Statistical analysis of LDMX data

Author
Noah Ritter

Supervisors
Lene Kristian Bryngemark
Ruth Pöttgen

FYSK03 FOR 15HP
EXAMINED MAY 2023
A BACHELOR OF SCIENCE THESIS

LUND UNIVERSITY

DEPARTMENT OF PHYSICS
DIVISION OF PARTICLE AND NUCLEAR PHYSICS

Abstract

This thesis explores a statistical framework for analysing LDMX data. It uses pre-built functions in the ROOT framework to calculate likelihoods and confidence levels (CL). Data, background and signal were generated in Monte Carlo simulations with a loose veto on signal in the Hcal with simple systematic uncertainties (for the background and signal) created by varying this selection. The data sets consisted of a 4000 event background and 25 event signal (a strong signal). Four hypotheses were tested with dark photon masses of 1, 10, 100, and 1000 MeV. The pre-bult functions did not work as intended hence another method to find CL_s values needs to be found. Signs of signals were found in three of four hypotheses with the most promising being the 1000 MeV followed by the 100 MeV which where observed to a 95% CL, the 10 MeV was observed to a 68% CL. Lastly, the 1 MeV signal was not observed at a 68% CL.

Acknowledgements

I would like to thank my supervisor Lene for the continuous support and help they have provided me during my thesis work. Without your help I would not have gotten anywhere with this project. I would also like to thank Ruth (co-supervisor) for her help and her role in getting me interested in particle physics through the introductory particle course. In addition, I would also like to thank the particle and nuclear physics division for having an atmosphere where a student, such as I, can barge in several times a day for whatever stupid reason and feel welcome.

Further, I would like to thank my fantastic group of friends whose unyielding support through the project have been invaluable. So thank you Al, Jordis, Kim, Mel, Paul, and Sora – without you I would never have made it.

Lastly, I would like to thank my mother. She has been amazingly supportive, without her I would never have reached the my third year. Whenever I have needed help for whatever reason she has always listen and helped – thank you!

Contents

List of Acronyms

BPC - Bayesian Posterior Calculator
CL - Confidence Level
CL$_s$ - Signal Confidence Level
CL$_b$ - Background Confidence Level
CL$_{s+b}$ - Signal plus background Confidence Level
Ecal - Electromagnetic calorimeter
Hcal - Hadronic calorimeter
LDMX - Light Dark Matter eXperiment
LR - Likelihood Ratio
LLR - Log Likelihood Ratio
PLC - Profile Liklihood Calculator
pdf - Probability Density Function
SM - Standard Model
μ- Signal Strength

1 Introduction

Our current notion of the particle content of the Universe is incomplete. Only ∼5% of the energy in the Universe is composed of standard model (SM) matter; the rest is composed of hypothesised dark matter (∼ 27%) and dark energy (∼ 68%). The Light Dark Matter eXperiment (LDMX) is an experiment aiming to solve this mystery by finding a potential dark matter candidate. At the moment, the experiment is still in its design stage and there is work continuing on all components, physical and theoretical [1]. This report focuses on a statistical analysis of the expected data. After all, an understanding of how assumptions and performance estimations affect the overall sensitivity is vital for its ability to detect the elusive dark matter.

LDMX aims to discover light dark matter, i.e. dark matter particles in the mass range of MeV to a few GeV. The experiment itself consists of a 4-8 GeV (16 GeV in later iterations) beam of electrons hitting a thin sheet of tungsten with the products of the reaction being observed by a specialised tracker, electromagnetic calorimeter (Ecal), and hadronic calorimeter (Hcal). Still, the dark matter particle producing reaction is expected to be extremely rare, as it should otherwise already have been discovered. Hence the discrimination and elimination of the background is very important [1]. The dark matter is expected to be produced in a so-called dark bremsstrahlung (dark braking radiation) reaction which differs from a normal bremsstrahlung event as either a dark matter particle or a dark photon is radiated from a standard model (SM) particle. Given this, the experiment relies on the dark matter being unreactive whilst the standard model particle is detected by the detectors, thus creating the appearance of missing energy and momentum. Detection of the dark matter can be thought of in two steps: 1) When totalling the amount of energy deposited in the detector there should be less energy than expected if there was no dark matter reaction, this is a promising sign; 2) the unexpected absence of momentum (specifically in the transverse direction), i.e. an apparent violation of momentum conservation. To conserve momentum there must be a corresponding particle that travels undetected through the experiment, this particle should be a type of dark matter. The resulting spectrum (after background elimination) will show the existence as well as indicate the mass of the dark mediator [1].

The main issue with this setup is the vast amount of background produced from other processes in the tungsten target, the trackers and the calorimeters. Some backgrounds can be easily eliminated whilst others are more challenging, such as neutral kaon and neutrino-producing events. The neutrino events, for example, are an irreducible background. Fortunately, the rate of neutrino reactions is expected to be several orders of magnitudes rarer than the dark matter reactions meaning they can be neglected. Unfortunately, the other backgrounds listed are significant and are expected to have similar or higher frequencies than the dark matter reaction. This leads to complications, as the experiment aims for zero background. The kaon background (neutral and charged) is expected to be especially hard to discriminate. The neutral kaon does not interact in the tracker or Ecal, so the only place for detection is the Hcal. Moreover, as a neutral kaon can remain intact over the time it is in the detector it may not decay in said detector at all (though one can still detect it via other means). A solution to single elements being responsible for the rejection of large portions of the background has been to require very precise calorimeters as well as excellent tracking capabilities. Still, without a statistical analysis it is unclear how deviations from the specifications and predictions will impact the performance of the experiment. Furthermore, since the energy regime is not a thoroughly investigated region, the simulations that have been relied upon to construct the experiment have not been validated by extensive data [1]. Consequently, there is some uncertainty in the performance of LDMX. A statistical analysis would allow for quantitative estimate how much signal is needed in the presence of a background to still discover dark matter for differing mass hypotheses [1]. This thesis hopes to provide a statistical framework that can provide this quantitative estimate.

A statistical analysis of LDMX data will entail accounting for statistical as well as systematic uncertainties. Monte Carlo (MC) simulated data estimating the performance of LDMX as well as the background production will be used for testing several different possible cases and checking whether statistically significant results can be extracted (more on significance in 2.3). Without a statistical analysis, it is impossible to ascertain whether the results are significant enough to warrant discovery. In particle physics, a so-called 5 sigma result is required (one false positive in, roughly, 1.7 million results) to claim a discovery. A statistical analysis is necessary to back up the 5 sigma claim. Consequently, it will be of vital importance when data starts being analysed. Hence the development of a statistical framework from which LDMX's performance can be analysed is important so that it is known whether the current expected performance is acceptable.

2 Theory

2.1 LDMX

LDMX has four main components (see Figure 1), namely the tracker (tagging and recoil), the target, the Ecal, and the Hcal. The tagging tracker is 60 cm long and positioned inside a 1.5T magnetic field whilst the recoil tracker is 18 cm long and in the fringe of the magnet. The tagging tracker uses silicon microstrips modules placed every 10 cm to ensure good momentum resolution. The recoil tracker uses the same modules spaced irregularly to optimise pattern recognition for low momentum particles whilst keeping good momentum resolution over the entire range. Between the trackers the tungsten target is placed. The Ecal is required to have radiation hardness, fast readout, plus good energy and spatial resolution. Lastly, the Hcal is a scintillator sampling calorimeter which main function is to detect neutral hadrons as well as showers escaping the Ecal and minimum ionising particles [2].

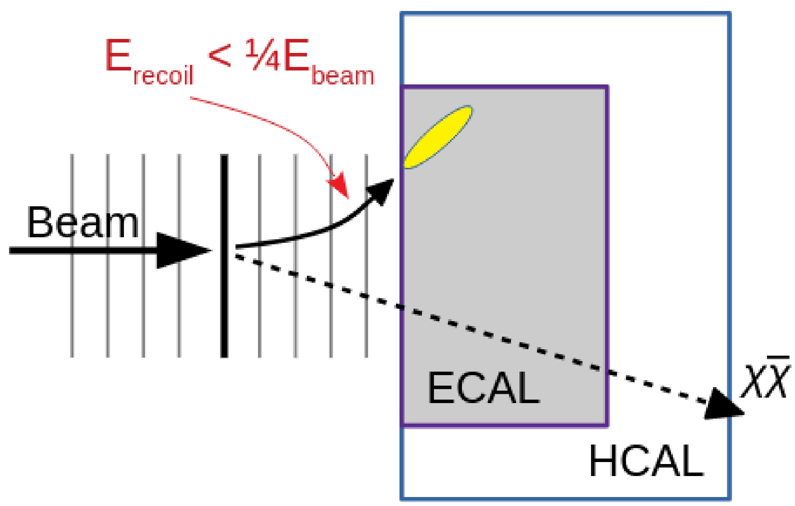

Figure 1: Illustration of the LDMX experiment. Shown is the electron beam incident from the left travelling trough the tagging tracker onto a thin tungsten target, then passing through the recoil tracker before reaching the Ecal and Hcal. [1]

2.2 Thermal Relic Dark Matter and Light Dark Matter

The thermal relic dark matter theory supposes that the dark matter arose in the Universe as a thermal relic of the early Universe, which is thought to have had a very hot and dense thermal period. The dark matter abundance is then theorised to have stabilised when a thermal freeze-out (expansion rate of the Universe overtook the rate of interaction) happened. This is an attractive model for the origin of dark matter as it is simple. It only requires that the rate of non-gravitational interaction between dark matter and standard model matter was larger than the Hubble expansion rate at some point in time. Moreover, the model is compatible with almost all models that allows for detection[3].

In the context of dark matter mass this gives a mass region spanning from MeV to several TeV. The upper range from GeV to TeV encompass the Weakly Interacting Massive Particles and has been the subject of most of dark matter particle research. The other region's dark matter candidate – light dark matter – has historically been difficult to investigate due the backgrounds involved. This is unfortunate as most stable ordinary particles have masses in this region which may be true for dark matter too. Light dark matter requires that there exist a light force carrier that can mediate annihilation for the thermal freeze-out. A minimal model, in this scheme only requires that the dark matter as well as its force carrier are singlets under the SM in a new U(1) gauge group[3].

2.3 Hypothesis testing

The following sections (section 2.3, 2.4, 2.5, and 2.6) follow the discussion in on statistics in the appendix of [4]. Hypothesis testing is a process of evaluating the validity of a hypothesis based on an experiment. The idea is that the prediction provided by the hypothesis is compared to the results of the experiment to see how unlikely the result is, assuming that the hypothesis is correct. If the result is sufficiently unlikely the hypothesis is rejected. A common way to quantify how well the results agree with the expected result is through p-values – p-values are a statement of how likely is it that an at least as extreme result is obtained. What can be considered extreme depends on the specifics of the experiment. Consider an experiment where testing the saying of dropping buttered toast lands buttered side down more commonly; in this case the hypothesis is that the number of drops which lands butter side down should be significantly higher than 50%. Agreement with the hypothesis would then be a high percentage of butter side down landings, whilst disagreement would be anything else. This, in terms of p-values, would be a *one-sided* probability with the p-value close to one for agreement with the hypothesis and p-values close to zero for disagreement. Consider the same experiment, but this time not acknowledging the old saying, we hypothesise that the toast should be equally likely to land on either side. In this case agreement is only found for roughly equal numbers of butter side up and down landings whilst a large divergence from this in either direction would result in disagreement. This is a *two sided* probability as both large and small count of buttered side down landings means poor agreement with the hypothesis [4].

The statement "If the result is sufficiently unlikely the hypothesis is rejected" is unacceptably vague to be used in any practical instance. One needs to define exactly what constitutes as sufficiently unlikely before an experiment is performed as well as defining what type of p-value test is to be used. The p-value is connected to the integral of an α-distribution, thereby the number of standard deviations.

$$Z = \begin{cases} \sqrt{2}\mathrm{erf}^{-1}(1-2p) & \text{if p is one sided} \\ \sqrt{2}\mathrm{erf}^{-1}(1-p) & \text{if p is two sided,} \end{cases} \tag{1}$$

where Z is the number of standard deviation and p is the p-value [4].

A common threshold for something to be deemed sufficiently unlikely is that it has to have a p-value lower than 0.05 (roughly 2 sigma for a two sided test and 1.65 sigma for a one sided test) with regards to the hypothesis. An alternative way of phrasing it is to say that a 95% *confidence level* (CL) is required [4]. Particle physics demands a 5 σ deviation from the SM (p-value of $6 \cdot 10^{-7}$ for a two sided test and $3 \cdot 10^{-7}$ for a one sided) to claim a discovery.

When conducting an experiment a null hypothesis is considered; the SM in particle physics. This leads to four possible cases: the null hypothesis is true and is not rejected; the null hypothesis is true but is rejected called a type I error; the null hypothesis is false and is rejected; and the null hypothesis is false but is not rejected called a type II error (see Fig. 2).

	Null Hypothesis is true	Alternate Hypothesis is true
Null Hypothesis is not rejected	Right decision	Type II Error
Null Hypothesis is rejected	Type I Error	Right decision

Figure 2: Illustration of the different possible error that can be committed whilst performing an hypothesis testing. Note that CL are usually concerned with the chance of committing a Type I error.

The confidence levels described above (in 2.3) correspond to a Type I error with respect to the SM (if the SM is the null hypothesis), i.e the chance of incorrectly retaining the null hypothesis. The confidence level is a measure of how many false negatives are accepted. The other error that can be committed is the Type II error, which describes the risk of incorrectly not rejecting the null hypothesis.

In general, when conducting an experiment using hypothesis testing two hypotheses are tested at the same time. Consequently, an experiment can be designed to optimise the testing of the alternate hypothesis as well as the null hypothesis. That is the same pre-experiment analysis can be done for the alternate hypothesis to define the region where the alternate hypothesis is best to be tested. Still the confidence level is defined with respect to the null hypothesis and judgement of the alternate hypothesis is conditional with respect to it.

In the interest of understanding hypothesis testing as a method of analysis, consider how it is applied. A confidence level is set and the experiment designed around this value, as this CL value tends to be rather harsh, in excess a 50% likelihood, this leads to experiments being optimised for minimising Type I errors and not Type II errors. In other words this is an inherently conservative method; an established theory is more likely to remain than an unestablished one. This should be kept in mind when using hypothesis testing.

2.4 Test Statistics

When conducting an experiment an outcome is obtained, for the purpose of this thesis this is a histogram with binned events. To interpret this a test statistic is used, in this thesis this is a probability density function (pdf). This is a function that helps distinguishing the null and alternate hypotheses from each other and allows for calculation of a p-value. From this the hypotheses can be either rejected or retained. There are two different kinds of test statistics: a *simple* test statistic which has no free parameters and a *composite* test statistic which can have free parameters. Of note is that from a *composite* test statistic confidence intervals of the different parameters can be obtained. As the composite version is the one used in this thesis it is the one which is focused on.

To obtain a test statistic the first step is to define the likelihood function, $\mathcal{L}$. This is a function that gives the probability of a certain outcome given the parameters of the model. Consider a model where the number of signal and background events are s and b respectively. Then the total number of observed events, λ, is the following:

$$\lambda(\mu) = s\mu + b, \tag{2}$$

where μ is the signal strength. $\mu = 0$ thus corresponds to a background only hypothesis whilst $\mu = 1$ to the signal hypothesis. Keep in mind that μ is just a scaling factor and that $\mu = 1$ (as a result) means that the assumptions made regarding the cross-sections for interactions were very good. In case of LDMX the signal strength is related to the cross section of dark bremstrahlung, the probability of the electron interacting with the target and causing dark bremsstrahlung. As the events are independent of each other and fluctuating they will conform to a Poisson distribution around an expectation value. In this case, the likelihood function becomes:

$$\mathcal{L}(n|\mu) = \mathcal{P}(n|\lambda(\mu)) = \frac{e^{\lambda}\lambda^{n}}{n!}, \tag{3}$$

where n is the number of events observed and $\mathcal{P}$ is a Poisson distribution. This function can be used in two ways. Firstly, it can be used to find how likely an outcome is given a set of parameter values (μ is the only parameter in eq. 3, but the number will be expanded when we introduce *nuisance parameters*). Secondly, it can be used to find the most likely values for the parameters, i.e. maximising $\mathcal{L}$ with respect to the data. Still, this is a simplistic model assuming only one background. In a more realistic situation there will be different backgrounds thus introducing 'free parameters'. The *parameter of interest*, the parameter which we want, is μ. The rest of the parameters, whose exact values are uninteresting are collectively called *nuisance parameters*. Do note that the only difference between the *parameter of interest* and the *nuisance parameters* is what we want to measure, thus the grouping is purview of the aims of the experiment. Furthermore, the nuisance parameters are not free, they are defined to maximise the likelihood for a given μ. The *nuisance parameters*, $\boldsymbol{b}$, are introduced to account for the different backgrounds changing the likelihood function to the following:

$$\mathcal{L}(n|\mu) \longrightarrow \mathcal{L}(n|\mu,\boldsymbol{b}) = \mathcal{P}(n|\lambda(\mu,\boldsymbol{b})). \tag{4}$$

Information on $\boldsymbol{b}$ can be obtained via usage of the control region(s) – regions without any signal. Maximising the likelihood with respect to the data will then give the most likely values for the nuisance paramenters. The last piece of the likelihood function is a modelling of the systematic uncertainties of the experiment. Through normalisation this can be done as a set of Gaussians with mean of zero and a standard deviation of one (denoted $\mathcal{G}(0|\theta,1)$), thus introducing another set of nuisance parameters $\boldsymbol{\theta}$. This leads to a new, more realistic likelihood function that accounts for more than just the randomness of the parameter of interest.

$$\mathcal{L}(\boldsymbol{n}|\mu,\boldsymbol{b},\boldsymbol{\theta}) = \mathcal{P}(n_s|\lambda_s(\mu,\boldsymbol{b},\boldsymbol{\theta})) \prod_i \mathcal{P}(n_i|\lambda_i(\mu,\boldsymbol{b},\boldsymbol{\theta})) \prod_j \mathcal{G}(0|\theta_j,1), \tag{5}$$

where $\boldsymbol{n}$ are the observed events in the signal and control regions, i denotes the set of control regions and j the set of systematic introduced uncertainties. Maximising the likelihood will give the most likely value of the parameter of interest (μ), as well as the optimal estimators from the nuisance parameters, notated as $\overline{\mu}, \overline{\boldsymbol{b}}$ and $\overline{\boldsymbol{\theta}}$. Now consider that there are several acceptable values given by the CL value – let $\hat{\boldsymbol{b}}$ and $\hat{\boldsymbol{\theta}}$ be the values that maximise the likelihood for a given $\hat{\mu}$ with $\hat{\mu} \neq \overline{\mu}$. From this a likelihood ratio, **LR**, can be made which is a type of test statistic:

$$\mathbf{LR} = \frac{\mathcal{L}(\boldsymbol{n}|\hat{\mu},\hat{\boldsymbol{b}},\hat{\boldsymbol{\theta}})}{\mathcal{L}(\boldsymbol{n}|\overline{\mu},\overline{\boldsymbol{b}},\overline{\boldsymbol{\theta}})}, \tag{6}$$

This ratio is commonly modified by taking the natural log of it and multiplying it by negative two to make it more suitable for computers to handle. This transformation makes no change in the information contained in the expressions. Computation time usually benefits from this as this transforms a product into a sum and changes the maximisation to a minimisation is also beneficial as minimisation algorithms are usually more efficient than maximisation algorithms. This gives the log-likelihood ratio, **LLR**.

$$\mathbf{LLR} = -2\log\frac{\mathcal{L}(\boldsymbol{n}|\hat{\mu},\hat{\boldsymbol{b}},\hat{\boldsymbol{\theta}})}{\mathcal{L}(\boldsymbol{n}|\overline{\mu},\overline{\boldsymbol{b}},\overline{\boldsymbol{\theta}})}. \tag{7}$$

Note that in the case of LDMX a positive result with respect to the alternate hypothesis only adds events. This means that the **LLR** expression by itself does not adequately describe the situation. Consequently, the test statistic q is defined bellow:

$$q = \begin{cases} -2\log\frac{\mathcal{L}(\boldsymbol{n}|\hat{\mu},\hat{\boldsymbol{b}},\hat{\boldsymbol{\theta}})}{\mathcal{L}(\boldsymbol{n}|0,\overline{\boldsymbol{b}},\overline{\boldsymbol{\theta}})}, & \overline{\mu} < 0 \\ -2\log\frac{\mathcal{L}(\boldsymbol{n}|\hat{\mu},\hat{\boldsymbol{b}},\hat{\boldsymbol{\theta}})}{\mathcal{L}(\boldsymbol{n}|\overline{\mu},\overline{\boldsymbol{b}},\overline{\boldsymbol{\theta}})}, & \overline{\mu} \geq 0 \end{cases}, \tag{8}$$

replacing the $\overline{\mu}$ with zero in the case of a negative signal ensures that the test statistics will never reject the standard model if there is a deficit in events. In regards to LDMX, the chances of a deficit of events causing an outlier is unlikely if the zero background or something close to it is achieved. A one sided test statistics can be created by demanding that $q = 0$ for either $\mu < \overline{\mu}$ or $\mu > \overline{\mu}$, the former being for the upper interval and the latter being for the lower interval.

2.5 Limits

Consider that an experiment has been conducted with one parameter of interest, μ, and one nuisance parameter, θ. Maximisation of the likelihood ratio with respect to the data gives the values of $\overline{\mu}$ and $\overline{\theta}$, but this does not mean that all other values can be neglected out of hand. Instead we say that they are excluded at a certain confidence level, i.e. they are disfavoured by the data. Indeed, at some point the likelihood becomes so small that it is excluded, but the type of test statistic defined decides what types of limits can be set. Consider the test statistic q, defined in 2.4, due to it being non-zero for both $\mu > \overline{\mu}$ and $\mu < \overline{\mu}$ it can be used to set both an upper and lower limit on the parameter of interest. This is in contrast to the one sided test statistics which can only set one sided limits. To set specific limits one relies on p-values which are defined as the integral of a pdf. The p-value defined by using pdfs is:

$$p = \int_{q^{point}}^{\infty} f(q|\mu)\mathrm{d}q, \tag{9}$$

where p is the p-value, f is the pdf given μ, and q^{point} is the value of q in the point of interest. By requiring a certain p-value, it is possible to find the allowed intervals for an observable given a dataset. In this way, if $p > 0.05$ the calculations will give an interval from which all values not included are excluded at a 95% CL.

To find the pdf there are two primary methods: the first is to utilise Monte Carlo simulations under the signal hypothesis; the second is to use the Asimov approximation. The Monte Carlo approach is perfect in the limit of infinite experiments, but has the drawback of being computationally intensive. The Asimov approximation on the other hand is much quicker than using Monte Carlo simulations and is accurate as long as one only requires a 95% CL.

2.6 $\mathrm{CL_s}$ method

The $\mathrm{CL_s}$ method is an extension to the CL formalism, its basic premise being that if the difference between the theoretical prediction and the data in the control region is significant, the confidence level of the result in the signal region should be punished. Consider a significant downwards fluctuation. Depending on its relative strength to the signal strength, it will allow the experiment to impose arbitrarily strong limits on the parameter of interest. This is an undesirable property of the CL method, as outliers are guaranteed to occur eventually. This issue must be addressed as not doing so means that statistical outliers could work as evidence against new physics. Consequently the $\mathrm{CL_s}$ method has been developed to ensure that models are harder to exclude if not significantly different from the null hypothesis.

To perform the $\mathrm{CL_s}$ method we first need to obtain the pdfs of the signal + background hypothesis as well as the background hypothesis. This, as mentioned in section 2.5, can be done using a large amount of Monte Carlo simulations for $\mu = 0$ and $\mu = 1$. From this, the p-value for the signal + background hypothesis called $\mathrm{CL_{s+b}}$ can be calculated.

$$\mathrm{CL_{s+b}} = \int_{q^{point}}^{\infty} f(q|\mu)\mathrm{d}q. \tag{10}$$

The p-value of the background only hypothesis can similarly be defined as 1-$\mathrm{CL_b}$.

$$1 - \mathrm{CL_b} = \int_{q^{point}}^{\infty} f(q|0)\mathrm{d}q. \tag{11}$$

The idea of $\mathrm{CL_s}$ is to ensure that the model is not discarded if the result is also extreme with regards to the background hypothesis. Thereby, $\mathrm{CL_{s+b}}$ is normalised to the 1-$\mathrm{CL_b}$ value and we define the resulting value as the $\mathrm{CL_s}$ value:

$$\mathrm{CL_s} = \frac{\mathrm{CL_{s+b}}}{1 - \mathrm{CL_b}}. \tag{12}$$

Whilst this value is used in the same way as a p-value to determine significance there are several things to note about this quantity that are different than the standard p-value. Firstly, a CL_s value is not a p-value – it is a ratio of p-values, hence its range is not limited to (0,1] but all positive numbers. Consequently, it is much harder to obtain small CL_s, it will only happen when the signal hypothesis is much less likely than the background hypothesis. Consider an experiment where there is a large downward fluctuation in the background so that it cancels the signal. Using a standard p-value the test can reject the signal hypothesis whilst a CL_s test would, in essence, say: the background is unexpected hence the signal measurement can not be trusted and the signal hypothesis can not be rejected. The trade off for this extra caution is that the limits set via CL_s are weaker than those of p-values. Lastly it is important to remember that CL_s can not reject the background hypothesis as its validity is implicitly assumed.

3 Method

Firstly, the pdfs of the null and alternate hypotheses were created – as described in section 2.5 – via Monte Carlo simulation using ldmx-sw (see [5] for GitHub). This was done for both signal and background (see Figure 3 & 4), with dark photon masses assumed to be 1, 10, 100, and 1000 MeV respectively to create four different signal hypotheses. The background comes from photonuclear reactions taking place in the Ecal which is expected to be one of the most challenging background forms in the final analysis. When taking data the observable of interest is p_T, the transverse momentum of the electrons from generator level. In the experiment the momentum is to be obtained from LDMX's tracker, but this is not implemented in the simulation. Hence this information is obtained from the outgoing electron four-momentum. LDMX will employ a sequence of vetoes to diminish the background as much as possible or, eliminate it, but for the purposes of this investigation much loser requirement where introduced: there can not be more than 200 photoelectrons[1] detected in the event in a bar in the Hcal. This construction allowed the creation of toy systematic uncertainties by varying the cut up and down by the arbitrary choice of 50 photoelectrons. The signal hypotheses where created with a signal strength of one corresponding to 2 signal events [6].

The processing of the simulated data consists of a series of several automated code analyses (see [7] for codes used). The code is written in C++ in an extensive package called ROOT [8]. More specifically, RooStats and HistFactory are the main sub packages used, the specific functions and methods used will be described more in detail when discussing their use. In broad strokes the analysis is done via four different codes (see [7] for codes):

[1] this is expected to be of O(10) in the real analysis

the first creates ROOT workspace and Model configuration; the second a profiled likelihood curve of the parameter of interest; the third a posterior of the parameter of interest; the fourth code calculated the CL_s value as well as plotting the differing hypotheses prediction's likelihoods to visually show if a hypothesis conforms to the data (see [9] for the original codes).

After the pdfs of the hypotheses and the histogram of the data were created systematic uncertainties were included via error propagation (systematic and statistical errors where added as squares under the square-root) instead of via a Gauss distributions described in section 2.4. Luminosity error, Lumi error, was also included in the model as it is required for the analysis functions. This was implemented with as little error as possible (1 millionth) as the data was created assuming no uncertainty in luminosity. Moreover, as the functions used for the analysis struggled or crashed when dealing with empty bins with no error it was ensured that the histogram's bins had a minimum value of one billionth with a minimum error of one billionth (See Figure 3). These modified pdfs where then organised into a model configuration which is a way to store and set data up for the functions and methods used in the analyses. This configuration was then folded into a ROOT workspace (an environment that contains files and variables global to the workspace) using RooStats::HistFactory::MakeModelAndMeasurementFast to be imported into the other files.

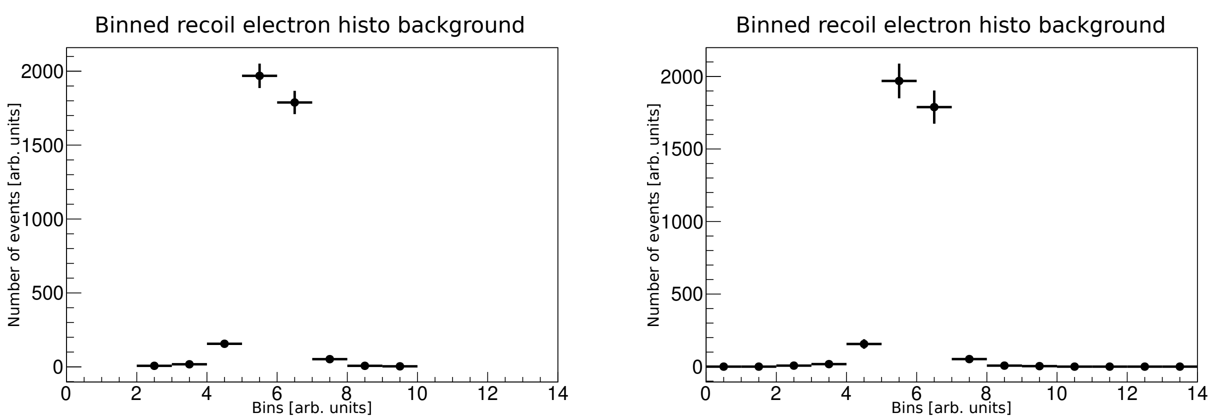

Figure 3: Histograms of background generated without a signal strength. Unmodified on the left and modified, i.e. no completely empty bins, on the right. X-axis is unevenly binned p_T (transverse momentum).

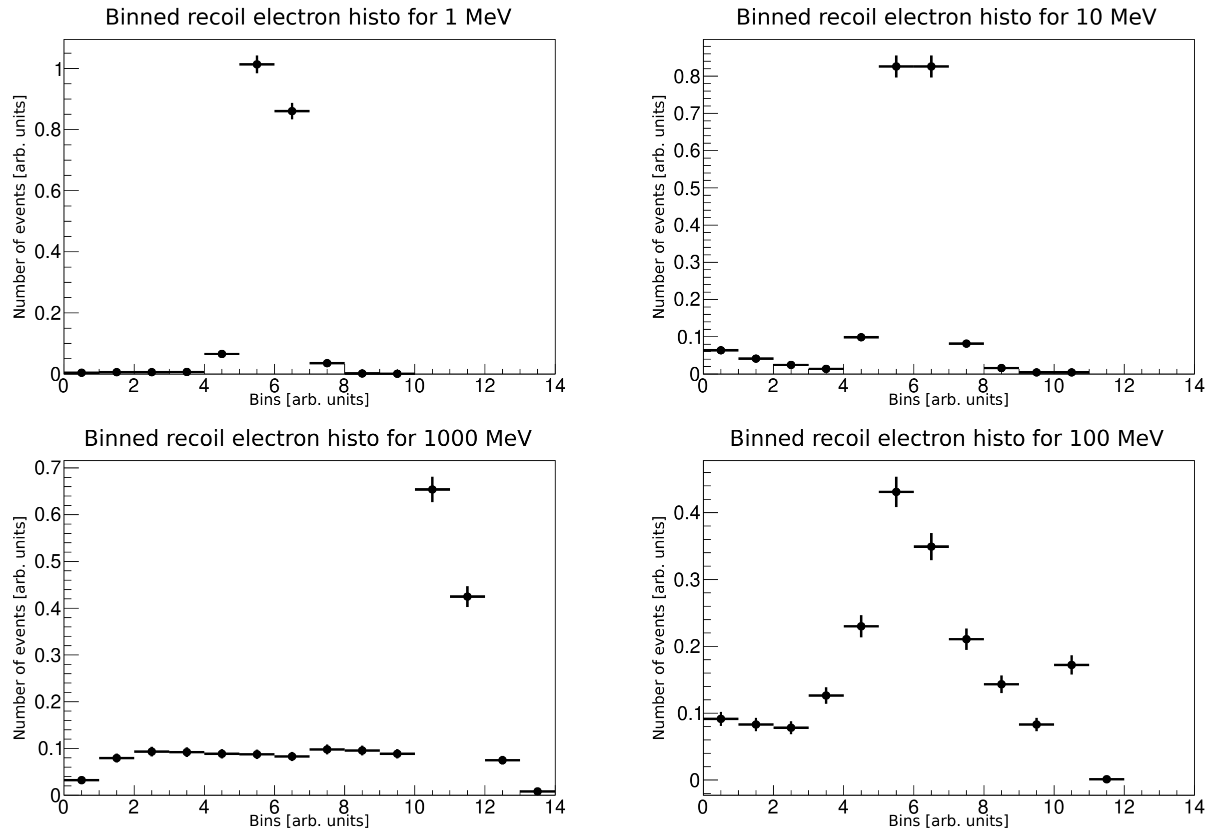

Figure 4: Histograms of signal hypotheses with signal strength of 1. In clockwise order: 1, 10, 100, and 1000 MeV hypothesis. X-axis is unevenly binned pT (transverse momentum).

The model configuration was then fed to the RooStats::ProfileLikelihoodCalculator method, with a specified significance interval. This outputs a plot of the likelihood as a function of the parameter of interest, the central interval required to achieve the specified confidence level (significance), and the most likely value of μ [10].

The model configuration was also entered into the RooStats::MCMCCalculator method which is a Bayesian model to obtain a graph of illustrating the most likely value for the signal strength by using the posterior distribution of the variable as well as the right handed interval needed to satisfy a specific confidence level. It works by using the Markov-Chain Monte Carlo method to find the posterior. This is done via a Metropolis-Hastings algorithm to create a Markov chain which is then used to integrate the likelihood with the prior distribution of the variable thus obtaining the posterior [11].

To calculate the CL_s value of the signal hypothesis with respect to the data, the model configuration is fed to the method RooStats::AsymptoticCalculator. It conducts hypothesis testing by using the asymptotic profile likelihood formula. It then creates an Asimov dataset, where the observed data values equal the expected ones, and uses it to calculate significance [12]. The alternative method, using of RooStats::FrequentistCalculator, conducts a frequentist analysis [13]. The profile likelihood calculator (PLC) and Bayesian posterior calculator (BPC) as well as the asymptotic and frequentist methods' results are cross checked with each other to ensure that they agree.

4 Results and Discussion

4.1 Signal strength analysis

To see a signal, or signs of it the observed result must be sufficiently different from the null or background hypothesis. This is done in accordance with hypothesis testing as described in section 2.3 and 2.5. Thereby, a confidence level can be constructed by taking the signal strength's likelihood distribution via integration as in eq. 9. Doing this with respect to the background only hypothesis allows for the determination of limits at certain confidence levels (at 68% and 95% in the case of this investigation) for each hypothesis. This helps determine if signs of a signal can be seen in the signal seeded datasets. The most likely value of μ, given by the PLC for a given dataset, can then be compared to these levels. It is also required that the PLC value and BPC interval for the most likely signal strength are compatible, i.e. the PLC's value plus minus uncertainty overlaps with the interval given by the BPC. The error on the signal strength from the PLC is required to be compatible with the 68% central interval. Lastly, it is required that the observed signal strength is within the uncertainty of the seeded signal strength as if this is not the case the analysis has failed as it can not be determined which method is correct.

The datasets used are a series of datasets sampled from distributions of background and known signal hypothesis as well as signal strength. The datasets, D1-4, were created with a μ of 12.5 whilst dataset D0 was created with a signal strength of zero hence corresponds to a background only hypothesis (henceforth the datasets are refered to as D0-4). D1 was generated with a signal hypothesis of a 1 MeV mediator, D2 with a 10 MeV mediator, D3 with a 100 MeV mediator, and D4 with a 1000 MeV mediator. Bellow is the analysis of a background only dataset with the signal hypotheses.

Table 1: Background only dataset, D0, computed with signal hypotheses

Signal hypothesis	1 Mev	10 MeV	100 MeV	1000 MeV
Lumi error	1 ± 10^{-6}	$1 \pm 7.98 \cdot 10^{-4}$	$1 \pm 7.94 \cdot 10^{-4}$	$1 \pm 7.77 \cdot 10^{-4}$
PLC most likely μ	$5.67 \cdot 10^{-12}$ ± 72.5	$1.54 \cdot 10^{-9}$ ± 5.02	$3.29 \cdot 10^{-7}$ ± 1.53	$1.27 \cdot 10^{-10}$ $\pm 4.02 \cdot 10^{-1}$
PLC central 68% interval	[0, 23.6]	[0, 4.88]	[0, 1.49]	[0, 0.396]
BPC left handed 95% interval	[0, 56.9]	[0, 24.9]	[0, 7.98]	[0, 2.40]

Observing in Table 1, it can be noted that the PLC's intervals reach zero hence they are effectively identical to a left handed interval in this case. The Lumi error deviation from 1 is next to non existent for all of the hypotheses tested which is as expected. Concerning the values of μ, it can be seen that the most likely signal strength is very small which is what was expected as the D0 set is modelled on a background only hypothesis. The error, on the other hand, is quite significant for 3 out of the 4 hypotheses, larger than $\mu = 1$. Especially the 1 MeV hypothesis as it has an error of 72.5, which is much larger than both the 68% and 95% CL which is strange. As for the other hypotheses the uncertainty corresponds to the upward range from the most likely value of the 68% CL. The 10 and 100 MeV hypotheses also have significant error on the order of unity with upper limits on their 68% confidence intervals at 5.02 and 1.53 respectively as well as 24.9 and 7.98 at 95%. The 1000 MeV hypothesis gave the strongest bounds on the signal strength with it being closer to zero, its confidence intervals are also the closest to zero with upper bounds at 0.396 (68%) and 2.40 (95%).

The 68% and 95% bounds give an indication of the limits on the sensitivity of this analysis. The analysis is not sensitive to a 12.5 μ signal in the 1 MeV hypothesis; it is sensitive to a 68% CL but not a 95% CL in the 10 MeV case; it is sensitive to both the 100 and 1000 MeV hypotheses at the 95% CL. This is to be expected if one looks at the background distribution (see Figure 4 & 3 in section 3 for signal & background); there is much more background and error in bins 6 and 7 which is where most of the predicted signal from the 1, 10, 100 Mev hypotheses is. Consequently precision is lost, but the larger the mass the more signal is expected in the higher p_T bins which is why the 10 and 100 MeV hypotheses far much better than the 1 MeV. The 1000 MeV fares even better as the majority of its signal is not in the sixth and seventh bins but in the eleventh and twelth bin. Hence there are limits for how strong an observed signal needs to be incompatible with the background hypothesis: if it is outside the 68% interval it there is maybe a signal, whilst if it is outside the 95% interval it is probably a signal. Overall, the tests against the background only dataset are reasonable and the analysis can proceed.

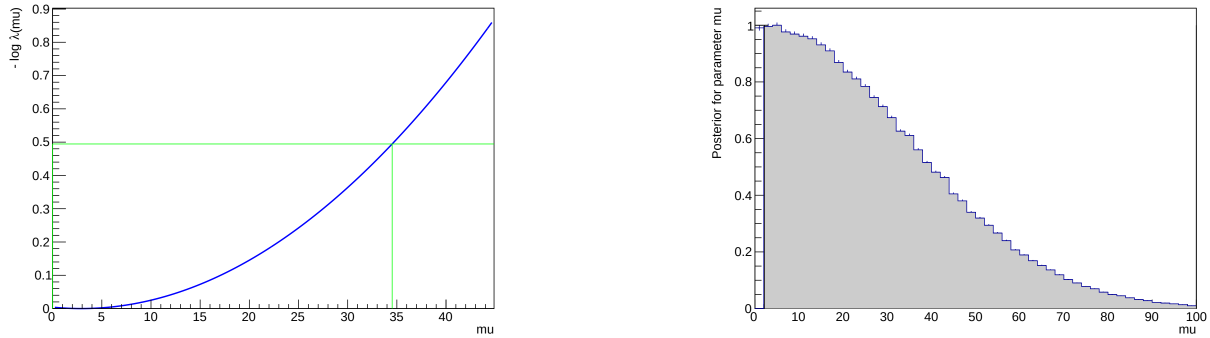

Figure 5: D1 analysed with the PLC and the BPC using the 1 MeV hypothesis. **Left**: Profile likelihood distribution of the signal strength μ, the blue line. A lower value of the negative log indicates a more likely value of the signal strength. The green lines indicate the 68% confidence level centred around the most likely signal strength. **Right**: Normalised posterior distribution of the signal strength, μ, with respect to its most likely value. The grey coloured area corresponds to the right handed 95% confidence level.

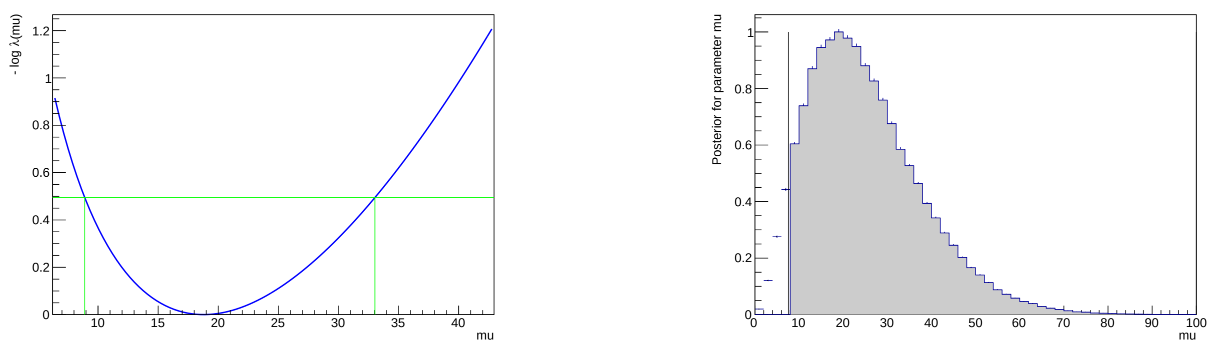

Figure 6: D2 analysed with the PLC and the BPC using the 10 MeV hypothesis. **Left**: Profile likelihood distribution of the signal strength μ, the blue line. A lower value of the negative log indicates a more likely value of the signal strength. The green lines indicate the 68% confidence level centred around the most likely signal strength. **Right**: Normalised posterior distribution of the signal strength, μ, with respect to its most likely value. The grey coloured area corresponds to the right handed 95% confidence level.

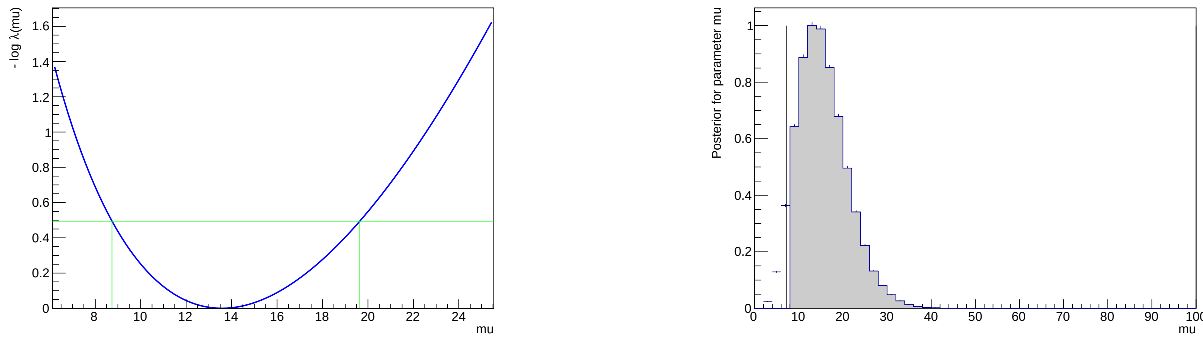

Figure 7: D3 analysed with the PLC and the BPC using the 100 MeV hypothesis. **Left**: Profile likelihood distribution of the signal strength μ, the blue line. A lower value of the negative log indicates a more likely value of the signal strength. The green lines indicate the 68% confidence level centred around the most likely signal strength. **Right**: Normalised posterior distribution of the signal strength, μ, with respect to its most likely value. The grey coloured area corresponds to the right handed 95% confidence level.

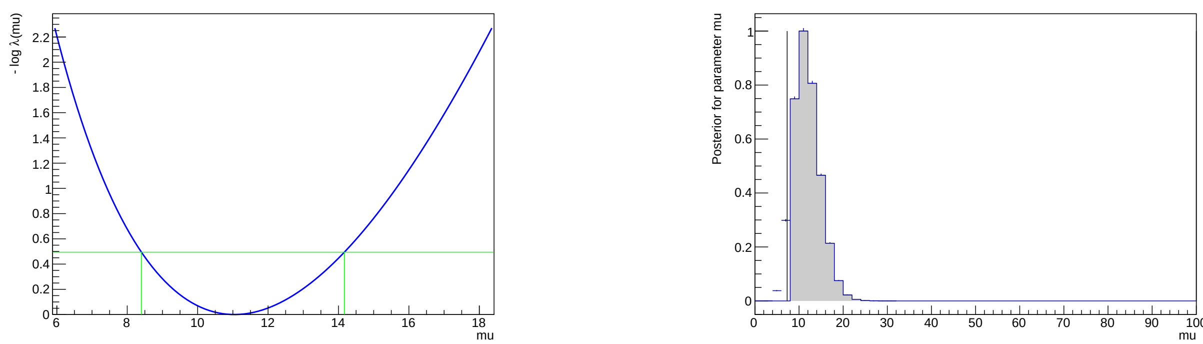

Figure 8: D4 analysed with the PLC and the BPC using the 1000 MeV hypothesis. **Left**: Profile likelihood distribution of the signal strength μ, the blue line. A lower value of the negative log indicates a more likely value of the signal strength. The green lines indicate the 68% confidence level centred around the most likely signal strength. **Right**: Normalised posterior distribution of the signal strength, μ, with respect to its most likely value. The grey coloured area corresponds to the right handed 95% confidence level.

When observing the posteriors (right graphs) in Figures 5-8 note that the grey area colours bin by bin and as the bins have a width of two the grey area may be slightly inaccurate. To compensate for this inaccuracy the grey vertical lines are placed on the edges of the confidence interval.

Bellow is the results of analysis on the datasets with a seeded signal.

Table 2: Datasets D1-4 analyses with the signal hypothesis that corresponds to their seeded signal

Signal hypothesis	D1 dataset & 1 Mev	D2 dataset & 10 MeV	D3 dataset & 100 MeV	D4 dataset & 1000 MeV
Lumi error	1 ± 10^{-6}	$1 \pm 7.31 \cdot 10^{-4}$	$1 \pm 6.69 \cdot 10^{-4}$	$1 \pm 5.64 \cdot 10^{-4}$
PLC most likely μ	3 ± 73	19 ± 12	13.6 ± 5.5	11 ± 3
BPC most likely μ	[4, 6]	[18, 20]	[12, 14]	[12, 14]
PLC central 68% interval	[0, 34.5]	[8.89, 33.0]	[8.74, 19.6]	[8.40, 14.2]
BPC right handed 95% interval	[2.15, 100]	[7.58, 100]	[7.27, 100]	[7.31, 100]

Observing Table 2 it can be concluded that there is good agreement between the PLC and the BPC regarding the most likely value of the signal strength, a graphical visualisation can be seen in Figures 5-8. Continuing with the most likely value of μ from the PLC it can be seen that all of them are compatible with the seeded value of μ within their uncertainty. Focusing on the uncertainties it can be seen that the precision of the results varies significantly. The error on the 1 MeV hypothesis, like observed from the background only test, is especially large and it is not compatible with the central 68% bound, thus the larger of the two are taken as the uncertainly. The other hypotheses' uncertainties roughly correspond to the arithmetic mean of the upward and downward swing of the 68% bound from the most likely value. Still, the uncertainty ranges from more than 70 to 3. This can be attributed to, as discussed previously, the larger number of events where the signal was expected (see Figure 4 & 3 in section 3 for signal & background). Simple comparison logic yields that a signal of 25 events will not be easily distinguished against a background in the thousands or even hundreds.

Connecting back to the results from the testing against the background only model the following is found: the 1 MeV hypothesis satisfy neither the 95% nor the 68% bound; the 10 MeV hypothesis does not satisfy the 95% bound, but does satisfy the 68% bound; the 100 and 1000 MeV hypotheses satisfy both the 68% and 95% bounds.

Taking a holistic view of the results, it can be seen that the sensitivity of the experiment decreases as the expected mass of the dark photon decreases. Consequently, the experiment may be limited in the ability to detect signals in the low mass range, though the much smaller background is expected in the real experiment due to stronger selection requirements together with a possible increase in statistics may remedy the situation even with a smaller signal. Conversely, these changes will make the experiment very sensitive to any fluctuation in the upper mass range. This is a promising result for the methodology of LDMX as it seems possible for it to detect the theorised dark matter.

4.2 CL_s computations

The two methods used in this thesis for calculating the CL_s value of the data with respect to the different signal hypothesis did not work properly. When testing them against the the datasets D0-4 both the asymptotic and frequentist calculators gave varying results for the CL_b values when testing different signal hypotheses. This is problematic as the background p-value (CL_b) should be the same for all signal hypothesis tests as the expected background is the same for each of the signal hypotheses. Furthermore, the two methods do not agree on CL_b and CL_{s+b} (and thus CL_s) values, they may vary by several orders of magnitude. Although the CL_{s+b} value can not be discounted as faulty, it is very possible that it is also incorrect, considering the other faulty values calculated. It is not possible to extract useful information from the comparison of the two methods by considering the trend of the CL values – test all different hypotheses and look at which one is the best for both – as they do not agree on the ordering of the hypotheses.

The frequentist calculator also has the problem of not always being able to calculate the CL_s value. Instead it would return infinity and nan. It was attempted to address this issue by drastically increasing the number of Toy MCs used in its calculation by two orders of magnitude to see it became more reliable, but this was not successful.

It should be considered that the source of the error could be something other than the functions themselves, it could come from the HistFactory::MakeModelAndMeasurementFast, but this is unlikely as the profile likelihood calculator and Bayesian posterior are seemingly correct. A more likely source of the error would be how the signal hypotheses are setup to be inputted into the asymptotic and frequentist calculators. This has the logical advantage over the idea that both the frequentist calculator and asymptotic calculators are wrong because it constitutes a single point of failure instead of two. It is also possible that the source of the issue is the small dataset which seems unlikely and would be problematic for LDMX considering the hope of zero background. There was already an issue with them not working with empty bins; it is possible that the filling of those bins with a small value and error did not fix the issue (see the method 3). Still, this is just speculation as the source of the error is still unknown.

5 Conclusion and Outlook

To summarise the investigation yielded that there is a significant decrease in sensitivity as the expected mass of the dark photon decreases. So much so that the analysis conducted was insensitive to the low mass hypotheses completely to the 1 MeV and partially to the 10 MeV hypothesis. Signs of signals where observed for the higher mass hypotheses. Overall, there is a trend with higher mass hypotheses being easier to observe due to the smaller background where they create a signal. Moreover, the results are promising for the real experiment as it is expected to have a much smaller background.

There are several issues with the method undertaken for this investigation. Firstly, consider the handling of the systematic uncertainties, which where added according to standard error propagation. The issue with this is that systematic uncertainties may be further constrained via fitting to a control region as described in section 2.4. Furthermore, this effect will be amplified further as a more complex model is used with several sources of systematic uncertainties. Another issue is the lack of a working calculator for CL_s values. The pre-programmed functions do not work as intended (RooStats::AsymptoticCalculator and RooStats::FrequentistCalculator). As stated before this was under the assumption that it ordered the likelihoods of the hypotheses correctly to then apply the PLC and BPC. For the future a working way of calculating CL values with regards to a specific hypotheses should be implemented. Possibly a pre-programmed hybrid calculator (RooStats::HybridCalculator) from ROOT could be used. There is also the issue of adding the statistics to empty bins. Adding statistics to empty bins introduces unnecessary error which should fixed in a future analysis. It could also be attempted to automate a search for a signal in the range 1 MeV to 1000 Mev. That is, making the code calculate the CL_s values for several signal hypotheses within a range and returning the most likely signal candidate hypothesis as the result.

6 References

[1] Torsten Åkesson et al. *Light Dark Matter eXperiment (LDMX)*. 2018. DOI: 10.48550/ARXIV.1808.05219. URL: https://arxiv.org/abs/1808.05219.

[2] Torsten Åkesson et al. *A High Efficiency Photon Veto for the Light Dark Matter eXperiment*. 2019. arXiv: 1912.05535 [physics.ins-det].

[3] Sophie Charlotte Middleton. *LDMX: The Light Dark Matter eXperiment*. [Online; accessed 10. May 2023]. 2021. URL: https://sophiemiddleton.people.caltech.edu/documents/19300/APS_v5.pdf.

[4] Anders Floderus. *Luminosity determination and searches for supersymmetric sleptons and gauginos at the ATLAS experiment*. 2014. URL: https://lup.lub.lu.se/search/files/5523057/4858083.pdf.

[5] Jeremy McCormick tomeichlersmith Omar Moreno. *ldmx-sw*. [Online; accessed 10. May 2023]. May 2023. URL: https://github.com/LDMX-Software/ldmx-sw.

[6] Lene Kristian Bryngemark. Private communication. May 2023.

[7] Noah Ritter. *Statistical-analysis-of-LDMX-data—-code*. [Online; accessed 10. May 2023]. May 2023. URL: https://github.com/Kanteno/Statistical-analysis-of-LDMX-data----code/tree/main.

[8] B. Brun and F. Rademakers. *ROOT – An object oriented data analysis framework*. Version 6.26/10. May 2023. URL: http//root.cern/.

[9] LorenzoMoneta. *RooFit/RooStats tutorials for INFN School of Statistics 2013*. [Online; accessed 10. May 2023]. May 2023. URL: https://twiki.cern.ch/twiki/bin/view/RooStats/RooStatsTutorialsJune2013.

[10] *ROOT: RooStats::ProfileLikelihoodCalculator Class Reference*. [Online; accessed 10. May 2023]. May 2023. URL: https://root.cern.ch/doc/master/classRooStats_1_1ProfileLikelihoodCalculator.html.

[11] *ROOT: RooStats::MCMCCalculator Class Reference*. [Online; accessed 8. May 2023]. May 2023. URL: https://root.cern.ch/doc/master/classRooStats_1_1MCMCCalculator.html.

[12] *ROOT: RooStats::AsymptoticCalculator Class Reference*. [Online; accessed 8. May 2023]. May 2023. URL: https://root.cern.ch/doc/master/classRooStats_1_1AsymptoticCalculator.html.

[13] *ROOT: RooStats::FrequentistCalculator Class Reference*. [Online; accessed 8. May 2023]. May 2023. URL: https://root.cern.ch/doc/master/classRooStats_1_1FrequentistCalculator html.