Constraints on axion-like particles from a combined analysis of blazars observed with MAGIC and LST-1

I. Batkovic[a,b], G. D'Amico[c], M. Doro[a,b], F. Schiavone[d]

[a]*Department of Physics and Astronomy, University of Padova, via Marzolo 8, 35131 Padova, Italy*
[b]*INFN Sezione di Padova, via Marzolo 8, 35131 Padova, Italy*
[c]*Institut de Fisica d'Altes Energies (IFAE), Barcelona, Spain*
[d]*INFN Sezione di Bari, Via Amendola 173, 70125, Bari, Italy*

Abstract

The mixing of high-energy photons with axion-like particles (ALPs) in external magnetic fields can imprint observable irregularities on the gamma-ray spectra of blazars. We analyse four blazars — Mrk 421, Mrk 501, BL Lacertae and 1ES 1959+650 — observed with the Major Atmospheric Gamma-ray Imaging Cherenkov (MAGIC) telescopes between 2014 and 2024 and with the first Large-Sized Telescope (LST-1) of the Cherenkov Telescope Array Observatory (CTAO) between 2020 and 2024, for a total of 980.7 h and the largest dataset assembled to date for a ground-based ALP search. MAGIC data are reconstructed with the automated pipeline `autoMAGIC`, and both instruments are analysed identically after reconstruction.

The central element of this work is a statistical framework that combines data at the likelihood level within and across instruments. The dataset is segmented into 442 Bayesian Blocks (BBs), each an interval of statistically constant flux state, and each block contributes a profile test statistic and a Monte-Carlo-calibrated number of degrees of freedom evaluated over the ALP parameter grid. Because these quantities are computed after profiling over all nuisance parameters, blocks are additive: further sources, observation periods or instruments can be incorporated by summation, without recomputing the full likelihood. We find that few high-flux BBs dominates the constraining power, so that rare flaring episodes dominate the combined result.

No statistically significant evidence for ALPs is found. We report upper limits on the ALP–photon coupling for masses between 1 and 1000 neV, derived under a specified model of the jet, Galactic and intergalactic magnetic fields, which are consistent with existing bounds in this range. We quantify the associated systematic uncertainties and discuss their implications for ALP searches towards variable gamma-ray sources.

Keywords: Axion-like particles, Blazars, Gamma-ray astronomy, Bayesian Blocks, CTA, MAGIC

1. Introduction

Axion-like particles (ALPs) are light spin-0 bosons predicted in various extensions of the Standard Model, typically arising from the spontaneous breaking of global symmetries. The first realization of this idea was the so-called *visible* axion, proposed in the 1970s (Peccei and Quinn, 1977; Wilczek, 1978; Weinberg, 1978) as a solution to the Strong CP Problem. Although this first prediction was ruled out by experiments, more general ALP scenarios have gained increasing attention since, and even more so recently, since ALPs could comprise a significant fraction—or even the entirety—of Dark Matter (DM) in the Universe (Arias et al., 2012), and may help explain anomalies in the expected absorption of astrophysical TeV gamma rays due to pair production with the Extragalactic Background Light (EBL) (see, e.g., Franceschini and Rodighiero, 2017), for instance from sources at $z \simeq 1$ (Galanti et al., 2023).

ALPs can be searched for experimentally thanks to their predicted couplings to photons, electrons, neutrons, protons, and quarks. These interactions are explored in many laboratories worldwide—both surface and underground—in accelerators, and through a wide range of astrophysical limits (Particle Data Group, 2024). A continuously updated collection of experimental constraints is maintained by O'Hare (2020)[1].

In this work, we investigate ALP–photon mixing in astrophysical magnetic fields. High-energy photons and strong magnetic fields are present in the ultra-relativistic jets of active galactic nuclei (AGNs), where the activity of a supermassive black hole (BH) generates a powerful ejection of collimated particles and radiation perpendicular to the accretion disk (Urry and Padovani, 1995). When a jet is oriented toward the Earth, relativistic beaming enhances the visibility of these distant objects. The collimation of the jet and the acceleration of particles within it are sustained by structured magnetic fields, whose exact nature remains under active investigation (Pudritz et al., 2012). Magnetic fields inside the jet are expected to be of order 10^{-3}– 1 G (see, e.g., Potter and Cotter, 2015). Magnetic fields are also observed around AGNs—for example, if the host galaxy lies within the core of a galaxy cluster (intra-cluster medium)—with strengths of order μG (see, e.g. Govoni and Feretti, 2004, for a review) (see, e.g., MAGIC,

[1]Web: `https://cajohare.github.io/AxionLimits/`

2024a, for an ALP oriented search). Intergalactic magnetic fields (IGMFs), the subject of intense ongoing research, are expected to have strengths between 10^{-18} G and 1 nG (Durrer and Neronov, 2013). Finally, μG magnetic fields are present in the Milky Way halo (see, e.g., Unger and Farrar, 2023). All these magnetic-field environments contribute, with different strengths, to photon–ALP mixing.

The photon–ALP conversion depends on the intensity and geometry of these fields with respect to the photon propagation direction, and is a function of the ALP mass m_a and coupling $g_{a\gamma}$. The interaction becomes efficient above a critical energy (Hooper and Serpico, 2007):

$$E_{\rm crit} \sim 2.5\ {\rm GeV}\ \frac{\left|m^2_{a,{\rm neV}} - \omega^2_{pl,{\rm neV}}\right|}{g_{11}\ B_{\mu{\rm G}}}, \qquad (1)$$

where $\omega_{pl,{\rm neV}}$ is the plasma frequency in units of neV, $B_{\mu{\rm G}}$ is the magnetic-field strength in μG, and $g_{11} = g_{a\gamma} \cdot 10^{11}$ GeV. For $E < E_{\rm crit}$ mixing is negligible; for $E \simeq E_{\rm crit}$ the system enters an oscillatory regime; and for $E > E_{\rm crit}$ the mixing approaches the strong regime. For ∼1 TeV photons—the energies investigated in this work—the critical energy corresponds to $m_a \sim$ 1–1000 neV.

At these photon energies, ALP–photon mixing competes with another process: the absorption of high-energy photons via pair production $\gamma\gamma \to e^+e^-$ on the diffuse EBL permeating the intergalactic medium (Franceschini and Rodighiero, 2017). The EBL consists of direct starlight and light reprocessed by interstellar dust, with a spectral distribution spanning UV to IR. The resulting optical depth $\tau(E, z_s)$ depends on the photon energy E and the source redshift z_s.

To date, searches for spectral irregularities consistent with ALP-induced modulations have been performed using H.E.S.S. observations of PKS 2155–304 (Abramowski et al., 2013), *Fermi*-LAT studies of NGC 1275 (Ajello et al., 2016), as well as MAGIC and VERITAS observations of the same source in the strongly magnetized environment of the Perseus cluster core (MAGIC, 2024a; Adams et al., 2025). A review of gamma-ray–based limits can be found in Batković et al. (2021). In particular, MAGIC has shown that, unless the spectrum is very finely sampled (as in the case of LAT), sensitivity is greater to sudden spectral modulations induced by ALP mixing than to smaller-scale wiggles (MAGIC, 2024a).

In this work, we search for spectral irregularities in the GeV–TeV gamma-ray spectra of four blazars: BL Lac, Mrk 421, Mrk 501, and 1ES 1959+650 observed with the Major Atmospheric Gamma-ray Imaging Cherenkov (MAGIC) telescopes from 2014 to 2024 and the first Large-Sized Telescope (LST-1) (Cortina, 2019), the first installment of the 13-telescope CTAO array at the Northern site (CTAO-N) (LST, 2023b), between 2020 and 2024, for a total of 980.7 h of data – the largest dataset assembled to date for a ground-based gamma-ray ALP search (Batković et al., 2021). MAGIC and LST-1 are instruments of the Imaging Atmospheric Cherenkov Telescope (IACT) class, that detect cosmic gamma-rays through the Cherenkov light generated within extended atmospheric particle showers (Chadwick, 2021; Mirzoyan, 2022). We distribute data over consecutive time-varying Bayesian Blocks (BBs, Scargle et al., 2013), and combine limits among BBs and across experiments, with a newly-introduced, tailored statistical framework.

This article is organized as follows. In Sect. 2 we report the methodology: in Sect. 2.1 we present the analyzed datasets, their basic properties (variability, etc.), and the experimental setups; Sect. 2.2 introduces photon–ALP mixing and its interplay with EBL absorption, while Sect. 2.3 details the magnetic-field environments traversed by the gamma-ray beam; Sect. 2.4 completes the signal model by combining these ingredients with the intrinsic astrophysical spectra into the observed flux; and in Sect. 2.5 we discuss the statistical framework used to infer ALP properties from wiggles in the gamma-ray spectra. Our results are shown in Sect. 3. We discuss our findings and conclude in Sect. 4. Appendices focus on technical aspects of the BB-sampling and study of systematics.

2. Methodology

In Sect. 2.1 we present our data preparation in Bayesian Blocks for the two instruments. Sect. 2.2 introduces photon–ALP mixing and its interplay with EBL absorption, while Sect. 2.3 details the magnetic-field environments traversed by the gamma-ray beam, which together determine the photon survival probability. Sect. 2.4 completes the signal model by specifying the intrinsic astrophysical spectra and combining them with this survival probability into the observed flux. Finally, Sect. 2.5 presents the statistical framework used to test the ALP hypothesis against the data.

2.1. Experimental Data Preparation

We analyzed four bright, well-studied blazars (see Tab. 1 for redshifts, exposure and global statistics): Mrk 421 and Mrk 501, two nearby TeV BL Lacs with frequent high states and hard spectra; 1ES 1959+650, an HBL known for intense flaring and "orphan" TeV activity; and BL Lacertae (BL Lac), the class prototype with a typically softer VHE spectrum. Throughout, we adopt the standard SED-based blazar taxonomy (Padovani et al., 2017): low-synchrotron-peaked (LSP) sources have a synchrotron peak frequency $\nu_s \lesssim 10^{14}$ Hz, intermediate-synchrotron-peaked (ISP) sources fall in $10^{14} \lesssim \nu_s \lesssim 10^{15}$ Hz, and high-synchrotron-peaked (HSP) sources have $\nu_s \gtrsim 10^{15}$ Hz. This classification correlates with the inverse-Compton peak: LSPs typically peak at MeV–GeV, whereas HSPs often extend into the GeV–TeV band, making them prime targets for IACT observations and for searches of ALP-induced spectral modulations. In our sample, Mrk 421, Mrk 501, and 1ES 1959+650 are HSPs, while BL Lac is an LSP.

The observations used in this study combine MAGIC data taken in years 2014-2024 and data from LST-1, overall between 2020–2024. We restrict our analysis to optimal quality data, that is data obtained under dark-sky conditions and with atmospheric transmission above 0.8. Tab. 1 summarizes

Target	Class	z	MAGIC				LST			
			Period	Duration	$\sigma_{\rm LM}$	BBs	Period	Duration	$\sigma_{\rm LM}$	BBs
Mrk 421	HSP	0.031	2014.12–2024.12	263.6 h	438.5	175	2020.12–2024.02	71.7 h	69.5	36
Mrk 501	HSP	0.034	2015.07–2024.09	235.0 h	85.2	55	2020.03–2023.06	55.7 h	17.1	29
BL Lac	LSP	0.069	2015.06–2024.11	90.9 h	101.9	43	2021.08–2022.11	31.8 h	31.6	17
1ES 1959+650	HSP	0.048	2015.03–2024.12	217.1 h	224.2	78	2020.08–2023.09	14.9 h	5.2	9

Table 1: Summary of MAGIC and LST observations. Classes: HSP/LSP denote High/Low Synchrotron Peak blazars. z is the redshift. $\sigma_{\rm LM}$ is the Test Statistic (global significance in standard deviations relative to the background-only hypothesis), computed following Li and Ma (1983). BBs is the number of Bayesian Blocks (see text).

the complete dataset along with some statistics. This comprises most of the data from the selected targets observed by each experiment, constituting the largest dataset analyzed to date, both for each target and for ground-based gamma-ray ALP searches (Batković et al., 2021). Data sets partially overlapping with those considered here have been used in previous astrophysical studies by the MAGIC and LST-1 collaborations, including, for MAGIC, (e.g. MAGIC, 2015a, 2018, 2020b, 2019, 2020a), and, for LST-1, (LST, 2025).

The data for each instrument are reconstructed up to DL3 (Data Level 3), following a uniform (non-tailored) pipeline, with the aim of generating an algorithm to extend ALP-oriented analysis to increasingly larger datasets. DL3 is a community data standard (gamma-astronomy data format; GADF) defined within the CTAO framework (Nigro et al., 2021; Gamma-ray astronomy community, 2018) and now increasingly adopted by MAGIC as well as by other gamma-ray instruments (H.E.S.S., VERITAS, HAWC, *Fermi*-LAT, etc.). Its adoption and tooling within MAGIC are detailed in MAGIC (2024c). DL3 files provide event lists together with the corresponding instrument response functions, enabling fully reproducible scientific analysis. These products are directly compatible with the open-source Python package `Gammapy` (Deil et al., 2017), which has become the reference tool for IACT data analysis.

In the present work, reconstruction is instrument-specific; the ALP analysis pipeline from DL3 onward is identical for MAGIC and LST-1.

MAGIC. MAGIC is an IACT system operating since 2003, but only since 2009 in stereoscopic mode (two telescopes) (MAGIC, 2016b,c). Our datasets for BL Lac, Mrk 421, Mrk 501, and 1ES 1959+650 encompass all stereoscopic observations with sufficient instrument stability to perform massive unsupervised reconstruction. This is done using the recent automatic reconstruction pipeline called `autoMAGIC`. This is an automated, database-driven workflow and open-source software presented in Schubert (2023); Walther et al. (2025). The pipeline automates the execution of the MAGIC Analysis and Reconstruction Software (MARS; Zanin et al., 2013) by wrapping each step in Python, tracking jobs in a relational database, and submitting them to the Port d'Informació Científica (PIC) cluster. This enables reproducible analyses and the production of standardized DL3 data products with minimal human intervention (see Appendix A for details).

Starting from single-telescope camera events, calibrated and image-cleaned, the analysis applies Hillas parameterization (Hillas, 1985) to each shower image. The pipeline then combines stereo information, matches the corresponding Monte Carlo response functions, and reconstructs the event energy, arrival direction, and time. A Random Forest classifier assigns a *gammaness* parameter, quantifying the probability that an event is a gamma-ray rather than hadronic background. The resulting DL3 event lists and instrument response functions are compliant with community standards (Gamma-ray astronomy community, 2018; MAGIC, 2024c). Before this step, the proprietary MAGIC `melibea` files are converted into DL3 format using the `magic-dl3` tool developed by the collaboration (MAGIC, 2024c). More details are provided in Appendix A.

LST-1. This is the first installation of the CTAO Northern Array (Cortina, 2019). The final array will comprise four LSTs, of 23 m diameter, and 9 Medium Size Telescope (MSTs) of 12 m diameter. We use all available LST-1 data after 2020 and the LST proprietary reconstruction pipeline `lstchain` (LST, 2023a), which is still under development. The algorithm is very similar to that presented for MAGIC, except for the stereoscopic combination, which is currently not applied since LST-1 operates as a single telescope. We make use of official DL3-releases from the LST collaboration. The format follows the same GADF DL3 conventions (Gamma-ray astronomy community, 2018) and `Gammapy` can be agnostic as to the specific origin of DL3 as long as they are correctly formatted. A subset of these data is studied for astrophysical interpretation in LST (2025).

2.1.1. *Bayesian Blocks definition*

Blazars are highly variable sources, typically observed in a baseline (low) state but capable of sudden *flares* with flux increases by factors of 10–100 that last from minutes to days, sometimes preceded or followed by pre-/post-flare activity (see, e.g., Peñil et al., 2024, for a recent discussion). During flares, a blazar can outshine the Crab Nebula, the standard candle of high-energy gamma-ray astronomy; fluxes are therefore often expressed in Crab Units (C.U.)[2] above 300 GeV. Targets are observed in a sparse cadence, with intensive campaigns over consecutive nights interleaved with periods of no observations (e.g., during flares or multi-instrument campaigns). As a result, the datasets are biased toward specific source states. The TeV portion of a blazar's gamma-ray spectral energy distribution (SED) typically exhibits a smooth yet steep profile (Gréaux

[2] 1 C.U. $\equiv F_E(> 300\ \mathrm{GeV}) \simeq 7 \cdot 10^{-11}$ erg cm^{-2} s^{-1} (MAGIC, 2015b)

et al., 2023) that can be modeled with different functions, as described below. The optimal spectral function can vary between targets and between different *states* of the same target. Consequently, time binning plays a central role.

In general, data are obtained in *runs* (both MAGIC and LST runs are ∼20 min) or grouped *night-wise*. In the following, we operate on a list of run-wise data units to improve the granularity of our dataset. Because we are particularly interested in spectral features superimposed on a smooth astrophysical continuum, we use an adaptive binning scheme based on Bayesian Blocks (BBs) that accounts for both source activity and mean spectral hardness. We used the `measures` option in the `astropy` implementation. The algorithm groups consecutive data into blocks, considering both data uncertainties and time separation, controlling the departure from a 'constant flux' hypothesis. The penalty term is controlled by a user parameter p_0 that represents the false positive rate for introducing change points. In this work we set $p_0 = 0.0027$, corresponding to the two-sided 3σ tail probability of a Gaussian distribution. In order to better compare BBs across targets, datasets and experiments, we define a common energy-binning scheme. We tested different binning choices, e.g. 5, 10, 15 logarithmically spaced bins per decade, and selected 10 log-spaced bins per decade, as the results are essentially insensitive to this choice. In addition, we split BBs spanning over long periods to avoid individual BBs to include extended gaps without observations. See Sect. B.1 for more details on the BB methodology.

We identify the following numbers of Bayesian Blocks (BBs) and exposure per source (see Tab. 1). Mrk 421: 211 BBs (MAGIC 175 / 263.6 h; LST-1 36 / 71.7 h), total 335.3 h; Mrk 501: 84 BBs (55 / 235.0 h; 29 / 55.7 h), total 290.7 h; BL Lac: 60 BBs (43 / 90.9 h; 17 / 31.8 h), total 122.7 h;; 1ES 1959+650: 87 BBs (78 / 217.1 h; 9 / 14.9 h), total 232.0 h. Altogether, the dataset amounts to 980.7 h. Each BB is treated as an independent analysis unit: For each BB we select the best-fitting spectral function among EPWL and LP (see Eq. 3 below), accounting for absorption by the extragalactic background light (EBL). Depending on the source state, different models may be preferred. The resulting best-fit SED parameters are treated as nuisance parameters in the maximum-likelihood search for ALP signatures.

2.2. ALP propagation and EBL

Photon–ALP propagation in an external magnetic field is described by a Schrödinger-like evolution for a three-state system (two photon polarizations plus one ALP state) (Raffelt and Stodolsky, 1988). The off-diagonal term responsible for mixing is proportional to $g_{a\gamma}B$, while the diagonal terms include the ALP mass contribution ($\propto m_a^2/2E$) and photon dispersion in the medium (plasma frequency and QED vacuum-polarization effects). The competition among these terms introduces an energy-dependent mixing angle and oscillation wavenumber, leading to an *oscillatory* photon survival/conversion probability.

The transition between regimes is conveniently characterized by the critical energy introduced in Eq. 1: for $E \ll E_{\rm crit}$ mixing is suppressed; for $E \simeq E_{\rm crit}$ the system develops energy-dependent oscillations; and for $E \gg E_{\rm crit}$ mixing approaches the strong regime, where the beam tends toward equipartition between photons and ALPs. The oscillations when $E \simeq E_{\rm crit}$ imprint wiggle-like deviations on top of the smooth intrinsic blazar spectra and motivate our spectral search strategy. Photon survival probability computed for the 4 sources under scrutiny, and for different oscillatory regimes, are shown in Fig. 3.

In realistic lines of sight, the beam traverses multiple magnetic environments with different field strengths and orientations (jet, host galaxy/cluster, intergalactic medium, and the Milky Way). We compute the total photon survival probability $P_{\gamma\gamma}(E)$ numerically by propagating through these domains and including EBL absorption, using the open-source package `gammaALPs` (Meyer et al., 2021b,a) with the magnetic-field and EBL models specified in the next section.

Photon–ALP oscillations do not act in isolation: TeV gamma rays from cosmological distances are simultaneously attenuated by pair production on the Extragalactic Background Light (EBL), the diffuse radiation field of starlight and dust-reprocessed infrared emission that fills the intergalactic medium. The two effects combine multiplicatively in the photon survival probability (Eq. 4): EBL absorption imprints a smooth, energy- and redshift-dependent suppression of the spectrum, on top of which ALP mixing can superimpose the oscillatory features that are the target of this search. We model the EBL opacity using Domínguez et al. (2011), as implemented in `gammaALPs`. A more recent EBL model has since been proposed (Banerjee et al., 2026), but is not yet implemented in `gammaALPs`. Given the relatively low redshifts of our targets ($z \lesssim 0.07$; see Tab. 1), where EBL attenuation is modest and differences between models are correspondingly small, we do not expect this choice to have a sizeable impact on our results.

2.3. Magnetic fields along the line of sight

Modeling the magnetic fields from the emission region to Earth is essential. While the propagation of gamma rays through astrophysical environments is not directly affected by magnetic fields, they become crucial in the presence of ALPs, as they mediate photon–ALP conversion seen in Sect. 2.2. This makes a detailed description of the magnetic structures encountered along the line of sight necessary.

2.3.1. Ultra-relativistic jet magnetic field

For blazars – whose relativistic jets are closely aligned with our line of sight – the magnetic field at the source plays a particularly significant role, and is expected to dominate photon–ALP mixing over the other propagation environments (host galaxy, intergalactic medium, Milky Way) discussed below.

In our study, we adopt the model developed by Davies et al. (2021), based on the jet structure proposed by Potter and Cotter (2012a,b, 2013a,b, 2015) (PC), which assumes a parabolic, accelerating base near the BH that transitions into a conical, slowly decelerating jet at larger distances (see Fig. 1). The transition region, located at $r_T = r_0$, is also the point of equipartition between magnetic and particle energy densities, and is identified with the very-high-energy emission region, $r_T = r_{VHE}$.

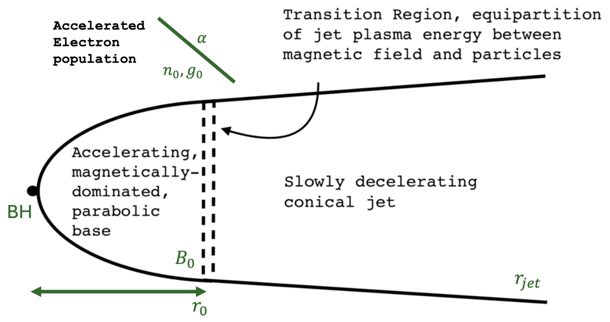

Figure 1: Schematic structure of a relativistic jet (not to scale). The innermost region near the BH is magnetically dominated and parabolic, where the plasma is accelerated. At the transition region, located at a distance x_T from the BH, the jet plasma reaches equipartition between magnetic and particle energy densities, with bulk Lorentz factor $\gamma_T = \gamma_{\rm VHE}$ and magnetic field strength of order G. Beyond this point, the flow expands into a slowly decelerating conical jet. Adapted from Davies et al. (2021).

Jet parameters for each target, derived from SED modeling by Potter and Cotter (2015) and Tavecchio et al. (2010), are reported in Tab. 2. The field is generally described as helical, with a poloidal component dominating the parabolic base and a toroidal component dominating the conical regime. Superimposed on the ordered helical field, the model includes a tangled component. We assign a fraction $f_t = 0.7$ of the magnetic energy density to this component and set its coherence length equal to the local jet width (the `jetwidth` option of `gammaALPs`), so that it varies along the jet rather than being held fixed; the helical-to-toroidal transition radius is $r_T = r_0$ and therefore differs between sources, as listed in Tab. 2. Most of the magnetic energy in our jet field is therefore in a disordered component. This is the value adopted by Davies et al. (2021) in their study of Mrk 501, and the range $f_t = 0$–0.7 that they consider is itself taken from Murphy et al. (2013), who fit the tangled fraction to rotation-measure maps of that source. The degree of disorder of the field is thus set by radio polarimetry rather than assumed, and our choice lies at the most disordered end of what those data allow.

Target	Mrk 421	Mrk 501	BL Lac	1ES 1959+650
r_0 [pc]	6.02	0.3	0.12	0.10
B_0 [G]	0.03	0.8	2.68	1.88
n_0 [cm^{-3}]	$8.5 \cdot 10^3$	$4.5 \cdot 10^4$	$8 \cdot 10^5$	$7 \cdot 10^2$
r_{jet} [10^3 pc]	9.72	3.24	16.20	32.41
g_{min}	9	2	5	4
g_{max}	12	9	8	8
α	1.55	1.68	1.95	1.60

Table 2: Parameters of the jet magnetic field model; radius r_0 where B field is equal to B_0, electron density n_0 at r_0, jet length r_{jet}, jet Lorentz factors g_{min} and g_{max} at r_{jet} and r_0, respectively, power-law index of electron energy distribution function α, radius at which helical field becomes toroidal r_T and distance of gamma-ray emission region from BH r_{VHE}. In this work we assume $r_t = r_{VHE} = r_0$. Values are taken from Potter and Cotter (2015), with exception of n_0 which are provided in Tavecchio et al. (2010).

This jet magnetic-field model is, among all propagation environments considered in this work, the one with the largest impact on the derived ALP constraints, and it is subject to significant theoretical and observational uncertainty.

Two features of this model are worth stressing, since they limit the freedom available in the description of the field. First, the PC model is extended and inhomogeneous rather than one-zone: the emission is computed by integrating along the whole jet and fitted to the radio data simultaneously with all other wavelengths, so that the optically-thick-to-thin synchrotron break constrains the radius of the brightest synchrotron-emitting region, identified with r_0 (Potter and Cotter, 2015). Within this framework r_0 is therefore measured rather than assumed. Second, what governs photon–ALP mixing is, in general, the integral of the magnetic field across the propagation length (e.g. Marsh et al., 2022). Thus ALP-photon mixing does not depend on B_0 or r_0 separately but on their product: for a toroidal field the mixing integral is

$$\int B\,\mathrm{d}l \;=\; B_0\, r_0\, \ln\!\left(\frac{r_{jet}}{r_0}\right), \tag{2}$$

and the invariant $B_0\, r_0$ is the quantity probed by VLBI core-shift observations, with which our adopted parameters are compatible. The residual uncertainty therefore concerns the field *structure* — the balance between ordered and tangled components along the escape path — rather than its overall normalisation. We discuss both points, and the comparison with analyses based on one-zone modelling, in Sect. C.2.

We further assume a static, large-scale field configuration throughout each source's observation period. This choice is motivated by the fact that photon–ALP mixing accumulates over the kpc-scale path traversed by gamma rays within the jet before they escape into the intergalactic medium, and is therefore governed by the large-scale, ordered field structure rather than by the sub-parsec perturbations thought to drive individual flares. A fully time-dependent treatment, linking the field configuration to the flux state of each Bayesian Block, would in principle be preferable, but requires quasi-simultaneous multiwavelength SED coverage for each state, which is available only for limited periods of our targets (see Sect. C.2). We discuss the validity of this assumption, and quantify the impact of the resulting uncertainty on our results, in Sect. C.2.

We finally point out that, while spectral irregularities are generally dependent on the chosen magnetic-field model, a spectral hardening at TeV energies is a universal effect of cosmological ALP-photon conversions, which would represent a clear hint of the existence of ALPs if detected (Sanchez-Conde et al., 2009).

2.3.2. *Magnetic field in the host galaxy cluster and the intergalactic medium*

There is no evidence that any of the targets in this work reside in gas-rich galaxy clusters hosting strong magnetic fields. The magnetic field associated with the relativistic jet is therefore expected to be the dominant factor governing photon–ALP conversion, and we neglect any contribution from the intracluster medium hereafter. This assumption is also adopted by most previous ALP studies including these sources, disregarding the possible impact of the host cluster magnetic field on photon–ALP propagation (Li et al., 2021; Gao et al., 2024; Li et al., 2025, 2024).

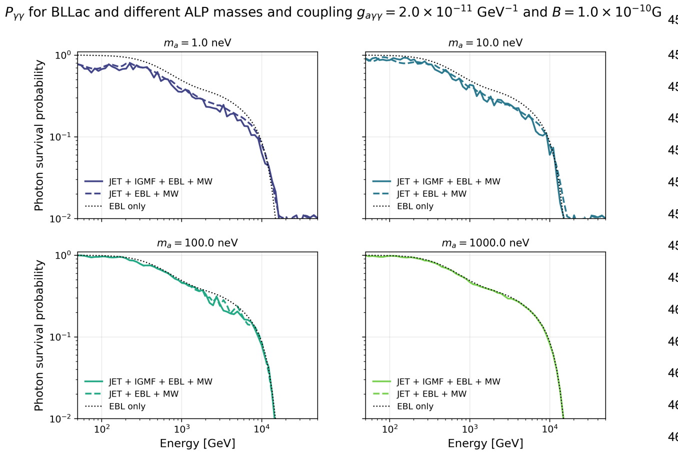

Figure 2: Photon survival probability in the case of BL Lac, assuming $g_{a\gamma\gamma} = 2.0 \cdot 10^{-11}\,\mathrm{GeV}^{-1}$, four different ALP masses and three magnetic-field configurations. The solid lines correspond to photon–ALP mixing in the jet, the IGMF, and the Milky Way, where EBL attenuation is included in the IGMF modeling. The dashed lines represent the case including the jet magnetic field, EBL attenuation, and the Milky Way magnetic field, but without the IGMF contribution. For comparison, the finely dashed line shows the photon survival probability considering EBL attenuation only.

Current observational constraints place lower limits on the strength of the IGMF 10^{-18} G and upper limits around 10^{-9} G inferred from cosmic microwave background (CMB) anisotropies and Faraday rotation measurements (Durrer and Neronov, 2013; Ade et al., 2016; Alves Batista and Saveliev, 2021).

To assess the possible impact of the IGMF on photon-ALP mixing for BL Lac, we perform a test assuming a cell-like structure of the IGMF with present-day coherence length $l_0 = 1$ Mpc and intensity $B_0 = 10^{-10}$ G, compatible with current constraints (e.g. Durrer and Neronov, 2013; Abdalla et al., 2021). Several combinations of m_a and $g_{a\gamma\gamma}$ are considered and compared with $p_{\gamma\gamma}$ obtained when only EBL attenuation is included in the calculation (see Fig. 2). One can see that the IGMF introduces some additional spectral irregularities in the 1 neV range, while these effects are subleading in comparison to those introduced by the magnetic fields of the jet and the Milky Way (see Fig. 2) for higher masses. We have tested that our results are not significantly affected by the IGMF above 10 neV. However, further exclusion curves appear between 1 and 10 neV. Because the value of IGMF is ultimately not known, we prefer not to include this contribution.

This finding aligns with previous studies (e.g. Ajello et al., 2016; Davies et al., 2021, 2023; MAGIC, 2024b; Gao et al., 2024), which similarly report that IGMFs of such strength or weaker have a minimal effect on photon-ALP conversions in the considered part of ALP parameter space. The IGMF constitutes the dominant astrophysical uncertainty in extracting ALPs limits from VHE gamma-ray observations of distant sources such as GRB 221009A (Han et al., 2026) or PKS 2155−304 (Wouters and Brun, 2012; Abramowski et al., 2013), as its strength, coherence length, and stochastic properties can substantially shift the derived exclusion contours.

2.3.3. *Magnetic field of the Milky Way*

Following the studies performed up to date, we use the Galactic Magnetic Field (GMF) model with a turbulent component and a regular component. In several previous works (MAGIC, 2024a; Ajello et al., 2016), the model developed by Jansson and Farrar (2012b,a, JF12) was used. Recently some of those authors compiled a new model based on more recent data (Unger and Farrar, 2023, UF23). In fact, UF23 model includes eight distinct realizations, all of which are compatible with current Faraday rotation and synchrotron observations. These realizations were systematically compared by Landgraf (2024) to JF12 and by Pshirkov et al. (2011) in terms of their impact on photon–ALP propagation and conversion probabilities. Since all UF23 variants are consistent with observational constraints, one cannot *a priori* state which is the "correct" model. However, for studies aiming at continuity with previous earlier GMF-based ALP analyses, Landgraf (2024) notes that the neCL realization provides the most consistent comparison, whereas the base and expX models can be regarded as providing an estimate of the maximal systematic range.

In this work we therefore make use of the neCL model, to better compare with earlier results, and study the effect of uncertainties on the GMF by bracketing them with the base and expX models, in Sect. C.3.

2.4. *Signal model*

The signal model is determined by the intrinsic astrophysical emission, discussed in Sect. 2.4.1, combined with the photon survival probability introduced in Sect. 2.2 and Sect. 2.3 above, which encodes the effects of EBL absorption, photon–ALP mixing, and the magnetic-field environments traversed by the beam. Sect. 2.4.2 brings these ingredients together into the observed differential flux.

2.4.1. *Astrophysical spectra of targets*

We model the intrinsic TeV spectra of the blazar targets with one of the following functions:

$$\begin{aligned}
\mathrm{EPWL}: \quad \phi(E) &= \phi_0 \left(\frac{E}{E_0}\right)^{-\Gamma} \exp\!\left(-\frac{E}{E_c}\right), \\
\mathrm{LP}: \quad \phi(E) &= \phi_0 \left(\frac{E}{E_0}\right)^{-\alpha-\beta\log(E/E_0)} .
\end{aligned} \tag{3}$$

Here, ϕ_0 is the normalization at reference energy E_0, Γ the photon index, and E_c the cutoff energy for the EPWL model. For the log-parabola model, the photon index varies with energy according to $\alpha + \beta \log(E/E_0)$.[3]

For each BB for any of the targets in Tab. 1, the intrinsic spectra of Eq. 3 are tested during the regression with unbound spectral parameters. We select the spectral fit with the best χ^2 after including EBL absorption and ALP mixing, as discussed in the next section.

[3]We remark that a widely used fit is the simple power-law (PL): $\phi(E) = \phi_0\,(E/E_0)^{-\Gamma}$. We did not include it because PL is a special case of of EPWL with $E_c = \infty$ or LP with $\beta = 0$

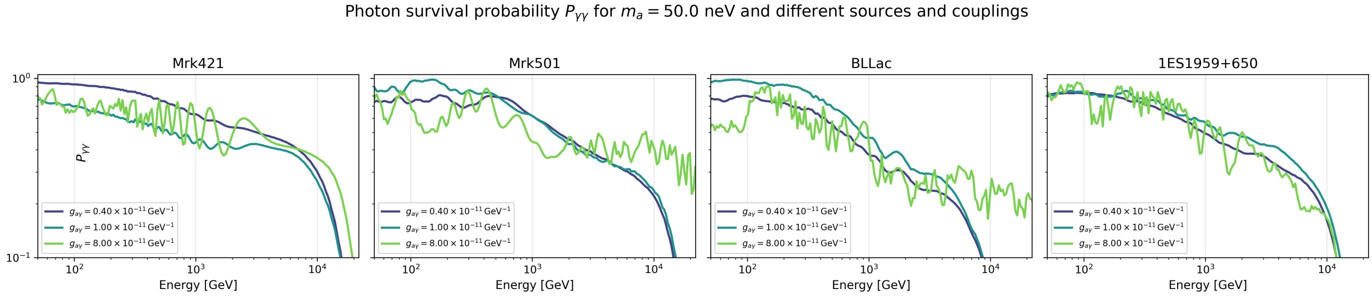

Figure 3: Photon survival probability for the four sources considered in this work, shown for $m_a = 50\,\mathrm{neV}$ and three values of the ALP-photon coupling constant $g_{a\gamma}$. From left to right, the chosen $g_{a\gamma}$ values represent: (i) a point where ALP-induced effects are barely noticeable, (ii) a point where the effects become significant, and (iii) a point where strong modulations appear, representative of the region that is most strongly excluded by our analysis. The comparison illustrates both the dependence of the survival probability on the ALP parameters and the differences arising from the intrinsic jet magnetic-field modeling and source distance.

2.4.2. *Photon survival probability and observed flux*

The observed differential flux is related to the intrinsic source emission through the photon survival probability:

$$\frac{d\Phi_{\rm est}}{dE} = \frac{d\Phi_{\rm int}}{dE} \cdot P_{\gamma\gamma}\big(E_\gamma; m_a, g_{a\gamma}, B, z\big), \qquad (4)$$

where $P_{\gamma\gamma}$ accounts for both ALP–photon mixing along the line of sight and absorption due to gamma rays scattering on EBL photons. Here, B denotes the set of magnetic-field environments traversed by the beam.

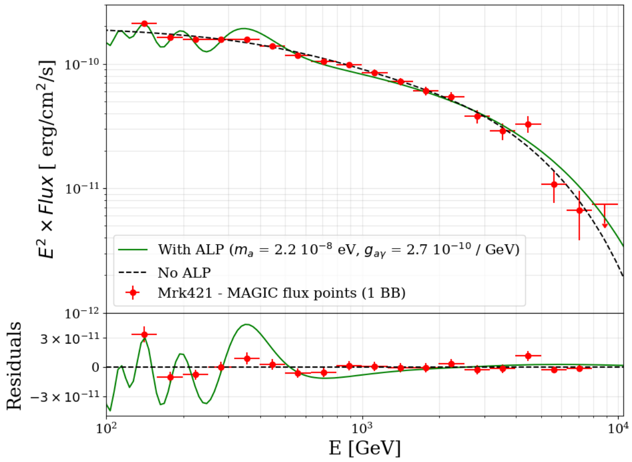

Figure 4: Spectral energy distribution of a representative Bayesian Block from the MAGIC observations of Mrk 421. The red points show the reconstructed flux measurements with their statistical uncertainties. The black dashed line represents the best-fit intrinsic spectral model of Eq. (3) after EBL attenuation, assuming the EBL model of Dominguez et al. (2011) and no ALPs. The green curve shows the corresponding observed spectrum of Eq. (4) including photon–ALP mixing for the benchmark point $(m_a,\, g_{a\gamma}) = (2.2\ 10^{-8}\ \mathrm{eV},\ 2.7\ 10^{-10}\ /\mathrm{GeV})$, corresponding to the ALP hypothesis yielding the largest discrepancy with respect to the null hypothesis for this Bayesian Block. The lower panel displays the residuals with respect to the null-hypothesis model.

The spectral energy distribution (SED) is shown in Fig. 4 for a representative BB from the MAGIC observation of Mrk 421.

2.5. *Statistical Analysis*

Our analysis aims to determine whether ALP signatures are present in the observed gamma-ray spectra. To investigate these potential signatures, for each i-th data block (see Sect. 2.1.1) with intrinsic SED described by a parametric model with parameters $\boldsymbol{\mu}_i$, we build the Poisson ON/OFF likelihood

$$\mathcal{L}_i(g_{a\gamma}, m_a; \boldsymbol{\mu}_i) = \prod_{j=1}^{N_i} \left[\frac{(s_{ij} + \alpha_i b_{ij})^{n_{ij}}}{n_{ij}!}\, e^{-(s_{ij}+\alpha_i b_{ij})} \times \frac{b_{ij}^{m_{ij}}}{m_{ij}!}\, e^{-b_{ij}} \right], \qquad (5)$$

which is, by definition, the probability of observing the counts n_{ij} (ON) and m_{ij} (OFF) in the energy bin j of block i when the expected signal and background count are s_{ij} and b_{ij} respectively. Here N_i is the number of energy bins for each i-th block and α_i is the exposure ratio (ON/OFF). The dependence on $(g_{a\gamma}, m_a)$ and on the SED parameters $\boldsymbol{\mu}_i$ enters via the signal expectation

$$s_{ik} = \int_{\Delta E_j} dE \int dE'\ \Phi^i_{\rm int}(E'; \boldsymbol{\mu}_i)\, P_{\gamma\gamma}(E'; g_{a\gamma}, m_a)\, \mathrm{IRF}^i(E \mid E'), \qquad (6)$$

where, in order of appearance:

- $\Phi^i_{\rm int}(E'; \boldsymbol{\mu}_i)$ is the intrinsic energy spectrum (see Sect. 2.4.1);
- $P_{\gamma\gamma}(E'; g_{a\gamma}, m_a)$ is the photon survival probability that includes the EBL attenuation and photon–ALP mixing (see Sect. 2.2);
- $\mathrm{IRF}^i(E \mid E') = A^i(E')\, G^i(E \mid E')$ is the instrument response for block i, with effective area A^i and $G^i(E|E')$ is the probability density for a gamma ray with true energy E' to be assigned an estimated energy E.

The likelihood in Eq. 5 reaches its maximum value $\hat{\mathcal{L}}_i$ for $s_{ij} = n_{ij} - \alpha_i m_{ij}$ and $b_{ij} = m_{ij}$, while for any *given* s_{ij}, it is maximized by $b_{ij} = \hat{b}_{ij}(s_{ij})$.[4]

[4]Dropping the indices i, j for readability, one finds Cowan et al. (2011); D'Amico (2022)

$$\hat{b}(s) = \frac{n + m - (1 + 1/\alpha)s + \sqrt{(n + m - (1 + 1/\alpha)s)^2 + 4(1 + 1/\alpha)s\, m}}{2(1 + \alpha)}\ .$$

Putting $\boldsymbol{\mu}_i$ equal to $\hat{\boldsymbol{\mu}}_i(g_{a\gamma}, m_a)$, i.e., the SED parameters that maximize $\mathcal{L}_i$ at fixed $(g_{a\gamma}, m_a)$,[5] we can define the following variable (sometimes referred to as `Wstat` (Knödlseder et al., 2019))

$$W_i(g_{a\gamma}, m_a) \equiv -2 \log \mathcal{L}_i(g_{a\gamma}, m_a; \hat{\boldsymbol{\mu}}_i) + 2 \log \hat{\mathcal{L}}_i = \\ = 2 \sum_j^{N_i} (s_{ij} + (1+\alpha_i)\hat{b}_{ij} - n_{ij} \log(s_{ij} + \alpha_i \hat{b}_{ij}) \\ - m_{ij} \log(\hat{b}_{ij}) - C) \quad (7)$$

with

$$C = n_{ij} + m_{ij} - n_{ij} \log(n_{ij}) - m_{ij} \log(m_{ij}) \quad (8)$$

The statistics defined in Eq. 7 has the advantage of being asymptotically a $\chi^2_{d_i}$ variable with effective number d_i of degrees of freedom (d.o.f.) approximately

$$d_i \approx N_i - p_i, \qquad p_i = 3, \quad (9)$$

where p_i is the number of free SED parameters fitted in block i (e.g. $\{\phi_0, \Gamma, E_c\}$ for EPWL or $\{\phi_0, \alpha, \beta\}$ for LP; see Eq. 3). In practice, low counts, empty or nearly empty bins can induce non-negligible deviations from the assumption in Eq. 9. Therefore we calibrate the d.o.f. d_i of W_i for each BB over the $(g_{a\gamma}, m_a)$ plane by dedicated Monte Carlo simulations; details of this calibration are provided in Sect. B.2.

After computing the block-wise profile statistics $W_i(g_{a\gamma}, m_a)$, we form the global test statistic by summation over all data blocks:

$$W(g_{a\gamma}, m_a) = \sum_i W_i(g_{a\gamma}, m_a), \quad (10)$$

which can be treated as χ^2_d-distributed with number of d.o.f.,

$$d(g_{a\gamma}, m_a) = \sum_i d_i(g_{a\gamma}, m_a). \quad (11)$$

A given point in the ALP parameter space $(g_{a\gamma}, m_a)$ can be excluded with a statistical significance quantified by the p-value

$$p(g_{a\gamma}, m_a) = Q\left(\frac{d(g_{a\gamma}, m_a)}{2}, \frac{W(g_{a\gamma}, m_a)}{2}\right), \quad (12)$$

where

$$Q(s, x) = \frac{1}{\Gamma(s)} \int_x^\infty t^{s-1} e^{-t} dt \quad (13)$$

is the *regularized upper incomplete gamma function* [6]. Smaller p-values indicate stronger exclusion of the corresponding ALP parameters. This p-value can be directly converted into an equivalent Gaussian significance σ, defined as the number of standard deviations corresponding to the same tail probability in a standard normal distribution, through the inverse of the error function: $\sigma = \sqrt{2}\mathrm{erf}^{-1}(1 - \mathrm{p})$.

[5]The values $\hat{\boldsymbol{\mu}}_i(g_{a\gamma}, m_a)$ are obtained by fitting the data with `gammapy` Deil et al. (2017), where the photon survival probabilities are computed with `gammaALPs` Meyer et al. (2021b,a).

[6]This function is implemented in the Python library `scipy` as `gammaincc`.

An important advantage of this χ^2-block-wise approach is its modularity. For each data block, one only needs to compute and store the values of $W_i(g_{a\gamma}, m_a)$ and the corresponding calibrated degrees of freedom $d_i(g_{a\gamma}, m_a)$ over a grid in the ALP parameter space *after the marginalization (or profiling) over all nuisance parameters*, such as the magnetic-field configuration. Once these nuisance parameters have been treated at the single-block level, each block provides a self-contained statistical contribution that depends only on $(g_{a\gamma}, m_a)$. This block design, together with the likelihood analysis, allows for great flexibility and scalability: if additional data blocks become available (e.g., from new sources or future observations), they can be seamlessly incorporated into the analysis by computing their individual W_i and d_i values and summing them into the global test statistic W and total degrees of freedom d as in Eq. 10 and Eq. 11, respectively. This avoids the need to recompute likelihoods from scratch for the entire dataset, enabling efficient updates and iterative refinement of the exclusion limits.

3. Results and Discussion

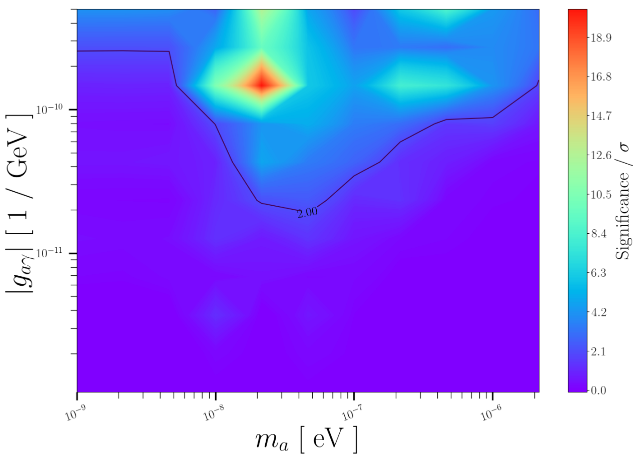

Figure 5: Exclusion significance, expressed in Gaussian standard deviations, over the ALP parameter space obtained by combining all MAGIC and LST-1 observations listed in Tab. 1. The solid black line marks the 95% confidence-level exclusion contour. The significance is evaluated on a grid of 121 ALP parameter points, while the color map is obtained with `Matplotlib contourf` function by interpolation between the sampled points.

In Fig. 5 we show the exclusion significance, expressed in Gaussian standard deviations σ, over the ALP parameter space explored in this work. The result is obtained by combining all MAGIC and LST-1 observations listed in Tab. 1, corresponding to the four blazars considered in this analysis and a total exposure of 980.7 h. The solid black contour indicates the boundary of the 95% confidence-level exclusion region. The scan covers ALP masses in the range $m_a \in [10^{-9}, 2\times10^{-6}]$ eV and photon–ALP couplings $g_{a\gamma} \in [10^{-12}, 5\times10^{-10}]$ GeV^{-1}. From Fig. 5 we can see that ALP models with $m_a \in [10^{-8}, 3\times10^{-7}]$ eV are excluded with large significance down to interaction strength $g_{a\gamma} \geq 2\times10^{-11}$ GeV^{-1}.

It is instructive to look at how this global result is built up from the individual data blocks. Fig. 6 shows the contribu-

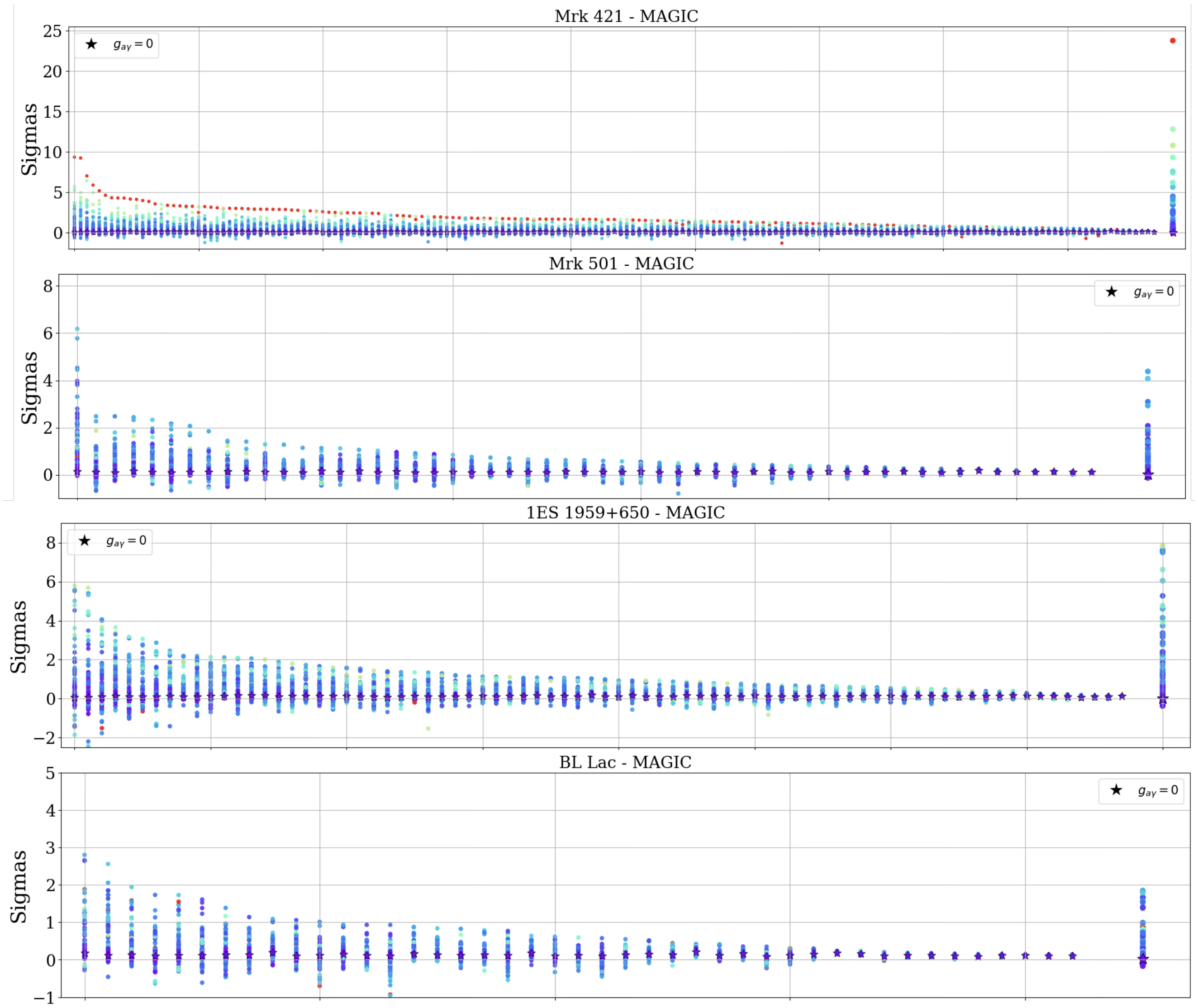

Figure 6: Contribution of the individual Bayesian Blocks (BBs) to the ALP exclusion obtained with the MAGIC dataset. Each panel corresponds to one source. Along the x axis, the BBs are ordered by decreasing constraining power, quantified as the maximum exclusion significance over the 121 sampled points in the $(m_a, g_{a\gamma})$ parameter space. The y axis shows, for each BB, the exclusion significance (in Gaussian standard deviations) of each individual ALP hypothesis. The color of each point identifies the corresponding ALP parameter point and follows the same colormap used in Fig.5: blue points correspond to ALP hypotheses not excluded in the combined analysis, orange points lie within the $\sim 2\sigma$ exclusion region, and red points correspond to the most strongly excluded hypotheses. The black star denotes the no-ALP hypothesis ($g_{a\gamma} = 0$). The rightmost column shows the exclusion significance obtained after combining all BBs of the corresponding source.

tion of each BB separately for the MAGIC datasets, with one panel per source. Every vertical column of points corresponds to a single BB and contains 121 points, one for each ALP hypothesis of the grid, whose colour identifies the corresponding $(m_a, g_{a\gamma})$ pair following the same colour scale as Fig. 5; the vertical axis gives the exclusion significance in Gaussian standard deviations, and the black star marks the no-ALP hypothesis $g_{a\gamma} = 0$. The columns are ordered along the horizontal axis by decreasing constraining power of the block, defined as the largest significance that block reaches over the whole grid, so that the leftmost column is the single most constraining BB of that source. The rightmost, detached column of each panel is the result of combining all blocks of that source through Eq. 10.

Fig. 6 shows a steep distribution in all four cases. The most constraining block reaches $\sim 9\sigma$ for Mrk 421 and $\sim 6\sigma$ for both Mrk 501 and 1ES 1959+650, after which the significance decreases rapidly, with the majority of blocks showing little or no exclusion for either the ALP or no-ALP hypotheses (black star). For Mrk 421, the combined significance reaches $\sim 24\sigma$, substantially exceeding that of any individual block. This demonstrates that, for this source, the block-wise combination in Eq. 10 genuinely accumulates information rather than simply inheriting the constraint from the single most constraining block. For the weaker sources, the improvement from combining blocks is much more limited, or even absent [7]

To better understand how the different instruments, sources,

[7] This is expected: the per-block maximum is taken independently over the

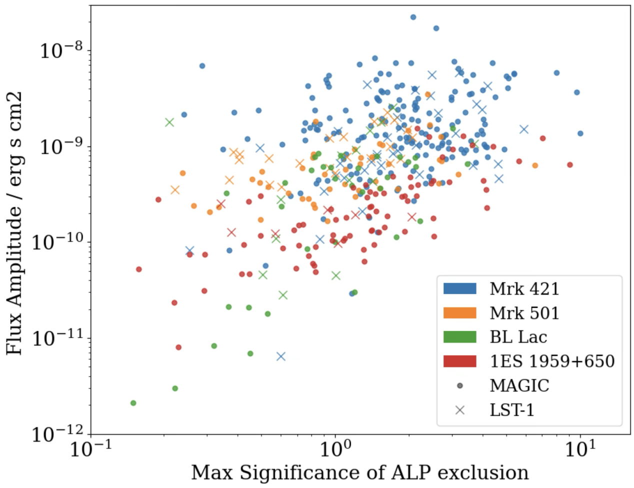

Figure 7: Distribution of the Bayesian Blocks (BBs) used in this analysis as a function of their intrinsic flux normalization and their constraining power on the ALP parameter space. Each point corresponds to one BB. The vertical axis shows the best-fit intrinsic flux normalization ϕ_0 (at the reference energy $E_{\rm ref}$ = 0.5 TeV) obtained by fitting the data with the intrinsic spectral model including only EBL attenuation. The horizontal axis reports the maximum exclusion significance, expressed in Gaussian standard deviations, reached by that BB over the explored ALP parameter space. Marker colors identify the source, while circles and crosses denote MAGIC and LST-1 observations, respectively.

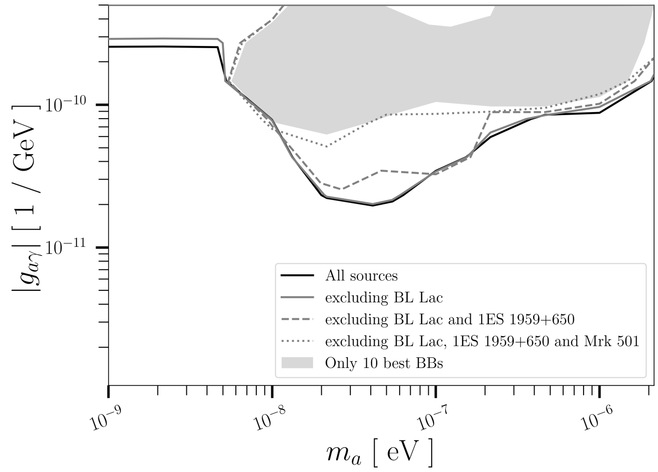

Figure 8: Comparison of the 95% CL exclusion obtained with different subsets of the dataset. The solid black curve shows the exclusion derived using all MAGIC and LST-1 observations. The gray curves show the result after progressively removing the least constraining sources, ultimately leaving only Mrk 421. The shaded gray region corresponds to the exclusion obtained using only the ten most constraining Bayesian Blocks of Mrk 421. These ten BBs provide a good approximation of the full Mrk 421 exclusion and are therefore used in the systematic studies presented in the appendix, where repeating the full analysis on all 442 BBs would be computationally prohibitive.

and observing periods contribute to the combined exclusion shown in Fig. 5, we investigate the constraining power of the individual Bayesian Blocks (BBs). For each BB we define its sensitivity as the maximum exclusion significance reached over the explored ALP parameter space, and compare it with the best-fit intrinsic flux normalization ϕ_0 (at the reference energy $E_{\rm ref}$ = 0.5 TeV) obtained from the preferred intrinsic spectral model (EPWL or LP) including only EBL attenuation. The result is shown in Fig. 7, where circles and crosses denote MAGIC and LST-1 observations, respectively, while colors identify the four blazar targets.

A clear positive correlation is observed between the intrinsic flux normalization and the maximum ALP exclusion significance. This is explained by the fact that brighter emission states provide larger photon statistics and, therefore, more precise spectral measurements (see Fig. 4), that in turn provide stronger sensitivity to ALP signatures. Consequently, the overall exclusion is not driven uniformly by the 442 BBs considered in this work (see Tab. 1), but is instead dominated by a relatively small number of high-flux BBs. These typically correspond to bright flaring episodes, which individually provide substantially greater constraining power than the majority of lower-flux observations.

To further quantify the relative contribution of the different targets to the final exclusion, we repeated the analysis while progressively removing sources from the combined dataset. Starting from the full dataset, we first removed the least constraining source (BL Lac), followed by 1ES 1959+650, and finally Mrk 501, leaving only Mrk 421, which is the most sensitive target in this analysis. We additionally performed the analysis using only the ten most constraining Bayesian Blocks (BBs) of Mrk 421. The resulting exclusions are shown in **??**. Three features are worth noting. First, discarding BL Lac leaves the limit essentially unchanged over the whole mass range: despite contributing 60 BBs and 122.7 h, this target adds almost nothing to the combination, consistent with its softer VHE spectrum and lower flux states (see Fig. 7). Second, removing 1ES 1959+650 as well degrades the limit moderately, while further removing Mrk 501 — leaving Mrk 421 alone — weakens it appreciably in the $m_a \simeq 10^{-8}$–10^{-7} eV range, where the combination is strongest. The constraining power is therefore shared mainly between Mrk 421 and Mrk 501, with the two remaining targets playing a marginal role. Third, and consistently with the correlation discussed above, the ten most constraining BBs alone — roughly 2% of the sample — already reproduce the shape of the full exclusion.

This observation further supports the conclusion drawn from Fig. 7 that the exclusion is driven primarily by a relatively small number of bright flaring states. It also motivates the systematic studies presented in the appendix, where, due to the substantial computational cost of repeating the full ALP analysis with modified likelihoods, we perform the tests only on these ten BBs.

4. Conclusions

We have searched for ALP-induced spectral irregularities in the GeV–TeV spectra of four blazars — Mrk 421, Mrk 501, BL Lacertae and 1ES 1959+650 — observed with MAGIC

entire $(m_a, g_{a\gamma})$ grid, so different blocks can reach their maximum significance at different ALP parameter values. The global combination, instead, sums W_i coherently at fixed ALP parameters.

between 2014 and 2024 and with LST-1 between 2020 and 2024, for a total of 980.7 h. To our knowledge this is the largest dataset assembled to date for a ground-based ALP search. MAGIC data were processed with the newly developed automated reconstruction pipeline `autoMAGIC`, while LST-1 data were reduced with the standard `lstchain` pipeline; from DL3 onward the two datasets were analyzed with an identical, instrument-independent procedure.

The methodological contribution of this work is the statistical framework of Sect. 2.5. The dataset is segmented into 442 Bayesian Blocks, each an interval of statistically constant flux state, and each block contributes a profile test statistic W_i together with a Monte-Carlo calibrated number of degrees of freedom d_i, evaluated over the ALP parameter grid. Because each block contributes independently once its nuisance parameters are profiled out, the method is naturally modular and readily extensible: additional data, sources, or instruments can be incorporated by simply adding their likelihood contributions, without repeating the analysis of the existing dataset.

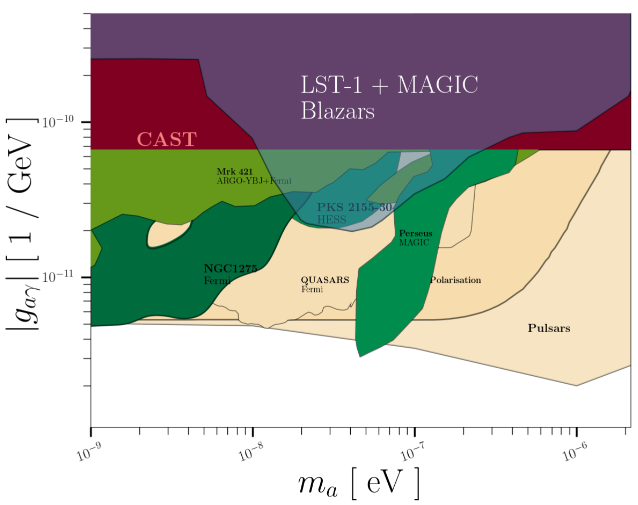

Figure 9: The 95% CL exclusion contour on the ALP-photon coupling $g_{a\gamma}$ as a function of the ALP mass m_a obtained in this work (solid black line, with the excluded region shaded) combining all MAGIC and LST-1 observations of BL Lac, Mrk 421, Mrk 501, and 1ES 1959+650, compared with existing constraints from laboratory experiments and astrophysical probes. This includes limits from the CAST helioscope, star cooling, and SN 1987A, together with exclusion contours from previous gamma-ray ALP searches with H.E.S.S. (PKS 2155−304), Fermi-LAT (NGC 1275), and earlier MAGIC analyses.

We find no statistically significant evidence for photon–ALP mixing, and therefore derive 95% confidence-level upper limits on the ALP-photon coupling $g_{a\gamma}$ for $m_a \in [1, 1000]$ neV. The combined exclusion exceeds 2σ significance over the region of the ALP parameter space $m_a \in [\sim 10^{-8}, \sim 10^{-7}]$ eV for couplings down to $g_{a\gamma} \gtrsim 2 \times 10^{-11}$ GeV^{-1}, broadly consistent with previous gamma-ray searches (Fig. 9).

Two features of the combined analysis are worth highlighting. The constraining power is concentrated in a small number of high-flux blocks: the ten most constraining BBs, approximately, 2% of the sample, already reproduce the shape of the full exclusion, and rare flaring episodes, therefore, matter more than long integrations in the baseline state. Furthermore, the combined constraint is driven predominantly by Mrk 421 and Mrk 501, with the two remaining targets contributing marginally.

Despite the unprecedented size of the dataset, this exclusion does not substantially extend current astrophysical constraints – a limitation of blazar-based ALP searches rather than of the statistical power of the present analysis. Unlike searches based on galaxy clusters or the Galactic magnetic field, photon–ALP conversion in blazars is governed by the magnetic-field structure of the relativistic jet, which cannot be directly measured and constitutes the dominant source of uncertainty in the analysis. The limits derived here, as in previous blazar-based ALP searches, should therefore be read as conditional on the adopted jet field model (e.g. Davies et al. 2021; Gao et al. 2025; Gao et al. 2024). The dominant systematic remains the jet magnetic-field model, as it is for every blazar-based ALP search. We have quantified it in Sect. C.2, together with the absolute energy scale, the effective collection area and the Galactic field model, and we have argued that the quantity governing the mixing, $B_0 r_0$, is more tightly constrained by VLBI core-shift measurements than B_0 and r_0 are individually. The residual uncertainty concerns the structure of the field rather than its normalization, and we have measured the associated realization-to-realization dispersion explicitly. The limits reported here should be read as conditional on the adopted field model, as is the case for all published constraints of this type.

While the exclusion derived in this work is ultimately limited by the current understanding of AGN jet magnetic fields, the statistical framework presented here is independent of any particular dataset or instrument. Its modular, likelihood-level combination of Bayesian Blocks provides a scalable approach for incorporating new observations without reanalyzing existing data. With CTAO-N approaching completion, the natural continuation of this work is to extend the block list to further sources and observatories. Given that a handful of flaring blocks carries most of the sensitivity, the most effective gain will come not from longer exposures but from catching more of the rare, bright states.

Closing Statements

Acknowledgments. We would like to highlight discussions with J. Biteau, S. Jonaas, A. Tramacere for the jet magnetic field modeling.
[MD: Acknowledgments to be rewritten.]

Author Contributions. Conceptualization: MD, IB. Statistical framework, Bayesian Block analysis and final limits: GDA. `autoMAGIC` pipeline and MAGIC data reconstruction: FS. Magnetic-field modelling and `gammaALPs` simulations: IB. Writing — original draft: MD. Writing — review and editing: all authors.

Data Availability. The DL3 data analysed in this work are proprietary to the MAGIC and LST collaborations and are not publicly available; requests for access should be addressed to the respective collaborations. The analysis chain used from DL3 onwards — comprising the interface to `gammaALPs`, the

Bayesian Block segmentation, and the implementation of the statistical framework described in Sect. 2.5 — is publicly available at [**MD:** placeholder: Zenodo/GitHub DOI].

References

Abdalla, H., et al. (CTA), 2021. Sensitivity of the Cherenkov Telescope Array for probing cosmology and fundamental physics with gamma-ray propagation. JCAP 02, 048. doi:10.1088/1475-7516/2021/02/048, arXiv:2010.01349.

Abramowski, A., et al. (H.E.S.S.), 2013. Constraints on axionlike particles with H.E.S.S. from the irregularity of the PKS 2155-304 energy spectrum. Phys. Rev. D 88, 102003. doi:10.1103/PhysRevD.88.102003, arXiv:1311.3148.

Acciari, V.A., et al. (MAGIC, FACT), 2021. Investigation of the correlation patterns and the Compton dominance variability of Mrk 421 in 2017. Astron. Astrophys. 655, A89. doi:10.1051/0004-6361/202141004, arXiv:2106.05516.

Adams, C.B., et al. (VERITAS), 2025. Constraints on axionlike particles from VERITAS observations of a flaring radio galaxy in the Perseus cluster. Phys. Rev. D 112, 103044. doi:10.1103/jydc-npy6, arXiv:2510.19010.

Ade, P.A.R., et al. (Planck), 2016. Planck 2015 results. XIX. Constraints on primordial magnetic fields. Astron. Astrophys. 594, A19. doi:10.1051/0004-6361/201525821, arXiv:1502.01594.

Ajello, M., et al. (Fermi-LAT), 2016. Search for Spectral Irregularities due to Photon–Axionlike-Particle Oscillations with the Fermi Large Area Telescope. Phys. Rev. Lett. 116, 161101. doi:10.1103/PhysRevLett.116.161101, arXiv:1603.06978.

Alves Batista, R., Saveliev, A., 2021. The Gamma-ray Window to Intergalactic Magnetism. Universe 7, 223. doi:10.3390/universe7070223, arXiv:2105.12020.

Arias, P., Cadamuro, D., Goodsell, M., Jaeckel, J., Redondo, J., Ringwald, A., 2012. WISPy cold dark matter. J. Cosmol. 2012, 013. doi:10.1088/1475-7516/2012/06/013, arXiv:1201.5902.

Banerjee, A., Finke, J.D., Ajello, M., Domínguez, A., Desai, A., Baxter, J., Hartmann, D., Paliya, V.S., 2026. A new measurement of the extragalactic background light using 15 yr of Fermi-large area telescope data. Astrophys. J. 1003, 86. URL: https://iopscience.iop.org/article/10.3847/1538-4357/ae581a, doi:10.3847/1538-4357/ae581a.

Batković, I., De Angelis, A., Doro, M., Manganaro, M., 2021. Axion-Like Particle Searches with IACTs. Universe 7, 185. doi:10.3390/universe7060185, arXiv:2106.03424. [Erratum: Universe 8, 74 (2022)].

Bernete, J., García-Soto, S., Hassan, T., Gueta, O., Linhoff, M., Sinha, A., Maier, G., 2026. Enhancing the Cherenkov Telescope Array Observatory high-level performance through an event-type-based analysis. Astropart. Phys. 182, 103277. doi:10.1016/j.astropartphys.2026.103277, arXiv:2607.09286.

Chadwick, P., 2021. 35 Years of Ground-Based Gamma-ray Astronomy. Universe 7, 432. doi:10.3390/universe7110432.

Cortina, J., 2019. Status of the large size telescopes of the cherenkov telescope array. arXiv preprint arXiv:1907.10146.

Cowan, G., Cranmer, K., Gross, E., Vitells, O., 2011. Asymptotic formulae for likelihood-based tests of new physics. The European Physical Journal C 71, 1–19.

D'Amico, G., 2022. Statistical Tools for Imaging Atmospheric Cherenkov Telescopes. Universe 8, 90. doi:10.3390/universe8020090, arXiv:2202.04590.

Davies, J., Meyer, M., Cotter, G., 2021. Relevance of jet magnetic field structure for blazar axionlike particle searches. Phys. Rev. D 103, 023008. doi:10.1103/PhysRevD.103.023008, arXiv:2011.08123.

Davies, J., Meyer, M., Cotter, G., 2023. Constraints on axionlike particles from a combined analysis of three flaring Fermi flat-spectrum radio quasars. Phys. Rev. D 107, 083027. doi:10.1103/PhysRevD.107.083027, arXiv:2211.03414.

Deil, C., Zanin, R., Lefaucheur, J., Boisson, C., Khelifi, B., Terrier, R., Wood, M., Mohrmann, L., Chakraborty, N., Watson, J., Lopez-Coto, R., Klepser, S., Cerruti, M., Lenain, J.P., Acero, F., Djannati-Ataï, A., Pita, S., Bosnjak, Z., Trichard, C., Vuillaume, T., Donath, A., Consortium, C., King, J., Jouvin, L., Owen, E., Sipocz, B., Lennarz, D., Voruganti, A., Spir-Jacob, M., Ruiz, J.E., Arribas, M.P., 2017. Gammapy - A prototype for the CTA science tools, in: 35th International Cosmic Ray Conference (ICRC2017), p. 766. arXiv:1709.01751.

Dominguez, A., et al., 2011. Extragalactic Background Light Inferred from AEGIS Galaxy SED-type Fractions. Mon. Not. Roy. Astron. Soc. 410, 2556. doi:10.1111/j.1365-2966.2010.17631.x, arXiv:1007.1459.

Domínguez, A., Primack, J.R., Rosario, D.J., Prada, F., Gilmore, R.C., et al., 2011. Extragalactic background light inferred from AEGIS galaxy-SED-type fractions. Mon. Not. R. Astron. Soc. 410, 2556–2578. URL: https://doi.org/10.1111/j.1365-2966.2010.17631.x, doi:10.1111/j.1365-2966.2010.17631.x, arXiv:https://academic.oup.com/mnras/article-pdf/410/4/2556/629525

Durrer, R., Neronov, A., 2013. Cosmological magnetic fields: their generation, evolution and observation. Astronomy & Astrophysics Review 21, 62. URL: https://doi.org/10.1007/s00159-013-0062-7, doi:10.1007/s00159-013-0062-7, arXiv:arXiv:1303.7121.

Franceschini, A., Rodighiero, G., 2017. The extragalactic background light revisited and the cosmic photon-photon opacity. Astron. Astrophys. 603, A34. doi:10.1051/0004-6361/201629684, arXiv:1705.10256.

Galanti, G., Nava, L., Roncadelli, M., Tavecchio, F., Bonnoli, G., 2023. Observability of the Very-High-Energy Emission from GRB 221009A. Phys. Rev. Lett. 131, 251001. doi:10.1103/PhysRevLett.131.251001, arXiv:2210.05659.

Gamma-ray astronomy community, 2018. Gamma astro data formats (gadf) specification. https://gamma-astro-data-formats.readthedocs.io/en/v0.3/. Version 0.3. Accessed: 2025-02-06.

Gao, L.Q., Bi, X.J., Guo, J.G., Lin, W., Yin, P.F., 2024. Constraints on axionlike particles from observations of mrk 421 using the $\mathrm{cl_s}$ method. Phys. Rev. D 109, 063003. URL: https://link.aps.org/doi/10.1103/PhysRevD.109.063003, doi:10.1103/PhysRevD.109.063003.

Gao, L.Q., Bi, X.J., Li, J., Yin, P.F., 2025. Impact of parameters in the toroidal blazar jet magnetic field model on axion-like particle constraints. J. Cosmol. Astropart. Phys. 2025, 031. doi:10.1088/1475-7516/2025/01/031, arXiv:2407.20118.

Govoni, F., Feretti, L., 2004. Magnetic field in clusters of galaxies. Int. J. Mod. Phys. D 13, 1549–1594. doi:10.1142/S0218271804005080, arXiv:astro-ph/0410182.

Gréaux, L., Biteau, J., Hassan, T., Hervet, O., Nievas Rosillo, M., Williams, D.A., 2023. STeVECat, the Spectral TeV Extragalactic Catalog. PoS ICRC2023, 751. doi:10.22323/1.444.0751.

Han, C., Lei, Z., Yang, J., Zhao, S., 2026. Line-of-sight magnetic-field propagation effects on axion-like particle constraints from GRB 221009A arXiv:2606.15269.

Hillas, A.M., 1985. Cerenkov light images of eas produced by primary gamma rays and by nuclei, in: Proceedings of the 19th International Cosmic Ray Conference, La Jolla, USA. pp. 445–448.

Hooper, D., Serpico, P.D., 2007. Detecting Axion-Like Particles With Gamma Ray Telescopes. Phys. Rev. Lett. 99, 231102. doi:10.1103/PhysRevLett.99.231102, arXiv:0706.3203.

Jansson, R., Farrar, G.R., 2012a. A New Model of the Galactic Magnetic Field. The Astrophys. Journal 757, 14. doi:10.1088/0004-637X/757/1/14, arXiv:1204.3662.

Jansson, R., Farrar, G.R., 2012b. The Galactic Magnetic Field. Astrophys. J. Lett. 761, L11. doi:10.1088/2041-8205/761/1/L11, arXiv:1210.7820.

Knödlseder, J., Tibaldo, L., Tiziani, D., Specovius, A., Cardenzana, J., Mayer, M., Kelley-Hoskins, N., Di Venere, L., Bonnefoy, S., Ziegler, A., et al., 2019. Analysis of the hess public data release with ctools. Astronomy & Astrophysics 632, A102.

Landgraf, B.A., 2024. Impact of Galactic Magnetic Field Models on Photon–Axion-Like Particle Mixing. Master's thesis. RWTH Aachen University. URL: https://astro.uni-bonn.de/~ungar/BA_Landgraf.pdf. supervisors: K. Unger and G. Farrar.

Li, H.J., 2022. Relevance of VHE blazar spectra models with axion-like particles. J. Cosmol. Astropart. Phys. 02, 025. doi:10.1088/1475-7516/2022/02/025, arXiv:2112.14145.

Li, H.J., Bi, X.J., Yin, P.F., 2022. Searching for axion-like particles with the blazar observations of MAGIC and Fermi-LAT *. Chin. Phys. C 46, 085105. doi:10.1088/1674-1137/ac6d4f, arXiv:2110.13636.

Li, H.J., Chao, W., Tan, X.H., Zhou, Y.F., 2025. Axion effects on gamma-ray spectral irregularities. II: Implications of EBL absorption. Phys. Lett. B 862, 139334. doi:10.1016/j.physletb.2025.139334, arXiv:2410.14791.

Li, H.J., Chao, W., Zhou, Y.F., 2024. Upper limit on the axion-photon coupling from Markarian 421. Phys. Lett. B 858, 139075. doi:10.1016/j.physletb.2024.139075, arXiv:2406.00387.

Li, H.J., Guo, J.G., Bi, X.J., Lin, S.J., Yin, P.F., 2021. Limits on axionlike particles from mrk 421 with 4.5-year period observations by argo-ybj and fermi-lat. Phys. Rev. D 103, 083003. URL: https://link.aps.org/doi/10.1103/PhysRevD.103.083003, doi:10.1103/PhysRevD.103.083003.

Li, T.P., Ma, Y.Q., 1983. Analysis methods for results in gamma-ray astronomy. Astrophys. J. 272, 317–324. doi:10.1086/161295.

Liu, N., Yang, J., Cheng, X., Zeng, A.L., Chen, W., Yang, X.L., Hong, X., Zhang, X.X., Liu, J.C., Zhu, Z., 2026. Detections of nearly bias-free core shifts with 5–30 μas precisions at 8–43 ghz in bl lacertae. Astrophys. J. 1003, 54. URL: https://doi.org/10.3847/1538-4357/ae63b4, doi:10.3847/1538-4357/ae63b4.

LST, 2023a. cta-observatory/cta-lstchain. URL: https://doi.org/10.5281/zenodo.8377093, doi:10.5281/zenodo.8377093.

LST (CTA-LST Project), 2023b. Observations of the Crab Nebula and Pulsar with the Large-sized Telescope Prototype of the Cherenkov Telescope Array. Astrophys. J. 956, 80. doi:10.3847/1538-4357/ace89d, arXiv:2306.12960.

LST, 2025. Vhe γ-ray observations of bright bl lacs with the large-sized telescope prototype. submitted to Astronomy & Astrophysics .

MAGIC, 2015a. The 2009 multiwavelength campaign on mrk 421: variability and correlation studies. Astronomy & Astrophysics 576, A126. doi:10.1051/0004-6361/201424216, arXiv:1502.02650.

MAGIC, 2015b. Measurement of the crab nebula spectrum over three decades in energy with the magic telescopes. Journal of High Energy Astrophysics 5-6, 30–38. doi:10.1016/j.jheap.2015.01.002, arXiv:1406.6892.

MAGIC (MAGIC), 2016a. The major upgrade of the magic telescopes, part ii: A performance study using observations of the crab nebula. Astroparticle Physics 72, 76–94. doi:https://doi.org/10.1016/j.astropartphys.2015.02.005.

MAGIC (MAGIC), 2016b. The major upgrade of the MAGIC telescopes, Part I: The hardware improvements and the commissioning of the system. Astropart. Phys. 72, 61–75. doi:10.1016/j.astropartphys.2015.04.004, arXiv:1409.6073.

MAGIC (MAGIC), 2016c. The major upgrade of the MAGIC telescopes, Part II: A performance study using observations of the Crab Nebula. Astropart. Phys. 72, 76–94. doi:10.1016/j.astropartphys.2015.02.005, arXiv:1409.5594.

MAGIC, 2018. Extreme hbl behaviour of markarian 501 during 2012. Astronomy & Astrophysics 620, A181. doi:10.1051/0004-6361/201833704, arXiv:1810.05513.

MAGIC, 2019. A fast, very-high-energy γ-ray flare from bl lacertae during a multiwavelength campaign. Astronomy & Astrophysics 623, A175. doi:10.1051/0004-6361/201834010, arXiv:1901.10900.

MAGIC, 2020a. Broadband characterisation of the very intense tev flares of 1es 1959+650 in 2016. Astronomy & Astrophysics 638, A14. doi:10.1051/0004-6361/201935450, arXiv:2004.02308.

MAGIC, 2020b. Study of the variable broadband emission of mrk 501 during the most extreme swift x-ray activity. Astronomy & Astrophysics 637, A86. doi:10.1051/0004-6361/201834603, arXiv:1906.11658.

MAGIC, 2024a. Constraints on axion-like particles with the perseus galaxy cluster with magic. Physics of the Dark Universe 44, 101425. doi:https://doi.org/10.1016/j.dark.2024.101425.

MAGIC, 2024b. Constraints on axion-like particles with the perseus galaxy cluster with magic. Physics of the Dark Universe 44, 101425.

MAGIC (MAGIC), 2024c. Standardised formats and open-source analysis tools for the MAGIC telescopes data. JHEAp 44, 266–278. doi:10.1016/j.jheap.2024.09.011, arXiv:2409.18823.

Marsh, M.C.D., Matthews, J.H., Reynolds, C., Carenza, P., 2022. Fourier formalism for relativistic axion-photon conversion with astrophysical applications. Phys. Rev. D 105, 016013. URL: https://link.aps.org/doi/10.1103/PhysRevD.105.016013, doi:10.1103/PhysRevD.105.016013.

Meyer, M., Davies, J., Kuhlmann, J., 2021a. gammaALPs: An open-source python package for computing photon-axion-like-particle oscillations in astrophysical environments. PoS ICRC2021, 557. doi:10.22323/1.395.0557, arXiv:2108.02061.

Meyer, M., Davies, J., Kuhlmann, J., 2021b. gammalps. "" URL: https://doi.org/10.5281/zenodo.6344566, doi:10.5281/zenodo.6344566.

Meyer, M., Montanino, D., Conrad, J., 2014. On detecting oscillations of gamma rays into axion-like particles in turbulent and coherent magnetic fields. J. Cosmol. 09, 003. doi:10.1088/1475-7516/2014/09/003, arXiv:1406.5972.

Mirzoyan, R., 2022. Technological Novelties of Ground-Based Very High Energy Gamma-Ray Astrophysics with the Imaging Atmospheric Cherenkov Telescopes. Universe 8, 219. doi:10.3390/universe8040219, arXiv:2411.09815.

Murphy, E., Cawthorne, T.V., Gabuzda, D.C., 2013. Analysing the transverse structure of the relativistic jets of active galactic nuclei. Mon. Not. Roy. Astron. Soc. 430, 1504–1515. doi:10.1093/mnras/sts561, arXiv:1302.0186.

Nigro, C., Hassan, T., Olivera-Nieto, L., 2021. Evolution of Data Formats in Very-High-Energy Gamma-Ray Astronomy. Universe 7, 374. doi:10.3390/universe7100374, arXiv:2109.14661.

O'Hare, C., 2020. cajohare/axionlimits: Axionlimits. https://cajohare.github.io/AxionLimits/. doi:10.5281/zenodo.3932430.

O'Sullivan, S.P., Gabuzda, D.C., 2009. Magnetic field strength and spectral distribution of six parsec-scale active galactic nuclei jets. Mon. Not. Roy. Astron. Soc. 400, 26–42. URL: https://doi.org/10.1111/j.1365-2966.2009.15428.x, doi:10.1111/j.1365-2966.2009.15428.x, arXiv:https://academic.oup.com/mnras/article-pdf/400/1/26/18690573

Padovani, P., Alexander, D.M., Assef, R.J., De Marco, B., Giommi, P., Hickox, R.C., Richards, G.T., Smolčić, V., Hatziminaoglou, E., Mainieri, V., Salvato, M., 2017. Active galactic nuclei: what's in a name? Astronomy and Astrophysics Reviews 25, 2. doi:10.1007/s00159-017-0102-9, arXiv:1707.07134.

Particle Data Group, 2024. Axions and other similar particles. Review of Particle Physics Review article in the Review of Particle Physics.

Peñil, P., Otero-Santos, J., Ajello, M., Buson, S., Domínguez, A., Marcotulli, L., Torres-Albà, N., González, J.B., Acosta-Pulido, J.A., 2024. Multiwavelength variability analysis of Fermi-LAT blazars. Mon. Not. Roy. Astron. Soc. 529, 1365–1385. doi:10.1093/mnras/stae594, arXiv:2403.01520.

Peccei, R.D., Quinn, H.R., 1977. CP conservation in the presence of pseudoparticles. Phys. Rev. Lett. 38, 1440–1443. doi:10.1103/PhysRevLett.38.1440.

Potter, W.J., Cotter, G., 2012a. Synchrotron and inverse-Compton emission from blazar jets – I. A uniform conical jet model. Mon. Not. R. Astron. Soc. 423, 756–765. URL: https://doi.org/10.1111/j.1365-2966.2012.20918.x, doi:10.1111/j.1365-2966.2012.20918.x.

Potter, W.J., Cotter, G., 2012b. Synchrotron and inverse-Compton emission from blazar jets – II. An accelerating jet model with a geometry set by observations of M87. Mon. Not. R. Astron. Soc. 429, 1189–1205. URL: https://doi.org/10.1093/mnras/sts407, doi:10.1093/mnras/sts407.

Potter, W.J., Cotter, G., 2013a. Synchrotron and inverse-Compton emission from blazar jets – III. Compton-dominant blazars. Mon. Not. R. Astron. Soc. 431, 1840–1852. URL: https://doi.org/10.1093/mnras/stt300, doi:10.1093/mnras/stt300.

Potter, W.J., Cotter, G., 2013b. Synchrotron and inverse-Compton emission from blazar jets – IV. BL Lac type blazars and the physical basis for the blazar sequence. Mon. Not. R. Astron. Soc. 436, 304–314. URL: https://doi.org/10.1093/mnras/stt1569, doi:10.1093/mnras/stt1569.

Potter, W.J., Cotter, G., 2015. New constraints on the structure and dynamics of black hole jets. Mon. Not. Roy. Astron. Soc. 453, 4070–4088. doi:10.1093/mnras/stv1657, arXiv:1508.00567.

Pshirkov, M.S., Tinyakov, P.G., Kronberg, P.P., Newton-McGee, K.J., 2011. Deriving global structure of the Galactic Magnetic Field from Faraday Rotation Measures of extragalactic sources. Astrophys. J. 738, 192. doi:10.1088/0004-637X/738/2/192, arXiv:1103.0814.

Pudritz, R.E., Hardcastle, M.J., Gabuzda, D.C., 2012. Magnetic fields in astrophysical jets: From launch to termination. Space Science Reviews 169, 27–72. URL: https://doi.org/10.1007/s11214-012-9895-z, doi:10.1007/s11214-012-9895-z.

Pushkarev, A.B., Hovatta, T., Kovalev, Y. Y., Lister, M. L., Lobanov, A. P., Savolainen, T., Zensus, J. A., 2012. Mojave: Monitoring of jets in active galactic nuclei with vlba experiments - ix. nuclear opacity. Astron. Astrophys. 545, A113. URL: https://doi.org/10.1051/0004-6361/201219173, doi:10.1051/0004-6361/201219173.

Raffelt, G., Stodolsky, L., 1988. Mixing of the Photon with Low Mass Particles. Phys. Rev. D 37, 1237. doi:10.1103/PhysRevD.37.1237.

Sahakyan, N., Vardanyan, V., Giommi, P., Bégué, D., Israyelyan, D., Harutyunyan, G., Manvelyan, M., Khachatryan, M., Dereli-Bégué, H., Gasparyan, S., 2024. Markarian multiwavelength data center (mmdc): A tool for retrieving and modeling multitemporal, multiwavelength, and multimessenger data from blazar observations. Astron. J. 168, 289. URL: https://doi.org/10.3847/1538-3881/ad8231, doi:10.3847/1538-3881/ad8231.

Sanchez-Conde, M.A., Paneque, D., Bloom, E., Prada, F., Dominguez, A., 2009. Hints of the existence of Axion-Like-Particles from the gamma-ray spectra of cosmological sources. Phys. Rev. D 79, 123511. doi:10.1103/PhysRevD.79.123511, arXiv:0905.3270.

Scargle, J.D., Norris, J.P., Jackson, B., Chiang, J., 2013. Studies in astronomical time series analysis. vi. bayesian block representations. The Astrophysical Journal 764, 167. URL: https://doi.org/10.1088/0004-637X/764/2/167, doi:10.1088/0004-637X/764/2/167, arXiv:arXiv:1207.5578.

Schubert, J.L., 2023. Pulsar analyses with the MAGIC experiment: Studies of the Crab and Dragonfly Pulsars with an automated, database-driven workflow and open-source software. Ph.D. thesis. Technische Universität Dortmund. URL: https://eldorado.tu-dortmund.de/items/227a1062-31e9-45a5-9f9e-01154a3c1bd0.

Tavecchio, F., Ghisellini, G., Ghirlanda, G., Foschini, L., Maraschi, L., 2010. TeV BL Lac objects at the dawn of the Fermi era. Mon. Not. Roy. Astron. Soc. 401, 1570–1586. doi:10.1111/j.1365-2966.2009.15784.x, arXiv:0909.0651.

Unger, K., Farrar, G.R., 2023. New galactic magnetic field models from faraday rotation and synchrotron emission. The Astrophysical Journal URL: https://arxiv.org/abs/2311.12138, arXiv:2311.12138. arXiv:2311.12138.

Urry, C.M., Padovani, P., 1995. Unified Schemes for Radio-Loud Active Galactic Nuclei. Publ. Astron. Soc. Pac. 107, 803. doi:10.1086/133630, arXiv:astro-ph/9506063.

Walther, C., et al., 2025. Tools and advancements towards data standardization of the magic collaboration, in: 39th International Cosmic Ray Conference (ICRC2025), Geneva, Switzerland. URL: https://pos.sissa.it/501/875/pdf. poS(ICRC2025) 875.

Weinberg, S., 1978. A New Light Boson? Phys. Rev. Lett. 40, 223–226. doi:10.1103/PhysRevLett.40.223.

Wilczek, F., 1978. Problem of Strong P and T Invariance in the Presence of Instantons. Phys. Rev. Lett. 40, 279–282. doi:10.1103/PhysRevLett.40.279.

Wouters, D., Brun, P., 2012. Irregularity in gamma ray source spectra as a signature of axionlike particles. Phys. Rev. D 86, 043005. doi:10.1103/PhysRevD.86.043005, arXiv:1205.6428.

Zanin, R., Carmona, E., Sitarek, J., Colin, P., Frantzen, K., Lombardi, S., Lopez-Oramas, A., Moralejo, A., Terzic, T., 2013. Mars, the magic analysis and reconstruction software, in: Proceedings of the 33rd International Cosmic Ray Conference (ICRC2013), p. 773. URL: https://pos.sissa.it/175/773.

Zhou, T., Gao, L.Q., Bi, X.J., Yin, P.F., Lin, W., 2025. Exploring constraints on axion-like particles with the observations for blazar Mrk 501. J. Cosmol. Astropart. Phys. arXiv:2501.18860.

Appendices

A. The `autoMAGIC` Framework

The MAGIC data analyses presented in this work employ `autoMAGIC`, a database–driven pipeline for automated MAGIC data reduction, reported in Schubert (2023); Walther et al. (2025). Its design philosophy is to minimize human intervention while ensuring full reproducibility of results. At the core of the system lies a relational SQL database that manages all analysis information through approximately eighty interlinked tables. These include job tables for each MARS executable, data tables for observation runs and Monte Carlo simulations, and logic tables to enforce dependencies and consistency across the workflow.

Analyses are defined by a lightweight configuration file in TOML format, where the user specifies the target, observation time range, quality cuts, and analysis parameters. From this configuration, `autoMAGIC` queries the database to identify the relevant runs, generates the required jobs in a bottom–up fashion, and submits them to the Port d'Informació Científica (PIC) computing cluster[8]. Each reconstruction step, performed with MARS proprietary code (Zanin et al., 2013), is wrapped by a Python layer that handles execution, error reporting, and communication with the database. This design ensures that jobs are only executed when all prerequisites are satisfied, avoiding redundant computation and allowing efficient scaling to large data volumes.

The output consists of standardized DL3 data products (Gamma-ray astronomy community, 2018; MAGIC, 2024c), complemented by meta–information such as observing conditions and quality flags. These are fully compatible with open–source high–level analysis frameworks like `Gammapy` (Deil et al., 2017). The integration of a relational database with automated job management makes `autoMAGIC` a powerful tool for reproducible, scalable, and transparent MAGIC data analysis. See also Nigro et al. (2021) for a broader review.

Some samples of the MAGIC data analyzed in this work using `autoMAGIC` were used to produce SEDs, whose spectral parameters were cross-checked against those obtained by conventional MARS analysis in order to ensure consistency within statistical errors.

B. Statistics studies

B.1. Details on the Bayesian Block binning

DL3 data from MAGIC and LST-1 are structured in units of runs of $\lesssim$ 20 minutes duration. Shorter units (sub-runs) can appear if fractions of runs are excluded. If the intrinsic source variability is slower than this timescale—which is typically the case—sub-runs and even full runs are often grouped into night-bins. This approach must be done carefully because the instrument acceptance changes continuously during the night, most notably due to the changing zenith angle of the target, which influences the energy reconstruction. With the adoption of a common DL3 formalism (Nigro et al., 2021; Gamma-ray astronomy community, 2018), the community is therefore shifting toward analysis based on individual *events*, each associated with its own IRF (Bernete et al., 2026).

To handle variability without imposing fixed night-bins, we apply the Bayesian Blocks (BB) algorithm to the time series of N original runs

$$\{(t_j, y_j, \sigma_j)\}_{j=1}^{N},$$

where t_j is the run center, y_j the measured flux (or counts), and σ_j its uncertainty. The algorithm partitions the index set

[8] https://www.pic.es/

$\{1, \ldots, N\}$ into contiguous "blocks." A block B is thus a group of M consecutive original runs, within which the signal is modeled as constant with mean value μ. We did not consider subruns.

For a candidate block B, the Gaussian log-likelihood is

$$\ln \mathcal{L}_B(\mu) = -\frac{1}{2} \sum_{j \in B} \frac{(y_j - \mu)^2}{\sigma_j^2} \; + \; \text{const},$$

which is maximized at the weighted average

$$\hat{\mu}_B = \frac{\sum_{j \in B} w_j y_j}{\sum_{j \in B} w_j}, \qquad w_j \equiv \sigma_j^{-2}.$$

Substituting back gives the block fitness

$$\mathcal{F}_B = \frac{1}{2} \frac{\left(\sum_{j \in B} w_j y_j \right)^2}{\sum_{j \in B} w_j}.$$

The Bayesian Blocks algorithm then selects the partition that maximizes

$$\sum_B \mathcal{F}_B \; - \; K \, \text{ncp\_prior},$$

where K is the number of blocks. The penalty term ncp_prior regulates the odds against introducing a new change point. In the `astropy.stats.bayesian_blocks` implementation used by `gammapy`, this is controlled by the user parameter `p0`, the false positive rate for change points. Smaller values of `p0` yield fewer splits and thus larger blocks (containing more original bins), while larger values of `p0` lead to more frequent splits and finer resolution. Although the closed-form calibration of ncp_prior was originally derived for event data (Scargle et al., 2013), it is widely and effectively applied as a practical heuristic to Gaussian flux measurements as well.

Compared to the event-based fitness (appropriate for unbinned photon arrival times) or the `regular_events` option (equal-width Poisson bins), the `measures` fitness used here directly incorporates heteroscedastic uncertainties through the weights w_j. This makes it particularly well suited to gamma-ray light curves with flux measurements and per-point errors, and ensures compatibility with community-standard DL3 analyses.

B.2. Calibration of the Effective Degrees of Freedom

The statistic $W_i(g_{a\gamma}, m_a)$ defined in Eq. 7 is asymptotically χ^2-distributed with $N_i - p_i$ degrees of freedom (see Eq. 9). However, low counts and empty or nearly empty bins can induce non-negligible deviations from the asymptotic expectation. For this reason, the number of degrees of freedom $d_i(g_{a\gamma}, m_a)$ of each Bayesian Block (BB) must be determined by a dedicated Monte Carlo (MC) calibration.

For a given data block i and for a fixed point $(g_{a\gamma}, m_a)$ in the ALP parameter space, the calibration proceeds as follows:

1. **Fit to real data.** We perform the fit described in App. 2.5 for the chosen $(g_{a\gamma}, m_a)$, obtaining the expected signal counts s_{ij} and background estimates $\hat{b}_{ij}(s_{ij})$. These represent the model prediction under the hypothesis that the ALP parameters $(g_{a\gamma}, m_a)$ are the true ones.

2. **Generate MC realizations.** We generate synthetic ON/OFF observations by drawing

$$n_{ij}^{\rm MC} \sim \text{Poisson}(s_{ij} + \alpha_i \hat{b}_{ij}), \qquad m_{ij}^{\rm MC} \sim \text{Poisson}(\hat{b}_{ij}),$$

i.e. we simulate what would be observed if the ALP hypothesis were true.

3. **Refit each MC realization.** For each simulated dataset, we refit the SED parameters $\boldsymbol{\mu}_i$ and compute the corresponding statistic $W_i^{\rm MC}(g_{a\gamma}, m_a)$.

4. **Extract the effective d.o.f.** The resulting distribution of $W_i^{\rm MC}$ is well described by a χ^2 distribution, as illustrated in Fig. B.10, where both the null and an ALP hypothesis are shown for comparison. Its effective degrees of freedom $d_i(g_{a\gamma}, m_a)$ are given by the expected value of the distribution,

$$d_i(g_{a\gamma}, m_a) \; \equiv \; \mathrm{E}\Big[W_i^{\rm MC}(g_{a\gamma}, m_a) \Big],$$

as predicted for a χ^2 random variable where $\mathrm{E}[\chi^2] = d$ is the number of degrees of freedom.

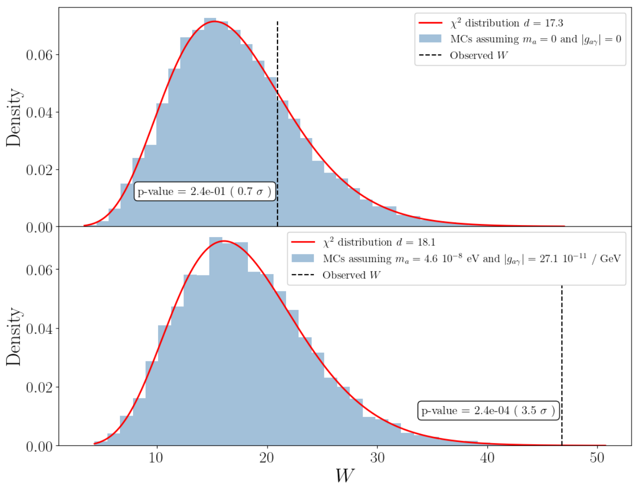

Figure B.10: Distribution of the MC-simulated test statistic $W_i^{\rm MC}$ for a given BB under two different hypotheses. **Top:** Null hypothesis with $m_a = 0$ and $g_{a\gamma} = 0$. **Bottom:** ALP hypothesis with $m_a = 4.6 \cdot 10^{-8}\,\text{eV}$ and $|g_{a\gamma}| = 27.1 \cdot 10^{-11}\,\text{GeV}^{-1}$. The blue histograms show the $W_i^{\rm MC}$ values obtained from pseudo-experiments, while the red curves represent the best-fit χ^2 distributions. The vertical dashed lines indicate the observed W_i value from real data. The reported p-values and corresponding Gaussian significances quantify the level of disagreement between the model and data.

This procedure is repeated for every BB and point of the grid in the ALP parameter space for which exclusion limits are computed.

Treatment of blocks with poor intrinsic SED fits. In some BBs, the intrinsic SED models (EPWL or LP) may provide a poor fit under the *null* hypothesis (no ALPs), leading to a global W value (see Eq. 10) significantly larger than the expected χ^2 distribution. First, we verify that ALPs do not improve the fit by checking that

$$W(g_{a\gamma} = 0, m_a = 0) \;\lesssim\; W(g_{a\gamma}, m_a), \tag{B.1}$$

i.e., the mismatch is due to the spectral model, not to missing ALP effects.

Rather than discarding such blocks, we adjust the Monte Carlo simulations to reflect this disagreement. Let $\bar{W}^{\mathrm{MC}}(0,0)$ be the mean of the simulated values W under the null hypothesis and $W^{\mathrm{data}}(0,0)$ the value obtained from the real data. If

$$W_i^{\mathrm{data}}(0,0) > \bar{W}^{\mathrm{MC}}(0,0), \tag{B.2}$$

we shift the simulated distribution by

$$\Delta = W^{\mathrm{data}}(0,0) - \bar{W}^{\mathrm{MC}}(0,0), \tag{B.3}$$

so that for each point of the grid in the ALP parameter space

$$W^{\mathrm{MC}}(g_{a\gamma}, m_a) \rightarrow W^{\mathrm{MC}}(g_{a\gamma}, m_a) + \Delta. \tag{B.4}$$

This shift (which is also visually described in Fig. B.11) ensures that the MC-calibrated degrees of freedom match the observed discrepancy, allowing BBs that are not well described [9] by the intrinsic spectral model to contribute consistently to the global exclusion analysis.

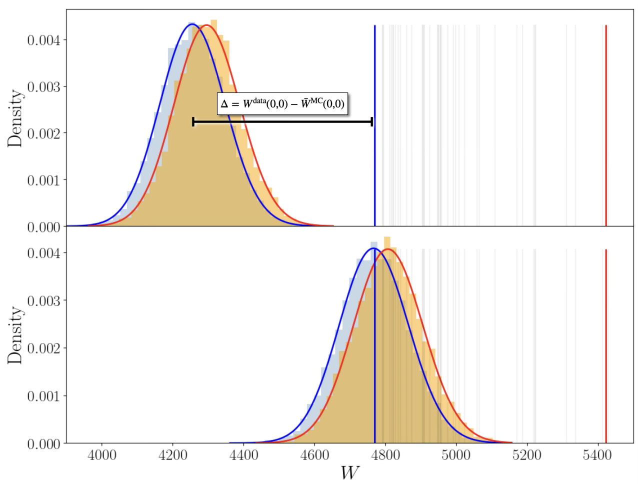

Figure B.11: Illustration of the correction applied to MC-simulated distributions of the test statistic W. The histograms show the distribution of W^{MC} values obtained under two hypotheses: the null hypothesis (blue) with $m_a = 0$ and $g_{a\gamma} = 0$, and an ALP benchmark hypothesis (red) with $m_a = 2.2 \cdot 10^{-8}$ eV and $|g_{a\gamma}| = 0.4 \cdot 10^{-11}\,\mathrm{GeV}^{-1}$. Solid curves represent the best-fit χ^2 distributions, while vertical lines indicate the corresponding W^{data} values for each hypothesis (blue for null, red for benchmark ALP). Light grey vertical lines show observed W values for other ALP hypotheses. **Top:** Original (uncorrected) MC distributions. **Bottom:** Same distributions after applying a shift $\Delta = W^{\mathrm{data}}(0,0) - \bar{W}^{\mathrm{MC}}(0,0)$ that aligns the simulated null distribution with the observed value.

[9] For completeness, we also repeated the analysis by *removing* BBs whose null-hypothesis fit showed a discrepancy larger than 2σ with respect to the χ^2 expectation. The resulting exclusion limits were consistent with those obtained using the shift prescription in Eq. B.4. We nevertheless prefer the shift approach because discarding BBs requires defining what constitutes a "bad" block (e.g. adopting a 2σ threshold), which is an intrinsically arbitrary choice that can affect the outcome. The shift method avoids this ambiguity and incorporates the observed discrepancy in a well-defined and continuous manner.

Validation of the shift prescription. The validity of the shift prescription in Eq. B.4 was tested with dedicated toy MC simulations. We first generated ON/OFF counts under the null hypothesis of no ALP signal and performed the complete analysis on this fake dataset, obtaining the corresponding exclusion region in the $(m_a, g_{a\gamma})$ parameter space. We then repeated the procedure after artificially increasing or decreasing the signal counts in selected energy bins, thereby introducing a spectral mismatch analogous to that of a BB with a poor intrinsic SED fit. For this modified dataset, the calibration was performed using the shift prescription described above. The resulting exclusion regions were found to be consistent with those obtained from the unmodified dataset. This procedure was repeated, yielding each time consistent results. This test provide an empirical validation that the shift prescription does not introduce a significant bias in the derived ALP exclusion and can therefore be safely applied to BBs with a poor intrinsic SED fit.

C. Systematic studies

C.1. *Errors in the energy scale*

An imperfect reconstruction of the gamma-ray energy modifies the position of the ALP-induced spectral features with respect to the observed spectrum and may therefore affect the derived exclusion limits. The absolute energy-scale uncertainty is estimated to be of order ±15% for both MAGIC (MAGIC, 2016a) and LST-1 over the energy range relevant to this work. To estimate the corresponding systematic uncertainty, we recompute the photon survival probability assuming that the true photon energy is shifted by ±15%, while leaving the reconstructed data unchanged. This effectively shifts the position of the ALP-induced oscillations in energy and provides a conservative estimate of the impact of the energy-scale uncertainty on the exclusion limits.

Since repeating the full analysis for the complete dataset of several hundred Bayesian blocks would be computationally prohibitive, we perform this study on the largest and most constraining individual dataset, namely the MAGIC observations of Mrk 421. Within this dataset, we select the 10 most constraining BB (ranked according to its individual constraining power on the ALP parameter space), which provides a representative and computationally tractable benchmark for evaluating the impact of this systematic uncertainty.

Fig. C.12 illustrates the effect of the energy-scale uncertainty for this representative set of BBs. The upper and lower panels compare the ALP exclusion limits obtained without applying any energy correction (blue contour) with the ones (in orange) obtained with a positive and negative shift, respectively. Overall, the limits remain stable under the assumed energy-scale uncertainty. The largest variation is observed for a negative energy shift (−15%), which slightly weakens the exclusion.

C.2. *Modeling of the relativistic jet magnetic field*

As discussed in Sect. 2.3.1, the jet magnetic field is expected to have the largest impact on photon–ALP mixing among the

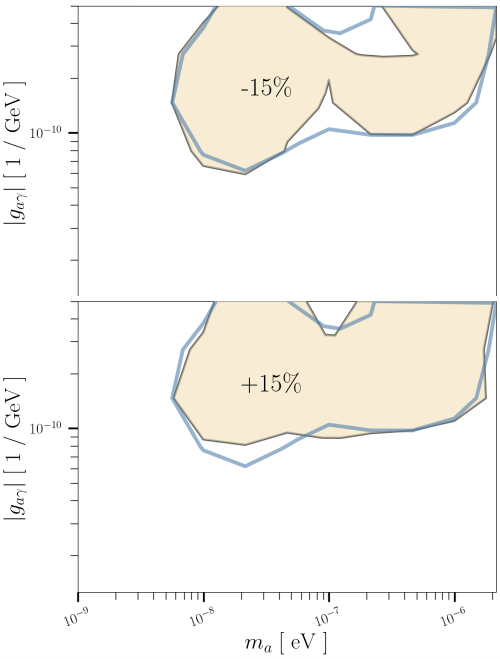

Figure C.12: Impact of the uncertainty in the absolute energy scale on the derived ALP exclusion limits. The blue contour shows the nominal exclusion obtained without applying any energy correction, while the orange region corresponds to the exclusion obtained after shifting the true photon energy by −15% (top) and +15% (bottom), respectively.

propagation environments considered in this work, and its modeling is subject to both theoretical and observational uncertainty. Here we discuss the origin of this uncertainty and quantify its impact on the derived ALP constraints.

Comparison with one-zone based analyses. A number of previous blazar ALP analyses adopt a simple toroidal field $B \propto r^{-1}$ whose normalisation is taken from one-zone synchrotron self-Compton (SSC) fits to the SED, in which the field strength B_0 at the emission site and its distance r_0 from the BH are only loosely constrained, or largely degenerate with one another in some cases (Zhou et al., 2025): Li et al. (2022) and Gao et al. (2024) find that varying r_0 over 10^{16}–10^{18} cm shifts the exclusion contour by one to two orders of magnitude in $g_{a\gamma}$. In a one-zone model the fit determines the field only within a single compact emission region of size $\sim 10^{15}$–10^{16} cm, while the field further out — the region that dominates photon–ALP mixing — follows from an assumed power-law extrapolation. Other works, similarly to ours, adopt the extended PC jet structure (Davies et al., 2021, 2023), in which the emission is computed by integrating along the whole jet and fitted to the radio data simultaneously with all other wavelengths. This matters here because the optically-thick-to-thin synchrotron break, observed at radio frequencies, tightly constrains the radius of the brightest synchrotron-emitting region, which the model identifies with the transition region r_0 (Potter and Cotter, 2015): within this framework r_0 is measured rather than assumed.

Validity of the static, large-scale field assumption. The PC model was originally developed to fit quiescent SEDs of blazars, without attempting to reproduce short-timescale variability (Potter and Cotter, 2012b, 2013a,b). This is reflected in the values of R_T (the size of the parabolic-to-conical transition region) reported by Potter and Cotter (2015), which typically correspond to light-crossing times longer than the variability observed in our targets. Furthermore, the multiwavelength data used to constrain the model are often not strictly simultaneous, so that the resulting parameters cannot be expected to capture the time-dependent behavior of any single source. Nevertheless, the inferred magnetic-field strengths and emission-region distances are compatible with independent estimates: assuming a toroidal field scaling as $B(r) = B_0(r/r_0)^{-1}$ (e.g. Meyer et al., 2014), the resulting field intensities at $r = 1$ pc from the central BH agree with values inferred from core-shift VLBI measurements (e.g. O'Sullivan and Gabuzda, 2009; Pushkarev et al., 2012; Liu et al., 2026). It has also been noted (Potter and Cotter, 2013a) that the PC model remains compatible with fast flares, provided these originate in the accelerating parabolic base of the jet – where light-crossing times are shorter – rather than in the transition region itself, which is the region relevant to our ALP search. That statement was made for Compton-dominant blazars rather than for BL Lac objects, but its geometrical content — a smaller jet radius, and hence a shorter light-crossing time, in the parabolic base than at the transition region — does not depend on the source class.

For what concerns ALP searches specifically, after emission gamma rays travel several kpc within the jet magnetic field before escaping into intergalactic space. While rapid flares can plausibly be produced by sub-pc perturbations of the field, we expect the large-scale structure described by the PC model to be comparatively stable, and therefore suitable to describe the jet field integrated over the gamma-ray escape path during different activity states. A fully time-dependent model, linking the field configuration to the flux state of each Bayesian Block, remains desirable. The Bayesian Block segmentation already prevents spectra taken in different source states from being averaged together, so that ALP-induced features are not smeared out by construction; what remains is that the same static field is assumed for all blocks, whereas the true configuration may differ between them. The direction of the resulting bias is unfavourable: if the assumed field is more ordered than the true one, the predicted spectral features fall where the data do not show them, the ALP hypothesis fits poorly, and the corresponding parameter point is excluded for the wrong reason. Estimating field parameters for each activity period would require quasi-simultaneous multiwavelength coverage and a broadband SED fit for each state; tools such as the Markarian Multiwavelength Data Center (MMDC; Sahakyan et al., 2024)[10] exist for this purpose, but such an attempt remains challenging given non-simultaneous variability across bands.

Constraints on the field normalisation. This core-shift comparison is a solid constraint. Although B_0 and r_0 are individually poorly determined — which is the origin of the large uncertainty reported by Gao et al. (2025) — photon–ALP mixing

[10] https://mmdc.am/

depends on them only through the product appearing in Eq. 2, and it is precisely this product that core-shift measurements determine, being conventionally quoted as $B_{1\,\mathrm{pc}} = B_0\, r_0/(1\ \mathrm{pc})$. The parameters of Tab. 2 illustrate this: $B_0\, r_0$ equals 0.18, 0.24, 0.32 and 0.19 G pc for Mrk 421, Mrk 501, BL Lac and 1ES 1959+650 respectively, so that although r_0 varies by a factor ~ 60 and B_0 by a factor ~ 90 across the sample, their product varies by less than a factor of two.

BL Lacertae is the only target of our sample with direct core-shift determinations, which give $B_{1\,\mathrm{pc}} = 0.139$ G (O'Sullivan and Gabuzda, 2009) and 0.134–0.201 G depending on the assumed core-position index k_r (Liu et al., 2026), against a median $B_{1\,\mathrm{pc}} = 0.4^{+0.3}_{-0.1}$ G for 18 BL Lac objects (Pushkarev et al., 2012). Our value of 0.32 G pc lies between the two. For the remaining three targets only this class-level distribution applies, and their values (0.18–0.24 G pc) fall within it. The factor ~ 3 between individual and class determinations, and the comparable dispersion among independent determinations for the same object, reflect the assumptions on k_r, equipartition and jet kinematics entering the derivation: the comparison constrains $B_0\, r_0$ at the level of a factor of two rather than more precisely. Two caveats apply. For Mrk 421, whose $r_0 = 6.02$ pc exceeds the fiducial radius, quoting $B_{1\,\mathrm{pc}}$ formally requires extrapolating inwards into the parabolic base, although the invariant $B_0\, r_0$ remains well defined. And while the parameters of Tab. 2 are obtained from SED fitting (Potter and Cotter, 2015) without reference to core-shift data, the underlying jet geometry of the PC model was itself partly constrained using core-shift relations (Potter and Cotter, 2013a); the agreement is therefore a consistency check on the field normalisation rather than a fully independent determination.

Coverage of this kind does exist for limited periods. The 2017 multiwavelength campaign on Mrk 421 (Acciari et al., 2021) provides SSC parameters for individual activity phases, and Li (2022) used them to repeat an ALP analysis adopting a distinct jet field for each phase, finding that the resulting constraints did not degrade substantially with respect to the static case. Comparable coverage is, however, not available for the large majority of the 442 BBs analysed here. We leave a full time-dependent treatment to future work, and instead bracket the impact of this uncertainty through the parameter variation described below.

Sensitivity to the emission-region distance r_0. Gao et al. (2025) showed that, within this model, the ALP-mixing uncertainty is dominated by the accuracy with which the distance r_0 of the emission region from the central BH is known – the location where particle acceleration and high-energy gamma-ray production peak, and which we identify with the transition radius $r_T = 10^5 r_g$ ($r_g = 2GM/c^2$; Davies et al. 2021). As noted above, however, the core-shift comparison constrains the product $B_0\, r_0$, so that at fixed $B_0\, r_0$ the residual dependence of the mixing integral on r_0 is only logarithmic. A moderate variation is therefore sufficient to bracket this parameter, and so we vary r_0 by ±30% for each source in our sample to assess the sensitivity of our results to this parameter, using the same representative BB sample adopted for the energy-scale systematic in App. C.1. The impact on the derived ALP exclusion limit is seen in Fig. C.13 and confirms the importance of r_0 in determining the final constraints. The largest overall variation is observed for smaller values of r_0, although the exclusion becomes stronger in the most sensitive mass range. In contrast, increasing r_0 produces an exclusion contour that remains broadly similar to the nominal case, with only a slight weakening of the limits in the most constraining mass range.

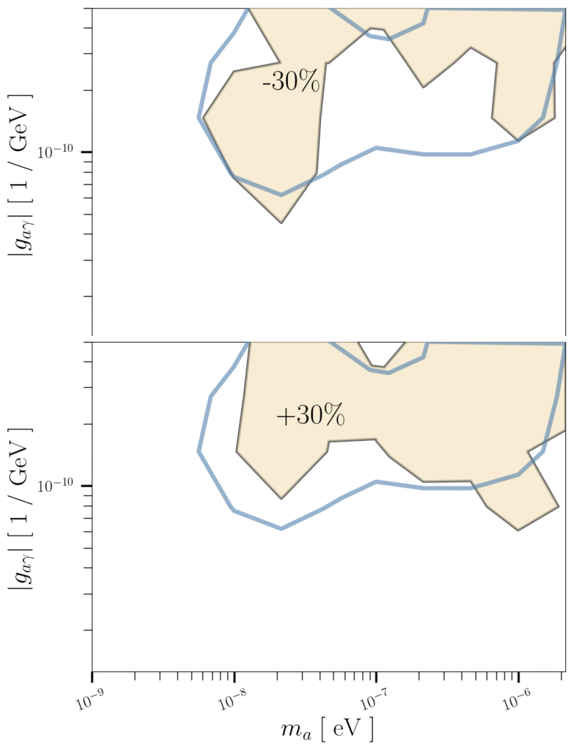

Figure C.13: Impact of a ±30% variation of the jet transition-region distance r_0 on the derived ALP exclusion limits, for the same representative BB sample used in Fig. C.12. Blue contour shows the nominal exclusion obtained without applying any energy correction, while the orange region corresponds to the exclusion obtained after shifting the jet transition-region distance r_0 by −30% (top) and +30% (bottom), respectively.

Field structure. Once the product $B_0\, r_0$ is fixed by the core-shift comparison, the residual uncertainty on the jet field concerns its *structure* rather than its normalisation, that is, the degree to which the field remains ordered along the escape path. This question was addressed by Davies et al. (2021) in the work that introduced the model adopted here, and we follow their prescription. Importantly, the degree of disorder is not a free parameter: the tangled fraction adopted here, $f_t = 0.7$, is the value used by Davies et al. (2021) for Mrk 501, and the range $f_t = 0$–0.7 they consider derives from Murphy et al. (2013), who fit this quantity to rotation-measure maps of that source. Our analysis therefore already assumes a strongly tangled field, at the most disordered end of what radio polarimetry allows, rather than a coherent one. Zhou et al. (2025) examined the impact of this choice for Mrk 501 by comparing a purely toroidal field with a helical-plus-tangled configuration, finding transverse field strengths that differ by nearly an order of magnitude beyond ∼ 1 pc, with the toroidal model yielding the more stringent limits in the mass range $m_a \sim 10^{-7}$–10^{-6} eV. The helical-plus-tangled configuration adopted here is therefore the more conservative of the two over that range.

We have also examined how the tangled component affects the ALP signature itself, using 100 realisations of the field computed for Mrk 421 at $f_t = 0, 0.3$ and 0.7. Contrary to what might be expected, a larger tangled fraction does not smear the spectral features out: the RMS amplitude of the modulation grows with f_t over most of the mass range, by some 30% from $f_t = 0.3$ to 0.7 at $m_a \simeq 100$ neV. The reason is that the tangled component contributes a transverse field everywhere along the path, whereas in the ordered helical configuration the poloidal part does not drive mixing; a more disordered field therefore produces more conversion, not less. Our choice of f_t is consequently not conservative in this respect, and we do not attempt an independent scan of the structural parameters here.

Stochasticity of the tangled component. At fixed f_t the tangled field is a random realisation, controlled in `gammaALPs` by a seed: different realisations of the same physical configuration yield different $P_{\gamma\gamma}(E)$. We quantify this using the 100 realisations mentioned above. Fig. C.14 shows, as a function of ALP mass, the ratio between the realisation-to-realisation dispersion of $P_{\gamma\gamma}$ and the amplitude of the ALP signal itself, both evaluated as RMS averages over the energy range of the analysis. Within the mass range excluded in this work the ratio is of order 0.35–0.45 for the adopted $f_t = 0.7$: the choice of realisation is therefore a non-negligible but sub-dominant contribution, comparable to the other systematics considered in this appendix. The $f_t = 0$ curve, for which no tangled component is present, provides the numerical floor set by the seed-dependent domain subdivision; subtracting it in quadrature, between 80% and 95% of the dispersion is attributable to the tangled field itself rather than to discretisation.

The limits reported in this work correspond to a single, fixed realisation. The complete statistical treatment consists in repeating the spectral fit for every realisation and adopting the 0.95 realisation quantile, as done by Davies et al. (2023); we do not apply it here, and the dispersion quantified above should be taken as the associated uncertainty.

C.3. Modeling of the Galactic magnetic field

GMF model is constantly under development, with several models available and regularly updated with new observations of Faraday rotation measures and polarized synchrotron emission. The most recent model is proposed by Unger and Farrar (2023) and consists of eight different realisations, each representing a different combination of disk-field geometry, halo-field morphology, thermal electron density model, and cosmic-ray electron distribution. All eight variants are consistent with the present observational data and thus jointly span the current systematic uncertainty on the coherent GMF. In the context of photon–ALP propagation, Landgraf (2024) showed that while all UF23 models produce qualitatively similar behavior, they differ in the magnitude of the conversion probability, especially at low energies. The neCL realization, which employs the

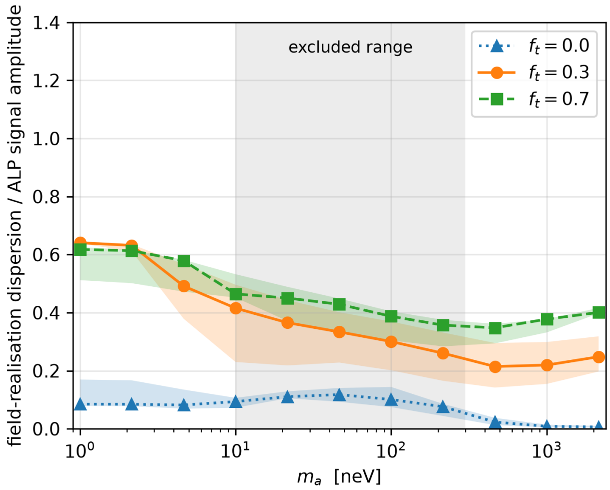

Figure C.14: Ratio between the dispersion of $P_{\gamma\gamma}$ over 100 realisations of the tangled field and the amplitude of the ALP signal, as a function of ALP mass, for three values of the tangled fraction f_t. Lines show the median over the ALP-photon couplings that drive the exclusion ($g_{a\gamma} \geq 4.3 \times 10^{-11}$ GeV^{-1}), bands the interquartile range. The shaded vertical strip marks the mass range excluded in this work. The $f_t = 0$ curve carries no tangled component and measures the numerical floor due to the seed-dependent domain subdivision.

NE2001 thermal electron distribution—also used in the Jansson and Farrar (2012b) and Pshirkov et al. (2011) models — was found to yield conversion probabilities most similar to those obtained with the Jansson and Farrar (2012b) model used so far in our studies. Landgraf (2024) notes that the neCL realization provides the most consistent comparison, whereas the base and expX models can be regarded as providing an estimate of the maximal systematic range.

For this reason, and to ensure continuity with previous photon–ALP mixing studies, we adopt the UF23–neCL realisation as our default GMF model. A comparison of the ALPs constraints obtained with the base, as well as expX realisation of the GMF model, seen in Fig. C.15 confirms that both produce similar results.

D. Systematic uncertainty in the effective collection area

The instrument response functions (IRFs) used to compute the expected signal counts are derived from Monte Carlo simulations and are therefore affected by finite statistical uncertainties. These uncertainties propagate into the effective collection area and consequently into the expected signal counts in each energy bin. Since the ALP search relies on detecting energy-dependent oscillatory features in the gamma-ray spectrum, statistical fluctuations in the effective area from bin to bin may partially mimic or absorb such signatures. To account for this possibility, we introduce an additional nuisance parameter ε_{ij} for each energy bin, representing a fractional uncertainty on the expected signal yield,

$$s_{ij} \longrightarrow s_{ij}(1 + \varepsilon_{ij}), \tag{D.1}$$

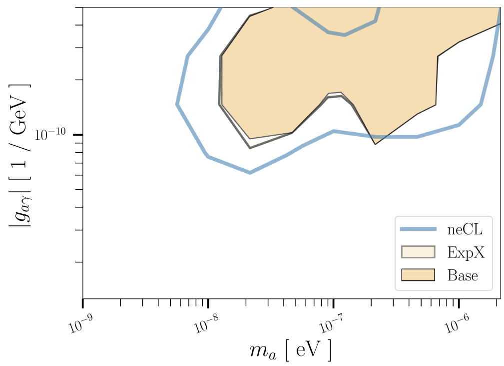

Figure C.15: Impact of the choice of the GMF model realisation on the derived ALP exclusion limit. The blue contour shows the nominal exclusion obtained with the UF23–neCL realisation, while the beige and orange regions correspond to the exclusions obtained with expX and base realisations of the GMF model, respectively.

where each ε_{ij} is assumed to follow an Gaussian prior,

$$\epsilon_{ij} \sim \mathcal{N}(0, \sigma_\varepsilon), \tag{D.2}$$

with σ_ε denoting the relative uncertainty on the effective area.

The ON/OFF likelihood introduced in Eq. 5 is then extended to

$$\mathcal{L}_i(g_{a\gamma}, m_a; \boldsymbol{\mu}_i) = \prod_{j=1}^{N_i} \left[\frac{\left(s_{ij}(1+\varepsilon_{ij}) + \alpha_i b_{ij}\right)^{n_{ij}}}{n_{ij}!} e^{-s_{ij}(1+\varepsilon_{ij}) - \alpha_i b_{ij}} \times \frac{b_{ij}^{m_{ij}}}{m_{ij}!} e^{-b_{ij}} \right] \exp\left(-\frac{\varepsilon_{ij}^2}{2\sigma_\varepsilon^2}\right), \tag{D.3}$$

where the exponential term corresponds to the Gaussian constraint on the nuisance parameter.

As in the nominal analysis, the likelihood is first profiled with respect to the background expectation b_{ij}, which still admits an analytic solution after replacing s_{ij} by $s_{ij}(1+\varepsilon_{ij})$. The likelihood is then profiled over ε_{ij} by solving the stationarity condition

$$\frac{\partial \ln \mathcal{L}}{\partial \varepsilon_{ij}} = 0. \tag{D.4}$$

After substituting the profiled background estimate, this condition reduces to a cubic equation (dropping the index i and j for simplicity),

$$\alpha\, \hat{\varepsilon}^3 + A\, \hat{\varepsilon}^2 + B\, \hat{\varepsilon} + C = 0, \tag{D.5}$$

where the coefficients A, B and C depend on the observed counts, the expected signal s, the exposure ratio α, and the assumed uncertainty σ_ε,

$$A = s\sigma^2(\alpha - 1) + \alpha \tag{D.6}$$

$$B = \sigma^2 \left(\alpha(s - m - n) - s - s^2\sigma^2\right) \tag{D.7}$$

$$C = \sigma^4 s\,(n - \alpha\, m - s)\,. \tag{D.8}$$

The physically admissible real solution $\hat{\varepsilon}$ is obtained analytically using Cardano's formula and substituted back into the likelihood.

Finally, the corresponding profile likelihood is used to define the modified test statistic

$$W_\epsilon(g_{a\gamma}, m_a) = -2 \ln \mathcal{L}\left(g_{a\gamma}, m_a; \hat{\boldsymbol{\mu}}, \hat{\epsilon}\right) + 2 \ln \hat{\mathcal{L}}, \tag{D.9}$$

which reduces to the standard statistic of Eq. 7 in the limit $\sigma_\varepsilon \to 0$.

To assess the impact of uncertainties in the effective collection area, we repeat the ALP analysis using the modified likelihood of Eq. D.3, assuming a relative uncertainty of $\sigma = 0.1$, corresponding to a conservative 10% uncertainty on the expected signal yield in each energy bin. As discussed above, independent fluctuations of the effective area across energy bins may introduce oscillatory distortions in the reconstructed spectrum that could mimic, or partially absorb, the spectral signatures expected from photon-ALP oscillations. As in the study of the other systematic uncertainties, the analysis is performed on a representative subset of the data, namely, the 10 most constraining BBs of the MAGIC observations of Mrk 421.

The resulting exclusion contour is shown in Fig. D.16. The nominal exclusion obtained with the standard likelihood is shown in blue, while the exclusion derived with the modified likelihood including the nuisance parameter ε is shown in orange. As expected, allowing independent variations of the expected signal yield in each energy bin reduces the sensitivity to the oscillatory signatures induced by photon-ALP mixing, leading to a noticeable shrinkage of the excluded region. Nevertheless, the overall morphology of the exclusion contour remains quite unchanged in the mass range between $10^{-8} - 10^{-7}$ neV, indicating that the constraining power of the analysis is degraded but not qualitatively altered in this mass range by this source of systematic uncertainty.

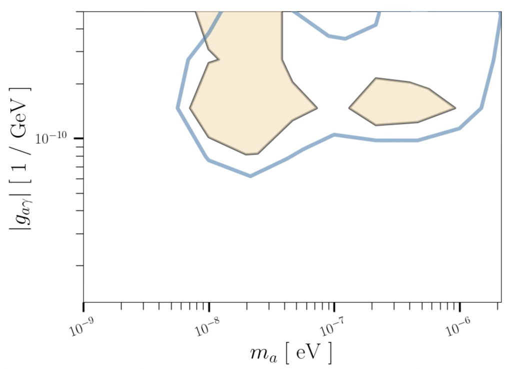

Figure D.16: The blue contour shows the nominal exclusion obtained with the standard likelihood, while the orange contour is obtained using the modified likelihood of Eq. (D.3), assuming an independent 10% uncertainty ($\sigma = 0.1$) on the expected signal yield in each energy bin.