (bottom left) requiring events to pass the three-muon trigger in the 2015–17 dataset, and (bottom right) requiring events to pass the three-muon trigger with additional charge and invariant mass requirements in the 2018 dataset. A signal at the $\Upsilon(1S)$ mass is observed, indicating the presence of di-$\Upsilon(1S)$ production. The solid curves represent the results of the signal-plus-background fits while the dotted ones are signal and background components. The $\Upsilon(1S)$ resonance is modelled by a Gaussian function and the mass of the $\Upsilon(1S)$ distribution is a free parameter in the fit. Backgrounds are modelled by fourth-order Chebyshev polynomials.

For di-$\Upsilon(1S)$ this is a fully-efficient requirement, but for general $\Upsilon(1S) + \mu^+\mu^-$ signatures this induces an additional inefficiency, estimated from MC simulations to be $\sim 10\%$.

The distribution of di-$\Upsilon(1S)$ candidates in $m_{4\mu}$ was studied and determined to not be the cause of the 18 GeV excess itself, having a relatively uniform distribution in $m_{4\mu}$ beginning from masses of ~ 18.6 GeV and extending to higher $m_{4\mu}$ invariant masses.

8 Signal interpretation and limit setting

The observations at 8 TeV and in the two 13 TeV datasets at 18 GeV in four-muon invariant mass are used to derive constraints on the upper limit of the total production cross-section times branching fraction of a resonance decaying to a four muon final state via one intermediate $\Upsilon(1S)$ resonance.

Upper limits are defined at a 95% confidence level using the modified frequentist construction CL_s [42, 47] with a signal model of fixed four-muon invariant mass and width. Background-only pseudo-experiments are performed to determine the median expected 95% CL_s upper limit for each dataset as well as the 68% and 95% inclusion intervals. This procedure is performed for a variety of signal models, described in Section 3. The relative efficiencies predicted by the MC samples between the three data periods have been corrected for the observations made from di-$\Upsilon(1S)$ production in Section 7.

The derived limits depend strongly on the MC signal model tested, due to small differences in kinematics between the signal models that modify predicted efficiencies and geometric acceptance. Models relating to the production of a scalar Higgs-like state resulted in the highest overall efficiencies (at $\sim$ 7% in 8 TeV data-taking) and thus lowest cross-section limits, whereas those with pseudoscalar tetraquark state production resulted in the lowest efficiencies of models studied (at $\sim$ 3% in 8 TeV data-taking) resulting in larger values of cross-section limits. The median expected and observed limits are illustrated in Figure 7 and are summarised in Table 4 for the maximum and minimum efficiency scenarios across the models studied, for the three distinct data-taking periods at ATLAS and the model-dependent statistical combination of the full 13 TeV dataset. We emphasise that the range of limits and cross-section best-estimates derived from this set of models does not necessarily span all possible variations of efficiency and kinematic acceptance, but simply illustrate the large model-dependence of any interpretation. The cross-section interpretations at 13 TeV are derived using efficiency corrections from di-$\Upsilon(1S)$ production studies described earlier (found to be compatible with expectations from signal MC simulations), and thus may be up to 50% higher or lower than reported due to systematic uncertainties.

Table 4: Expected and observed upper limits on the production cross-section times branching fraction limits for a particle X with an invariant mass of 18 GeV decaying to a $\Upsilon(1S) + \mu^+\mu^- \rightarrow \mu^+\mu^-\mu^+\mu^-$ final state in the three distinct data-taking periods at ATLAS (as well as the statistical combination of the full 13 TeV dataset). 'Low ε' and 'high ε' refer to the limits derived from signal models with lowest (Higgs-like scalar) and highest (pseudoscalar tetraquark) predicted selection plus reconstruction efficiencies, respectively.

		Dataset			
		8 TeV	13 TeV 2015–17	13 TeV 2018	13 TeV comb.
Low ε (fb)	Expected	36	81	101	58
	Observed	145	154	94	90
High ε (fb)	Expected	21	41	51	30
	Observed	85	78	48	46

Due to the significant excess in 8 TeV data, the observed limits are necessarily much weaker than the median expected limits. In this case we additionally derive a total production cross-section estimate for the excess, interpreted as the production of a new state decaying to four muons, equal to between 61 ± 12 fb and 105 ± 20 fb, dependent on the model considered (where the uncertainty quoted is purely statistical).

Cross-section limits at 8 and 13 TeV are not directly comparable due to model-dependent uncertainties and parton luminosity scaling of the production rates. To allow for visual comparison of the results at two centre-of-mass energies we indicate (with a purple dotted line) the 8 TeV upper limit corresponding to a scaling of the 13 TeV combined observed limit, as predicted by each of the extreme models under study.

The observed limits are consistent with the existing results from the CMS Collaboration [15] and the LHCb Collaboration [14], although neither of these searches observed significant excesses.

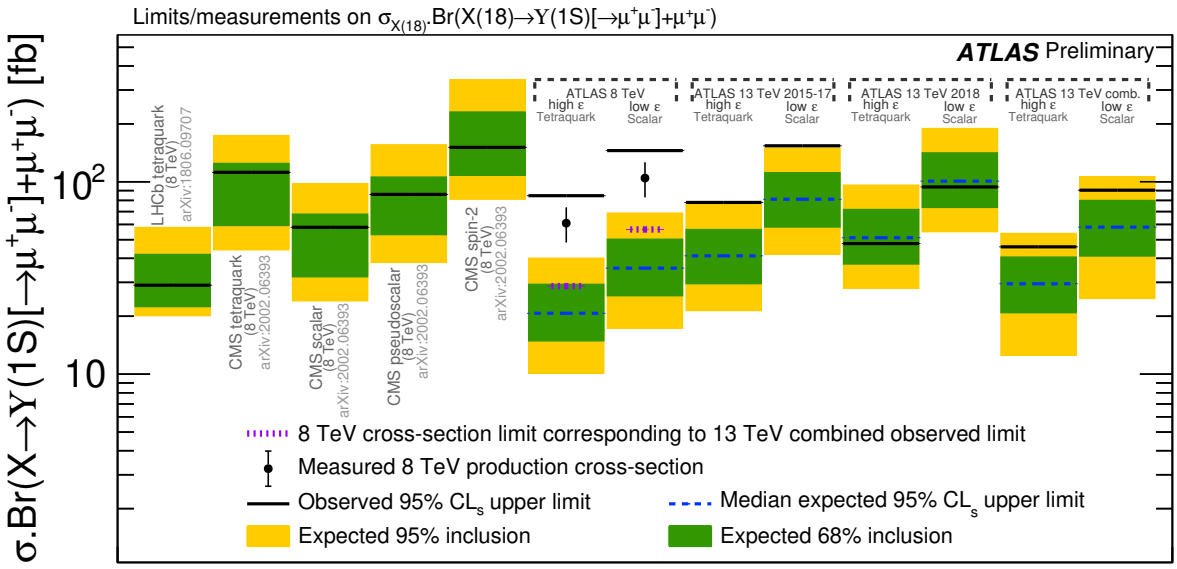

Figure 7: Expected and observed production cross-section times branching fraction limits for a particle X with an invariant mass of 18 GeV decaying to a $\Upsilon(1S) + \mu^+\mu^- \to \mu^+\mu^-\mu^+\mu^-$ final state in the three distinct data-taking periods at ATLAS (as well as the statistical combination of the full 13 TeV dataset). Due to the large theoretical uncertainty on the interpretation of the results, we present results for a range of signal models. 'Low ε' and 'high ε' refer to the limits derived from signal models with lowest (Higgs-like scalar) and highest (pseudoscalar tetraquark) predicted selection plus reconstruction efficiencies, respectively. Where an excess is significant, cross-section estimates are indicated with data points (with statistical uncertainties). The purple dotted line indicates the 8 TeV cross-section limit corresponding to the 13 TeV combined observed limit, once parton luminosity scaling predicted by the model under study is accounted for, for the specific models under study at leading-order. Comparison is made to limits from CMS and LHCb for a variety of models.

9 Conclusion

A search for new low-mass resonances in the four-muon final state is performed using proton-proton collision data corresponding to integrated luminosities of 20.3 fb^{-1} at $\sqrt{s}$ = 8 TeV and 51.5 fb^{-1} and 58.5 fb^{-1} collected in 2015–2017 and 2018, respectively, at 13 TeV by the ATLAS experiment at the LHC. With the 8 TeV data, typical selection criteria are applied but the analysis was not blinded. An excess consistent with a narrow-width particle is observed at 18 GeV in the four-muon invariant-mass distribution of $\Upsilon(1S) + \mu^+\mu^- \to \mu^+\mu^-\,\mu^+\mu^-$ events in the 8 TeV dataset. The stability of the observed excess is confirmed by an extensive variation of the cuts. The range of local significances of the excess in the 8 TeV data from a background-only hypothesis is estimated to vary between 3.6 σ and 6.3 σ for the 10 – 50 GeV search range considered corresponding to global significances of 1.9 σ to 5.4 σ depending on the selection choice. A statistically independent study with the selection requirements fixed to those from the 8 TeV analysis is subsequently performed, using the 13 TeV data. In early 13 TeV data, collected with an equivalent three-muon trigger to the one available at 8 TeV, a structure is observed at the same mass with a reduced significance of 1.9 σ. In later 13 TeV data-taking with more restrictive trigger selections, no excess is observed.

Performance studies using signal MC simulations and di-$\Upsilon(1S)$ production signals suggest reductions in expected sensitivity across datasets, due to the more restrictive trigger selections. The MC and data-driven

estimates yield consistent sensitivity losses although data-driven di-$\Upsilon(1S)$ studies, which are at a different mass, suggest that there is loss in sensitivity in the 13 TeV dataset relative to that at 8 TeV that is slightly beyond that predicted by the signal MC models, but still compatible. For the MC models, the excess at 8 TeV is excluded by the 13 TeV data at up to 2.7σ. Analyses by LHCb and CMS have not reported an excess at a mass of 18 GeV.

Acknowledgements

We thank CERN for the very successful operation of the LHC, as well as the support staff from our institutions without whom ATLAS could not be operated efficiently.

We acknowledge the support of ANPCyT, Argentina; YerPhI, Armenia; ARC, Australia; BMWFW and FWF, Austria; ANAS, Azerbaijan; CNPq and FAPESP, Brazil; NSERC, NRC and CFI, Canada; CERN; ANID, Chile; CAS, MOST and NSFC, China; Minciencias, Colombia; MEYS CR, Czech Republic; DNRF and DNSRC, Denmark; IN2P3-CNRS and CEA-DRF/IRFU, France; SRNSFG, Georgia; BMBF, HGF and MPG, Germany; GSRI, Greece; RGC and Hong Kong SAR, China; ISF and Benoziyo Center, Israel; INFN, Italy; MEXT and JSPS, Japan; CNRST, Morocco; NWO, Netherlands; RCN, Norway; MEiN, Poland; FCT, Portugal; MNE/IFA, Romania; MESTD, Serbia; MSSR, Slovakia; ARRS and MIZŠ, Slovenia; DSI/NRF, South Africa; MICINN, Spain; SRC and Wallenberg Foundation, Sweden; SERI, SNSF and Cantons of Bern and Geneva, Switzerland; MOST, Taiwan; TENMAK, Türkiye; STFC, United Kingdom; DOE and NSF, United States of America. In addition, individual groups and members have received support from BCKDF, CANARIE, Compute Canada and CRC, Canada; PRIMUS 21/SCI/017 and UNCE SCI/013, Czech Republic; COST, ERC, ERDF, Horizon 2020 and Marie Sk lodowska-Curie Actions, European Union; Investissements d'Avenir Labex, Investissements d'Avenir Idex and ANR, France; DFG and AvH Foundation, Germany; Herakleitos, Thales and Aristeia programmes co-financed by EU-ESF and the Greek NSRF, Greece; BSF-NSF and MINERVA, Israel; Norwegian Financial Mechanism 2014-2021, Norway; NCN and NAWA, Poland; La Caixa Banking Foundation, CERCA Programme Generalitat de Catalunya and PROMETEO and GenT Programmes Generalitat Valenciana, Spain; Göran Gustafssons Stiftelse, Sweden; The Royal Society and Leverhulme Trust, United Kingdom.

The crucial computing support from all WLCG partners is acknowledged gratefully, in particular from CERN, the ATLAS Tier-1 facilities at TRIUMF (Canada), NDGF (Denmark, Norway, Sweden), CC-IN2P3 (France), KIT/GridKA (Germany), INFN-CNAF (Italy), NL-T1 (Netherlands), PIC (Spain), ASGC (Taiwan), RAL (UK) and BNL (USA), the Tier-2 facilities worldwide and large non-WLCG resource providers. Major contributors of computing resources are listed in Ref. [48].

References

[1] A. V. Berezhnoy, A. V. Luchinsky and A. A. Novoselov, *Tetraquarks Composed of 4 Heavy Quarks*, Phys. Rev. **D86** (2012) 034004, arXiv: 1111.1867 [hep-ph] (cit. on p. 2).

[2] A. V. Berezhnoy, A. K. Likhoded and A. A. Novoselov, *Υ-meson pair production at LHC*, Phys. Rev. **D87** (2013) 054023, arXiv: 1210.5754 [hep-ph] (cit. on p. 2).

[3] W. Chen, H.-X. Chen, X. Liu, T. Steele and S.-L. Zhu, *Hunting for exotic doubly hidden-charm/bottom tetraquark states*, Phys. Lett. B **773** (2017) 247, arXiv: 1605.01647 [hep-ph] (cit. on p. 2).

[4] M. Karliner, S. Nussinov and J. L. Rosner, *$QQ\bar{Q}\bar{Q}$ states: masses, production, and decays*, Phys. Rev. D **95** (2017) 034011, arXiv: 1611.00348 [hep-ph] (cit. on p. 2).

[5] D. Curtin, R. Essig, S. Gori and J. Shelton, *Illuminating Dark Photons with High-Energy Colliders*, JHEP **02** (2015) 157, arXiv: 1412.0018 [hep-ph] (cit. on p. 2).

[6] J. D. Wells, *How to Find a Hidden World at the Large Hadron Collider*, (2008), arXiv: 0803.1243 [hep-ph] (cit. on p. 2).

[7] S. Gopalakrishna, S. Jung and J. D. Wells, *Higgs boson decays to four fermions through an abelian hidden sector*, Phys. Rev. **D78** (2008) 055002, arXiv: 0801.3456 [hep-ph] (cit. on p. 2).

[8] U. Ellwanger, C. Hugonie and A. M. Teixeira, *The Next-to-Minimal Supersymmetric Standard Model*, Phys. Rept. **496** (2010) 1, arXiv: 0910.1785 [hep-ph] (cit. on p. 2).

[9] D. Curtin et al., *Exotic decays of the 125 GeV Higgs boson*, Phys. Rev. **D90** (2014) 075004, arXiv: 1312.4992 [hep-ph] (cit. on p. 2).

[10] M. N. Doroshenko, V. G. Kartvelishvili, E. G. Chikovani and S. M. Esakiya, *Vector quarkonium in decays of heavy Higgs particles*, Yad. Fiz. **46** (1987) 864, [Sov. J. Nucl. Phys.46,493(1987)] (cit. on p. 2).

[11] V. G. Kartvelishvili, E. G. Chikovani and S. M. Esakiya, *The Production and Decays of Heavy Quark Bound States in Strong and Electroweak Interactions. (In Russian)*, Fiz. Elem. Chast. Atom. Yadra **19** (1988) 139 (cit. on p. 2).

[12] V. Kartvelishvili, A. V. Luchinsky and A. A. Novoselov, *Double vector quarkonia production in exclusive Higgs boson decays*, Phys. Rev. **D79** (2009) 114015, arXiv: 0810.0953 [hep-ph] (cit. on p. 2).

[13] G. T. Bodwin, F. Petriello, S. Stoynev and M. Velasco, *Higgs boson decays to quarkonia and the $H\bar{c}c$ coupling*, Phys. Rev. **D88** (2013) 053003, arXiv: 1306.5770 [hep-ph] (cit. on p. 2).

[14] LHCb Collaboration, *Search for beautiful tetraquarks in the $\Upsilon(1S)\mu^+\mu^-$ invariant-mass spectrum*, JHEP **10** (2018) 086, arXiv: 1806.09707 [hep-ex] (cit. on pp. 2, 16).

[15] CMS Collaboration, *Measurement of the $\Upsilon(1S)$ pair production cross section and search for resonances decaying to $\Upsilon(1S)\mu^+\mu^-$ in proton-proton collisions at $\sqrt{s}$ = 13 TeV*, Phys. Lett. B **808** (2020) 135578, arXiv: 2002.06393 [hep-ex] (cit. on pp. 2, 16).

[16] LHCb Collaboration, *Observation of structure in the J/ψ-pair mass spectrum*, Sci. Bull. **65** (2020) 1983, arXiv: 2006.16957 [hep-ex] (cit. on p. 2).

[17] ATLAS Collaboration, *Observation of an excess of di-charmonium events in the four-muon final state with the ATLAS detector*, (2023), arXiv: 2304.08962 [hep-ex] (cit. on pp. 2, 5).

[18] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003 (cit. on p. 2).

[19] ATLAS IBL Collaboration, *Production and Integration of the ATLAS Insertable B-Layer*, JINST **13** (2018) T05008, arXiv: 1803.00844 [physics.ins-det] (cit. on p. 2).

[20] ATLAS Collaboration, *The ATLAS Collaboration Software and Firmware*, ATL-SOFT-PUB-2021-001, 2021, URL: https://cds.cern.ch/record/2767187 (cit. on p. 2).

[21] E. Eichten and Z. Liu, *Would a Deeply Bound $b\bar{b}b\bar{b}$ Tetraquark Meson be Observed at the LHC?*, (2017), arXiv: 1709.09605 [hep-ph] (cit. on p. 3).

[22] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **07** (2014) 079, arXiv: 1405.0301 [hep-ph] (cit. on p. 3).

[23] T. Sjostrand, S. Mrenna and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, arXiv: 0710.3820 [hep-ph] (cit. on p. 3).

[24] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. A **462** (2001) 152 (cit. on pp. 3, 4).

[25] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, ATL-PHYS-PUB-2014-021, 2014, URL: https://cds.cern.ch/record/1966419 (cit. on p. 3).

[26] R. D. Ball et al., *Parton distributions for the LHC Run II*, JHEP **04** (2015) 040, arXiv: 1410.8849 [hep-ph] (cit. on p. 3).

[27] T. Sjostrand, S. Mrenna and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **05** (2006) 026, arXiv: hep-ph/0603175 [hep-ph] (cit. on p. 3).

[28] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau and A. D. Polosa, *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, JHEP **07** (2003) 001, arXiv: hep-ph/0206293 [hep-ph] (cit. on p. 3).

[29] T. Melia, P. Nason, R. Rontsch and G. Zanderighi, W^+W^-, *WZ and ZZ production in the POWHEG BOX*, JHEP **11** (2011) 078, arXiv: 1107.5051 [hep-ph] (cit. on p. 3).

[30] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010) 043, arXiv: 1002.2581 [hep-ph] (cit. on p. 3).

[31] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **11** (2004) 040, arXiv: hep-ph/0409146 [hep-ph] (cit. on p. 3).

[32] GEANT4 Collaboration, S. Agostinelli et al., *GEANT4: A Simulation toolkit*, Nucl. Instrum. Meth. A **506** (2003) 250 (cit. on p. 4).

[33] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C **70** (2010) 823, arXiv: 1005.4568 [physics.ins-det] (cit. on p. 4).

[34] ATLAS Collaboration, *Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC proton-proton collision data*, Eur. Phys. J. **C74** (2014) 3130, arXiv: 1407.3935 [hep-ex] (cit. on p. 4).

[35] ATLAS Collaboration, *Identification of very-low transverse momentum muons in the ATLAS experiment*, tech. rep. ATL-PHYS-PUB-2020-002, CERN, 2020, URL: https://cds.cern.ch/record/2710574 (cit. on p. 4).

[36] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at* $\sqrt{s} = 13$ *TeV*, Eur. Phys. J. C **76** (2016) 292, arXiv: 1603.05598 [hep-ex] (cit. on p. 4).

[37] K. A. Olive et al., *Review of Particle Physics*, Chin. Phys. **C38** (2014) 090001 (cit. on p. 5).

[38] ATLAS Collaboration, *Measurement of the differential cross-sections of prompt and non-prompt production of* J/ψ *and* $\psi(2S)$ *in* pp *collisions at* $\sqrt{s} = 7$ *and 8 TeV with the ATLAS detector*, Eur. Phys. J. C **76** (2016) 283, arXiv: 1512.03657 [hep-ex] (cit. on p. 5).

[39] ATLAS Collaboration, *Measurement of Upsilon production in 7 TeV pp collisions at ATLAS*, Phys. Rev. D **87** (2013) 052004, arXiv: 1211.7255 [hep-ex] (cit. on p. 5).

[40] L. Moneta et al., *The RooStats Project*, PoS **ACAT2010** (2010) 057, arXiv: 1009.1003 [physics.data-an] (cit. on p. 9).

[41] W. Verkerke and D. P. Kirkby, *The RooFit toolkit for data modeling*, eConf **C0303241** (2003) MOLT007, [186 (2003)], arXiv: physics/0306116 [physics] (cit. on p. 9).

[42] G. Cowan, K. Cranmer, E. Gross and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. C **71** (2011) 1554, [Erratum: Eur.Phys.J.C 73, 2501 (2013)], arXiv: 1007.1727 [physics.data-an] (cit. on pp. 9, 16).

[43] E. Gross and O. Vitells, *Trial factors or the look elsewhere effect in high energy physics*, Eur. Phys. J. **C70** (2010) 525, arXiv: 1005.1891 [physics.data-an] (cit. on p. 9).

[44] ATLAS Collaboration, *Search for resonances in diphoton events at $\sqrt{s}$=13 TeV with the ATLAS detector*, JHEP **09** (2016) 001, arXiv: 1606.03833 [hep-ex] (cit. on p. 10).

[45] M. Oreglia, 'A Study of the Reactions $\psi' \to \gamma\gamma\psi$', PhD thesis: SLAC, 1980 (cit. on p. 10).

[46] P. D. Dauncey, M. Kenzie, N. Wardle and G. J. Davies, *Handling uncertainties in background shapes*, JINST **10** (2015) P04015, arXiv: 1408.6865 [physics.data-an] (cit. on p. 10).

[47] A. L. Read, *Presentation of search results: the CL_s technique*, J. Phys. G **28** (2002) 2693, ed. by M. R. Whalley and L. Lyons (cit. on p. 16).

[48] ATLAS Collaboration, *ATLAS Computing Acknowledgements*, ATL-SOFT-PUB-2021-003, 2021, URL: https://cds.cern.ch/record/2776662 (cit. on p. 18).