

On the evidence for axionlike particles from active galactic nuclei

Guido Walter Pettinari* and Robert Crittenden

*Institute of Cosmology and Gravitation, University of Portsmouth, Dennis Sciama Building,
Burnaby Road, Portsmouth PO1 3FX, United Kingdom*

(Received 2 July 2010; published 1 October 2010)

Burrage, Davis, and Shaw [1] recently suggested exploiting the correlations between high and low energy luminosities of astrophysical objects to probe possible mixing between photons and axionlike particles (ALP) in magnetic field regions. They also presented evidence for the existence of ALP's by analyzing the optical/UV and x-ray monochromatic luminosities of active galactic nuclei. We extend their work by using the monochromatic luminosities of 320 unobscured active galactic nuclei from the Sloan Digital Sky Survey/Xmm-Newton Quasar Survey [2] which allows the exploration of 18 different combinations of optical/UV and x-ray monochromatic luminosities. However, we do not find compelling evidence for the existence of ALPs. Moreover, it appears that the signal reported by Burrage *et al.* is more likely due to x-ray absorption rather than to photon-ALP oscillation.

DOI: 10.1103/PhysRevD.82.083502

PACS numbers: 98.80.Es, 14.80.Va, 95.85.Nv, 98.54.Cm

I. INTRODUCTION

Very light scalar fields could have had a significant impact on cosmology, potentially acting as dark matter (e.g., axions or axionlike particles) or explaining the recent accelerated expansion (e.g., a quintessence field). In both cases, their roles are primarily gravitational, either to provide additional gravitational clustering on galactic or cluster scales, or to drive the overall expansion of the Universe. However, this does not rule out nongravitational interactions, which should also exist if not explicitly forbidden by some symmetry [3].

The strengths of such interactions are well constrained for axionlike particles (ALPs). These interactions could lead to more efficient stellar cooling, and the limits from solar axions observed on Earth constrain the couplings to be $g \lesssim 1 \times 10^{-10} \text{ GeV}^{-1}$ [4]. Additionally, ALPs can be emitted from the core of supernovae at a significant rate [5]; the lack of evidence for such outburst from SN 1987A yields $g \lesssim 1 \times 10^{-11} \text{ GeV}^{-1}$ for a very light ALP ($m \lesssim 10^{-9} \text{ eV}$) [4]. Similarly in the quintessence case, such interactions between photons with a slowly rolling field would lead to time variations in the fine structure constant which could be observed in stellar lines [3].

One way of avoiding these constraints is the so-called chameleon model [6], where nonminimal couplings to gravity lead to the effective mass or coupling of the scalar field being dependent on the local mass density. In this way, many of the constraints on the mass of the scalar fields can be satisfied, while still allowing for a significant interaction strength in regions where the Universe is less dense [7,8].

It is worth trying to constrain couplings of the axionlike particles in low density regions where they are not masked by chameleon effects. Typical interactions couple the axionlike particle to two photons; this can lead to photons

decaying into axions as they pass through magnetic fields. If the probability is small, this leads to “tired light” scenarios, where objects at large distances are progressively dimmer than expected. These ideas have been investigated in the context of Type Ia supernovae, where the effect is similar to that of grey dust [9–12] but still cannot explain their apparent dimming without introducing cosmic acceleration (see Ref. [13] and references therein).

In the opposite strong-mixing regime, the photons could convert to axions rarely, but with high probability, as they pass through the magnetic fields around a galaxy or cluster. On average, the mixing will result in one-third of the photons being converted into ALP's, but the exact amount will depend on the magnetic field orientations along the individual photon paths. Given the relatively short coherence lengths associated with such magnetic fields, the mixing is expected to vary from source to source.

The average suppression of such strong mixing is difficult to detect without well calibrated sources at cosmological distances. However Burrage, Davis, and Shaw [1] (hereafter, BDS) recently proposed using the distribution of the fluxes of cosmological sources as a means of constraining the mixing. Since the mixing is dependent of the photon energy, the low energy fluxes are used to calibrate the brightness of the sources. After considering a number of potential sources, BDS analyzed the distribution of x-ray fluxes of active galactic nuclei (AGNs), normalized by their optical fluxes and found significant evidence of such mixing, up to the 5σ level in their most recent analysis [14].

Such strong evidence for ALP mixing is tantalizing, and the purpose of this paper is to re-examine and extend their analysis with larger data sets to see how robust the signal is. In Sec. II, we review the dynamics of the mixing model and the expected signal. In Sec. III, we examine a number of possible ways to evaluate the statistical significance of such a signal, and in Sec. IV we apply these methods, re-examining the original claim of BDS and then extending

*Guido.Pettinari@port.ac.uk

the analysis to the larger data set of Young *et al.* [2], before concluding in Sec. V.

II. MODEL ASSUMPTIONS

A. Scatter from mixing to axionlike particles

For our purposes, an ALP is a light but massive scalar or pseudoscalar field that couples with the kinetic term of the Lagrangian of the photons. Depending on whether the ALP is a scalar or a pseudoscalar, this interaction term has one of the following forms:

$$\mathcal{L}_{\text{int}}^S = \frac{\phi}{M}(\mathbf{B}^2 - \mathbf{E}^2), \quad \mathcal{L}_{\text{int}}^{PS} = \frac{\phi}{M}\mathbf{E} \cdot \mathbf{B},$$

where ϕ is the axion field and $M \equiv 1/g$ sets the scale of the strength of the photon-ALP coupling. As a result, when a photon of energy E travels through a magnetic domain of length L and intensity B , there is a nonzero probability that it oscillates into an ALP [15–17]:

$$P_{\gamma \leftrightarrow \phi} = \frac{1}{1 + E_{\text{crit}}^2/E^2} \sin^2\left(\frac{\mu L}{2} \sqrt{1 + \frac{E_{\text{crit}}^2}{E^2}}\right), \quad (1)$$

where $E_{\text{crit}} = m_\phi^2/2\mu$ is the characteristic energy scale, $\mu \equiv B/M$, and m_ϕ is the ALP mass.

We are most interested in ALP-photon mixing in an astrophysical context where the propagating medium is an electron plasma. Therefore, we have to substitute an effective mass for the mass of the ALP, $m_\phi \rightarrow m_{\text{eff}}$, where,

$$m_{\text{eff}}^2 \equiv |m_\phi^2 - \omega_{\text{pl}}^2 - \epsilon\mu^2|;$$

here, $\omega_{\text{pl}}^2 = 4\pi\alpha_{\text{em}}n_e/m_e$ is the plasma frequency, α_{em} is the fine structure constant, n_e is the free electron number density and m_e is the mass of the electron. The parameter ϵ can be either 1 (scalar ALP) or 0 (pseudoscalar ALP); in the following discussion we shall always assume $\epsilon = 0$, i.e. a pseudoscalar axionlike particle. In the typical environments we are interested in, the free electron densities are of order 10^{-2} – 10^{-3} cm $^{-3}$, leading to plasma frequencies of order 10^{-11} eV. For very light masses ($m_\phi < 10^{-12}$ eV), the effective mass will be dominated by this plasma frequency.

Here, we focus on mixing which would occur as photons transverse a typical intracluster medium, where there are magnetic domains of coherent length $L \sim 1$ – 100 Kpc and intensity $B \sim 1$ – 10 μG [18]. Passing through a whole cluster of length 1 Mpc, the photons will encounter a number ($N \gg 1$) of independent magnetic domains. Initially we shall assume that every light path either crosses a cluster while travelling towards us or is originated inside a cluster. This assumption is optimistic and we shall discuss what happens when it is relaxed in Sec. III.

We are interested in the *strong-mixing limit* and when the mixing is independent of the photon energy. Strong mixing occurs in a single domain when $BL/2M \geq 1$; that is, if the magnetic fields are sufficiently strong, are coherent over a large enough region or the coupling to the ALP is

high enough. For typical cluster magnetic fields, strong mixing in a single coherent region requires $M < 10^{11}$ GeV. However, even if the mixing probability is small in a given region, strong mixing over the whole path still occurs as long as $NP_{\gamma \leftrightarrow \phi} = N(BL/2M)^2 \gg 1$. If $BL/2M \geq 1$, energy independence holds if $E \gg E_{\text{crit}}$; for weaker mixing, the energy independence extends to lower energies ($E \gg E_{\text{crit}}(BL/2M)$).

BDS estimate that the frequency-independent and strong-mixing limits are both reached when $E \geq 0.3$ – 3 keV if one assumes $m_\phi \lesssim 10^{-12}$ eV, $M \lesssim 3 \times 10^{11}$ GeV and typical properties for the intracluster medium. Thus, one expects the effects of mixing to be most significant for x-rays and γ -rays, and be relatively small for softer photons such as those in the optical or UV bands. The test we describe will exploit this, by taking the optical luminosities as a direct indicator of the true luminosity in order to normalize the luminosities in harder bands, where mixing could be significant.

In the strong-mixing and frequency-independent regimes, BDS found that, beginning with a pure photon beam, its intensity along a given line of sight will be decreased by a random factor given by

$$C \equiv \frac{I_\gamma}{I_{\text{tot}}} = 1 - \frac{1 + p_0}{2} K_1^2 - \frac{1 - p_0}{2} K_2^2, \quad (2)$$

where p_0 is the initial degree of polarization and K_1 and K_2 are uniformly distributed random variables over the interval $[-1, 1]$. The resulting probability distribution of this ratio is given by

$$f_C(c; p_0) = \frac{1}{\sqrt{1 - p_0^2}} \left\{ \tan^{-1} \left[\sqrt{a} \left(1 - \frac{2c_+}{1 + p_0} \right)^{-1/2} \right] - \tan^{-1} \left[\sqrt{a} \left(1 - \frac{2c_-}{1 - p_0} \right)^{1/2} \right] \right\}, \quad (3)$$

where $a \equiv (1 - p_0)/(1 + p_0)$, $c_\pm \equiv \min(c, (1 \pm p_0)/2)$ and $p_0 \in [0, 1)$ is the amount of linear polarization.¹ This unusual PDF is shown in Fig. 1 for three different values of p_0 . The expectation value of C given by Eq. (3) is independent of p_0 and amounts to $\bar{C} = 2/3$, meaning that, on average, one-third of emitted photons that cross $N \gg 1$ magnetic domains is turned into axions. Its standard deviation increases with p_0 and is in the range 0.2–0.3.

The curves are perhaps simplest to understand in the fully polarized case, where K_2 drops out of Eq. (2), leaving the PDF described by the Jacobian of the transformation from K_1 to C , giving $f_C(c) = \frac{1}{2}(1 - c)^{-1/2}$. In the other cases, one must marginalize over the residual degree of freedom; the hard boundaries on the distributions of the K 's lead to the features at $c = (1 \pm p_0)/2$, as seen in Fig. 1.

¹For details on the polarization induced by photon-ALP mixing, refer to Burrage *et al.* [19].

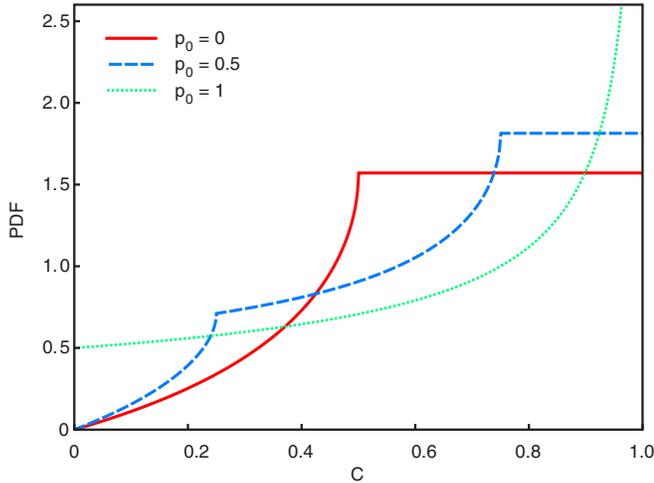


FIG. 1 (color online). Probability distribution function of C , the ratio between the photon and the total intensity, when $p_0 = 0., 0.5, 1.$

B. Other sources of scatter

If we had perfectly calibrated sources, where we knew their distances, it would be straightforward to tell whether they were dimmed due to photon-ALP mixing with some scalar field. In the absence of such calibrated sources, BDS proposed to exploit empirical relations between the luminosities in different bands of certain classes of astrophysical objects. The basic idea is that we have observations of objects in a large range of frequency bands; low frequency luminosities are assumed not to be affected by mixing, and so are taken to be an indicator of the true high frequency luminosity of the object, assuming some empirical relation. We then examine how the observed high frequency luminosities relate to that predicted; we focus on the characteristic scatter caused by the photon-ALP mixing, as any average decrease in the luminosity is absorbed in the empirical relations.

We require a relation between a quantity that is affected by photon-ALP mixing (Y), such as x-ray or γ -ray energy, and one that is unaffected (X), such as the optical luminosity or some other low energy feature of the light curve. A number of such empirical laws exist for AGNs [2,20–26], Blazars [27,28] and γ -ray bursts [29]. BDS focused on those in the form of a power law:

$$\log_{10}(Y) = a + b \log_{10}(X); \quad (4)$$

here, the definition of Y and X depends on the empirical law we are considering. If photon-ALP mixing occurs, however, we never observe Y but rather its “dimmed” counterpart CY (the same does not apply for X which is assumed to be unchanged by photon-ALP mixing.)

Even in the absence of photon-ALP mixing, it is unreasonable to assume that low frequency and high frequency luminosities are perfectly correlated, as there could be many factors affecting these luminosities which vary

from object to object. The origin of this scatter depends on the physics of the emission of the different energy photons, which varies according to the type of object under consideration. (See Sec. IVA for some specific examples in the AGN case.) The scatter is usually assumed to be a Gaussian distributed random variable with zero mean, $S = \sigma_{\text{in}} N(0, 1)$. Thus, our final data model is

$$\log_{10}(Y) = a + b \log_{10}(X) + \log_{10}(C) + S. \quad (5)$$

Here, C 's distribution is given by f_C in Eq. (3) and included only when photon-ALP mixing is assumed. The resulting probability distribution can be shown to be,

$$L_f(z; p_0) = \frac{1}{\sqrt{2\pi}\sigma_{\text{in}}} \int_0^1 dc \exp\left(-\frac{z^2}{2}\right) f_C(c; p_0), \quad (6)$$

where $z \equiv \log_{10}(Y) - a - b \log_{10}(X) - \log_{10}(C)$.

To determine whether photon-ALP mixing actually occurred, we compare two different models: the *Gaussian model*, where the scatter from the empirical relation is simply Gaussian, and the *ALP-mixing model*, where the dimming due to photon-ALP mixing is superimposed on the Gaussian scatter. Note that σ_{in} of the Gaussian scatter is empirically determined, and is normalized to match the observed scatter. In the mixing model, it is assumed there is less intrinsic scatter, in order to keep the total scatter constant. In fact, there is a minimum amount of scatter predicted by the photon-ALP mixing model, and if a probe were found with less scatter than this, we could rule out the possibility of strong mixing for that probe. The typical scatter coming from mixing alone is $\sigma_{\text{mix}} = 0.2$, but this varies with the degree of initial polarization.

Another factor to consider is the fraction of light paths which cross sufficiently magnetized regions to experience

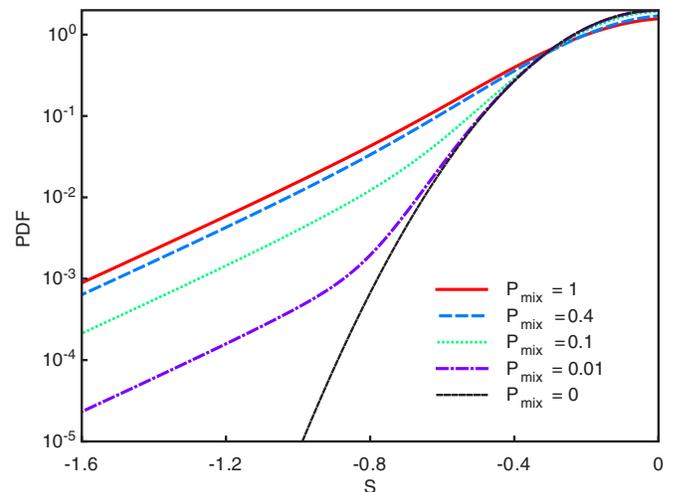


FIG. 2 (color online). Probability distribution of the scatter in the ALP-mixing model for various P_{mix} values. We assumed $p_0 = 0.1$ and $\sigma_{\text{in}} = 0.2$. For small P_{mix} , the distribution is nearly Gaussian, apart from the low intensity tail. Here and below we focus on the low end tail, where the relative deviations are largest.

strong mixing. There is no guarantee that any given source will live in a cluster environment, or that its light will pass through such an environment on its way to us. Obviously, the less likely this is, the harder it will be to constrain the coupling to axionlike particles. The final distribution will be a linear combination of the mixed distribution and the intrinsic distribution, weighted by the fraction of photon paths which experience mixing (P_{mix}). For small P_{mix} the observed variance will be dominated by the intrinsic variance; the resulting distribution is very nearly Gaussian, but the mixing will significantly increase the low luminosity tail. The dependence of the likelihood from P_{mix} is shown in Fig. 2.

C. Shot noise

Finally, another potentially important contribution to the intrinsic scatter could be from shot noise in the numbers of x-ray or γ -ray photons detected. This is of interest because it follows a Poisson distribution rather than Gaussian, and could have similar effects as the photon-ALP mixing on the scatter. The inferred x-ray luminosity is proportional to N , while the intrinsic luminosity is proportional to $\langle N \rangle$, so that we have an additional contribution to the luminosity ratio which is $\log_{10} N / \langle N \rangle$, where N is assumed to have a Poisson distribution.

A very rough idea of the impact of the shot noise can be estimated by comparing its variance to that arising from the photon-ALP mixing. The photon-ALP mixing variance increases with the initial polarization, but its minimum value is $\sigma_A^2 = 0.033$ when $p_0 = 0$. The variance from shot noise depends on the number of photons observed, and the associated variances are shown in Table I. As can be seen, the variances are only comparable to σ_A^2 for $N < 50$, while typical surveys exceed this. In the catalogs we consider below, all the objects have more than 50 net counts, with an average which is greater than 1000. Thus, we do not expect naively that shot noise will be a major issue for these observations. On the other hand, objects in the catalogs considered by BDS have photon counts as low

TABLE I. Variance of the shot-noise contribution to the scatter for different average photon counts. The variance of the ALP contribution to the scatter is at least $\sigma_A^2 = 0.033$. In order to calculate σ_s^2 we did not consider the case in which $N = 0$ photons are collected, otherwise the variance would have been infinite. This is reasonable since $N = 0$ implies no measurement at all.

$\langle N \rangle$	σ_s^2
5	0.048
10	0.023
50	0.0038
100	0.0019
1000	0.00019
10000	1.9×10^{-5}

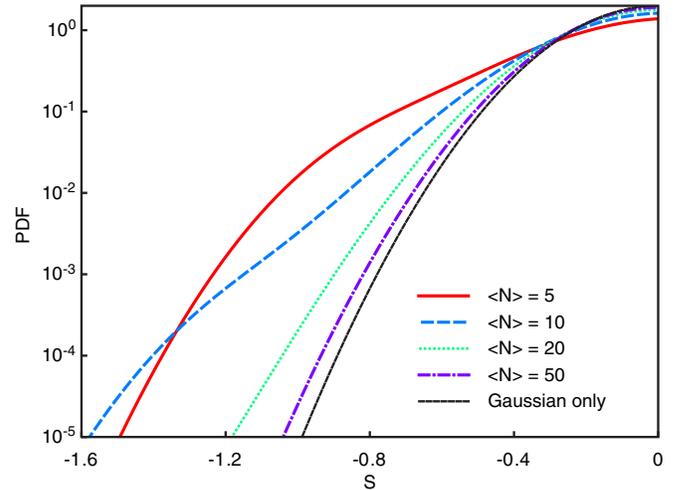


FIG. 3 (color online). Probability distribution of the scatter taking into account shot noise for various values of collected x-ray photons. We assumed $\sigma_{\text{in}} = 0.2$ for the intrinsic Gaussian noise. Like the ALP-mixing model, shot noise can increase the low luminosity tail.

as 10, with an average of 120; half of these sources have less than 50 net x-ray counts.

However, the contribution to the variance is only a rough proxy for the true effect, which is sensitive to the full distribution and, in particular, to the tail at low luminosities. Also, if the fraction of sources which are strongly mixed is relatively low, then the shot noise can be a more significant issue. See Fig. 3 for the probability distribution of the scatter in presence of shot noise for various $\langle N \rangle$ values. In Sec. V C we quantify this effect further for the statistical tests we consider.

III. STATISTICAL PROCEDURE

A. Bayes ratio

We look at a number of statistical tests to examine whether the data support the ALP-mixing model. The Bayesian approach to comparing two models is to compare their Bayesian evidences; the evidence is the likelihood of the observed data given a model (A), integrated over the model parameters (\mathbf{p}_A) and accounting for the prior distributions of the parameters:

$$\mathcal{P}(\mathbf{D}|A) = \int d^n \mathbf{p}_A \mathcal{P}(\mathbf{D}|A, \mathbf{p}_A) \mathcal{P}(\mathbf{p}_A|A). \quad (7)$$

Here we wish to compare the evidences for the Gaussian model and the ALP-mixing model; for simplicity we fix the initial polarization p_0 and the mixing fraction, P_{mix} . The resulting models have the same parameters a , b and σ_{in} , and we assume the same flat prior distribution for these parameters.

If we make the further simplifying assumption that the shapes of the likelihoods of the models are the same

(which must be established to first order by comparing the a , b and σ_{in} error bars of the models), the Bayes ratio of the models is simply given by the ratio of the peaks of the likelihoods:

$$\frac{\mathcal{P}(\mathbf{D}|\text{ALP})}{\mathcal{P}(\mathbf{D}|G)} \simeq \frac{\mathcal{P}(\mathbf{D}|\text{ALP}, \hat{\mathbf{p}}_{\text{ALP}})}{\mathcal{P}(\mathbf{D}|G, \hat{\mathbf{p}}_G)}. \quad (8)$$

Here, $\hat{\mathbf{p}}$ are the parameters which maximize the likelihoods of the two models. This ratio is effectively equivalent to the r quantity used by BDS:

$$r(p_0, P_{\text{mix}}) = 2 \ln \left[\frac{\hat{L}_{\text{ALP}}(p_0, P_{\text{mix}})}{\hat{L}_G} \right]. \quad (9)$$

If $r > 0$, then the ALP-model is preferred over the Gaussian model; if $r < 0$ the opposite is true. The absolute value of r is distributed to a good approximation as a χ^2 random variable with one degree of freedom. For example, $|r| = 9$ corresponds to a $\sim 3\sigma$ preference for one model over the other.

Below we will be comparing the likelihoods of the various model parameters for different sets of data. For the Gaussian model, the parameters include the amplitudes a and b , and the intrinsic variance σ_{in} . For the mixing model, the parameters also include p_0 and P_{mix} . The data are simply a collection of N observations of intensities of various sources in two different bands, X_i and Y_i , which are taken to be independent. The full resulting likelihood of the parameters given the data is

$$\begin{aligned} L(a, b, \sigma_{\text{in}}; p_0, P_{\text{mix}}) &= \prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma_{\text{in}}} \left[P_{\text{mix}} \int dc \exp\left(-\frac{z_i^2}{2\sigma_{\text{in}}^2}\right) f_C(c; p_0) \right. \\ &\quad \left. + (1 - P_{\text{mix}}) \exp\left(-\frac{s_i^2}{2\sigma_{\text{in}}^2}\right) \right], \end{aligned} \quad (10)$$

where $s_i(a, b) = \log_{10}(Y_i) - a - b \log_{10}(X_i)$, $z_i(a, b) = s_i(a, b) - \log_{10}(C)$ and we assumed P_{mix} does not depend on redshift.

As discussed above, we then perform a maximum likelihood estimate of the parameters (a, b, σ_{in}) for the Gaussian model ($P_{\text{mix}} = 0$) and the ALP-mixing model, leaving p_0 and P_{mix} fixed. As a result, we obtain (i) two sets of parameters: $(\hat{a}, \hat{b}, \hat{\sigma}_{\text{in}})_G$ and $(\hat{a}, \hat{b}, \hat{\sigma}_{\text{in}})_{\text{ALP}}$ and (ii) the respective maximized likelihoods: \hat{L}_G and $\hat{L}_{\text{ALP}}(p_0, P_{\text{mix}})$.

B. Goodness-of-fit tests

The Bayes ratio is the best means of comparing two models, but it does not examine whether either model provides a good fit to the data. BDS also looked at bootstrap [30] resamplings of the data sets, and showed plots of the variance versus skewness of the data (their so-called ‘‘fingerprints’’), comparing these to what is expected in the Gaussian and ALP models. In Sec. IV A we shall reproduce their analysis. In particular, we will show that much of the

structure in these fingerprints arises from resampling of a few outliers multiple times and the resulting significance of such plots is hard to quantify.

Instead, we focus on the 1D cumulative distributions of the scatter around the mean behavior which contain all of the relevant information. For this kind of unbinned data, a standard goodness-of-fit statistic is the Kolmogorov-Smirnov (KS) test, which looks at the maximum difference of the cumulative distributions. As we are fitting for parameters of the distribution, we simulate the process to see how often the observed KS statistic occurs in the two models.

As we shall see, statistics like the Bayes ratio are dominated by a few sources where the x-ray intensity is much lower than expected. The KS test is not greatly sensitive to the tails of the distribution, so we also examine some related statistical tests, the Kuiper test and the Anderson-Darling (AD) statistic. (A description of these can be found in Press *et al.* [30].) Briefly, the Kuiper statistic is the sum of the largest positive and negative difference in the observed and theoretical cumulative distributions, while the AD statistic is a renormalized version which gives more importance to the tails of the distribution. Again, these tests are calibrated using simulations of the full process.

As is evident in the cumulative plots below, while the presence of the outliers can strongly favor the ALP distribution over the Gaussian, often many more outliers are seen than is expected by either model. This most likely suggests that neither model is correct and that another explanation could be required for the low x-ray luminosity of some sources. One strong candidate is that the soft x-rays are sometimes strongly absorbed.

IV. DATA ANALYSIS

A. Previous analyses

To discriminate between the Gaussian model and the ALP-mixing model, BDS analyzed several classes of astrophysical objects where an empirical law of the form in Eq. (4) is valid. Examples of such objects are Blazars, γ -ray bursts and AGNs. They obtained relevant results using a well-known correlation between the 2500 Å and 2 keV monochromatic luminosities of AGNs [2,20–26]:

$$\log_{10}(L_{2500 \text{ \AA}}) = a + b \log_{10}(L_{2 \text{ keV}}). \quad (11)$$

BDS took into consideration 77 optically selected AGN with redshifts less than 2.7 taken from Steffen *et al.* [24]. Of these, 32 are from the COMBO-17 survey [31] and 45 from the Bright Quasar Survey [32]. These two sets are matched with x-ray measurements coming, respectively, from the Extended Chandra Deep Field-South survey [33] and the ROSAT experiment [34,35]. We will refer to this set of AGNs as the BDS-77 catalog. By analyzing it and assuming $P_{\text{mix}} = 1$, BDS obtained

$$r(p_0, P_{\text{mix}} = 1) \simeq 14$$

for $0 \leq p_0 \leq 0.4$ and $r(p_0) \geq 11$ for $p_0 > 0.4$. This corresponds, respectively, to a 3.7σ and 3.3σ evidence in favor of the ALP-mixing model.

Burrage *et al.* [14] recently extended their sample to 203 AGNs by including 126 more AGNs with redshifts less than 3.8 from Strateva *et al.* [23], optically selected from the Sloan Digital Sky Survey [36] and matched for the most part with *ROSAT* x-ray data [34,35]. The net x-ray counts are in the range 10–1500 with an average of 120 counts and 13% of AGNs below 20 counts. We will refer to this set of AGNs as the BDS-203 catalog. Using this catalog and assuming $P_{\text{mix}} = 1$, they obtained the following result:

$$r(p_0 \lesssim 0.5, P_{\text{mix}} = 1) \simeq 25,$$

which corresponds to a 5σ evidence in favor of the ALP-mixing model.

Below, we will quote only results where $p_0 = 0.1$ since (i) this is the amount of linear polarization predicted for AGN light [37] at $E = 2$ keV and (ii) the likelihood is almost insensitive to changes in p_0 .

1. The problem of x-ray absorption

In addition to photon-ALP mixing, there are many other possible sources of scatter in the intrinsic relation between an AGN's optical or UV luminosity and its x-ray luminosity. The x-rays are thought to arise from the hot coronal gas, while the lower energy photons are believed to radiate from the AGN accretion disk. X-ray emission can be enhanced if there are jets, or suppressed if the coronae are absent or disrupted, or if there is significant absorption, which might occur as a result of outflows [26]. In addition, variation may occur because the time dependent x-ray and optical luminosities are measured at different epochs,

though recent data have suggested that most of the scatter remains even for coeval observations [38].

Here we focus primarily on the effects of mixing, but if evidence suggests that there may be mixing, other sources of scatter must also be considered to explain the data. The r test employed by BDS is a simple likelihood ratio test. The significant preference shown by the BDS-77 and BDS-203 catalogs for the ALP-mixing model over the Gaussian model only demonstrates the relative fits of the models, but does not imply that either is actually a good fit to the data. One way to evaluate the goodness of fit is to look at the cumulative distribution functions (CDF). In Fig. 4 we plot the CDF of the scatter according to the two models against the empirical CDF coming from these catalogs. For both AGN sets, neither of the theory curves manages to reproduce the features of the scatter, though the Gaussian model is a much worse fit. The empirical CDF is much larger than the model CDFs at the low end of the scatter axis, meaning that the catalogs contain several objects with an x-ray to UV/optical luminosity which is much lower than what is likely via the x-ray dimming due to photon-ALP mixing.

This effect can be naturally explained if we assume that x-ray light from these AGNs has been absorbed. Steffen *et al.* [24] attempted to remove this possible contamination by excluding AGNs with flat x-ray spectra according to the effective x-ray power-law photon index Γ (some details on this procedure can be found in Sec. IV B 2); however, this was only possible for the Chandra subset, leaving some potentially contaminated AGN which can dominate the statistics. Unfortunately, the r statistic is easily dominated by outliers, so that even a small contamination can significantly bias the result.

To emphasize the effect of the unaccounted for x-ray absorption, we exclude from the BDS-77 and BDS-203 catalogs one $z = 0.067$ AGN (PG2214 + 139, also known

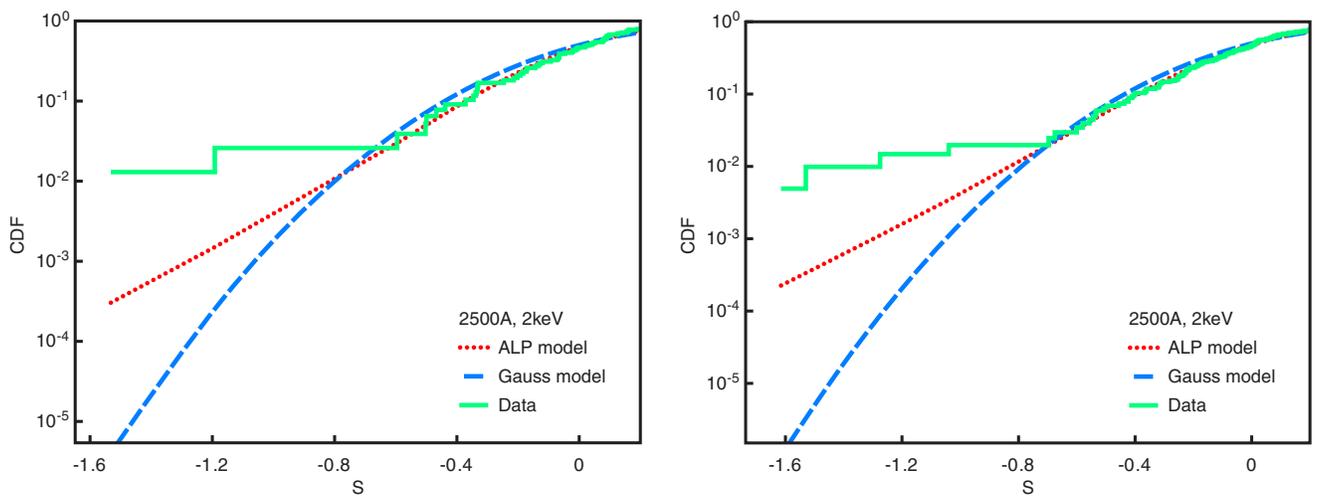


FIG. 4 (color online). Cumulative distribution functions of the scatter for the BDS-77 catalog (left panel) and BDS-203 (right panel). Each empirical CDF is plotted along with the ALP-mixing model and Gaussian model theoretical CDFs. We assumed $p_0 = 0.1$ and $P_{\text{mix}} = 1$.

as MKN 304), which is known to be heavily obscured in the x-ray [39,40]. This source has a very flat x-ray spectrum dominated by strong absorption features due to ionized gas. The gas is well modeled by a two-component warm gas, which yields one of the highest column density in XMM-Newton and Chandra. PG2214 + 139 is the biggest outlier in both data sets and has a scatter ~ 4.5 standard deviations below the average. It is actually visible in Fig. 4 as the leftmost point in both plots. By repeating the r test without PG2214 + 139, we obtain a drop in the r statistic of 64% for the BDS-77 catalog and of 36% for the BDS-203 catalog.

Another AGN in the BDS-203 catalog that deserves special attention is TGN336Z208, known in SDSS as SDSSJ134 351.12 + 000434, which has been spectroscopically classified [41] as a broad absorption line (BAL) AGN. It is thought that BALs are AGNs viewed through the nonspherical wind that surrounds the accretion disk of the supermassive black hole powering AGN emission (see Murray *et al.* [42]). The x-ray absorption resulting from this obscuration renders BAL AGNs unsuited to study the intrinsic correlation between their UV and x-ray emissions [43,44].² Moreover, Strateva *et al.* [23] only report 24 net x-ray counts for TGN336Z208, a number which is too low to permit model fitting on the spectrum. By removing both SDSSJ134 351.12 + 000434.8 and PG2214 + 139 from the BDS-203 catalog, the r statistic drops from $r \approx 25$ to $r \approx 7$, a $\sim 70\%$ decrease.

Inevitably, our focus was drawn to these particular AGN because they were less x-ray bright and dominate the Bayes ratio test; this means that some care must be taken to treat all the data consistently to avoid a posterior bias which would occur from arguing selectively to omit those AGN with the highest scatter. However, we have shown that there are independent reasons, based on their spectroscopic properties, for excluding these from the sample. In any case, these results emphasize the great sensitivity of this test to outliers and the need to ensure a sample free from absorption.

2. Bootstrap resamplings

As further evidence of the ALP-mixing model, BDS introduced the concept of “fingerprints,” which are plots based on bootstrap resamplings of the AGN data set. They created 10^5 data sets derived from the BDS-77 catalog, each of the same size as the original catalog, by sampling it with replacement. For each of these resampled data sets, they calculated the moments of the distributions, defined as

$$k_m = \left(\frac{1}{N} \sum_{i=1}^N S_i^m \right)^{1/m},$$

where N is the number of AGNs in the sample and S_i is the scatter of the i -th AGN. The fingerprints are scatter plots of the moments (e.g. the variance, k_2 , versus the skewness, k_3) for all of the resampled data sets. We have reproduced this analysis and it can be seen in Fig. 5(a).

BDS showed that there were similarities between the scatter plots generated from the data and those generated from a sample simulated with the ALP model which were not seen in the simple Gaussian case. Example simulations can be seen in Figs. 5(b) and 5(c). The data and the simulations both share a similar tail to large variance and negative skewness, which are rare in Gaussian simulations. In addition, the data and Fig. 5(c) also have a similar substructure in this tail.

However, these fingerprints are similarly sensitive to outliers in the data set which is being resampled. For example, the substructure apparent in Fig. 5(c) is due to a single outlier more than 4σ away from the mean. In some resamplings, this outlier does not appear, resulting in two islands of low variance and small skewness (which can be positive or negative); in other resamplings it can appear once, resulting in another island with higher variance and more negative skewness. The outlier can be resampled multiple times, and islands of substructure arise associated with the outlier appearing two, three or even 4 times. The actual data is similar, but there is one large outlier and one moderate outlier which provide somewhat finer structure.

Substructure, since it results from outliers, is more likely to arise from the ALP-mixing model than in the Gaussian model because of its higher tail (Fig. 2.) However, even in the ALP-mixing model such substructure is not common in resamplings based on typical simulations; a more typical result is shown in Fig. 5(b). Even in the data, when the data sets get bigger, the impact of a single data point is smaller and the substructure is less apparent. (For example, significantly less substructure was seen in the BDS-203 catalog.)

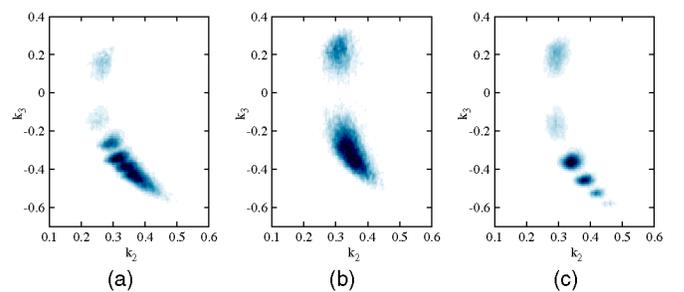


FIG. 5 (color online). Fingerprint plots of variance versus skewness. Each point represents a bootstrap resampling of 77 data points (a) from the BDS-77 catalog of AGNs, (b) simulated with the ALP-mixing model and (c) simulated with the ALP-mixing model, but where there happens to be a large outlier.

²The selection criteria in forming the BDS-203 catalog [23] included removal of BALs according to the properties of the C_{IV} and Mg_{II} absorption lines. However, these can be seen in SDSS spectra only for sources with $z > 1.55$ and $z > 0.45$ respectively, while TGN336Z208 has $z = 0.0736$.

The tail to large variance and negative skewness does remain in the data and the simulations of the ALP-mixing model and is absent from the Gaussian models. This reflects the fact that both distributions are skewed, having larger tails on the side with lower luminosities. However, the similarities are qualitative, and the fingerprints have not been used to quantify the size of the tails, where the data and the ALP-mixing model differ significantly (as was seen in Fig. 4.)

B. Defining our AGN samples

Here we take advantage of a new AGN sample, using more than 300 AGNs from the Fifth Data Release Sloan Digital Sky Survey/Xmm-Newton Quasar Survey (Young *et al.* [2,45]). This sample contains three optical monochromatic luminosities and six x-ray monochromatic luminosities. The multiband data allow us to compute the r statistic for 18 different combinations of optical/UV and x-ray monochromatic luminosities. In addition, spectral fits of each AGN aid in excluding x-ray obscured AGNs from our sample. The harder x-ray luminosities available in this sample should also be less subject to absorption.

1. The SDSS/XMM-Newton catalog

The SDSS/XMM catalog contains 792 AGNs. We first select 340 AGNs by applying the same cuts performed in Young *et al.* [2]:

- (1) we exclude all radio loud quasars;
- (2) we exclude all broad absorption line quasars;
- (3) we consider only AGNs with x-ray detections characterized by a signal-to-noise ratio (S/N) greater than 6. The objects we are left with have net x-ray counts in the range 51–39 300, with an average of 1336 counts (so that shot noise is minimal). This allows for spectral fits to be made over the 0.5–10 keV band for each source;
- (4) we select AGNs whose preferred x-ray spectral fit is a single power-law with no intrinsic absorption;

- (5) between the remaining AGNs, we choose only those with a good spectral fit, that is with a reduced χ^2 statistic (i.e. $\chi^2/\text{d.o.f.}$) smaller than 1.2.

The resulting set of AGNs lies in the redshift range 0.1–4.4. We will refer to it as the Full catalog.

2. Removing obscured AGNs

We consider also a further cut on the effective x-ray power-law photon index Γ (column 10 in Table II of Young *et al.* [45]). The photon spectral index is defined as: $\Gamma = -\alpha + 1$, where $-\alpha$ is the exponent obtained by fitting the x-ray part of the AGN spectrum with a single power-law model. A low Γ value implies a flat spectrum, that is a spectrum where the soft x-ray component (i.e. the 0.5–2 keV band) is much weaker than the hard x-ray component (i.e. the 2–10 keV band). There is strong evidence [26,40,43,44] that x-ray absorption is the primary cause of soft x-ray weakness. Intuitively, this happens because the low energy x-rays are more easily absorbed than high energy ones.

Following Steffen *et al.* [24], we select only those AGNs where $\Gamma > 1.6$, thus reducing our sample to 320 AGNs. We will refer to this as the High- Γ catalog. One can further reduce the impact of x-ray absorption by imposing cuts on the signal-to-noise ratio; it is possible, for example, to form AGN catalogs where $S/N > 10$ or $S/N > 20$. We statistically analyzed these samples as well and we ended up with results very similar to those obtained from the High- Γ catalog. Moreover, we applied cuts on the Galactic column density in the direction of the sources and did not find any significant correlation with the r statistic.

3. Multiwavelength data

A significant advantage of the sample of Young *et al.* is its many frequency bands. This consists in the rest-frame monochromatic luminosities at the following frequencies: 1500 Å, 2500 Å, 5000 Å for the optical/UV side and 1 keV, 1.5 keV, 2 keV, 4 keV, 7 keV, 10 keV for the x-ray

TABLE II. Outcome of the r test applied to the Full (upper line) and High- Γ (lower line) catalogs at the various wavelengths and for different mixing fractions. Positive values imply that the ALP-mixing model is preferred; negative values imply that the Gaussian model is preferred. We assumed $p_0 = 0.1$, but the results do not change greatly for different values.

	1 keV	1.5 keV	2 keV	4 keV	7 keV	10 keV
$P_{\text{mix}} = 1.0$	63	45	26	-0.91	-2.6	0.39
	9.2	-1.9	-5.4	-7.8	-3.9	-0.20
$P_{\text{mix}} = 0.7$	62	43	25	-1.3	-2.8	0.28
	8.6	-2.8	-6.2	-8.2	-4.0	-0.34
$P_{\text{mix}} = 0.4$	60	44	26	0.60	-1.5	0.86
	10	0.87	-2.4	-5.6	-2.6	0.38
$P_{\text{mix}} = 0.1$	42	32	19	2.4	-0.12	0.66
	9.2	4.7	2.6	-0.90	-0.50	0.55

side. By applying the statistical analysis outlined in Sec. III to multiwavelength data, we can check whether the r statistic varies as the x-ray frequency goes from the soft to the hard side of the spectrum. The ALP-mixing model does not predict any variation since, as long as we are in the frequency-independent limit (see Sec. I), the oscillation probability in Eq. (1) is insensitive to the photon energy. On the other hand, soft x-rays are more easily absorbed than hard x-rays. It is therefore clear that a multiwavelength analysis can help discriminating between photon dimming due to mixing and x-ray absorption.

C. Results

In Table II we report the results we obtained by analyzing the Full and High- Γ catalogs using the method outlined in Sec. III at the various wavelengths and for four P_{mix} values. Figure 6 shows the same results in a r vs x-ray energy plot. In the following discussion we shall always assume an optical/UV wavelength of 2500 Å since the dependence on optical wavelength is barely noticeable.

1. Full catalog

The outcome of the r test significantly depends on the analyzed x-ray frequency. Regardless of P_{mix} , the ALP-mixing model is strongly preferred in the soft x-ray band (1 keV, 1.5 keV and 2 keV). For example, for $E = 1$ keV and $P_{\text{mix}} = 1$ we have $r \approx 63$, a $\sim 8\sigma$ preference for the ALP-mixing model. However, as soon as we move to the hard x-ray bands (4 keV, 7 keV and 10 keV), r quickly approaches zero: neither model is preferred. The energy dependence suggests that x-ray absorption, rather than photon-ALP mixing, represents the main contribution to the scatter at low energies. Where the r value is high, the

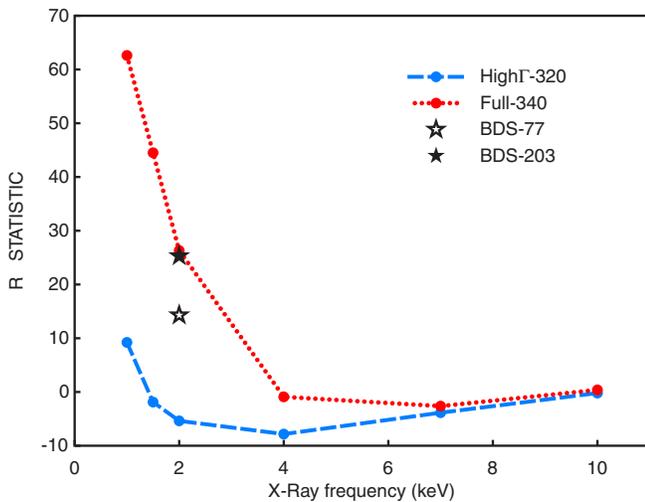


FIG. 6 (color online). Value of the r statistic for the various wavelengths and catalogs ($P_{\text{mix}} = 1$). The evidence is significantly lower in those bands where the effects of absorption are expected to be less.

PDFs are not especially consistent with the ALP model or the Gaussian model.

In Fig. 7(a) we plot the CDFs for the ALP-mixing model and the Gaussian model together with the empirical CDF of the Full catalog. In the soft x-ray band both the ALP-mixing model and the Gaussian model fail to reproduce the features of the scatter. The preference for the ALP-mixing model over the Gaussian model given by the r test at these energies is due to a very bad performance of the latter rather than to a good performance of the former. Moreover, both models systematically overestimate the x-ray luminosities: x-ray absorption, again, could be the culprit of this discrepancy. In the hard x-ray band, where absorption is less likely to affect the photons, the empirical CDF stays in between the two theoretical curves: no model is to be preferred.

2. High- Γ catalog

The signal for the ALP-mixing model dramatically drops upon removing the 20 AGNs which have lower photon spectral index. As shown in Table II, this is true for all the considered P_{mix} values. These are explained looking at the High- Γ CDFs in Fig. 7(b).

At the lowest energies, the distribution is very much like the expected ALP prediction, especially in the 1 keV case. However, we do not expect the photon-ALP mixing signal to be frequency dependent, while we see the x-ray absorption is. This suggests also that the agreement with the ALP model is likely the accidental effect of a small amount of residual x-ray intrinsic absorption, rather than due to photon-ALP mixing.

V. SIMULATED DATA SETS

It is useful to understand the sensitivity to the question of mixing which we expect to find given the catalogs we consider. Using simulated samples drawn from the mixing distribution, we can measure the distribution of r values which would typically occur and then compare these to the r values obtained in Sec. IV C. If we measure a significantly different value, this could indicate that there may be an alternative explanation of the signal.

The same simulated data sets can be used to derive the statistical significance of the goodness-of-fit tests introduced in Sec. III, namely, the KS, Kuiper, and AD tests. We will do so in Sec. V D.

For the models we analyzed in Sec. II—the ALP-mixing model, the Gaussian model and the shot-noise model—the PDF's are known and it is straightforward to produce samples consistent with each. For each assumed scatter model, we adjust the parameters to match those of the observed samples; in particular, we match the total variance σ_{tot}^2 of the scatter distribution for each sample, as well as the number N_{AGN} of AGNs in each sample. We consider the three combinations of N_{AGN} and σ_{tot}^2 corresponding to

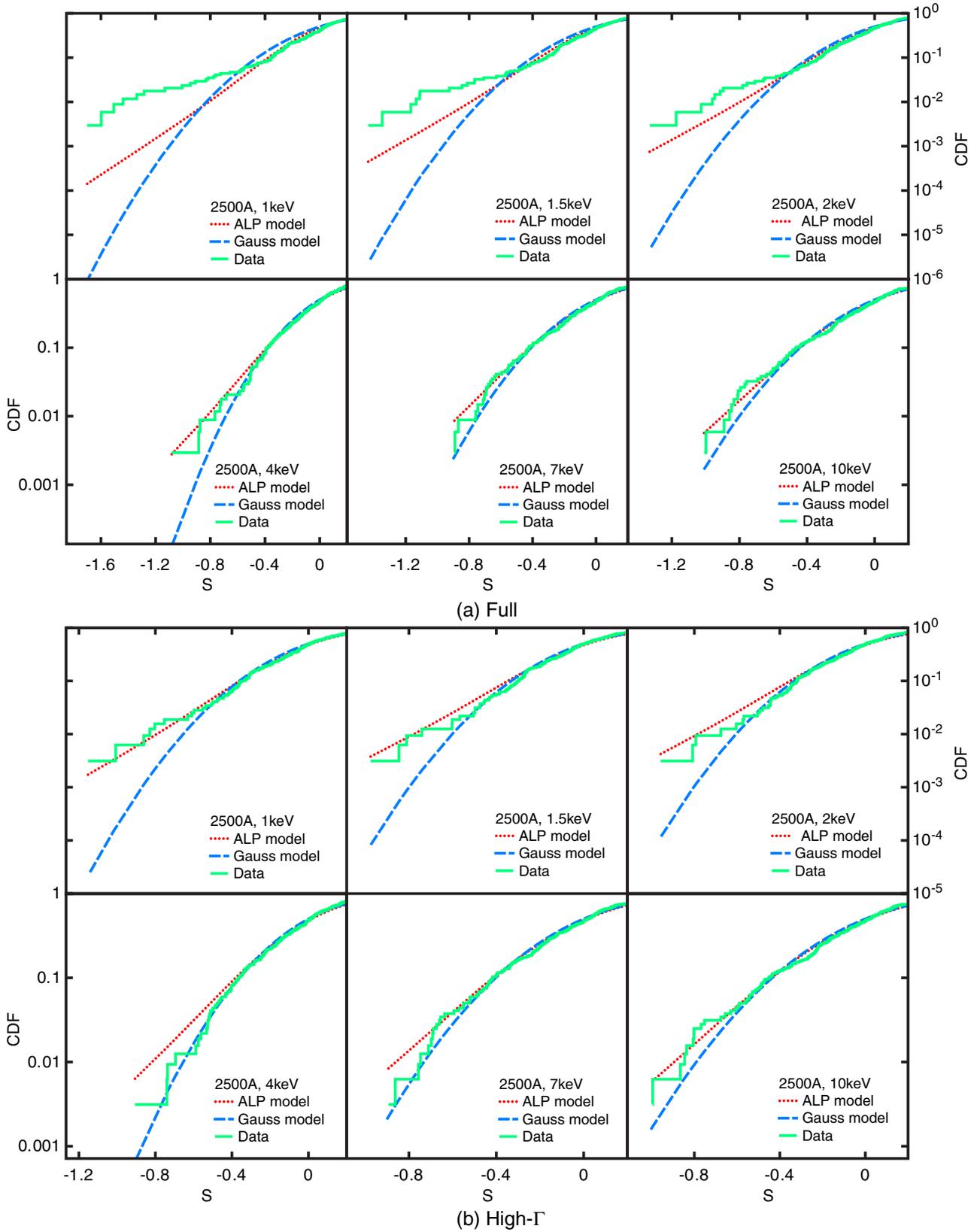


FIG. 7 (color online). Cumulative distribution functions of the scatter for the Full (a) and High- Γ (b) catalogs. The empirical CDF is plotted along with the ALP-mixing model and Gaussian model theoretical CDFs. For the ALP-mixing model we assumed $p_0 = 0.1$ and $P_{\text{mix}} = 1$. When $P_{\text{mix}} < 1$, the data curve stays the same while the CDF of the ALP-mixing model tends to that of the Gaussian model.

TABLE III. Combinations of N_{AGN} and σ_{tot} used to generate simulated scatter samples of the AGN catalogs.

Catalog	N_{AGN}	$\sigma_{1 \text{ keV}}$	$\sigma_{1.5 \text{ keV}}$	$\sigma_{2 \text{ keV}}$	$\sigma_{4 \text{ keV}}$	$\sigma_{7 \text{ keV}}$	$\sigma_{10 \text{ keV}}$
BDS-203	203			0.340			
Full	340	0.357	0.314	0.300	0.296	0.319	0.344
High- Γ	320	0.282	0.259	0.260	0.282	0.315	0.340

the BDS-203, Full and High- Γ catalogs. All values of N_{AGN} and σ_{tot} are shown in Table III.

For each model and each data set, we generate 10000 scatter samples, and apply the r test outlined in Sec. III to each generated sample. Histograms of these results are shown in the figures below. Each figure reports also the measured signal as a vertical black line and its statistical significance (i.e. the value of the empirical CDF) in the legend.

A. ALP-mixing model results

In Figs. 8–10, we show the r distributions obtained for the various catalogs for $E = 2 \text{ keV}$. Before comparing them with the measured signals, let us point out some general properties of the r test which stem from these distributions.

First, the r test may not be sufficient to provide a conclusive preference for photon-ALP-mixing even if the latter is actually happening. Consider the expectation value of the r statistic for an AGN sample with $N_{\text{AGN}} = 203$ and $\sigma_{\text{tot}} = 0.34$ (i.e. the BDS-203 catalog), in the best scenario of $P_{\text{mix}} = 1$. This amounts to $\langle r \rangle = 3.2$, which corresponds to an evidence for the ALP-mixing model of less than 2σ . Even if photon-ALP mixing were happening, the BDS-203

catalog would usually not be enough to provide a detection. The statistical significance attainable with the r test increases with N_{AGN} and P_{mix} and decreases with the intrinsic variance of the data set σ_{in}^2 . For the Full and High- Γ catalogs, we have $\langle r \rangle \sim 9$ and $\langle r \rangle \sim 21$ which correspond to a preference for the ALP-mixing model of 3σ and 4.5σ respectively.

When $P_{\text{mix}} < 1$, we expect it to be harder to discriminate between the ALP-mixing model and the Gaussian model and, therefore, the average of the r statistic should be smaller. The simulations confirm this behavior. For example, if $P_{\text{mix}} = 0.4$, then the average r values for the Full and High- Γ data sets shrink to $\langle r \rangle \sim 6.6$ and $\langle r \rangle \sim 13$, that is a 2.6σ and 3.6σ preference, respectively.

The figures also demonstrate another issue: the probability of a false-negative result, i.e. the r test yielding a preference for the Gaussian model when the scatter comes from the ALP-mixing model, is not negligible even when $P_{\text{mix}} = 1$. We define the false-negative probability as the value of the empirical CDF at $r = 0$. For the BDS-203 catalog, this amounts to 0.27: there is a 27% probability that the r test gives a false-negative result when applied to a data set similar to the BDS-203 catalog. The false-negative probability for the Full and High- Γ catalogs is smaller since both have a smaller scatter variance than BDS-203, but is still significant: 11% and 4%, respectively.

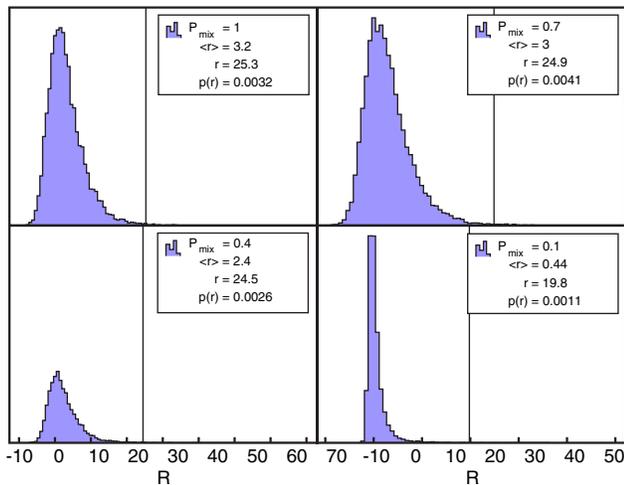


FIG. 8 (color online). r distribution for the BDS-203 catalog at $E = 2 \text{ keV}$ when the scatter is distributed according to the ALP-mixing model. The measured signal, r , is represented by the vertical black line, while its statistical significance is quoted in the legend as $p(r)$.

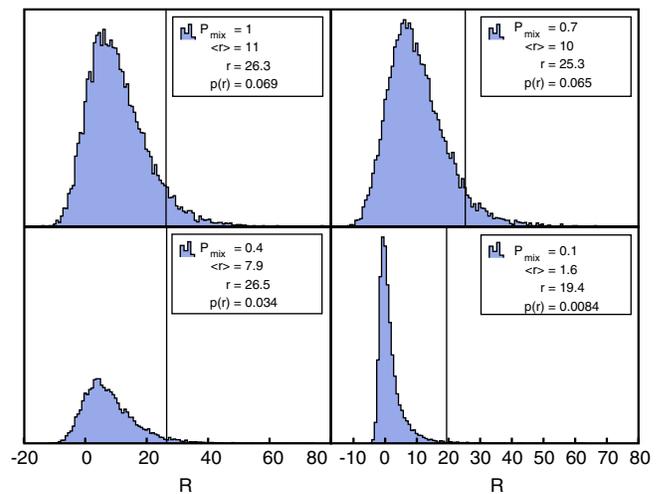


FIG. 9 (color online). r distribution for the Full catalog at $E = 2 \text{ keV}$ when the scatter is distributed according to the ALP-mixing model.

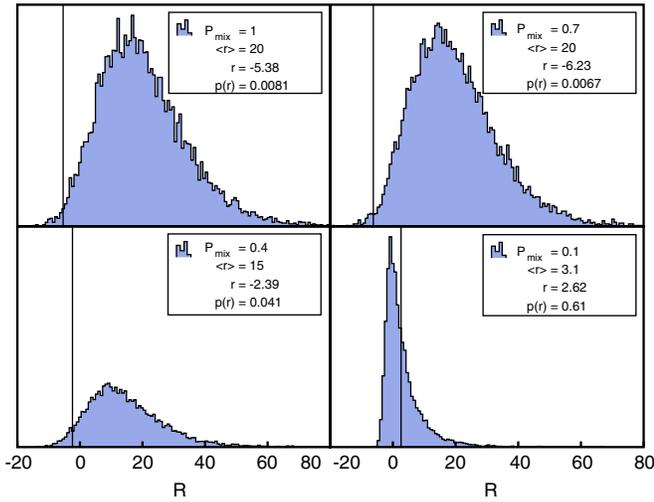


FIG. 10 (color online). r distribution for the High- Γ catalog at $E = 2$ keV when the scatter is distributed according to the ALP-mixing model.

Clearly the limiting factors are the number N_{AGN} of AGNs and the intrinsic variance σ_{in}^2 of the scatter of the empirical relation taken into consideration. The contribution to the scatter from mixing is fixed (for a given P_{mix}), so if the total scatter is larger, the greater will be the relative contribution from the intrinsic scatter and the harder it is to distinguish photon-ALP mixing. Statistically significant results require using data sets larger than those considered in this paper. For example, 500 AGNs with $\sigma_{\text{tot}} = 0.26$, in the best scenario of $P_{\text{mix}} = 1$, lead to an average evidence 5.7σ , while 1000 objects yield an average evidence of 7.7σ .

1. Comparison with the measured signal

The measured r value for the BDS-203 catalog is much larger than what would typically occur when applying the r test to a scatter sample distributed according to the ALP-mixing model. As can be seen by Fig. 8, regardless of the considered P_{mix} , we find the probability of the observed signal to be smaller than 0.4%. We obtain a similar result for the BDS-77 catalog.

A similar situation is found for the Full catalog, where the measured signal is systematically higher than the expected one—see Fig. 9. The probability of such measurements reaches the 7% level for $P_{\text{mix}} = 1$.

The BDS-77, BDS-203, and Full catalogs all suggest that something different than photon-ALP mixing is affecting the x-ray-to-optical luminosity ratio in a way that mimics the mixing effect. As we already pointed out in Sec. IV A, x-ray absorption could be the cause.

The High- Γ catalog, due to its low scatter variance, can potentially provide a statistically significant detection of photon-ALP mixing. This is evident from Fig. 10, where the average outcome of the r test can be as high as $\langle r \rangle = 20$. However, the measured signal never shows a significant preference for the ALP-mixing model.

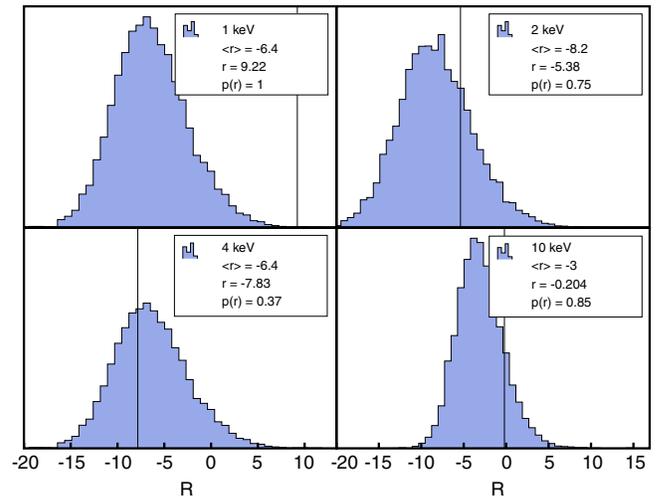


FIG. 11 (color online). r distribution for the High- Γ catalog at $P_{\text{mix}} = 1$ for four different frequencies when the scatter is distributed according to the Gaussian model.

In fact, it favors the Gaussian model in most of the cases ($P_{\text{mix}} = 1, 0.7, 0.4$).

B. Gaussian model results

The measured r values are compatible with what expected from a Gaussian scatter only in the following cases:

- (1) for the High- Γ catalog, when $E \geq 2$ keV;
- (2) for the Full catalog, when $E \geq 4$ keV.

This means that the Gaussian model performs well according to the r test where we expect x-ray absorption to be negligible. An example of this is shown in Fig. 11, where we show the r distribution for the High- Γ catalog for four different frequencies. For frequencies higher than 2 keV, the measured signal lies comfortably within the expected distribution of r values.

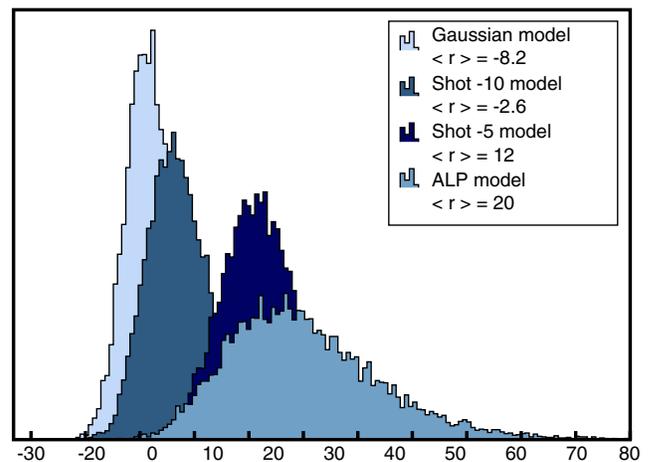


FIG. 12 (color online). r distribution for the High- Γ catalog at $E = 2$ keV and $P_{\text{mix}} = 1$ for four different models. The shot-noise models stay in between the Gauss and ALP-mixing models.

TABLE IV. Statistical significance of the KS, Kuiper, and AD statistics obtained for the various catalogs. For the ALP-mixing model we assumed $p_0 = 0.1$ and $P_{\text{mix}} = 1$; for the shot-noise model we used $\langle N \rangle = 5$.

		1 keV			1.5 keV			2 keV			4 keV			7 keV			10 keV		
ALP	High- Γ	0.17	0.15	0.0021	0.13	0.3	6.2e-4	0.061	0.22	4.2e-4	0.078	0.028	6.6e-4	0.061	0.18	0.0051	0.033	0.023	0.017
	Full	0	0	0.03	2e-4	8e-5	0.064	0.028	0.0042	0.014	0.082	0.22	0.0026	0.024	0.067	0.008	0.036	0.046	0.026
	BDS-77							0.27	0.11	0.067									
	BDS-203							0.056	0.008	0.067									
Gauss	High- Γ	0.005	2.2e-4	0.015	0.035	0.004	0.017	0.005	7e-4	0.011	0.043	0.011	0.014	0.089	0.03	0.035	0.051	0.0039	0.071
	Full	0	0	3.6e-4	0	0	0.001	4e-5	0	0.0031	0.06	0.0068	0.031	0.039	0.0053	0.046	0.073	0.0083	0.099
	BDS-77							0.062	0.029	1.6e-4									
	BDS-203							0.012	0.0035	3.2e-4									
Shot	High- Γ	0.064	0.039	0.0016	0.026	0.042	0.001	0.01	0.041	9e-4	0.019	0.11	7.8e-4	0.024	0.057	0.0047	0.014	0.0054	0.013
	Full	0	0	0.0019	8e-5	0	0.0044	0.0084	7.6e-4	0.011	0.026	0.12	0.0026	0.0087	0.016	0.0059	0.018	0.014	0.021
	BDS-77							0.19	0.062	0.0019									
	BDS-203							0.032	0.0023	0.0021									

C. Shot-noise model results

The introduction of shot-noise on top of a Gaussian-distributed scatter can significantly alter the outcome of the r test if less than $\langle N \rangle = 20$ x-ray photons are collected. The impact on the r distribution is a shift of its peak towards higher r values and a broadening of its width. Moreover, when the number of collected photons is smaller than ~ 10 , shot noise can even trick the r test into showing a preference for the ALP model (i.e. a positive r value). For example, for the High- Γ catalog and when $\langle N \rangle = 5$, the r test yields an average value of 12, that is a $\sim 3.5\sigma$ preference for the ALP-mixing model. All of this is clear from Fig. 12, where we show the r distributions for the High- Γ catalog for four different models: ALP-mixing model, Gaussian model, shot-noise model with $\langle N \rangle = 5$ and shot-noise model with $\langle N \rangle = 10$.

We do not expect the Full and High- Γ catalogs to be affected by shot noise since all their AGNs are detected with far more than 20 x-ray counts. However, a non-negligible number of AGNs in the BDS-77 and BDS-203 catalogs—around the 10% of the total, see Sec. IVA—have photon counts below 20. Therefore, shot noise may contribute to the high r values measured for those catalogs. Nevertheless, the impact of shot noise is likely to be much smaller than that due to x-ray absorption.

D. Goodness-of-fit tests

We tested the measured scatter against the ALP-mixing, Gaussian, and shot-noise models by making use of the Kolmogorov-Smirnov, Kuiper and Anderson-Darling tests. The Kolmogorov-Smirnov and Kuiper tests tend to be more sensitive to the centers of the distribution, while the Anderson-Darling is weighted to emphasize the tails. For every data set, x-ray frequency and scatter model, we calibrated the goodness-of-fit tests by applying them to 50000 simulated scatter samples. Table IV shows the statistical significances obtained by comparing the measured statistics to the simulated distributions. Values lower than 0.05 imply a rejection of the model at the 2σ level.

The most obvious result of the goodness-of-fit tests is that none of the distributions does that well in reproducing the data. Regardless of the analyzed model or data set, most of the measured signals have a statistical significance smaller than 10%. Every test fails at the 1σ level, and only one data set is accepted in all three tests at the 2σ level (the BDS-77 set for the ALP-mixing model).

The r test is most sensitive to the tails of the distributions, so we expect the AD test to be the best predictor of the r value. Indeed, this is the case: where the r value is high, the AD test gives a very low value for the Gaussian distribution, and higher values for the ALP-mixing model. However, though the values are higher for the ALP-mixing model, they remain far lower than would typically be expected. This is consistent with what can be seen from

the distributions in Figs. 4 and 7. (Note that the figures show only the left tail.)

In the soft x-ray band, most of the tests yield very low significances, particularly in the Full data set (the Full catalog is never fit by any model for $E \leq 2$ keV.) This suggests that none of the distributions well reflects the data and that something, such as x-ray absorption, is missing in the models. This is generally improved when going to the harder x-ray bands or the better cleaned High- Γ data set. This is not surprising, as the scatter in these cases tends to fall in between the ALP-mixing model and Gaussian distributions.

These tests generally confirm what we have seen above: preference for the ALP-mixing model, as reflected in the r statistic, should not be confused with the data being consistent with this model. Indeed, there are strong indications that the simple models we consider fail to explain the data, particularly in the soft x-ray band.

VI. CONCLUSIONS

The ratios of luminosities of high redshift objects, introduced by Burrage *et al.* [1], offers a new avenue to explore the possibility of strong mixing between photons and axions, which in principle can have significant power to constrain these models. However, much effort must be taken to ensure a homogenous sample of objects, so that any scatter is due to the coupling to axions.

Here, we have reproduced the analysis of Burrage *et al.* studying scatter of the empirical relation between the optical/UV and x-ray monochromatic luminosities of AGNs. In addition to the AGN catalogs already analyzed by them — BDS-77 and BDS-203 —, we considered two samples from the SDSS/XMM-Newton Quasar Survey [2]: Full-340 and High- Γ -320. These data sets have multiwavelength information which has allowed a more thorough investigation of the model constraints.

We ran statistical analyses on 18 combinations of optical/UV and x-ray monochromatic luminosities and found no compelling evidence for the presence of ALPs. Whenever we found a signal compatible with the presence of ALPs, it was coming either from the soft x-ray band, where absorption is more likely to happen, or from the Full catalog, which is thought to include more absorbed AGNs. Moreover, the ALP-mixing model fails to reproduce the features of the scatter even where it is preferred over its competing model, the simple Gaussian model.

By means of simulations, we calculated the distribution of the r statistic used by Burrage *et al.* to discriminate between the ALP-mixing model and the Gaussian model. We found out that actual data yields values of the r statistic which are either significantly higher (BDS-203 and Full catalogs) or lower (High- Γ catalog) than what we would typically expect if photon-ALP mixing were taking place. Stated differently, we see ALP evidence where there should not be sensitivity, and we do not see

evidence in the majority of data sets which should be most sensitive.

This behavior suggests that another source of scatter, such as x-ray absorption, is taking place and casts doubts on the suitability of the r statistic alone to estimate photon-ALP mixing. This is supported by a detailed examination of the low luminosity sources which dominate the statistical tests. We found that the BDS-77 and BDS-203 catalogs, respectively, contain one and two AGNs which are known to be x-ray-weak independently of photon-ALP mixing. Upon removing these sources, the value of the r statistic significantly drops in both cases, as does the preference for the ALP-mixing model over the Gaussian model.

Thus, while this new method for observing photon-ALP mixing is in principle very powerful, considerable care must be taken. Given the many ways that scatter can be introduced into the relations between the low and high energy luminosities of objects, evidence for mixing must be closely examined that it fits the expected signal. In particular, the scatter should follow the expected PDF

and be independent of the energy of the high energy photons. Ideally, it should also be observed in multiple classes of objects, where their intrinsic scatter have different physical origins.

This also highlights the importance of finding classes of objects where the intrinsic scatter is small or well understood. Since the scatter from mixing is well understood and fixed, if any sample is observed with less scatter than expected from mixing, then mixing can be ruled out, or at the very least, the probability of mixing, P_{mix} , can be constrained.

ACKNOWLEDGMENTS

We thank M. Young for providing us with the monochromatic luminosities of the Quasars in the SDSS/Xmm-Newton Quasar Survey and D. Shaw for many useful remarks. In addition, we acknowledge many useful conversations with C. Burrage, H. Campbell, T. Clemson, A. Martinez-Sansigre, B. Nichol, K. Romer and D. Thomas.

-
- [1] C. Burrage, A. Davis, and D. J. Shaw, *Phys. Rev. Lett.* **102**, 201101 (2009).
 - [2] M. Young, M. Elvis, and G. Risaliti, *Astrophys. J. Suppl. Ser.* **183**, 17 (2009).
 - [3] S. M. Carroll, *Phys. Rev. Lett.* **81**, 3067 (1998).
 - [4] C. Amsler *et al.* (Particle Data Group), *Phys. Lett. B* **667**, 1 (2008).
 - [5] G. G. Raffelt, *Lect. Notes Phys.* **741**, 51 (2008).
 - [6] J. Khoury and A. Weltman, *Phys. Rev. D* **69**, 044026 (2004).
 - [7] P. Brax, C. van de Bruck, and A. Davis, *Phys. Rev. Lett.* **99**, 121103 (2007).
 - [8] D. F. Mota and D. J. Shaw, *Phys. Rev. D* **75**, 063501 (2007).
 - [9] C. Csáki, N. Kaloper, and J. Terning, *Phys. Rev. Lett.* **88**, 161302 (2002).
 - [10] C. Csáki, N. Kaloper, and J. Terning, *Phys. Lett. B* **535**, 33 (2002).
 - [11] B. A. Bassett and M. Kunz, *Phys. Rev. D* **69**, 101305 (2004).
 - [12] B. A. Bassett and M. Kunz, *Astrophys. J.* **607**, 661 (2004).
 - [13] A. Mirizzi, G. G. Raffelt, and P. D. Serpico, *Lect. Notes Phys.* **741**, 115 (2008).
 - [14] C. Burrage, A. Davis, and D. J. Shaw, arXiv:0912.1716.
 - [15] D. Hooper and P. D. Serpico, *Phys. Rev. Lett.* **99**, 231102 (2007).
 - [16] P. Sikivie, *Phys. Rev. Lett.* **51**, 1415 (1983).
 - [17] G. Raffelt and L. Stodolsky, *Phys. Rev. D* **37**, 1237 (1988).
 - [18] C. L. Carilli and G. B. Taylor, *Annu. Rev. Astron. Astrophys.* **40**, 319 (2002).
 - [19] C. Burrage, A. Davis, and D. J. Shaw, *Phys. Rev. D* **79**, 044028 (2009).
 - [20] G. Zamorani *et al.*, *Astrophys. J.* **245**, 357 (1981).
 - [21] Y. Avni and H. Tananbaum, *Astrophys. Lett.* **262**, L17 (1982).
 - [22] C. Vignali, W. N. Brandt, D. P. Schneider, S. F. Anderson, X. Fan, J. E. Gunn, S. Kaspi, G. T. Richards, and M. A. Strauss, *Astron. J.* **125**, 2876 (2003).
 - [23] I. V. Strateva, W. N. Brandt, D. P. Schneider, D. G. Vanden Berk, and C. Vignali, *Astron. J.* **130**, 387 (2005).
 - [24] A. T. Steffen, I. Strateva, W. N. Brandt, D. M. Alexander, A. M. Koekemoer, B. D. Lehmer, D. P. Schneider, and C. Vignali, *Astron. J.* **131**, 2826 (2006).
 - [25] D. Just, W. Brandt, O. Shemmer, A. Steffen, D. Schneider, G. Chartas, and G. Garmire, *Astrophys. J.* **665**, 1004 (2007).
 - [26] R. R. Gibson, W. N. Brandt, and D. P. Schneider, *Astrophys. J.* **685**, 773 (2008).
 - [27] S. D. Bloom, in *The First GLAST Symposium*, edited by S. Ritz, P. Michelson, and C. A. Meegan AIP Conf. Proc. No. 921 AIP, New York, (2007).
 - [28] G. Z. Xie, Y. H. Zhang, and J. H. Fan, *Astrophys. J.* **477**, 114 (1997).
 - [29] B. E. Schaefer, *Astrophys. J.* **660**, 16 (2007).
 - [30] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C++: The Art of Scientific Computing, 2002* (Cambridge University Press, Cambridge England, 2002).
 - [31] C. Wolf *et al.*, *Astron. Astrophys.* **421**, 913 (2004).
 - [32] M. Schmidt and R. F. Green, *Astrophys. J.* **269**, 352 (1983).
 - [33] B. D. Lehmer *et al.*, arXiv astro-ph/0506607v2.
 - [34] W. Voges *et al.* *Astron. Astrophys.* **349**, 389 (1999).

- [35] E. Pfeffermann *et al.*, in *Soft X-ray Optics and Technology; Proceedings of the Meeting, Berlin, Federal Republic of Germany, Dec. 8-11, 1986*, Bellingham, WA, edited by E.-E. Koch and G. Schmahl, Vol. 733 (Society of Photo-Optical Instrumentation Engineers, Bellingham, WA, 1987), p. 519.
- [36] D.G. York *et al.*, *Astrophys. J.* **120**, 1579 (2000).
- [37] G. Bao, P. Hadrava, P. J. Wiita, and Y. Xiong, *Astrophys. J.* **487**, 142 (1997).
- [38] F. Vagnetti, S. Turriziani, D. Trevese, and M. Antonucci, [arXiv:1005.0144](https://arxiv.org/abs/1005.0144).
- [39] E. Piconcelli, E. Jimenez-Bailón, M. Guainazzi, N. Schartel, P.M. Rodríguez-Pascual, and M. Santos-Lleó, *Mon. Not. R. Astron. Soc.* **351**, 161 (2004).
- [40] E. Piconcelli, E. Jimenez-Bailón, M. Guainazzi, N. Schartel, P.M. Rodríguez-Pascual, and M. Santos-Lleó, *Astron. Astrophys.* **432**, 15 (2005).
- [41] A. Georgakakis, I. Georgantopoulos, M. Vallbé, V. Kolokotronis, S. Basilakos, M. Plionis, G.C. Stewart, T. Shanks, and B.J. Boyle, *Mon. Not. R. Astron. Soc.* **349**, 135 (2004).
- [42] N. Murray, J. Chiang, S.A. Grossman, and G.M. Voit, *Astrophys. J.* **451**, 498 (1995).
- [43] W.N. Brandt, A. Laor, and B.J. Wills, *Astrophys. J.* **528**, 637 (2000).
- [44] S.C. Gallagher, W.N. Brandt, G. Chartas, and G.P. Garmire, *Astrophys. J.* **567**, 37 (2002).
- [45] M. Young, M. Elvis, and G. Risaliti, *Astrophys. J. Suppl. Ser.* **185**, 250 (2009).