



Physics Beyond the Standard Model with Future X-Ray Observatories: Projected Constraints on Very-light Axion-like Particles with Athena and AXIS

Júlia Sisk-Reynés¹ , Christopher S. Reynolds^{1,2} , Michael L. Parker¹ , James H. Matthews^{1,3} , and M. C. David Marsh⁴

¹Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK; jms332@cam.ac.uk

²Department of Astronomy, University of Maryland, College Park, MD 20742, USA

³Department of Physics, Astrophysics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK

⁴The Oskar Klein Centre, Department of Physics, Stockholm University, Stockholm SE-106 91, Sweden

Received 2022 November 9; revised 2023 April 12; accepted 2023 April 21; published 2023 June 27

Abstract

Axion-like particles (ALPs) are well-motivated extensions of the Standard Model of Particle Physics and a generic prediction of some string theories. X-ray observations of bright active galactic nuclei (AGNs) hosted by rich clusters of galaxies are excellent probes of very-light ALPs, with masses $\log(m_a/eV) < -12.0$. We evaluate the potential of future X-ray observatories, particularly Athena and the proposed AXIS, to constrain ALPs via observations of cluster-hosted AGNs, taking NGC 1275 in the Perseus cluster as our exemplar. Assuming perfect knowledge of the instrument calibration, we show that a modest exposure (200 ks) of NGC 1275 by Athena permits us to exclude all photon–ALP couplings $g_{a\gamma} > 6.3 \times 10^{-14} \text{ GeV}^{-1}$ at the 95% confidence level, as previously shown by Conlon et al., representing a factor of 10 improvement over current limits. We then proceed to assess the impact of realistic calibration uncertainties on the Athena projection by applying a standard Cash likelihood procedure, showing the projected constraints on $g_{a\gamma}$ weakened by a factor of 10 (back to the current most sensitive constraints). However, we show how the use of a deep neural network can disentangle the energy-dependent features induced by instrumental miscalibration and those induced by photon–ALP mixing, allowing us to recover most of the sensitivity to the ALP physics. In our explicit demonstration, the machine learning applied allows us to exclude $g_{a\gamma} > 2.0 \times 10^{-13} \text{ GeV}^{-1}$, complementing the projected constraints of next-generation ALP dark matter birefringent cavity searches for very-light ALPs. Finally, we show that a 200 ks AXIS/on-axis observation of NGC 1275 will tighten the current best constraints on very-light ALPs by a factor of 3.

Unified Astronomy Thesaurus concepts: Particle astrophysics (96); Dark matter (353); X-ray active galactic nuclei (2035); Magnetic fields (994); X-ray telescopes (1825); Galaxy clusters (584)

1. Introduction

Astrophysical observations probe physical extremes that provide a window onto physics beyond the Standard Model (SM). Axions and axion-like particles (ALPs) in particular arise from well-motivated extensions to the SM and are amenable to astrophysical study. The quantum chromodynamics (QCD) axion is the leading solution to the so-called “strong CP problem” in the SM. This problem arises due to the unexpectedly small value of the electric dipole moment for the neutron, $<10^{-9} e \text{ cm}$ (Barton & White 1969), yielding experimental evidence for the conservation of charge conjugation parity (CP) symmetry by the strong force, which is not predicted by SM theory. Such a tension can be solved through the introduction of a new $U(1)$ symmetry in the SM Lagrangian which enables replacing the CP-violating phase by a pseudoscalar, dynamical field a (Peccei & Quinn 1977). This symmetry, known as Peccei–Quinn symmetry, is both spontaneously broken at a high scale and is anomalous, resulting in a pseudo-Nambu–Goldstone boson (pNGB) of mass m_a , the QCD axion (Weinberg 1978; Wilczek 1978). The dynamics and interactions of such particle can be derived from its mass m_a and decay constant, whose product is set by that of the mass and decay constant of the neutral pion.

ALPs are pNGBs that arise in many “Beyond the SM” (BSM) frameworks, such as string compactifications (Green et al. 1988; Conlon 2006; Svrcek & Witten 2006; Cicoli et al. 2012). Excitingly, some of these theories predict sizeable ALP–photon couplings that are within reach of astrophysical observations (Halverson et al. 2019; Demirtas et al. 2021; Mehta et al. 2021). ALPs do not couple to QCD, but can couple to other vector bosons and have derivative interactions with matter fields. Another driver for the rising interest in ALPs and axion searches is their potential to be dark matter (DM; Abbott & Sikivie 1983; Dine & Fischler 1983; Preskill et al. 1983; Irastorza & Redondo 2018), or, for sufficiently light ALPs, dark energy (Carroll 1998). In the low-energy effective theory relevant for observations, the ALP mass m_a and ALP coupling to the electromagnetic field $g_{a\gamma}$ are assumed to be independent, regardless of their cosmological role.

Most relevant to our work is the study of the interaction of ALPs with electromagnetism. At a fundamental level, this interaction is described by the Lagrangian term

$$\mathcal{L}_{a\gamma} = a g_{a\gamma} \mathbf{B} \cdot \mathbf{E}, \quad (1)$$

where a is the ALP field, \mathbf{B} and \mathbf{E} are the magnetic and electric fields, respectively, and $g_{a\gamma}$ is the photon–ALP coupling constant. In most astrophysical searches for ALPs, the relevant electric field is that of the photon beam, whereas the magnetic field is imposed externally as a “background” field. The resulting expressions for photon–ALP mixing in the simple case of a uniform magnetic field are outlined in Appendix A.

Table 1Top: List of the Current Most Sensitive Bounds on Very-light ALPs ($m_a < 10^{-12}$ eV) Based on Single-source Spectroscopy Studies (see the References in Section 1)

Source	Instrument	Current Bounds on $g_{a\gamma}$ (not excluded) at the 95% level	Reference
NGC 1275	Chandra/HETG	$g_{a\gamma} < 4.0 \times 10^{-13} \text{ GeV}^{-1}$	(Reynolds et al. 2020; Model B)
A1795Sy1 [†]	Chandra	$g_{a\gamma} \lesssim 0.6 \times 10^{-12} \text{ GeV}^{-1}$	Schallmoser et al. (2021) + ML
H1821+643	Chandra/LETG+HETG	$g_{a\gamma} < 5.0 \times 10^{-13} \text{ GeV}^{-1}$	Sisk-Reynés et al. (2021)
Projected bounds $g_{a\gamma}$ (not excluded) at the 95% level from next-generation X-ray telescopes			
NGC 1275	Athena/X-IFU ^{*,†}	$g_{a\gamma} < 6.3 \times 10^{-14} \text{ GeV}^{-1}$	This work, no calibration (+ML)
NGC 1275	Athena/X-IFU [†]	$g_{a\gamma} < 2.0 \times 10^{-13} \text{ GeV}^{-1}$	This work, including calibration with ML
NGC 1275	AXIS/on-axis [*]	$g_{a\gamma} < 2.0 \times 10^{-13} \text{ GeV}^{-1}$	This work, no calibration

Note. Schallmoser et al. (2021) used a 3D field model and machine-learning classifiers[†] to improve on the previous bound from A1795Sy1 found by Conlon et al. (2017b). Bottom: projected bounds on very-light ALPs from the next-generation X-ray telescopes Athena/X-IFU and AXIS (when used for on-axis observations), based on their optimal calibration^{*} (see Section 5). The upper bound on $g_{a\gamma}$ quoted for Athena/X-IFU under the assessment of detector calibration[†] is found by a deep neural network trained on simulated spectra of NGC 1275 with injected ALPs at different $g_{a\gamma}$ across a library of conservative detector responses (see Section 7).

Table 2

Overview of the Designs of Athena, AXIS (for its Average Field-of-view Response) and Chandra/HETG (Quoted for Its High Energy Grating, HEG, After the First Five Years of Performance—see Canizares et al. 2005)

X-ray Mission	Spectral Resolution	Angular Resolution	Collecting Area at 1 keV	Broadband Energy Range
Athena	2.5 eV [†] at 1 keV	5''	1 m ²	0.5–12 keV
AXIS	150 eV at 6 keV	1''	0.56 m ²	0.2–12 keV
Chandra/HETG	150 eV at 1 keV	0.5''	0.04 m ²	2.0–12 keV

Note. All Athena specifications result from a combination of the capabilities of both of its instruments, the WFI and X-IFU, unless flagged with a dagger (†) symbol, in which case they correspond to those of the X-IFU only.

Importantly, photon–ALP mixing requires the external \mathbf{B} -field component perpendicular to the beam propagation direction to be nonzero.

Photon–ALP mixing will induce energy-dependent features in the spectrum of a photon beam propagating through an ionized, magnetized plasma as the photons mix with sufficiently light ALPs. Photon–ALP mixing is an oscillatory function with energy that is strongly dependent on the properties of the plasma, as well as m_a , $g_{a\gamma}$. It is the presence (or absence) of such energy-dependent modulations in the spectra of astrophysical sources that allows us to constrain ALPs, that is, to set an upper bound on $g_{a\gamma}$ for a given m_a .

Bounds on anomalous stellar cooling provided the earliest astrophysical bounds on ALPs of $m_a < 10^{-10}$ eV (Raffelt & Stodolsky 1988; Raffelt 1996; Carena et al. 2020). Competitive constraints on $g_{a\gamma}$ for such ALP masses come from gamma-ray observations of SN 1987A (Payez et al. 2015) and other core-collapse supernovae (Meyer & Petrushevska 2020; Crnogorčević et al. 2021; Meyer et al. 2022). These limits have been complemented in a broader mass range ($m_a < 0.02$ eV) by the CERN Axion Solar Telescope (CAST; Anastassopoulos et al. 2017), which aimed to reprocess ALPs generated at the core of the Sun back into X-ray photons within a magnetized chamber located at CERN and excluded all photon–ALP couplings $g_{a\gamma} > 6.6 \times 10^{-11} \text{ GeV}^{-1}$ at the 95% level. We refer to Armengaud et al. (2019) and Hong et al. (2019) for projected constraints from solar axions from future ground-based and spaced-based, next-generation missions.

Many astrophysical settings are both photon rich and magnetized, providing the opportunity to study ALPs. We now proceed to focus on the bounds on $g_{a\gamma}$ inferred from X-ray observations of AGNs hosted by galaxy clusters. Clusters are permeated by the intracluster medium (ICM), known to be magnetized up to $\sim 10 \mu\text{G}$ within cool-core cluster cores. While the 3D structure of the ICM field cannot directly be mapped,

the strength and typical coherence length of the ICM field can be inferred from measurements of Faraday rotation measures (RMs) at radio wavelengths. RMs have also indicated that ICM fields are not regular but turbulent on scales of 100 pc–10 kpc (Govoni 2012).

In many clusters, the central brightest cluster galaxy (BCG) hosts an X-ray luminous AGN. As X-rays from the AGN traverse the magnetized ICM, photon–ALP interconversion would imprint distortions on the intrinsic AGN spectrum. The absence of observed distortions can be used to exclude regions of the $(m_a, g_{a\gamma})$ parameter space for $m_a < 10^{-12}$ eV. The current most sensitive bounds on these very-light ALPs were inferred with this technique using Chandra transmission grating spectroscopy of the luminous cluster-hosted AGN H1821+643 (Sisk-Reynés et al. 2021; Table 1). We refer the reader to Wouters & Brun (2013), Conlon et al. (2017b), Berg et al. (2017), and Marsh et al. (2017) for previous astrophysical ALP bounds inferred from X-ray observations of AGNs hosted by or located behind clusters. The purpose of this paper is to assess the capabilities of future X-ray missions, particularly the Athena and AXIS observatories, to probe ALP physics.

The Athena X-ray Observatory is the European Space Agency’s next flagship X-ray observatory and will offer spatially resolved spectroscopic capabilities exceeding those of current and next-generation missions expected to launch in the near future (Barret et al. 2023). The leap in the science Athena will provide is due to the unprecedented spectral resolution (2.5 eV) of its X-ray Integral Field Unit (the Athena/X-IFU or X-IFU), Athena’s large effective area (see Table 2), and an arcsecond-scale spatial resolution across a broadband coverage (0.5–12 keV; see Barret et al. 2020, 2023). These factors provide Athena with exciting prospects for cluster–ALP searches, given that an X-ray telescope with a fine energy resolution such as the X-IFU will be able to discriminate

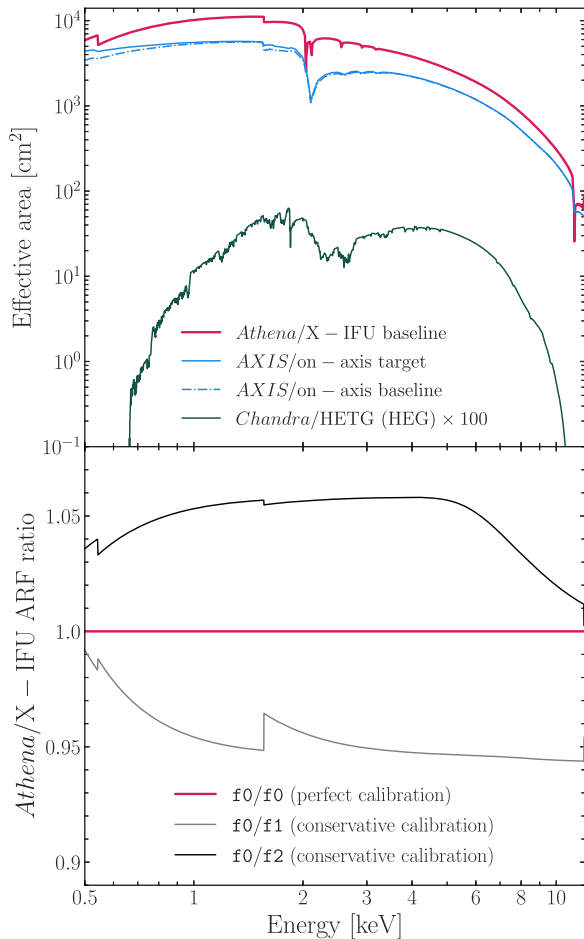


Figure 1. Top panel: optimal (baseline) effective area responses of Athena/X-IFU, Chandra/HETG (when its HEG is read on ACIS-S), and AXIS (for on-axis observations under its baseline and target requirements). Bottom panel: ratio of the effective areas of Athena/X-IFU’s optimal response (in magenta) divided by two realistic conservative responses (f_1 , f_2). We refer the reader to Barret & Cappi (2019) for a description on how the latter were produced.

further among photon–ALP mixing models suited to describe the residuals in high-quality AGN spectra.

The capabilities of undertaking ALP studies with Athena under optimal detector calibration were first studied by Conlon et al. (2017a), who considered the projected bounds on very-light ALPs from a simulated observation of the central AGN in the Perseus cluster, NGC 1275, with 200 ks of exposure. They found projected bounds on very-light ALPs of $g_{a\gamma} < 1.5 \times 10^{-13} \text{ GeV}^{-1}$ at 95% confidence.

In this paper, we revisit the potential of Athena to constrain ALPs by extending the previous considerations of Conlon et al. (2017a) to include the effects of instrumental miscalibration. Experience with all previous X-ray spectroscopes has highlighted the difficulty of obtaining (relative) calibration errors below the 1%–2% level. Importantly, the performance requirements of the X-IFU are subject to knowing both the broadband shape and normalization (within 0.5–10 keV and at 1 keV, respectively) of the effective area curve with above 97% precision. We refer to Section 5 of Barret & Cappi (2019) and Section 2 of Barret et al. (2023) for detailed descriptions of these requirements, which imply that any high-quality spectrum of a sufficiently bright AGN seen by Athena/X-IFU will be dominated by systematic uncertainties, namely detector miscalibration. Generically, this would not

apply to high-quality AGN spectra inferred by Chandra or XMM-Newton since these would be dominated by statistical uncertainties due to their smaller correcting areas.

Using a library of realistic (conservative) Athena/X-IFU detector responses (such as in Barret & Cappi 2019), we show that the application of the standard analysis techniques in the presence of calibration errors hobbles the ALP constraints from Athena/X-IFU data. However, spectral features resulting from instrumental miscalibration are generally quite distinct from those induced by ALP–photon mixing. We show that the application of machine-learning techniques can, to a very large extent, circumvent calibration uncertainties when evaluating Athena’s projected bounds on ALPs. We conclude that the use of such techniques may be key in unveiling the potential of other next-generation X-ray observatories to probe ALP parameter space further.

We also present the first projected bounds on ALPs from the Advanced X-ray Imaging Satellite (AXIS), a probe-class concept that will be proposed to NASA for a 2032 launch. While AXIS has a smaller collecting area and significantly poorer spectral resolution than Athena, its superior spatial resolution of 1'' half-power diameter (HPD) offers its own advantages for studying cluster-embedded AGNs.

This paper is organized as follows. Section 2 summarizes the Athena and AXIS missions. We proceed by simulating a 200 ks Athena/X-IFU and AXIS observation of NGC 1275 under the assumption of optimal detector responses (i.e., optimal calibration, Section 3). The grid of photon–ALP mixing models employed in our analysis pipeline is described in Section 4. The projected bounds on very-light ALPs from such missions are presented in Section 5. In Section 6, we address ALP constraints with Athena in the presence of realistic calibration errors by marginalizing over a set of nonoptimal detector responses. In Section 7, we introduce the exciting prospect of using machine-learning techniques—in particular, deep neural networks—to disentangle the effects of detector calibration and photon–ALP signatures for Athena and next-generation X-ray observatories. In Section 8, we present the projected recovery of several “injected” ALP signals by Athena/X-IFU under the effects of detector calibration and the magnetic field model. We discuss our results and limitations and conclude in Sections 9 and 10, respectively.

2. The Athena and AXIS Observatories

The Athena X-ray Observatory is the second large (L)-class mission selected by the European Space Agency within the Cosmic Vision Program to address the Hot and Energetic Universe scientific theme (Nandra et al. 2013). In this work, we consider the baseline Athena design and capabilities as described by Barret et al. (2020). Athena will consist of a single X-ray telescope constructed from iridium-coated silicon pore optics with a focal length of 12 m, permitting imaging with a half energy width (HEW) of 5 arcsec and a mirror effective area at 1 keV of 1.4 m². A hexapod mechanism allows the focused X-ray beam to be directed at one of two focal plane instruments: the Wide Field Imager (WFI), consisting of an array of active pixel sensors based of DePFET, and the X-IFU, consisting of a cryogenically cooled microcalorimeter. Given its unprecedented spectral resolution and effective area (Table 2), the X-IFU will provide the high-quality view of the intrinsic AGN emission and nearby ICM needed for sensitive ALP constraints.

AXIS is a response to the NASA’s Astrophysics Probe Explorer Program and answers the call by the Astro2020 Decadal Survey for a probe-class X-ray or mid-Infrared Satellite to launch in 2032 (National Academies of Sciences, Engineering, and Medicine 2021). AXIS will be a high-spatial-resolution X-ray observatory that will provide powerful synergy and complementarity with other facilities expected in the 2030s, including Athena.

AXIS achieves 1 arcsec HPD image quality across a 24×24 arcmin² field of view and a mirror effective area of 7000 cm² via a novel mirror design that employs precision cut monocrystalline silicon foils. The detector consists of a cooled fast-readout digital CCD, and is protected from both optical light and molecular contamination by a set of filters. Including the filter transmission curves and the detectors’ quantum efficiency, the effective area of the observatory is 5700 cm² at 1 keV. That is, an order of magnitude greater than Chandra when used in imaging mode with the Advanced CCD Imaging Spectrometer (ACIS).

The energy-dependent effective areas of Athena/X-IFU, Chandra/HETG, and AXIS are shown in Figure 1. All responses present a narrow drop at 1.5–2.0 keV, corresponding to the absorption edge of the material coating the mirror. For the baseline ARF scenario, AXIS’s effective area (for on-axis observations of point-like sources) would differ from that in the target scenario due to a limited performance at energies below 1 keV due to a significantly thicker optical blocking filter.

3. Simulating Athena and AXIS Observations of NGC 1275

NGC 1275 is the central BCG of the Perseus cluster, the most massive ($M_{200} = (6.6 \pm 0.4) \times 10^{14} M_{\odot}$; Simionescu et al. 2012) nearby cluster (at redshift $z = 0.017284$; Hitomi Collaboration et al. 2018a) whose ICM structure and dynamics have been extensively studied over the years with numerous X-ray missions (e.g., Fabian et al. 2006; Simionescu et al. 2012; Zhuravleva et al. 2014). NGC 1275 hosts a luminous AGN that is a prime target for ALP searches.

We refer to Reynolds et al. (2021) for a recent X-ray study of this AGN and its circumnuclear environment. Reynolds et al. (2021) employed the same 490 ks Chandra/HETG observation previously used to set tight bounds on ALPs (see Table 1 and Reynolds et al. 2020). Formerly, the study of Perseus and NGC 1275 had been the first science target for the Hitomi Soft X-ray Spectrometer, an X-ray microcalorimeter which provided unprecedented details on the dynamics and thermodynamics of the ICM emission as well as the Fe K α region within the core of NGC 1275 (Hitomi Collaboration et al. 2016, 2018b, 2018).

Below, we present the spectral models we assumed to simulate Athena/X-IFU and AXIS observations of NGC 1275. Throughout the entirety of our analysis, we assume a flat ($\Omega_{\kappa} = 0.0$) universe compatible with Λ -cold dark matter (Λ CDM) cosmology, with $H_0 = 70$ km s⁻¹/Mpc, $\Omega_m = 0.3$, and $\Omega_{\Lambda} = 0.7$ (Aghanim et al. 2020). At a redshift of $z = 0.017284$, this places NGC 1275 at a luminosity distance of 75 Mpc. We assume a hydrogen column density local to the Milky Way of $N_H = 1.32 \times 10^{21}$ cm⁻² (Kalberla et al. 2005) and use the element abundance ratios of Wilms et al. (2000).

3.1. The Baseline Model

Using the 490 ks Chandra/HETG observation of NGC 1275, Reynolds et al. (2020) found the AGN to be well characterized

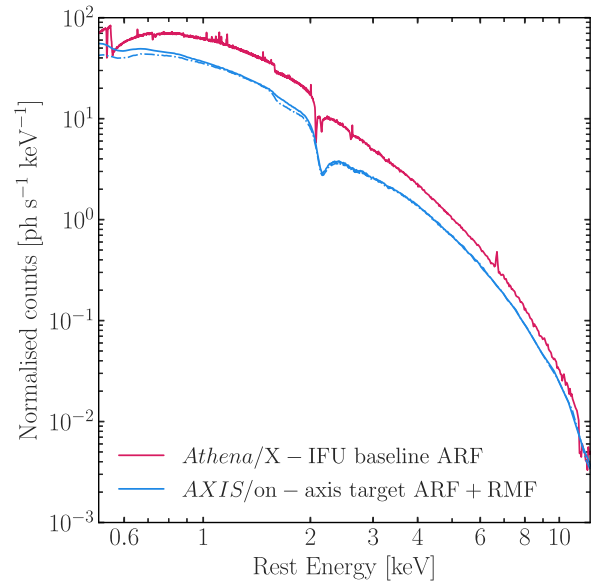


Figure 2. Effective-area-corrected simulated spectra of NGC 1275 seen by Athena/X-IFU and AXIS (used for on-axis observations under its target ARF+RMF specifications) when read on their optimal detector responses (ARFs). AXIS’s baseline and target RMFs differ only due to the improved spectral energy resolution of the latter at lower energies. The dashed-dotted line shows the predicted count rate for a 200 ks AXIS/on-axis observation of NGC 1275 under its baseline ARF+RMF requirements. For plotting purposes, an energy binning scheme of a target signal-to-noise ratio of 100 and 25, with the restriction that no more than 200 and 75 spectral bins were coadded, was applied to the Athena/X-IFU and AXIS spectra, respectively.

by an unabsorbed power law with photon index $\Gamma \approx 1.9$ modified by the effects of Galactic absorption, yielding residuals only below the 3% level outside of the iron band. In Reynolds et al. (2021), the description of the spectrum was found to improve slightly with the use of a partial covering intrinsic absorber, interpreted as a composite X-ray source with an absorbed jet working surface and an unabsorbed accretion disk corona. This interpretation also brings the X-ray data in line with the detection of molecular absorption in Atacama Large Millimeter/submillimeter Array (ALMA) observations (Nagai et al. 2019).

However, Matthews et al. (2022) found that the upper bound on $g_{a\gamma}$ inferred from Chandra data of NGC 1275 (see Table 1) remains insensitive to whether a partial covering absorber is accounted for in the spectral fits to the AGN spectrum. Thus, for simplicity, in this work we adopt the simple power-law model in order to generate simulated (fake) Athena and AXIS spectra of NGC 1275. These simulated spectra were produced with PYXSPEC’s *fakeit* command under the consideration of Poisson noise statistics and neglecting background emission. In all cases, we consider an on-source exposure of 200 ks for comparison with Conlon et al. (2017a).

3.2. Simulated AXIS Observations of NGC 1275

Given AXIS’s superior spatial resolution (see Table 2), we assume that it can capture a spectrum of NGC 1275 free from the contributions of the surrounding ICM. Thus, the baseline model for our AXIS simulations is just an absorbed power law ($tbabs * po$); see Figure 2. We note that we have not included the influence of photon pileup in our simulated AXIS spectra, as follows. In this work, we operate under the assumption that AXIS will have bright-source modes which permit accurate

Table 3
Top: Model Parameters Used to Fit the Optimal Simulated AXIS (Section 3.2) and Athena/X-IFU (Section 3.3) Observations of NGC 1275

Model Parameters: Simulated Spectra of NGC 1275 Under Optimal ARFs				
Component	Parameter	Description	Athena/X-IFU	AXIS
tbabs	N_{H}	ISM column density	$1.32 \times 10^{21} \text{ cm}^{-2}$	-
pow	Γ_X	Photon index	1.89	-
pow	A_X	Power-law normalization at 1 keV	$8.28 \times 10^{-3} \text{ phot s}^{-1} \text{ keV}^{-1}$	-
bvvapec	kT_{plasma}	Plasma temperature	3.969 keV	-
bvvapec	σ_{broad}	Doppler velocity broadening	156 km s^{-1}	-
bvvapec	A_{plasma}	Plasma normalization	$2 \times 10^{-3} \text{ cm}^{-3}$	-
bvvapec	z	Redshift	0.017284	-
bvvapec	$Z_{\text{Si}}, Z_{\text{S}}, Z_{\text{Ar}}, Z_{\text{Ca}}, Z_{\text{Mn}}, Z_{\text{Fe}}, Z_{\text{Ni}}$	Element abundances	$\neq Z_{\odot}$	-
bvvapec	Z_{other}	Element abundances	Z_{\odot}	-
Free parameters in fits to the simulated spectra of NGC 1275				
pow	Γ_X	Photon index	✓	✓
pow	A_X	Power-law normalization at 1 keV	✓	✓
bvvapec	A_{plasma}	Plasma normalization	✓	-

Note. The *tbabs*, *po*, and *bvvapec* parameter values were adopted from Kalberla et al. (2005), Reynolds et al. (2020), and Hitomi Collaboration et al. (2018; see their Table 1), respectively. Bottom: layout of the free parameters of the ALP-containing and fiducial astrophysical model simulated spectra of NGC 1275.

CCD spectroscopy of sources with fluxes at least as high as a few milliCrab.

3.3. Simulated Athena/X-IFU Observation of NGC 1275

Given an angular resolution of $5''$ HEW, an Athena/X-IFU observation of a cluster-hosted AGN will necessarily suffer from ICM contamination. However, excellent spectral separation of the AGN and ICM emission should be possible given the combination of the Athena/X-IFU effective area and spectral resolution.

We simulate an Athena/X-IFU observation of NGC 1275 using the *bvvapec* single-temperature plasma model to model the velocity-broadened ICM emission lines. All model parameters were taken from the reference *apec-v.3.0.8* model that had been fitted to data from Hitomi Collaboration et al. (2018; see their Table 1), with the exception of its normalization parameter (A_{plasma} in Table 3), for reasons specified below. The element abundances within *bvvapec* of the elements not listed in Table 1 of Hitomi Collaboration et al. (2018) were assumed to be 1.0 relative to the protosolar metallicities of Lodders & Palme (2009).

Given Hitomi’s $0.5'$ angular resolution, the normalization of the *apec-v3.0.8* model fitted to the Hitomi Collaboration et al. (2018) data modeled the ICM emission across a large volume within Perseus. Considering Athena’s improved angular resolution compared to Hitomi’s, a suitable value of A_{plasma} for our analysis was found as follows.

We first identified the single longest observation of NGC 1275 taken with Chandra. This observation (ObsID 4952; exposure 164.24 ks) had started in October 2004 and had been read on the ACIS-S detector. We reprocessed these data with the CIAO software⁵ v.4.14. We extracted a spectrum of the ICM emission from an annuli centered on NGC 1275 with inner/outer radii of $1''/10''$ and $2''/10''$ in an attempt to isolate the ICM emission in a plausible Athena extraction region. To first order, these spectra were found to be well described by an *apec* model with a normalization component of $2 \times 10^{-3} \text{ cm}^{-3}$. We therefore adopted the value of

$A_{\text{plasma}} = 2 \times 10^{-3} \text{ cm}^{-3}$ in our simulated Athena/X-IFU spectrum of NGC 1275 (see Table 3).

Our simulated Athena/X-IFU spectra of NGC 1275 were generated using a spectral model consisting of this ICM component added to an absorbed power-law model, that is, via *tbabs* * (*po* + *bvvapec*). The resulting Athena/X-IFU view of NGC 1275 is shown by Figure 2, where the narrow emission lines arising from the cluster ICM are immediately apparent when compared to the AXIS view of the AGN.

4. Grid of Photon–ALP Mixing Models

In order to find projected bounds on ALPs from Athena and AXIS, we generate a library of photon–ALP mixing models for photons propagating from NGC 1275 out to the virial radius of the Perseus cluster, estimated to be $R_{200} = 1.8 \text{ Mpc}$ (Simionescu et al. 2012).

To produce the photon–ALP mixing models, we must assume a magnetic field structure. We use a cell-based approach to model the turbulent cluster field, specifically adopting “Model B” from Reynolds et al. (2020) to determine the radial profile of the cluster magnetic field and distribution of cell sizes (“coherence lengths”). This model is chosen so as to reproduce a value of the thermal-to-magnetic pressure ratio of $\beta_{\text{plasma}} = 100$ up to 1.8 Mpc from the cluster center. “Model B” from Reynolds et al. (2020) broadly reproduces the Faraday RMs within Perseus presented by Taylor et al. (2006), where the cluster acts as a Faraday screen. Having been motivated by RM observations of cool-core clusters (Taylor et al. 2006; Böhringer et al. 2016), the value $\beta_{\text{plasma}} = 100$ has been adopted in previous cluster–ALP searches (Reynolds et al. 2020; Sisk-Reynés et al. 2021; Matthews et al. 2022). The derived constraints on $g_{\text{a}\gamma}$ can readily be scaled to other choices of β_{plasma} (see Section 5.1 of Sisk-Reynés et al. 2021). We also refer the reader to Matthews et al. (2022) for a thorough exploration of the effects of different assumptions of the turbulent magnetic field structure (e.g., Gaussian random fields (GRFs) and “cell-based” models) on cluster–ALP bounds.

⁵ See <https://cxc.cfa.harvard.edu/ciao/>

We generate 216 realizations of the turbulent ICM field, where each realization reproduces a cluster field profile, $B(r)$, consistent with the notion that $\beta_{\text{plasma}} = 100$ up to the virial radius. Under this assumption, the Fabian et al. (2006) pressure profile within Perseus yields an expected magnetic field strength of $6.5 \mu\text{G}$ at 25 kpc from the cluster center. The expected field profile up to 1.8 Mpc from the inner core is found by approximating the cluster as isothermal. The field strength at comoving distance r to the cluster center (shown in Figure 4 of Sisk-Reynés et al. 2021) is therefore given by $B(r) \sim 1/\sqrt{n_e(r)}$ and normalized according to its value at 25 kpc from the cluster center. Here, $n_e(r)$ is the analytic Churazov et al. (2003) electron number density profile found from fitting XMM-Newton observations of Perseus.

For each realization, the cluster line of sight is split into coherence lengths in the 3.5–10 kpc range under a power-law probability distribution (Vacca et al. 2012; Berg et al. 2017). We highlight that, as specified in Section 3 of Reynolds et al. (2020), their “Model B” accounted for the expected increase in field coherence lengths away from the cluster center. In each coherence length (or cell), the components of the magnetic field perpendicular to the line of sight (i.e., the field components which mix with ALPs as given by Equation (1)) are set independently and randomly.

We employ the Axion-Like PROpagation (ALPRO⁶) public-release code to construct a library of photon–ALP mixing curves for ALPs with $\log(m_a/\text{eV}) = -14.0$ across a grid of $\log(g_{a\gamma}/\text{GeV}^{-1}) \in [-14.0, -11.0]$, in steps of 0.1 dex, for 216 configurations of the turbulent cluster field. Although these curves are computed for a specific ALP mass, these models will accurately capture photon–ALP mixing for “effectively” massless ALPs with $\log(m_a/\text{eV}) < -12.0$. Although the extrapolation to higher (but still very light) ALP masses will generally depend on the underlying density structure of the host ICM, this does not impact our results.

For each magnetic field realization, the photon–ALP propagation problem described in Appendix A is solved for an initially unpolarized beam propagating from the inner core to 1.8 Mpc from the cluster center, at all coherence cells within the line of sight. The contribution from all cells results in a (total) energy-dependent photon “survival” probability $P_{\gamma \rightarrow \gamma}$ for photons mixing with massless ALPs of a given $g_{a\gamma}$. All survival curves are sampled at 1.25 eV resolution (half of the spectral resolution of Athena/X-IFU) and computed across 0.1–15 keV. The details of this calculation are outlined in Sections 2 and 3 of Matthews et al. (2022).

The survival photon–ALP mixing curves for $g_{a\gamma} = 6.3 \times 10^{-13} \text{ GeV}^{-1}$ for two different configurations of the turbulent cluster field are shown by Figure 3. This corresponds to the upper bound on $g_{a\gamma}$ inferred at the 99.7% level by the most sensitive cluster–ALP single-source spectroscopy study to date (Sisk-Reynés et al. 2021).

5. Optimal Bounds from Next-generation X-Ray Telescopes

Equipped with the grid of photon–ALP mixing models introduced above, which we refer to as $\{\text{ALP}(i, g_{a\gamma})\}$, where i and $g_{a\gamma}$ label the field realization and the photon–ALP coupling, respectively, we first aim to find projected bounds on ALPs from Athena and AXIS under the assumption that the energy-dependent area of the detectors were perfectly

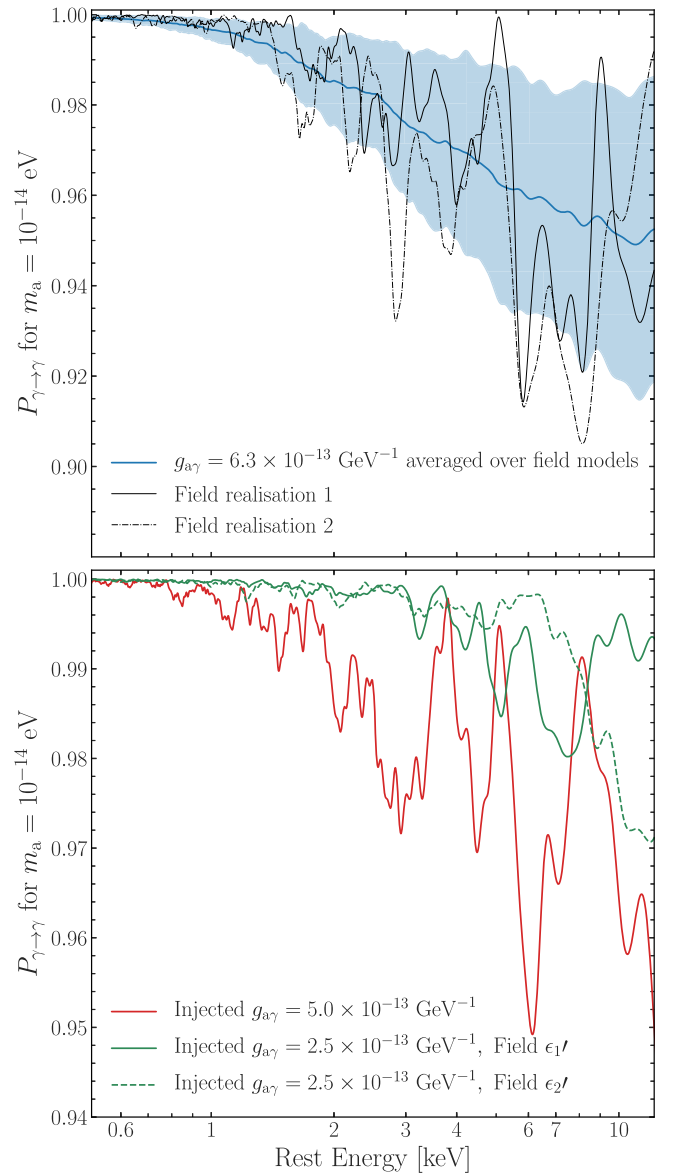


Figure 3. Top: average photon survival probability (over 216 field realizations) for quanta propagating through the Perseus cluster when mixing with massless ALPs ($m_a = 10^{-14} \text{ eV}$) of $g_{a\gamma} = 6.3 \times 10^{-13} \text{ GeV}^{-1}$. The black solid and dotted-dashed lines show the survival curves for two different turbulent field realizations (respectively referred to as ϵ_1 and ϵ_2 in the text). The shaded region shows the standard deviation of 216 curves in each energy bin, computed by using 11,920 bins between 0.1 – 15 keV. Bottom: photon–ALP mixing models showing the injected signals in the Athena/X-IFU analyses discussed in Section 8. The solid and dashed green curves illustrate ALP models generated under two different turbulent field models (ϵ_1' and ϵ_2' , respectively).

known. We highlight that Athena’s projected bounds on ALPs were first reported in Sections 4 and 5 of Conlon et al. (2017a), albeit upon a slightly different methodology to the one we follow (see their Section 4).

To derive the projected constraints on ALPs, we analyze simulated ALP-free Athena and AXIS spectra following the procedure described in Section 4 of Sisk-Reynés et al. (2021). We fit the simulated spectra in the observed energy range 0.5–12.0 keV. For each ALP($i, g_{a\gamma}$) curve in our photon–ALP mixing grid, we fit a spectral model that includes photon–ALP mixing (via $\text{tbabs}^*(\text{po} + \text{bvvapec})^*\text{ALP}(i, g_{a\gamma})$) to the simulated Athena spectrum (Section 3.3), and $\text{tbabs}^*\text{po}^*\text{ALP}(i, g_{a\gamma})$ to the (separate target and baseline

⁶ See <https://github.com/jhmatthews/alpro/tree/v1.1>

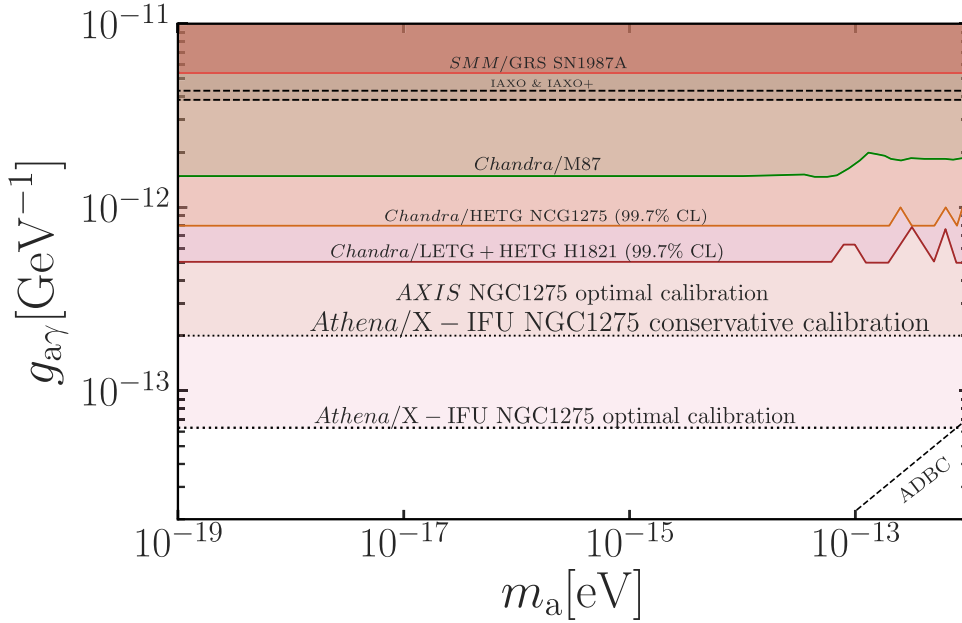


Figure 4. Projected (dotted and dashed lines) and current observational (solid lines) upper bounds (all shown at the 95% exclusion level unless otherwise explicitly stated) on very-light ALPs inferred from this work and previous astrophysical searches. We include our projected upper bounds on $g_{a\gamma}$ from both the Athena/X-IFU and the AXIS observatories under the assumption of a “perfect” detector calibration scenario, where, in both cases, we have marginalized over 216 realizations of the turbulent field within Perseus. We also show the projected bounds on ALPs from Athena/X-IFU under the assessment of detector calibration (refer to Section 7), which overlap with the AXIS exclusion under its optimal calibration scenario. We show the upper bounds inferred from the SN 1987A event (Payez et al. 2015), complemented by Fermi/LAT observations of extragalactic supernovae (Meyer & Petrushevska 2020); projected bounds from the next-generation axion helioscope, the International AXion Observatory (IAXO) and its upgrade (IAXO+; Armengaud et al. 2019); and existing AGN/cluster bounds from Chandra observations of M87/Virgo (Marsh et al. 2017), NGC 1275/Perseus (Reynolds et al. 2020, their Model B), and H1821+643 (Sisk-Reynés et al. 2021). The shaded regions above all curves underlie the parameter space excluded by any of the studies shown. We include projected bounds from the future birefringent cavity experiment ADBC (dashed black line), a future axion interferometer which will constrain ALP DM (Liu et al. 2019).

ARF+RMF) AXIS spectra. The best-fit regime of each spectral fit was found by minimizing the Cash (C) statistic (Cash 1979; Kaastra 2017), appropriate for Poisson-distributed data. We find posteriors on $g_{a\gamma}$ for $m_a < 10^{-12}$ eV by using the Bayesian framework described in Section 4 of Sisk-Reynés et al. (2021). This procedure derives a normalized posterior distribution on $(m_a, g_{a\gamma})$ by marginalizing over all magnetic field models. This normalization is achieved via Equation (C1) and computed across $\log(m_a/\text{eV}) \in [-30.0, -12.0] \cup \log(g_{a\gamma}/\text{GeV}^{-1}) \in [-19.0, -10.1]$. Operationally, all posteriors for $\log(m_a/\text{eV}) = -14.0$ are identical to those for ALPs of $\log(m_a/\text{eV}) \in [-30.0, -12.0]$. Moreover, all $p_{\text{post}}(m_a, g_{a\gamma})$ for $\log(g_{a\gamma}/\text{GeV}^{-1}) \in [-19.0, -14.0]$ were assigned by taking those for $g_{a\gamma} = 1.0 \times 10^{-14}$ GeV^{-1} as a proxy. That is, in this regime, the energy-dependent modulations induced by $g_{a\gamma}$ are indistinguishable from those induced by Poisson noise.

Our priors on the ALP parameter space were chosen as follows. First, the mass range considered embeds the phenomenology of “effectively” massless ALPs, that is, where the relevant photon–ALP mixing equations (see Appendix A) are mass independent. In general, the latter would hold for ALPs of $m_a \lesssim 10^{-12}$ eV. If such ALPs were to play a cosmological role, they could indeed comprise DM. At masses $\lesssim 10^{-30}$ eV, however, their de Broglie wavelength will exceed the particle horizon of the universe and ALPs will then act as dark energy (Carroll 1998). Second, for photon–ALP couplings $g_{a\gamma}^{-1} \gtrsim 10^{19}$ GeV , the energy scale at which ALPs couple to electromagnetism will exceed the Planck scale, which is not expected to be realizable in consistent theories of quantum gravity.

The 95% (2σ) confidence level (CL) constraints on $g_{a\gamma}$ we infer from the simulated 200 ks Athena/X-IFU and AXIS observations of NGC 1275 under the “perfect calibration” scenario are listed in Table 1 and shown in Figure 4. These projected Athena bounds significantly improve on the current best constraints (Table 1), with Athena pushing a full order of magnitude deeper than Chandra. We refer to Section 9.1 for a comparison with the results in Conlon et al. (2017a). The marginalized posterior distribution (over all magnetic field realizations) for ALPs of $m_a = 10^{-14}$ eV is expected to be dominated by low values of photon–ALP coupling, i.e., $g_{a\gamma} \lesssim 6.3 \times 10^{-14}$ GeV^{-1} (refer to Appendix B for a discussion).

In the perfect calibration scenario, AXIS improves upon current constraints by 0.3–0.4 dex for both its baseline and target area responses. The latter is shown in Figure 4.

6. The Effects of Calibration on Athena’s Projected Bounds on ALPs

Following on from Conlon et al. (2017a), we have shown that Athena/X-IFU, with its unprecedented spectral energy resolution of 2.5 eV and effective area of about 1 m^2 at 1 keV, could push down the current most sensitive bounds on $g_{a\gamma}$ by an order of magnitude provided we have perfect knowledge of its energy-dependent area. However, no X-ray observatory has perfect calibration, with the practicalities of both ground and in-orbit calibration leading to irreducible uncertainties in the (relative) effective area curves at the level of 1%–2%. In view of this reality, we now proceed to reassess Athena’s power to constrain very-light ALPs under an imperfect calibration. The importance of such an exercise is discussed in Sections 6.1 and

in Appendix B. In Section 6.2, we then quantify the effects of detector calibration in Athena’s projected bounds on ALPs using a Cash statistic likelihood method that extends from that employed in the “optimal” calibration scenario. In Section 7, we introduce the exciting prospect of disentangling calibration residuals and ALP features via the use of machine learning, thus leading to tighter bounds on $g_{a\gamma}$.

6.1. Overview of the Effects of a Nonoptimal Calibration

One may worry that the energy-dependent residuals induced by the calibration of a given detector (encapsulated in the area response file, ARF) could mimic the energy-dependent oscillations induced by photon–ALP mixing. We examine this issue using a library of sampled effective area curves (ARFs) for Athena/X-IFU, which was produced as described in Barret & Cappi (2019). These ARFs capture the realistic effects of instrumental modeling uncertainties on detector calibration. The effects of modeling uncertainties of the mirror response on detector calibration were not taken into account and are therefore beyond the scope of this paper. These mirror-induced calibration effects were recently explored by Guainazzi et al. (2022). For reference, Figure 1 illustrates a subsample of the realistic “nonoptimal” ARFs employed in our analysis.

Together with the energy-response matrix, the ARFs relate the observed photon fluxes to the actual source spectrum. A given detector’s ARF is sensitive to multiple design features, for instance, the level of impurity and thickness of the material coating the mirror, the orientation of the mirror, etc. Therefore, in general, a representative set of ARFs for next-generation X-ray observatories must be constructed in order to assess the extent to which these missions will be able to meet their science objectives.

We assess the effects of detector calibration on Athena’s projected bounds on ALPs using the library of ARFs outlined in Section 5 of Barret & Cappi (2019). Each ARF in the library was generated so that its normalization remained unchanged at 1 keV and throughout 0.5–10 keV compared to the nominal ARF, and was subjected to a change in the effective area shape of up to 3% with respect to the nominal ARF. As the end product, each ARF in the library captured the uncertainty in the thickness of the instrument filters and absorbers without accounting for uncertainties around the edge of the response. The latter will be considered in a future paper.

We can obtain a first estimate of the effect of calibration expected from pure Poisson statistics by assuming that the “true” ARF is an unknown member of the library of ARFs. Following Drake et al. (2006), one can then proceed by (i) either simulating a single spectrum with the nominal ARF and fitting it across all ARFs in the library, (ii) or by generating a single spectrum per ARF in the library and subsequently fitting it with the nominal ARF. Following Section 5 of Barret & Cappi (2019), we use the former method as it is able to capture more closely the perturbative nature of detector miscalibration in the recovery of best-fit parameter values; and we refer to Cucchetti et al. (2018) for an implementation of the latter method. In both approaches, the realistic effects of detector miscalibration can be inferred from a Monte Carlo approach by recording the best-fit parameter values when a spectral model is fitted to the simulated spectra. We refer to Figure 5 and Appendix B for a qualitative assessment on the realistic impact of calibration in Athena’s sensitivity to constraining ALPs.

6.2. Revisiting Athena’s Projected Bounds on ALPs Under the Assessment of Detector Calibration

In this section, we examine the impact of realistic imperfections in the calibration of the ARF on our ability to set bounds on the ALP–photon coupling constant using our standard C -statistic likelihood approach. We simulate an ALP-free Athena/X-IFU spectrum of NGC 1275 using a particular choice of ARF. We then fit the simulated spectrum across both our library of ALP–photon survival curves (including scans over both the ALP–photon coupling constant and the magnetic field realizations) as well as a subset of the library of calibration responses (i.e., ARFs). The normalized posterior distribution on ALPs must then be found by marginalizing over magnetic field models and ARFs. The Bayesian framework we follow is described in Appendix C.

Marginalizing over 216 ARFs and over 216 magnetic field realizations, we see that our standard C -statistic likelihood technique for setting bounds on $g_{a\gamma}$ is significantly influenced by imperfect knowledge of the calibration, as follows. The excluded region weakens to $g_{a\gamma} \geq 6.3 \times 10^{-13} \text{ GeV}^{-1}$ both at the 95% and 99.7% CLs compared to the value of $g_{a\gamma} > 6.3 \times 10^{-14} \text{ GeV}^{-1}$ excluded at 95% CL in the “perfect” calibration scenario, implying a weakening of Athena’s bounds on ALPs by an order of magnitude.

The essential reason for this is that our standard technique for assessing bounds on the ALP–photon coupling constant simply examines the significance of X-ray spectral distortions away from the astrophysical model, and *not* the nature of the distortions. However, as we see from a comparison of Figure 3 and Figure 1, the typical characteristics of instrumental-calibration-related features are very distinct from those of the ALP–photon conversion, respectively, with the latter having significant quasiperiodic structure. In the next section, we show that the application of machine learning can use these differences to disentangle ALP-induced and calibration-induced effects partially, thereby restoring to a large degree the ability of the Athena/X-IFU to constrain ALPs.

7. Circumventing Detector Calibration Limits with Neural Networks

Because artificial neural networks (ANNs) do not rely on a fit statistic to derive a detection confidence, it is possible that they may be able to circumvent the reduction in confidence associated with calibration uncertainties. For a proof-of-concept test we adapt the parameter estimation scheme from Parker et al. (2022), and train ANNs to recover $g_{a\gamma}$ from raw spectra. If a strong ALP signal is present in the spectrum, the ANN should recover the value of $g_{a\gamma}$ of the input spectrum, and a detection limit can be derived from the point where this recovery breaks down.

First, we tested several different network architectures, including the two designs from Parker et al. (2022). In the first of these architectures, the spectra were normalized and directly input into the first layer of the ANN. In the second architecture, a principal component analysis (Pearson 1901) was first applied to reduce the dimensionality of the spectroscopic data (counts per energy bin), making it easier for the ANN to learn the relevant features and to dismiss the redundant features of the spectroscopic data while reducing the risk of overfitting. However, PCA preprocessing was found not to be

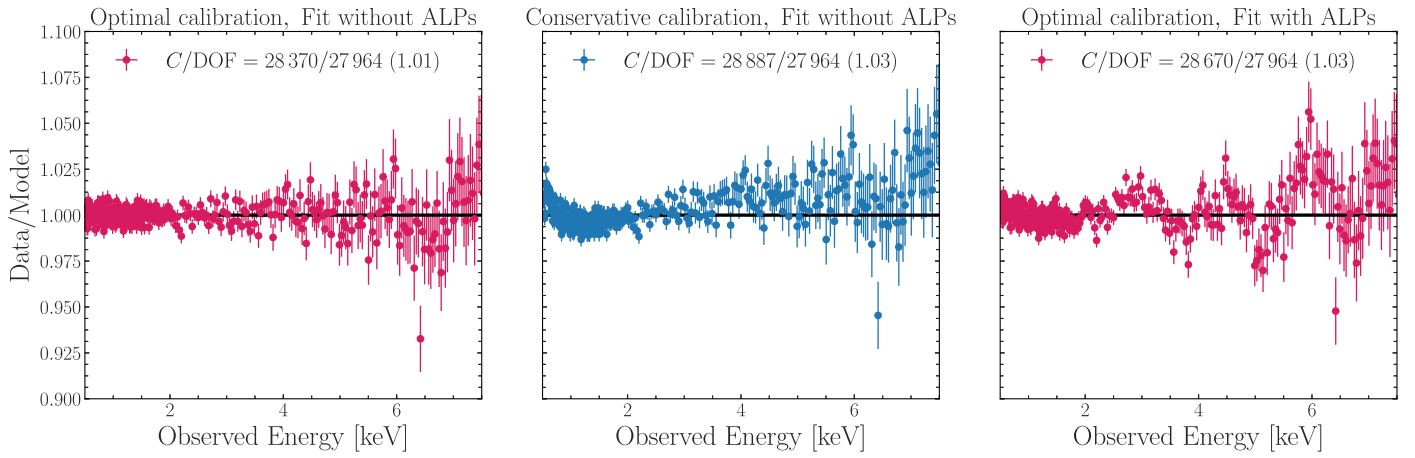


Figure 5. Residuals from fitting the synthetic Athena/X-IFU spectrum presented in Section 3 with a variety of models. The different panels illustrate the effects of detector calibration and of photon–ALP mixing when fitting the synthetic data as follows. Left panel: a “fiducial” (astrophysics only) model is fitted under an optimal detector calibration scenario. Central panel: a “fiducial” (astrophysics only) model is fitted to the data under a conservative calibration scenario. Right panel: a multiplicative ALP model (featuring $g_{a\gamma} = 5.0 \times 10^{-13} \text{ GeV}^{-1}$) is fitted under an optimal calibration scenario. The simulated data were fitted within the observed energies 0.5–12 keV. For plotting purposes, a binning scheme of a target signal-to-noise ratio of 250 with the restriction that no more than 100 spectral bins were coadded was applied, and we show the energy range 0.5–7.5 keV. The best-fit C -statistic of each fit, the number of DOF, and the reduced C -statistic ($\equiv C/\text{DOF}$) are quoted (with the latter in parenthesis). The ALP fit corresponds to $\log(g_{a\gamma}/\text{GeV}^{-1}) = -12.3$, i.e., close to the current exclusion on $g_{a\gamma}$ (refer to Table 1).

effective given the complexity of the ALP-induced features in our synthetic Athena/X-IFU spectra.

Finally, we tested a convolutional neural network (CNN), where the spectroscopic data pass through a series of convolutional layers. In the more common use of CNNs for image recognition, where the convolutional layers identify structure within the data, such as edges or lines, CNNs determine the important features of the data and optimize for such features themselves. As the data pass through multiple layers, this information becomes more abstracted and more condensed. In our CNN implementation, after each convolutional layer we include a max pooling layer, which down-samples the data, taking the highest value from each N data points (where N is the degree of pooling). This compresses the data, and allows CNNs to be invariant to small shifts in the position of features. We find that the CNN generally outperforms the two ANN architectures described above, for which we focus on the CNN for the remainder of this section.

The CNN was thereby trained on 10,000 simulated Athena/X-IFU spectra, with values of $\log(g_{a\gamma}/\text{GeV}^{-1}) \in [-15.0, -11.0]$ in steps of 0.1 dex, and randomly allocated response files and magnetic field realizations from a library of 1001 representative response files and a set of 300 configurations of the turbulent cluster field. The data are truncated to the first 20,000 spectral bins and rebinned by a factor of 5. We then normalized and detrended the spectra using the `SCIPY.SIGNAL.DETREND` function. The latter implements a linear detrending of the data. We find our results (shown by Figure 6) to be insensitive to whether linear or a higher-order polynomial detrending is employed. This applies to both the 95% and 99.7% exclusion levels on $g_{a\gamma}$ inferred by the CNN and to the distributions illustrated by Figure 6. We used the KERAS deep-learning API (Chollet 2015) for PYTHON, running on the TENSORFLOW machine-learning platform (Abadi et al. 2015). We used the Adam optimizer (Kingma & Ba 2014), the rectified linear unit (ReLU) activation function (Nair & Hinton 2010), the mean squared error loss function, and early stopping (Prechelt 1998) with a patience of 50. We followed each convolutional layer with a max pooling layer, downsampling the feature map. Our CNN design is summarized in Figure 7.

We performed a basic hyperparameter optimization, randomly selecting 200 combinations of values of the number of convolutional layers, number and breadth of dense layers, number of filters, learning rate, batch size, and pooling size. In each case, we evaluated the performance of the network on a test set of 2000 spectra and selected the best-performing network based on the mean squared error (Figure 7). Finally, we used the network to predict the value of $g_{a\gamma}$ from a set of 1001 validation spectra (one for each response file) which were not used to train the network or evaluate the performance during hyperparameter optimization. We note that the latter were generated with an independent library of 204 magnetic field configurations within Perseus.

The recovered value of the photon–ALP coupling as a function of its simulated value is shown by the main panel in Figure 6. We refer to these variables as $\log(g_{\text{sim}})$ and $\log(g_{\text{fit}})$, respectively, where both g_{sim} and g_{fit} are expressed in GeV^{-1} . The absolute value of these variables is shown in Figure 6, where the photon–ALP coupling value decreases along both axes.

We interpret the CNN regression method by referring the reader to the main panel of Figure 6. For sufficiently high values of $g_{a\gamma}$, the photon–ALP-induced oscillations are distinguishable from the effects of detector calibration. Clearly, the CNN is successful at identifying the presence of ALPs when $\log(g_{\text{sim}}) \gtrsim -12.3$, albeit with a bias of 0.2–0.4 dex in $\log(g_{\text{rec}})$. The flattening of g_{fit} at $|\log(g_{\text{sim}})| \in [11.0, 11.2)$ arises due to the averaging of $P_{a\gamma}$ to 0.5 at hard energies from an unpolarized photon beam (see Section 4 of Marsh et al. 2017).

In the intermediate range of $\log(g_{\text{sim}}) \in [-12.7, -12.4]$, the CNN is only partially successful at detecting the presence of ALPs in the input spectra. In such successful scenarios, one would expect a favorable magnetic field model that enables the CNN to disentangle correctly between the effects of detector calibration, ALP-induced features, and Poisson noise in the synthetic data. However, for simulated values of the photon–ALP coupling below the intermediate range, i.e., for $\log(g_{\text{sim}}) \leq -12.8$, the CNN is unable to recover the ALP signal as the amplitude and shape of the ALP-induced

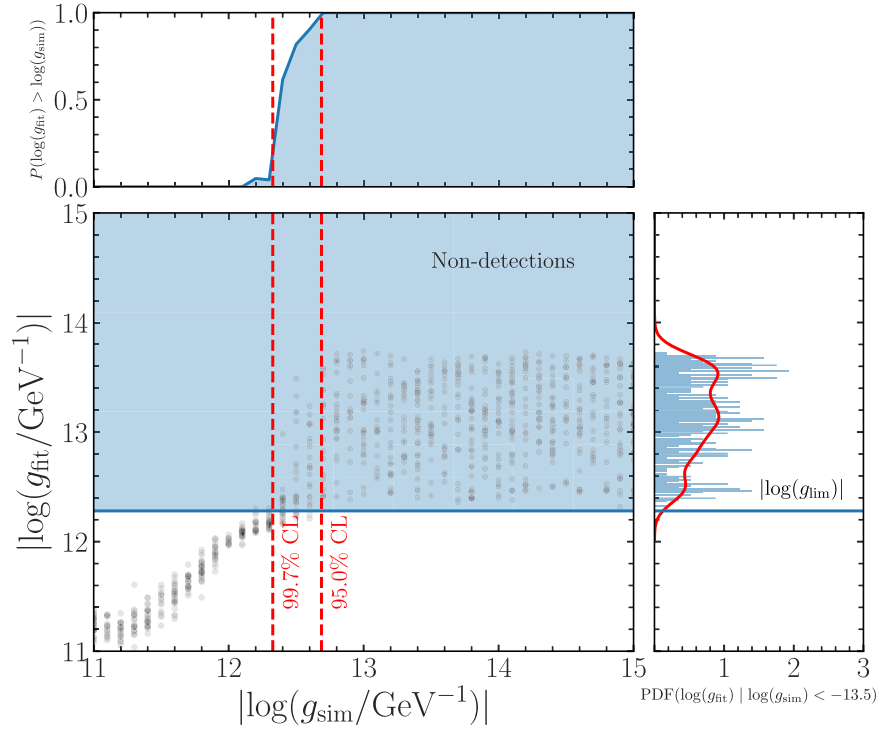


Figure 6. Main panel: recovery of the value of $g_{a\gamma}$ as inferred by the artificial neural network (ANN) introduced in Section 7. The x -axis of the main panel shows the absolute simulated value of the photon–ALP coupling, while the y -axis shows its parameter estimation by the ANN. The horizontal line shows the detection threshold $|\log(g_{\text{lim}})|$, defined as the 99% limit of the distribution of $\log(g_{\text{fit}})$ for $\log(g_{\text{sim}}) < -13.5$. This is shown in the right panel, where the red line shows a Gaussian kernel density estimation. The dashed lines show the limits derived from the posterior distribution of $|\log(g_{\text{sim}})|$ for all spectra that do not have ALP detections (shown in the top panel). We refer the reader to Section 7 for a discussion.

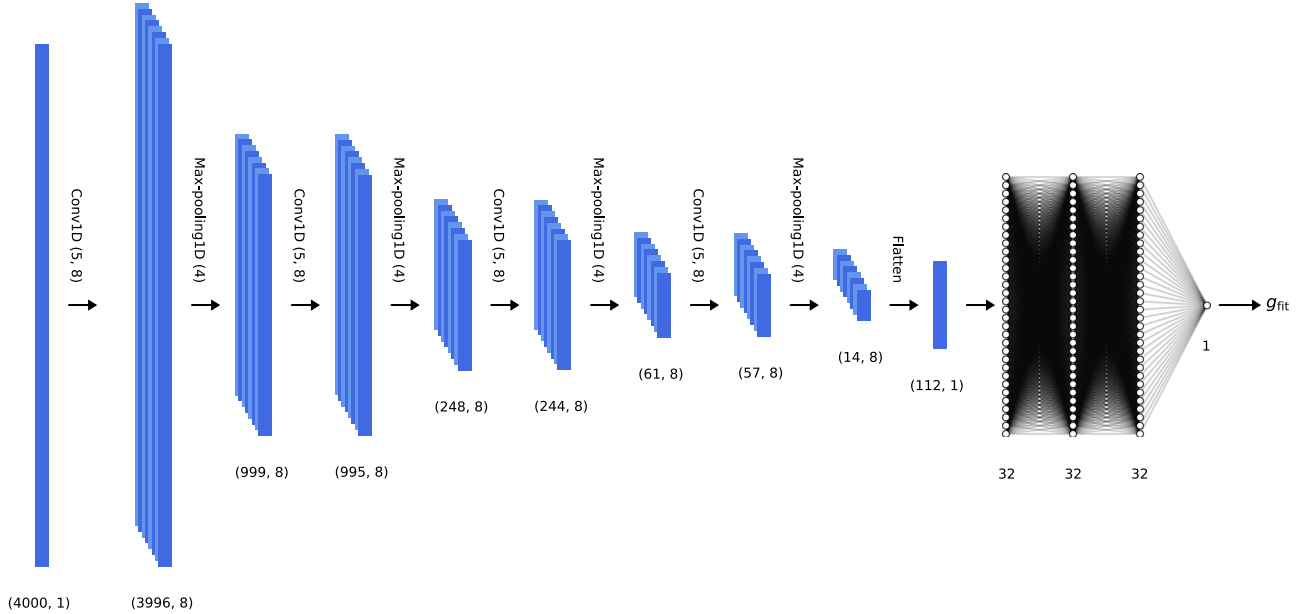


Figure 7. Architecture of the CNN used to estimate the value of $g_{a\gamma}$. The data input into the CNN are represented by the first blue column on the left. These pass through a series of 1D convolution and pooling layers, which extract features from the data and then downsample it by a factor of 4 (see first paragraph of Section 7). The final feature map is then flattened, resulting in the final single column, before being input into a series of fully connected layers which estimate $g_{a\gamma}$ from the feature map. The numbers in brackets below the columns represent the size of the data or feature map at each stage in the convolutional layers, and the numbers below the fully connected layers give the number of nodes in each layer.

perturbations are comparable and indistinguishable to those of the Poisson noise in the data. In this regime, the CNN “guesses” a value of g_{fit} within the range of g_{sim} which the CNN cannot adequately recover in order to minimize the loss function.

We compute the 95% and 99.7% CLs on $g_{a\gamma}$ inferred by the CNN as follows. To begin with, we consider the probability of obtaining a given value of $\log(g_{\text{fit}})$ from $\log(g_{\text{sim}})$. First, we define a detection threshold. We use the points $|\log(g_{\text{sim}})| \geq 13.5$ as proxies for spectra with no injected

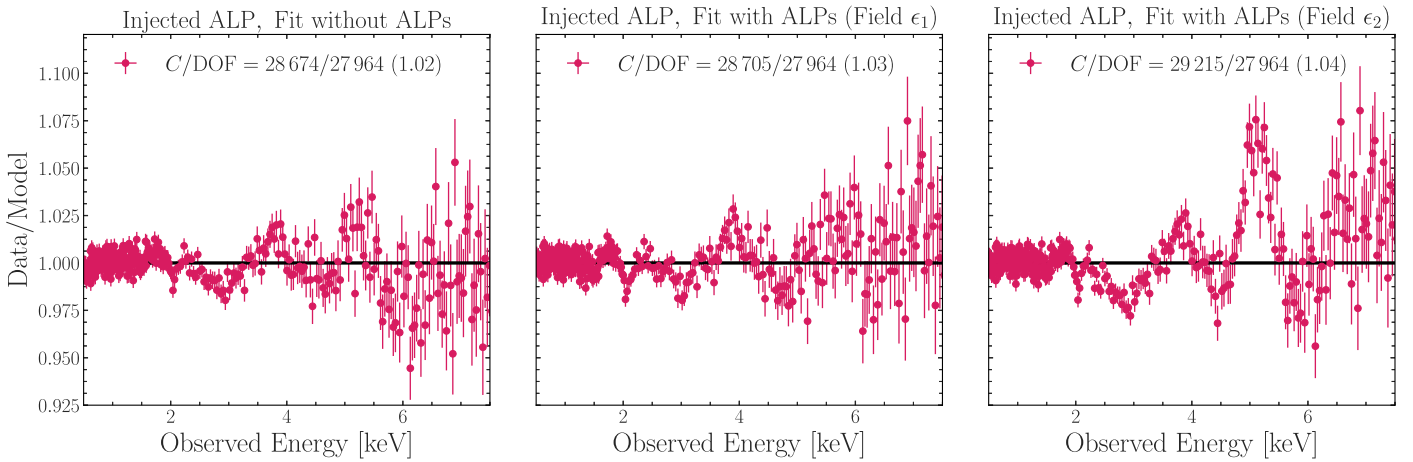


Figure 8. Residuals from fitting one of the Athena/X-IFU-injected ALP spectra presented in Section 8 with a set of spectral models. For illustration purposes, all panels show the residuals under the optimal calibration scenario. The photon–ALP model employed as the injected signal is shown in the bottom panel of Figure 1, and features the “true” value of $g_{a\gamma} = 5.0 \times 10^{-13} \text{ GeV}^{-1}$. The different panels illustrate the importance of the field model when employing a cell-based approach to describe the spectrum, as follows. Left panel: a “fiducial” (astrophysics only) model is fitted. Central and right panels: a multiplicative ALP model (featuring $g_{a\gamma} = 5.0 \times 10^{-13} \text{ GeV}^{-1}$) is fitted under a specific turbulent field geometry (referred to as ϵ_1 and ϵ_2 , respectively). The simulated data were fitted within the observed energies 0.5–12 keV. For plotting purposes, a binning scheme of a target signal-to-noise ratio of 250 with the restriction that no more than 100 spectral bins were coadded was applied, and we show the energy range 0.5–7.5 keV. The best-fit C -statistic of each fit, the number of DOF, and the reduced C -statistic ($\equiv C/\text{DOF}$) are quoted (with the latter in parenthesis).

ALP signal. We calculate the distribution of $\log(g_{\text{fit}})$ for these spectra. The latter is shown by the right panel of Figure 6. We then define a detection (“limit”) threshold, $\log(g_{\text{lim}})$, such that there is a 1% false positive rate. We then find the posterior distribution of $\log(g_{\text{sim}})$ for all spectra that have $\log(g_{\text{fit}}) > \log(g_{\text{lim}})$. The probability that this distribution corresponds to is shown by the top panel of Figure 6. Adopting the same priors in ALP parameters described in Section 5, we extrapolate this out to $|\log(g_{a\gamma}/\text{GeV}^{-1})| = 19.0$, and then find the 95% and 99.7% CLs of the distribution.

In general, one would expect the CNN performance to improve upon training it with a larger sample of synthetic spectra and with the use of alternative network architectures. This will be the subject of a future publication. We highlight that training the CNN on residuals from fitting synthetic data—rather than on spectra—could also help improve the CNN’s performance (see Schallmoser et al. 2021).

Clearly, the CNN permits inferring tighter projected bounds on ALPs from Athena/X-IFU under the assessment of detector calibration when compared to the C -statistic likelihood approach (Section 6.2). We refer to Section 9.3 for further discussion and for an overview of other machine-learning implementations in the context of cluster/ALP searches. We highlight that machine-learning classifiers such as support vector machines, decision tree, and random forest classifiers were used by Schallmoser et al. (2021) to tighten the bounds on ALPs previously found by Conlon et al. (2017b).

8. Injected Signal: Athena Recovery

A useful exercise that facilitates comparison with Section 7 within the C -statistic likelihood framework we have employed (Sections 5 and 6.2) is to assess to which extent Athena would be able to recover a true ALP signal under the effects of detector calibration, as follows.

We first simulated an Athena/X-IFU 200 ks observation of NGC 1275 by loading a multiplicative ALP spectrum with $m_a = 10^{-14} \text{ eV}$ and $g_{a\gamma} = 5.0 \times 10^{-13} \text{ GeV}^{-1}$ (whose photon

survival curve is shown in the lower panel of Figure 3). As in Section 3, we considered Poisson noise statistics without the assignment of a background spectral file. This injected spectrum was generated under a turbulent cluster field configuration which is not part of the 216 set of models present in our analysis pipeline (Section 4).

The effects of the turbulent magnetic field model in the recovery of the ALP signal are shown by Figure 8. Its left panel shows the residuals resulting from fitting the injected spectrum with an astrophysics-only model under a given detector calibration setup. Clearly, the injected ALP induces features in the spectrum such that residuals up to $\sim 2\%$ appear in the spectrum at high energies.

The central and right panels of Figure 8 illustrate the residuals resulting from fitting the injected spectrum with two multiplicative photon–ALP mixing models (ϵ_1 and ϵ_2 , see the upper panel of Figure 3) with $m_a = 10^{-14} \text{ eV}$ and $g_{a\gamma} = 5.0 \times 10^{-13} \text{ GeV}^{-1}$. Interestingly, the geometry of field ϵ_2 (Figure 8) is better suited to describing the injected ALP spectrum at energies 4–6 keV (Figure 8).

The injected spectrum was thereby fitted across our library of 216 possible detector responses, each of which was fitted across our library of 216 field models, following our previous analysis. The posteriors on the ALPs were inferred following the Bayesian framework described in Appendix C. The injected signal was found to be well recovered at both the 95% and 99.7% CLs, that is, all couplings $g_{a\gamma} \geq 5.0 \times 10^{-13} \text{ GeV}^{-1}$ are excluded at the 2σ level for $m_a < 10^{-12} \text{ eV}$.

We note that, given the input spectrum contains an ALP in the “massless regime”, no knowledge on the ALP mass would be recovered and would be limited to $\log(m_a/\text{eV}) \in [-30.0, -12.0]$. Nevertheless, if there were an effectively massless ALP with $g_{a\gamma} = 5.0 \times 10^{-13} \text{ GeV}^{-1}$ in the spectrum, it would indeed be detected by Athena/X-IFU, even under the assessment of detector calibration. Figure 9 shows the posterior on the recovery of such an injected signal under the effects of detector calibration, where the inferred posterior on $g_{a\gamma}$ is Gaussian distributed with a mean corresponding to the true $g_{a\gamma}$ and a standard deviation of 0.1–0.2 dex.

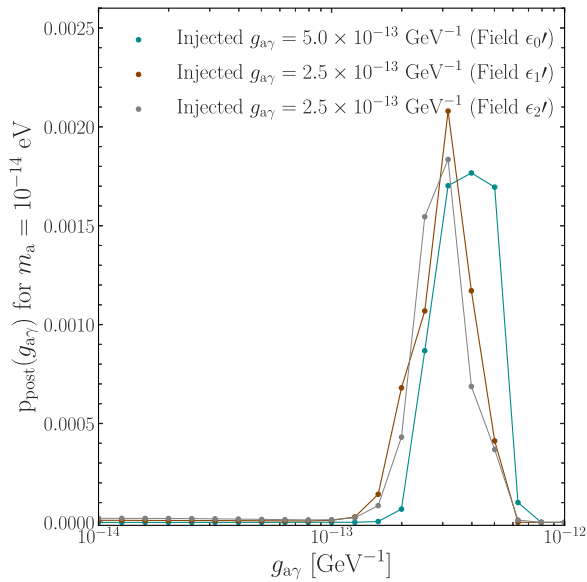


Figure 9. Normalized posterior distribution on $g_{a\gamma}$ for ALPs of $m_a = 10^{-14}$ eV (see Equation (C1)). Each color represents the posterior inferred when fitting each of the simulated Athena/X-IFU spectra of NGC 1275 which contain a specific ALP signal (illustrated in the lower panel of Figure 8). In all cases, the posteriors were inferred by marginalizing over 216 magnetic field realizations and over 216 conservative detector responses.

On the other hand, the upper bound on $g_{a\gamma}$ inferred in the previous section (see Figure 6) suggests that one could expect to recover all true $g_{a\gamma}$ up to $\log(g_{a\gamma}/\text{GeV}^{-1}) \sim -12.6$. We thereby generated two simulated Athena/X-IFU spectra, with 200 ks of exposure, by loading multiplicative ALP models with $m_a = 10^{-14}$ eV and $\log(g_{a\gamma}/\text{GeV}^{-1}) = -12.6$ using two configurations of the cluster field not embedded in the photon–ALP mixing grid introduced in Section 4. Their corresponding photon survival functions are shown in the lower panel of Figure 3. Both of these spectra were then fed into our analysis scripts, resulting in the posterior distributions shown in Figure 9. Although biased to higher photon–ALP coupling values, both signals would be detected and recovered. This suggests that, even if marginally outside of the detection threshold, a true ALP signal may eventually be recovered if existing in a favorable underlying cluster field geometry. We refer to Section 9.4 for further discussion.

9. Discussion

9.1. Bounds on ALPs from Athena and AXIS: Optimal Calibration Scenarios

Following from the work of Conlon et al. (2017a), we have shown that Athena/X-IFU, with its unprecedented effective area and spectral resolution, can in principle improve on the current ALP constraints (Table 1) by an order of magnitude for $\log(m_a/\text{eV}) < -12.0$, if its calibration is well understood.

In this work, we simulated a 200 ks Athena/X-IFU observation of NGC 1275. The cluster emission component which would be included in the Athena beam was modeled with a Doppler-broadened single-plasma temperature component, with reference free parameters taken from Hitomi Collaboration et al. (2018; see Section 9.5 for further details). We highlight that, for a 1 Ms Athena/X-IFU observation of NGC 1275, all photon–ALP couplings

$g_{a\gamma} > 4.0 \times 10^{-14} \text{ GeV}^{-1}$ would be excluded for $m_a \leq 10^{-12}$ eV under an optimal calibration scenario. Following from Reynolds et al. (2020), we have employed a fiducial model for the cluster magnetic field (summarized in Section 9.4). In our analysis pipeline, we consider 216 realizations of the turbulent cluster field.

We have also presented projected bounds on effectively massless ALPs from AXIS, under its current target and baseline on-axis response files. In both cases, we generated a fake 200 ks AXIS observation of NGC 1275, finding an upper bound of $g_{a\gamma} \sim 2.0 \times 10^{-13} \text{ GeV}^{-1}$ at the 95% CL for both responses. With its more modest spectral resolution compared with that expected for Athena/X-IFU, AXIS’s superb spatial resolution will permit a clear extraction of the intrinsic AGN emission. This will provide AXIS with exciting prospects to constrain ALPs, in particular, given the improvement compared to the current most sensitive constraints from single-source Chandra AGN observations (Table 1).

Our projected AXIS bounds on ALPs also highlight the prospects of next-generation telescopes with intrinsically different setups to probe the physics BSM such as Arcus, LEM (Sisk-Reynés et al. 2023), STROBE-X (Ray et al. 2019), and Lynx (Gaskin et al. 2019).

9.2. Bounds on ALPs from Athena: Conservative Calibration Scenario

Knowledge of detector calibration will undoubtedly make an impact on the potential of next-generation observatories to constrain the physics BSM and, in particular, very-light ALPs. In our work, we have solely focused on reassessing Athena’s projected bounds under the effects of calibration, although such a consideration will need be made for any upcoming X-ray mission.

We first employed a library of conservative calibration files and fitted the simulated 200 ks Athena/X-IFU observation of NGC 1275 across this library, and 216 models for the turbulent field for each detector response. Using the standard C-statistic likelihood procedure (see Appendix C), marginalizing over field models and detector responses, all couplings $g_{a\gamma} > 6.3 \times 10^{-13} \text{ GeV}^{-1}$ for ALPs of masses $m_a \leq 10^{-12}$ eV are excluded at the 95% and 99.7% CLs. This would imply a weakening of Athena’s projected bounds on massless ALPs by an order of magnitude, compared to the optimal calibration scenario. We note that, for a 1 Ms Athena/X-IFU observation of NGC 1275, all photon–ALP couplings $g_{a\gamma} > 4.0 \times 10^{-13} \text{ GeV}^{-1}$ would be excluded at the 95% CL on the basis of the standard C-statistic likelihood procedure.

With the aim of disentangling between the effects of cross-calibration residuals and those of photon–ALP mixing in order to acquire more sensitive bounds on ALPs, we proceeded by reevaluating Athena’s projected bounds on ALPs through the use of machine learning.

9.3. The Effect of Calibration Assessed via Machine Learning

Machine-learning techniques have increasingly been used in cluster–ALP searches (e.g., Conlon & Rummel 2019; Day & Krippendorf 2020). One of the most relevant studies pertinent to our work is that of Schallmoser et al. (2021), who employ machine-learning classifiers to find updated bounds on very-light ALPs from previous Chandra observations of AGNs centered or behind clusters, following from the work of

Conlon et al. (2017b). Under a 3D field model (see Section 9.4 for a further discussion) and for the multiclass classification method, Schallmoser et al. (2021) were able to tighten the previous upper bound on $g_{a\gamma}$ from Chandra observations of A1795Sy1 by a factor of 4, excluding $g_{a\gamma} \gtrsim 6 \times 10^{-13} \text{ GeV}^{-1}$ at the 95% CL (Table 1). In this case, the machine-learning classifier had been trained on the residuals from fitting simulated Chandra spectra of the source divided by the average survival probability across a range of $g_{a\gamma}$ values while being fitted by an astrophysics-only model. Finally, we note that Schallmoser et al. (2021) were the first to use approximate Bayesian computation (ApBC) methods, a specific type of likelihood-free inference, to find updated bounds on $g_{a\gamma}$ from their analysis (Sisson et al. 2018).

Our machine-learning study mostly differs from that of Schallmoser et al. (2021) as follows. Schallmoser et al. (2021) use a wide range of machine-learning classifiers to distinguish between spectra with and without an injected ALP signal. The classifiers had been trained on spectra with a specific coupling, and were then tested on spectra with a range of couplings. We instead test a CNN regression approach, training CNNs on data with a wide range of coupling factors to estimate the coupling as accurately as possible. Our parameter estimation approach more closely mirrors conventional spectral fitting.

The advantage of a convolutional approach is that it enables the network to “learn” effectively the important features of the data for itself, for example by detecting edges in the data, and also makes it invariant to small shifts of spectral features due to the pooling of features from adjacent bins. We speculate that the convolutional layers are able to distinguish between the shapes of residual features caused by any instrumental miscalibration and residual features due to an ALP signal based on their shapes. That is, ARF features tend to be either sharp edges or gradual broadband slopes, whereas ALP features are smoother (that is, more absorption-like). For instance, the amplitude of the ALP signal tends to increase monotonically with energy (Figure 3).

The biggest limitation of the machine-learning approach is likely the requirement for synthetic training data. This means that the accuracy and performance of the CNN is dependent on how comprehensive the simulated spectra are, as the network cannot be relied upon outside the scope of the training set. The ANNs used here are relatively simple compared to many used in deep-learning applications, so they can be trained and tested in a short timeframe (~ 5 minutes running on a laptop CPU) once the synthetic training set has been constructed.

9.4. Modeling Cluster Magnetic Fields

Our inferred upper bounds on $g_{a\gamma}$ are based upon the assumption that the magnetic field strength is scaled by the electron number density for an isothermal ICM. The inferred field profile broadly reproduces a value of the magnetic-to-pressure ratio of $\beta_{\text{plasma}} = 100$ up to 1.8 Mpc up from the cluster center. This fiducial field model, which had already been adopted in “Model B” of Reynolds et al. (2020), is supported by RM observations of cool-core clusters (Taylor et al. 2006; Govoni 2012) and by measurements of the turbulent velocities of the ICM in the Perseus cluster (Zhuravleva et al. 2014; Hitomi Collaboration et al. 2018c). Moreover, magnetohydrodynamic (MHD) simulations of galaxy clusters predict the existence of turbulent magnetic fields in the ICM (Vazza et al. 2017; Donnert et al. 2018). However, some of these studies have predicted a

radial decrease of β_{pl} across the cluster volume, e.g., due to an increase in the nonthermal pressure support triggered by mergers. We refer to Matthews et al. (2022) and Marsh et al. (2022) for discussions of the impact of magnetic field modeling assumptions on photon–ALP bounds inferred from cluster/AGN searches under cell-based and GRF approaches. Specifically, Matthews et al. (2022) found that a radially dependent β_{pl} can systematically shift bounds on $\log(g_{a\gamma}/\text{GeV}^{-1})$ from cell-based and GRF studies by 0.3 dex (depending on the exact radial profile of β_{pl}). The same study also finds that the limits are fairly insensitive to whether cell-based or GRF models are used.

The use of the Fourier formalism (Marsh et al. 2022), accurate to leading order in the coupling, may provide a fairly computationally inexpensive route to computing photon–ALP conversion. This formalism is valid for sufficiently low values of $g_{a\gamma}$ such that the ALP-induced distortions are of the order $\lesssim 5\%–10\%$, which is the appropriate regime for our simulated observations. This formalism, embedded in ALPRO (Matthews 2022), can also be applied to the massive ALP regime (see Section III of Marsh et al. 2022). Furthermore, as noted by Marsh et al. (2022), if an ALP signal is present, the residuals of the data may be able to be transformed directly to obtain the autocorrelation function of the line-of-sight magnetic field. Such an exercise is challenging, requiring very good quality data, but the high spectral resolution and large effective area of Athena offer exciting prospects if this formalism were to be applied in future.

Ideally, a 3D description of the cluster field, constrained by observational RM and pressure profile data, would be adopted to unveil its true underlying structure (for recent studies, see Schallmoser et al. 2021; Carena et al. 2022). The turbulent nature of the field can create departures from predictions using a GRF, as shown by Carena et al. (2022). They compute the conversion probabilities using a high-resolution 3D MHD simulation, finding that non-Gaussian, local spikes in the MHD magnetic field produce “heavy-tailed” distributions of the conversion probability. This work therefore suggests that, for certain sightlines, stronger ALP signals could be observed.

Finally, we note that if an effectively massless ($m_a \lesssim 10^{-12} \text{ eV}$) ALP signal with sufficiently high $g_{a\gamma}$ (compared to the Poisson noise) were to be discovered by Athena/X-IFU, the extent to which the former signal would be recovered would depend on the underlying field structure. This is shown by Figures 8 and 9; the results are encouraging, in the sense that the ALP coupling is recovered quite reliably even when the exact magnetic field structure is not known. However, in this effectively massless ALP regime, no information about the true ALP mass would be retained, given that all ALPs of masses $\log(m_a/\text{eV}) \in [-30.0, -12.0]$ induce near-identical spectral distortions. In addition, if the true magnetic field model intersected by the line of sight included the non-Gaussian structure described by Carena et al. (2022), then using a cell-based or GRF model for the inference of $g_{a\gamma}$ would result in a bias, leading to a higher $g_{a\gamma}$ being estimated than the true photon–ALP coupling value.

Overall, the above considerations regarding magnetic field modeling suggest a fairly encouraging outlook for the next generation of X-ray ALP searches, but they do highlight the importance of high-frequency-resolution RM studies of clusters, combined with improved theoretical understanding of the ICM magnetic field structure from the cluster core to the virial radius.

9.5. Other Systematics

9.5.1. Initially Unpolarized Photon Beam

In our work, we have solved the photon–ALP propagation problem for an initially unpolarized photon beam propagating through the host cluster. One would expect such an assumption to yield conservative photon–ALP mixing probabilities compared to an initially polarized photon beam. We refer the reader to Day & Krippendorf (2018), Dessert et al. (2022), Galanti et al. (2023), and Galanti (2022) for explorations on how one can use source polarization and ALP-induced polarization to probe ALPs.

9.5.2. Spectral Model for NGC 1275

Other than for its normalization, all model parameters of the `bvvapec` model component applied to describe the simulated Athena/X-IFU spectrum of NGC 1275 were frozen (see Table 3). Our inferred posterior on ALPs remains insensitive to this choice. Furthermore, one would expect our results to remain insensitive to whether partial covering is accounted for in the fits to our Athena and AXIS simulated spectra of NGC 1275 (see Figure 6 and the discussion in Matthews et al. 2022).

Finally, we note that throughout our work, we ignored the effects of relativistic reflection from slowly moving material surrounding the immediate vicinity of NGC 1275. The latter has frequently been modeled to explain the 6.4 keV reflection signature frequently detected in high-resolution X-ray spectra of NGC 1275 (e.g., Hitomi Collaboration et al. 2018b; Reynolds et al. 2021). We would expect this to not affect the photon–ALP bounds inferred in our work.

10. Conclusions

We have conducted an analysis that revisits Athena’s projected bounds on very-light ALPs, following from Conlon et al. (2017a), under the effects of detector calibration. We have also explored the potential of AXIS for BSM searches. In both cases, we employ 200 ks simulated observations of NGC 1275 as the subject of our study; for concreteness, we consider detector setups broadly consistent with those proposed in Barret et al. (2020) and Mushotzky & AXIS Team (2019), respectively. Our main findings are:

1. Bright cluster-hosted AGNs located within magnetically rich galaxy clusters are excellent probes of massless ALPs, $\log(m_a/\text{eV}) < -12.0$. With its unprecedented spectral energy resolution (2.5 eV for its X-IFU), Athena will be able to exclude all photon–ALP couplings $g_{a\gamma} \geq 2.0 \times 10^{-13} \text{ GeV}^{-1}$ at 95% CL. These bounds have been inferred under a conservative assessment of detector calibration.
2. Athena/X-IFU will be able to improve on the current limits on $g_{a\gamma}$ by an order of magnitude (excluding $g_{a\gamma} \geq 6.3 \times 10^{-14} \text{ GeV}^{-1}$ at 95% CL) provided accurate knowledge of its detector-induced residuals. For a 1 Ms exposure of NGC 1275, the upper bound on $\log(g_{a\gamma}/\text{GeV}^{-1})$ would further tighten by 0.2 dex in both the optimal and conservative calibration scenarios.
3. Machine learning will provide an automated route to disentangling the effects of detector calibration and photon–ALP-induced residuals in next-generation

cluster/ALP searches. This will be particularly relevant for the AXIS CCD mission.

4. Despite having a fundamentally different design to Athena/X-IFU, AXIS’s more moderate spectral resolution will be circumvented by its unprecedented angular resolution, being able to offer high-quality AGN spectra free from cluster emission. In the optimal calibration scenario, AXIS would exceed the current best bounds on light ALPs from a 200 ks on-axis observation of NGC 1275 by a factor of 3. Excitingly, for a 1 Ms observation of NGC 1275, AXIS will exclude all couplings $g_{a\gamma} > 4 \times 10^{-14} \text{ GeV}^{-1}$ in the optimal calibration scenario, exceeding the current best bounds on massless ALPs by more than an order of magnitude.
5. Next-generation X-ray observatories may play a key role in constraining axions arising in string theories which predict low values of $g_{a\gamma}$ such as type-IIB string theories (Halverson et al. 2019; Demirtas et al. 2021). Some of these theories will additionally make predictions for a favored number of ultralight axion fields. The parameter space of these theories may additionally be probed by complementary astrophysical observations, e.g., with population studies of the spin–mass distribution of black holes over cosmic time with next-generation gravitational wave detectors such as LISA (Mehta et al. 2021).

Overall, our work shows the potential of next-generation X-ray observatories to constraining very-light ALPs, even under fundamentally different designs. Our analysis suggests that, in the future, these X-ray missions may complement future laboratory-based ALP DM searches (e.g., in birefringent cavities such as ADBC) down to very-light masses.

Acknowledgments

We thank the anonymous referee for a constructive and helpful report. We also thank Didier Barret for providing access to the library of Athena/X-IFU response files from Barret & Cappi (2019) as well as for providing useful comments on this paper. We thank Francesca Chadha-Day, Andy Fabian, and Stefan Heimersheim for helpful discussions. J.S.R. acknowledges the support from the Science and Technology Facilities Council (STFC) under grant ST/V50659X/1 (project reference 2442592). C.S.R. thanks the STFC for support under the Consolidated Grant ST/S000623/1, as well as the European Research Council (ERC) for support under the European Union Horizon 2020 research and innovation program (grant 834203). M.L.P. is supported by European Space Agency (ESA) Research Fellowships. J.H.M. acknowledges a Herchel Smith Fellowship while at Cambridge and support from the Royal Society at Oxford. M.C.D.M. is supported by the European Research Council under Grant No. 742104 and by the Swedish Research Council (VR) under grants 2018-03641 and 2019-02337. This work was performed using resources provided by the Cambridge Service for Data Driven Discovery (CSD3) operated by the University of Cambridge Research Computing Service (www.csd3.cam.ac.uk), provided by Dell EMC and Intel using Tier-2 funding from the Engineering and Physical Sciences Research Council (capital grant EP/P020259/1), and DiRAC funding from the Science and Technology Facilities Council (www.dirac.ac.uk). This work was performed using resources provided by the Cambridge Service for Data Driven Discovery (CSD3) operated by the

University of Cambridge Research Computing Service (www.csd3.cam.ac.uk), provided by Dell EMC and Intel using Tier-2 funding from the Engineering and Physical Sciences Research Council (capital grant EP/P020259/1), and DiRAC funding from the Science and Technology Facilities Council (www.dirac.ac.uk). We gratefully acknowledge the use of the following software packages: astropy (Astropy Collaboration et al. 2013, 2018), matplotlib (Hunter 2007), pandas (McKinney 2010; pandas development team 2020), OpenMPI (Gabriel et al. 2004), and XSPEC and PYXSPEC (Arnaud 1996). As introduced in Section 4, we have employed the public-source code ALPRO v1.0 (Matthews 2022).

Data Availability

The simulated X-ray spectra of NGC 1275, which this study is based on, were generated from the optimal response matrix and ARFs of the Athena (Spiga et al. 2017) and AXIS⁷ telescopes. The nonoptimal ARFs over which we marginalize (Section 6.2) were kindly provided by D. Barret. Moreover, the raw X-ray data employed to estimate the normalization factor of the single-temperature cluster emission component (Section 3.2) in our simulated Athena/X-IFU spectrum of NGC 1275 are available from the public data archives of the Chandra Science Center. Our limits data will be made publicly available in the ALPRO⁸ repository. The reduced data products used in this work may be shared on reasonable request to the authors.

Appendix A Theory of Photon–ALP Mixing

At a fundamental level, the interaction between ALPs and electromagnetism is specified by Equation (1). Here, we present the relevant equations derived from Equation (1) for a simplified external field geometry. This is for illustration purposes and context only. We refer to Sections 2 and 3 of Matthews et al. (2022) for a description of the calculations used to generate the photon–ALP mixing grid employed throughout our analysis.

We proceed to consider a scenario where the electromagnetic wave is traversing an ionized and magnetized medium such that its self-induced magnetic field can be ignored relative to the latter. We refer to Galanti & Roncadelli (2022) for an overview of the relevant equations inferred when the self-induced electric and magnetic fields are not negligible.

We consider an initially unpolarized beam propagating through an external homogeneous field perpendicular to the direction of propagation, \mathbf{B}_0 (of strength B_0). Most generally, however, the magnetic field geometry of a given external medium will be complex, for which the photon–ALP mixing problem (derived from Equation (1)) must be solved numerically.

In the field geometry considered, a wave-like equation can be derived for $\mathcal{L}_{a\gamma}$ which, at energies $E \gg m_a$, for $m_a \leq 10^{-10}$ eV, simplifies to a Schrödinger-type equation that can be solved through the eigenvalue problem. The solution to such problem can be quantified via a nonunity “survival” probability for the photon beam as it interconverts into ALPs of mass m_a and coupling $g_{a\gamma}$ after having traveled distance L along

the magnetized medium. Following Section 4.1 of Marsh et al. (2017), $P_{\gamma \rightarrow \gamma}$ is given by:

$$P_{\gamma \rightarrow \gamma} = 1 - \frac{1}{4} \frac{1}{1 + \theta^{-2}} \sin^2(L \Delta_{\text{eff}} \sqrt{1 + \theta^2}), \quad (\text{A1})$$

where θ and Δ_{eff} are defined via:

$$\theta = 4\pi \left(\frac{B_0 g_{a\gamma} E}{m_a^2 - \omega_{\text{pl}}^2} \right), \quad (\text{A2})$$

and

$$\Delta_{\text{eff}} = \frac{\omega_{\text{pl}}^2 - m_a^2}{8\pi E}, \quad (\text{A3})$$

where ω_{pl} is the plasma frequency of the ionized medium and $m_{\text{eff}}^2 = m_a^2 - \omega_{\text{pl}}^2$ is the effective photon mass. For effectively massless ALPs ($m_a \ll \omega_{\text{pl}}$, i.e., $m_{\text{eff}}^2 \rightarrow -\omega_{\text{pl}}^2$), we now proceed to consider the regime where Equations (A2) and (A3) satisfy $\theta \gg 1$ and $\Delta_{\text{eff}} \ll 1$, respectively. In a suitable environment (i.e., that provided by rich cool-core clusters and for a suitable location of ALP parameter space), the survival function will be:

$$P_{\gamma \rightarrow \gamma} \sim 1 - 10^{-3} \left(\frac{B_0}{1 \mu\text{G}} \right)^2 \left(\frac{g_{a\gamma}}{10^{-10} \text{ GeV}^{-1}} \right)^2 \times \left(\frac{L}{1 \text{ kpc}} \right)^2 g(E), \quad (\text{A4})$$

where we have adopted a similar notation to Conlon et al. (2017a) and where $g(E) \in [0, 1]$ is an oscillatory function with energy of order unity. We highlight that the product $(B_0 L)$ in Equation (A4) can attain large values in rich cool-core clusters compared to other astrophysical systems, which underlies why rich clusters should be efficient photon–ALP interconverters.

Appendix B Effects of a Nonoptimal Calibration on Athena/X-IFU Revisited

The left panel of Figure 5 shows the residuals resulting from fitting the simulated Athena/X-IFU spectrum of NGC 1275 with an astrophysics-only spectral model (see Table 3), i.e., without ALPs. The input spectrum is then interpreted under an “optimal” calibration (simulating and fitting with the same ARF) and an example “nonoptimal” calibration (simulating and fitting with different ARFs) scenario (left and central panels, respectively). Clearly, the miscalibrated spectrum seems to retain most of the curvature present in the residuals of the optimal case with the exception of acquiring additional curvature at energies < 2 keV. This is likely due to a miscalibration of instrumental effects at such energies. We highlight that an even more conservative approach to assessing the effects of detector calibration would involve convolving the miscalibrated responses for the mirror and microcalorimeters, but is, however, beyond the scope of this paper (see the discussion in Barret & Cappi 2019). The right panel of Figure 5 shows the residuals resulting from fitting a multiplicative ALP model in our photon–ALP mixing grid (Section 4) to the simulated Athena/X-IFU spectrum of NGC 1275 under the optimal calibration scenario. For the specific turbulent field realization and value of $g_{a\gamma}$ chosen, the ALP-induced features result in well-defined peaks in the residuals whose amplitudes

⁷ See <https://axis.astro.umd.edu/>

⁸ See <https://github.com/jhmatthews/alpro/tree/v1.1>

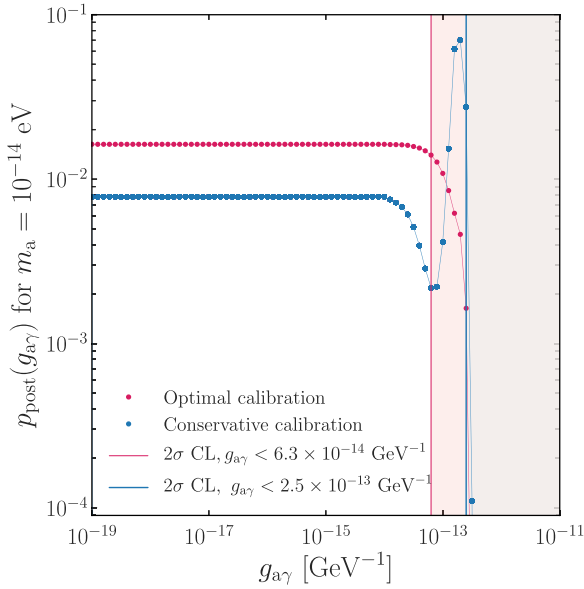


Figure 10. Normalized posterior distribution on $g_{a\gamma}$ for ALPs of $m_a = 10^{-14}$ eV (see Equation (C1)), inferred when fitting our simulated Athena/X-IFU spectrum of NGC 1275 in an “optimal” and a representative “conservative” detector response scenario (Section 6.1 for a discussion). The posterior is inferred by marginalizing over 216 magnetic field realizations. The shaded regions delimit regions of parameter space that are excluded at 95% (i. e. 2σ) confidence, respectively.

increase with energy. Interestingly, the C -statistic of such ALP fit will still more accurately describe the simulated spectrum of NGC 1275 compared to a miscalibrated non-ALP-containing model. This is illustrated by comparing the C -statistics of the central and right panels to that in the left panel of Figure 5.

For further comparison, we fit the miscalibrated Athena/X-IFU spectrum (Section 3) across our grid of photon-ALP mixing models $\{\text{ALP}(\dot{i}, g_{a\gamma})\}$, following the procedure described in Section 5. The marginalized posterior (over 216 field models) is compared to that for the optimal calibration case in Figure 10. Importantly, we see that the former is now dominated by large values of the photon-ALP coupling $g_{a\gamma}$ (i.e., above the 95% exclusion level in the optimal calibration case) suited to describe the energy-dependent features induced by instrumental detector miscalibration. The analysis of such a miscalibrated spectrum would translate into a weakening of the 95% CL on $g_{a\gamma}$ by a factor of 4, compared to the optimal calibration scenario. In Section 6.2, we thereby proceed to assess the potential effects of detector miscalibration on the resulting posteriors on ALPs by considering a library of conservative ARFs.

Appendix C Marginalizing Over Response Files

In Section 5, we present the optimal bounds on ALPs from Athena/X-IFU and AXIS by using the Bayesian framework presented in Section 4 of Sisk-Reynés et al. (2021). We hereby proceed to compute Athena’s bounds on massless ALPs by marginalizing over a set of 216 ARFs and 216 realizations of the turbulent field within Perseus. Our results are presented in Section 6.2.

Equipped with the library of photon-ALP mixing curves introduced in Section 4, we proceed as follows. We iteratively fit the optimal Athena/X-IFU spectrum of NGC 1275 (Figure 2) with the spectral model $\text{tbabs}^*(\text{po} +$

$\text{bvvapec})^* \text{ALP}(m_a, g_{a\gamma}, \dot{i})$ across a library of conservative detector responses, $\{\mathbb{f}\}$. Here, $\text{ALP}(m_a, g_{a\gamma}, \dot{i})$ quantifies the energy-dependent survival function of quanta mixing with ALPs of parameters $(m_a, g_{a\gamma})$ for a given field realization \dot{i} . We refer to the goodness-of-fit statistic (C -statistic) of each fit as $C(m_a, g_{a\gamma}, \dot{i}, \mathbb{f})$. The free parameters of each fit are outlined in Table 3. Each multiplicative table ALP model has a redshift parameter attributed to it, frozen to that of the galaxy ($z = 0.01278$; Hitomi Collaboration et al. 2018a).

We now introduce the Bayesian framework that permits converting $C(m_a, g_{a\gamma}, \text{mathiti}, \mathbb{f})$ into a normalized posterior $p_{\text{post}}(m_a, g_{a\gamma})$. The latter can then be used to find CLs on ALP parameter space.

Ultimately, we seek to find a posterior distribution on ALP parameters $p_{\text{post}}(m_a, g_{a\gamma})$ normalized according to the condition:

$$\sum_{\log(m_a/\text{eV})} \sum_{\log(g_{a\gamma}/\text{GeV}^{-1})} p_{\text{post}}(m_a, g_{a\gamma}) = 1, \quad (\text{C1})$$

where we assume flat priors on the ALP parameters. The summations involved comprise the parameter space $\log(m_a/\text{eV}) \in [-30.0, -12.0] \cup \log(g_{a\gamma}/\text{GeV}^{-1}) \in [-19.0, -11.0]$, where the choice of lower limits in m_a and $g_{a\gamma}$ is justified in Section 5.

Operationally, $p_{\text{post}}(m_a, g_{a\gamma})$ is found by marginalizing over all magnetic field models $\dot{i} \in [1, 216]$ and detector responses $\mathbb{f} \in [0, 215]$, where $\mathbb{f}0$ is the optimal Athena/X-IFU ARF. We broadly follow the description in Appendix 2 of Marsh et al. (2017), where we marginalize over the added DOF encapsulated by the set of ARFs (\mathbb{f}). For a given ARF \mathbb{f} , the unnormalised posterior on a given set of ALP parameters $(m_a, g_{a\gamma})$ is given by:

$$p_{\text{post}}(m_a, g_{a\gamma}, \mathbb{f}) \propto \sum_{\dot{i}=1}^{216} \exp \left\{ \frac{-(C(m_a, g_{a\gamma}, \dot{i}, \mathbb{f}) - C_{\text{fid}}(\mathbb{f}))}{2} \right\}, \quad (\text{C2})$$

where the summation over \dot{i} marginalizes over magnetic field configurations, on which we assume flat priors (as previously done in Sisk-Reynés et al. 2021; Matthews et al. 2022). Here, $C_{\text{fid}}(\mathbb{f})$ is the “fiducial” best-fit C -statistic of the simulated X-IFU spectrum read on the \mathbb{f} th ARF when fitted with the $\text{tbabs}^*(\text{po}+\text{bvvapec})$ model (see Table 3).






Clearly, for a given (\mathbb{f} th) ARF, the posterior distribution on ALPs will be dominated by the ALP models suited to describing the underlying residuals in the spectrum of NGC 1275 once its calibration-induced features have been corrected. Similar to the case of the magnetic field realization, the dependence on detector response can be eliminated through:

$$p_{\text{post}}(m_a, g_{a\gamma}) \equiv \mathcal{L}(m_a, g_{a\gamma}) = \sum_{\mathbb{f}=0}^{215} p_{\text{post}}(m_a, g_{a\gamma}, \mathbb{f}), \quad (\text{C3})$$

where, again, we assume flat priors on the ARF nuisance parameter. Operationally, we find the constant of proportionality in Equation (C2) through the normalization condition of $p_{\text{post}}(m_a, g_{a\gamma})$, i.e., Equation (C1). In the equation above, we have explicitly stated the equivalence of the normalized posterior and the marginalized likelihood $\mathcal{L}(m_a, g_{a\gamma})$ over ARFs and field configurations.

Finally, we note the ALP constraints inferred in our work are insensitive to the choice of binning scheme chosen throughout the data processing discussed in Sections 5 and 6.

ORCID iDs

Júlia Sisk-Reynés  <https://orcid.org/0000-0003-3814-6796>
 Christopher S. Reynolds  <https://orcid.org/0000-0002-1510-4860>
 Michael L. Parker  <https://orcid.org/0000-0002-8466-7317>
 James H. Matthews  <https://orcid.org/0000-0002-3493-7737>
 M. C. David Marsh  <https://orcid.org/0000-0001-7271-4115>

References

- Abadi, M., Agarwal, A., Barham, P., et al. 2015, TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, <https://www.tensorflow.org/>
- Abbott, L. F., & Sikivie, P. 1983, *PhLB*, **120**, 133
- Aghanim, N., Akrami, Y., Ashdown, M., et al. 2020, *A&A*, **641**, A6
- Anastassopoulos, V., Aune, S., Barth, K., et al. 2017, *NatPh*, **13**, 584
- Armengaud, E., Attie, D., Basso, S., et al. 2019, *JCAP*, **06**, 047
- Arnaud, K. A. 1996, in ASP Conf. Ser. 101, *Astronomical Data Analysis Software and Systems V*, ed. G. H. Jacoby & J. Barnes (San Francisco, CA: ASP), 17
- Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, *AJ*, **156**, 123
- Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *A&A*, **558**, A33
- Barret, D., Albuys, V., Herder, J.-W. D., et al. 2023, *ExA*, **55**, 373
- Barret, D., & Cappi, M. 2019, *A&A*, **628**, A5
- Barret, D., Decourchelle, A., Fabian, A., et al. 2020, *AN*, **341**, 224
- Barton, G., & White, E. D. 1969, *PhRv*, **184**, 1660
- Berg, M., Conlon, J. P., Day, F., et al. 2017, *ApJ*, **847**, 101
- Böhringer, H., Chon, G., & Kronberg, P. P. 2016, *A&A*, **596**, A22
- Canizares, C., Davis, J., Dewey, D., et al. 2005, *PASP*, **117**, 1144
- Carenza, P., Sharma, R., Marsh, M. C. D., Brandenburg, A., & Müller, E. 2022, arXiv:2208.04333
- Carenza, P., Straniero, O., Döbrich, B., et al. 2020, *PhLB*, **809**, 135709
- Carroll, S. M. 1998, *PhRvL*, **81**, 3067
- Cash, W. 1979, *ApJ*, **228**, 939
- Chollet, F. 2015, Keras Development Team <https://keras.io>
- Churazov, E., Forman, W., Jones, C., & Böhringer, H. 2003, *ApJ*, **590**, 225
- Cicoli, M., Goodsell, M. D., & Ringwald, A. 2012, *JHEP*, **2012**, 146
- Conlon, J. P. 2006, *JHEP*, **2006**, 078
- Conlon, J. P., Day, F., Jennings, N., Krippendorf, S., & Muia, F. 2017a, *MNRAS*, **473**, 4932
- Conlon, J. P., Day, F., Jennings, N., Krippendorf, S., & Rummel, M. 2017b, *JCAP*, **07**, 005
- Conlon, J. P., & Rummel, M. 2019, *MNRAS*, **484**, 3573
- Crnogorčević, M., Caputo, R., Meyer, M., Omodei, N., & Gustafsson, M. 2021, *PhRvD*, **104**, 103001
- Cucchetti, E., Eckart, M. E., Peille, P., et al. 2018, *Proc. SPIE*, **10699**, 106994M
- Day, F., & Krippendorf, S. 2018, *Galax*, **6**, 45
- Day, F., & Krippendorf, S. 2020, *JCAP*, **2020**, 046
- Demirtas, M., Gendler, N., Long, C., McAllister, L., & Moritz, J. 2021, arXiv:2112.04503
- Dessert, C., Dunsky, D., & Safdi, B. R. 2022, *PhRvD*, **105**, 103034
- Dine, M., & Fischler, W. 1983, *PhLB*, **120**, 137
- Donnert, J., Vazza, F., Brüggem, M., & ZuHone, J. 2018, *SSRv*, **214**, 122
- Drake, J. J., Ratzlaff, P., Kashyap, V., et al. 2006, *Proc. SPIE*, **6270**, 62701I
- Fabian, A. C., Sanders, J. S., Taylor, G. B., et al. 2006, *MNRAS*, **366**, 417
- Gabriel, E., Fagg, G. E., Bosilca, G., et al. 2004, in Proc. 11th European PVM/MPI Users' Group Meeting, 3241, ed. D. Kranzlmüller, P. Kacsuk, & J. Dongarra (Berlin: Springer), 97
- Galanti, G. 2022, *PhRvD*, **107**, 043006
- Galanti, G., & Roncadelli, M. 2022, *Univ*, **8**, 253
- Galanti, G., Roncadelli, M., Tavecchio, F., & Costa, E. 2023, *PhRvD*, **107**, 103007
- Gaskin, J. A., Swartz, D., Vikhlinin, A. A., et al. 2019, *JATIS*, **5**, 021001
- Govoni, F. 2012, IAU Proc. Vol. 10, Highlights H16: Highlights of Astronomy (Cambridge: Cambridge Univ. Press), 404
- Green, M. B., Schwarz, J. H., & Witten, E. 1988, *Superstring Theory* (Cambridge: Cambridge Univ. Press)
- Guainazzi, M., Willingale, R., Brenneman, L. W., et al. 2022, *JATIS*, **8**, 044002
- Hitomi Collaboration, Aharonian, F., Akamatsu, H., et al. 2018a, *PASJ*, **70**, 9
- Hitomi Collaboration, Aharonian, F., Akamatsu, H., et al. 2016, *Natur*, **535**, 117
- Hitomi Collaboration, Aharonian, F., Akamatsu, H., et al. 2018b, *PASJ*, **70**, 13
- Hitomi Collaboration, Aharonian, F., Akamatsu, H., et al. 2018c, *PASJ*, **70**, 9
- Hitomi Collaboration, Akamatsu, H., Akimoto, F., et al. 2018, *PASJ*, **70**, 12
- Halverson, J., Long, C., Nelson, B., & Salinas, G. 2019, *PhRvD*, **100**, 106010
- Hong, J., Romaine, S., Kenter, A., et al. 2019, *Proc. SPIE*, **11118**, 1111810
- Hunter, J. D. 2007, *CSE*, **9**, 90
- Irastorza, I. G., & Redondo, J. 2018, *PrPNP*, **102**, 89
- Kaastra, J. S. 2017, *A&A*, **605**, A51
- Kalberla, P. M. W., Burton, W. B., Hartmann, D., et al. 2005, *A&A*, **440**, 775
- Kingma, D. P., & Ba, J. 2014, arXiv:1412.6980
- Liu, H., Elwood, B. D., Evans, M., & Thaler, J. 2019, *PhRvD*, **100**, 023548
- Lodders, K., & Palme, H. 2009, *M&PSA*, **72**, 5154
- Marsh, M. C. D., Matthews, J. H., Reynolds, C., & Carenza, P. 2022, *PhRvD*, **105**, 016013
- Marsh, M. D., Russell, H. R., Fabian, A. C., et al. 2017, *JCAP*, **2017**, 036
- Matthews, J. 2022, jhmatthews/alpro: Alpro v1.1, Zenodo, doi:10.5281/zenodo.6137185
- Matthews, J. H., Reynolds, C. S., Marsh, M. C. D., Sisk-Reynés, J., & Rodman, P. E. 2022, *ApJ*, **930**, 90
- McKinney, W. 2010, in Proc. 9th Python in Science Conf., ed. S. van der Walt & J. Millman (Austin, TX: SciPy), 56
- Mehta, V. M., Demirtas, M., Long, C., et al. 2021, *JCAP*, **2021**, 033
- Meyer, M., Davies, J., & Kuhlmann, J. 2022, *ICRC (Berlin)*, **37**, 557
- Meyer, M., & Petrushevska, T. 2020, *PhRvL*, **124**, 231101
- Mushotzky, R. & AXIS Team 2019, in The Space Astrophysics Landscape for the 2020s and Beyond, **2135**, 5025
- Nagai, H., Onishi, K., Kawakatu, N., et al. 2019, *ApJ*, **883**, 193
- Nair, V., & Hinton, G. E. 2010, in Proc. 27th Int. Conf. on Machine Learning (Madison, WI: Omnipress), 807
- Nandra, K., Barret, D., Barcons, X., et al. 2013, arXiv:1306.2307
- National Academies of Sciences, Engineering, and Medicine 2021, Pathways to Discovery in Astronomy and Astrophysics for the 2020s (Washington, DC: The National Academies Press)
- Pandas Development Team, Reback, J., jbrockmendl, et al. 2020, pandas-dev/pandas v1.4.2, Zenodo, doi:10.5281/zenodo.3509134
- Parker, M. L., Lieu, M., & Matzeu, G. A. 2022, *MNRAS*, **514**, 4061
- Payez, A., Evoli, C., Fischer, T., et al. 2015, *JCAP*, **2015**, 006
- Pearson, K. 1901, *Lond. Edinb. Dublin Philos. Mag. J. Sci.*, **2**, 559
- Peccei, R. D., & Quinn, H. R. 1977, *PhRvL*, **38**, 1440
- Prechelt, L. 1998, *Neural Networks: Tricks of the Trade* (Berlin: Springer), 55
- Preskill, J., Wise, M. B., & Wilczek, F. 1983, *PhLB*, **120**, 127
- Raffelt, G., & Stodolsky, L. 1988, *PhRvD*, **37**, 1237
- Raffelt, G. G. 1996, *Stars as laboratories for fundamental physics: The astrophysics of neutrinos, axions, and other weakly interacting particles* (Chicago: Univ. Chicago Press)
- Ray, P. S., Arzoumanian, Z., Ballantyne, D., et al. 2019, arXiv:1903.03035
- Reynolds, C. S., Marsh, M. C. D., Russell, H. R., et al. 2020, *ApJ*, **890**, 59
- Reynolds, C. S., Smith, R. N., Fabian, A. C., et al. 2021, *MNRAS*, **507**, 5613
- Schallmoser, S., Krippendorf, S., Chadha-Day, F., & Weller, J. 2021, *MNRAS*, **514**, 329
- Simionescu, A., Werner, N., Urban, O., et al. 2012, *ApJ*, **757**, 182
- Sisk-Reynés, J., Matthews, J. H., Reynolds, C. S., et al. 2021, *MNRAS*, **510**, 1264
- Sisk-Reynés, J. M., Reynolds, C. S., & Matthews, J. H. 2023, arXiv:2304.08513
- Sisson, S. A., Fan, Y., & Beaumont, M. A. 2018, arXiv:1802.09720
- Spiga, D., Ferreira, D. D. M., Shortt, B., et al. 2017, *Proc. SPIE*, **10399**, 103990H
- Svrcek, P., & Witten, E. 2006, *JHEP*, **2006**, 051
- Taylor, G. B., Gugliucci, N. E., Fabian, A. C., et al. 2006, *MNRAS*, **368**, 1500
- Vazza, V., Murgia, M., Govoni, F., et al. 2012, *A&A*, **540**, A38
- Vazza, F., Brunetti, G., Brüggem, M., & Bonafede, A. 2017, *MNRAS*, **474**, 1672
- Weinberg, S. 1978, *PhRvL*, **40**, 223
- Wilczek, F. 1978, *PhRvL*, **40**, 279
- Wilms, J., Allen, A., & McCray, R. 2000, *ApJ*, **542**, 914
- Wouters, D., & Brun, P. 2013, *ApJ*, **772**, 44
- Zhuravleva, I., Churazov, E., Schekochihin, A. A., et al. 2014, *Natur*, **515**, 85