Photon Timing and Preliminary Pulse Shape Discrimination in LUX^{*}

Rose Baunach Whitman College, Department of Physics Walla Walla, WA 99362

(UC Davis Physics REU Program)

Advisors: Mani Tripathi and Brian Lenardo University of California Davis, Department of Physics, Davis, CA 95616

(Dated: September 4, 2015)

This paper includes the following: A) a review of the computational framework used currently to identify photons and calculate their arrival times in order to perform Pulse Shape Discrimination (PSD) analysis on experimental data from the Large Underground Xenon Experiment (LUX); B) a review of improvements made to the analysis code for identifying and timing photons, including 1) the addition of area constraint probabilities, and 2) the capacity for the code to identify 5 photons per PMT instead of 3; and, finally C) a review of our efforts to track and verify improvements made to our analysis code via a PSD test-run on CH_3T and DD calibration data. The research detailed in this paper supports the UC Davis branch of the LUX PSD analysis effort.

I. INTRODUCTION

From Zwicky's initial analysis of the Coma cluster in 1933, to the 2006 observation of the Bullet Cluster merger, we have ample evidence of the existence of dark matter [1]. The current problem is to identify the particle that makes up the majority of dark matter in the universe. The leading candidates are WIMPs, or weakly interacting massive particles, a category of particles that are both cold and non-baryonic. Direct detection of these WIMPs is the main goal of the Large Underground Xenon Experiment (LUX).

Direct detection of WIMPs is based on the principal that most normal radiation, such as γ and β , interact electromagnetically producing electronic recoils (ER). WIMPs, in contrast, only interact via weak and gravitational forces, and so only produce nuclear recoils (NR). The ability to accurately discriminate between ER and NR events is therefore of great importance in successfully detecting potential WIMP signals in LUX.

The current method of data discrimination in LUX uses a charge/light ratio (S2/S1), resulting from ionization and excitation of the xenon atoms. Pulse Shape Discrimination (PSD) is intended to be a complementary data discrimination method for LUX. PSD presents the possibility of two-dimensional ER/NR discrimination, which would aid in further constraining the dark matter search window (see Fig. 1).

Within this paper, we present a review of the computational framework being used to identify photons and calculate their arrival times in order to perform PSD analysis on LUX experimental data. We additionally discuss improvements made to our analysis code, and some preliminary results from a PSD test-run on calibration data.



FIG. 1. The effect of adding a PSD cut to the dark matter search window. Horizontal axis is current charge/light (S2/S1) discrimination parameter, and the vertical is the PSD parameter (graphic credit: B. Lenardo).

II. BACKGROUND

A. WIMPs

The most accepted theory of dark matter is that it is made up of weakly interacting massive particles, called WIMPs. The motivation to assume a weakly interacting massive particle comes from such evidence as density fluctuations in the Cosmic Microwave Background and the age of galaxies (for details see [1]). However, no particle in the Standard Model matched the WIMP profile,

^{*} Any specific questions about the UC Davis branch of the PSD effort in LUX can be addressed to Brian Lenardo (bglenardo@ucdavis.edu) from September 2015 to May 2016.

so various particle candidates came from supersymmetry [1] and other theories beyond the Standard Model. The collection of various particles that could potentially satisfy the accepted cosmological and astronomical predictions and observations are collectively called WIMPs.

B. The LUX detector

The LUX detector is a dual-phase (liquid-gas) time projection chamber located 4850 ft underground at the Sanford Underground Research Facility in Lead, South Dakota.

The core of the detector is a large cylinder which contains 350 kg of liquid xenon (LXe), with an array of 61 photomultiplier tubes (PMTs) at each end (see Fig. 2).

LXe is a logical detection medium for WIMPs because its high Z content yields high interaction potential for heavy nuclear recoils [2]. It also has a high light yield which makes detection of NR and ER signals easier than in other mediums (see the next section for details on signals in LUX), and it is relatively easy to purify [2].

The PMTs in the LUX detector operate by the following principal: a photon is incident on a transmitting photocathode, which by the photoelectric effect releases a photoelectron (phe). This photoelectron is then accelerated through a series of dynodes that exponentially multiply the number of photoelectrons until a signal can be digitally measured from the PMT.

Among other precautions, the LUX detector is submerged in a large cylindrical water tank to shield it from background radiation.



FIG. 2. The core of the LUX detector, shown are the LXe cylinder and the two hexagonal PMT arrays. Production of S1 and S2 signals is also depicted.

1. Signals in LUX

Each scattering interaction in LUX produces two types of signal: primary scintillation light (S1) and secondary electroluminescence (S2), both of which are measured by the PMT arrays (see Fig. 2). S1 signals occur when an incident particle deposits energy in the liquid xenon, which causes various xenon atoms to excite and then deexcite emitting photons (scintillation light), which are measured by the PMT arrays. The photons measured in a S2 signal are proportional to the charge released at the interaction site due to xenon ionization. The electrons that comprise the S2 signal are drifted up to the top of the detector by an applied E-field, where they collide with xenon gas and emit photons from this secondary interaction, which are then detected by the top PMT array.

Since a typical incident particle produces a projectile in the LXe with some energy in the keV range, the complete S1 or S2 signal we measure for an ER event is the product of many xenon excitations and ionizations, and hence many photons. WIMPs, due to their low cross section, won't cascade through the detector like other particles, but when they interact with a xenon nucleus they impart some energy to the nucleus in the keV range. This energy transfer from the WIMP allows the xenon nucleus to cascade throughout the detector, producing lots of scintillation light and ionization electrons, and hence many photons in the resulting overall S1 and S2 signals.

2. Calibration Data

For this paper, there are two relevant LUX calibration data sets, Tritium (CH_3T) and DD (neutron generator).

When released into the LXe, CH_3T mimics the typical ER signal. CH_3T is ideal because it decays in the energy region where we most expect to find WIMPs (0-18 keV), it is easy to dissolve throughout the detector, and it is easy to remove from the LXe post-calibration.

Similarly, the DD (neutron generator), being a source of neutrons, mimics a typical NR signal. We only utilize the elastic scatters from the DD data, as these are exactly the signals expected of WIMPs. Inelastic neutron scatters are easy to reject, since these only occur at energies higher than our region of interest.

These calibration data sets are useful for testing our analysis in that they mimic the signals of real data, yet are already defined as NR and ER, and so can provide good validation for our analysis.

C. A Short Overview of PSD

Pulse Shape Discrimination is an analysis method for detectors that produce scintillation light, and has been in use for over forty years [3]. However, the research this paper supports represents some of the first attempts to apply PSD to LUX data.

There are many types of PSD, such as the rise time technique, the charge method, and the optimal method, which is also called Gatti's method [3]. In this paper, we will apply a form of the charge method which uses a prompt to total time ratio. It has been shown that the charge method is closest to Gatti's method in discrimination potential [3], so a test-run with the charge method is useful groundwork for future LUX PSD analysis using Gatti's method.

1. PSD in LUX

In contrast to the typical discrimination in LUX data analysis, which uses a charge/light (S2/S1) ratio, PSD in LUX only considers S1 signals and exploits a property of the excitation states of xenon for its discrimination.

LXe has two excited states, a singlet and a triplet state, and their associated lifetimes are on the order of 4 and 24 ns, respectively. It is strongly predicted that ER and NR will populate these states differently, with S1 signals from NR events generating significantly more photons from the singlet state and relatively few from the triplet state, while photons from ER S1's are more evenly distributed across both states (see Fig.3). Therefore, a time cut in S1 events presents itself as a logical starting point for PSD.

But before we move on to discussing an example of PSD, some terminology clarifications are in order regarding event measurement. For each ER or NR event in our analysis, we only consider 1 pulse per event, which is the S1 signal. This overall S1 pulse, (and hence all its constituent photons), is divided up and measured by the 122 PMT channels in the detector. Each PMT channel is capable of measuring a certain number of photons. As we will discuss later in this paper, our analysis framework is currently capable of identifying up to 5 photons per PMT channel (versus 3 before our code modifications). So at 5 photons per channel, with 122 channels, we are in principle sensitive up to 610 photons per overall S1 pulse, and hence per ER or NR event. Typically a S1 signal for an ER or NR event will fall in the range of 0-100 photons across all PMT channels, as shown in Fig.3 on the left. As an additional note, when you begin to consider entire data sets, these are composed of many S1's and hence proportionally many more photons.

Now, as mentioned previously, in our analysis we will apply a version of the charge PSD method for ER and NR events using a prompt/total time ratio, with the intent of laying groundwork for eventual implementation of Gatti's method. To illustrate how we will later apply the charge PSD method, see Fig.3. On the left, Fig.3 shows a typical distribution of photons in time for a S1 signal across all 122 PMT channels for 1 ER and 1 NR event. To apply a prompt/total discrimination, simply make a uniform time cut (shown by a vertical line through both distributions), define everything to the left of the line as "prompt" photons, and then calculate your prompt/total time ratio for both ER and NR events. If you repeat this ratio calculation procedure for all the ER and NR events in a data set, you will eventually obtain a distribution of ER and NR prompt/total ratios, which can then be graphed, as shown in Fig.3, on the right. These ratio distributions will then generate a PSD cut. In Fig.3 we could say, for example, that all events which generate a prompt/total ratio greater than 0.17-0.2 are likely NR events.

The prompt/total ratio is effective and relatively sim-

time

0.2

0.4

Prompt/Total Ratio

0.6

FIG. 3. A schematic of photons in time for 1 ER and 1 NR event, and some associated potential PSD prompt/total time ratio distributions. The left is a typical example of the distribution of photons in time for a S1 signal across all 122 PMT channels for ER and NR events with a uniform time cut applied; the right is an example of the prompt/total ratio distributions that can be obtained from many such events after applying the charge PSD method. The several overlaid ER and NR ratio distributions show the effect of increasing number of photons. The figure at the right was modified from G. Ranucci [3].

ple to calculate. However, in order to perform PSD analysis, it is necessary to first assign photon arrival times (the time it takes for the photons to be measured by a PMT) to all the photons within an overall S1 pulse. And in order to assign arrival times, you must first accurately determine the number of photons measured in a pulse shape. For the remainder of this paper, Photon Timing will be the term used to describe these steps of identifying photons and assigning their arrival times. It is this photon timing process that is the subject of the analysis of the next section.

III. PHOTON TIMING: COMPUTATIONAL ANALYSIS FRAMEWORK

A. The PhotonTiming Module

Our photon timing analysis is performed with a set of C++ and ROOT executable C++ programs. The PhotonTiming Module is a set of C++ programs with ROOT capabilities enabled, executed in a python shell. The module is written to take one or more LUX data sets as input, process them into *.root files containing several useful quantities such as the area of each pulse in the data set, and then identify the number of photons in each pulse in the data set and their corresponding arrival times (before writing all the information to the root file the module created).

The part of the module this paper will concentrate on will be the identifying of photons and assigning their arrival times, as this is the part of the module we have made significant improvements to.

As previously mentioned, for each ER or NR event we

consider one signal, the overall S1 pulse. This pulse is divided and measured by 122 PMT channels. Before our improvements, the general process by which the module identified the number of photons in a S1 pulse on the per PMT level was as follows: Given a set of average pulse shapes for 1-3 photons incident on a PMT, a bestlikelihood fit was calculated for how well a pulse in our data matched each average pulse. The number of photons in a specific PMT pulse shape was then assigned based on whichever likelihood fit was the largest, and the location of each photon in the overall pulse and a corresponding arrival time was then calculated.

As a side note, photon arrival times are measured relative to the start of the S1 signal using the 25% area time. The 25% are time, which we will mention again later, is the time at which 25% of the overall S1 pulse has been integrated over. In our analysis, the integral over the overall S1 pulse across time is the proxy for energy.

But going back to our original identification method, an example schematic of the likelihood fitting procedure is depicted in Fig. 4, where on the bottom you can clearly see a waveform with two distinct peaks, corresponding to two photons within the overall PMT pulse shape. The bottom right shows an example of the process the program goes through to fit two photons to this pulse shape, in order to determine each photon's arrival time.



FIG. 4. An example schematic of how the PhotonTiming Module fits photons to pulse shapes on a per PMT level. The bottom is an example of fitting photons to a pulse with two distinct peaks, the top shows a pulse shape where the photons are closer than two distinct peaks. (Graphic credit: B. Lenardo).

This method of assigning photons to pulses based on a likelihood fit is effective when there are distinct peaks, however it is not sufficiently accurate to correctly identify photons when photon peaks are not clearly defined, as in the top of Fig.4. This initial insufficiency in our code led to the first of the photon timing improvements within our module, which was the addition of an area constraints probability in how the module decides the number of photons to assign to a pulse on the per PMT level.

B. Improvements to the PhotonTiming Module

1. Addition of Area Constraints

The idea of the area constraints probability comes from a division of statistical analysis called Bayesian Model Comparison, detailed in [4]. Consider the following equation:

$$P(H_i|D) \propto P(D|H_i)P(H_i) \tag{1}$$

 $P(H_i|D)$ is the probability of a specific model H_i given the data D. So for our purposes, $P(H_i|D)$ is the probability of i photons being contained in a pulse, given the data of the pulse itself. $P(D|H_i)$ is the likelihood fit that we mentioned earlier, the best fit for our data given an averaged pulse shape for i photons. So the original method the module used to determined how many photons were in a pulse at the per PMT level was to calculate the probability based on this equation:

$$P(H_i|D) \propto P(D|H_i) \tag{2}$$

and then assign the number of photons to a pulse based on whichever probability, $P(H_1|D)$ or $P(H_2|D)$ or $P(H_3|D)$, was the greatest.

However, if you look again equation 1, you see another factor, $P(H_i)$, or the probability of your specific model H_i . We have interpreted $P(H_i)$ to be the area constraints probability. This area constraints probability works like this: given the PMT pulse the module is analyzing, the module calculates the pulse area and determines the probability that the pulse area came from an area probability distribution of 1 or more incident photons. In other words, the module contains equations for specific probability distribution functions for one to the total number of photons it is capable of assigning to a pulse shape. The code takes the pulse area as an input "x-value" and outputs the corresponding function value, or $P(H_i)$, for each of the probability distributions. $P(H_i)$ is then taken into account when the module calculates the probabilities $P(H_i|D)$ for each of the possible numbers of photons H_i it could assign to the PMT pulse being analyzed.

This modification of adding the area constraints probability is significant in that the module now determines how many photons it assigns to a pulse on the per PMT level based on pulse shape and pulse area, which, as we will discuss later, drastically improves our ability to resolve photons closer than distinct spikes (see the top of Fig.4).

2. Creating the Area Constraint Probability Distributions

To create the area constraint probability distributions mentioned in the previous section, the first step was to create average pulse area distributions for incident single photons (single photoelectrons) on each PMT in the LUX detector, following the work done by A. Manalaysay [5]. We know that creating unique single photoelectron (sphe) area distributions for each PMT in the detector is necessary due to each PMT having its own resolution idiosyncrasies that affect the distribution of possible pulse areas, among other things [5]. An example of an average sphe pulse area distribution for PMT 60 is shown in Fig.5.



FIG. 5. Average sphe pulse area distribution for PMT 60.

To create the average sphe pulse area distributions for all 122 LUX PMTs, the following process was utilized for a ROOT executable C++ code:

- Input data: 5 CH_3T data sets (December Run 03 LUX data)
- Select only S1 pulses
- Select only S1 pulses with overall pulse area less than 10 phe
- For each PMT, histogram all pulse areas that passed the selection criteria across all 5 data sets

The pulse area below 10 phe cut insures that we're only considering S1 pulses that have a low probability of having 2 photons incident on the same PMT, which would bias the sphe distributions. Since we have 122 PMT channels, and if we assume all channels are equally likely, at high energies (and hence higher photon count) we are likely to see multiple photons on the same PMT. Conversely, if we only had 10 or less photons (less than 10 phe), we are very unlikely to have 2 photons incident on the same PMT.

The process to turn these sphe PMT area distributions into specific area probability distribution functions for each PMT included the following: Another ROOT executable C++ program was written to utilize Minuit, the FORTRAN multi-parameter capable fitting program adapted for ROOT, in order to fit each area distribution with the following constrained double Gaussian function, using a Chi-squared fit.

$$y = ae^{\frac{-(x-\mu)^2}{2\sigma^2}} + be^{\frac{-(x-2\mu)^2}{2(\sqrt{2}\sigma)^2}}$$
(3)

We fit a double Gaussian function because when 1 photon is incident on a PMT, there is a probability of measuring 1 or 2 corresponding photoelectrons in your signal. We expect the means of the function, corresponding to pulse areas, to be distributed around 1 and 2 phe as a result. The other stipulation in the function, that the second standard deviation be $\sqrt{2}$ of the first, is assumed for both mathematical reasons and due to previous experience.

The results of fitting PMT 60's sphe area distribution with this double Gaussian are shown in Fig.6. We constrain the fit to be between 0.4 and 3 phe to account for edge effects in the detector.





FIG. 6. Double Gaussian fitted sphe area distribution for PMT 60.

The area distribution fitting code then extracts the 4 fit parameters (a, b, μ , σ) in equation 3, and then uses them to specify a unique double Gaussian function for each PMT. These double Gaussian functions are then normalized to create the sphe area constraint probability functions for each PMT in LUX.

The above process creates the area constraint probability function for single photons incident on a PMT, which yields $P(H_1)$; but to create probability distribution functions for more than 1 incident photon, we convolve the general sphe double Gaussian function (eq. 3) multiple times. For example, to create the area constraint function for 2 photons incident on a PMT, simply convolve the general sphe double Gaussian function (eq.3) with itself, and then normalize the result to obtain a dphe area constraint probability distribution function. We assume we can take the analytical convolution rather than a numerical approach because the tails on the resulting Gaussians all go to zero fast enough so the differences in approach would be extremely negligible.

The PhotonTiming module stores all the PMT specific sphe parameters, and when it calculates $P(H_i)$ for 1 or more incident photons, it takes the PMT number as one of its inputs, and then inserts the specific sphe parameters into whichever general area constraint probability distribution function is being called for.

3. 5 photon capability

Besides the addition of area constraint probabilities, which as previously discussed is extremely useful in helping us correctly identify photons closer than two distinct peaks, another modification we made to our original code was introducing the capability of identifying up to 5 photons per PMT channel. Previously when analyzing our PMT pulse shapes, the code was only capable of identifying and timing up to 3 photons in a waveform on the per PMT level. We enabled the logical decisions in our code so that the module is now able to identity and time up to 5 photons per PMT pulse. A test of our code on the S1 signals in one of our 5 CH_3T data sets shows that we are indeed identifying 1-5 photons in logical proportions to each other (see Fig.7).



FIG. 7. Test of 5 photon enabled PhotonTiming Module on 1 CH_3T data set. Single photons are counted most often, and the number of 2-5 photons counted decrease in expected proportions.

Further diagnostics were then run on the same CH_3T data set to investigate the accuracy of the module's photon identification with the 5 photon capability enabled. Figure 8 displays a plot of Photon Count/Pulse Area vs Photon Count for the data set, with the mean of the distribution noted in blue squares. Pulse area here refers to the pulse area of individual photons contained in the per PMT pulses, from all the S1 signals in the data set. This distribution is useful because it gives us a clue as to how effectively our module is counting photons when it considers their associated pulse areas. A relatively narrow distribution with a mean around y = 1 would be ideal, as that would signify we are reconstructing 100 % of the photons we identify correctly (with a one-to-one Photon Count to Pulse Area ratio) with no loss of identification power at higher photon count (which corresponds to higher energies).

The actual mean of this distribution is plotted in Fig.9, where the linear fit is 98.03 - 0.1623x%. This signifies a successful reconstruction of 98% of our initial photons, with a gradual loss of reconstruction power less than 0.2%. The fit was performed between 15-80 photons, in order to discount "turn-up" and "drop-off" effects. The



FIG. 8. Photon Count/Pulse Area vs Photon Count distribution for 1 CH_3T data set. Mean of the distribution shown in the blue dotted line.



FIG. 9. Mean fit line of the Photon Count/Pulse Area vs Photon Count distribution shown in figure 8. The fit was calculated between 15 - 80 photons to discount "turn-up" and "drop-off" effects. A fit line of 98.03 - 0.1623x % was calculated with the 5 photon and area constraint enabled module.

turn-up effect has been observed throughout the use of our module, it refers to the tendency of the mean of the distribution in fig. 8 to be greater than 1 for low photon count. Drop-off is the opposite, it is the tendency for the mean of the distribution to drift towards much less than y = 1 at high photon count, contrary to the mean behavior of the bulk of the distribution. We believe drop-off is simply caused by a lack of high statistics within the high energy regime in our current data sets, and is not a symptom of the success of the module itself.

IV. RESULTS AND DISCUSSION

A. Results of PhotonTiming Module Improvements

The real significance of the area constraint and 5 photon capability modifications to the PhotonTiming Module becomes apparent when you make a comparison between the state of the module before and after those improvements. Figure 10 shows the mean fit lines of the same CH_3T data set as in Fig.8 and Fig.9, both with and without the area constraints and 5 photon per PMT capability included in the module. It is apparent from the fit lines that we have made noticeable improvements, as we are now reconstructing 98% of our initially measured photons, instead of 97%, and we now have a loss of photon reconstruction power of 0.1623% instead of 0.1744%.



FIG. 10. Mean fit line comparison of the original and current state of the PhotonTiming Module's reconstruction capabilities for the same CH_3T data set. As denoted in the legend, the blue line is the mean fit for the data set with the area constraints and 5 photon per PMT capability, whereas the yellow is the data set without those improvements.

B. Preliminary PSD on Calibration Data

After the success of the improvements to the Photon-Timing Module, we ran a preliminary PSD analysis over ER and NR event calibration data (CH_3T and DD, respectively) using the form of the charge PSD method that utilizes a prompt/total time ratio (as detailed earlier in this paper). We ran the PhotonTiming Module over the same 5 CH_3T data sets (Run 03 December) that have been previously utilized in this paper, as well as 20 DD (Run 03) data sets. The larger number of NR data sets is motivated by the fact that NR data typically has lower statistics.

Figure 11 shows the average S1 ER and NR pulses for many events across all the data sets we processed. As shown on the x-axis, to obtain each averaged pulse shape we subtracted a 25 % area time from each photon's arrival time in order to line up the average pulse distributions for better comparison. It is apparent in Fig.11 that the average NR pulse is both slightly narrower and distributed



FIG. 11. Average ER and NR S1 pulse shapes for our prompt/total PSD test-run.

After validating that the average overall S1 pulse shapes for ER and NR were as expected, we calculated a prompt-to-total ratio for all the S1 pulses across all 25 data sets, the results of which are shown in Fig.12. While the PSD ratio distributions of Fig.12 are not as clearly distinguishable as the sample distributions of Fig.3, the fact that they are distinct from one another is encouraging for the future of this analysis method.



FIG. 12. PSD test-run ratio distributions for a prompt/total time ratio.

V. CONCLUSIONS

This paper has been a review of the noticeable progress made in our PSD analysis capabilities. This progress is due to code changes made to the PhotonTiming Module, such as the addition of area constraints and the capability to identify 5 (versus 3) photons per PMT (see Fig.10), which have significantly improved how accurately we identify photons and assign their arrival times. Further research, however, is still necessary to understand and mitigate small idiosyncrasies in the PhotonTiming Module such as the turn-up effect.

Furthermore, the data gained from the preliminary PSD prompt/total test-run on calibration data is encouraging for the future of LUX PSD research. The fact that our module identifies noticeably different average S1 pulse shapes for ER and NR (Fig.11), and that the PSD ratio distributions for the same set of S1 pulses are distinct from one another (Fig.12), shows that our module can discriminate between ER and NR events. These preliminary results warrant further testing of the Photo-Timing Module on actual data, as well as utilizing even more accurate PSD methods, such as Gatti's method, to see if we can realize even greater discrimination potential.

VI. ACKNOWLEDGEMENTS

Thank you to Brian Lenardo, Dr. Mani Tripathi, and the rest of the UC Davis LUX/LZ group for all their help and guidance with this project. Thank you to my fellow REU students for a very fun summer, and to Dr. Rena Zieve for organizing the program. And thank you to the NSF for the financial funding that made this experience possible.

VII. REFERENCES

- 1. Sanders, R. (2010). The Dark Matter Problem: A Historical Perspective. Cambridge University Press.
- 2. Aprile, E. et all. (2006). Noble Gas Detectors. Wiley-VCH. p.116.
- Ranucci, G. (2004). A Review of the Statistical Foundations of the Classical Pulse Shape Discrimination Techniques in Scintillation Applications. IEEE.
- 4. MacKay, D. (2003). Information Theory, Inference, and Learning Algorithms. Cambridge University Press.
- 5. Manalaysay, A. (2014). Size of Single Detected Photons. LuxDB00000315.