

**Search for Axion Like Particles with the *BABAR* detector
and
photon hadron separation using Zernike moments at Belle II**

by

Alon Hershenhorn

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL
STUDIES

(Physics)

The University of British Columbia

(Vancouver)

May 2021

© Alon Hershenhorn, 2021

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the thesis entitled:

Search for Axion Like Particles with the *BABAR* detector and photon hadron separation using Zernike moments at Belle II

submitted by **Alon Hershenhorn** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy** in **Physics**.

Examining Committee:

Christopher Hearty, Physics, UBC
Supervisor

Douglas Bryman, Physics, UBC
Supervisory Committee Member

Alison Lister, Physics, UBC
University Examiner

Lutz Lampe, Electrical and Computer Engineering, UBC
University Examiner

Additional Supervisory Committee Members:

Gary Hinshaw, Physics, UBC
Supervisory Committee Member

Thomas Mattison, Physics, UBC
Supervisory Committee Member

Mark Van Raamsdonk, Physics, UBC
Supervisory Committee Member

Abstract

Even though the Standard Model of particle physics is a very successful model, we know that it is incomplete. For example, it does not explain why the world we see around us is made of matter rather than anti-matter and it does not incorporate the force of gravity. One extension of the Standard Model is the introduction of Axion Like Particles, ALPs. ALPs appear in string theories and supersymmetry and they might explain some astrophysical anomalies. ALPs can be produced in electron-positron colliders and detected in the specialized detectors built around their interaction point, like the PEP-II collider and the *BABAR* detector at the Stanford Linear Accelerator Center. This work presents an un-blinded search for an ALP that couples exclusively to photons in 5% of the *BABAR* data. We search for an excess in the invariant mass distribution of ALP candidates over a smooth background. The results are consistent with the data being composed only of Standard Model background. 90% credible interval upper limits are set on the ALP production cross section and coupling constant. These limits exclude previously unexplored regions of the phase space in the mass range $0.29 \text{ GeV}/c^2$ to $5 \text{ GeV}/c^2$.

In searches involving photons, it is important to be able to efficiently detect them while rejecting other types of particles. Many high energy particle detectors detect photons in electromagnetic calorimeters that are made up of many cells. A photon interacting with the calorimeter typically leaves a different energy distribution in the cells than some other particle types, hadrons, for example. Discriminating variables for photons, based on Zernike moments, are developed in order to improve the photon identification at Belle II. One of the new variables is found to be the best at identifying photons among all other such variables used at Belle II for photons with energies in the energy range most relevant to $e^+e^- \rightarrow B\bar{B}$ events.

Lay Summary

The Standard Model of particle physics describes the most basic constituents of matter and how they interact with each other. Even though it is very successful and has been experimentally validated to a high degree of accuracy in a multitude of ways, we know that it is not complete. There exist extensions to the Standard Model. One of these extensions predicts a new type of particle that interacts with photons, the particles of light. This work describes a search for this new particle in data collected by the *BABAR* experiment. *BABAR* collided electrons and their anti-particles, positrons. It was built and run by an international scientific collaboration. No significant signal is found in the data.

In addition, new variables for identifying photons are implemented for the Belle II experiment, the scientific successor of *BABAR*. One of these is found to perform better than all available alternatives.

Preface

The idea for the ALP search, described in Chapter 3, was suggested to me by my supervisor, Prof. Christopher Hearty. The design of the analysis was done by both Prof. Hearty and me. The production of signal MC and all the data analysis was done by me. The *BABAR* collaboration designed, built, commissioned, and ran the *BABAR* experiment. It also generated all the background MC. The colliding beams were provided by colleagues at the PEP-II collider. The data was collected between 1999 and 2008. I started my Ph.D. after the end of data taking, and did not participate in all of the above. Figures 3.65 and 3.67 are reproduced with permission from [1], of which I am an author but I did not participate in the analysis.

Prof. Hearty also suggested the idea for the analysis described in Chapter 4. We designed the analysis together. I produced all the MC and performed all other parts of the analysis, except for Figures 4.1 and 4.2 which are modified versions of ones appearing in [2], Figure 4.3 which is reproduced from [3], Figure 4.25 which is taken from [4], and Figures 4.26 to 4.29 which were produced by Torben Ferber. The information in this chapter is summarized in an internal Belle II note - A. Hershenthorn, T. Ferber, and C. Hearty, *ECL shower shape variables based on Zernike moments*, BELLE2-NOTE-TE-2017-001 (2017).

Figures 2.1, 2.2, 2.4 and 2.5 are reproduced with permission from [5]. Figure 2.3 is taken with permission from [6].

Table of Contents

Abstract	iii
Lay Summary	iv
Preface	v
Table of Contents	vi
List of Tables	ix
List of Figures	x
List of Symbols	xiv
Glossary	xv
Acknowledgments	xvi
1 Introduction	1
2 The <i>BABAR</i> experiment	8
2.1 The PEP-II collider	9
2.2 The <i>BABAR</i> detector	9
2.2.1 Silicon vertex tracker	11
2.2.2 Drift chamber	12
2.2.3 Detector of internally reflected Cherenkov light	13
2.2.4 Electromagnetic calorimeter	13

2.2.5	Instrumented flux return	15
3	Search for axion like particles	17
3.1	Analysis overview	17
3.1.1	Disclaimer	19
3.2	Event samples	20
3.2.1	Signal MC simulation	20
3.2.2	Background MC	21
3.2.3	Data samples	22
3.3	Event selection	24
3.3.1	Candidate formation and preliminary selection cuts	24
3.3.2	Trigger	25
3.3.3	Main selection cuts	28
3.3.4	Optimization of selection cut values	32
3.4	Data-MC comparison	47
3.5	Toy experiments creation	50
3.6	Signal extraction	55
3.6.1	Signal shape function	55
3.6.2	Standard model resonance contributions	56
3.6.3	Background shape	62
3.6.4	Fit windows	64
3.6.5	Background only MC fits	67
3.6.6	Signal plus background MC fits	70
3.6.7	Signal efficiency	75
3.7	Search of the data for evidence of an ALP	78
3.7.1	Local significances	78
3.7.2	Global significances	79
3.8	Cross section and coupling constant upper limits	80
3.8.1	CS upper limits with statistical uncertainties only	81
3.8.2	Systematic uncertainties	83
3.8.3	CS upper limits with statistical and systematic uncertainties	85
3.8.4	Coupling constant upper limits	86
3.9	Phase space exclusion	88

3.9.1	Comparison of the Belle II and <i>BABAR</i> limits	90
3.10	Future analysis of the blind data	93
4	Photon hadron separation using Zernike moments	95
4.1	Introduction	95
4.2	Mathematical background	99
4.2.1	Zernike polynomials	99
4.2.2	Zernike moments	100
4.2.3	Zernike moments of a Dirac comb	102
4.3	Plane perpendicular to the shower direction	103
4.4	Calculation of Zernike moments for ECL showers	105
4.5	Selection of best Zernike moments to distinguish between photons and K_L^0 mesons	106
4.5.1	MC samples	107
4.5.2	Selection cuts	108
4.5.3	Zernike moment distributions	110
4.5.4	Selection of the best Zernike moments and the parameters with which to calculate them	112
4.6	MVAs using eleven Zernike moments for <code>basf2</code>	119
4.6.1	MC sample	121
4.6.2	Over training check	122
4.6.3	Results	122
4.7	ECL shower shape module	122
4.8	Comparison of zernikeMVA in data and MC	125
4.9	Shower shape variables in $e^+e^- \rightarrow B\bar{B}$ events	125
4.10	Usage of the zernikeMVA variable at Belle II	127
5	Conclusions	132
	Bibliography	136

List of Tables

Table 3.1	Number of MC background events.	22
Table 3.2	Background MC CSs in nb.	23
Table 3.3	Ratio of background MC luminosity to data luminosity	23
Table 3.4	Number of data events used	23
Table 3.5	Scale factors for the $e^+e^- \rightarrow \gamma\gamma(\gamma)$ MC	50
Table 3.6	Number of entries in a toy experiment from each MC sample	54
Table 3.7	SM resonances cross sections and branching ratios	58
Table 3.8	Correction factor for the $e^+e^- \rightarrow \gamma\phi \rightarrow \gamma\gamma\pi^0$ CS.	61
Table 3.9	Background component Chebyshev polynomial degrees for dif- ferent ALP mass ranges.	67
Table 4.1	Number of events that pass selection cuts	111
Table 4.2	Generated events and those that passed selection cuts	122

List of Figures

Figure 2.1	<i>BABAR</i> schematic in longitudinal direction	10
Figure 2.2	<i>BABAR</i> schematic end view	10
Figure 2.3	Schematic view of the SVT	12
Figure 2.4	Schematic view of the path of the Cherenkov radiation travels inside the DIRC	14
Figure 2.5	Schematic view of the EMC along the beam pipe direction . .	16
Figure 3.1	Feynman diagrams for the ALP production mechanisms. . . .	18
Figure 3.2	Data-MC inconsistency due to trigger effects	27
Figure 3.3	Trigger efficiency after preliminary selection cuts	28
Figure 3.4	2D histogram of $ \Delta\theta $ versus $m_{\gamma\gamma}$	31
Figure 3.5	Di-photon mass distributions with and without the $ \Delta\theta $ cut . .	31
Figure 3.6	Di-photon mass distributions in signal MC	34
Figure 3.7	E^{CM} distributions for 0.5 GeV/c ² ALP before applying the se- lection cut	35
Figure 3.8	E^{CM} distributions for 0.5 GeV/c ² ALP after applying the selec- tion cut	36
Figure 3.9	S/\sqrt{B} vs E^{CM} for 0.5 GeV/c ² ALP	36
Figure 3.10	χ^2 distributions for 0.5 GeV/c ² ALP before applying the selec- tion cut	37
Figure 3.11	χ^2 distributions for 0.5 GeV/c ² ALP after applying the selec- tion cut	38
Figure 3.12	S/\sqrt{B} vs χ^2 for 0.5 GeV/c ² ALP	38

Figure 3.13	$ \Delta\theta $ distributions for 0.5 GeV/c ² ALP before applying the selection cut	39
Figure 3.14	$ \Delta\theta $ distributions for 0.5 GeV/c ² ALP after applying the selection cut	39
Figure 3.15	S/\sqrt{B} vs $\Delta\theta_{\min}$ for 0.5 GeV/c ² ALP	40
Figure 3.16	E^{CM} distributions for 10.3 GeV/c ² ALP before applying the selection cut	40
Figure 3.17	E^{CM} distributions for 10.3 GeV/c ² ALP after applying the selection cut	41
Figure 3.18	S/\sqrt{B} vs E^{CM} for 10.3 GeV/c ² ALP	41
Figure 3.19	χ^2 distributions for 10.3 GeV/c ² ALP before applying the selection cut	42
Figure 3.20	χ^2 distributions for 10.3 GeV/c ² ALP after applying the selection cut	43
Figure 3.21	S/\sqrt{B} vs χ^2 for 10.3 GeV/c ² ALP	43
Figure 3.22	$ \Delta\phi^{\text{CM}} - 180^\circ $ distributions of 10.3 GeV/c ² ALP before applying the selection cut	44
Figure 3.23	$ \Delta\phi^{\text{CM}} - 180^\circ $ distributions for 10.3 GeV/c ² ALP after applying the selection cut	44
Figure 3.24	S/\sqrt{B} vs $ \Delta\phi^{\text{CM}} - 180^\circ $ for 10.3 GeV/c ² ALP	45
Figure 3.25	Optimal values of χ^2_{\min} and smoothed function	45
Figure 3.26	Optimal values of E_{\min}^{CM} and smoothed function	46
Figure 3.27	Optimal values of $\Delta\theta_{\min}$ and smoothed function	46
Figure 3.28	Optimal values of $\Delta\phi_{\min}^{\text{CM}}$ and smoothed function	47
Figure 3.29	Di-photon mass distributions with three sets of E_{\min}^{CM} selection cuts	48
Figure 3.30	Two dimensional histogram of E_{\min}^{CM} versus the ALP candidate mass	48
Figure 3.31	Data-MC comparison	49
Figure 3.32	Data-MC comparison at $\Upsilon(2S)$ CME after MC scaling	51
Figure 3.33	Data-MC comparison at $\Upsilon(3S)$ CME after MC scaling	52
Figure 3.34	Data-MC comparison at $\Upsilon(4S)$ CME after MC scaling	53
Figure 3.35	$e^+e^- \rightarrow e^+e^-\gamma$ distribution approximation	53

Figure 3.36	Toy experiments	54
Figure 3.37	Signal shape component	56
Figure 3.38	η shape	57
Figure 3.39	ω shape	59
Figure 3.40	Signal entries uncertainty when fitting with floating signal and ω components	61
Figure 3.41	ϕ shape	62
Figure 3.42	Percentage of bad fits for $\Upsilon(2S)$ CME	64
Figure 3.43	Percentage of bad fits for $\Upsilon(3S)$ CME	65
Figure 3.44	Percentage of bad fits for $\Upsilon(4S)$ CME	66
Figure 3.45	Di-photon mass distribution for 0.135 GeV/c ² signal MC . . .	68
Figure 3.46	Example background only fit	69
Figure 3.47	Example of number of signal entries distribution	69
Figure 3.48	Example of N_S pull distribution	71
Figure 3.49	Standard deviation of background MC	72
Figure 3.50	Example signal plus background MC fit	73
Figure 3.51	Residuals of N_S for a 4.7 GeV/c ² ALP	73
Figure 3.52	Adjusted residual of a 4.7 GeV/c ² ALP	74
Figure 3.53	Adjusted pull means	75
Figure 3.54	Adjusted pull standard deviations vs. ALP mass	76
Figure 3.55	Adjusted local significances vs. ALP mass	77
Figure 3.56	Signal efficiency as a function of ALP mass	79
Figure 3.57	Local significance of data	80
Figure 3.58	Fits to data with highest significance	81
Figure 3.59	Global significance of data	82
Figure 3.60	Ratio of systematic uncertainties on the number of signal en- tries to statistical uncertainties	85
Figure 3.61	Upper limits on ALP production cross section	87
Figure 3.62	Ratio of CS with and without systematic uncertainties	88
Figure 3.63	Couplin constant upper limit with different priors	89
Figure 3.64	Combined upper limit on coupling constant	89
Figure 3.65	Phase space exclusion	90
Figure 3.66	Comparison of Belle II and <i>BABAR</i> limits	92

Figure 3.67	Phase space exclusion with predicted limits using all the data	94
Figure 4.1	Example energy deposited by a photon	97
Figure 4.2	Example energy deposited by a photon	98
Figure 4.3	Zernike polynomial surface plots	101
Figure 4.4	Perpendicular plane to shower direction	104
Figure 4.5	Radial distance in perpendicular plane	104
Figure 4.6	Calculation of c_p	106
Figure 4.7	Generated K_L^0 momentum in $e^+e^- \rightarrow Y(4S)$ events.	108
Figure 4.8	Generated photon energies distributions	109
Figure 4.9	Reconstructed photon energy distributions	110
Figure 4.10	Reconstructed shower time distributions	112
Figure 4.11	Angle between reconstructed and generated momentum	113
Figure 4.12	Normalized Zernike moments	114
Figure 4.13	Normalized Zernike moments with $\rho_{0;N1} = 20\text{cm}$	115
Figure 4.14	Zernike moments as a function of shower energy	116
Figure 4.15	Zernike moments as a function of shower θ	117
Figure 4.16	Zernike moments as a function of shower ϕ	118
Figure 4.17	Larges radial distance in the perpendicular plane	119
Figure 4.18	ROC comparison	120
Figure 4.19	MVA over training	120
Figure 4.20	K_L^0 rejection at 90 % γ efficiency	121
Figure 4.21	Over training plot of forward endcap MVA.	123
Figure 4.22	Over training plot of barrel MVA.	123
Figure 4.23	Over training plot of backward endcap MVA.	124
Figure 4.24	Final BDT ROC curves	124
Figure 4.25	comparison of zernikeMVA results in data and MC	126
Figure 4.26	ROC for the energy range 25 MeV to 30 MeV	128
Figure 4.27	ROC for the energy range 40 MeV to 50 MeV	129
Figure 4.28	ROC for the energy range 100 MeV to 150 MeV	130
Figure 4.29	ROC for the energy range 500 MeV to 1000 MeV	131
Figure 4.30	True photon energies in $e^+e^- \rightarrow Y(4S)$ events	131

List of Symbols

a	Axion like particle
θ	Polar angle
B^0	Neutral B meson
B^\pm	Charged B meson
B	Charged or neutral B meson
CP	Charge parity symmetry
e^\pm	Electron or positron
η	Eta meson
η'	Eta prime meson
K_L^0	Long lifetime kaon meson
K_S^0	Short lifetime kaon meson
ω	Omega meson
γ	Photon
π^0	Neutral pion meson
ϕ	Azimuthal angle or phi meson
Υ	Upsilon meson

Glossary

ALP	axion like particle
BDT	boosted decision tree
CI	credible interval
CM	center of mass
CME	center of mass energy
CS	cross section
DCH	drift chamber
DIRC	detector for internally reflected Cherenkov light
ECL	electromagnetic calorimeter
EMC	electromagnetic calorimeter
IFR	instrumented flux return
IP	interaction point
ISR	initial state radiation
LST	limited streamer tube
MC	Monte Carlo
MVA	multivariate analysis
PDF	probability density function
PID	particle identification
PMT	photo multiplier tube
QCD	quantum chromodynamics
ROC	receiver operating characteristic
RPC	resistive plate chamber
SM	standard model
SVT	silicon vertex tracker

Acknowledgments

I would like to thank my supervisor, Prof. Christopher Hearty. Your constant availability, both through emails and in face to face meetings, made me feel I could always receive your input on issues I was facing. I really appreciate that. You are extremely intelligent and knowledgeable, and you always have creative ideas. Your slow, methodical approach to problem solving is one I try to imitate. I have learned a lot from you over the years.

Chapter 1

Introduction

The physics model describing our world in the subatomic level is the standard model (SM) of particle physics. It is a quantum field theory that describes the elementary particles that build up all other matter and three out of the four fundamental forces that the particles interact through, the electromagnetic, weak, and strong forces. The SM is a very successful model that manages to explain almost all of the experimental results collected thus far in particle physics. The most recent confirmation of the SM was the discovery of the predicted Higgs boson, independently by the ATLAS and CMS collaborations, in 2012 [7, 8]. For their theoretical prediction of the Higgs boson, Peter Higgs and François Englert were awarded the 2013 Nobel prize in physics.

Even though the SM is so successful, there are things it does not explain and so is not a “theory of everything”. We know that there exist *new physics*, physics beyond the standard model.

One example of an unexplained phenomena is the amount of baryon asymmetry. Baryons are composite particles made up of three quarks or three anti-quarks. The difference between the number of baryons and the number of anti-baryons is called the *baryon asymmetry*. The world we see around us is made up of baryons rather than anti-baryons. There is therefore a baryon asymmetry in nature. A necessary condition for baryon asymmetry is *CP* violation [9], where *C* is charge conjugation and *P* is parity. The SM contains some *CP* violating effects in the weak interactions, but these are not enough to explain the baryon asymmetry we see

around us.

The SM also does not explain what can constitute dark matter. Dark matter is matter that does not emit or absorb light [10]. One of the first evidence for its existence was from galaxy rotation curves, plots of star velocities versus their distance from the center of the galaxy they belong to. These rotation curves showed that the star velocities seemed to deviate from predictions of Newtonian gravity. One of the explanations for this deviation is that the galaxies contain additional matter, dark matter, that does not emit light but has a mass that affects the star velocities. None of the SM particles can take the role of dark matter [11].

Furthermore, the SM does not address the strong CP problem, which will be discussed later in this chapter, and it does not incorporate the force of gravity into the model.

One of the places new physics can be searched for is in the data collected by the B Factories, the $BABAR$ and Belle experiments. The PEP-II accelerator for the $BABAR$ experiment at SLAC and the KEKB accelerator for the Belle experiment at KEK were e^+e^- colliders [12], that operated between the late 1990s to the late 2000s [13]. They had successful physics programs which culminated with the first observation of direct CP violation in the B meson system [14, 15]. For predicting the existence of this phenomena, the 2008 Nobel prize in physics was awarded to Makoto Kobayashi and Toshihide Maskawa.

Even though the B Factories discovered CP violation in the B system, it is not enough to explain the matter anti-matter asymmetry we see in nature. An additional possible source of CP violation is the strong force. The quantum chromodynamics (QCD) Lagrangian has a CP violating term $-(\alpha_s/8\pi) \bar{\Theta} G_{\mu\nu} \tilde{G}^{\mu\nu}$, where α_s is the strong coupling constant, G is the gluonic field strength, and $-\pi \leq \bar{\Theta} \leq \pi$ is an effective parameter after diagonalizing the quark masses [10]. Even though there is no theoretical reason $\bar{\Theta}$ should be small, experimental limits on the neutron electric dipole moment [16] imply that $|\bar{\Theta}| \lesssim 10^{-10}$. The fact that the strong force lacks CP violating effects, even though they are allowed in principle is termed the *strong CP problem*.

A possible solution to this problem is that CP in the strong force is protected by a new global $U(1)_{PQ}$ symmetry [10]. This is known as the Peccei-Quinn mechanism [17, 18]. The new field ϕ_A is spontaneously broken creating a pseudo

Nambu-Goldstone boson, the axion. The axion's coupling to gluons adds a new term to the Lagrangian $(\phi_A \alpha_s)/(8\pi f_A) G_{\mu\nu} \tilde{G}^{\mu\nu}$, where f_A is the axion's decay constant. If fluctuations in the gluon fields create a potential for ϕ_A with a minimum at $\phi_A = \bar{\Theta} f_A$, the new term exactly cancels the CP violating term in the QCD Lagrangian making it CP conserving. For the axion $m_A f_A \approx f_\pi m_\pi$, where $m_\pi = 135 \text{ MeV}/c^2$ is the pion mass and $f_A \approx 92 \text{ MeV}$ its decay constant. A more precise calculation produces [10]

$$m_A = 5.7 \times 10^{-3} \left(\frac{\text{GeV}}{f_A} \right) \text{ GeV}/c^2. \quad (1.1)$$

The axion's coupling constant to photons $g_{A\gamma\gamma}$ depends on the choice of electromagnetic and color anomalies of the axial current associated with the axion [10], but is roughly $g_{A\gamma\gamma} \sim 10^{-3} (\text{GeV}/f_A) \text{ GeV}$ [19]. Because heavier axions have a stronger coupling to photons, axions heavier than $\sim 10 \text{ meV}/c^2$ have been experimentally excluded [19]. The axion is a dark matter candidate and, depending on its mass, can account for part or all of the dark matter content of the universe. Thus, it may simultaneously solve the strong CP problem and account for dark matter at the same time [20].

Axions are searched for both in dedicated experiments and by analyzing astrophysical observations. Light shining through walls experiments shine photons into a strong magnetic field divided into two regions by an optically opaque wall. The photons from the laser can interact with the virtual photons of the magnetic field and convert to axions, but they can not pass through the wall. The axion can pass through the wall and interact with the second part of the magnetic field converting back into a photon. The OSQAR experiment set limits on the axion's coupling to photons of $g_{A\gamma\gamma} < 3.5 \times 10^{-8} \text{ GeV}^{-1}$ in the sub meV range using two 9 T Large Hadron Collider dipole magnets and an 18.5 W laser [10].

If axions are produced in the sun they carry energy and momentum with them and so require more nuclear energy production compared to the scenario in which they are not produced in order to account for the same photon flux. The increased energy production increases the neutrino flux. Measurements of the all-flavor neutrino flux and of the photon flux are used to set limits on the axion's coupling to photons of $g_{A\gamma\gamma} < 4.1 \times 10^{-10} \text{ GeV}^{-1}$ at the 3σ level [10]. Even tighter limits can

be set by analysing stars in globular clusters, spherical groups of stars tightly bound by gravity. Stars in the cluster that have a burning helium core are recognizable because they have a brightness that depends only weakly on their mass. They are called *horizontal branch* stars. If an axion exists, it increases the helium consumption in the core, reducing the lifetime of the horizontal branch stars. Red giants are a different class of stars in globular clusters in which a helium core exists but is not burning. The red giant's lifetime is less sensitive to the existence of the axion. By counting the number of horizontal branch stars and the number of red giants in thirty-nine globular clusters a limit on the axion's coupling to photons is set to $g_{A\gamma\gamma} < 6.6 \times 10^{-11} \text{ GeV}^{-1}$ at 95 % confidence level [10].

Another class of experiments tries to detect axions produced in the sun by detectors on earth. The CAST experiment uses a decommissioned large hadron collider dipole magnet on a mount tracking the sun. Axions interacting with the magnetic field convert to x-ray photons which can be detected by solid-state detectors. CAST set limits on the coupling constant of $g_{A\gamma\gamma} < 6.6 \times 10^{-11} \text{ GeV}^{-1}$ [10]. The ADMX experiment searches for axions from the galactic halo. It uses a high quality factor microwave resonant cavity in an external static 8 T magnetic field. Axions interacting with the magnetic field can convert to photons inside the cavity producing a spike in the power, which can be detected [10].

Axion like particles (ALPs) are a generalization of axions where the mass and coupling constant are independent of each other [10] and they do not necessarily couple to gluons. ALPs appear in string compactifications and in models with broken supersymmetry as the R-axion. Similarly to axions, they can influence the evolution of stars [21]. The ALP is also a dark matter candidate, but it has to have a lifetime greater than the age of the universe in order to account for the dark matter in galaxies [22]. It would therefore decay outside the *BABAR* detector. ALPs below the MeV/c^2 scale can play a role in astrophysics and cosmology. They can affect big bang nucleosynthesis which is the production of nuclei in the early stages of the universe [21]. The ALP can modify the cosmic microwave background by rotating the polarization of the photons [23].

ALPs might explain some astrophysical anomalies, for example, the observation that some stellar systems like red giants, horizontal branch stars and white dwarfs show a weak tendency for non-standard cooling mechanisms compared to

theoretical models. These could be explained by the production of ALPs, which introduces a new way of cooling these objects [24]. White dwarfs are stars where fusion has ended and they have an inactive, degenerate core. ALPs might also explain why the universe is surprisingly transparent to γ -rays with energies above 100 GeV. These high energy photons are expected to interact with the extragalactic background light, the starlight integrated over all epochs and starlight absorbed and re-emitted by dust in galaxies, and produce an electron-positron pair. If the source of these γ -rays is sufficiently far and the energy high enough, the universe is expected to be opaque to them. But observations seem to suggest that the universe is more transparent to these γ -rays than what the models predict. If an ALP exists, the γ -rays can interact with the extragalactic background light and produce an ALP. The ALP can then travel cosmological distances before converting back to a γ -ray. This mechanism would decrease the opacity of the universe to these high energy photons [25].

ALPs spanning a mass range tens orders of magnitude have been considered, from 10^{-32} GeV/ c^2 [26] up to 10^3 GeV/ c^2 [27]. ALPs are typically thought to couple to gauge bosons, mostly to photons. This is due to the mixing of the ALP with the SM pseudo-scalars, the π^0 and η , which predominately couple to photons [10]. ALPs that couple to fermions can lead to flavor changing processes and are thus more strongly constrained experimentally [28]. Models where the ALP couples exclusively to photons include two parameters. The ALP mass m_{ALP} and the ALP-photon coupling constant $g_{\gamma\gamma}$ [21].

For *BABAR*, the relevant ALP mass range is roughly 0.1 GeV/ c^2 to 10 GeV/ c^2 . Previous limits around this mass range have been set with beam dump experiments and with colliders. Published results from proton [29] and electron [21] beam dumps have been re-interpreted to set limits on the ALP phase space. NA64 is an electron beam dump experiment that performed a search for an ALP with the data they collected [30]. In these beam dump experiments, the proton or electron radiates a photon which interacts with a nucleus of a target and converts into an ALP via the Primakoff effect [31]. The ALP passes through an absorber, designed to absorb background SM particles, before finally decaying in a decay volume. The resulting photons can then be detected. Photons from a photon beam can interact with the nucleus of a target via the Primakoff process and produce an ALP. PrimEx [32]

analyzed such data to measure the π^0 photoproduction differential cross section. An example di-photon invariant mass histogram for one of the angular bins in the PrimEx paper is reinterpreted in [33] to set limits on the ALP parameter space.

Results from collider experiments have been used to set limits in the relevant ALP mass range. Published LEP2 results of $e^+e^- \rightarrow \gamma\gamma$ events are reinterpreted in order to search for an ALP [34]. The center of mass energy (CME) at LEP2 was much higher than the ALP masses considered here, causing the ALP to be highly boosted. The two photon daughters of the ALP hit the calorimeter close to each other and are reconstructed as a single photon. If the ALP is long lived and decays outside the detector, the signature of the event is a single photon and missing energy and momentum. This fact is used in [21] to reinterpret a *BABAR* search for dark photons $e^+e^- \rightarrow \gamma + \text{invisible}$ to set limits on the ALP [21]. Recently, the Belle II collaboration published [1] a very similar analysis to the one described in Chapter 3 which set limits on the ALP in the same mass range as the one considered in this work. The size of the dataset collected by Belle II is smaller than the one used in Chapter 3 and therefore the limits set here are more restrictive than the Belle II limits. At the highest energies relevant to this search, the CMS collaboration set limits on the ALP by analyzing fourteen events involving light-by-light ($\gamma\gamma \rightarrow \gamma\gamma$) scattering in lead ion collisions [35]. Light-by-light in the SM at leading order involves box diagrams with virtual charged particles. The cross section (CS) for these is suppressed because of the α^4 term, where α is the quantum electrodynamics coupling constant.

The *BABAR* detector and the PEP-II collider will be described in Chapter 2. Chapter 3 will describe a search for an ALP with the data collected by *BABAR*.

In searching for an ALP coupled to photons in a particle accelerator, it is important to be able to discriminate between photons and other particles. In most high energy particle physics detectors photons are measured in electromagnetic calorimeters which are typically made up of multiple cells, each one measuring the energy deposited by the electromagnetic shower initiated by the interaction of the incident particle with the detector material. The pattern of energies deposited in the cells of the calorimeter by a photon can be different from the energy pattern deposited by a different particle, a hadron, for example. This fact is used to develop discriminating variables between photons and hadrons, based on Zernike

moments and machine learning techniques, at the Belle II experiment. The Belle II detector, together with the SuperKEKB collider, are upgrades to the Belle detector and KEKB collider, respectively. The instantaneous luminosity of SuperKEKB is planned to reach almost 40 times higher than that of KEKB [36]. Belle II started taking data in 2018. Chapter 4 will describe work done to develop these new discriminating variables.

Chapter 5 presents the conclusions of this thesis.

Chapter 2

The *BABAR* experiment

The *BABAR* experiment operated between 1999 to 2008 at the SLAC National Accelerator Laboratory. The experiment is composed of the PEP-II collider and the *BABAR* detector. The experiment is described in detail in [5, 6, 37, 38].

The main goal of the *BABAR* experiment is the study of *CP* asymmetries in *B* meson decays. A secondary goal is to study other *B* physics, charm physics, tau physics, two photon physics, and *Y* physics. The detector is optimized for *CP* asymmetry studies, but also allows to perform research into the other types of physics.

Most of the data taken by *BABAR* was at a CME of 10.58 GeV corresponding to the mass of the $\Upsilon(4S)$. The $\Upsilon(4S)$ decays almost exclusively to $B^0\bar{B}^0$ and B^+B^- . This ensures a unique opportunity to study these *B* mesons. About 0.7×10^9 $B\bar{B}$ mesons pairs were collected throughout the life of the experiment. 6 % of the data was taken at a CME of the $\Upsilon(3S)$, and 3 % at a CME of the $\Upsilon(2S)$.

For each resonance, about 90 % of the data was taken at a CME as close as possible to the resonance mass. This is termed *On Peak*, or *on-resonance*. About 10 % of the data was taken *Off Peak*, or *off-resonance*, that is 25 MeV, 30 MeV, 40 MeV below the $\Upsilon(2S)$, $\Upsilon(3S)$, $\Upsilon(4S)$ masses, respectively. Off-resonance data was taken mostly for background studies and to study non-resonant processes.

2.1 The PEP-II collider

The PEP-II machine [39, 40] was an asymmetric electron-positron collider, meaning the electron and positron had different energies in the lab frame. The accelerator achieved the design luminosity in its first year of operation. Following a series of upgrades, the instantaneous luminosity increased to $1.2 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$, four times the designed luminosity. The integrated luminosity collected each day exceeded design by a factor of seven. A big factor in this increase was the transition to *trickle injection*, whereby the beam is replenished at the same time the detector collects data.

The collider operated at a CME range of 9.99 GeV to 11.2 GeV. For the $\Upsilon(4S)$ CME, the electron beam energy was 9 GeV and the positron beam energy was 3.1 GeV. The center of mass (CM) frame was thus boosted in the lab frame. This allows to reconstruct the decay vertexes of the B and \bar{B} mesons and measure their decays rates as a function of time. This type of measurement makes it possible to measure CP violation in $B^0\bar{B}^0$ mixing.

The accelerator ring was built out of six straight sections and six arc sections. The beam pipe contained two separate rings, one for the electrons and one for the positrons. Each ring was able to maintain about 1700 bunches, each bunch with a maximum charge of 8×10^{10} particles which amounts to 3 A of current.

2.2 The BABAR detector

In order to obtain the physics goals, the detector was designed to have a large and uniform acceptance. It was required to have good reconstruction efficiency for charged particles down to a momentum of 60 MeV/c. Photons were required to be efficiently reconstructed down to a momentum of 20 MeV/c. Good momentum resolution and good energy and angular resolution for photons in the range 20 MeV to 4 GeV is needed in order to be able to reconstruct π^0 mesons and η mesons. Additional requirements were good vertex resolution, and efficient muon, electron, and hadron identification. Figure 2.1 shows a schematic view of the BABAR detector along the beam direction and Figure 2.2 across the beam direction.

The detector is composed of a number of sub detectors. The silicon vertex tracker (SVT) and drift chamber (DCH) are at the innermost part of the detector.

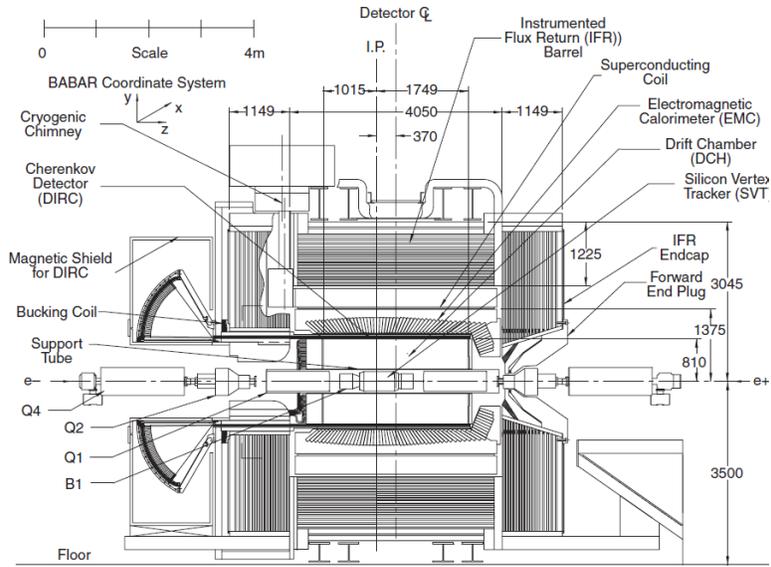


Figure 2.1: BABAR detector schematic along the beam direction. Numbers are distances in mm. Figure from [5], page 13, Copyright Elsevier (2002).

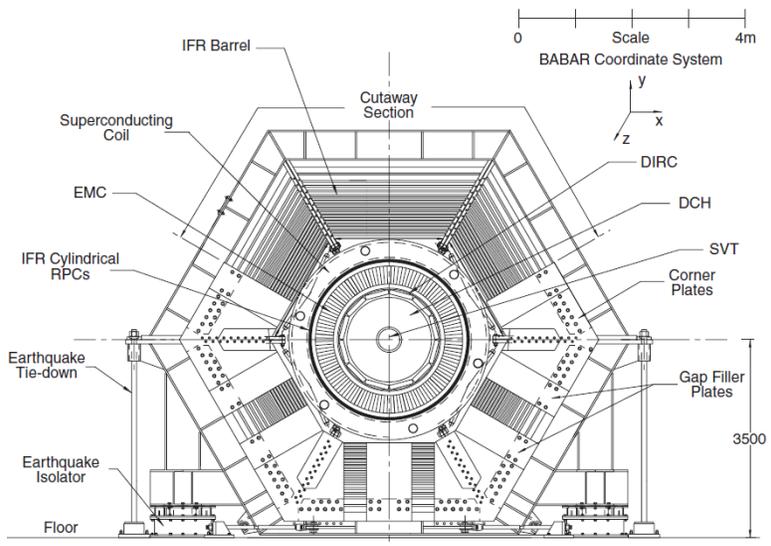


Figure 2.2: BABAR detector schematic across the beam direction. The detector center is 3500 mm above the floor. Figure from [5], page 14, Copyright Elsevier (2002).

They form the tracking system which is responsible for reconstructing tracks of charged particles. Further out are the detector for internally reflected Cherenkov light (DIRC), that separates pions from kaons and the electromagnetic calorimeter (EMC) that measures electromagnetic showers. These sub detector are surrounded by a superconducting solenoid magnet that produces a 1.5 T magnetic field. The steel flux return of the magnet is part of the instrumented flux return (IFR) sub detector intended to measure muons and neutral hadrons.

The *BABAR* coordinate system is anchored to the DCH. The z axis coincides with the DCH principal axis and points in the direction of travel of the electrons. The y axis points upward and the x axis points away from the center of the accelerator. The beam axis is offset from the z axis of *BABAR* by 20 mrad in order to minimize the effect of the solenoid magnetic field on the beam. The polar angle θ is measured from the z -axis. The azimuthal angle ϕ is measured from the x -axis in the $x - y$ plane.

2.2.1 Silicon vertex tracker

The SVT is used to measure charged particles close to the beam pipe. It provides the best measurement of the charged track angles. Figure 2.3 shows a schematic view of the SVT [6].

The SVT has five layers of double sided, 300 μm thick, silicon strip detectors. The inner three layers are used to measure the position and angle information of charged particles, while the outer two layers are primarily used to link SVT and DCH tracks. The spatial resolution of the inner three layers is 10 μm to 15 μm , and the resolution of the outer layers is 40 μm .

Each layer is made up of a number of modules. The modules of the three inner layers are planar shaped. The modules of the two outer layers are arch shaped. Each module has strips running parallel to the beam pipe on one side and perpendicular to the beam pipe on the other. The modules provide geometrical acceptance that covers 90 % of the solid angle in the CM frame.

Each strip is a p-n junction with a bias voltage of 10 V above the depletion voltage of 25 V to 35 V. When a charged particle passes through the strip, it ionizes the material forming electron-hole pairs. The electron-hole pairs travel in opposite

directions, because of the bias voltage applied to the strips, forming a current which is measured by the detector.

The hit efficiency, one of the most important metric of the SVT, was above 95 %.

2.2.2 Drift chamber

Surrounding the SVT is the drift chamber. As it is further away from the interaction point (IP), it can therefore measure the curvature of charged tracks and thus their momentum. It is the only detector that can reconstruct the interaction vertexes outside the SVT volume. This is useful for reconstructing K_S^0 , for example.

The DCH is composed of forty layers of hexagonal cells. Each cell has one sense wire in the center, held at a high voltage, and six field wires surrounding it, at ground potential. The DCH volume is filled with a 80:20 mixture of helium:isobutane. This decreases multiple scattering. The hit resolution for a single cell is $100\ \mu\text{m}$ for charged tracks passing near the middle of the cell. In order to provide longitudinal track information, the wires in 24 of the layers were at a small angle to the z axis.

A charged particle passing through the DCH volume ionizes the gas, creating free electrons and heavy ions. Due to the electric field, the electrons travel toward the sense wires and the ions travel to the field wires. When the electrons get close enough to the sense wire so that they gain enough kinetic energy between collisions with the gas molecules to ionize additional atoms, an avalanche is formed. The

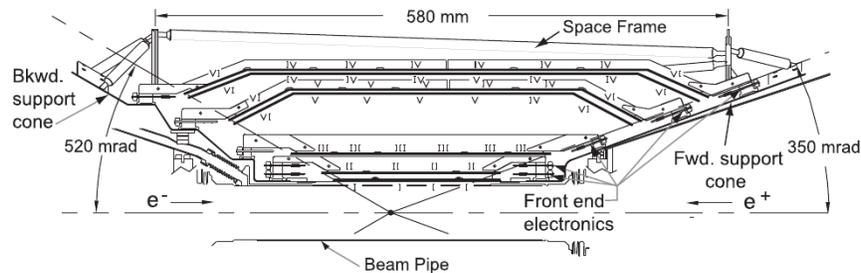


Figure 2.3: Schematic view of the SVT. Figure from [6], page 654, Copyright Elsevier (2013).

avalanche amplifies the electric charge by 5×10^4 at a voltage of 1960 V.

The DCH provides particle identification (PID) information by measuring the ionization loss, dE/dx , which is based on the total charge deposited in each cell. The DCH information is combined with the DIRC information to achieve better PID in the barrel region.

2.2.3 Detector of internally reflected Cherenkov light

The DIRC detector helps separate pions from kaons. Its principle of operation is the detection of Cherenkov radiation. Charged particles travelling faster than the speed of light in material emit Cherenkov light. The angle between the emitted photons and the direction of travel, the Cherenkov angle, depends on the particle speed and the refracting index of the material. Therefore, measuring the Cherenkov angle is equivalent to measuring the particle speed. Combining the particle's speed with its momentum allows to deduce its mass. The DIRC performs PID in the momentum range 0.5 GeV/c to 4.5 GeV/c. PID of particles with momentum below 700 MeV/c is mostly done by measurements of dE/dx in the DCH and SVT.

The DIRC is composed of 4.9 m long, 17 mm by 35 mm rectangular bars of fused silica. The bars act as emitters of Cherenkov light, as well as light guides to guide the light toward photo multiplier tubes (PMTs) via total internal reflection. The Cherenkov light is measured by a closely packed array of PMTs, located 1.2 m from the end of the bars. The PMTs measure the position and time of arrival of the photons. Figure 2.4 shows a schematic view of the path the Cherenkov radiation travels in the detector.

The Cherenkov angle is inferred from the location of the measured photons as well as from their time of arrival. The single photon Cherenkov angle resolution in the DIRC was 7 mrad.

2.2.4 Electromagnetic calorimeter

The EMC detects electromagnetic showers in the energy range 0.02 GeV to 9 GeV. This allows the EMC to detect low energy photons from π^0 mesons and η mesons and high energy photons from electromagnetic, weak, and radiative processes.

The EMC is composed of 6580 Thallium doped Cesium Iodide crystals. The

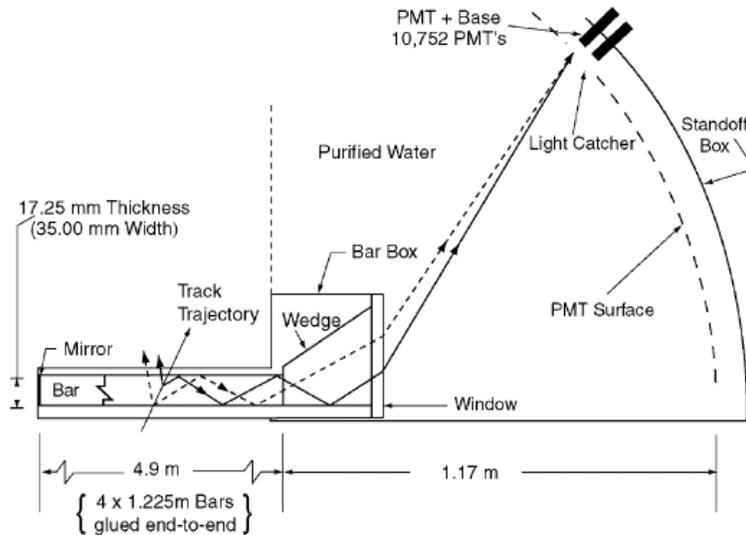


Figure 2.4: Schematic view of the path Cherenkov radiation travels inside the DIRC. Figure from [5], page 60, Copyright Elsevier (2002).

crystals form two separate structures. The barrel is a cylindrical structure concentric with the beam pipe made up of 5760 crystals arranged in 48 rings. The endcap is shaped as a circle, perpendicular to the beam pipe, in the direction of the electron beam. It is made up of 820 crystals arranged in 8 rings. Together, the EMC covers the polar angle range 15.8° to 141.8° in the lab frame, which translates to 90 % of the solid angle in the CM frame. Figure 2.5 shows a schematic view of the EMC along the beam pipe direction.

Each crystal has a trapezoidal cross section and a length of 30 cm. The cross section lengths vary, but are typically $4.7 \text{ cm} \times 4.7 \text{ cm}$ at the front and $6.1 \text{ cm} \times 6.0 \text{ cm}$ at the back. These dimensions are similar to the molière radius. Similarly to the DIRC, the crystals act as light guides. In addition, the crystals are wrapped by two layers of diffuse white reflector which reflect any transmitted light back into the crystals. Aluminum foil is wrapped around the reflector to create a Faraday cage. In order to isolate the aluminum from the metal support, the crystals were further wrapped by an insulating Mylar layer.

Light from the electromagnetic showers were measured by two $2 \text{ cm} \times 1 \text{ cm}$ silicon PIN diodes at the center of the back face of each crystal. To collect as much

information about the interacting particle, the amount of inactive material in front of and between the crystals was minimized.

Electromagnetic showers are created when a photon interacts with the crystal material and either creates an electron-positron pair directly, or causes the crystal atoms to scintillate and release scintillation light that then creates an electron-positron pair. These pairs radiate additional photons via the Bremsstrahlung process. The new photons produce more electron-positron pairs, and this process continues until the photons do not have enough energy to produce additional pairs.

The energy resolution depended on the incident photon energy. It varied from $\sigma_E/E = 5.0 \pm 0.8\%$ at an energy of 6.13 MeV, to $\sigma_E/E = 1.90 \pm 0.07\%$ at an energy of 7.5 GeV [5]. The angular resolution varied between 12 mrad at low energies to 3 mrad at high energies [5].

2.2.5 Instrumented flux return

The IFR was designed to identify muons and detect neutral hadrons, primarily K_L^0 and neutrons. It used the magnet's steel flux return as a muon filter and a hadron absorber. The steel was sectioned into layers that ranged in thickness between 2 cm to 10 cm. The nominal gap between the plates was 3.5 cm in the inner layers of the barrel and 3.2 cm in other parts of the IFR. Single gap resistive plate chambers (RPCs) were placed between the steel layers. Nineteen were placed in the barrel and eighteen in the end caps. Two additional layers of cylindrical RPCs with four readout planes were placed outside of the EMC in order to detect particles exiting the calorimeter.

The RPCs consisted of two planer electrodes, separated by a 2 mm gap, with a high voltage of about 8 kV applied to them. The gap was filled with a gas mixture of 57 % argon, 39 % freon and 4.5 % isobutane. Copper strips were attached to the outside of both sides of the electrodes. Electric charges inside the gap were read out capacitively via the strips. A particle passing through the RPC would ionize the gas, and similarly to the DCH, avalanches were formed. But because the electric field was high, during the formation of the avalanche photons were emitted with high enough energy to produce additional avalanches, amplifying the signal. This eliminated the need for costly electronic amplification and produced fast signals,

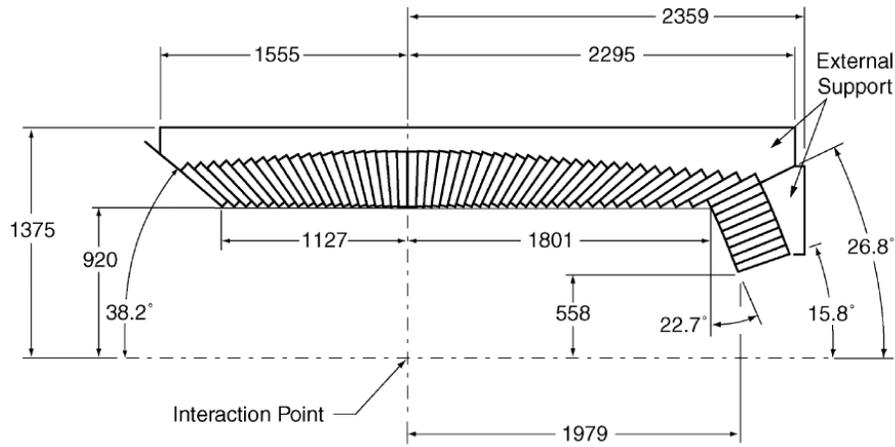


Figure 2.5: A schematic view of the top half of the EMC along the beam pipe direction. Distances are in mm. Figure from [5], page 72, Copyright Elsevier (2002).

but the signal amplitude depended only weakly on the incident particle energy.

The muon detection efficiency was 90 % in the momentum range 1.5 GeV/c to 3.0 GeV/c. The K_L^0 detection efficiency increased roughly linearly from 20 % to 40 % in the momentum range 1 GeV/c to 4 GeV/c.

The IFR underwent a number of upgrades over the years, because of aging effects that affected the chambers. In 2001 the RPCs in the forward endcap were replaced with an improved version, and in 2004 the barrel RPCs were replaced with limited streamer tubes (LSTs). The LSTs were square tubes with an anode wire passing through their center. The wire was parallel to the beam direction and measured the azimuthal coordinate, while strips on the outside of the tube, perpendicular to the beam, measured the z coordinate. In addition, six RPCs in the barrel were replaced with brass plates in order to increase the total absorber thickness.

Chapter 3

Search for axion like particles

3.1 Analysis overview

This chapter will describe a search for an ALP coupled exclusively to photons using the *BABAR* detector. The goal of the work described in this chapter is to check whether the *BABAR* data contains events involving an ALP, and if not, to set upper limits on the cross section and coupling constant of the ALP.

Two leading order Feynman diagrams contribute to the production of events with three photons in the final state. Figure 3.1a shows the ALP-strahlung process, where the ALP is radiated from a photon. Figure 3.1b shows the production mechanism through the decay of an Υ meson. The latter is important only in the $\Upsilon(2S)$ and $\Upsilon(3S)$ CMEs, where the branching ratios to ALPs of these resonances can be non negligible. The production CS through Υ is estimated to be 17 % of the combined production CS through ALP-strahlung and through the $\Upsilon(2S)$, at the $\Upsilon(2S)$ CME. For the $\Upsilon(3S)$, the ratio is 14 %. These assume the SuperKEKB machine parameters [41]. Nonetheless, the production mechanism through Υ is not taken into account in this work.

Section 3.2 describes the signal simulation, background simulation and data samples used in the analysis.

BABAR has a two back-to-back photon trigger and a trigger requiring at least one energetic photon. The trigger, together with the event selection process is described in Section 3.3. We look for events with three photons that add up to the

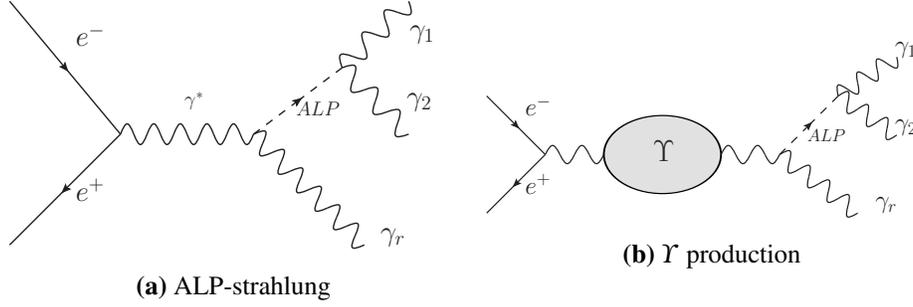


Figure 3.1: Feynman diagrams for the ALP production mechanisms.

CME. Selection cuts are applied in order to optimize the sensitivity of the search. Throughout the text, cut is used in the sense of retention. 5% of the data is used as a background sample. Several of the selection cut threshold values depend on the ALP candidate invariant mass. These threshold values are optimized for a number of ALP masses by counting signal and background candidates inside a counting window and maximizing the sensitivity to the signal. The thresholds are then generalized to any ALP mass with a smoothing function.

The selection cuts are applied and the ALP candidate mass distributions of data and simulations are compared in Section 3.4. It is found that the number of ALP candidates in data is around 80% that in simulations. The simulations are scaled to match the data and this scaling factor is incorporated into the signal efficiency and its associated systematic uncertainty.

Section 3.5 describes the production of toy experiments. These are used for validation, for determining what function should be used to describe the background shape, and for evaluating the global significance.

The signal extraction procedure is described in Section 3.6. We search for a bump on top of a smooth background in the ALP candidate invariant mass distribution. A binned maximum likelihood fit is performed to extract the number of signal entries inside a fit window that is about five times wider than the counting window. The Section describes the fit function used and how the shape of the background is described. Parts of the mass spectrum where the invariant mass distribution varies rapidly and which require too many degrees of freedom to describe properly are not used for the search. Fits to background only toy experiments are performed

in order to validate the fit method, to account for a bias because of how the toy experiments are produced, to determine what function is needed to describe the background, and to select which ALP mass hypotheses to use. In addition, fits to background toys together with entries from signal simulations are described. These are done in order to verify that the fit can pick up a signal bump, if it is present in the data, and to evaluate the signal efficiency.

Section 3.7 describes how the data is searched for evidence of an ALP. The Section describes the results of the local and global significances.

Bayesian credible interval (CI) upper limits on the ALP production CS and its coupling constant to photons are described in Section 3.8.4. This Section also describes systematic uncertainties incorporated into the limits.

The upper limit on the coupling constant is compared to limits from other experiments in a phase space exclusion plot in Section 3.9. Specifically, the results are compared with results from Belle II.

Finally, Section 3.10 describes how this analysis can be expanded in the future.

3.1.1 Disclaimer

The analysis was designed as a blind analysis which here means a measurement performed on a data sample that is not seen by the experimenter prior to finalizing the analysis procedure. Analyses that are not blind are prone to experimenter's bias. This bias can lead to the optimization of the selection cuts on statistical fluctuations of the dataset in a way that artificially enhances the significance of the signal producing a false signal peak. They might also produce artificially restrictive upper limits [42].

Initially, the measurement was meant to be performed on 95 % of the *BABAR* data, without looking at the sample before all other aspects of the analysis are agreed on and finalized. The remaining 5 % was meant to be used as a sample to help optimize the analysis and perform sanity checks. It was meant to be used as a background sample for the selection cut optimization, to examine how well the background Monte Carlo (MC) agrees with the data, to test the signal extraction procedure, and to test the limit setting procedure. The plan was to then request permission from the *BABAR* collaboration to perform the measurement on the re-

maintaining 95 % of the data. Such a request requires a collaboration wide review of the analysis details. *BABAR* stopped collecting data in 2008 and over the years fewer and fewer collaborators have been actively involved in the collaboration. Therefore, collaboration wide reviews of new results take a considerable amount of time, on the order of a full year. Because of these time considerations, the measurement using the blinded data is not performed here. Rather, all the results are based on the 5 % of the data.

It should be stressed, however, that this data sample is the same sample that is used to optimize the selection cuts. The analysis is therefore not a blind analysis and this fact should be kept in mind when evaluating the results presented here. Furthermore, because the analysis is not a blind analysis, the publication of these results was not pursued.

A further point to keep in mind is that because the analysis did not go through a *BABAR* collaboration wide review, the results are not results endorsed by the *BABAR* collaboration. Rather, they are results of data collected by *BABAR*.

3.2 Event samples

3.2.1 Signal MC simulation

Signal MC is generated with `MadGraph5_aMC@NLO` [43] version 2.6.1. `MadGraph5_aMC@NLO` does not simulate initial state radiation (ISR), therefore an ISR plugin for Madgraph is used [44]. The plugin modifies the momenta of the colliding e^+e^- such that they simulate the ISR process. It does not, however, produce the ISR photon itself. This is acceptable, as the ISR photon is typically produced at low or high polar angles, and is lost along the beam-pipe.

Two sets of signal MC are simulated. The first is used for optimizing the selection cuts. It includes ALPs with masses in the range $0.1 \text{ GeV}/c^2$ to $1 \text{ GeV}/c^2$, every $0.1 \text{ GeV}/c^2$, and in the range $1 \text{ GeV}/c^2$ to $10.5 \text{ GeV}/c^2$ every $0.5 \text{ GeV}/c^2$. In addition, ALPs with masses of $0.135 \text{ GeV}/c^2$, $0.548 \text{ GeV}/c^2$, and $0.958 \text{ GeV}/c^2$ are simulated. These are the masses of the SM π^0 , η , and η' , respectively. The number of simulated events is 2500, 2497, 49965 for the $\Upsilon(2S)$, $\Upsilon(3S)$, $\Upsilon(4S)$ CMEs, respectively. This set of signal MC simulations have no restriction on the polar angle

of the final state photons.

The second signal MC set is used to study the invariant mass distribution of the ALP photon daughters. It includes ALPs simulated every $0.01 \text{ GeV}/c^2$ in the mass range $0.1 \text{ GeV}/c^2$ to $10.02 \text{ GeV}/c^2$, $0.1 \text{ GeV}/c^2$ to $10.35 \text{ GeV}/c^2$, $0.1 \text{ GeV}/c^2$ to $10.49 \text{ GeV}/c^2$, for the $\Upsilon(2S)$, $\Upsilon(3S)$, $\Upsilon(4S)$ CMEs, respectively. The step size is chosen to be roughly five times smaller than the size of the base width for the ALP mass with the smallest base width. The base width is the width of the ALP candidate mass distribution at 10 % of its maximum, and is described in detail in Section 3.6.4. The number of simulated events is 10000, 9996, 49965 for the $\Upsilon(2S)$, $\Upsilon(3S)$, $\Upsilon(4S)$ CMEs, respectively. These numbers are selected to ensure that the number of ALP events that pass the selection cuts is at least five times larger than the expected number of ALP events in the full *BABAR* data set, assuming the current limits on the ALP coupling to photons. To reduce the computation time, the polar angle of the final state photons for this set of signal MC is limited to the range 24° to 156° in the CM frame. This range ensures that the photons are inside the acceptance of the EMC.

All of the signal MC is simulated with a coupling constant of $g_{\gamma\gamma} = 10^{-3} \text{ GeV}^{-1}$. This value is chosen as it is roughly the coupling constant that is expected to be excluded with the *BABAR* data. It is high enough to ensure that the ALP decays practically at the interaction point. With this coupling constant the ALP width is negligible compared to the detector resolution effects.

3.2.2 Background MC

Three background MC samples are used.

1. $e^+e^- \rightarrow \gamma\gamma(\gamma)$. This is the main background source for this analysis. It is generated with the BKQED [45] event generator.
2. $e^+e^- \rightarrow \omega\gamma$. This is the only SM resonance MC relevant to this analysis available at *BABAR*. The ω can decay to a $\pi^0\gamma$ and the π^0 can further decay to two photons. But as the π^0 is relatively boosted, its daughters can hit the EMC close to each other and be reconstructed as a single photon. Such events have four real photons, but only three reconstructed photons, the same number of

reconstructed photons of a signal event. This background is generated with the evtGen [46] event generator.

3. $e^+e^- \rightarrow e^+e^-\gamma$. If the electron and positron are not identified as charged tracks, they will be reconstructed as photons, mimicking a signal event. This background is generated with the BHWide [47] generator.

The response of the detector for all MC samples is simulated with Geant4 [48].

Table 3.1 shows the number of events for each of the background MC samples, and Table 3.2 lists the CSs. The CSs for the processes $e^+e^- \rightarrow \gamma\gamma(\gamma)$ and $e^+e^- \rightarrow e^+e^-\gamma$ are estimated by the respective generator. The CS for the process $e^+e^- \rightarrow \omega\gamma$ is estimated by formula 7 in [49] with updated mass and width values from [10]. Table 3.3 shows the ratio of the equivalent integrated luminosity of the different background MC samples and the integrated luminosity of the full *BABAR* dataset.

3.2.3 Data samples

This work uses 5% of the data collected with the *BABAR* detector. These events are semi-evenly interspersed throughout the complete dataset. A group of 10000 consecutive events is used, then 190000 consecutive events are skipped, then another 10000 events are used, and so on. The number of events analyzed and the associated integrated luminosity is detailed in Table 3.4.

Table 3.1: Number of MC background events.

Center of mass energy	$e^+e^- \rightarrow \gamma\gamma(\gamma)$	$e^+e^- \rightarrow \omega\gamma$	$e^+e^- \rightarrow e^+e^-\gamma$
$\Upsilon(2S)$ Off peak	10802×10^3	123×10^3	7500×10^3
$\Upsilon(2S)$ On peak	104846×10^3	1176×10^3	70988×10^3
$\Upsilon(3S)$ Off peak	27092×10^3	228×10^3	19000×10^3
$\Upsilon(3S)$ On peak	224158×10^3	2046×10^3	170412×10^3
$\Upsilon(4S)$ Off peak	21524×10^3	1079×10^3	38368×10^3
$\Upsilon(4S)$ On peak	214713×10^3	10707×10^3	403000×10^3

Table 3.2: Background MC CSs in nb.

Center of mass energy	$e^+e^- \rightarrow \gamma\gamma(\gamma)$	$e^+e^- \rightarrow \omega\gamma$	$e^+e^- \rightarrow e^+e^-(\gamma)$
$\Upsilon(2S)$ Off peak	3.32 ± 0.07	0.018	25.80 ± 0.07
$\Upsilon(2S)$ On peak	3.29 ± 0.07	0.018	25.80 ± 0.07
$\Upsilon(3S)$ Off peak	3.04 ± 0.06	0.018	25.77 ± 0.07
$\Upsilon(3S)$ On peak	3.01 ± 0.06	0.017	25.76 ± 0.07
$\Upsilon(4S)$ Off peak	2.87 ± 0.06	0.017	25.29 ± 0.07
$\Upsilon(4S)$ On peak	2.84 ± 0.06	0.017	25.07 ± 0.06

Table 3.3: Ratio of the equivalent integrated luminosities of background MC and the integrated luminosity of the full *BABAR* dataset.

Center of mass energy	$e^+e^- \rightarrow \gamma\gamma(\gamma)$	$e^+e^- \rightarrow \omega\gamma$	$e^+e^- \rightarrow e^+e^-(\gamma)$
$\Upsilon(2S)$ Off peak	2.29	4.71	0.20
$\Upsilon(2S)$ On peak	2.35	4.73	0.20
$\Upsilon(3S)$ Off peak	3.42	5.01	0.28
$\Upsilon(3S)$ On peak	2.67	4.22	0.24
$\Upsilon(4S)$ Off peak	0.17	1.47	0.03
$\Upsilon(4S)$ On peak	0.18	1.52	0.04

Table 3.4: Number of events in 5 % of the *BABAR* data and the associated integrated luminosity.

Center of mass energy	Number of events	Integrated luminosity (fb^{-1})
$\Upsilon(2S)$ Off peak	2200000	0.071
$\Upsilon(2S)$ On peak	25960000	0.68
$\Upsilon(3S)$ Off peak	4480000	0.13
$\Upsilon(3S)$ On peak	49060000	1.4
$\Upsilon(4S)$ Off peak	32160726	2.2
$\Upsilon(4S)$ On peak	335637695	21.0

3.3 Event selection

3.3.1 Candidate formation and preliminary selection cuts

The first set of selections cuts, the *preliminary selection cuts*, is applied while processing the events saved on the *BABAR* disks. The events that pass these selections are saved locally. The preliminary selection cuts are relatively loose and are meant to get rid of the bulk of the uninteresting events. Additional, tighter, selections are applied in later steps.

Each event is required to have no more than 3 tracks. The highest energetic photon in the event has to have an energy of at least 3 GeV in the CM frame. Candidates are formed from reconstructed photons. A candidate is a triplet of photons. Two out of these photons are assumed to be the ALP daughters and form the ALP candidate. The third photon is termed the *recoil photon*. There can be multiple candidates per event and the same three photons can produce three different candidates, each one with a different photon assuming the role of the recoil photon.

All the photons in a candidate are required to have a polar angle of more than 15.8° in the lab frame. This is to ensure that they are inside the acceptance of the electromagnetic calorimeter. The fact that *BABAR*'s EMC does not include a backward end cap, as discussed in Section 2.2.4, ensures that the reconstructed photons are sufficiently far from the beam pipe, and no additional selection cut is applied on the higher end of the reconstructed polar angle. The photons of the ALP candidate are required to have an energy greater than 0.7 GeV in the CM frame. The recoil photon needs to have an energy of at least 0.1 GeV in the lab frame. These are meant to reject beam background which is typically low energy photons. The invariant mass of the three photons needs to be in the range $8.5 \text{ GeV}/c^2$ to $12 \text{ GeV}/c^2$ in order to reject candidates which do not add up to the CME.

The three photons are assumed to originate from the IP and are kinematically fit [50]. The kinematic fit adjusts the photon's 4-vectors, taking into account their uncertainties, and constrains the corresponding candidate's 4-vector to the pre-collision 4-vector. In other words, the candidate's invariant mass is constrained to the CME, and its momentum is constrained to zero in the CM frame. In addition, the fitter takes into account the topology of the decay, i.e., it constructs the ALP

candidate out of two photons and then combines the ALP candidate with the third photon to form a candidate. This process improves the resolution of the ALP candidate mass, as it incorporates the information from the recoil photon as well as the CME of the collision. On the other hand, kinematic fitting “forces” some types of background candidates to look like signal candidates. For example, assume the event is a $e^+e^- \rightarrow \gamma\gamma$ with an unrelated additional beam background photon. Kinematically fitting this candidate will erroneously adjust the 4-vectors so that the invariant mass of the three photons is equal to the CME, thereby “forcing” it to resemble a signal candidate. Therefore, the un-adjusted 4-vectors, as well as the adjusted, constrained, 4-vectors after the kinematic fit are saved to disk. The unadjusted values are used for the selection cuts. The constrained ALP candidate mass is used for studies involving the di-photon mass distribution.

3.3.2 Trigger

Two types of triggers were used to select interesting events. During all of *BABAR*’s data taking a two photon trigger was available, named DigiFGammaGamma. This trigger requires exactly one pair of trigger clusters that each have 30 % of the CM energy and that are back-to-back in the CM frame, and no trigger tracks with transverse momentum greater than 0.25 GeV and consistent with originating from the IP.

In addition, two types of single photon triggers are used. These triggers are used only for data taken at the CME corresponding to the $\Upsilon(2S)$ and $\Upsilon(3S)$ and for the very end of the data taken at the $\Upsilon(4S)$ CME. The DigiFSingleGamma trigger requires at least one isolated cluster with at least 2 GeV of energy in the CM frame and no trigger tracks consistent with originating from the IP. The BGFSingleGammaInvisible is a high level trigger which has access to the EMC and DCH reconstructions and therefore to more accurate information. This trigger requires at most one cluster with at least 3 GeV energy in the CM frame and no tracks with momentum above 1 GeV/c. Events are selected if they pass any of the mentioned triggers.

Single photon trigger issue at high di-photon masses

There is a disagreement between MC and data for high invariant masses in the DigiFSingleGamma trigger. Due to an unknown reason, the trigger efficiency at high invariant masses drops in data, but not in MC. Figure 3.2 shows the di-photon invariant mass in the $e^+e^- \rightarrow e^+e^-\gamma$ MC sample and in data for the on-resonance $Y(2S)$ and $Y(3S)$ CMEs for events that pass the preliminary selection cuts discussed in Section 3.3.1. A possible explanation for this effect is that the DigiFSingleGamma trigger had a veto for high di-photon invariant mass events, presumably in order to prevent all DigiFGammaGamma triggers, which should be pre-scaled, to be accepted. If this veto existed, it is not properly documented at *BABAR*. To deal with this effect, this work only uses di-photon invariant masses up to 97.5 % of the CME for the $Y(2S)$ and $Y(3S)$ CMEs. There are no such issues for the other two triggers.

Trigger efficiency after preliminary selection cuts

The trigger efficiency after the preliminary selection cuts are applied is calculated by $n_{\text{trg}}/n_{\text{tot}}$, where n_{trg} is the number of triggered events after preliminary selection cuts, and n_{tot} is the total number of generated events. Figure 3.3 shows the trigger efficiency in signal MC after the preliminary selection cuts. As data taken at the $Y(2S)$ and $Y(3S)$ CMEs had the single photon trigger, the trigger efficiency is relatively constant down to low di-photon invariant masses. At low invariant masses, the trigger efficiency drops because the ALP is highly boosted and the two photons it decays to start to merge and are reconstructed as a single photon. In that case, there are not enough photons to properly reconstruct the decay. Therefore, the fall in efficiency is due to the preliminary selection cuts, not the trigger efficiency directly.

The efficiency of the DigiFGammaGamma trigger used for the $Y(4S)$ CME is more complicated. At very low di-photon invariant masses it has a similar shape to the trigger efficiency of the $Y(2S)$ and $Y(3S)$ CMEs. The efficiency rises as the invariant mass increases due to the ALP being less boosted and a lower chance of its daughter photons merging. At higher invariant masses the angle between the ALP daughter photons increases, and the efficiency decreases because the trigger

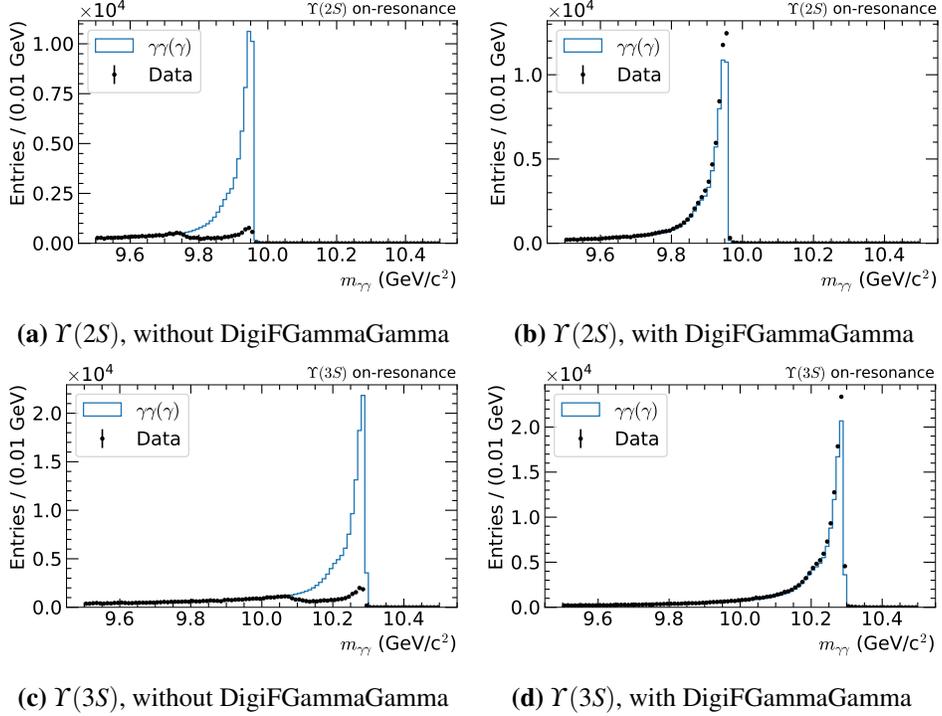


Figure 3.2: Data MC comparison for high invariant masses of candidates in $e^+e^- \rightarrow \gamma\gamma(\gamma)$ MC events that passed preliminary selection cuts. The top row shows $\Upsilon(2S)$ on-resonance, bottom row shows $\Upsilon(3S)$ on-resonance. The left column shows events where the DigiFGammaGamma trigger is not triggered. The right column shows events where DigiFGammaGamma is triggered.

reconstruction starts having 3 photons instead of 2 merged photons. At around a di-photon mass of $6 \text{ GeV}/c^2$ the trigger efficiency starts to drop off sharply, presumably due to some kinematic effect, as at this ALP mass the angle between the three photons is 120° . The efficiency drops down to around 1%. The efficiency starts to increase again at $6.5 \text{ GeV}/c^2$ in part due to one of the ALP daughter photons and the recoil photon being merged in the trigger reconstruction, leaving two back-to-back trigger clusters. These large variations in the efficiency for the $\Upsilon(4S)$ CME percolate throughout the analysis.

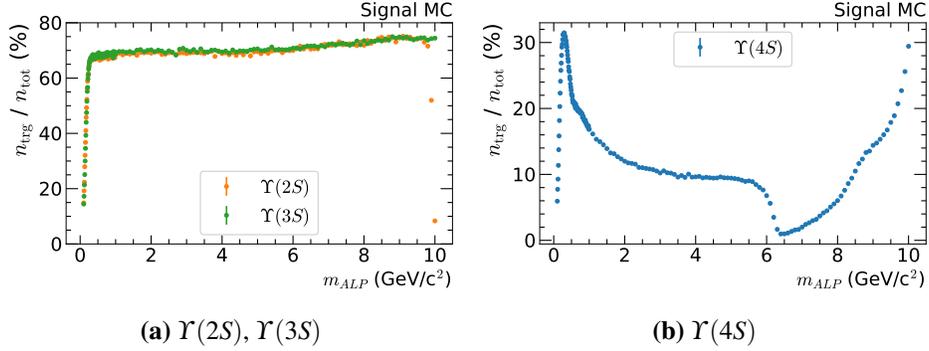


Figure 3.3: Trigger efficiency after preliminary selection cuts in signal MC of (a) $\Upsilon(2S)$ and $\Upsilon(3S)$ CMEs, (b) $\Upsilon(4S)$ CME. Note the different y-axis scales.

3.3.3 Main selection cuts

In addition to the preliminary selection cuts, additional selection cuts are applied in order to reject background events while keeping signal events. The following selections are used.

Number of charged tracks

The number of charged tracks in the event is required to be no more than one. One is allowed because there could be a beam background track from the beam-beam interactions or interactions of the beam with the beam pipe. Tightening the cut on the number of charged tracks compared to the preliminary selection increases the sensitivity at ALP masses below $0.5 \text{ GeV}/c^2$.

Polar angle of photons

A discrepancy is found between data and MC at low polar angles. This can be due to the detector material not being properly simulated in MC. To avoid this region, the polar angle of all photons is required to be larger than 22.5° . The fact that the EMC does not include a backward end cap ensures photons are reconstructed far enough away from the beam pipe in the backward region that no such discrepancies are seen for the higher values of the polar angles.

χ^2 of the kinematic fit

In order to reject candidates which are not consistent with the decay chain, or with the CME, the χ^2 of the kinematic fit involving the three photons is required to be smaller than a maximum value χ_{max}^2 , which depends on the constrained ALP candidate mass.

The reason the χ^2 is used and not the associated probability is that the photon energy distributions are not Gaussian, but rather have a long low energy tail. The long tail is mainly because of energy leakage out of the back of the crystals. The fitter assumes Gaussian uncertainties and therefore the χ^2 probability of the fit is practically zero for the large majority of fits.

ALP daughter photon energies

Beam background photons typically have low energies. To reject candidates involving beam background photons the energies of the ALP daughter photons in the CM frame are required to be larger than a minimum value $E_{\text{min}}^{\text{CM}}$, which depends on the constrained ALP candidate mass.

The energy of the recoil photon cannot be used for such a selection. This is because its energy in the CM frame can be shown to equal $(s - m_{\text{ALP}}^2)/(2\sqrt{s})$, where \sqrt{s} is the CME, and m_{ALP} is the ALP mass. Therefore, requiring a minimum value for the recoil photon energy imposes a maximum value for the ALP candidate mass. To be able to search for ALPs with a wide range of masses the selection cut on the recoil photon energy is kept loose.

Photons from beam background typically have reconstructed times that are not compatible with the collision time. Therefore, an additional method of reducing beam background is using a selection cut on the reconstructed photon times. Unfortunately, *BABAR* did not have photon timing information. Therefore, such a cut is not employed in this analysis.

Polar angle difference of the ALP daughter photons

The absolute value of the difference in polar angle of the two ALP daughter photons in the lab frame is required to be larger than a minimum value $\Delta\theta_{\text{min}}$. If any candidate fails this requirement, the whole event is rejected. This selection cut is

applied only to low mass ALP candidates.

This selection is meant to reject events where a photon converts into an e^+e^- pair via interaction with the detector material. If a photon converts and the resulting electron and positron are not identified as charged tracks for some reason, they will be reconstructed as photons. Because of the magnetic field of the *BABAR* detector, charged particles curve in the azimuthal angle. Therefore, these miss-reconstructed photons have different azimuthal angles, but a very similar polar angle. Requiring a minimum polar angle difference rejects this type of events.

Figure 3.4 shows a two dimensional histogram of the absolute value of the polar angle difference $|\Delta\theta|$, versus the di-photon invariant mass $m_{\gamma\gamma}$, of the ALP photon daughters after all selection cuts are applied, apart from the $|\Delta\theta|$ cut. There is a large population of ALP candidates with low $|\Delta\theta|$ angles, and invariant mass below $0.6 \text{ GeV}/c^2$. They are concentrated predominantly at masses of $0.15 \text{ GeV}/c^2$ and $0.3 \text{ GeV}/c^2$. The concentration of entries at those masses causes a peaking structure resembling a signal peak. These candidates occur due to conversions. The dashed blue line in Figure 3.4 shows the smoothed threshold cut value discussed later in this section. All entries below the line are discarded. The peaks can be seen in the blue curve of Figure 3.5 which shows the di-photon invariant mass after all selection cuts are applied, apart from the $|\Delta\theta|$ cut. The orange curve in Figure 3.5 shows the same distribution with the $|\Delta\theta|$ cut applied. The peaking structures are removed by this selection cut.

The entries are concentrated at these masses because of interactions with a support tube and with the inner wall of the drift chamber which are at 22 cm to 24 cm from the interaction point in the transverse direction. Due to the magnetic field, the electron positron pair splits by the same amount in the azimuthal angle. Therefore, the invariant mass of these is relatively constant and creates the peak.

No photon pair back-to-back in azimuthal angle

Candidates with any photon pair back-to-back in the azimuthal angle are rejected. The photons of the candidate are required to satisfy $|\Delta\phi^{\text{CM}} - 180^\circ| > \Delta\phi_{\text{min}}^{\text{CM}}$, where $\Delta\phi^{\text{CM}}$ is the azimuthal angle difference in the CM frame and $\Delta\phi_{\text{min}}^{\text{CM}}$ is the minimum threshold. This selection cut is meant to reject $e^+e^- \rightarrow \gamma\gamma\gamma$ events where one of

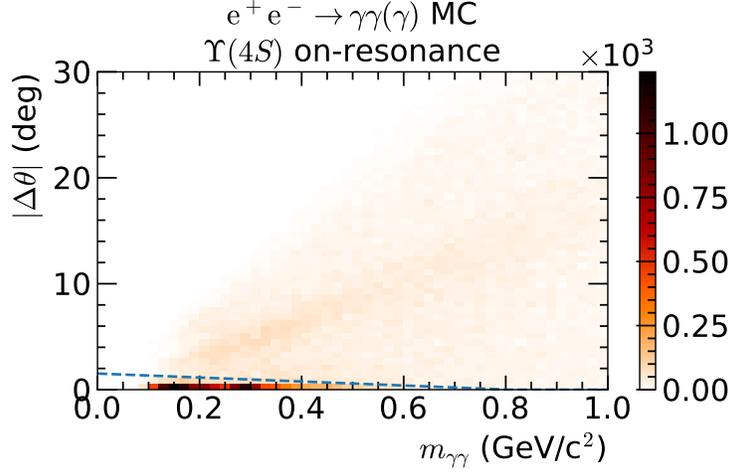


Figure 3.4: 2D histogram of the absolute value of the polar angle difference $|\Delta\theta|$ versus the di-photon invariant mass $m_{\gamma\gamma}$ of the ALP photon daughters after all selection cuts are applied, apart from the $|\Delta\theta|$ cut. The dashed blue line shows the smoothed threshold cut value. All entries below the dashed line are discarded. The plot is for the $e^+e^- \rightarrow \gamma\gamma(\gamma)$ MC.

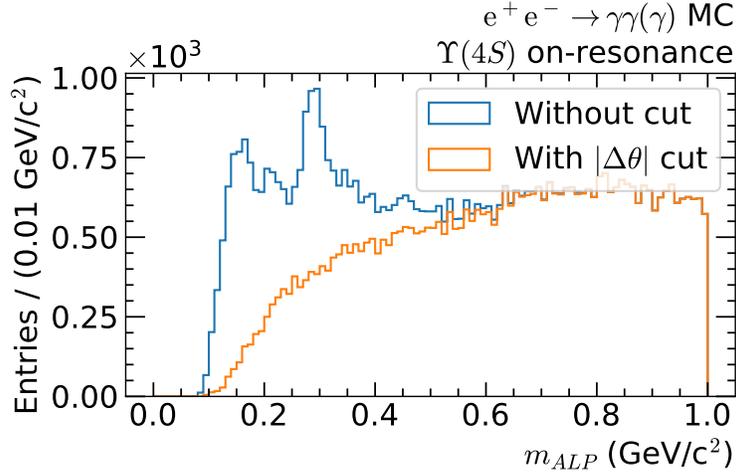


Figure 3.5: Distributions of $m_{\gamma\gamma}$ with and without the $|\Delta\theta|$ selection cut for the $e^+e^- \rightarrow \gamma\gamma(\gamma)$ MC.

the photons is lost along the beam pipe and the remaining photons are combined with a beam background photon to construct a candidate. As the beam direction is almost along the z axis, the two photons belonging to the collision are back-to-back in the azimuthal angle. This selection cut is only effective for high invariant mass ALP candidates, and is therefore only applied for those candidates.

Selection cuts that were not used

Other selection cuts that were assessed but did not improve the sensitivity are the maximum polar angle, optimizing the unconstrained three photon invariant mass limits, the helicity angle, and the maximum angle between pairs of photons in the CM frame.

3.3.4 Optimization of selection cut values

For the selection cut thresholds that depend on the ALP mass, the thresholds are optimized to maximize the sensitivity S/\sqrt{B} , where S is the number of signal candidates and B is the number of background candidates expected in the full *BABAR* dataset. S and B are counted in a counting window, i.e., in a range of di-photon masses. The counting window width and its center depend on the ALP mass. The 5% data sample is used as the background sample. This is done to avoid any discrepancies between MC and data.

One could be concerned that an ALP exists in the data and because the data is used as the background sample, the selection cut threshold optimization procedure somehow filters out ALP events from the data. In fact, for each selection cut variable, the optimization procedure only removes background components that are different from the signal distribution. In other words, if the signal and background distribution are identical for a particular selection variable, the optimization procedure will not remove any events.

If an ALP exists and has a mass inside one of the counting windows, it can affect the selection cut threshold values for one single ALP mass hypothesis. But, as discussed in Section 3.3.4, the threshold values are later smoothed. The smoothing is expected to wash out the effect of an ALP in data on the selection cut threshold values.

A disadvantage of this method is that it can bias the selection cut threshold value based on statistical fluctuations in areas with a low number of entries. This can be seen, for example, in Figure 3.10 which shows the distribution of the χ^2 of the kinematic fit for signal and background. The selected threshold is in the tail of the distributions, where there are a small number of entries per bin and the exact location of the threshold value depends on where exactly these statistical fluctuations appear. This effect will be adjusted, at least partially, by the smoothing of the selection cut threshold values discussed in Section 3.3.4.

Counting window

The number of signal and background entries are counted inside a counting window of the di-photon mass. The di-photon mass distribution varies depending on the ALP masses. Figure 3.6 shows the di-photon mass for a $0.5 \text{ GeV}/c^2$ and a $10.3 \text{ GeV}/c^2$ ALP. As can be seen, the distribution for the $10.3 \text{ GeV}/c^2$ ALP has a long high end tail, while the $0.5 \text{ GeV}/c^2$ ALP is more symmetrical, with a small low side tail.

The counting window is chosen so that at its edges the distribution values are 10 % of the maximum bin height. Values between bins are linearly interpolated. The dashed horizontal line in Figure 3.6 is at 10 % of the maximum bin height. The solid vertical lines show the counting window edges. For a Gaussian with standard deviation σ , 10 % of the maximum occurs at $\pm 2.15 \sigma$ from the mean.

The counting window is used only to optimize the selection cut threshold values. The actual number of signal entries is evaluated with a dedicated fit to the di-photon invariant mass, as will be discussed in Section 3.6.

Optimal selection cut threshold values

For each selection cut that is optimized, different threshold values are tested. For each threshold value, the selection cut is applied, and the sensitivity S/\sqrt{B} is calculated. The threshold value that maximizes the sensitivity is chosen. The next selection cut is then optimized, while applying the first selection cut as well. This process is continued until all threshold values are optimized. As applying one selection cut can change the optimal value of a previously optimized selection cut,

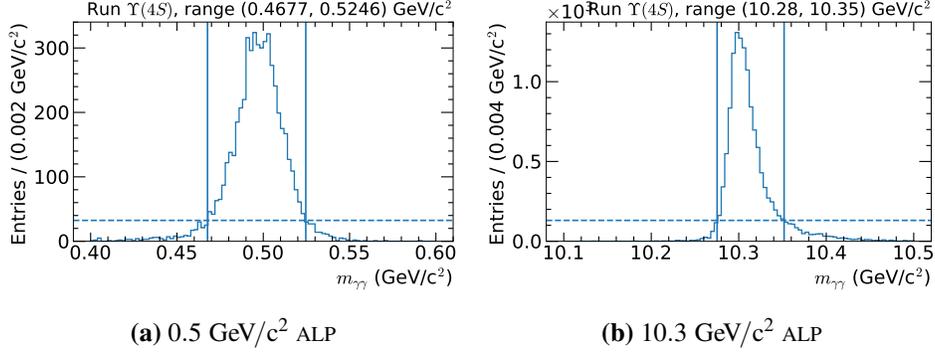


Figure 3.6: Signal MC ALP candidate mass distributions for 0.5 GeV/c² and 10.3 GeV/c² ALP masses. Preliminary and trigger selections are applied. The dashed horizontal line represent 10% of the maximum bin height. The solid vertical lines show the counting window range. They are at the interpolated intersection of the dashed line with the histogram. The counting ranges are shown in the titles.

after all selection cuts are optimized the whole procedure is repeated two more times, three times in total, while applying the optimal values that were found in the previous iteration. This whole procedure is applied for each generated ALP mass.

Figures 3.7 to 3.15 show the results for the 0.5 GeV/c² ALP and Figures 3.16 to 3.24 for the 10.3 GeV/c² ALP, both at the $\Upsilon(4S)$ CME. The 0.5 GeV/c² results include the selection cut on $|\Delta\theta|$ which is only applied to low di-photon invariant masses, while the results for the 10.3 GeV/c² ALP include the selection cut to prevent photon pairs being back-to-back in the azimuthal angle, which is only used for high di-photon invariant masses. For each selection cut there is a plot comparing the distributions of signal MC and data after all selection cuts are applied apart from the one being considered, distributions after all the selection cuts are applied and a scatter plot of the S/\sqrt{B} versus the threshold selection cut value. All the plots are after the final round of optimization.

The distributions of signal and background for the χ^2 selection cut of the kinematic fit are relatively similar to each other because the remaining background after the other selection cuts are applied is composed mainly of $e^+e^- \rightarrow \gamma\gamma\gamma$. The 4-vectors of the candidates of this background add to the pre-collision 4-vector, just like the signal.

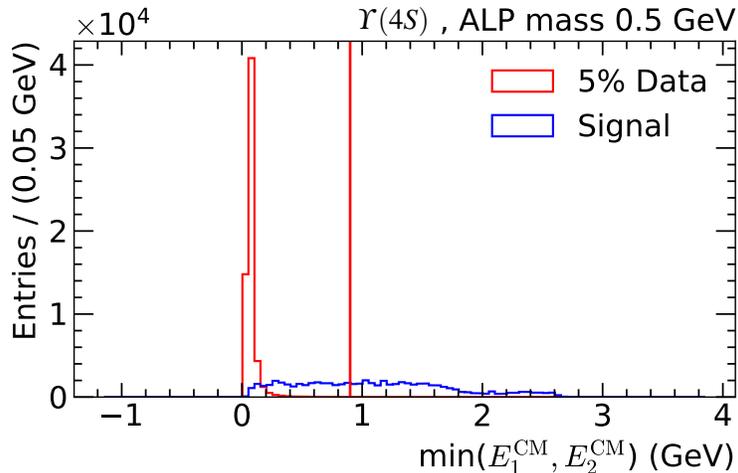


Figure 3.7: Distribution of the minimum energy of the ALP daughters in the CM frame for signal MC and data for a $0.5 \text{ GeV}/c^2$ ALP at the $Y(4S)$ CME after three rounds of optimization after all other selection cuts are applied. The signal distribution is normalized to have the same area as the data. The vertical red line signifies the optimal selection cut threshold E_{\min}^{CM} .

Smoothed threshold selection cut values

The procedure so far has obtained optimal selection cut values at specific generated ALP masses. In order to apply selection cuts for any ALP candidate mass, the optimal threshold cut values were generalized to any ALP candidate mass by a smoothing function. The smoothing functions are chosen to be piecewise linear.

ALP masses for which the background sample has less than 100 candidates in the counting window after all selection cuts are applied are not used.

Figures 3.25, 3.26, 3.27 and 3.28 show the optimal threshold values and the smoothed functions for χ_{\min}^2 , E_{\min}^{CM} , $\Delta\theta_{\min}$, $\Delta\phi_{\min}^{\text{CM}}$, respectively, for the combined $Y(2S)$, $Y(3S)$ runs and for the $Y(4S)$ runs separately. Figure 3.26 has two different smoothing functions which will be discussed below.

The smoothing functions are not chosen to be exact. They are not fitted. Rather, they are chosen to be close to the optimal values, while avoiding sharp transitions. Sharp transition, such as a step function, are found to make the di-photon mass

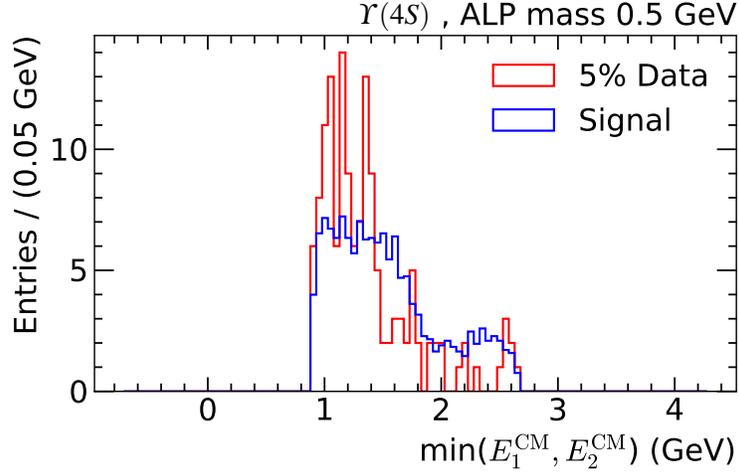


Figure 3.8: Distribution of the minimum energy of the ALP daughters in the CM frame for signal MC and data for a $0.5 \text{ GeV}/c^2$ ALP at the $\Upsilon(4S)$ CME after three rounds of optimization after all selection cuts are applied, including the minimum energy cut. The signal distribution is normalized to have the same area as the data.

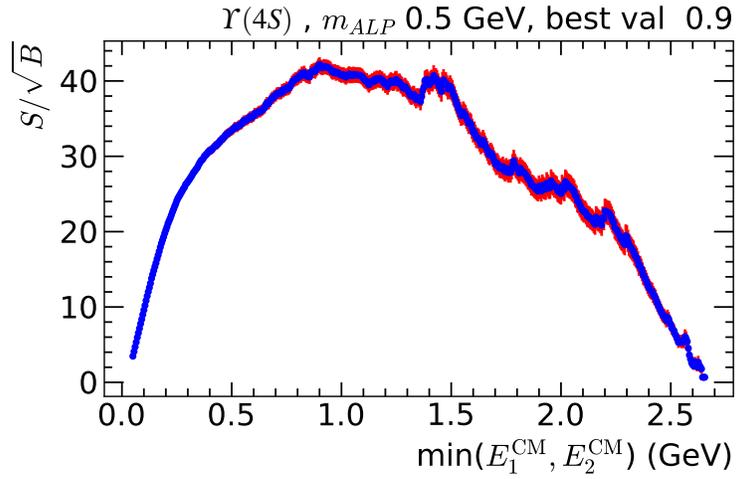


Figure 3.9: Scatter plot of S/\sqrt{B} versus E_{\min}^{CM} for $\Upsilon(4S)$ CME after three rounds of optimization for a $0.5 \text{ GeV}/c^2$ ALP. Error bars are colored red. The best threshold value is noted at the top of the plot.

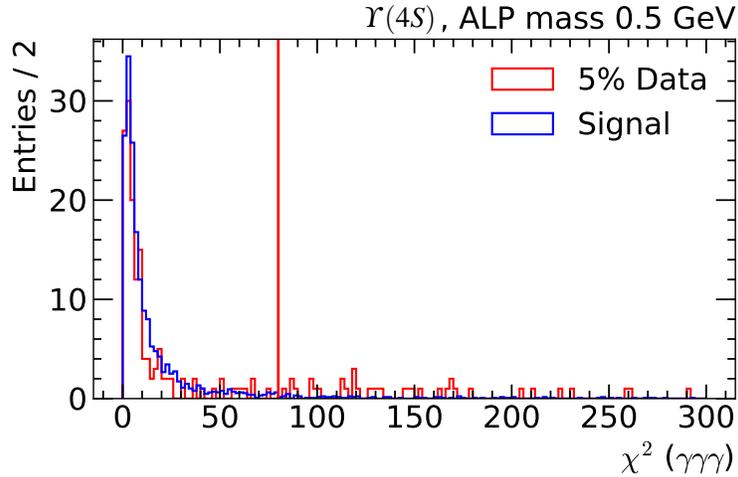


Figure 3.10: Distribution of the χ^2 of the kinematic fit, for signal MC and data for $0.5 \text{ GeV}/c^2$ ALP at the $\Upsilon(4S)$ CME after three rounds of optimization after all other selection cuts are applied. The signal distribution is normalized to have the same area as the data. The vertical red line signifies the optimal selection cut threshold χ_{\min}^2 .

distribution non-smooth. For example, the maximum smoothed function value for χ_{\min}^2 in Figure 3.25 is 100 as it is a nice round number close to the optimal ones.

A significant deviation of the smoothed function from the optimal values can be seen at high ALP masses in Figure 3.26b. The reason the optimal values increase rapidly with ALP mass above $9 \text{ GeV}/c^2$ is artificial. The signal and background distributions of the minimum energy of the ALP daughter photons is very similar for those ALP masses. Because of this, the optimized threshold value found during the optimization procedure is at the lower edge of the distributions, so that few candidates are rejected. This can be seen in Figure 3.16, for example. Therefore, the smoothed function does not follow the optimal values, but instead is chosen to continue with the same slope determined by lower ALP mass optimal values.

Very few mass points for the $|\Delta\theta_{\min}|$ variable for the combined $\Upsilon(2S)$, $\Upsilon(3S)$ dataset have enough background candidates left after the selection cuts are applied, as can be seen in Figure 3.27a. Therefore, the smoothed function for $|\Delta\theta_{\min}|$ for the $\Upsilon(2S)$, $\Upsilon(3S)$ CMEs is chosen to be the same as the function for the $\Upsilon(4S)$

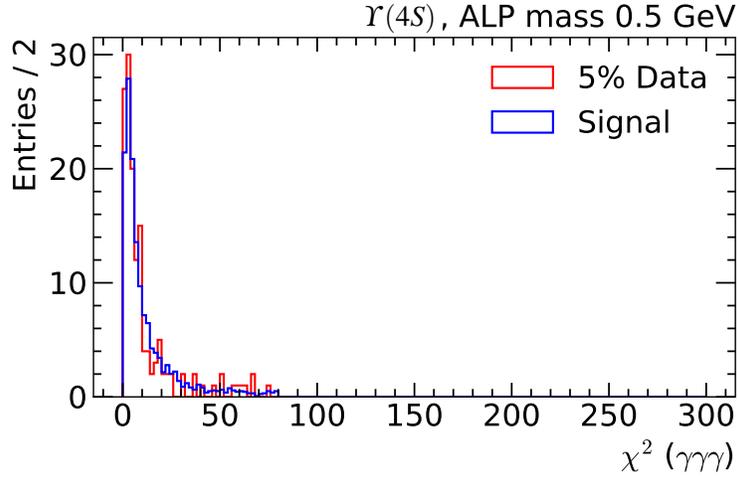


Figure 3.11: Distribution of the χ^2 of the kinematic fit, for signal MC and data for $0.5 \text{ GeV}/c^2$ ALP at the $Y(4S)$ CME after three rounds of optimization after all other selection cuts are applied, including the χ^2 selection cut. The signal distribution is normalized to have the same area as the data.

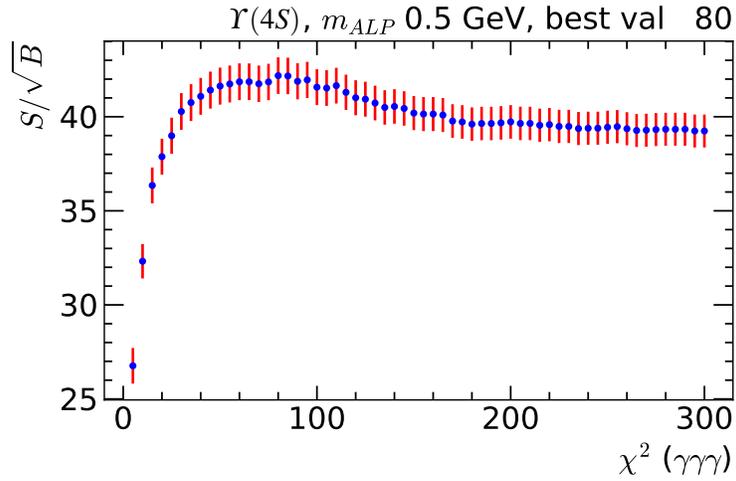


Figure 3.12: Scatter plot of S/\sqrt{B} versus χ^2_{\min} for $Y(4S)$ CME after three rounds of optimization for a $0.5 \text{ GeV}/c^2$ ALP. Error bars are colored red. The best threshold value is noted at the top of the plot.

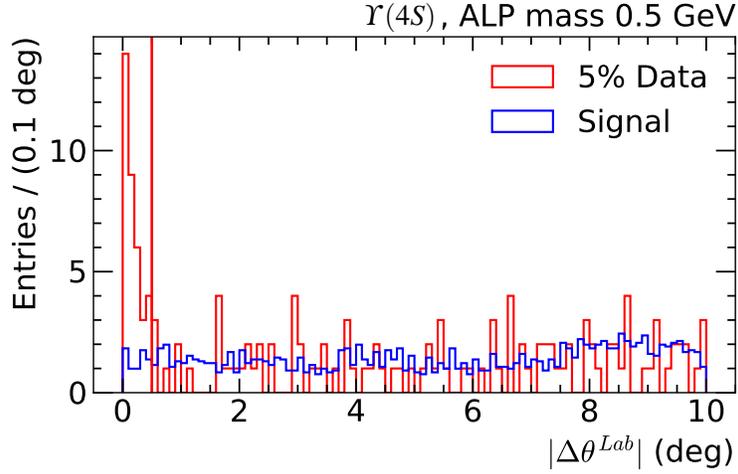


Figure 3.13: Distribution of the $|\Delta\theta|$, for signal MC and data for 0.5 GeV/c² ALP at the $\Upsilon(4S)$ CME after three rounds of optimization after all other selection cuts are applied. The signal distribution is normalized to have the same area as the data. The vertical red line signifies the optimal selection cut threshold $\Delta\theta_{\min}$.

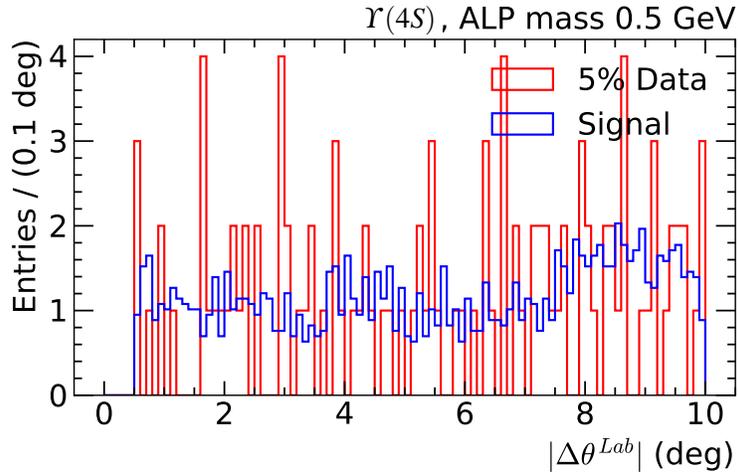


Figure 3.14: Distribution of the $|\Delta\theta|$, for signal MC and data for 0.5 GeV/c² ALP at the $\Upsilon(4S)$ CME after three rounds of optimization after all other selection cuts are applied, including the χ^2 selection cut. The signal distribution is normalized to have the same area as the data.

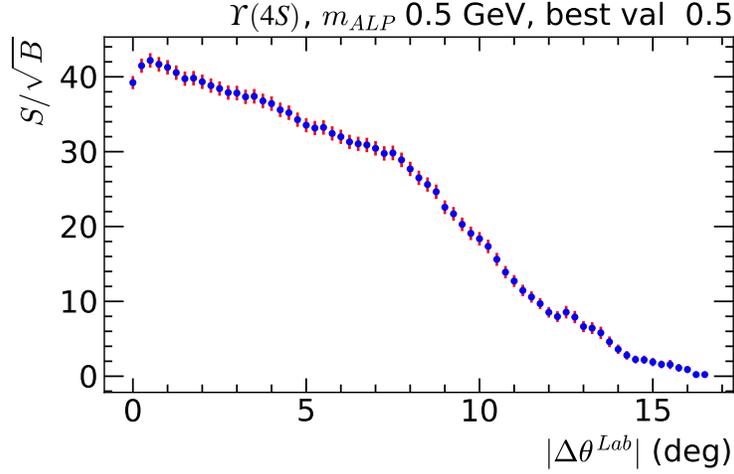


Figure 3.15: Scatter plot of S/\sqrt{B} versus $\Delta\theta_{\min}$ for $Y(4S)$ CME after three rounds of optimization for a $0.5 \text{ GeV}/c^2$ ALP. Error bars are colored red. The best threshold value is noted at the top of the plot.

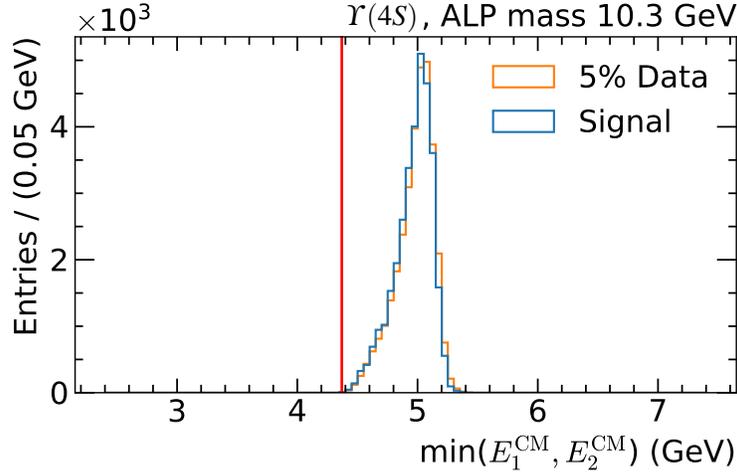


Figure 3.16: Distribution of the minimum energy of the ALP daughters in the CM frame for signal MC and data for a $10.3 \text{ GeV}/c^2$ ALP at the $Y(4S)$ CME after three rounds of optimization after all other selection cuts are applied. The signal distribution is normalized to have the same area as the data. The vertical red line signifies the optimal selection cut threshold.

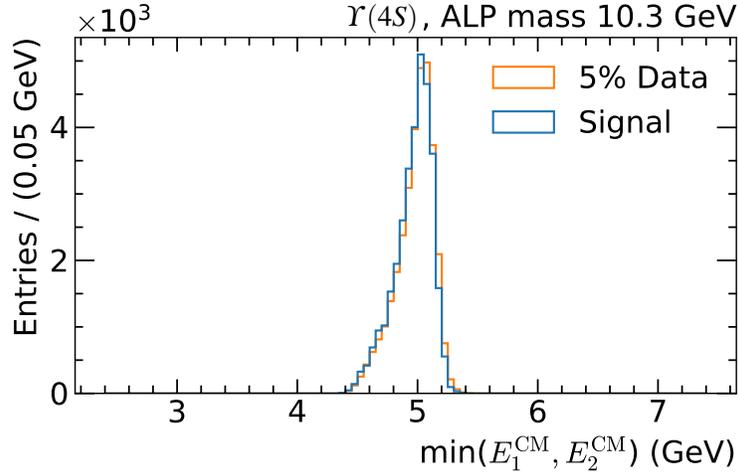


Figure 3.17: Distribution of the minimum energy of the ALP daughters in the CM frame for signal MC and data for a $10.3 \text{ GeV}/c^2$ ALP at the $\Upsilon(4S)$ CME after three rounds of optimization after all selection cuts are applied, including on the ALP daughter photon energies. The signal distribution is normalized to have the same area as the data.

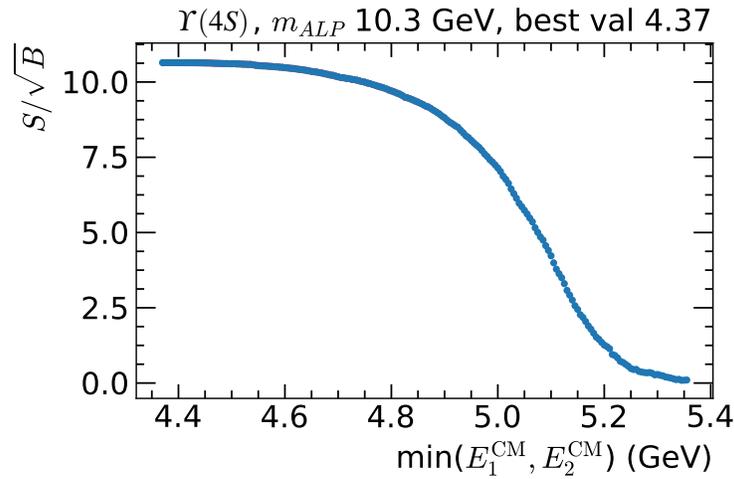


Figure 3.18: Scatter plot of S/\sqrt{B} versus E_{\min}^{CM} for $\Upsilon(4S)$ CME after three rounds of optimization for a $10.3 \text{ GeV}/c^2$ ALP. Error bars are colored red. The best threshold value is noted at the top of the plot.

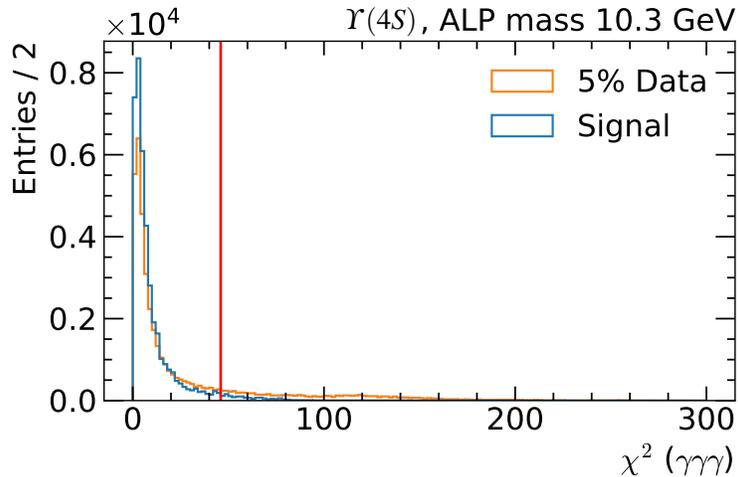


Figure 3.19: Distribution of the χ^2 of the kinematic fit, for signal MC and data for 10.3 GeV/c² ALP at the $Y(4S)$ CME after three rounds of optimization after all other selection cuts are applied. The signal distribution is normalized to have the same area as the data. The vertical red line signifies the optimal selection cut threshold χ_{\min}^2 .

CME.

The selection cut for rejecting pairs of photons that are back-to-back in the azimuthal angle are beneficial only for the $Y(4S)$ CME for ALP candidate masses above 10.3 GeV/c². To avoid sharp transitions, the smoothed function for $\Delta\phi_{\min}^{\text{CM}}$ is gradually increased, starting at a mass of 10 GeV/c², as can be seen in Figure 3.28.

Smoothing di-photon mass distribution for the $Y(2S)$, $Y(3S)$ CMEs

After applying all the selection cuts the ALP candidate mass distribution has a relatively broad peaking feature around 5.5 GeV/c² for the $Y(2S)$ and $Y(3S)$ CMEs. This can be seen in the blue curves in Figure 3.29. It is more pronounced for the $Y(3S)$ CME.

It is desirable that this mass distribution be as smooth as possible as will be discussed in Section 3.6.3. The peak is created because the slope of the E_{\min}^{CM} selection cut changes at 5.5 GeV/c². It turns out that the values of this selection cut that optimize S/\sqrt{B} reject too many ALP candidates above 5.5 GeV/c² which causes a

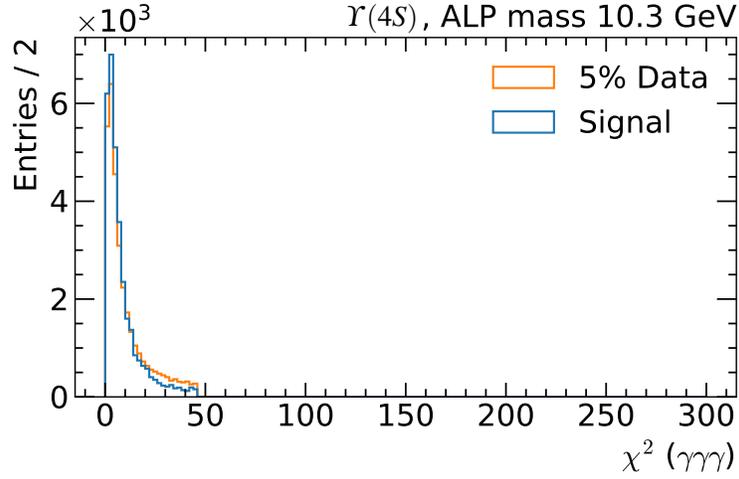


Figure 3.20: Distribution of the χ^2 of the kinematic fit, for signal MC and data for $10.3 \text{ GeV}/c^2$ ALP at the $\Upsilon(4S)$ CME after three rounds of optimization after all other selection cuts are applied, including the χ^2 selection cut. The signal distribution is normalized to have the same area as the data.

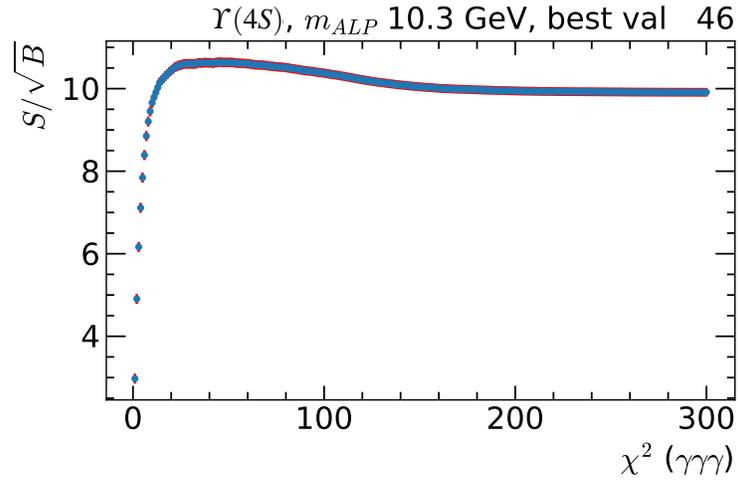


Figure 3.21: Scatter plot of S/\sqrt{B} versus χ^2_{\min} for $\Upsilon(4S)$ CME after three rounds of optimization for a $10.3 \text{ GeV}/c^2$ ALP. Error bars are colored red. The best threshold value is noted at the top of the plot.

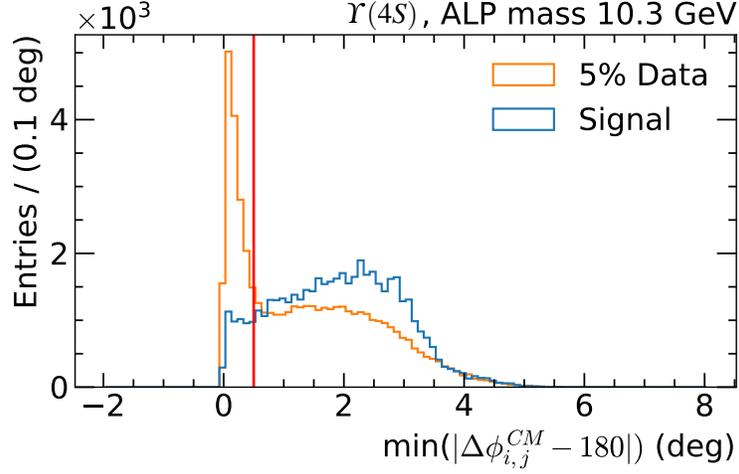


Figure 3.22: Distribution of the $|\Delta\phi^{\text{CM}} - 180^\circ|$, for signal MC and data for a $10.3 \text{ GeV}/c^2$ ALP at the $\Upsilon(4S)$ CME after three rounds of optimization after all other selection cuts are applied. The signal distribution is normalized to have the same area as the data. The vertical red line signifies the optimal selection cut threshold.

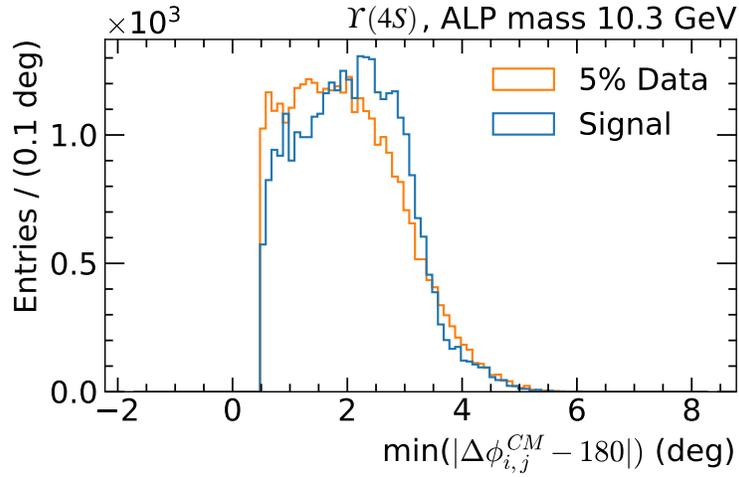


Figure 3.23: Distribution of the $|\Delta\phi^{\text{CM}} - 180^\circ|$, for signal MC and data for a $10.3 \text{ GeV}/c^2$ ALP at the $\Upsilon(4S)$ CME after three rounds of optimization after all selection cuts are applied. The signal distribution is normalized to have the same area as the data.

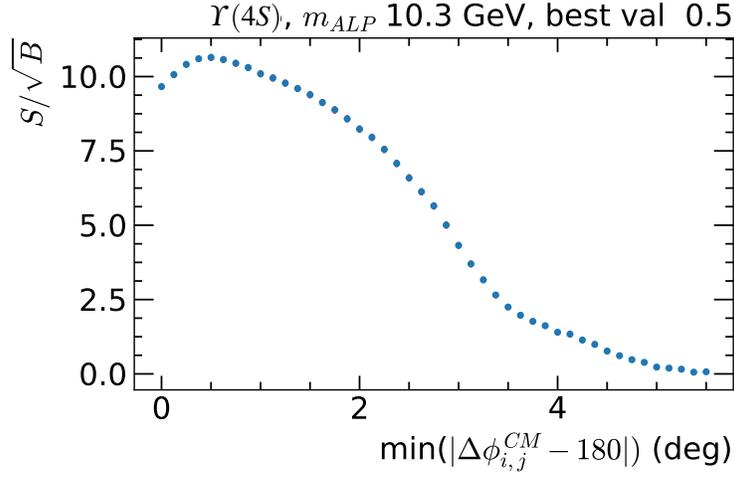


Figure 3.24: Scatter plot of S/\sqrt{B} versus $\Delta\phi_{\min}^{CM}$ for $\Gamma(4S)$ CME after three rounds of optimization for a $10.3 \text{ GeV}/c^2$ ALP. Error bars are colored red. The best threshold value is noted at the top of the plot.

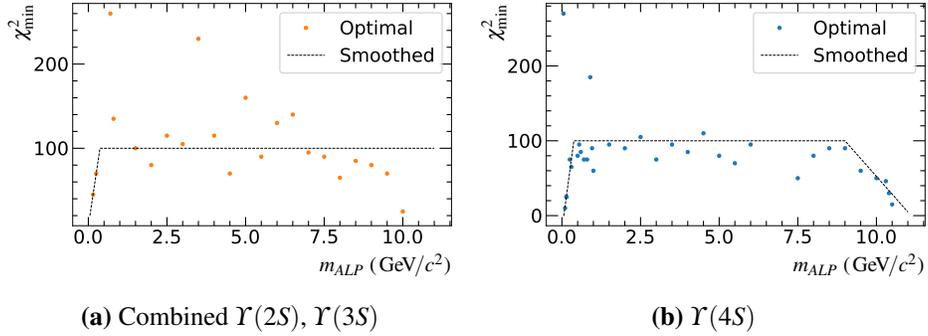


Figure 3.25: Optimal χ_{\min}^2 values and the chosen smoothed function for (a) combined $\Gamma(2S), \Gamma(3S)$ and (b) $\Gamma(4S)$ CMEs.

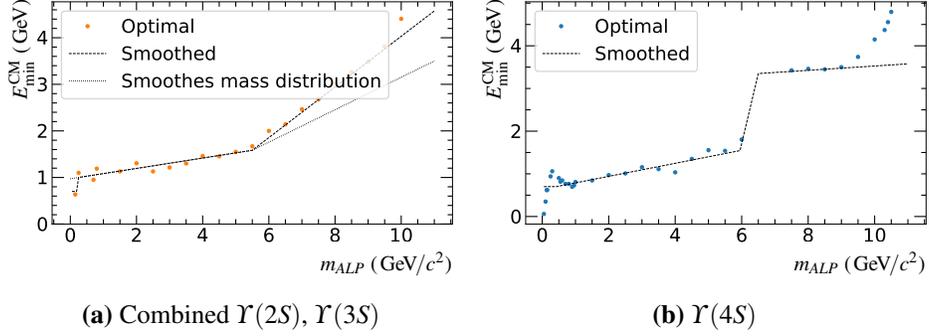


Figure 3.26: Optimal E_{\min}^{CM} values and the chosen smoothed function for (a) combined $Y(2S)$, $Y(3S)$ and (b) $Y(4S)$ CMEs.

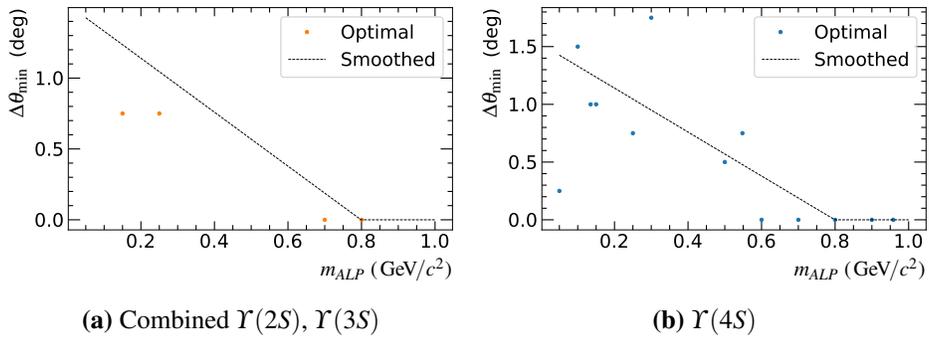


Figure 3.27: Optimal $\Delta\theta_{\min}$ values and the chosen smoothed function for (a) combined $Y(2S)$, $Y(3S)$ and (b) $Y(4S)$ CMEs.

trough in the mass range $5.5 \text{ GeV}/c^2$ to $8 \text{ GeV}/c^2$. A smoother mass distribution is preferred in this work over optimizing S/\sqrt{B} . Therefore, an alternate smooth function is used for E_{\min}^{CM} .

The alternate function can be seen in the dotted line of Figure 3.26a. This function is selected empirically because it creates a smoother ALP candidate mass distribution. The green curves in Figure 3.29 are obtained by applying the alternate smoothed function.

Figure 3.30 shows a two dimensional histogram of E_{\min}^{CM} versus the ALP candidate mass. Three sets of smoothed E_{\min}^{CM} functions are superimposed. The original in blue, the chosen alternate in green and a third alternate in orange. The orange

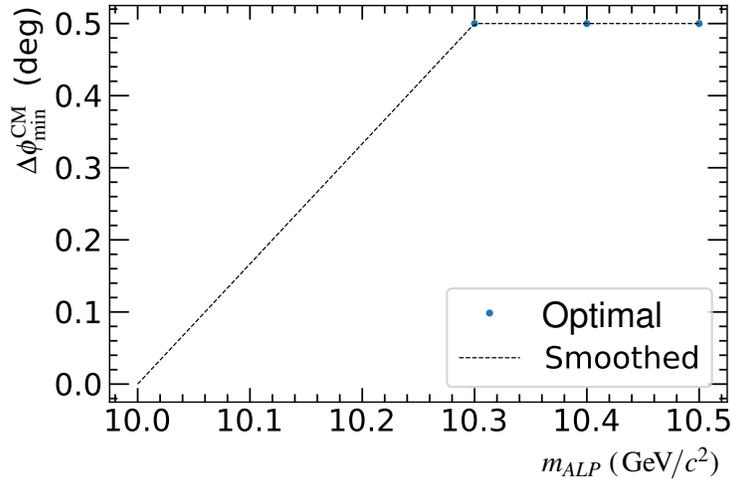


Figure 3.28: Optimal $\Delta\phi_{min}^{CM}$ values and the chosen smoothed function for $\Upsilon(4S)$ CME.

alternate was chosen to see what would happen if the E_{min}^{CM} threshold value is kept constant above $5.5 \text{ GeV}/c^2$. As can be seen by the orange curves in Figure 3.29, the third alternate function does not reject enough ALP candidates and creates a peak around $5.9 \text{ GeV}/c^2$. It turns out that the chosen alternate rejects just enough ALP candidates to keep the di-photon mass distribution smooth-ish.

3.4 Data-MC comparison

It is important to check how well the data and background MC agree, after applying all the selection cuts. Figure 3.31 shows a comparison of the ALP candidate mass distributions of MC and 5% of data for the $\Upsilon(2S)$, $\Upsilon(3S)$, and $\Upsilon(4S)$ CMEs. The MC is scaled to 5% of the full integrated luminosity. There is a relatively constant ratio between the number of MC candidates and the number of data candidates in all the mass range and in all the CMEs. The reason for this disagreement has not been found. It could be because the BKQED generator is relatively old and does not correctly simulate the $e^+e^- \rightarrow \gamma\gamma\gamma$ process.

To deal with this disagreement, the $e^+e^- \rightarrow \gamma\gamma(\gamma)$ MC is scaled so that the number of entries in the mass range $0 \text{ GeV}/c^2$ to $9 \text{ GeV}/c^2$ is equal to the number of

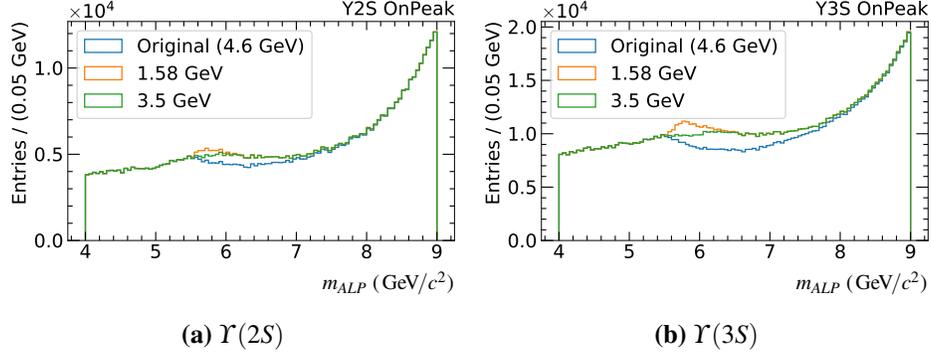


Figure 3.29: ALP candidate mass distributions for (a) $\Upsilon(2S)$ and (b) $\Upsilon(3S)$ CMEs with three sets of selection cuts for E_{\min}^{CM} . The blue distribution shows the original smoothed function. The green one is the one used in the rest of the analysis. The orange curve is the third alternate described in the text.

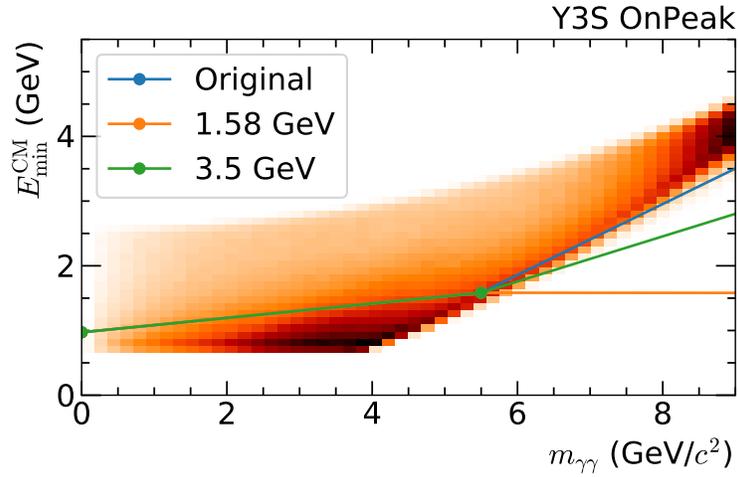


Figure 3.30: Two dimensional histogram of E_{\min}^{CM} versus the ALP candidate mass. The three sets of E_{\min}^{CM} are overlaid. The blue curve is the original smoothed function. The green one is the one used in the rest of the analysis. The orange curve is the third alternate described in the text.

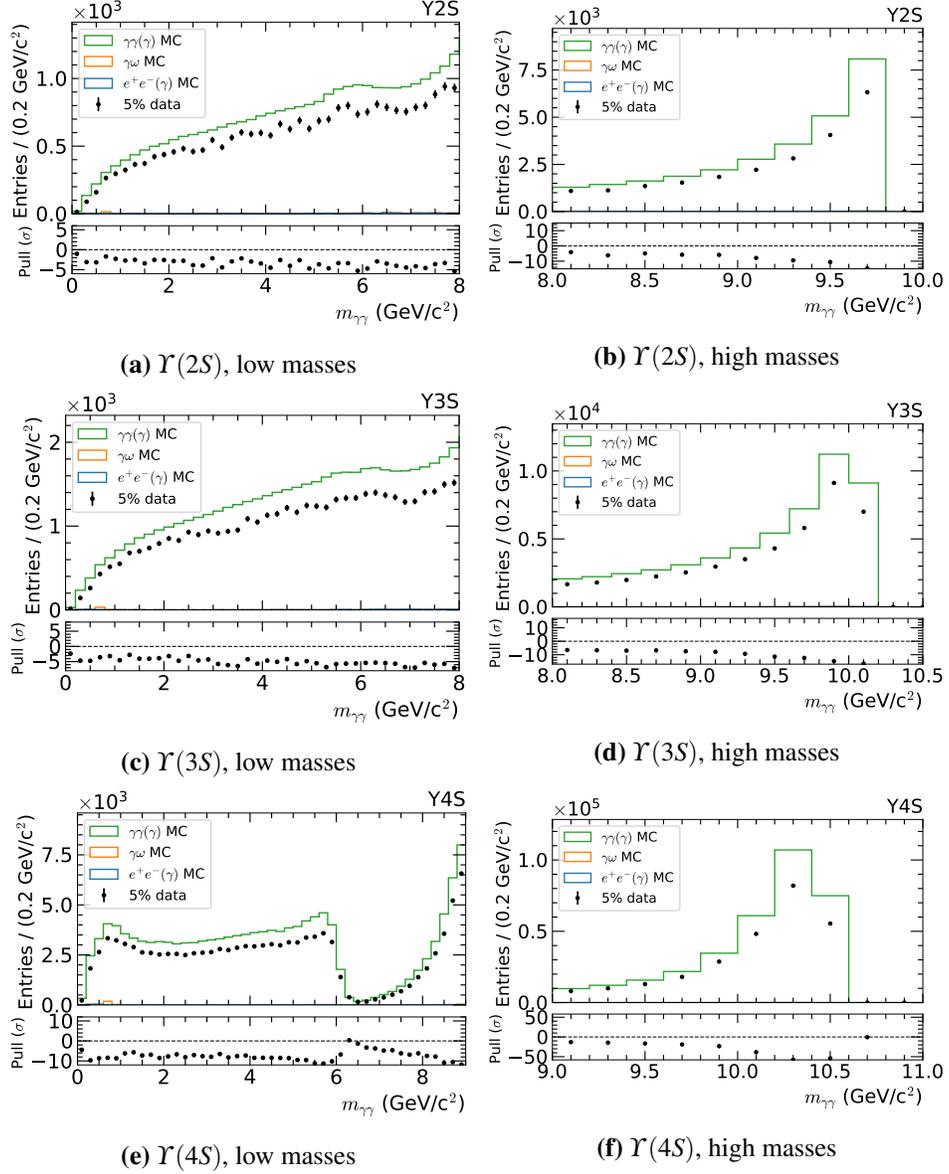


Figure 3.31: Data-MC comparison. The left column shows low ALP mass candidates while the right shows high mass candidates. The top row shows the $\Upsilon(2S)$ CME, the middle shows the $\Upsilon(3S)$ CME, and the bottom shows the $\Upsilon(4S)$ CME.

entries in the data for that range. The upper limit of this range is chosen to be high but still lower than the CME. The exact value is chosen arbitrarily. This procedure shoves all the correction into the $e^+e^- \rightarrow \gamma\gamma(\gamma)$ MC, but as the other MC processes contributing to the background are small, this should be adequate. For each CME, the on-resonance and off-resonance are scaled separately. Table 3.5 lists the scaling factors for each sample. This correction is later applied to the signal efficiency and contributes to the systematic uncertainty as discussed in Section 3.8.2.

Figures 3.32, 3.33, 3.34 show data-MC comparisons for the $\Upsilon(2S)$, $\Upsilon(3S)$, and $\Upsilon(4S)$ CMEs, respectively, after the MC scaling is applied. Figure 3.34a nicely shows the contribution of the $e^+e^- \rightarrow \omega\gamma$ process in the $\Upsilon(4S)$ CME sample. The $\Upsilon(2S)$ and $\Upsilon(3S)$ have lower statistics and this process is not as visible in them. Figure 3.34c shows that the MC simulates quite well the relatively sharp transitions in the ALP candidate mass distributions due to the trigger efficiency.

3.5 Toy experiments creation

The background MC is used to create 5000 toy experiment. Each toy experiment is a histogram of di-photon invariant masses expected if 5 % of the data are collected. A toy experiment is created by bootstrapping - randomly drawing ALP candidate masses from the different MC samples at the different CMEs. The number of entries drawn from each MC sample is the expected number of entries after scaling the MC to 5 % of the total integrated luminosity. For the $e^+e^- \rightarrow \gamma\gamma(\gamma)$ MC, the number of drawn entries is further scaled by the scale factor used to normalize the MC to

Table 3.5: Scale factors for the $e^+e^- \rightarrow \gamma\gamma(\gamma)$ MC sample that normalize the number of entries to that in data in the mass range 0 GeV/ c^2 to 9 GeV/ c^2 .

CME	On/Off resonance	Scale factor
$\Upsilon(2S)$	Off	0.83
	On	0.83
$\Upsilon(3S)$	Off	0.79
	On	0.82
$\Upsilon(4S)$	Off	0.87
	On	0.81

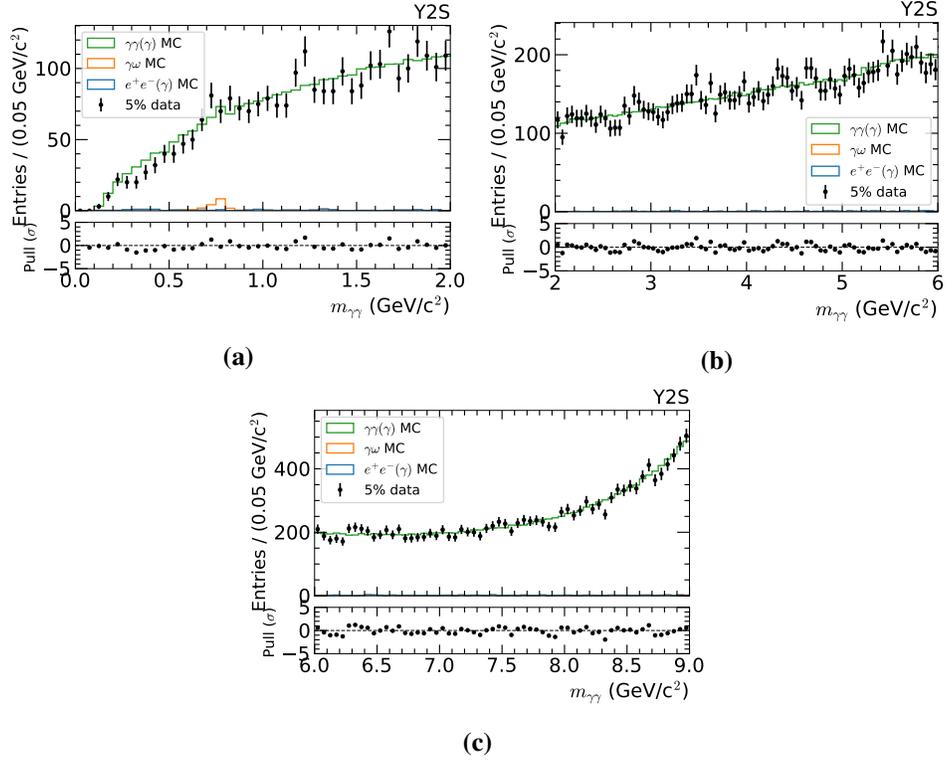


Figure 3.32: Data-MC comparisons at the $Y(2S)$ CME. $e^+e^- \rightarrow \gamma\gamma(\gamma)$ MC scaling is applied.

data, as discussed in Section 3.4. For the runs at the $Y(2S)$ and $Y(3S)$ CMEs, only masses up to 97.5 % of the CME are used because of the trigger issues discussed in Section 3.3.2.

The size of the $e^+e^- \rightarrow e^+e^-\gamma$ MC sample at the $Y(4S)$ CME is relatively small. The number of entries that pass the selection cuts are not enough to bootstrap from. Therefore, the di-photon invariant mass distribution for that MC sample is approximated by a function. The distributions and approximation functions can be seen in Figure 3.35. The off-resonance sample is approximated by a constant term and a Gaussian at the high mass peak. The on-resonance sample is approximated by a constant term and three Gaussians at the three peaks. The entries in the toy experiments for the $e^+e^- \rightarrow e^+e^-\gamma$ process at the $Y(4S)$ CME are drawn from the approximation functions instead of from the MC entries themselves. The actual

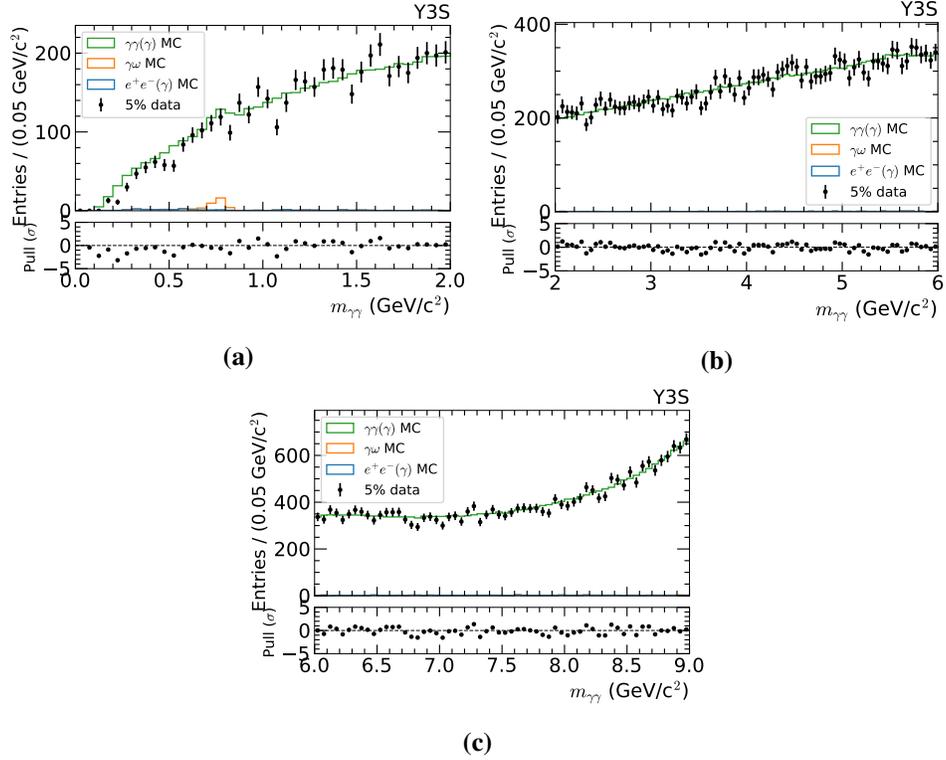


Figure 3.33: Data-MC comparisons at the $Y(3S)$ CME. $e^+e^- \rightarrow \gamma\gamma(\gamma)$ MC scaling is applied.

number of entries used for each MC sample and each CME is listed in Table 3.6.

Figure 3.36 shows the di-photon mass distributions of two toy experiments for the combined $Y(2S)$, $Y(3S)$ CMEs and the $Y(4S)$ CME together with the data. Small statistical fluctuations can be seen between the two toy experiments.

Discussion of the method

The method of creating the toy experiments by sampling from the MC entries themselves has a disadvantage. The toys are not sampled from a continuous distribution, but rather from a distribution already affected by statistical fluctuations. These statistical fluctuation in the MC manifest themselves in all the toy experiments drawn from that MC. The size of the fluctuations in MC compared to the ones in data is

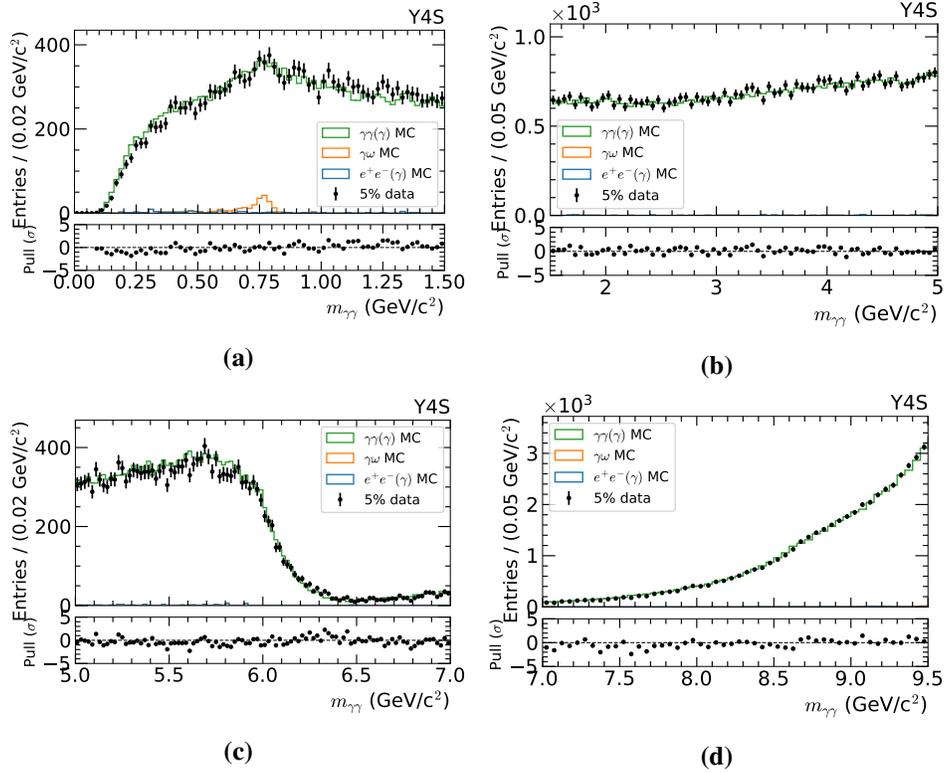


Figure 3.34: Data-MC comparisons at the $Y(4S)$ CME. $e^+e^- \rightarrow \gamma\gamma(\gamma)$ MC scaling is applied. Note the different bin sizes in the subplots.

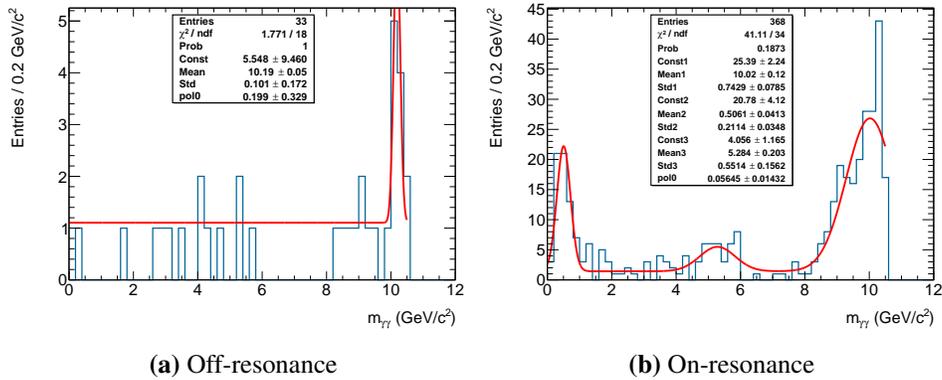


Figure 3.35: ALP candidate mass distributions for the $e^+e^- \rightarrow e^+e^-\gamma$ sample at the $Y(4S)$ CME, together with the functions approximating the distributions.

Table 3.6: Number of entries contributing to each toy experiment by each background process for each CME.

CME	On/Off resonance	$e^+e^- \rightarrow$		
		$\gamma\gamma(\gamma)$	$\omega\gamma$	$e^+e^-(\gamma)$
$\Upsilon(2S)$	Off	4587	6	18
	On	41546	57	181
$\Upsilon(3S)$	Off	7294	10	36
	On	76447	112	249
$\Upsilon(4S)$	Off	36932	57	47
	On	346328	565	485

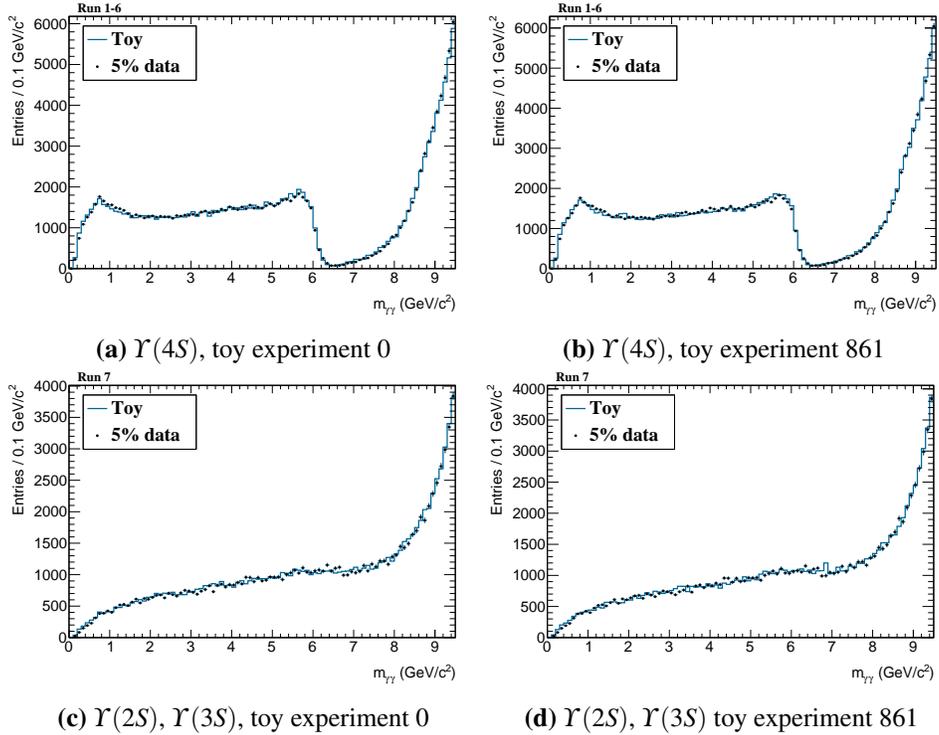


Figure 3.36: ALP candidate mass distributions of two toy experiments together with the data. The left column shows toy experiment 0. The right column shows toy experiment 861. The top row shows the $\Upsilon(4S)$ CME. The bottom row shows the combined $\Upsilon(2S)$, $\Upsilon(3S)$ CMEs. The data points in the two plots of the same toy experiment are identical.

determined by the ratio of the equivalent integrated luminosity of the MC and the integrated luminosity of the data. Table 3.3 lists this ratio for the full *BABAR* dataset.

If the MC has a downward fluctuation at a particular mass range then, on average, the toy experiments will also have a deficit in the number of entries in that mass range. For example, the sum of all the MC samples at the $\Upsilon(4S)$ CME has a deficit in the di-photon mass range $1.12 \text{ GeV}/c^2$ to $1.16 \text{ GeV}/c^2$. This can be seen as a two bin wide trough in Figure 3.34a. The number of entries in that mass range in all the toys will, on average, be smaller than the number of entries in ranges $1.08 \text{ GeV}/c^2$ to $1.12 \text{ GeV}/c^2$ and in $1.16 \text{ GeV}/c^2$ to $1.2 \text{ GeV}/c^2$. This poses a problem because the Poisson uncertainties associated with each bin do not accurately represent the average difference between the bin content and the true probability density function (PDF) describing the distribution. For this reason, fits to the toy experiments usually have a very low χ^2 probability. It should be noted that this effect only applies to MC toy experiments. The data is truly drawn from the PDF describing the data. The uncertainties of the data correctly describe the deviation of the measurement from the underlying PDF.

3.6 Signal extraction

A binned maximum likelihood fit to the ALP candidate mass distribution is used to extract the number of signal events. The fit is performed with the `root` framework version 6.18.00 [51]. The fit modifies the fit parameters and maximizes the likelihood that the data is described by the fit function. The fit function is composed of a signal component, a background component, and components for SM resonances.

3.6.1 Signal shape function

For each ALP mass hypothesis, the signal component is built using the ALP candidate mass distribution of the signal MC for that mass. The distribution is linearly interpolated between the bin contents to construct a continuous function. All entries that pass the selection cuts contribute, even those with the wrong combination of photons, i.e. ones in which an ALP daughter photon is combined with the recoil photon to produce an ALP candidate. Figure 3.37 shows examples of the histogram and the smooth function for two ALP masses at the $\Upsilon(4S)$ CME.

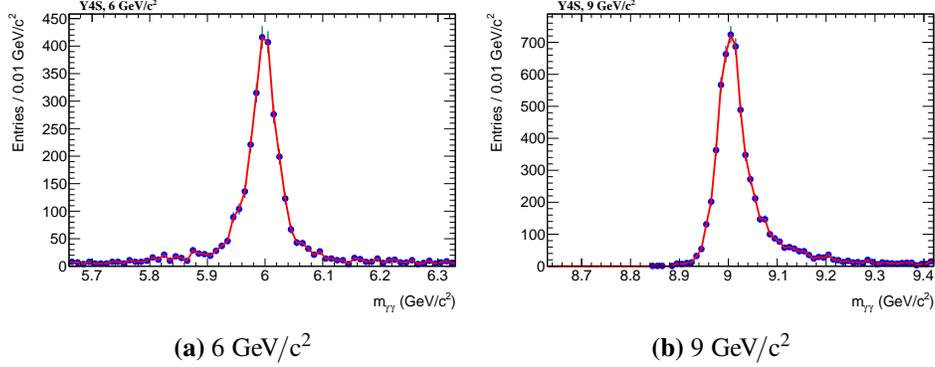


Figure 3.37: Signal MC shapes for (a) $6 \text{ GeV}/c^2$ and (b) $9 \text{ GeV}/c^2$ at the $Y(4S)$ CME. The figures show the histogram bin content in blue and the linearly interpolated function in red.

3.6.2 Standard model resonance contributions

Some SM resonances can contribute a peaking component to the background. We consider the π^0 , η , η' , ρ^0 , ω , and ϕ . Heavier SM resonances are neglected. The π^0 , η , and η' SM resonances with masses of $0.135 \text{ GeV}/c^2$, $0.548 \text{ GeV}/c^2$, $0.958 \text{ GeV}/c^2$ [10], respectively, can decay to two photons, just like an ALP, and therefore leave the same signature in the detector as an ALP.

The ρ^0 with a mass of $0.775 \text{ GeV}/c^2$, the ω with a mass of $0.784 \text{ GeV}/c^2$, and the ϕ with a mass of $1.02 \text{ GeV}/c^2$, can all decay to a $\gamma\pi^0$, where the π^0 is boosted enough that when it decays to two photons, the photons merge in the detector and are reconstructed as a single photon. Therefore, these SM resonances look like they decay into two photons.

π^0

The π^0 mesons mass is very close to the lower edge of the mass range this search is sensitive to. In order not to complicate the search, it was decided to avoid the mass range that the π^0 influences and start the search for the ALP above it. The method used to decide where the π^0 contribution is negligible and where to start the search is described in Section 3.6.4.

η

As *BABAR* does not have $e^+e^- \rightarrow \eta\gamma$ MC the shape of the η is evaluated from signal MC where the ALP is generated at the η mass. Figure 3.38 shows signal MC at the η mass. Each histogram is fitted with a Gaussian. The Gaussians are used as the η component in the fit function.

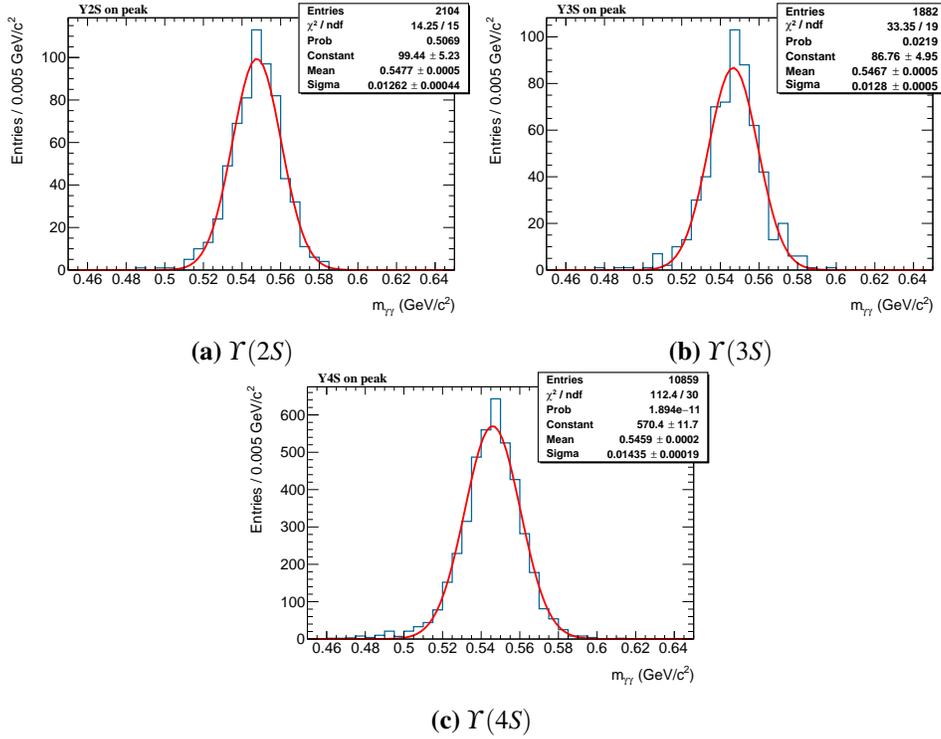


Figure 3.38: MC distributions used to evaluate the η resonance shape at the (a) $Y(2S)$, (b) $Y(3S)$, and (c) $Y(4S)$ CMEs. Each distribution is approximated by a Gaussian.

As the number of η events is too small to measure their contribution from 5 % of the data, its normalization in the fit function is fixed to the expected value. The CS of the process $e^+e^- \rightarrow \eta\gamma$ was measured at *BABAR* at the $Y(4S)$ CME [52]. In order to calculate the expected number of η events, the CS is multiplied by 5 % of the integrated luminosity, the $\eta \rightarrow \gamma\gamma$ branching fraction, and the signal efficiency at the η mass. Table 3.7 lists prompt production CSs and branching fractions to two

photons for the η and some of the other SM resonances appearing in this section. The values are for the $\Upsilon(4S)$ CME. The expected number of η events for the full *BABAR* dataset is 12, 22, 105, at the $\Upsilon(2S)$, $\Upsilon(3S)$, $\Upsilon(4S)$ CMEs, respectively.

It is assumed that the CS does not change rapidly with the CME. Therefore, the measured CS at the $\Upsilon(4S)$ CME is also used for the $\Upsilon(2S)$ and $\Upsilon(3S)$ CMEs. As the contributions at the $\Upsilon(2S)$ and $\Upsilon(3S)$ CMEs are small, it probably does not make much difference to the final calculated limits, even if this assumption is not true.

Table 3.7: CSs and branching ratios for SM resonances at the $\Upsilon(4S)$ on-resonance. The sources are listed in the table. The CSs for the $e^+e^- \rightarrow \omega\gamma$ and $e^+e^- \rightarrow \phi\gamma$ processes are calculated with formula 7 from [49].

X	CS of $e^+e^- \rightarrow X\gamma$ (fb)	$X \rightarrow \gamma\gamma$ branching fraction (%)
η	4.5 [52]	39.4 [10]
ω	17000	8.4 [10]
η'	5.4 [52]	2.2 [10]
ϕ	26000	0.1 [10]

ρ^0

Although the ρ^0 has a mass of 0.775 GeV/ c^2 which is in the search range, it has a relatively large width of 0.148 GeV. This width is an order of magnitude wider than the detector resolution at the ρ^0 mass and therefore the background function absorbs its contribution. The ρ^0 is therefore not part of the fit function.

ω

BABAR has prompt ω MC. This MC is used to produce a smooth function by linearly interpolating between the bin contents, in the same way that is done in the signal MC described in Section 3.6.1. Figure 3.39 shows the histograms of the prompt ω MC used to create the smooth function.

The ω peak is relatively large and can be measured in data. Therefore, a dedicated fit to extract the ω normalization is performed. The dedicated fit is done only around the ω mass range. The ω normalization in all the mass range is then fixed to the value obtained from the dedicated fit.

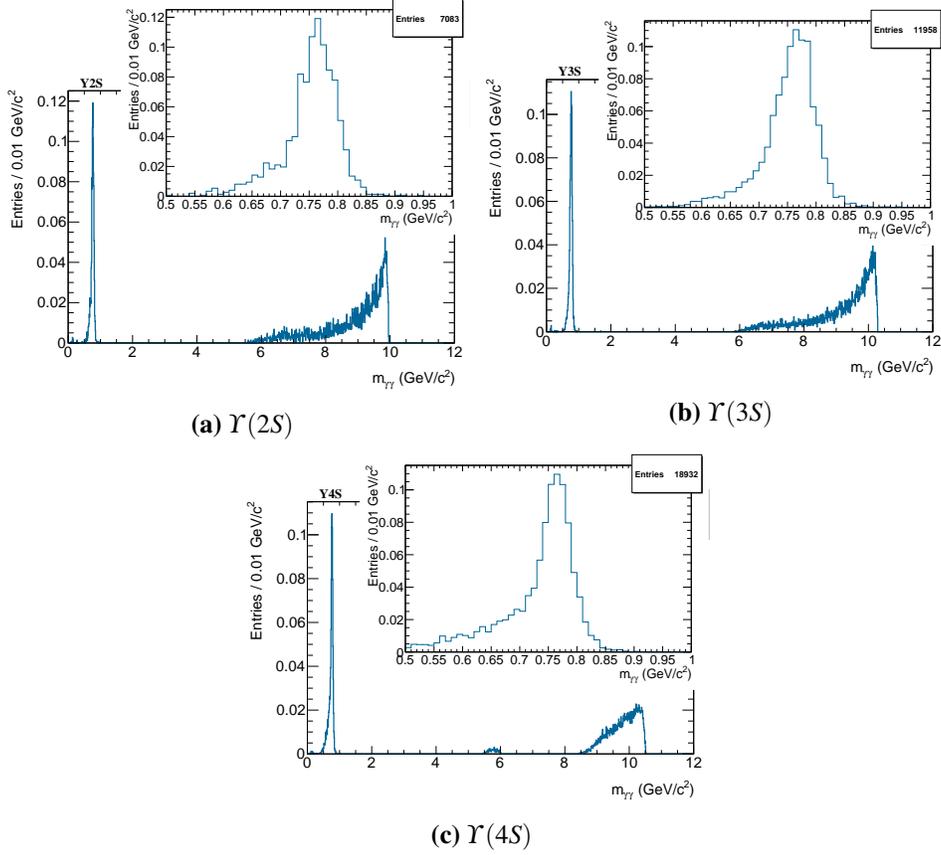


Figure 3.39: Distributions of the $e^+e^- \rightarrow \omega\gamma$ MC after all selection cuts are applied for. Each subplot includes an inset with a zoomed in version around the ω peak.

It was decided that even if an excess of entries around the ω mass is observed in the data, no ALP discovery will be claimed. To decide what range of masses around the ω mass to exclude, the mass range around the ω mass is fitted with floating omega component as well as floating signal component. Floating here means a variable which is adjusted by the fit. The uncertainty on the number of signal entries, averaged over all the toy experiments is plotted in Figure 3.40 for the $\Upsilon(4S)$ CME. The figure shows that around the ω mass, the fit cannot decide how to split the entries between the signal and ω components and therefore the uncertainty of the number of signal entries increases. Using the figure, the mass

range $0.72 \text{ GeV}/c^2$ to $0.81 \text{ GeV}/c^2$, where the uncertainty on the number of signal entries increases, is excluded. The same range is used to exclude the masses for the $\Upsilon(2S)$ and $\Upsilon(3S)$ CMEs.

η'

An identical analysis to that of the η is used to roughly estimate the η' contribution. The production CS is taken from [52]. The number of η' events is estimated to be less than six events in the complete $\Upsilon(4S)$ data sample. Therefore, the η' contribution is neglected.

ϕ

The shape of the ϕ component is obtained in exactly the same way as the η , described earlier in the section, using the $1 \text{ GeV}/c^2$ ALP signal MC. Figure 3.41 shows the signal MC at the ϕ mass, used to obtain the Gaussian parameters.

A naive calculation of the number of expected events as the product of the CS, integrated luminosity, and signal efficiency at the ϕ mass is not sufficient to estimate the ϕ normalization. It does not reflect the fact that the ϕ has total angular momentum of one, unlike the ALP which is a scalar. It also does not take into account that for a $\phi\gamma \rightarrow \pi^0\gamma\gamma$ event to be signal-like, the daughter photons of the π^0 must merge and be reconstructed as a single photon. The single merged photon reconstructed from the π^0 decay and the second photon of the ϕ decay combine to produce a peaking background at the ϕ mass. A correction factor is used to adjust the naive expected number of events to take these effects into account. As *BABAR* does not have prompt ϕ MC, the correction factor is calculated from the prompt ω MC. It is assumed that the correction factors for the ϕ and the ω are similar.

The correction factor is the ratio of the number of prompt ω MC events with an ALP candidate with a mass below $2 \text{ GeV}/c^2$ and the expected number of $e^+e^- \rightarrow \gamma\omega \rightarrow \gamma\gamma\pi^0$ events. The later is calculated by multiplying the theoretical CS of the $e^+e^- \rightarrow \omega\gamma$ process, integrated luminosity, $\omega \rightarrow \gamma\pi^0$ branching fraction, and efficiency estimated from signal MC at the ω mass. The CS of the $\omega \rightarrow \gamma\pi^0$ process is calculated using formula 7 in [49], with second order corrections and updated mass and width values from [10]. The correction factors are listed in Table 3.8.

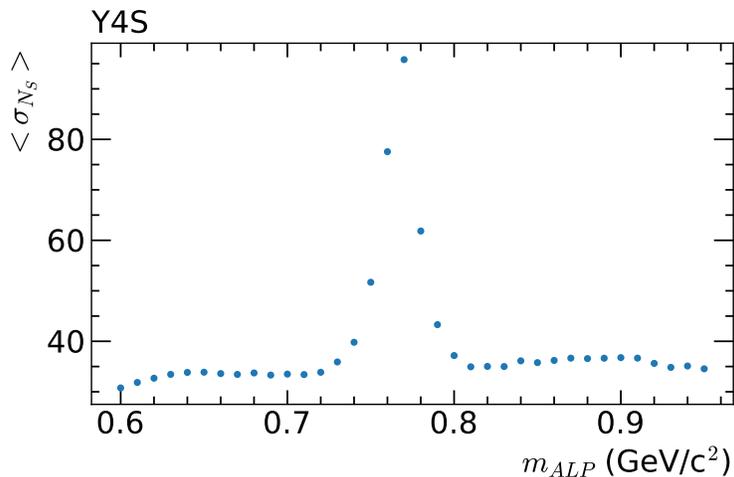


Figure 3.40: Uncertainty on the number of signal entries, averaged over all the toy experiments as a function of the ALP candidate mass, for the $\Upsilon(4S)$ CME. Both the signal component and the ω component are allowed to float in the fit.

Having obtained the correction factor, the theoretical CS of the process $e^+e^- \rightarrow \gamma\phi \rightarrow \gamma\gamma\pi^0$ is calculated in the same way. It is multiplied by the correction factor to produce the expected number of relevant ϕ events, which is used to fix the normalization of the ϕ component of the fit function. The corrected expected number of ϕ events in the full *BABAR* dataset are 8, 16, 87, for the $\Upsilon(2S)$, $\Upsilon(3S)$, $\Upsilon(4S)$ CMEs, respectively.

Table 3.8: Correction factor for the $e^+e^- \rightarrow \gamma\phi \rightarrow \gamma\gamma\pi^0$ CS.

Center of mass energy	Correction factor
Y2S	0.032
Y3S	0.034
Y4S	0.054

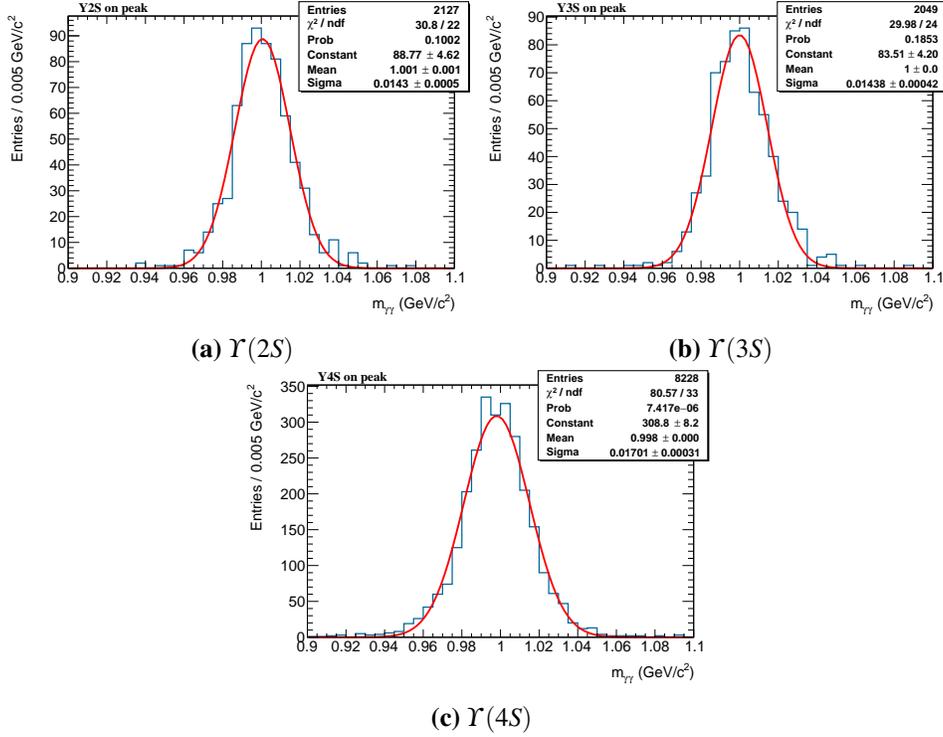


Figure 3.41: MC distributions used to evaluate the ϕ resonance shape. Each distribution is approximated with a Gaussian.

3.6.3 Background shape

The background component of the fitting function is modeled by a Chebyshev polynomial. Chebyshev polynomials are chosen as they are an orthogonal set, unlike Taylor polynomials.

In order to determine the polynomial degree required to describe the background, all the toy experiments are fitted with polynomials of different degrees. The signal component in these fits is fixed to zero and the fit is performed in the fit window described in Section 3.6.4. For each mass, the percent of toy experiments with bad fits is checked. Bad fits for this purpose are defined as fits with a χ^2 fit probability of less than 1%. The value is chosen arbitrarily. The polynomial degree chosen to describe the background is the degree that if it is increased, does not significantly decrease the percentage of bad fits. Each ALP mass hypothesis can

have a different polynomial degree.

In an ideal case, if the background is truly described by the Chebyshev polynomial that is fit and the uncertainties are properly modeled, the χ^2 probability of the fit is uniformly distributed between zero and one. Therefore, in such a case, the probability of a bad fit is 1 %. But as discussed in Section 3.5, the uncertainties of the toy experiments do not accurately reflect the deviation from the underlying distribution. Therefore, even if the background is properly modeled by a specific polynomial degree, the percentage of bad fits can be substantially higher than 1 %. For this reason, the absolute value of the percentage of bad fits is not a meaningful value. The change in the percentage of bad fits when modifying the polynomial degree is chosen as the metric for choosing the polynomial degree, rather than the absolute value.

Figure 3.42 and Figure 3.43 show the percentage of toy experiments with bad fits as a function of the ALP mass hypothesis for the $\Upsilon(2S)$ and $\Upsilon(3S)$ CMEs, respectively. Each figure has a zoomed version of the high mass region in an inset. As can be seen, in most of the mass range, up to around $8.5 \text{ GeV}/c^2$, the percentage of bad fits is the same whether the background is modeled by a first degree Chebyshev polynomial or a higher degree polynomial. At the higher masses, the ALP candidate mass distribution is sufficiently non-linear that a first degree polynomial is not enough to describe it. Increasing the polynomial degree decreases the percentage of bad fits. Therefore, in the high mass region, a higher polynomial degree is needed to describe the background.

Figure 3.44 shows the percentage of bad fits as a function of the ALP mass hypothesis for the $\Upsilon(4S)$ CME at different mass ranges. The lower end of the ALP candidate mass distribution changes sufficiently fast that a second degree polynomial is needed. The mass range up to $5.5 \text{ GeV}/c^2$ can be described by a first degree polynomial. The range $5.5 \text{ GeV}/c^2$ to $6.5 \text{ GeV}/c^2$ changes rapidly because of the large decrease in the trigger efficiency. This mass range is excluded. The highest mass range requires second and third degree polynomials to describe the background.

Table 3.9 lists the different ALP mass ranges and the polynomial degree used to describe the background.

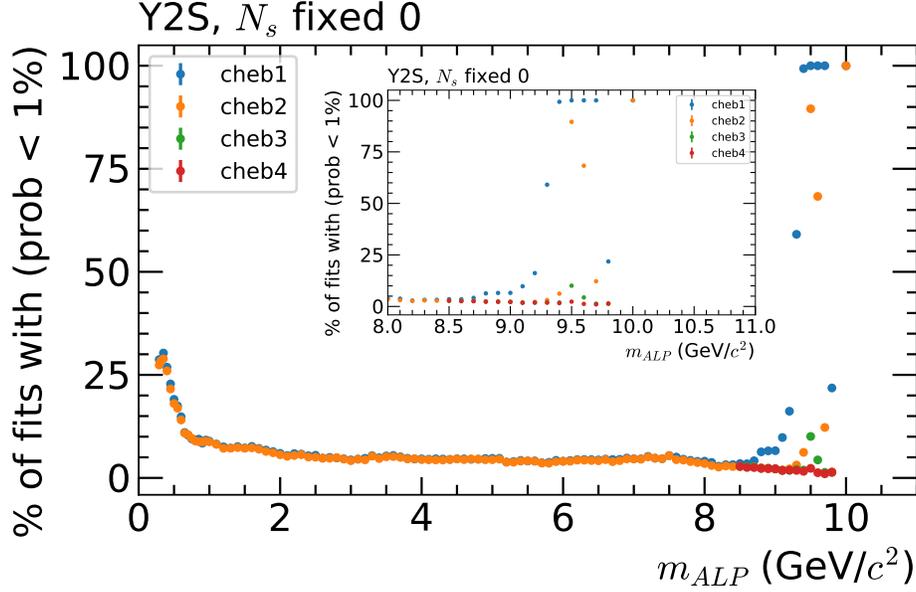


Figure 3.42: Percentage of bad fits to the background MC as a function of the ALP mass hypothesis for different Chebyshev polynomial degrees at the $\Upsilon(2S)$ CME. The inset photo shows a zoomed area of the high masses.

3.6.4 Fit windows

For each ALP mass hypothesis, the fit is done in a fit window containing the ALP mass being considered. To select the fit window width, first a base width is determined. The base width is calculated in the same way as the counting window in Section 3.3.4, only that for the base width, all the final selection cuts are applied. As each mass hypothesis has a different base width, adjacent mass hypotheses can have different widths. The base width is smoothed as a function of the ALP mass hypothesis and the edges are adjusted to coincide with bin edges. To form the actual fit window two base widths are added below and two above the base width containing the ALP mass. The total fit window is five times wider than the base width. If the counting window can be thought of as being 4.3σ wide, the fit window is 21.5σ wide, where σ is the equivalent Gaussian standard deviation. The fit window is wider than the base width to give the fit more information about the

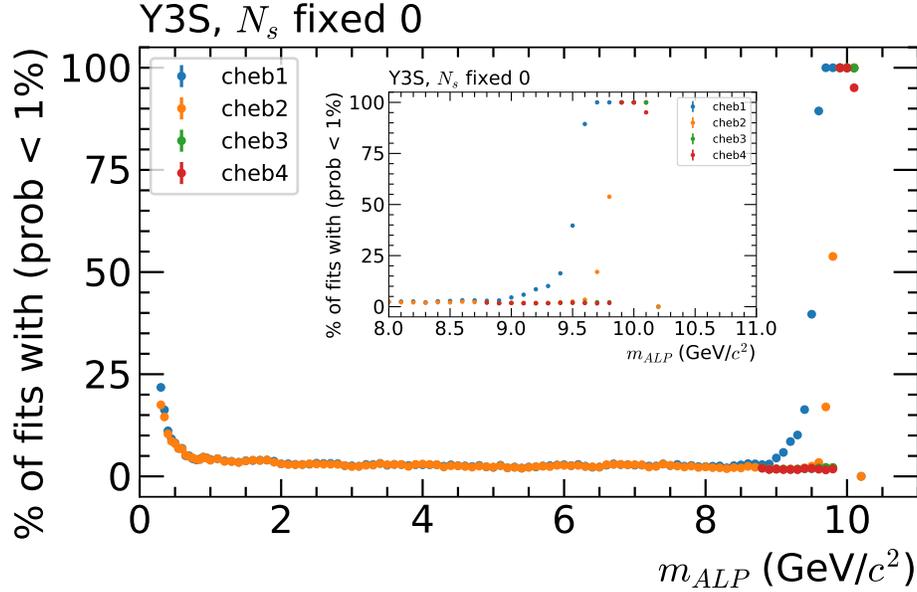


Figure 3.43: Percentage of bad fits to the background MC as a function of the ALP mass hypothesis for different Chebyshev polynomial degrees at the $\Upsilon(3S)$ CME. The inset photo shows a zoomed area of the high masses.

background so it can fit the background component of the fit function.

Selecting relevant ALP mass hypotheses

A few different methods are used to decide on the mass range in which to search for an ALP. The smallest ALP mass hypothesis is chosen to exclude the π^0 peak. The ALP candidate mass distribution of signal MC of an ALP with a mass of $0.135 \text{ GeV}/c^2$ is fitted with a Gaussian. Figure 3.45 shows the distribution and the fitted function. The ALP mass for which the lower end of the fit window is larger than two standard deviations above the Gaussian mean of the π^0 peak is the first ALP mass hypothesis used. This turns out to be $0.29 \text{ GeV}/c^2$, $0.31 \text{ GeV}/c^2$, $0.32 \text{ GeV}/c^2$, for the $\Upsilon(2S)$, $\Upsilon(3S)$, $\Upsilon(4S)$ CMEs, respectively.

For the $\Upsilon(2S)$ and $\Upsilon(3S)$ CMEs, the largest ALP mass is chosen such that the higher end of the fitting window is smaller than the mass for which the trigger

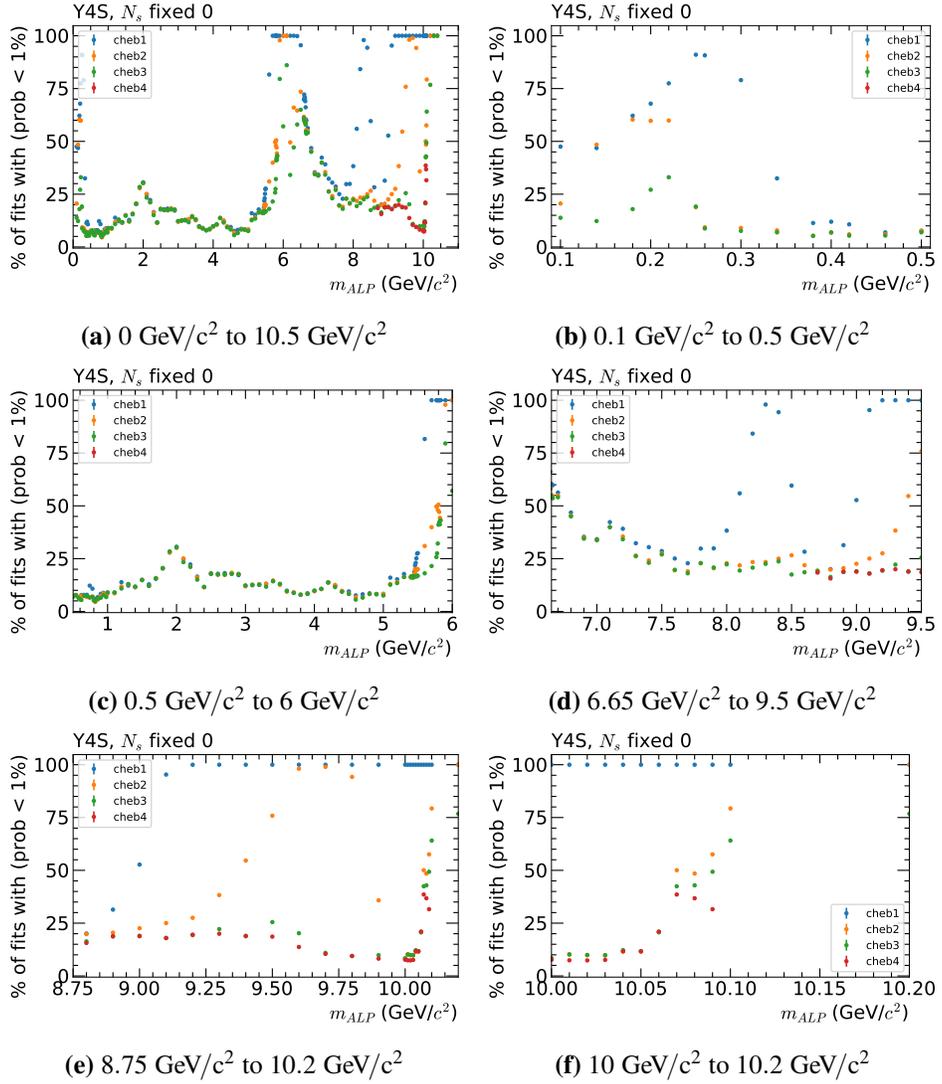


Figure 3.44: Percentage of bad fits to the background MC as a function of the ALP mass hypothesis for different Chebyshev polynomial degrees at the $\Upsilon(4S)$ CME.

Table 3.9: Background component Chebyshev polynomial degrees for different ALP mass ranges.

CME	Mass range (GeV/c ²)	Polynomial degree
$\Upsilon(2S)$	0.29 to 0.72	1
	0.81 to 8.59	1
	8.6 to 9.29	2
	9.3 to 9.4	3
$\Upsilon(3S)$	0.31 to 0.72	1
	0.81 to 8.89	1
	8.9 to 9.59	2
	9.6 to 9.8	3
$\Upsilon(4S)$	0.32 to 0.37	2
	0.38 to 0.72	1
	0.81 to 5.45	1
	6.65 to 7.19	1
	7.2 to 8.79	2
	8.8 to 10.05	3

issues arise, as discussed in Section 3.3.2. For the $\Upsilon(4S)$ CME, the largest ALP mass is the mass for which a 3rd degree Chebyshev polynomial is not sufficient to describe the background shape, as discussed below. The maximum fitted ALP mass is 9.4 GeV/c², 9.8 GeV/c², 10.05 GeV/c², for the $\Upsilon(2S)$, $\Upsilon(3S)$, $\Upsilon(4S)$ CMEs, respectively.

3.6.5 Background only MC fits

To check the fitting procedure each ALP mass hypothesis for each of the 5000 toy experiments is fitted with a floating signal component. Figure 3.46 shows an example fit for an ALP with a mass of 4.7 GeV/c² at the $\Upsilon(4S)$ CME. The result of the fit is the number of signal entries N_S and the uncertainty on the number of signal entries σ_{N_S} .

The distribution of the number of signal entries for the 4.7 GeV/c² ALP of all the toy experiments at the $\Upsilon(4S)$ CME is shown in in Figure 3.47. Ideally, the distribution should be centered on zero, as there are no signal entries added in the

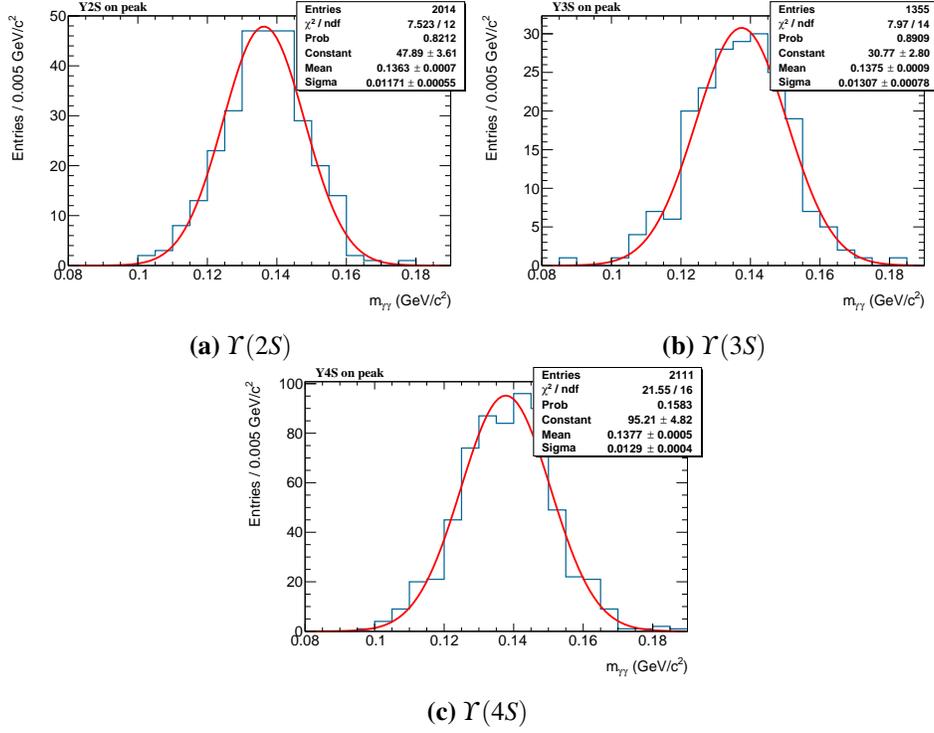


Figure 3.45: ALP candidate mass distribution for signal MC of an ALP with a mass of $0.135 \text{ GeV}/c^2$. The distributions are fitted with Gaussians.

fits. Nevertheless, the mean is highly inconsistent with zero. It is 97 standard deviations away. 90 % of the means for the mass hypotheses at the $\Upsilon(4S)$ CME are between -48 and 67 . They are also between -57 and 71 standard deviations away from zero. One of the more extreme examples is shown in Figure 3.47. As discussed in Section 3.5, this bias is due to the way the toy experiments are generated and should not appear in data. The bias information is still useful and will be used in later parts of the analysis.

Even though the number of signal entries returned by the fit is not consistent with zero, we can check whether the uncertainty on the number of signal entries that the fit returns is reasonable. This is checked using the pull distribution for N_S . As the expected number of signal entries is zero, the pull for N_S is $(N_S - 0) / \sigma_{N_S}$. If the uncertainties are correct, the standard deviation of the pull distribution is one.

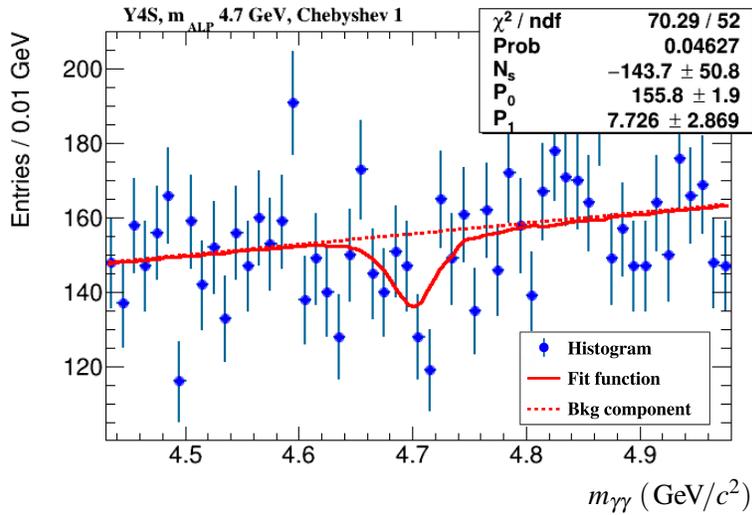


Figure 3.46: Example of a fit to a background only toy experiment, at a mass of $4.7 \text{ GeV}/c^2$ at the $\Upsilon(4S)$ CME. The solid red line is the complete fit function. The dashed red line is the fit function without the signal component.

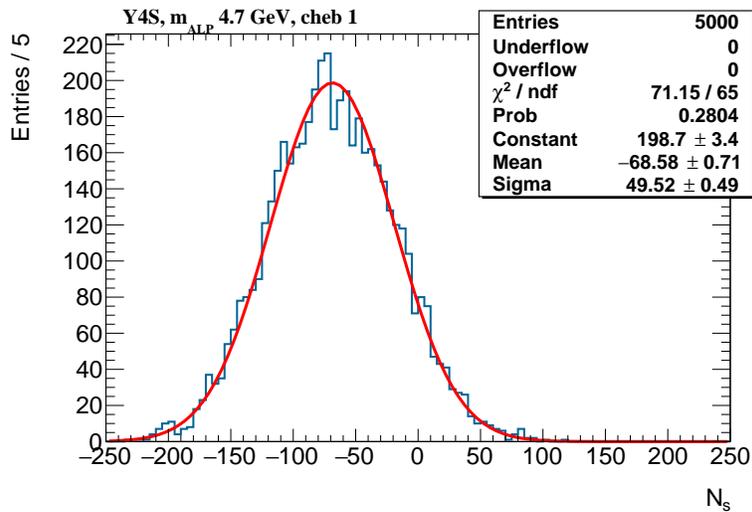


Figure 3.47: Distribution of the number of signal entries for all the toy experiment fits at a mass of $4.7 \text{ GeV}/c^2$ at the $\Upsilon(4S)$ CME. The distribution is fitted with a Gaussian.

Figure 3.48 shows the pull of the $4.7 \text{ GeV}/c^2$ ALP at the $Y(4S)$ CME. It shows that the uncertainties returned by the fit are a bit too large. The same can be done for all the other masses. Figure 3.49 shows the distribution of the standard deviations of the pulls of N_S for the different CMEs. Each entry in a histogram is the standard deviation of a pull of a specific ALP mass. As can be seen, on average, the uncertainties returned by the fit are too large by a few percent. This is not too concerning. First, the uncertainties are too large by a small amount. Second, uncertainties that are too large will make it harder to claim a discovery of an ALP and they increase the limit on the ALP coupling constant. If the fit had returned uncertainties that were too small, that would be more concerning.

3.6.6 Signal plus background MC fits

To check that the signal extraction works as intended in the presence of signal, signal MC entries are added to the toy experiments and the signal extraction procedure is repeated.

Adding signal MC entries

A fixed number of signal MC events is chosen for each ALP mass hypothesis. The number of events is chosen so that, on average, the total number of entries in the fit region is equal to five standard deviations of the number of signal entries of the background only MC fits to the toy experiments, discussed in Section 3.6.5. Five standard deviations is chosen because that is the accepted threshold for a new discovery. Each ALP mass hypothesis has its own fixed number of added signal events. The events for each toy are randomly chosen from among all the signal MC events that have at least one ALP candidate that passes the selection cuts.

Fits

Figure 3.50 shows a sample fit to a toy experiment with signal MC entries added, at the $4.7 \text{ GeV}/c^2$ ALP mass at the $Y(4S)$ CME. This is the same toy experiment as in Figure 3.46, the only difference is that in Figure 3.50 signal entries are present.

One way to check if the fits are biased is to examine the residuals. If the fit returns the correct number of signal entries, the residual distribution should be

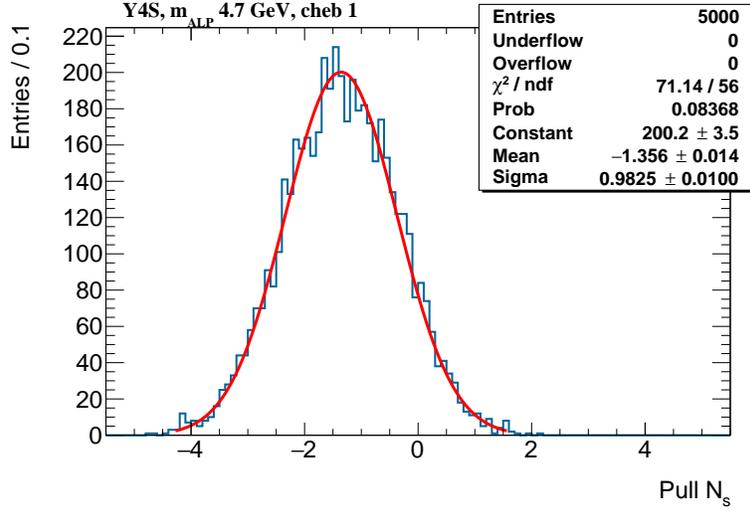


Figure 3.48: The N_S pull, N_S/σ_{N_S} , for the toy experiment fits at a mass of $4.7 \text{ GeV}/c^2$ at the $\Upsilon(4S)$ CME. The distribution is fitted with a Gaussian.

centered on zero. Figure 3.51 shows the residuals for the $4.7 \text{ GeV}/c^2$ ALP at the $\Upsilon(4S)$ CME for all the toy experiments. N_S^{True} is the true number of signal MC entries in the fit window added to each toy experiment. As can be seen, the mean residual has a large bias. As discussed in Section 3.6.5, the number of signal entries in the background only fits is also biased. Therefore, to get a meaningful residual, the number of signal entries from the background only fits is used as a baseline for the number of signal entries in the signal plus background case. This creates a new residual, the *adjusted residual*, and it is calculated by subtracting the mean number of signal entries of the background only fit $\langle N_S^{\text{0sig}} \rangle$ from the regular residual, where $\langle \dots \rangle$ signifies the mean. The bias disappears if the adjusted residual is used, as can be seen in Figure 3.52. The mean is now consistent with zero. It should be noted that this adjustment is needed to address the bias in the toy experiments. The data should not have a bias and therefore no adjustments are made to fits of the data.

A better way to evaluate the bias which takes into account the uncertainties on the number of signal entries is with the adjusted pull, $(N_S - \langle N_S^{\text{0sig}} \rangle - N_S^{\text{True}}) / \sigma_{N_S}$. Figure 3.53 shows the mean of the adjusted pull for the three CMEs. There is a

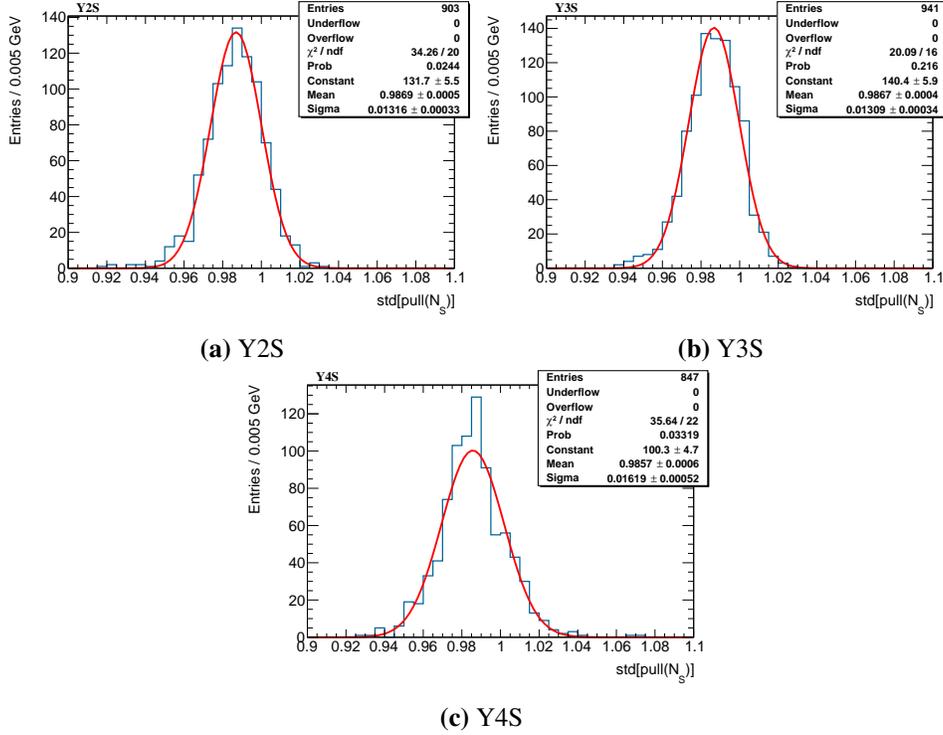


Figure 3.49: Distributions of the standard deviation of the pulls of the number of signal entries for fits to toy experiments for all the ALP mass hypotheses. Each ALP mass hypothesis is one entry in each of the histograms. Each plot is fitted with a Gaussian.

slight bias toward negative values, meaning the fit returns slightly less signal entries than there actually are. But like previously, the bias is small and it is toward the conservative direction. Figure 3.53 shows that each of distributions has a low edge tail. This tail is because the fits to the ALP masses around 1.1 GeV/ c^2 return a lower number of signal entries than the true number of signal entries added. This happens for all three CMEs. The reason why this happens is not clear.

The uncertainty on the number of signal events in the signal plus background MC fits is also checked. Figure 3.54 shows the standard deviation of the pull distributions as a function of the ALP mass hypothesis. Ideally, these should equal to one. For the $Y(2S)$ and $Y(3S)$ CMEs the standard deviations are relatively small at low ALP masses and increase for high masses. For the $Y(4S)$ CME, there is a large

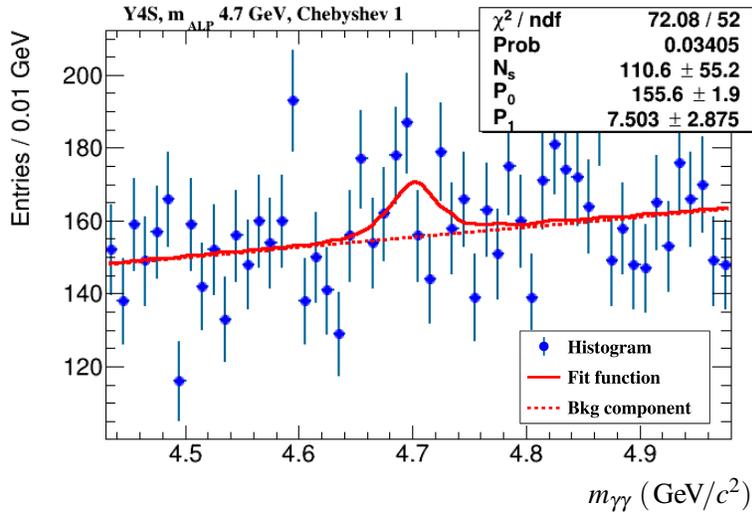


Figure 3.50: Example of a fit to a toy experiment with signal MC entries for an ALP with a mass of $4.7 \text{ GeV}/c^2$ at the $\Upsilon(4S)$ CME.

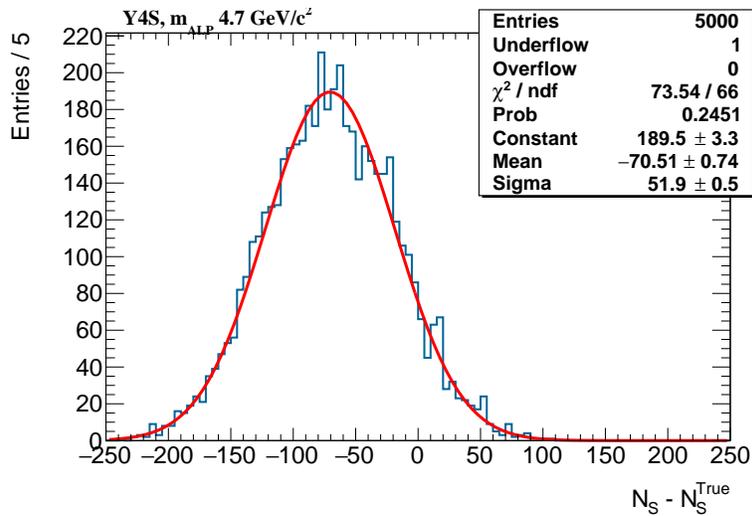


Figure 3.51: Residuals distribution for the number of signal entries for a $4.7 \text{ GeV}/c^2$ ALP at the $\Upsilon(4S)$ CME in signal plus background MC. The distribution is fitted with a Gaussian.

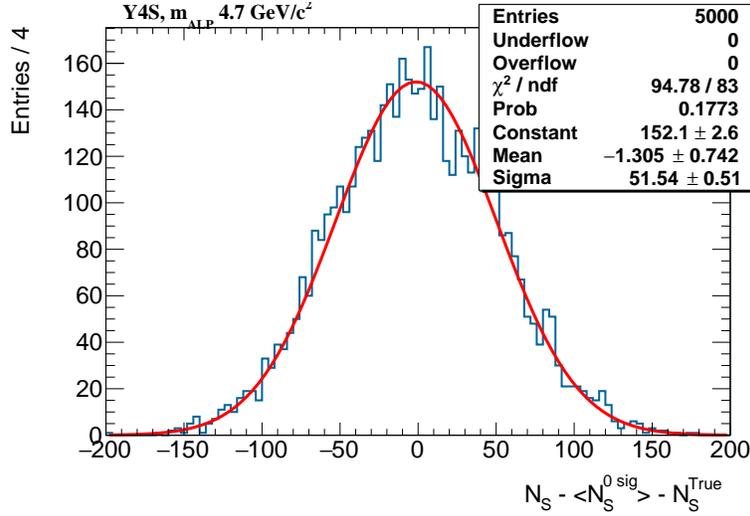


Figure 3.52: Adjusted residuals for an ALP with a mass of $4.7 \text{ GeV}/c^2$ at the $\Upsilon(4S)$ CME in signal plus background MC. The distribution is fitted with a Gaussian.

decrease in the standard deviations around $6.5 \text{ GeV}/c^2$, just on the edge of the excluded masses. Low standard deviations mean that, on average, the uncertainty on the number of signal entries returned by the fit is too large. The reason why the standard deviations are so different from one, and why they depend on the mass is not understood. But, like previously, the deviations from the ideal case are in the conservative direction.

Because adding signal entries increases the total number of entries in the signal region, the uncertainty on the number of signal entries in the signal plus background fits is higher than in the background only case. This causes the mean of the adjusted local significance, $\langle (N_S - \langle N_S^{0 \text{sig}} \rangle) / \sigma_{N_S} \rangle$, to decrease. This is demonstrated in Figure 3.55. The blue points show the mean of the adjusted local significance as a function of the ALP mass hypothesis, where the uncertainty is the uncertainty of the signal plus background fits. Remember that the number of signal entries added to the toy experiments is chosen so that the average local significance is five. The blue points show that for the signal plus background case, the local significance is lower than five. The orange points show the adjusted local significance with the uncertainties taken from the background only fits. In this case, the local

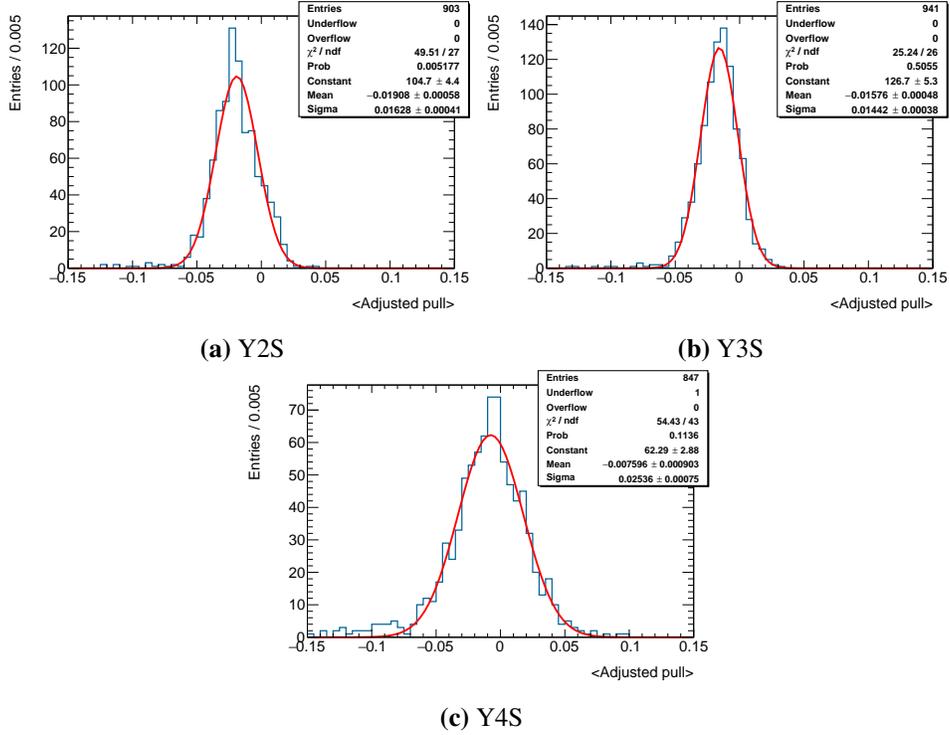


Figure 3.53: Distributions of the mean of the adjusted pull for signal plus background MC fits. Each ALP mass hypothesis contributes one entry to a histogram. Each plot is fitted with a Gaussian.

significances are close to five, as intended.

3.6.7 Signal efficiency

In this work, the signal efficiency is treated as a conversion factor between the fitted number of signal entries N_{fit} and the true number of ALP events n_{true} . n_{true} can be either generated MC events, or ALP particles produced in the accelerator. N denotes number of entries and n denotes number of events. The signal efficiency needs to reflect the fact that an event can have multiple ALP candidates in the fit window.

With this definition, the relation between N_{fit} , the efficiency ε , the integrated

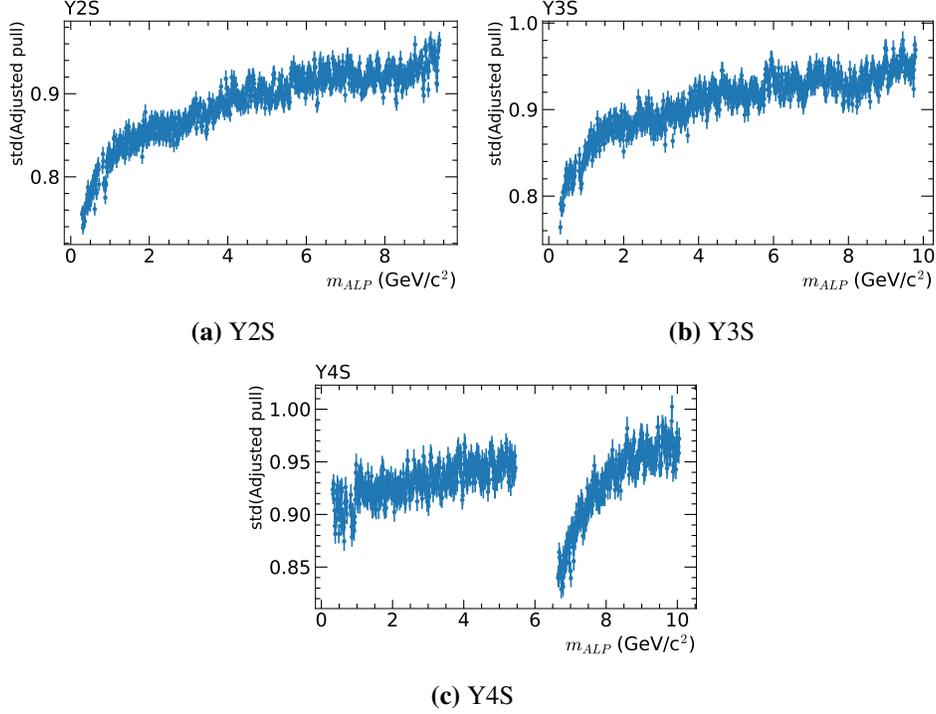


Figure 3.54: Scatter plots of the standard deviation of the adjusted pull as a function of the ALP mass hypothesis, in signal plus background MC fits.

luminosity \mathcal{L} and the cross section σ is

$$N_{\text{fit}} = \mathcal{L} \sigma \epsilon. \quad (3.1)$$

The efficiency is evaluated from the signal plus background fits, described in Section 3.6.6. As not all the signal MC events that passed the selection cuts are added to the toy experiments, the effective number of generated ALP events $n_{\text{gen}}^{\text{eff}}$ is

$$n_{\text{gen}}^{\text{eff}} = \frac{n_{\text{added}}}{n_{\text{cut}}} n_{\text{gen}}, \quad (3.2)$$

where n_{added} is the number of signal MC events added to the toy experiments, n_{gen} is the number of generated signal MC events, and n_{cut} is the number of signal MC events that pass the selection cuts. The events counted in n_{added} do not necessarily

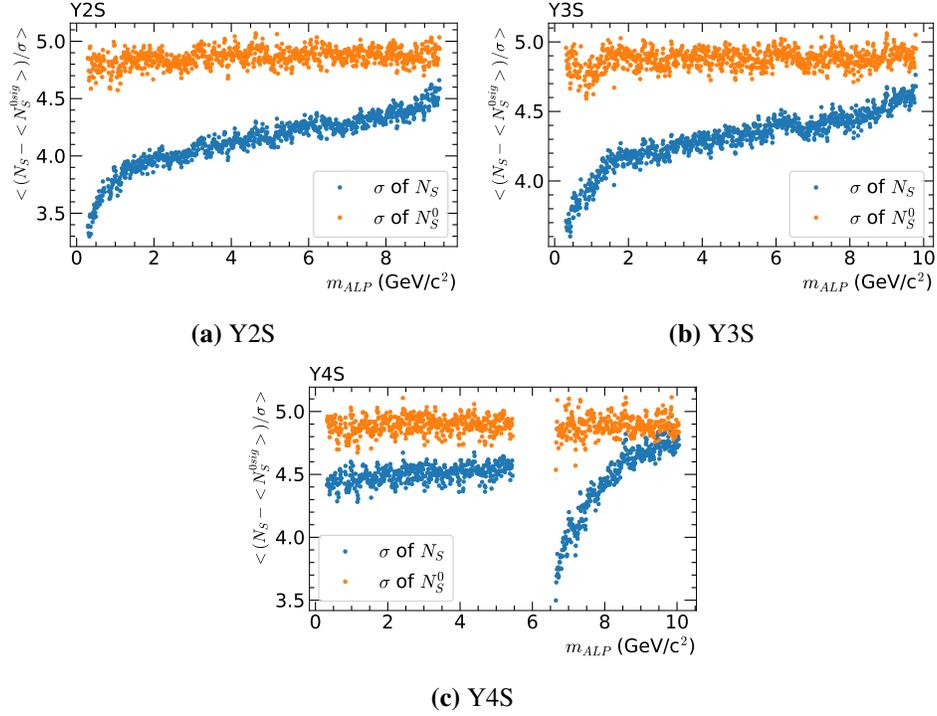


Figure 3.55: Scatter plots of the mean of the adjusted local significances as a function of the ALP mass hypothesis, in signal plus background MC fits. The blue points use the uncertainty on the number of signal entries of the signal plus background fits. The orange points use the uncertainty on the number of signal entries of the background only fits.

have an ALP candidate in the fit window.

The efficiency is then calculated as

$$\varepsilon = \frac{N_{\text{fit}}}{n_{\text{gen}}^{\text{eff}}} \kappa, \quad (3.3)$$

where κ is the scale factor that corrects for the disagreement between data and MC, as listed in Table 3.5 and is equal to about 0.8. Figure 3.56 shows the efficiency as a function of ALP mass. The efficiency for the $Y(2S)$ and $Y(3S)$ CMEs is relatively low at ALP masses below $3.75 \text{ GeV}/c^2$ because the number of entries per event is 0.6. The efficiency increases in the ALP mass range $3.75 \text{ GeV}/c^2$ to $7 \text{ GeV}/c^2$

because the number of “wrong” ALP candidates that fall within the fit window gradually increases. This has the effect of gradually increasing the average number of ALP candidates per event up to around 1.4. A “wrong” ALP candidate is any ALP candidate that is not formed from the two ALP daughter photons. For example, an ALP candidate formed from an ALP daughter photon and the recoil photon. Above $7 \text{ GeV}/c^2$ the number of “wrong” ALP candidates inside the fit window falls sharply to one, causing the efficiency to fall sharply. The gradual increase above $7.5 \text{ GeV}/c^2$ is due to an increase in the selection cut efficiency. The same effects happens for the $Y(4S)$ as well, but they are not as apparent because of the large variation of the trigger efficiency for this CME.

3.7 Search of the data for evidence of an ALP

In order to search the 5% of the *BABAR* data for evidence of an ALP, the fitting procedure is repeated on the data. This is the same data sample that is used for the selection cut optimization, therefore, as discussed in Section 3.1.1, the results might be biased. The local significance, as well as the global significance are evaluated.

3.7.1 Local significances

The local significance is the naive significance N_S/σ_{N_S} for each ALP mass hypothesis, without taking into account the fact that many ALP mass hypotheses are used. Even though the local significance is not the correct statistical quantity used for discovery, it is still interesting to evaluate it and inspect which ALP mass had the largest significance. Figure 3.57 shows histograms of the local significances for the three CMEs. Figure 3.58 shows the fits to the data that gave the maximum local significance for each of the CMEs.

These figures show that the largest local significances in 5% of the data do not approach the threshold for discovery of five standard deviations. Even if they were large enough, to actually claim a discovery the global significance should be used.

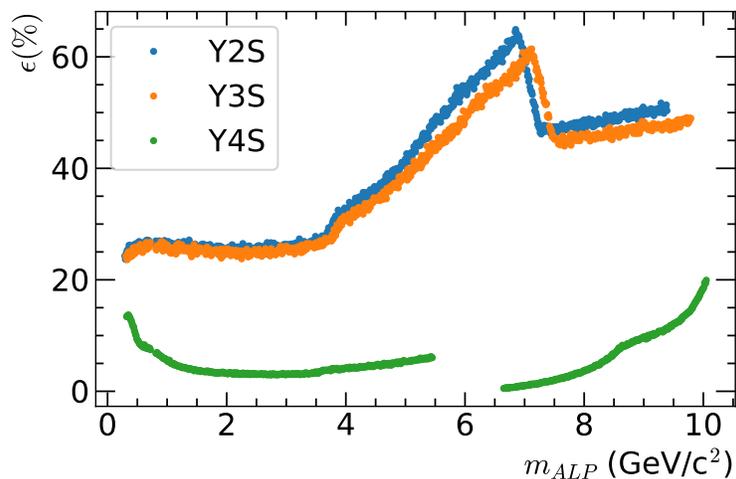


Figure 3.56: The efficiency ε as a function of the ALP mass hypothesis.

3.7.2 Global significances

The local significance does not take into account the number of ALP mass hypotheses. Just due to statistics, the more ALP mass hypotheses are used, the higher the chance that at least one of them will have a high local significance result. This is known as the look elsewhere effect.

To take into account the number of ALP mass hypotheses used, the histogram of the maximum adjusted local significance of each of the MC toy experiments is created. These are created by looking at all the local significances for a particular toy experiment, one value for each ALP mass hypothesis, after correcting for the bias on N_S observed in the toy MC. The maximum local significance value from each toy is used to fill the histogram, one entry per toy experiment. These histograms are treated as the PDF for the maximum local significance. The p-value, the probability of obtaining a maximum local significance that is greater or equal to the one measured in data, is then evaluated. Figure 3.59 shows the histograms, the maximum local significance observed in data, the p-value, and the p-value equivalent in Gaussian standard deviations. The p-values for the three CMEs are large, meaning there is a high probability that a random fluctuation of the background will produce a maximum local significance at least as large as observed in data. As the p-values

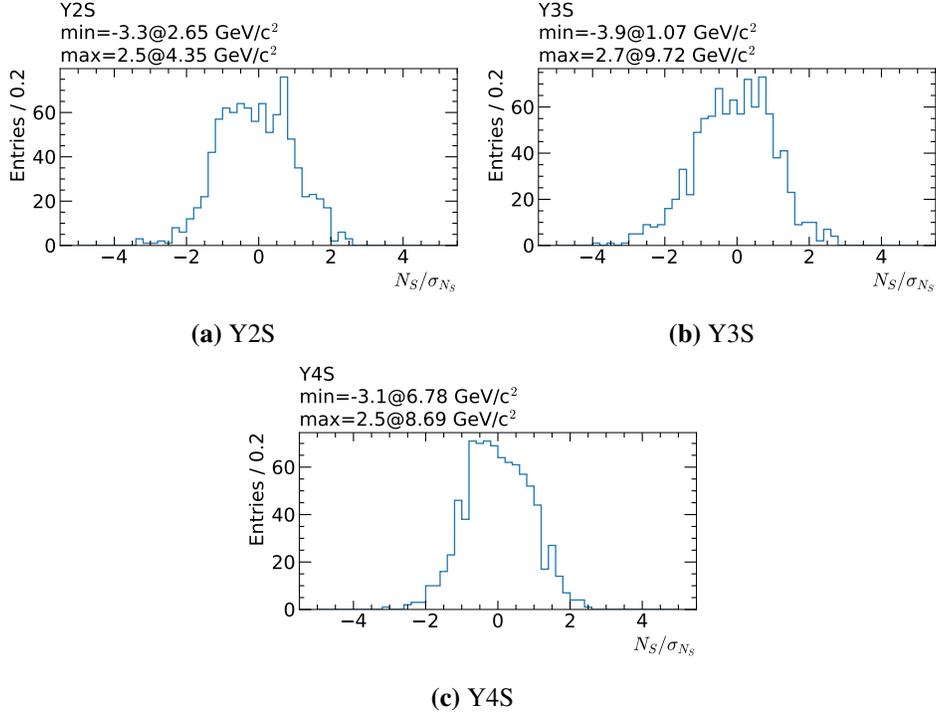


Figure 3.57: Histograms of the local significances of fits to 5 % of the data. The top of each subplot has the ALP masses that produced the minimum and maximum local significance.

for the three CMEs are all greater than one half, the equivalent Gaussian standard deviations are negative. Analyzing the global significance shows that 5 % of the data is completely consistent with the data being composed only of background, with no hints of an ALP. Even if the results had suggested that an ALP existed, a discovery could be claimed only by performing a blind analysis on the remaining 95 % of the data, without re-optimizing the selection cut threshold values.

3.8 Cross section and coupling constant upper limits

Failing to find indications for an ALP in the data, the data is used to set 90 % CI upper limits on the ALP production CS σ . As each CME has a different CS, the measurements at the three CMEs are later combined to produce upper limits on the

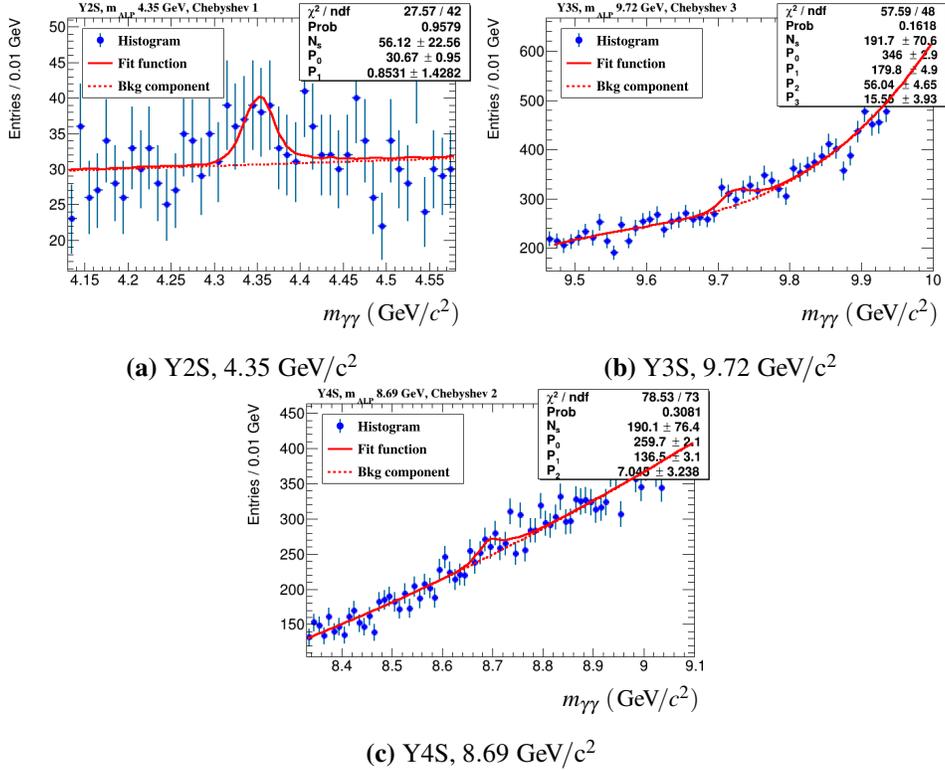


Figure 3.58: Fits to 5 % of the data for the ALP mass hypotheses that gave the maximum local significance for each CME. The solid red line is the complete fit function. The dashed red line is the fit function without the signal component.

ALP-photon coupling $g_{\gamma\gamma}$. Limits are set using Bayesian methods [53].

3.8.1 CS upper limits with statistical uncertainties only

As a first step, the limits are calculated taking into account statistical uncertainties only. The statistical uncertainty is the uncertainty on the number of signal entries returned by the fit.

The likelihood function, i.e., the probability of measuring the number of signal entries observed in data given the model, for the measured number of signal entries N_S and its uncertainty δ_S given σ , \mathcal{L} , and ε from Equation 3.1 is modeled by a

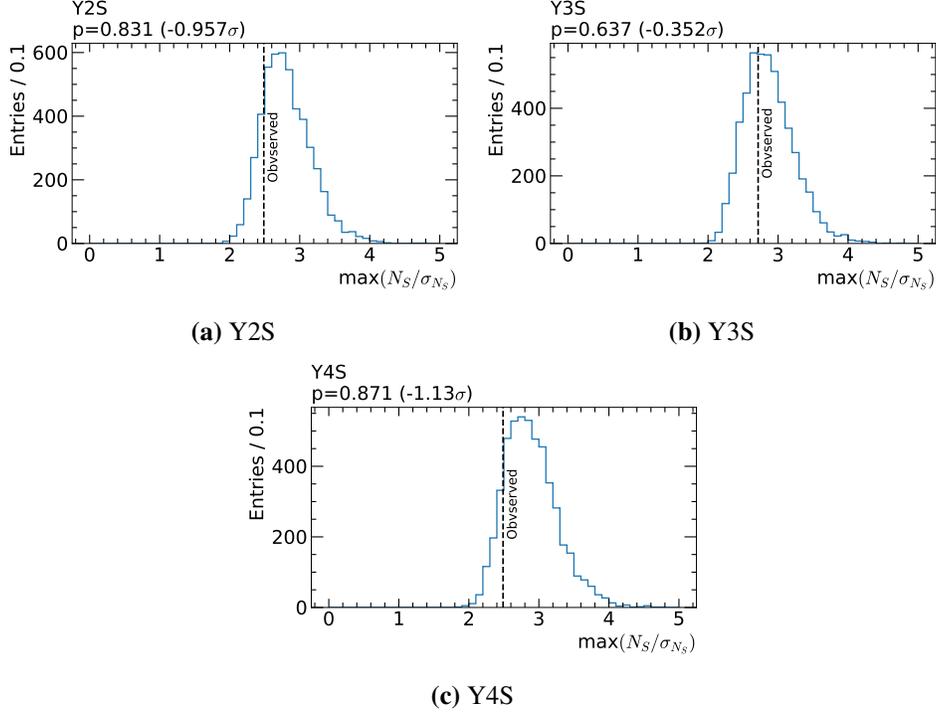


Figure 3.59: Histograms of the maximum adjusted local significances of the MC toy experiments. The vertical dashed line shows the maximum local significance in 5% of the data. Above each subplot are the p-value and the equivalent number of Gaussian standard deviations of the observed value.

Gaussian with mean N_S and standard deviation δ_S

$$P(N_S, \delta_S | \sigma) = \frac{1}{\delta_S \sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{\mathcal{L}\sigma\epsilon - N_S}{\delta_S} \right)^2 \right]. \quad (3.4)$$

A flat non-negative prior is used for the CS up to some unimportant large value σ_{\max} . The posterior, i.e., the PDF of the CS given the data, is

$$P(\sigma | N_S, \delta_S) = \frac{\frac{1}{\delta_S \sqrt{2\pi}} \exp \left(-\frac{1}{2} \left(\frac{\mathcal{L}\sigma\epsilon - N_S}{\delta_S} \right)^2 \right) \frac{1}{\sigma_{\max}}}{\int_0^{\infty} \frac{1}{\delta_S \sqrt{2\pi}} \exp \left(-\frac{1}{2} \left(\frac{\mathcal{L}\sigma\epsilon - N_S}{\delta_S} \right)^2 \right) \frac{1}{\sigma_{\max}} d\sigma}. \quad (3.5)$$

To calculate the 90 % CI upper limit on σ , σ_{UL} , the following equation needs to be solved

$$0.9 = \frac{\int_0^{\sigma_{\text{UL}}} \exp \left[-\frac{1}{2} \left(\frac{\sigma - N_S / \mathcal{L} \varepsilon}{\delta_S / \mathcal{L} \varepsilon} \right)^2 \right] d\sigma}{\int_0^{\infty} \exp \left[-\frac{1}{2} \left(\frac{\sigma - N_S / \mathcal{L} \varepsilon}{\delta_S / \mathcal{L} \varepsilon} \right)^2 \right] d\sigma}. \quad (3.6)$$

The analytical solution to this equation is

$$\sigma_{\text{UL}} = N_S / \mathcal{L} \varepsilon + \sqrt{2} \delta_S / \mathcal{L} \varepsilon \cdot \text{erf}^{-1} \left(0.9 + (0.9 - 1) \text{erf} \left(\frac{N_S}{\sqrt{2} \delta_S} \right) \right). \quad (3.7)$$

The green curve in Figure 3.61 shows the 90 % CI upper limit on the CS using statistical uncertainties only, for 5 % of the data.

3.8.2 Systematic uncertainties

Systematic uncertainties are calculated for the number of signal entries N_S and for the signal efficiency ε .

Number of signal entries

Two effects are taken into account when calculating the systematic uncertainty on the number of signal entries. The first is the background polynomial degree. Section 3.6.3 described how the background polynomial degree is selected. If the selected degree is smaller than the true degree needed to describe the background, the number of signal events returned by the fit will be wrong. This can happen, for example, due to limited statistics of the MC sample.

To evaluate the effect this has on the number of signal events returned by the fit, the polynomial degree is increased by one and the fit procedure is repeated. The absolute value of the difference between the number of signal events with the nominal polynomial degree and the increased polynomial degree is used as a systematic uncertainty.

Theoretically, it could be possible for the selected polynomial degree to be too high, instead of too low. But the statistics are high enough that decreasing the polynomial degree causes fits for many masses to fail. For example, if the background is truly described by a second degree polynomial and the statistics are high enough, fitting it with a first degree polynomial will fail. Therefore, the

systematic uncertainty is evaluated only by increasing the polynomial degree.

The second effect contributing to the systematic uncertainty on the number of signal entries is the fit window size. The nominal fit window size is five times the size of the base width, as described in Section 3.6.4. If the nominal fit window size is not wide enough to extract the true shape of the background, the fit will return the wrong number of signal entries. To evaluate how this affects the fitted number of signal events, the fit procedure is repeated with a fit window six times larger than the counting window. The absolute value of the difference between the number of signal events with the nominal fit window range and the wider fit window is used as a systematic uncertainty.

Figure 3.60 shows the distributions of the ratios of the systematic uncertainty on the number of signal entries because of the Chebyshev polynomial degree and because of the fit window size to the statistical uncertainty on the number of signal entries. These give a sense of the size of the systematic uncertainties on the number of signal entries.

Because the bin width is much smaller than the ALP candidate mass resolution, the systematic uncertainty that could result from changing the bin width is expected to be negligible and is not evaluated.

Signal efficiency

The systematic uncertainty on the signal efficiency is roughly 11 %. The vast majority of the uncertainty is because of the scaling of the MC to fit the data, κ , described in Section 3.4. This effect is included in the systematic calculation because it can cause the number of signal MC entries to be too large. The uncertainty on κ is calculated by $(1 - \kappa)/2$. Any systematic uncertainties on the integrated luminosities, the trigger, and on the photon detection efficiency are incorporated into κ .

Even though the selection cut threshold values affect the background shape, the analysis does not depend on the knowledge of the exact shape of the background. The polynomial coefficients describing the shape of the background are floating parameters in the fit function. Therefore, the exact values of the selection cut thresholds are not considered a source of systematic uncertainty.

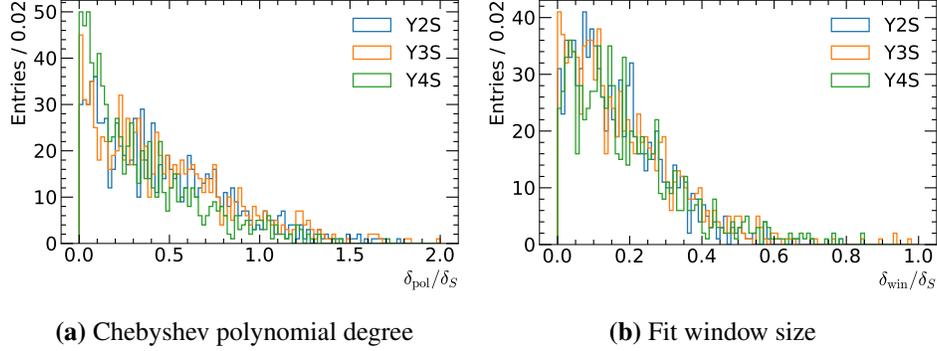


Figure 3.60: Distributions of the ratios of the systematic uncertainty on the number of signal entries because of the Chebyshev polynomial degree (a) and the fit window size (b) to the statistical uncertainty on the number of signal entries.

3.8.3 CS upper limits with statistical and systematic uncertainties

The systematic uncertainties on N_S and ϵ are added to the limit calculation in different ways.

The equation used to calculate the limit, Equation 3.5 already includes a term for the statistical uncertainty on N_S , δ_S . The systematic uncertainty on N_S is incorporated into the limit by modifying δ_S to be the square root of the quadratic sum of the statistical and systematic uncertainties

$$\delta_S^{\text{tot}} = \sqrt{\delta_S^2 + \delta_{\text{pol}}^2 + \delta_{\text{win}}^2}, \quad (3.8)$$

where δ_{pol}^2 is the systematic uncertainty due to modifying the background polynomial degree, and δ_{win}^2 is the systematic uncertainty due to modifying the fit window size. This assumes there is no correlation between δ_{pol}^2 and δ_{win}^2 .

In order to add the systematic uncertainty on ϵ , a Gaussian term for the probability of ϵ is added to Equation 3.5. The Gaussian has mean $\bar{\epsilon}$, which is the efficiency obtained in Section 3.6.7, and standard deviation δ_ϵ , which is the uncertainty on the obtained efficiency. The efficiency is then marginalized out. The

modified equation that includes the systematic uncertainties is

$$0.9 = \frac{\int_0^{\sigma_{\text{UL}}} d\sigma \int_0^1 d\epsilon \frac{1}{\delta_s^{\text{tot}} \sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\mathcal{L}\sigma\epsilon - N_S}{\delta_s^{\text{tot}}}\right)^2\right] \frac{1}{\delta_\epsilon \sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\epsilon - \bar{\epsilon}}{\delta_\epsilon}\right)^2\right]}{\int_0^\infty d\sigma \int_0^1 d\epsilon \frac{1}{\delta_s^{\text{tot}} \sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\mathcal{L}\sigma\epsilon - N_S}{\delta_s^{\text{tot}}}\right)^2\right] \frac{1}{\delta_\epsilon \sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\epsilon - \bar{\epsilon}}{\delta_\epsilon}\right)^2\right]}. \quad (3.9)$$

This equation is solved numerically. Figure 3.61 shows the upper limits obtained for the three CMEs. The green curve shows the limits calculated with statistical uncertainties only. The black curve shows the limits with the systematic uncertainties included.

To get a sense of the effect of the systematic uncertainties on the limit, Figure 3.62 shows the ratio of the limit with the systematic and statistical uncertainties to the limit calculated with only the statistical uncertainties. The systematic uncertainties degrade the upper limit by roughly 5%. The uncertainty on N_S produces the long high side tail in Figure 3.62. The uncertainty on ϵ increases the limit by roughly 5% for all mass hypotheses, shifting the distribution to the right.

3.8.4 Coupling constant upper limits

Instead of calculating the limit on the CS, the limit on the coupling constant can be calculated instead. A CS value can be converted into a coupling constant value. The CS is proportional to the coupling constant squared, $\sigma(g) \propto g^2$ [21]. The information provided by Madgraph while simulating the signal MC is used to relate the cross section to the coupling constant

$$\sigma = g^2 \sigma_{\text{mad}} / g_{\text{mad}}^2, \quad (3.10)$$

where σ_{mad} is the CS reported by Madgraph and g_{mad} is the coupling constant set in Madgraph, which in simulated signal MC is equal to $g_{\text{mad}} = 10^{-3} \text{ GeV}^{-1}$. The upper limit on the coupling constant is calculated in the same way to the one on the CS, with a flat prior for the coupling constant and a likelihood function for the coupling constant

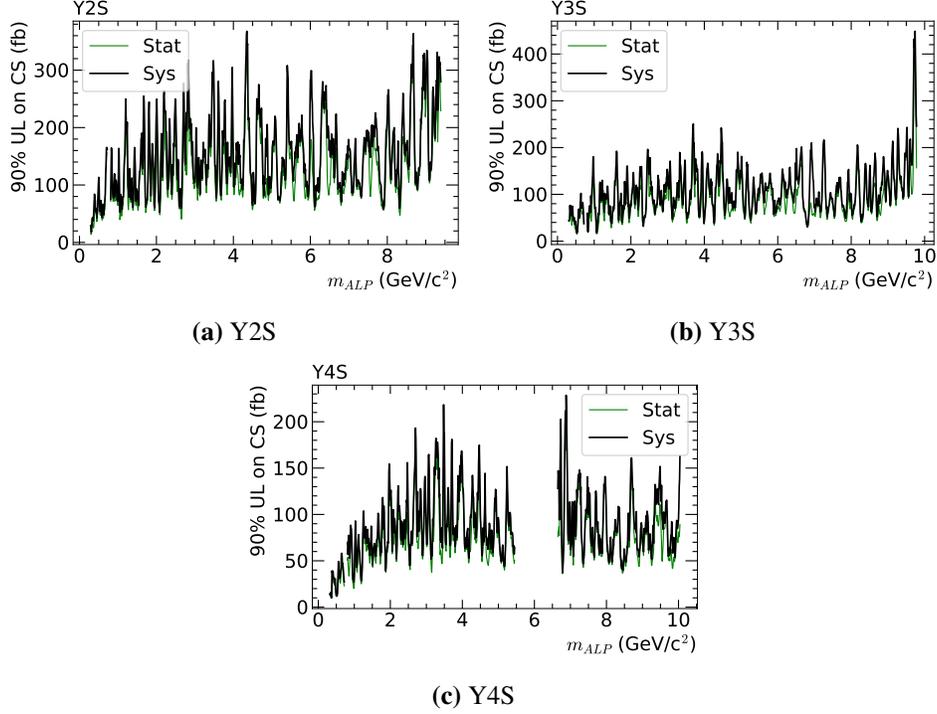


Figure 3.61: 90 % CI upper limits on the ALP production cross section, using 5 % of the data. The green curve shows the limit calculated with statistical uncertainties only. The black curve shows the limit calculated with both statistical and systematic uncertainties.

$$\begin{aligned}
 P(N_S, \delta_S | g) = & \frac{1}{\delta_S^{\text{tot}} \sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{g^2 - N_S g_{\text{mad}}^2 / \mathcal{L} \sigma_{\text{mad}} \epsilon}{\delta_S g_{\text{mad}}^2 / \mathcal{L} \sigma_{\text{mad}} \epsilon} \right)^2 \right] \cdot \\
 & \frac{1}{\delta_\epsilon \sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{\epsilon - \bar{\epsilon}}{\delta_\epsilon} \right)^2 \right].
 \end{aligned} \tag{3.11}$$

The upper limit on the coupling constant is calculated separately for each CME. Figure 3.63 shows the ratio of the limit on the coupling constant calculated with a flat prior on the coupling constant over the limit on the CS converted into a limit on the coupling constant. The limit on the CS is calculated with a flat prior on the CS. The figure shows that calculating the limit directly on the coupling constant, with

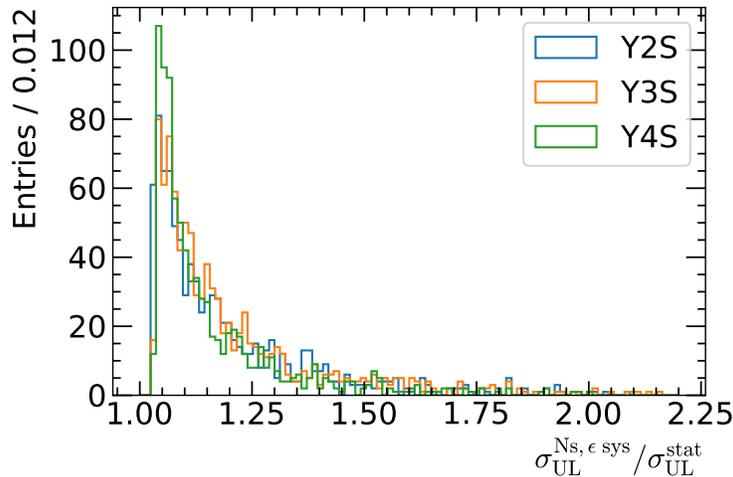


Figure 3.62: Ratio of the CS upper limit calculated with systematic and statistical uncertainties to the limit calculated with statistical uncertainties only. The plots use 5 % of the *BABAR* data.

a flat prior on the coupling constant, produces results better by about 15 %.

The main advantage of using the likelihood for the coupling constant is that it allows to combine the measurements at the three CMEs. As the cross section is different at each CME, the coupling constant is the quantity that is constant with energy. The combined likelihood is the product of the likelihoods in Equation 3.11 for each of the CMEs. Figure 3.64 shows the combined limit on the coupling constant using the three CMEs measurements.

3.9 Phase space exclusion

Figure 3.65, reproduced from [1], shows the ALP phase space with different regions excluded by different types of experiments. It also shows the result of this work, which uses 5 % of the data collected by *BABAR*. As can be seen, the results exclude regions of the phase space that have not been explored before in the mass range $0.29 \text{ GeV}/c^2$ to $5 \text{ GeV}/c^2$. It should be kept in mind, however, that the analysis is not a blind analysis, as discussed in Section 3.1.1.

The excluded coupling constants are large enough that the ALP's proper decay

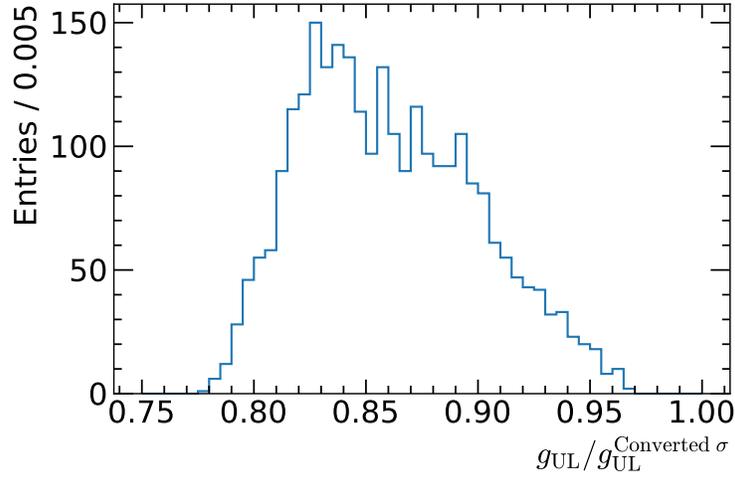


Figure 3.63: Histogram of the ratio of the coupling constant upper limit and the CS upper limit converted into a limit on the coupling constant. The limit on the coupling constant is calculated with a flat prior on the coupling constant and the limit on the CS is calculated with a flat prior on the CS. 5 % of the *BABAR* data is used.

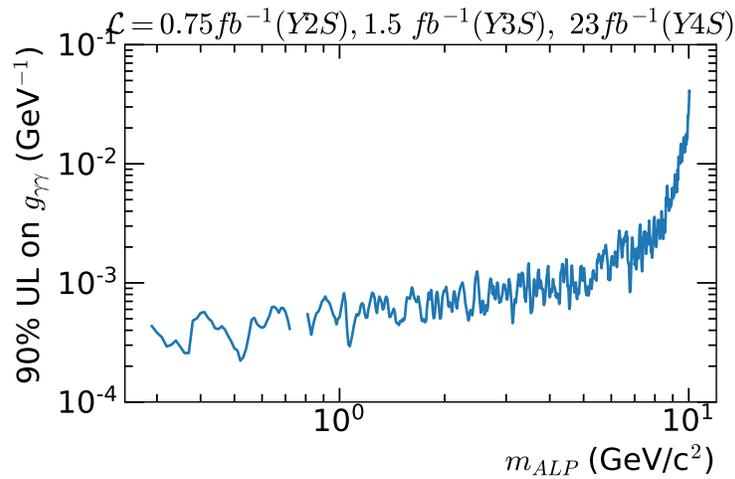


Figure 3.64: 90 % CI upper limit on the coupling constant, combining the measurements at the $\Upsilon(2S)$, $\Upsilon(3S)$, and $\Upsilon(4S)$ CMEs. 5 % of the *BABAR* data is used.

length in the lab frame is less than 0.2 mm for the lowest masses excluded here. This is short enough that the ALP can be assumed to be produced at the IP. This assumption should be re-evaluated after the full *BABAR* dataset is analysed and more stringent limits are produced.

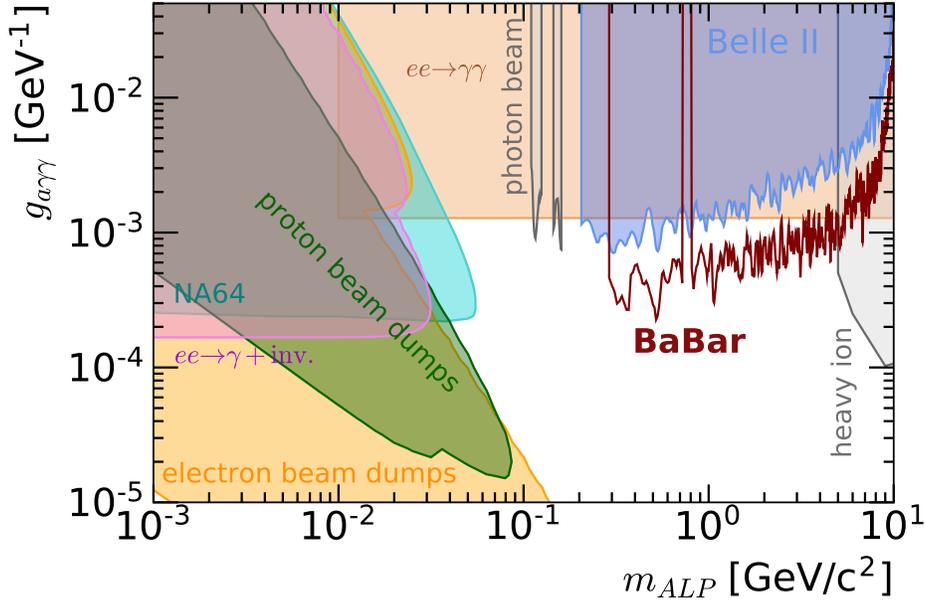


Figure 3.65: Phase space exclusion plot in the coupling constant - ALP mass plane. The plot is a reproduction of Figure 5 from [1]¹ with updated proton beam dump limits from [29]. Superimposed are the result of this work which are based on 5 % of the *BABAR* data.

¹© 2021 American Physical Society

3.9.1 Comparison of the Belle II and *BABAR* limits

The limits of *BABAR* and Belle II are calculated with different integrated luminosities. This analysis uses 25.5 fb^{-1} which is 5 % of the *BABAR* data. The published Belle II analysis is based on 445 pb^{-1} .

In order to compare the Belle II limits to the limits calculated in this work, while taking into account the differences in integrated luminosities, the Belle II

limits are scaled by $\sqrt[4]{\mathcal{L}}$ to a total integrated luminosity of 25.5 fb^{-1} . The limits on the coupling constant scale like $\sqrt[4]{\mathcal{L}}$ because the limit on the cross section scales like the uncertainty on the number of signal events which, assuming Poisson statistics, scales as the square root of the number of entries. The coupling constant scales like $g_{\gamma\gamma} \propto \sqrt{\sigma}$, as can be seen in Equation 3.10. Combining these produces the $\sqrt[4]{\mathcal{L}}$. Figure 3.66 shows the scaled Belle II limits and the limits calculated here. As can be seen, the limits are very similar to each other. In fact, a more detailed comparison shows that they are practically identical.

There are a number of differences between *BABAR* and Belle II that affect each experiment's sensitivity to an ALP. One advantage of Belle II over *BABAR* is its higher trigger efficiency. Most of *BABAR*'s data is at the $\Upsilon(4S)$ CME. As discussed in Section 3.3.2, *BABAR* had only a two back-to-back photon trigger for the $\Upsilon(4S)$ runs. For the $\Upsilon(2S)$ and $\Upsilon(3S)$ runs, *BABAR* had a single photon trigger requiring at least a single photon. The single photon trigger efficiency is up to nine times higher than the two photon trigger efficiency in the mass range excluding the ALP mass around $6.5 \text{ GeV}/c^2$, where the two photon trigger efficiency has a minimum. Close to the minimum, the trigger efficiency for the single photon trigger is up to eighty times higher than for the two photon trigger. Belle II has had a single photon trigger since it started recording data, and is not affected by the low trigger efficiencies like most of the *BABAR* data. Scaling the Belle II result to the effective *BABAR* integrated luminosity if *BABAR* had a trigger efficiency for the $\Upsilon(4S)$ runs like that of the $\Upsilon(2S)$, $\Upsilon(3S)$ runs shows that limits of *BABAR* and Belle II are comparable at the lowest and highest masses. For most of the mass range, though, *BABAR*'s limits are roughly one and a half times better. This indicates that *BABAR* has superior performance per integrated luminosity.

Another advantage of Belle II is that it has timing information for its reconstructed photons. This allows to reject photon candidates with times that are not compatible with the collision time. This is important in rejecting beam background photons, for example. Photon timing information is not available at *BABAR*.

On the other hand, *BABAR* has a number of advantages over Belle II. One of these is that *BABAR* has kinematic fitting which improves the ALP candidate mass resolution. The ALP candidate mass resolution is comparable between the experiments for ALP masses of $0.5 \text{ GeV}/c^2$ and $9 \text{ GeV}/c^2$. For ALP masses of $5 \text{ GeV}/c^2$

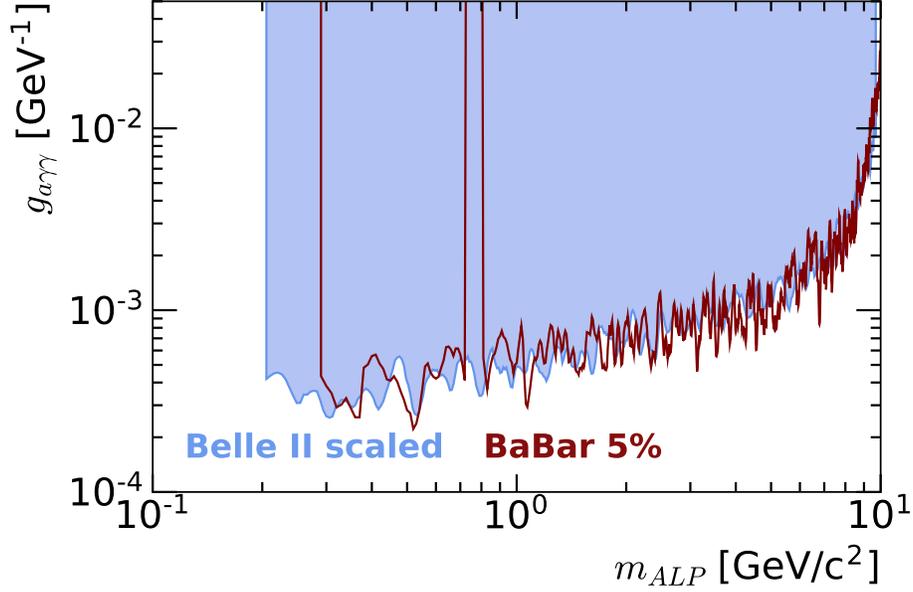


Figure 3.66: Comparison of the phase space exclusion of Belle II and *BABAR*. The Belle II result [1] is scaled by $\sqrt[4]{\mathcal{L}}$ to 5% of the integrated luminosity collected by *BABAR* and used in this work.

and $7.5 \text{ GeV}/c^2$, *BABAR*'s mass resolution is three to three and a half times better than Belle II's. The number of entries in the signal region is proportional to the ALP mass resolution. Therefore, the limits on the coupling constant scale like the fourth root of the ratio of mass resolutions. The published Belle II analysis [1] did not use kinematic fitting, as it was not available in the Belle II software when the analysis was performed. It is planned to be implemented at Belle II in the future, though. Even though Belle II has both a forward and a backward end cap calorimeter, the published Belle II analysis [1] requires all the reconstructed photons to be inside the calorimeter barrel. The *BABAR* analysis uses the barrel as well as the forward end cap. *BABAR* does not have a backward end cap. This means the angular acceptance used in the Belle II analysis is 80% that used in *BABAR* in the CM frame. As the number of candidates is proportional to the acceptance, the limits scale by the fourth root of the ratio of acceptances. If a future Belle II analysis will use both end caps, Belle II's acceptance will be greater than that of *BABAR*. Lastly, even though

Belle II may have had lower beam backgrounds than *BABAR* for their published analysis, they will have higher backgrounds as they gather larger datasets due to the higher instantaneous luminosities.

It should be noted that the effects of these differences are washed out because most of them scale as the fourth root of the ratio between the relevant quantity at Belle II and *BABAR*.

3.10 Future analysis of the blind data

If the remaining data will be analysed in the future, the analysis can follow the exact same procedure described in Sections 3.7 to 3.9 on the remaining 95 % of the data. The 5 % of the data used for optimizing the selection cuts should not be used for the final analysis because of the issues discussed in Section 3.1.1.

As Belle II is currently gathering data at high rate, at some point the Belle II dataset will be larger than that of *BABAR*. Using current projections, Belle II will gather 430 fb^{-1} to 490 fb^{-1} by the end of 2021. This amount of data is comparable to the *BABAR* dataset. This puts into context when a result using the full *BABAR* dataset might not be competitive with a Belle II result.

Figure 3.67 shows the phase space exclusion plot and includes the projected limits using 95 % of the *BABAR* data. These limits are calculated by scaling the limits from Figure 3.65 by $\sqrt[4]{\mathcal{L}}$. The limits are expected to improve by a factor of $\sqrt[4]{19} = 2.1$ when analysing 95 % of the *BABAR* data. The projected limit is smoothed by a rolling mean calculated inside a sliding window that is 30 mass points wide.

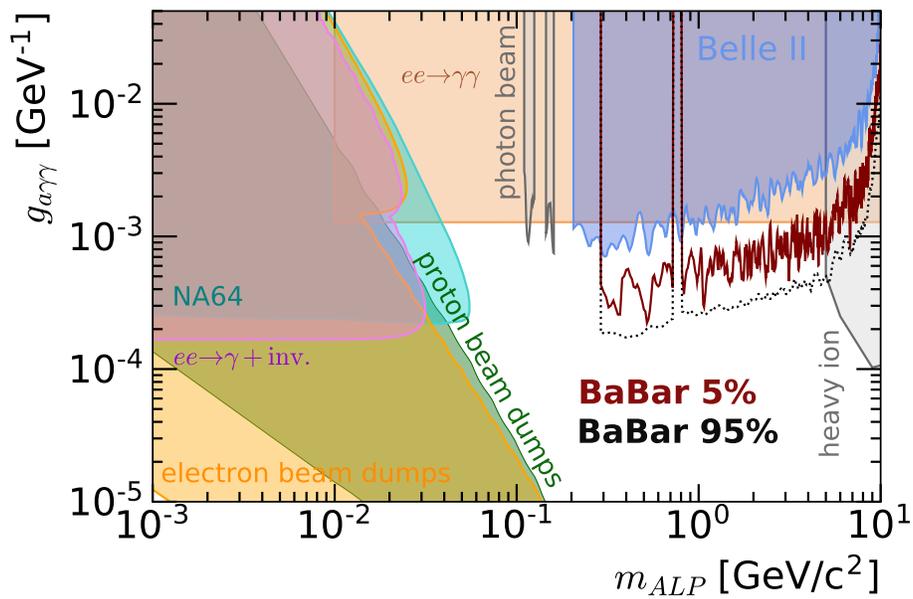


Figure 3.67: Phase space exclusion plot in the coupling constant - ALP mass plane. The plot is a reproduction of Figure 5 from [1]¹ with updated proton beam dump limits from [29]. The solid maroon line is the result of this work based on 5 % of the *BABAR* data. The dotted black line is the projected limits using 95 % of the *BABAR* data, smoothed by a rolling mean inside a sliding window that is 30 mass points wide.

¹© 2021 American Physical Society

Chapter 4

Photon hadron separation using Zernike moments

4.1 Introduction

Whereas Chapter 3 described an analysis using data collected by *BABAR* this chapter describes a project done for Belle II. In general, Belle II and *BABAR* are similar detectors. The main differences relevant to this project are between the two detectors' electromagnetic calorimeters and the higher beam backgrounds at Belle II.

The *BABAR* electromagnetic calorimeter, the EMC, is composed of a barrel and a forward end cap. Belle II's electromagnetic calorimeter (ECL) has, in addition, a backward end cap. This allows it to cover a larger fraction of the solid angle, 12.4° to 155.1° [36]. Another interesting difference is that *BABAR*'s EMC crystals are oriented such that they point at the IP along the polar angle direction. This allows some of the photons to pass without interacting with the EMC. In contrast, Belle II's ECL crystals are rotated in both axes so that there is no gap between the crystals for a particle originating at the IP.

In order to achieve the higher instantaneous luminosities required for SuperKEKB, the *nano beam scheme* is used, in which the width of the beams is decreased and the crossing angle between the beams is increased [54]. At SuperKEKB, the beam width is reduced by a factor of fifteen and the crossing angle is increased by

a factor of four compared to KEKB. Unfortunately, this change causes a large increase in the beam backgrounds. The higher luminosity produces more low angle radiative Bhabha events ($e^+e^- \rightarrow e^+e^-\gamma$) that shower on the beam focusing magnets close to the interaction point and deposit energy in the detector. This process is another source of higher backgrounds.

This chapter describes a project to develop shower shape variables based on Zernike moments that distinguish between electromagnetic and hadronic interactions in the ECL. Shower shapes are variables that depend on how energy is distributed between the crystals when an incident particle interacts with the calorimeter. Different particles produce different distributions of energies. For example, minimum ionizing particles usually leave energy in just one or two crystals [2]. Photons usually create symmetric showers with most of the energy in the center of the shower. K_L^0 mesons can create non-symmetric showers spread out over many crystals. Therefore, shower shape variables can be useful for particle identification [13, 55].

Figure 4.1 shows energies deposited by a photon with an energy of 1.5 GeV shot at the central crystal of a 15×15 crystal matrix. Each crystal is a box with dimensions of $6 \text{ cm} \times 6 \text{ cm} \times 30 \text{ cm}$. The matrix represents a simplified version of the ECL. In contrast, the ECL crystals have shapes of truncated pyramids with a front face typically measuring $5.5 \text{ cm} \times 5.5 \text{ cm}$ and a back face typically measuring $6.5 \text{ cm} \times 6.5 \text{ cm}$. A 1.5 T magnetic field points to the left. Figure 4.2 shows a similar plot, only that a K_L^0 with an energy of 1.5 GeV is shot at the crystal matrix. The differences between the energy distributions left by these two particles can be seen in these plots. The plots are modified versions of ones appearing in [2].

The ECL reconstruction analyses the energy in the ECL crystals and constructs *ECL showers* [2]. Each shower has an energy and position and is assumed to be the result of a single incident particle. The reconstruction begins by grouping all the ECL crystals into groups of adjacent crystals that are assumed to contain energy from one or more incident particles. Each group of crystals is called a *connected region*. To form a connected region, the reconstruction code loops over the crystals and searches for crystals with relatively high energy depositions of 10 MeV. These are the seeds of connected regions. It then searches crystals adjacent to the seed crystal and adds them to the connected region if their energy is above 1.5 MeV.

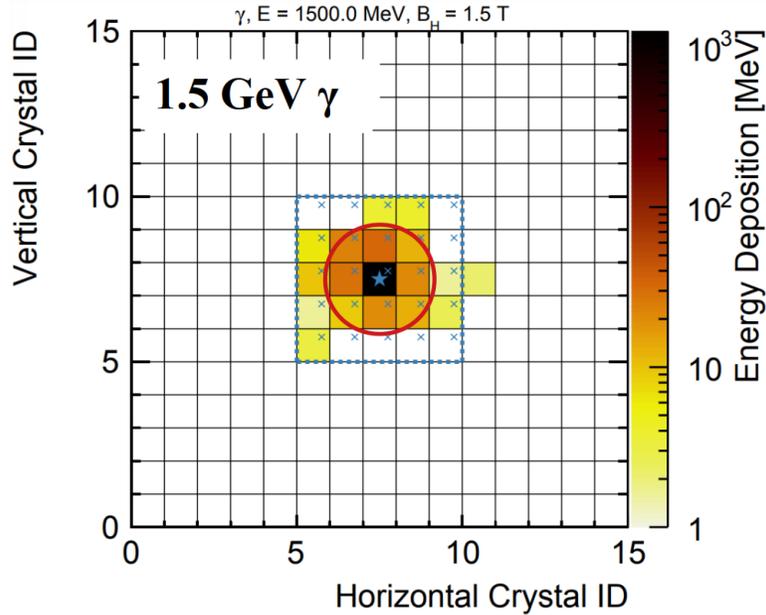


Figure 4.1: The energy deposited in a 15×15 crystal matrix by a photon with an energy of 1.5 GeV photon shot at the central crystal. The matrix is a simplified version of the ECL. Each crystal is a $6 \text{ cm} \times 6 \text{ cm} \times 30 \text{ cm}$ box. The red circle has a radius of 10 cm and is centered on the central crystal. The plot is a modified version of one appearing in [2].

The crystals adjacent to these are then added if their energy is high enough. This is repeated until no more crystals with high enough energy are found adjacent to the crystals in the connected region.

The reconstruction continues by forming ECL showers from the crystals in connected regions based on the hypothesis of which particle(s) interacted with the ECL in a particular connected region. In the hypothesis based reconstruction, each connected region is analysed twice, each with a different hypothesis of the incident particle(s).

In the *photon hypothesis*, N1, the connected region is analysed with the hypothesis that the incident particle(s) are photons, each creating a shower. There can be multiple photon showers in a connected region. Each photon shower is associated with, at most, twenty-five crystals arranged in 5×5 matrix. If a particular crystal is associated with more than one shower, the total energy in the crystal is

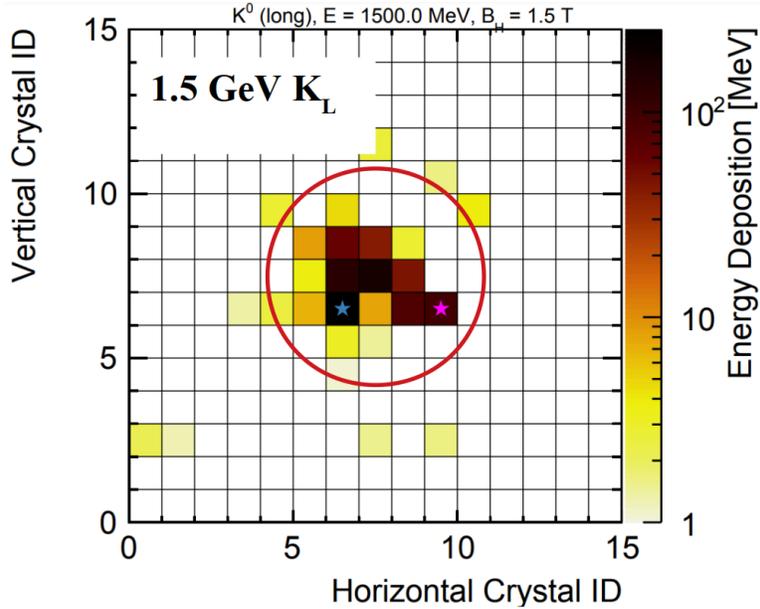


Figure 4.2: The energy deposited in a 15×15 crystal matrix by a photon with an energy of $1.5 \text{ GeV } K_L^0$ shot at the central crystal. The matrix is a simplified version of the ECL. Each crystal is a $6 \text{ cm} \times 6 \text{ cm} \times 30 \text{ cm}$ box. The red circle has a radius of 20 cm and is centered on the central crystal. The plot is a modified version of one appearing in [2].

split between the showers.

In the *neutral hadron hypothesis*, N2, the connected region is analysed with the hypothesis that the incident particle is a neutral hadron. Because hadron cascades are much less radially symmetric and can spread over a larger number of crystals, all the crystals in the connected region are associated with the shower and are used to infer the shower energy and position.

Zernike polynomials were first introduced by Zernike [56]. They are useful in the field of optics and play a vital role in the diffraction theory of optical aberrations [57]. They are also used in image recognition [58]. Zernike moments are used in *BABAR* for particle identification [13] as well as in *ZEUS* to distinguish between particles interacting electromagnetically and hadronically [55]. Zernike moments were not used in Belle.

In this work Zernike moments are introduced into the Belle II software frame-

work, basf2. Eleven Zernike moments are considered. Saving all eleven moments for each ECL shower to disk is not feasible, as it would take up too much space. Therefore, only three values are saved for each shower, two raw Zernike moment values and an output of a multivariate analysis (MVA) using all eleven Zernike moments. The calculation of Zernike moments for ECL showers depends on a number of parameters.

The goals set out for this work are:

- Decide which two raw Zernike moment values to save for each shower and with which parameters to calculate them.
- Decide which parameters to use in the calculation of the eleven Zernike moments for the MVA.

In order to make these decisions, the separation power between K_L^0 and photon events is evaluated using different combinations of Zernike moments calculated with different parameters. The relevant neutral hadrons that reach the ECL are neutrons and K_L^0 mesons. But as K_L^0 mesons are more common in B decays, they are the particle used to represent neutral hadrons in this work.

This chapter uses solely MC samples. There is no data analysed in this chapter.

The work described in this chapter is summarized in an internal Belle II note [59].

4.2 Mathematical background

4.2.1 Zernike polynomials

There are an infinite number of complete sets of complex polynomials in two real variables, $V(x, y)$, which are [57]:

1. Orthogonal for the interior of the unit circle,
2. Contain only polynomials which are invariant in form with respect to rotations of the axes by angle α about the origin. That is, for a polynomial $V(x, y)$, and x', y' rotated coordinates, $V(x, y) = G(\alpha)V(x', y')$, where $G(\alpha)$ is a continuous function with period 2π .

Such sets of polynomials in polar coordinates ($x = \rho \cos \alpha, y = \rho \sin \alpha$) can be shown to have the form [60]:

$$V_{nm}(\rho \cos \alpha, \rho \sin \alpha) = R_{nm}(\rho) e^{im\alpha}, \quad (4.1)$$

where m is an integer. $R_{nm}(\rho) = V(\rho, 0)$ is a polynomial of degree n in ρ , contains no power of ρ lower than $|m|$, and n, m have the same parity. The set of Zernike polynomials is unique in that

3. It contains a polynomial for each permissible pair of values n and m , i.e. for integer values of n and m such that $n \geq 0, n \geq |m|$, and $n - |m|$ is even.

The radial component can be shown to have the form

$$R_{nm}(\rho) = \sum_{s=0}^{\frac{n-|m|}{2}} (-1)^s \frac{(n-s)!}{s! \left(\frac{n+|m|}{2} - s\right)! \left(\frac{n-|m|}{2} - s\right)!} \rho^{n-2s}. \quad (4.2)$$

Figure 4.3 shows surface plots of the Zernike polynomials up to $n = 10$, taken from [3]. In the figure, surfaces with $m \geq 0$ are of the real part of the polynomial and surfaces with $m < 0$ are of the imaginary part. Some of the surfaces are labeled with the name associated with their classical optical aberration function.

4.2.2 Zernike moments

The moments of a function $f(\rho \cos \alpha, \rho \sin \alpha)$ with respect to Zernike polynomial V_{nm} are [58]

$$Z_{nm} = \frac{n+1}{\pi} \int_0^1 \int_0^{2\pi} V_{nm}^*(\rho \cos \alpha, \rho \sin \alpha) f(\rho \cos \alpha, \rho \sin \alpha) \rho d\rho d\alpha. \quad (4.3)$$

The factor before the integrals is a normalization factor which is due to the fact that the Zernike polynomials are orthogonal and not orthonormal.

A property of Zernike moments which will be useful in Section 4.3 is that the absolute value of the Zernike moments is invariant under rotations of the function f around the origin [58]. In other words, the value of $|Z_{nm}|$ is invariant under the transformation $f(\rho, \alpha) \rightarrow f(\rho, \alpha - \alpha_0)$ where α_0 is the angle of rotation.

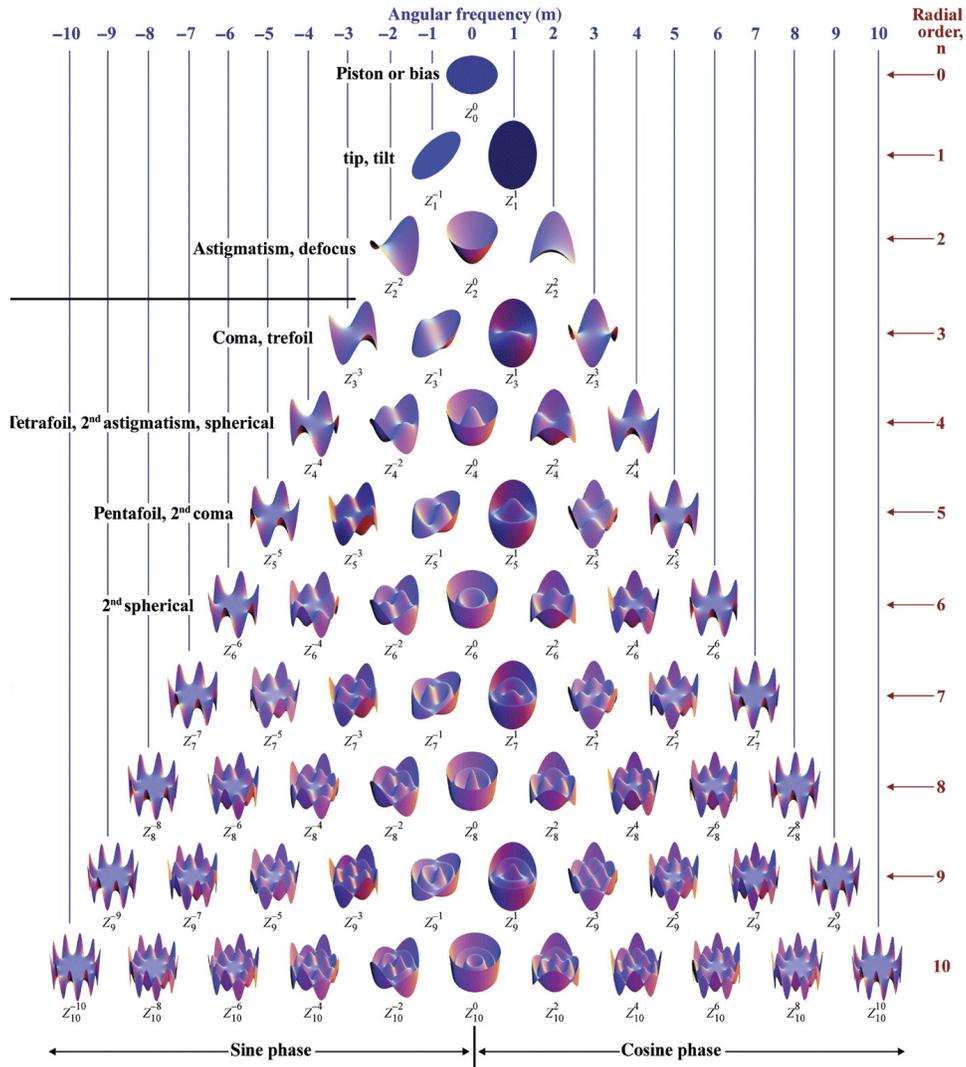


Figure 4.3: Surface plots of the Zernike polynomial with $n \leq 10$. Some surfaces are labeled with the name associated with their classical optical aberration function. The figure is from [3].

Another useful property of the absolute value of the Zernike moments is that $|Z_{n(-m)}| = |Z_{nm}|$ [58]. The analysis sections in this note will only use absolute values of Zernike moments. Therefore, only Zernike moments with positive m values will be discussed.

4.2.3 Zernike moments of a Dirac comb

As will be discussed in Section 4.4, the function $f(\mathbf{x})$ for which the Zernike moments will be calculated is a weighted Dirac comb, that models the discrete ECL crystal energy measurements. A Dirac comb is defined by

$$f(\mathbf{x}) = \sum_i \delta(\mathbf{x} - \mathbf{x}_i) A_i. \quad (4.4)$$

Boldface denotes a vector. In polar coordinates and for $\rho_i \neq 0$, the Dirac delta function can be written as:

$$\delta(\mathbf{x} - \mathbf{x}_i) = \frac{1}{\rho} \delta(\rho - \rho_i) \delta(\alpha - \alpha_i). \quad (4.5)$$

Thus, the function f becomes

$$\begin{aligned} f(\mathbf{x}) &= \sum_i \delta(\mathbf{x} - \mathbf{x}_i) A_i \\ &= \sum_i \frac{1}{\rho} \delta(\rho - \rho_i) \delta(\alpha - \alpha_i) A_i. \end{aligned} \quad (4.6)$$

The Zernike moments of such a function are:

$$\begin{aligned} Z_{nm} &= \frac{n+1}{\pi} \int_0^{2\pi} \int_0^1 V_{nm}^*(\rho \cos \alpha, \rho \sin \alpha) f(\rho \cos \alpha, \rho \sin \alpha) \rho \, d\rho \, d\alpha \\ &= \frac{n+1}{\pi} \int_0^{2\pi} \int_0^1 R_{nm}(\rho) e^{-im\alpha} \sum_i \frac{1}{\rho} \delta(\rho - \rho_i) \delta(\alpha - \alpha_i) A_i \rho \, d\rho \, d\alpha \\ &= \frac{n+1}{\pi} \sum_i R_{nm}(\rho_i) e^{-im\alpha_i} A_i. \end{aligned} \quad (4.7)$$

4.3 Plane perpendicular to the shower direction

Electromagnetic showers are typically symmetric around the incident particle direction. Thus, it will be convenient to calculate the Zernike moments in a plane perpendicular to the shower direction. This plane is termed the *perpendicular plane*. The shower direction used here is defined by the vector \mathbf{p} from the detector coordinate system origin to the measured shower position in the ECL. The perpendicular plane's origin is the measured shower position.

In the following, un-primed vectors, \mathbf{v} , will represent quantities in the detector coordinate system. Primed vectors, \mathbf{v}' , will represent vectors in the perpendicular plane. The perpendicular plane is spanned by two perpendicular unit vectors, $\hat{\mathbf{x}}'$ and $\hat{\mathbf{y}}'$ defined by

$$\begin{aligned}\hat{\mathbf{x}}' &= \frac{\hat{\mathbf{p}} \times \hat{\mathbf{z}}}{|\hat{\mathbf{p}} \times \hat{\mathbf{z}}|}, \\ \hat{\mathbf{y}}' &= \frac{\hat{\mathbf{x}}' \times \hat{\mathbf{p}}}{|\hat{\mathbf{x}}' \times \hat{\mathbf{p}}|},\end{aligned}\tag{4.8}$$

where $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$ are the unit vectors that define the detector coordinate system.

As was discussed in Section 4.2.2, the absolute value of the Zernike moments is invariant under rotations of the coordinate system around the origin. Thus, as long as $\hat{\mathbf{x}}'$ and $\hat{\mathbf{y}}'$ are perpendicular, their orientation in the perpendicular plane will not affect the absolute value of the Zernike moments.

Each crystal in a shower is projected onto the perpendicular plane. The projected point in the plane is represented in polar coordinates (ρ, α) , as shown in Figure 4.4.

Distances in the perpendicular plane are initially calculated with variables that have units. The distances are later scaled by a distance scale factor to obtain dimensionless quantities so that the results of Section 4.2 can be used. The non scaled radial distance r is defined by Equation 4.9 and depicted in Figure 4.5.

$$r = |\mathbf{p}| \tan \psi,\tag{4.9}$$

where ψ is the angle between \mathbf{p} and the vector pointing from the detector origin to the crystal center, \mathbf{c} .

α is the angle between the projected crystal center \mathbf{c}' and $\hat{\mathbf{x}}'$, as shown in Fig-

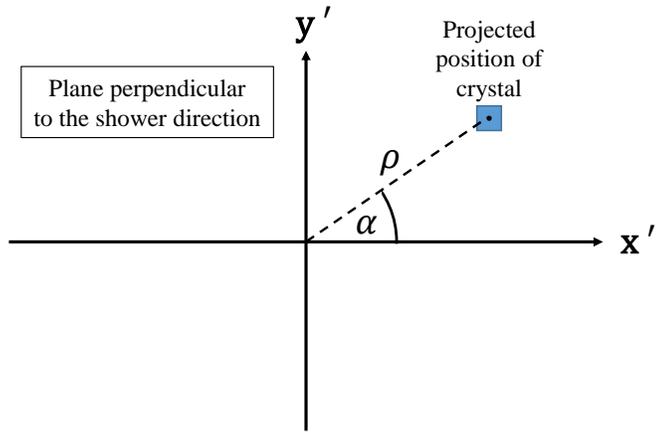


Figure 4.4: Perpendicular plane to shower direction. Each point in the plane is represented by polar coordinates - a radial distance ρ , and a polar angle α .

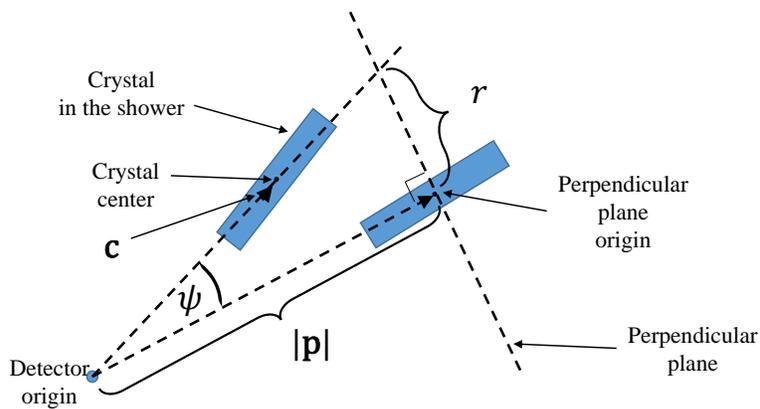


Figure 4.5: Cartoon diagram of the radial distance, r , of a crystal in the plane perpendicular to the shower direction. \mathbf{p} is the vector from the detector coordinate system origin to the shower position, and ψ is the angle between \mathbf{p} and the vector from the detector origin to the crystal center, \mathbf{c} .

ure 4.4. Instead of using \mathbf{c}' to calculate α , a vector parallel to \mathbf{c}' , \mathbf{c}_p , will be used. \mathbf{c}_p is defined by Equation 4.10

$$\mathbf{c}_p = \mathbf{c} - (\mathbf{c} \cdot \hat{\mathbf{p}})\hat{\mathbf{p}} \quad (4.10)$$

and depicted in Figure 4.6.

Note that in general $|\mathbf{c}_p| \neq r$. Nevertheless, as \mathbf{c}_p is only used to find the angle α , its magnitude does not matter.

4.4 Calculation of Zernike moments for ECL showers

In order to use the results of Section 4.2, the projected crystal positions in the perpendicular plane are scaled so they are dimensionless

$$\mathbf{x}' = \frac{\mathbf{X}'}{\rho_0}, \quad (4.11)$$

where \mathbf{X}' is the position of the crystal in the perpendicular plane before scaling, ρ_0 is a distance scale factor, and \mathbf{x}' is the dimensionless crystal position in the perpendicular plane.

Zernike moments are calculated with respect to a function. The function that will be used is a Dirac comb weighted by the normalized crystal energies as defined in Equation 4.12.

$$f(\mathbf{x}') = \sum_i \delta(\mathbf{x}' - \mathbf{x}'_i) \frac{w_i E_i}{\sum_i w_i E_i}, \quad (4.12)$$

where E_i is the energy of the i th crystal, and the sum runs over the crystals in the shower. As a crystal can be related to more than one shower, w_i is the fraction of the energy of the i th crystal associated with the shower in question.

The absolute value of the Zernike moments of ECL showers will be calculated by combining Equation 4.7 with Equation 4.12

$$|Z_{nm}| = \frac{n+1}{\pi} \frac{1}{\sum_i w_i E_i} \left| \sum_i R_{nm}(\rho_i) e^{-im\alpha_i} w_i E_i \right|, \quad (4.13)$$

where $\rho_i = |\mathbf{x}'_i|$ is the dimensionless radial distance.

As the Zernike polynomials are defined only inside the unit circle, only crystals

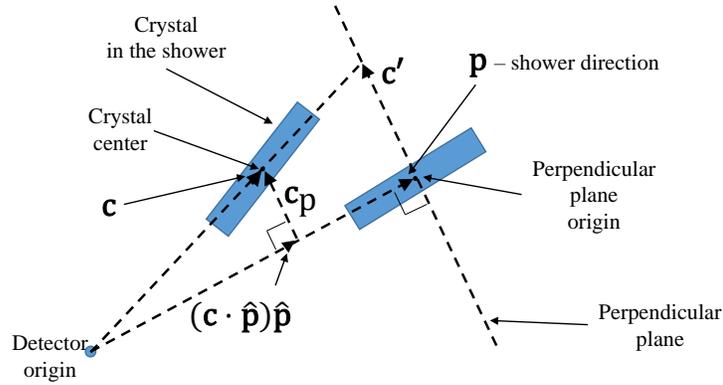


Figure 4.6: Cartoon diagram of how \mathbf{c}_p is calculated. \mathbf{p} is the vector from the detector coordinate system origin to the shower position, \mathbf{c} points from the detector origin to the crystal center, and \mathbf{c}_p is a vector parallel to the vector from the perpendicular plane origin to the projected crystal center.

with $\rho \leq 1$ can be used for the Zernike moment calculation. Crystals with $\rho > 1$ will be called *far crystals*.

Two ways to deal with far crystals are evaluated

- Ignore them in the Zernike moment calculation,
- Reduce their radial distance so that their $\rho = 1$ and include them in the Zernike moment calculation.

The two mentioned parameters that influence the calculation of the Zernike moments, ρ_0 and how to treat far crystals, will be optimized so that the calculated Zernike moments best distinguish between photon events and K_L^0 events.

4.5 Selection of best Zernike moments to distinguish between photons and K_L^0 mesons

In this work, the eleven Zernike moments with n between one to five are considered. Using fewer moments provides less information about the shower. Higher level moments are more susceptible to noise [61]. Using a maximum n of five is

chosen as a compromise. The V_{00} Zernike polynomial is a constant. Therefore, Z_{00} is proportional to the sum of the weighted energies in the shower and is not used in this work.

This section will describe how the best two Zernike moments that distinguish between photon and K_L^0 events are chosen, what are the optimal parameters with which to calculate them, and what are the optimal parameters to calculate the Zernike moments used as input to the MVA that uses all eleven Zernike moments. In order to perform these tasks, MC samples are generated.

4.5.1 MC samples

The MC samples consist of single particle events of either a K_L^0 or a photon. In order to choose at which energies to generate the particles, an initial sample of $1.01 \times 10^6 e^+e^- \rightarrow \Upsilon(4S)$ events is generated with EvtGen [46]. This sample is not fully reconstructed and is used only in order to choose the energies with which to generate the photons and K_L^0 mesons. Figure 4.7 shows the distribution of the momentum of all the K_L^0 particles in the generated sample.

Based on Figure 4.7, K_L^0 momentum ranges are chosen.

- Low momentum range - 100 MeV/c to 200 MeV/c,
- Medium momentum range - 450 MeV/c to 550 MeV/c,
- High momentum range - 1500 MeV/c to 1600 MeV/c.

For each momentum range a photon sample and a K_L^0 sample are generated. Each sample has a single initial particle created at the IP. Each initial particle is generated with a flat polar angle distribution in the range 17° to 150° , which is the acceptance range of the Belle II drift chamber. The predicted background for the full SuperKEKB instantaneous luminosity is used in the simulation. For each particle type and momentum range 5×10^4 events are generated.

For the K_L^0 sample, momenta are chosen from a uniform distribution corresponding to the momenta range. The energies of the photon sample are generated so that the distribution of reconstructed energies of the photons is similar to the distribution of reconstructed energies of the K_L^0 mesons. This is done so that in case the Zernike moments have an energy dependence, the reconstructed energy is

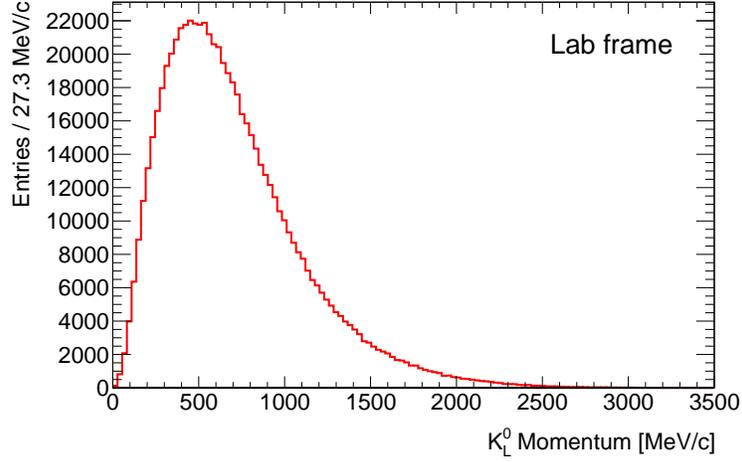


Figure 4.7: Generated K_L^0 momentum in $e^+e^- \rightarrow Y(4S)$ events.

not a discriminating factor between the event types. The generated photon energies are sampled from a distribution defined by a piecewise linear polynomial. The generated photon energies are tweaked until the photon and K_L^0 distributions are similar. Figure 4.8 shows the distributions of the generated photon energies in the lab frame for the three momentum ranges. Figure 4.9 shows the resulting distributions of reconstructed energies after the selection cuts described in Section 4.5.2 were applied for the photon hypothesis and hadron hypothesis showers for the three momentum ranges.

4.5.2 Selection cuts

Selection cuts are applied in order to choose a single N1 photon hypothesis shower and a single N2 neutral hadron hypothesis shower from each event. Showers reconstructed using each hypothesis have to pass all selection criteria. The cuts are

- $t_{\text{rec}} < 125 \text{ ns}$,
- $E_{\text{rec}} \geq 30 \text{ MeV}$,
- Angle between reconstructed shower momentum and primary generated particle momentum in the lab frame $\leq 10^\circ$,

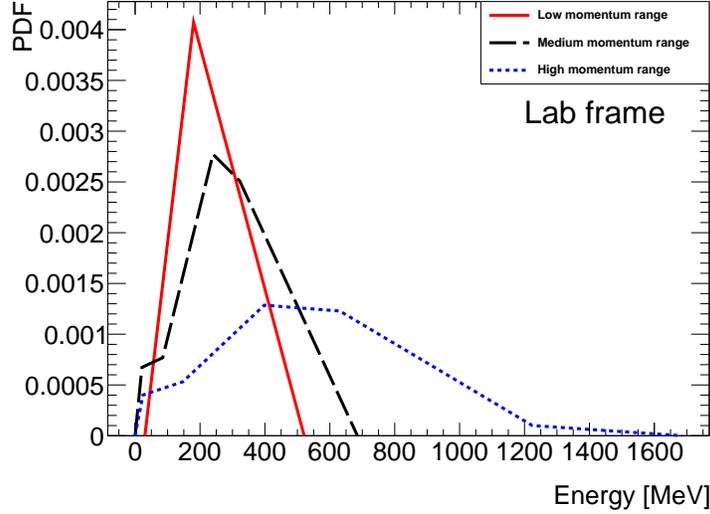


Figure 4.8: Distributions of generated photon energies for the low, medium and high momentum ranges.

where t_{rec} is the reconstructed shower time, and E_{rec} is the reconstructed shower energy in the lab frame. Only one photon hypothesis shower and one neutral hadron hypothesis shower are picked from each event. If more than one photon hypothesis shower passes the selection cuts, the most energetic shower is picked. The selected neutral hadron hypothesis shower is the one from the same connected region as the selected photon hypothesis shower. If an event does not have a pair of showers, one photon hypothesis shower and one neutral hadron hypothesis shower, that both pass the selection cuts, the event is discarded.

Table 4.1 lists the number of events that pass all the selection cuts. Some K_L^0 mesons pass through the ECL without interacting and the showers for some of those that do interact are not fully contained in the ECL. Therefore, the number of reconstructed showers is smaller than the number of generated events for the K_L^0 sample. Figures 4.9 to 4.11 show the reconstructed energy distributions, reconstructed time distributions, and the angle between the primary generated particle's momentum and reconstructed shower momentum, respectively in the three K_L^0 momentum ranges, separately for photon hypothesis and neutral hadron hypothesis showers after all selection cuts are applied.

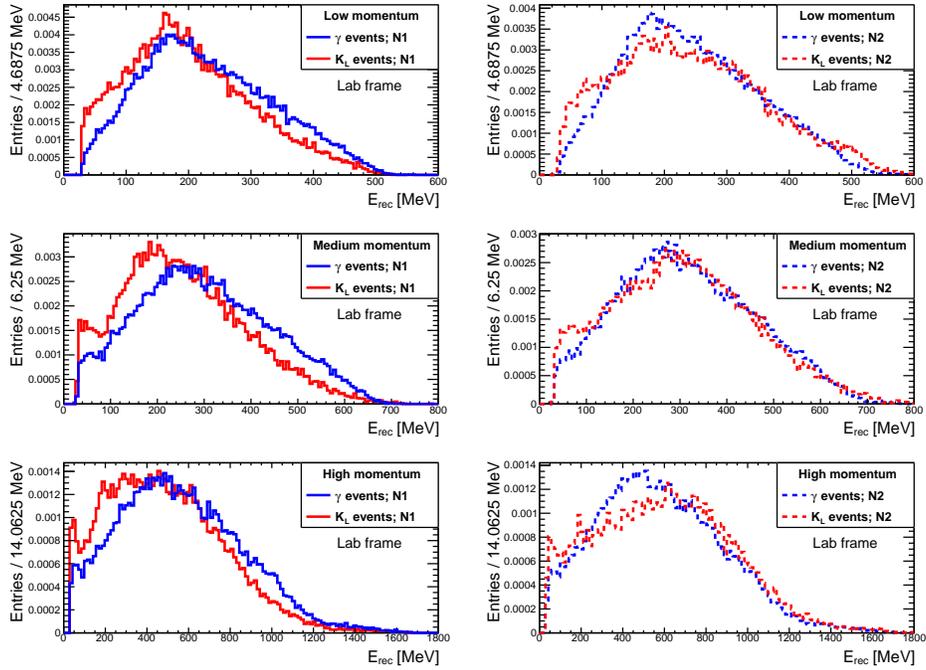


Figure 4.9: Reconstructed energy distributions of showers chosen as described in the text. Figures on the left are of photon hypothesis showers (N1), and figures on the right are of neutral hadron hypothesis showers (N2). The top, center, and bottom rows are for the low, medium, and high momentum ranges, respectively. Each momentum range has a different x-axis scale. The histograms in each plot have an area of one.

4.5.3 Zernike moment distributions

In order to increase statistics and to simplify the analysis, the events from all three momentum ranges are combined into one dataset for the rest of the analysis. Figure 4.12 shows normalized distributions of the absolute value of the Zernike moments for showers reconstructed in the barrel region. The Zernike moments are calculated with $\rho_{0;N1} = 10\text{cm}$, $\rho_{0;N2} = 20\text{cm}$, and using the far crystals. $\rho_{0;N1}$ ($\rho_{0;N2}$) is the ρ_0 used to calculate Zernike moments of N1 (N2) showers. Each plot has four curves. The blue curves are of events with a generated photon and the red curves are of events with a generated K_L^0 . The solid curves are for showers reconstructed under the photon hypothesis and the dotted curves are for showers

Table 4.1: Number of events that passed the selection cuts for the different detector regions, momentum ranges, and particle types.

		γ	K_L^0
Low momentum range	Forward endcap	5288	2528
	Barrel	34296	19852
	Backward endcap	6517	3052
	Entire ECL	46101	25432
Medium momentum range	Forward endcap	5209	2736
	Barrel	33892	18070
	Backward endcap	6609	3394
	Entire ECL	45710	24200
High momentum range	Forward endcap	5326	3335
	Barrel	34546	20866
	Backward endcap	6839	4113
	Entire ECL	46711	28314

reconstructed under the neutral hadron hypothesis.

The ρ_0 used in calculating the Zernike moments has a large impact on some of the distribution shapes. For example, Figure 4.13 shows the same events as in Figure 4.12 only that the Zernike moments of the photon hypothesis showers are calculated with $\rho_{0;N1} = 20$ cm.

The dependence of the Zernike moments on the reconstructed energy, polar angle, and azimuthal angle are plotted in Figure 4.14, Figure 4.15, and Figure 4.16, respectively. Some Zernike moments have a strong dependence on the reconstructed energy, especially in the $E_{\text{Rec}} < 500$ MeV range. In the barrel region, on average, the value of the Zernike moments does not depend on polar angle. There is some dependence on the polar angle in the gap between the barrel and the forward end cap, at a polar angle of 32° , and in the gap between the barrel and backward end cap and in the backward endcap itself, at a polar angle larger than 129° . The dependence of some Zernike moments on the azimuthal angle might be due to the dependence of the boost on the azimuthal angle. Plotting Figures 4.15, and 4.16 with finer binning (not shown) reveals that there is a periodic dependence of the Zernike moments on the reconstructed polar and azimuthal angles. The pe-

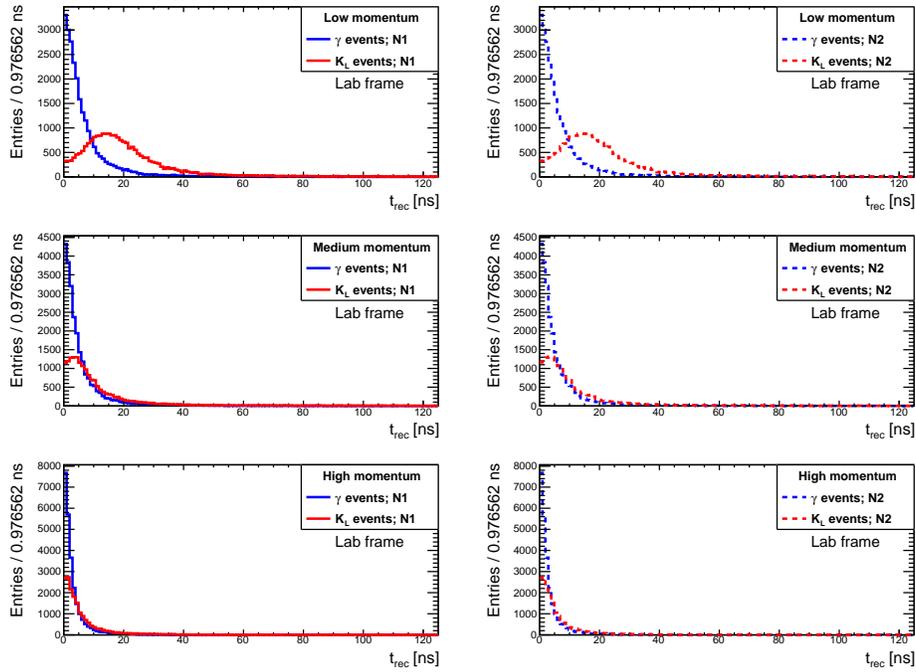


Figure 4.10: Reconstructed time of showers chosen as described in the text. Figures on the left are of photon hypothesis (N1) showers, and figures on the right are of neutral hadron hypothesis showers (N2). The top, center, and bottom rows are for the low, medium, and high momentum ranges, respectively.

riod is the same as the crystal period in each variable. This is due to the fact that in Equation 4.12 the energy is modeled to be deposited at a single point, the center of the crystal.

4.5.4 Selection of the best Zernike moments and the parameters with which to calculate them

In order to compare the usefulness of different combinations of Zernike moments, different MVAs are trained, each one with a different combination of Zernike moments. The ability of the output of each MVA to distinguish between photon and K_L^0 events is checked.

The following combinations of Zernike moments are tested

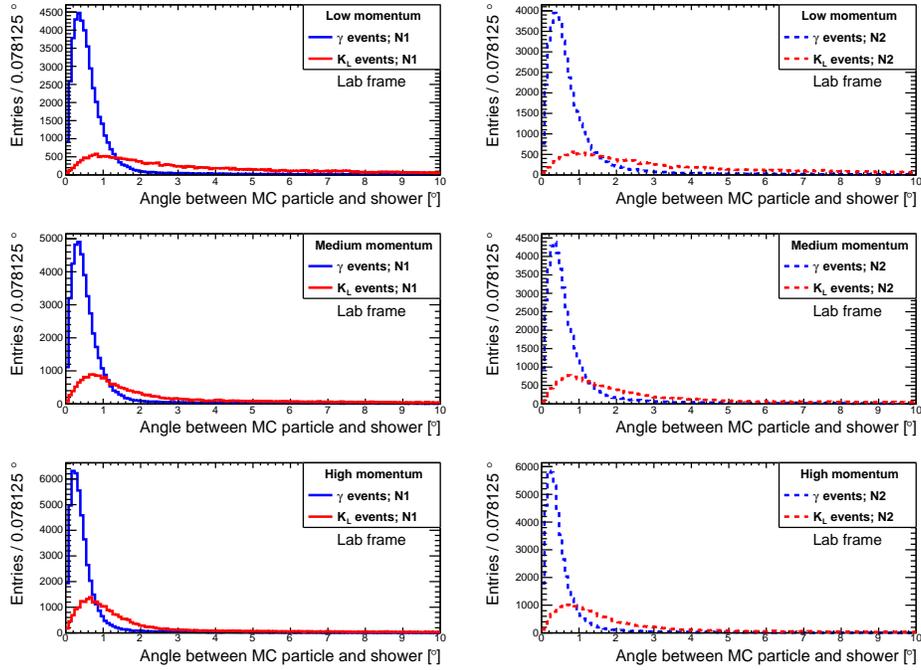


Figure 4.11: Angle between primary generated particle momentum and reconstructed shower momentum. Events chosen as described in the text. Figures on the left are of photon hypothesis showers (N1), and figures on the right are of neutral hadron hypothesis showers (N2). The top, center, and bottom rows are for the low, medium, and high momentum ranges, respectively.

- All single moments: $|Z_{11}|, |Z_{20}|, |Z_{22}|, \dots$
- All two moment combinations: $(|Z_{11}|, |Z_{20}|), (|Z_{11}|, |Z_{22}|), \dots, (|Z_{20}|, |Z_{22}|), (|Z_{20}|, |Z_{31}|), \dots$
- All eleven moments $(|Z_{11}|, |Z_{20}|, |Z_{22}|, |Z_{31}|, |Z_{33}|, |Z_{40}|, |Z_{42}|, |Z_{44}|, |Z_{51}|, |Z_{53}|, |Z_{55}|)$.

For each Zernike moment used, the input to the MVA includes the Zernike moment of the selected photon hypothesis shower and the Zernike moment of the selected neutral hadron hypothesis shower. In other words, the number of input variables to the MVA is twice the number of Zernike moments used. For example,

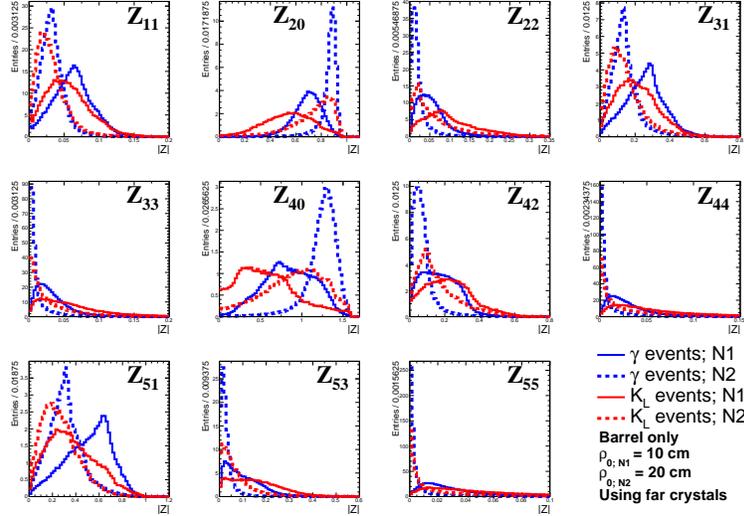


Figure 4.12: Distributions of the absolute value of the Zernike moments. For each particle type, the events of the three momentum ranges are combined. The plots are of events where the shower location is in the barrel region. The Zernike moments are calculated with $\rho_{0;N1} = 10\text{cm}$, $\rho_{0;N2} = 20\text{cm}$, and using the far crystals. Each histogram is normalized to have unit area.

when testing the combination of Zernike moments ($|Z_{40}|, |Z_{51}|$), four values are supplied as inputs to the MVA for each event

1. $|Z_{40}|$ of the photon hypothesis shower,
2. $|Z_{51}|$ of the photon hypothesis shower,
3. $|Z_{40}|$ of the neutral hadron hypothesis shower,
4. $|Z_{51}|$ of the neutral hadron hypothesis shower.

In addition to the different combinations of Zernike moments, different combinations of the parameters used in calculating the moments are evaluated. These parameters are discussed in Section 4.4

In order to choose which values of ρ_0 to evaluate, the distributions of r_{max} , the maximum radial distance of a crystal in the perpendicular plane of the selected

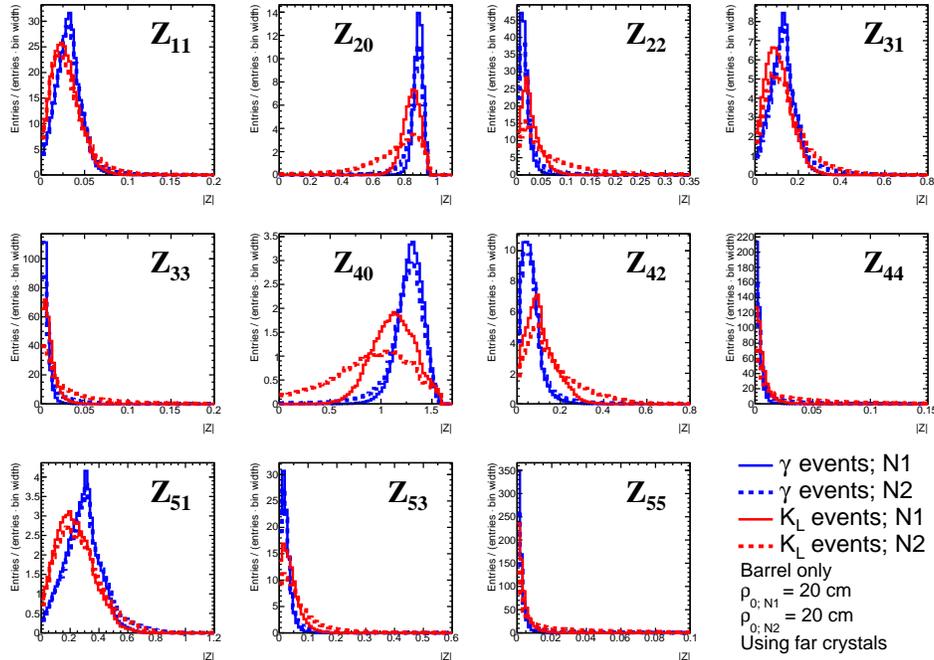


Figure 4.13: Distributions of the absolute value of the Zernike moments. This plot is the same as Figure 4.12, only that the Zernike moments are calculated with $\rho_{0;N1} = 20$ cm.

showers, are evaluated. The distributions are shown in Figure 4.17. Based on the results of Figure 4.17, five values of ρ_0 are selected, 10 cm, 15 cm, 20 cm, 30 cm, and 50 cm.

For each event, the Zernike moments of the selected showers are calculated using all possible combinations of the following three parameters

1. $\rho_{0;N1} \in \{10, 15, 20, 30, 50\}$ cm,
2. $\rho_{0;N2} \in \{10, 15, 20, 30, 50\}$ cm,
3. Using the far crystals or ignoring them.

A boosted decision tree (BDT) with the GradientBoost algorithm, BDTG, is used as the MVA method in the TMVA framework [62] using the default BDTG parameters. TMVA is chosen because it is an integral part of the ROOT framework

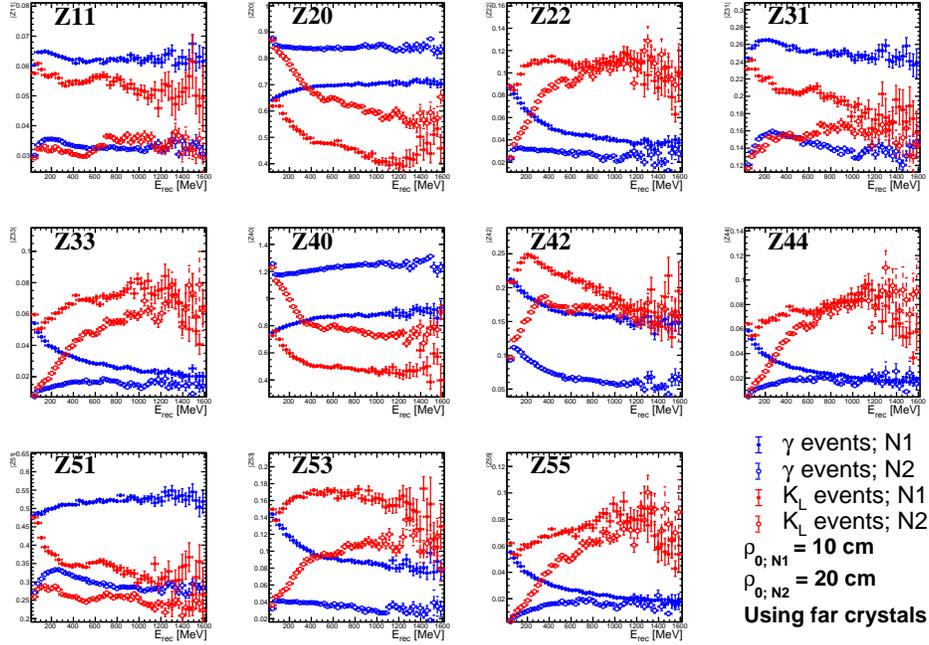


Figure 4.14: Mean absolute values of the Zernike moments as a function of the reconstructed shower energy. The Zernike moments are calculated with $\rho_{0;N1} = 10\text{ cm}$, $\rho_{0;N2} = 20\text{ cm}$, and using the far crystals.

[51] and provides easy access to different validation plots of the MVA training and testing.

The metric that is used for comparing the output of the different MVAs is the K_L^0 rejection at 90 % photon efficiency. The 90 % value is chosen arbitrarily. Figure 4.18 shows two receiver operating characteristic (ROC) curves for the outputs of two MVAs each using different Zernike moments. The intersection of the ROC curve with the blue vertical line representing the 90 % photon efficiency is used as the metric for the quality of the MVA. In the case depicted in the figure, the combination of ($|Z_{40}|, |Z_{51}|$) gives a better result than the combination of ($|Z_{31}|, |Z_{55}|$).

In order to check if the MC sample, described in Section 4.5.1, is too small, the over training plots produced by TMVA were examined. Over training happens when the training sample is too small and the MVA is tuned on the statistical fluctuations of the sample. Figure 4.19 shows examples of such plots for the MVA

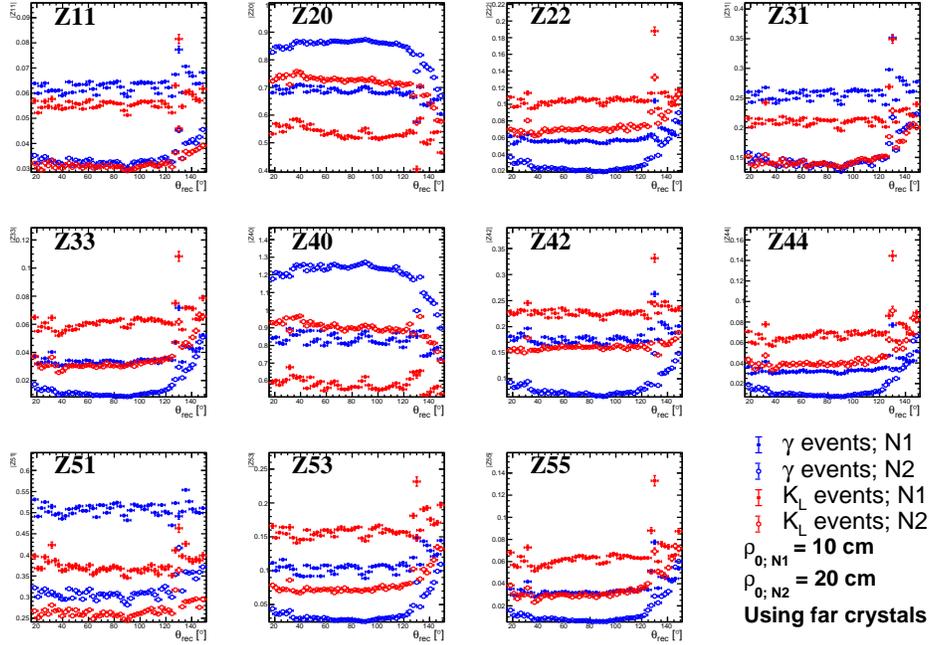


Figure 4.15: Mean absolute values of the Zernike moments as a function of the reconstructed shower polar angle. The Zernike moments are calculated with $\rho_{0;N1} = 10$ cm, $\rho_{0;N2} = 20$ cm, and using the far crystals.

using the Zernike moment pair ($|Z_{40}|, |Z_{51}|$) and the MVA using the eleven Zernike moments. Both MVAs use $\rho_{0;N1} = 10$ cm, $\rho_{0;N2} = 20$ cm, and the far crystals, and include only events where the showers are reconstructed in the barrel.

The Kolmogorov-Smirnov test is a statistical test designed to check if two samples originate from the same underlying distribution. Values close to zero imply that the samples are drawn from different distributions. Two out of the four Kolmogorov-Smirnov test results are close to zero. This suggests that their associated training and testing samples are from different parent distributions. This hints that the MVAs might be over trained on the sample used for the training. Nevertheless, the shapes of the distributions are similar and a larger sample was not produced.

All eight distributions have a peak around the BDT response value of -0.2 . A similar feature appeared in the BDT response discussed in Section 4.6.2. It is not

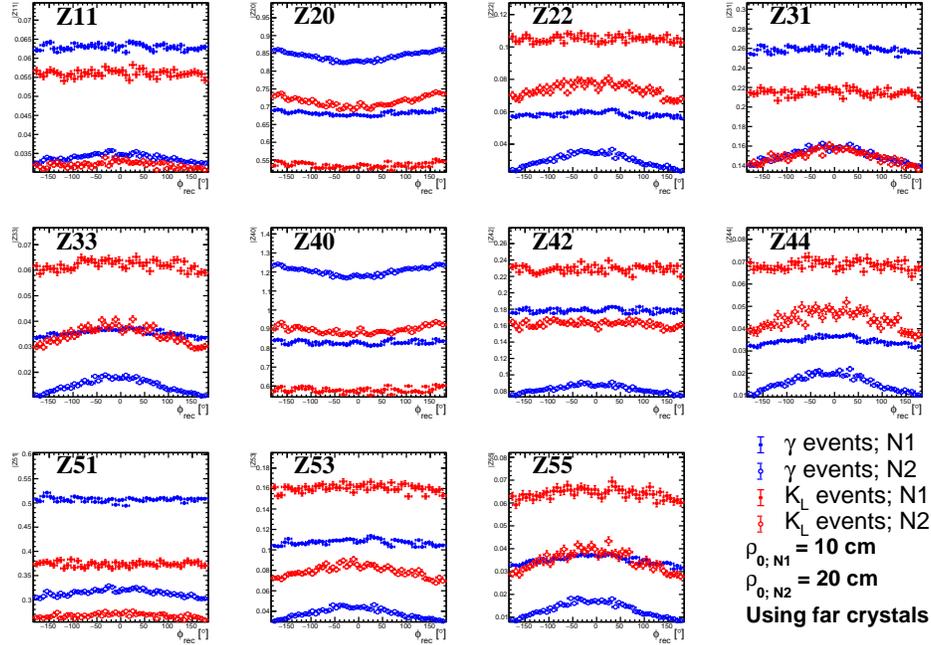


Figure 4.16: Mean absolute values of the Zernike moments as a function of the reconstructed shower azimuthal angle. The Zernike moments are calculated with $\rho_{0;N1} = 10\text{cm}$, $\rho_{0;N2} = 20\text{cm}$, and using the far crystals.

clear what causes this feature.

Figure 4.20 shows the K_L^0 rejection at 90% γ efficiency as a function of the number of Zernike moments used in the MVA for the MVAs that gave the best results using one, two, and eleven Zernike moments. Because the best results with the two and eleven Zernike moments are obtained using $\rho_{0;N1} = 10\text{cm}$ which is the smallest $\rho_{0;N1}$ value tested, values of $\rho_{0;N1} = 8\text{cm}$ and $\rho_{0;N1} = 12.5\text{cm}$ are also evaluated. They give inferior results.

As Z_{40} is the moment that gives the best results in the MVA with a single moment as input, and (Z_{40}, Z_{51}) is the pair of moments that gives the best results in the MVA with two moments as input, the raw values of $|Z_{40}|$ and $|Z_{51}|$ are the values that are selected to be kept for each shower. They are calculated with $\rho_{0;N1} = 10\text{cm}$ and $\rho_{0;N2} = 20\text{cm}$ and using the far crystals, as these parameters give the best results

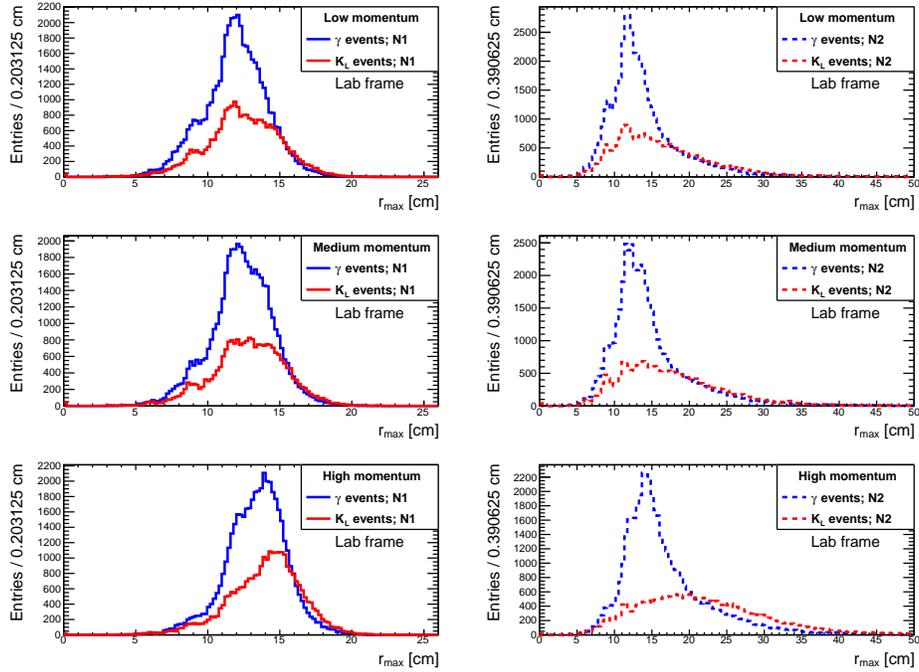


Figure 4.17: Largest radial distance of a crystal in the perpendicular plane, r_{\max} , for the selected showers. Figures on the left are of photon hypothesis (N1) showers, and figures on the right are of neutral hadron hypothesis showers (N2). The top, center, and bottom rows are for the low, medium, and high momentum ranges, respectively. Note the different x-axis scales.

for the MVA with two and eleven moments as input. In order to give a sense of the selected ρ values, Figure 4.1 has a red circle with a radius of 10 cm centered on the central crystal, and Figure 4.2 has a red circle with a radius of 20 cm. In `basf2`, ρ is measured from the reconstructed shower position.

4.6 MVAs using eleven Zernike moments for `basf2`

This section will describe the training of the MVAs using all eleven Zernike moments that are implemented in `basf2` version 00-08-00, which was released in January, 2018. Different MVAs are created for the forward endcap, barrel, and backward endcap. Each one is trained separately.

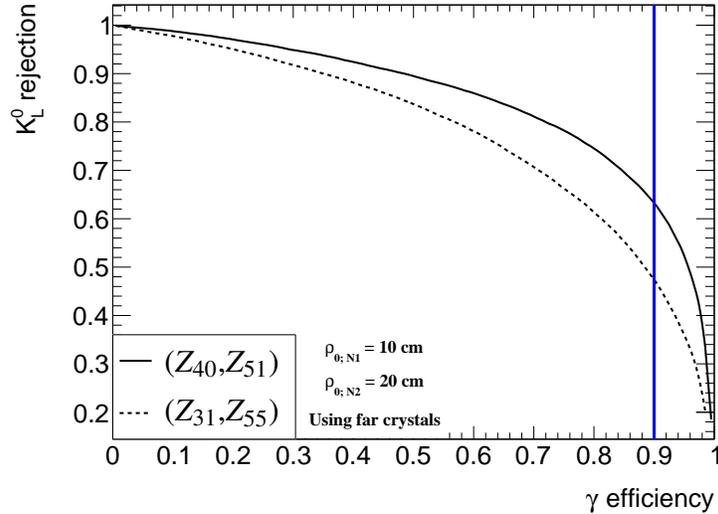
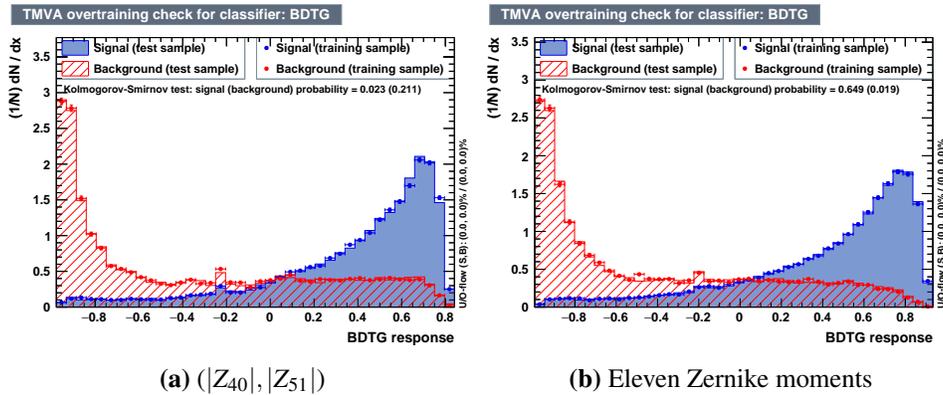


Figure 4.18: ROC curves for two MVAs. The first uses Zernike moments ($|Z_{40}|, |Z_{51}|$). The second uses ($|Z_{31}|, |Z_{55}|$). All Zernike moments are calculated with $\rho_{0;N1} = 10\text{cm}$, $\rho_{0;N2} = 20\text{cm}$, and using the far crystals. The intersection of the curves with the vertical blue line is used as the metric for comparing the quality of the MVAs.



(a) ($|Z_{40}|, |Z_{51}|$)

(b) Eleven Zernike moments

Figure 4.19: TMVA over training plots for (a) the MVA using Zernike moment pair ($|Z_{40}|, |Z_{51}|$) and (b) the MVA using eleven Zernike moments. Both MVAs use $\rho_{0;N1} = 10\text{cm}$, $\rho_{0;N2} = 20\text{cm}$, and the far crystals. Only showers reconstructed in the barrel region are considered.

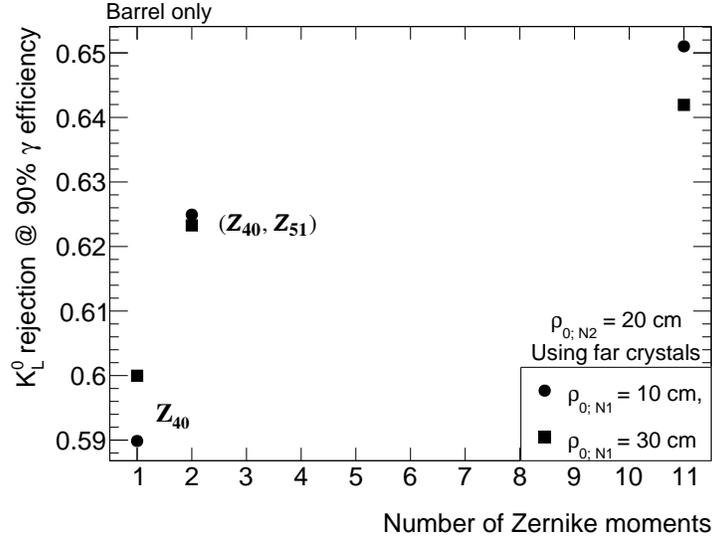


Figure 4.20: K_L^0 rejection at 90 % γ efficiency for the best MVAs using one, two, and eleven Zernike moments. Only events reconstructed in the barrel are used.

The input to the MVAs are the eleven Zernike moments of the photon hypothesis shower and the eleven Zernike moments of the neutral hadron hypothesis shower. The Zernike moments are calculated with the optimal parameters found in Section 4.5.4, $\rho_{0;N1} = 10\text{cm}$ and $\rho_{0;N2} = 20\text{cm}$ and using the far crystals. Although these parameters are found using showers reconstructed in the barrel region, they are used for the end caps as well.

The MVA method that is used in is `FastBDT` [63]. It is chosen for two reasons. First, it is the default MVA method in `basf2`. Second, the training is considerably faster than `BDTG` in `TMVA`. As the MVAs are used in production versions of `basf2`, a larger MC sample is generated in order to train the MVAs.

4.6.1 MC sample

The MC sample is generated using the same procedures as described in Section 4.5.1 and the same selection cuts as in Section 4.5.2 are used. Table 4.2 lists the number of generated events and the number of events that passed the selection cuts for each particle type and each ECL detector region.

Table 4.2: Number of generated events and number of events that passed selection cuts for the MC samples used for the training of the MVAs used in `basf2` release 00-08-00.

	Generated		Passed cuts	
	γ	K_L^0	γ	K_L^0
Forward endcap	895 000	1 695 000	772 785	757 852
Barrel	720 000	1 245 000	771 371	778 038
Backward endcap	735 000	1 860 000	602 348	758 957

4.6.2 Over training check

Figures 4.21, 4.22, and 4.23 show over training plots for the forward, barrel, and backward detector regions, respectively. All three MVAs exhibit a peak around a classifier output value of 0.2. This peak is similar to the one in Figure 4.19 even though the MC samples are completely different and much larger and a different MVA methods are used in each section. The origin of this feature is not investigated.

4.6.3 Results

Each `FastBDT` MVA for each of the three ECL detector regions is trained with the complete sample of events that passed the selection cuts. The ROC curves for the MVAs are presented in Figure 4.24. The results in the barrel are similar to the results obtained using the smaller sample and TMVA’s BDTG algorithm described in Section 4.5.

4.7 ECL shower shape module

The calculation of the Zernike moments is added to the `basf2` shower shape module. For each ECL shower, the module adds the values of $|Z_{40}|$ and $|Z_{51}|$ to the C++ data object representing an ECL shower. In addition, for a connected region containing at least one photon hypothesis shower and a neutral hadron hypothesis shower, the module adds a value called `zernikeMVA` to the showers. The value of `zernikeMVA` is based on the MVAs described in Section 4.6 and is calculated differently for each shower hypothesis. For a photon hypothesis shower `zernikeMVA`

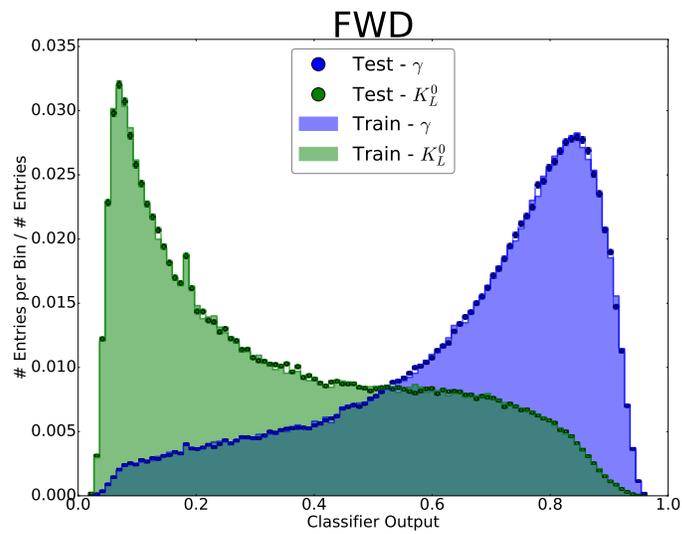


Figure 4.21: Over training plot of forward endcap MVA.

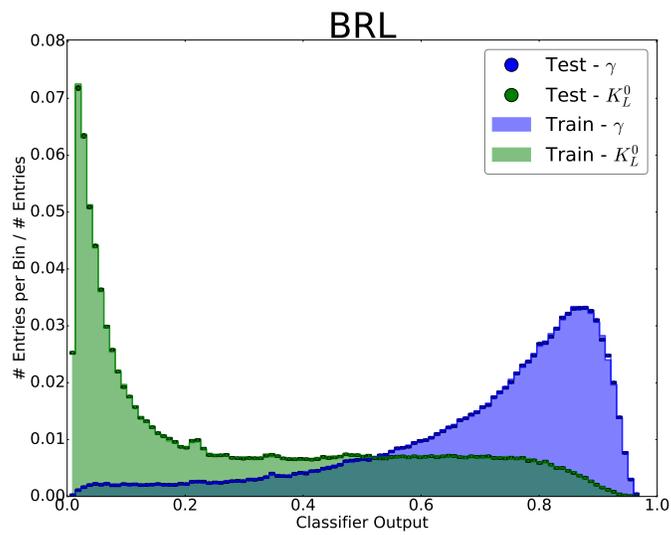


Figure 4.22: Over training plot of barrel MVA.

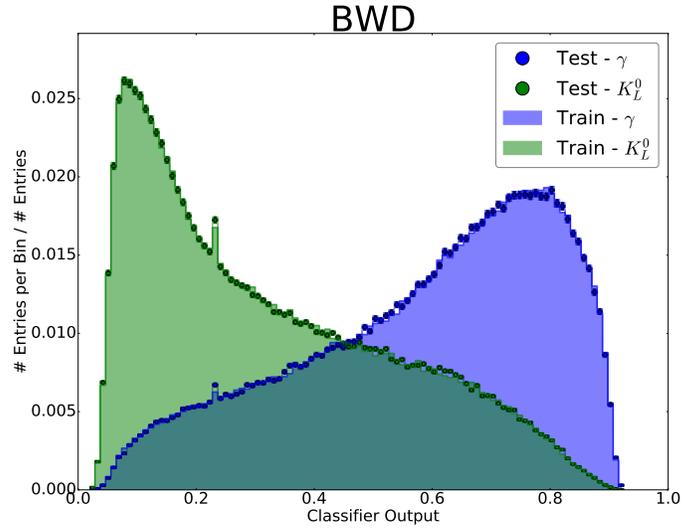


Figure 4.23: Over training plot of backward endcap MVA.

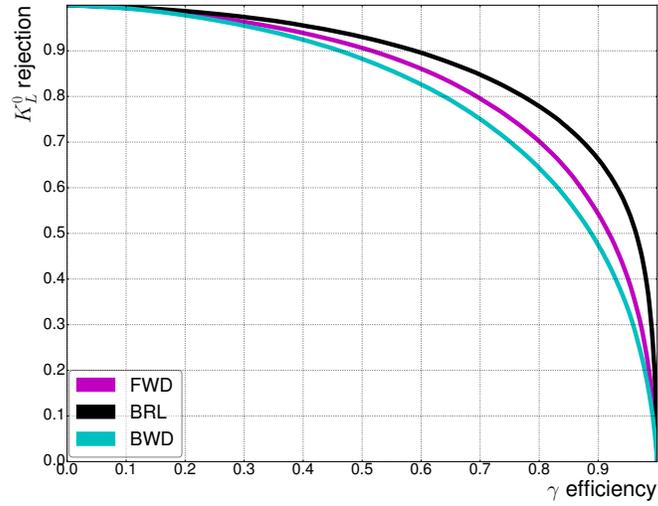


Figure 4.24: K_L^0 rejection as a function of γ efficiency for the three ECL detector regions. The input to the MVAs are the eleven Zernike moments of the photon hypothesis showers and the eleven Zernike moments of the neutral hadron hypothesis showers.

is set to the output of the MVA using the eleven Zernike moments of the photon hypothesis shower and the eleven Zernike moments of the neutral hadron hypothesis shower from the same connected region. For a neutral hadron hypothesis shower, `zernikeMVA` is set to the value of

$$1 - \prod_{i \in N1} (\text{zernikeMVA}_i), \quad (4.14)$$

where the product is over all the photon hypothesis showers in the same connected region as the neutral hadron hypothesis shower.

For a given connected region with one photon hypothesis shower and one neutral hadron hypothesis shower, the `zernikeMVA` variable acts as the likelihood that the photon hypothesis shower is created by a photon. This is because the `FastBDT` MVA output is between 0 (neutral hadron like) and 1 (photon like).

4.8 Comparison of `zernikeMVA` in data and MC

Comparison of the `zernikeMVA` in MC and data is performed in [4] with radiative muon pairs, $e^+e^- \rightarrow \mu^+\mu^-\gamma$. Events with exactly two tracks and one photon are selected. The tracks are required to have a momentum greater than 1 GeV/c, and to originate close to the IP. The photons need to have an energy greater than 0.1 GeV. Both the photons and the muons are required to have a polar angle in the range $37^\circ < \theta < 124^\circ$. This is the range of the ECL barrel without the 5° closest to the edges. The muon recoil momentum, p_{recoil} , is the difference between the pre-collision momentum and the sum of the muon momenta. The photons are required to be within 14° of p_{recoil} . No other photons with $E > 0.2$ GeV can be present in the event. All quantities are in the lab frame.

Figure 4.25 shows the `zernikeMVA` values in data and in MC for different ranges of p_{recoil} . The MC agrees well with the data for all vales of p_{recoil} .

4.9 Shower shape variables in $e^+e^- \rightarrow B\bar{B}$ events

In order to check if the three new Zernike moment variables are useful shower shape variables to identify photons, ROC curves for the different ECL shower shape variables in release 00-08-00 for $e^+e^- \rightarrow B\bar{B}$ events are produced. The plots are at

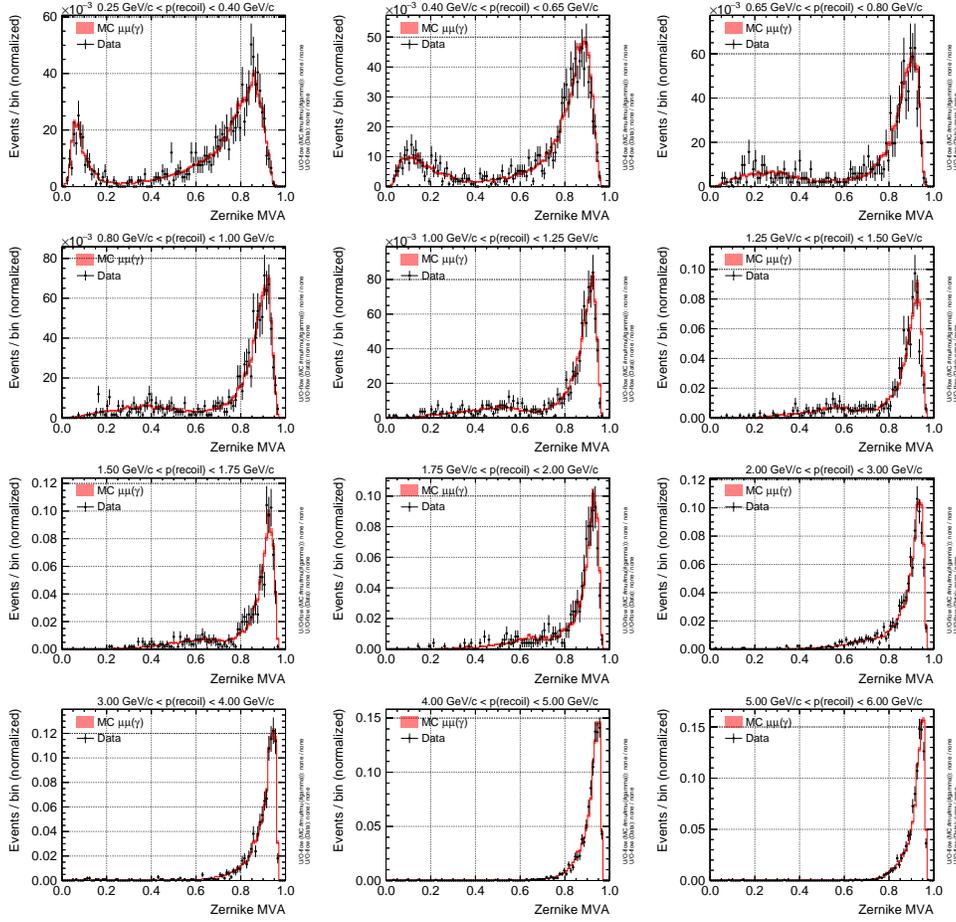


Figure 4.25: Comparison of zernikeMVA results in data and MC in radiative muon pair events, $e^+e^- \rightarrow \mu^+\mu^-\gamma$. The different plots are for different ranges of p_{recoil} , which is the difference between the pre-collision momentum and the sum of the muon momenta. The plot is from [4].

a few different energy ranges.

Signal in the ROC curves is defined as showers that are not associated with a charged track, contain at least 60 % of the energy of a photon originating from a $B\bar{B}$ pair, are in the polar angle range 17° to 150° in the lab frame, have a reconstructed energy significance between -4 and 3 , have a reconstructed polar and azimuthal angle significance between -5 and 5 , and have a reconstructed time within the timing range that includes 99 % of all true photons. The significance for a quantity in this context is the quantity over the uncertainty on the quantity. Background is defined as anything that is not signal.

Figures 4.26 to 4.29 show ROC curves for different energy ranges. At the lowest energy range, the Zernike moment variables are not the optimal variables to select photons. On the other hand, at the energy range of 0.1 GeV to 0.15 GeV, the zernikeMVA variable performs the best. At the highest energy range, the zernikeMVA gives the best results for signal efficiencies up to about 90 %.

Figure 4.30 shows the true photon energy spectrum of any photon with $E > 0.5$ MeV in the lab frame appearing in the 1.01×10^6 $\Upsilon(4S)$ MC sample discussed in Section 4.5.1. Most photons in these events are in the energy range of Figure 4.28 where the zernikeMVA variable is the best identifier of photons.

4.10 Usage of the zernikeMVA variable at Belle II

The zernikeMVA variable is used in at least three projects at Belle II. It is used to identify background photons which originate far away from the IP, when the beam hits the detector material and resulting photons travel back to interact with the ECL. In this case, the energy distribution in the crystals is not consistent with a photon originating at the IP and the zernikeMVA has some discriminating power to reject these events. The zernikeMVA is also used as a discriminating variable in the Belle II ALP search [1] to reject neutral hadrons and particles that do not originate at the IP. The zernikeMVA variable is also used in the Belle II trigger to select single photon event candidates.

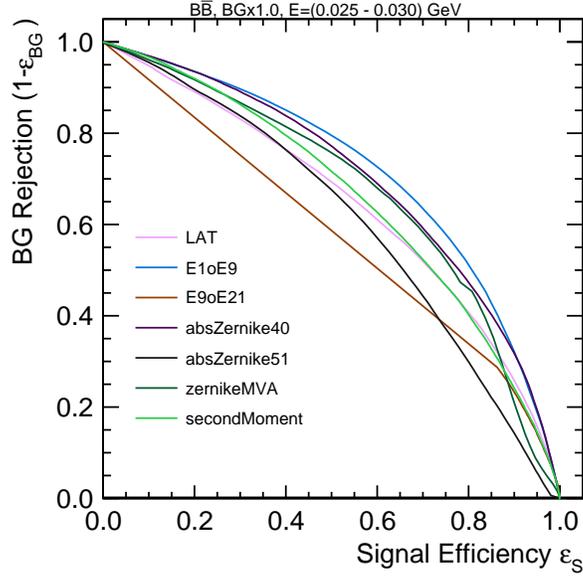


Figure 4.26: ROC curves for the different ECL shower shape variables in release 00-08-00 in $e^+e^- \rightarrow B\bar{B}$ events for the energy range 25 MeV to 30 MeV. LAT is the lateral moment described in [13], E1toE9 is the energy of the central crystal over the energy sum of 3×3 crystal matrix around the central crystal, E9toE21 is the energy sum of a 3×3 crystal matrix over the energy sum of a 5×5 crystal matrix without the corners, and second moment is defined as $\sum w_i E_i r_i^2 / \sum w_i E_i$, where E_i is the energy of the i th crystal in the shower, w_i is the weight of the i th crystal in the shower to account for crystals that belong to multiple showers, and r_i is the distance of the i th crystal to the shower center in the perpendicular plane.

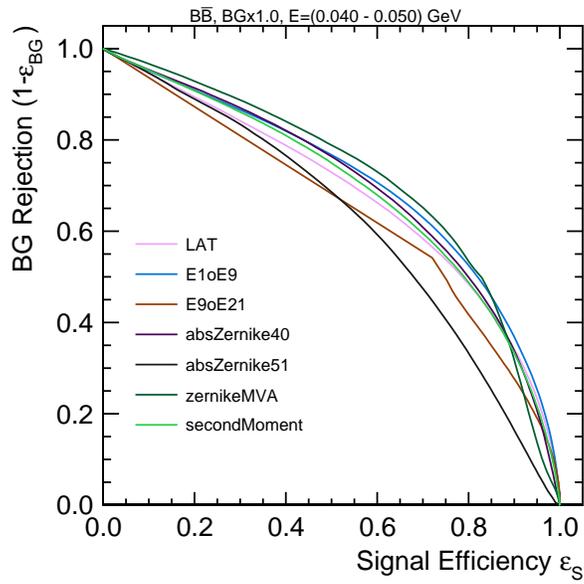


Figure 4.27: ROC curves for the different ECL shower shape variables in release 00-08-00 in $e^-e^+ \rightarrow B\bar{B}$ events for the energy range 40 MeV to 50 MeV.

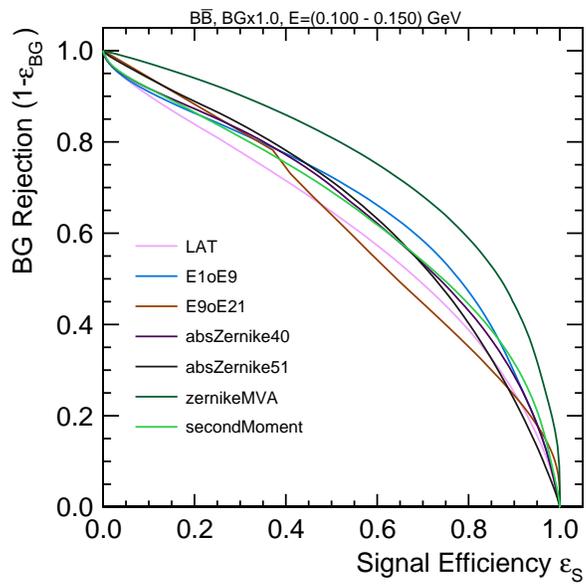


Figure 4.28: ROC curves for the different ECL shower shape variables in release 00-08-00 in $e^-e^+ \rightarrow B\bar{B}$ events for the energy range 100 MeV to 150 MeV.

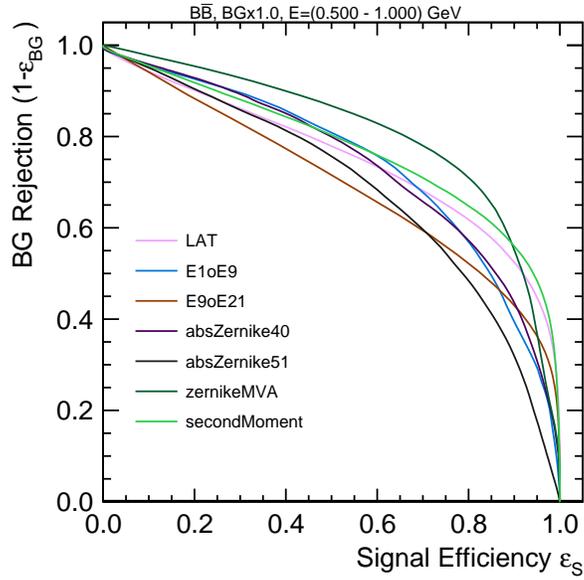


Figure 4.29: ROC curves for the different ECL shower shape variables in release 00-08-00 in $e^-e^+ \rightarrow B\bar{B}$ events for the energy range 500 MeV to 1000 MeV.

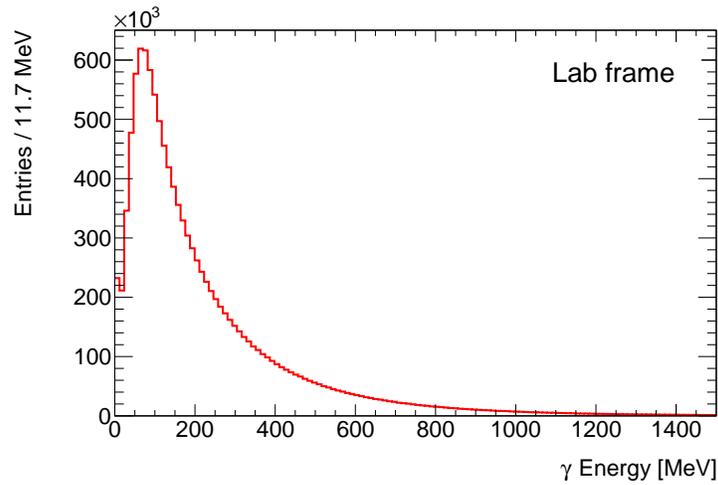


Figure 4.30: True photon energy of any photons in $e^+e^- \rightarrow Y(4S)$ events with $E > 0.5$ MeV in the lab frame.

Chapter 5

Conclusions

ALP search at *BABAR*

This chapter described a search for a new particle, an ALP that couples exclusively to photons in 5 % of the *BABAR* data. Depending on their mass, ALPs can influence a range physical phenomena, for example big bang nucleosynthesis and the cosmic microwave background. ALPs are dark matter candidates and they affect a number of astrophysical processes.

The goal of the chapter was to find evidence of an ALP or to set upper limits on the cross section and coupling constant of the ALP.

In the model considered here, an ALP event leaves a signature of three photons in the detector. The main background for this search are $e^+e^- \rightarrow \gamma\gamma\gamma$ events, which have an identical signature. Various selection variables and threshold values are optimized to maximize the sensitivity of the search.

The number of signal entries is extracted by a binned maximum likelihood fit to the di-photon invariant mass distribution. The signal shape component of the fit function is the signal MC di-photon invariant mass histogram. The background is modeled by Chebyshev polynomials.

The number of entries in data is only about 80 % that in the MC. This difference is incorporated into the signal efficiency and contributes to the systematic uncertainty on the signal efficiency, which is the main source of systematic uncertainty in this search. A more modern generator might produce better agreement between

data and MC, which will reduce the systematic uncertainty on the limits.

The results after analysing 5 % of the *BABAR* data are consistent with the data being composed only of background. Therefore, 90 % Bayesian CI upper limits on the cross section and coupling constant are set. Even with just 5 % of the *BABAR* data, this is the most stringent measurement in the mass range $0.29 \text{ GeV}/c^2$ to $5 \text{ GeV}/c^2$, and the limits exclude regions of the phase space that have not been explored before. The data sample used to optimize the selection cuts is the same sample used to perform the measurements and to set the upper limits. Therefore, the analysis is not blind and can be biased. The results should be interpreted in light of this.

More stringent limits might be set if the analysis is applied to the remaining 95 % of the *BABAR* data. This was not pursued here due to time constraints and the long collaboration review required in order to obtain permission to un-blind the rest of the data.

The effect of differences between the Belle II and *BABAR* analyses and detectors is explored. *BABAR* had a much lower trigger efficiency for the runs at the $Y(4S)$ CME. Scaling the Belle II limits to the effective integrated luminosity taking into account the lower trigger efficiency shows that the *BABAR* limits on the coupling constant are around one and a half times better for intermediate ALP masses. This shows that *BABAR* has superior performance per integrated luminosity. This conclusion should be re-evaluated when Belle II starts using kinematic fitting in their analyses.

Photon hadron separation using Zernike moments

Shower shape variables are useful for particle identification. Zernike moments are used in *BABAR* and *ZEUS* for this purpose. This chapter described a project to introduce shower shape variables into the Belle II software framework. It is one of the first projects to take advantage of the hypothesis based ECL reconstruction. The goal of this project was to study the use of Zernike moments to distinguish photon showers from K_L^0 showers. To choose the best moments, different MVAs were checked. Each one had a different combination of Zernike moments as input variables and they were calculated with different parameter values.

The best two Zernike moment combination was found to be $(|Z_{40}|, |Z_{51}|)$, calculated with $\rho_{0;N1} = 10\text{cm}$, $\rho_{0;N2} = 20\text{cm}$, and using far crystals. In `basf2` release 00-08-00, which was released in Jan 2018, the raw values of each of these moments are saved for each shower. In addition, `FastBDT` MVAs which use all the eleven Zernike moments considered in this project were trained for each of the three ECL detector regions. The output of these BDTs are used to fill the `zernikeMVA` variable, which is also saved for each shower. This might be the first time information from multiple Zernike moments is combined in order to improve particle identification results. Furthermore, it might also be the first time particle identification is performed with an MVA method based on information of Zernike moments.

ROC curves for the different ECL shower shape variables were produced for $e^+e^- \rightarrow B\bar{B}$ events for different energy ranges. The `zernikeMVA` variable was the best variable to identify photons in the energy range in which most photons are produced in $e^+e^- \rightarrow B\bar{B}$ events.

The `zernikeMVA` is used at Belle II to identify background photons which originate far away from the IP and travel back to interact in the ECL. This results in an energy distribution in the crystals that is not consistent with a photon originating at the IP. The `zernikeMVA` is also used as a discriminating variable in the Belle II ALP search, which was recently published. The variable is used in the Belle II trigger to select single photon event candidates.

The project described in this chapter was summarized in an internal Belle II note.

The process of converting the information about the energy deposited in the ECL crystals into a limited number of Zernike moments is lossy. That is, information is lost in this process. A different approach is to treat the ECL crystals as pixels of an image, where the color of each pixel is proportional to the energy deposited in the associated crystal. In such an approach Figures 4.1 and 4.2 can be interpreted as images composed of pixels. These images can be used to try to infer which particle interacted with the ECL with either image analysis tools, or by training an MVA method with the images themselves. The advantage of these methods is that they use the full information provided by the ECL, without compressing it into the Zernike moments.

The light produced in the ECL by a particle interacting with the crystal is con-

verted into an electronic signal by PIN diodes. Different particles interacting with the ECL produce slightly different electronic signal shapes. These differences provide a completely different approach to particle identification with the ECL [64].

Bibliography

- [1] F. Abudinén et al., Belle II Collaboration, *Search for Axionlike Particles Produced in e^+e^- Collisions at Belle II*, Phys. Rev. Lett. **125** (2020) 161806. → pages v, 6, 88, 90, 92, 94, 127
- [2] T. Ferber, C. Hearty, and M. Roney, *Design of the ECL Software for Belle II*, Belle II internal note BELLE2-NOTE-TE-2016-001 (2016) . → pages v, 96, 97, 98
- [3] V. Lakshminarayanan and A. Fleck, *Zernike polynomials: a guide*, Journal of Modern Optics **58** (2011) no. 7, 545–561. → pages v, 100, 101
- [4] M. D. Nuccio and T. Ferber, *Search for Axion-Like Particles produced in $e+e$ collisions at Belle II*, Belle II internal note BELLE2-NOTE-PH-2019-005, version 7.0 (2019) . → pages v, 125, 126
- [5] B. Aubert et al., BABAR Collaboration, *The BaBar detector*, Nucl. Instrum. Meth. A **479** (2002) 1–116. → pages v, 8, 10, 14, 15, 16
- [6] B. Aubert et al., BABAR Collaboration, *The BABAR Detector: Upgrades, Operation and Performance*, Nucl. Instrum. Meth. A **729** (2013) 615–701. → pages v, 8, 11, 12
- [7] G. Aad et al., ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1–29. → page 1
- [8] S. Chatrchyan et al., CMS Collaboration, *Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*, Phys. Lett. B **716** (2012) 30–61. → page 1
- [9] H. R. Quinn, *The Asymmetry between matter and antimatter*, Phys. Today **56N2** (2003) 30–35. → page 1

- [10] M. Tanabashi et al., Particle Data Group, *Review of Particle Physics*, Phys. Rev. D **98** (2018) 030001. → pages 2, 3, 4, 5, 22, 56, 58, 60
- [11] K. Garrett and G. Duda, *Dark Matter: A Primer*, Adv. Astron. **2011** (2011) 1–22. → page 2
- [12] T. Aushev et al., *Physics at Super B Factory*, KEK-REPORT-2009-12 (2010), arXiv:1002.5012 [hep-ex]. → page 2
- [13] A. Bevan et al., BABAR Belle Collaborations, *The Physics of the B Factories*, Eur. Phys. J. C **74** (2014) 3026. → pages 2, 96, 98, 128
- [14] K. Abe, Kazuo, et al., Belle Collaboration, *Observation of large CP violation in the neutral B meson system*, Phys. Rev. Lett. **87** (2001) 091802. → page 2
- [15] B. Aubert et al., BABAR Collaboration, *Observation of CP violation in the B⁰ meson system*, Phys. Rev. Lett. **87** (2001) 091801. → page 2
- [16] C. A. Baker et al., *An Improved experimental limit on the electric dipole moment of the neutron*, Phys. Rev. Lett. **97** (2006) 131801. → page 2
- [17] R. Peccei and H. R. Quinn, *CP Conservation in the Presence of Instantons*, Phys. Rev. Lett. **38** (1977) 1440–1443. → page 2
- [18] R. Peccei and H. R. Quinn, *Constraints Imposed by CP Conservation in the Presence of Instantons*, Phys. Rev. D **16** (1977) 1791–1797. → page 2
- [19] J. L. Hewett et al., *Fundamental physics at the intensity frontier. Report of the workshop held December 2011 in Rockville, MD.*, (2012). <https://www.osti.gov/biblio/1042577>. → page 3
- [20] J. Beacham et al., *Physics Beyond Colliders at CERN: Beyond the Standard Model Working Group Report*, J. Phys. G **47** (2020) no. 1, 010501. → page 3
- [21] M. Dolan et al., *Revised constraints and Belle II sensitivity for visible and invisible axion-like particles*, JHEP **12** (2017) 094, 1709.00009. → pages 4, 5, 6, 86
- [22] P. Arias et al., *WISPy Cold Dark Matter*, JCAP **06** (2012) 013. → page 4
- [23] T. Fujita, Y. Minami, K. Murai, and H. Nakatsuka, *Probing Axion-like Particles via CMB Polarization*, arXiv:2008.02473 [astro-ph.CO]. → page 4

- [24] M. Giannotti et al., *Cool WISPs for stellar cooling excesses*, JCAP **05** (2016) 057. → page 5
- [25] M. Meyer, D. Horns, and M. Raue, *First lower limits on the photon-axion-like particle coupling from very high energy gamma-ray observations*, Phys. Rev. D **87** (2013) 035027. → page 5
- [26] W. A. Terrano et al., *Constraints on axionlike dark matter with masses down to 10^{-23} eV/c²*, Phys. Rev. Lett. **122** (2019) no. 23, 231301. → page 5
- [27] M. Bauer et al., *Axion-Like Particles at Future Colliders*, Eur. Phys. J. C **79** (2019) no. 1, 74. → page 5
- [28] M. J. Dolan et al., *A taste of dark matter: Flavour constraints on pseudoscalar mediators*, JHEP **03** (2015) 171. [Erratum: JHEP 07, 103 (2015)]. → page 5
- [29] B. Döbrich, J. Jaeckel, and T. Spadaro, *Light in the beam dump - ALP production from decay photons in proton beam-dumps*, JHEP **05** (2019) 213. [Erratum: JHEP 10, 046 (2020)]. → pages 5, 90, 94
- [30] D. Banerjee et al., NA64 Collaboration, *Search for Axionlike and Scalar Particles with the NA64 Experiment*, Phys. Rev. Lett. **125** (2020) no. 8, 081801. → page 5
- [31] D. A. Dicus et al., *Astrophysical Bounds on the Masses of Axions and Higgs Particles*, Phys. Rev. D **18** (1978) 1829. → page 5
- [32] I. Larin et al., PrimEx, *A New Measurement of the π^0 Radiative Decay Width*, Phys. Rev. Lett. **106** (2011) 162303. → page 5
- [33] D. Aloni et al., *Photoproduction of Axionlike Particles*, Phys. Rev. Lett. **123** (2019) no. 7, 071801. → page 6
- [34] S. Knapen et al., *Searching for Axionlike Particles with Ultrapерipheral Heavy-Ion Collisions*, Phys. Rev. Lett. **118** (2017) no. 17, 171801. → page 6
- [35] A. M. Sirunyan et al., CMS Collaboration, *Evidence for light-by-light scattering and searches for axion-like particles in ultraperipheral PbPb collisions at $\sqrt{s_{NN}} = 5.02$ TeV*, Phys. Lett. B **797** (2019) 134826. → page 6
- [36] T. Abe et al., Belle II collaboration, *Belle II Technical Design Report*, arXiv:1011.0352 [physics.ins-det]. → pages 7, 95

- [37] D. Boutigny et al., *BABAR* Collaboration, *The BABAR physics book: Physics at an asymmetric B factory*, SLAC-R-0504 (1998) . → page 8
- [38] D. Boutigny et al., *BABAR* Collaboration, *BaBar technical design report*, SLAC-R-0457 (1995) . → page 8
- [39] *PEP-II: An Asymmetric B Factory. Conceptual Design Report. June 1993*, SLAC-418, LBL-PUB-5379, CALT-68-1869, UCRL-ID-114055, UCIIRPA-93-01, SLAC-0418, SLAC-R-0418, SLAC-R-418 (1994) . → page 9
- [40] J. Seeman, *Last Year of PEP-II B-Factory Operation*, Conf. Proc. C **0806233** (2008) TUXG01. → page 9
- [41] L. Merlo et al., *Revisiting the production of ALPs at B-factories*, JHEP **06** (2019) 091. → page 17
- [42] J. R. Klein and A. Roodman, *Blind analysis in nuclear and particle physics*, Ann. Rev. Nucl. Part. Sci. **55** (2005) 141–163. → page 19
- [43] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **7** (2014) 79. → page 20
- [44] Q. Li and Q.-S. Yan, *Initial State Radiation Simulation with MadGraph*, arXiv e-prints (2018), arXiv:1804.00125 [hep-ph]. → page 20
- [45] F. Berends and R. Kleiss, *Distributions for electron-positron annihilation into two and three photons*, Nucl. Phys. Bk **186** (1981) 22–34. → page 21
- [46] D. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. A **462** (2001) 152–155. → pages 22, 107
- [47] S. Jadach, W. Paczek, and B. Ward, *BHWIDE 1.00: $O(\alpha)$ YFS exponentiated Monte Carlo for Bhabha scattering at wide angles for LEP1/SLC and LEP2*, Phys. Lett. B **390** (1997) 298–308. → page 22
- [48] S. Agostinelli et al., *Geant4 a simulation toolkit*, Nucl. Instrum. Methods Phys. Res., Sect. A **506** (2003) 250 – 303. → page 22
- [49] M. Benayoun, S. I. Eidelman, V. N. Ivanchenko, and Z. Silagadze, *Spectroscopy at B-factories using hard photon emission*, Mod. Phys. Lett. A **14** (1999) 2605–2614. → pages 22, 58, 60

- [50] W. D. Hulsbergen, *Decay chain fitting with a Kalman filter*, Nucl. Instrum. Methods Phys. Res., Sect. A **552** (2005) no. 3, . → page 24
- [51] R. Brun and F. Rademakers, *ROOT: An object oriented data analysis framework*, Nucl. Instrum. Meth. A **389** (1997) 81–86. → pages 55, 116
- [52] B. Aubert et al., *BABAR Collaboration, Measurement of the η and η' transition form factors at $q^2 = 112 \text{ GeV}^2$* , Phys. Rev. D **74** (2006) 012002. → pages 57, 58, 60
- [53] L. Lyons, *Bayes and Frequentism: a particle physicists perspective*, Contemporary Physics **54** (2013) no. 1, 1–16. → page 81
- [54] K. Akai, K. Furukawa, and H. Koiso, SuperKEKB accelerator team, *SuperKEKB Collider*, Nucl. Instrum. Meth. A **907** (2018) 188–199. → page 95
- [55] R. Sinkus and T. Voss, *Particle identification with neural networks using a rotational invariant moment representation*, Nucl. Instrum. Meth. A **391** (1997) 360–368. → pages 96, 98
- [56] F. Zernike, *Beugungstheorie des schneidenver-fahrens und seiner verbesserten form, der phasenkontrastmethode*, Physica **1** (1934) no. 7, 689–704. → page 98
- [57] A. B. Bhatia and E. Wolf, *On the circle polynomials of Zernike and related orthogonal sets*, Proceedings of the Cambridge Philosophical Society **50** (1954) no. 1, 40–48. → pages 98, 99
- [58] A. Khotanzad and Y. H. Hong, *Invariant image recognition by Zernike moments*, IEEE Transactions on Pattern Analysis and Machine Intelligence **12** (1990) no. 5, 489–497. → pages 98, 100, 102
- [59] A. Hershenhorn, T. Ferber, and C. Hearty, *ECL shower shape variables based on Zernike moments*, Belle II internal note BELLE2-NOTE-TE-2017-001, version 1.0 (2017) . → page 99
- [60] M. Born et al., *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*. Cambridge University Press, 7 ed., 1999. → page 100
- [61] C. Teh and R. Chin, *On Image Analysis by the Methods of Moments*, IEEE Transactions on Pattern Analysis and Machine Intelligence **10** (1988) no. 04, 496–513. → page 106

- [62] A. Hocker et al., *TMVA - Toolkit for Multivariate Data Analysis*, arXiv:physics/0703039. → page 115
- [63] T. Keck, *FastBDT: A Speed-Optimized Multivariate Classification Algorithm for the Belle II Experiment*, *Comput. Softw. Big Sci.* **1** (2017) no. 1, 2. → page 121
- [64] S. Longo et al., *CsI(Tl) pulse shape discrimination with the Belle II electromagnetic calorimeter as a novel method to improve particle identification at electron–positron colliders*, *Nucl. Instrum. Meth. A* **982** (2020) 164562. → page 135