

# Improving Photoelectron Counting and Particle Identification in Scintillation Detectors with Bayesian Techniques

M. Akashi-Ronquest<sup>f</sup>, P.-A. Amaudruz<sup>s</sup>, M. Batygov<sup>e,d</sup>, B. Beltran<sup>a</sup>, M. Bodmer<sup>i</sup>, M.G. Boulay<sup>m</sup>, B. Broerman<sup>m</sup>, B. Buck<sup>g</sup>, A. Butcher<sup>m</sup>, B. Cai<sup>m</sup>, T. Caldwell<sup>h,\*</sup>, M. Chen<sup>m</sup>, Y. Chen<sup>a</sup>, B. Cleveland<sup>o,e</sup>, K. Coakley<sup>h</sup>, K. Dering<sup>m</sup>, F.A. Duncan<sup>o,e</sup>, J.A. Formaggio<sup>g</sup>, R. Gagnon<sup>m</sup>, D. Gastler<sup>b</sup>, F. Giuliani<sup>i</sup>, M. Gold<sup>i</sup>, V.V. Golovko<sup>m</sup>, P. Gorel<sup>a</sup>, K. Graham<sup>d</sup>, E. Grace<sup>n</sup>, N. Guerrero<sup>g</sup>, V. Guiseppe<sup>p</sup>, A.L. Hallin<sup>a</sup>, P. Harvey<sup>m</sup>, C. Hearns<sup>m</sup>, R. Henning<sup>j,r</sup>, A. Hime<sup>f,k</sup>, J. Hofgartner<sup>o</sup>, S. Jaditz<sup>g</sup>, C.J. Jillings<sup>o,e</sup>, C. Kachulis<sup>b</sup>, E. Kearns<sup>b</sup>, J. Kelsey<sup>g</sup>, J.R. Klein<sup>l</sup>, M. Kuźniak<sup>m</sup>, A. LaTorre<sup>l</sup>, I. Lawson<sup>o</sup>, O. Li<sup>o</sup>, J.J. Lidgard<sup>m</sup>, P. Liimatainen<sup>o</sup>, S. Linden<sup>b</sup>, K. McFarlane<sup>d</sup>, D.N. McKinsey<sup>t</sup>, S. MacMullin<sup>j,r</sup>, A. Mastbaum<sup>l</sup>, R. Mathew<sup>m</sup>, A.B. McDonald<sup>m</sup>, D.-M. Mei<sup>p</sup>, J. Monroe<sup>n</sup>, A. Muir<sup>s</sup>, C. Nantais<sup>o</sup>, K. Nicolics<sup>m</sup>, J.A. Nikkel<sup>m</sup>, T. Noble<sup>m</sup>, E. O'Dwyer<sup>m</sup>, K. Olsen<sup>a</sup>, G.D. Orebi Gann<sup>c</sup>, C. Ouellet<sup>d</sup>, K. Palladino<sup>o</sup>, P. Pasuthip<sup>m</sup>, G. Perumpilly<sup>p</sup>, T. Pollmann<sup>e</sup>, P. Rau<sup>m</sup>, F. Retière<sup>s</sup>, K. Rielage<sup>f</sup>, R. Schnee<sup>a</sup>, S. Seibert<sup>l</sup>, P. Skensved<sup>m</sup>, T. Sonley<sup>m</sup>, E. Vázquez-Jáuregui<sup>o</sup>, L. Veloce<sup>m</sup>, J. Walding<sup>n</sup>, B. Wang<sup>a</sup>, J. Wang<sup>i</sup>, M. Ward<sup>m</sup>, C. Zhang<sup>q</sup>

<sup>a</sup>Department of Physics, University of Alberta, Edmonton, Alberta, T6G 2R3, Canada

<sup>b</sup>Department of Physics, Boston University, Boston, MA 02215, USA

<sup>c</sup>Department of Physics, University of California, Berkeley, Berkeley, CA 94720, USA

<sup>d</sup>Department of Physics, Carleton University, Ottawa, Ontario, K1S 5B6, Canada

<sup>e</sup>Department of Physics and Astronomy, Laurentian University, Sudbury, Ontario, P3E 2C6, Canada

<sup>f</sup>Los Alamos National Laboratory, Los Alamos, NM 87545, USA

<sup>g</sup>Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

<sup>h</sup>National Institute of Standards and Technology, Boulder, CO 80305, USA

<sup>i</sup>University of New Mexico, Albuquerque, NM 87131, USA

<sup>j</sup>Department of Physics and Astronomy, University of North Carolina, Chapel Hill, NC 27599, USA

<sup>k</sup>Pacific Northwest National Laboratory, Richland, WA 99352, USA

<sup>l</sup>Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA

<sup>m</sup>Department of Physics, Engineering Physics, and Astronomy, Queen's University, Kingston, Ontario, K7L 3N6, Canada

<sup>n</sup>Department of Physics, Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK

<sup>o</sup>SNOLAB Institute, Lively, ON P3Y 1N2, Canada

<sup>p</sup>Department of Physics, University of South Dakota, Vermillion, SD 57069, USA

<sup>q</sup>Physics Department, Syracuse University, Syracuse, NY 13244, USA

<sup>r</sup>Triangle Universities Nuclear Laboratory, Durham, NC 27708, USA

<sup>s</sup>TRIUMF, Vancouver, British Columbia, V6T 2A3, Canada

<sup>t</sup>Department of Physics, Yale University, New Haven, CT 06520, USA

## Abstract

Many current and future dark matter and neutrino detectors are designed to measure scintillation light with a large array of photomultiplier tubes (PMTs). The energy resolution and particle identification capabilities of these detectors depend in part on the ability to accurately identify individual photoelectrons in PMT waveforms despite large variability in pulse amplitudes and pulse pileup. We describe a Bayesian technique that can identify the times of individual photoelectrons in a sampled PMT waveform without deconvolution, even when pileup is present. To demonstrate the technique, we apply it to the general problem of particle identification in single-phase liquid argon dark matter detectors. Using the output of the Bayesian photoelectron counting algorithm described in this paper, we construct several test statistics for rejection of backgrounds for dark matter searches in argon. Compared to simpler methods based on either observed charge or peak finding, the photoelectron counting technique improves both energy resolution and particle identification of low energy events in calibration data from the DEAP-1 detector and simulation of the larger MiniCLEAN dark matter detector.

**Keywords:** dark matter, neutrino, pulse-shape discrimination, liquid argon

## 1. Introduction

Scintillators are a key component of many neutrino and dark matter experiments due to their relatively low cost and high light yield. In addition, many scintillators can be used for particle identification when the time pro-

\*Corresponding author.

Email address: tcald@hep.upenn.edu (T. Caldwell)

file of the scintillation light is sensitive to the energy loss characteristics of different particles. Experiments that require low energy thresholds, high energy resolution, and/or high levels of background rejection often combine a scintillating target medium with an array of photomultiplier tubes (PMTs) and a data acquisition system with sensitivity to individual photoelectrons. In order to fully capture the scintillation time profile, the PMT pulses are typically recorded using a waveform digitizer.

Many common scintillators produce light on at least two characteristic time scales. Particle identification in such scintillation detectors employs two related features: the scintillation time scales, and the probability of populating the different time scales, which depends on the particle's energy loss characteristics. The canonical approach (examples include MicroCLEAN [1], DEAP-1 [2], XMASS [3], XENON-10 [4], GERDA [5], KamLAND [6]) to time-based particle identification with a digitized time-dependent voltage waveform  $V(t)$  is to estimate the fraction of the light produced on a fast timescale relative to the total amount of light produced in the event. This particle discriminant we refer to as the prompt-fraction, or  $f_p$ , and is defined as

$$f_p = \frac{\int_{T_i}^{\epsilon} V(t) dt}{\int_{T_i}^{T_f} V(t) dt}, \quad (1)$$

where  $T_i$  is some time before the prompt peak,  $T_f$  is the time defined by the end of the event window, and  $\epsilon$  depends on the timing characteristics of the scintillator. Typically the  $f_p$  parameter is used to place a cut or perform some likelihood-based analysis in order to select a certain class of interactions in the scintillator. In later sections, we refer to  $f_p$  *leakage* as the probability of events from a certain class of background leaking into the  $f_p$  signal region of interest. Although particle identification with  $f_p$  is robust in the sense that it is fairly insensitive to fluctuations in the scintillation light production or PMTs, its discrimination power breaks down at low energies beyond statistical effects because it loses information about the precise timing and charge of individual photoelectrons created by photons produced in the scintillator.

The best strategy to count photoelectron (PE) pulses in a waveform depends on the intensity and time structure of the light, as well as the characteristics of the PMT electronics. A typical single photoelectron pulse from a large area PMT, shown for cryogenic measurements of a Hamamatsu Photonics R5912-02-MOD<sup>1</sup> 8" PMT in Figure 1, spans 20 ns or more [7]. Additionally, there are large

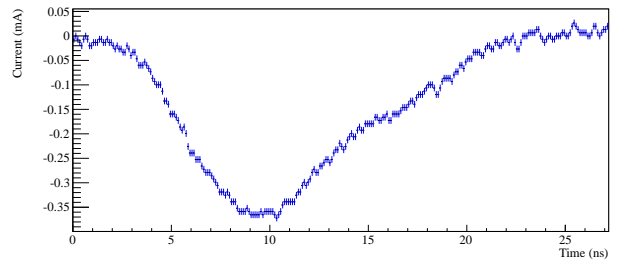


Figure 1: Single photoelectron pulse from a Hamamatsu R5912-02-MOD PMT, as digitized by a 10 GHz oscilloscope.

pulse-to-pulse variations in the amplification of a single photoelectron, resulting in a fairly broad charge distribution, shown again for the R5912-02-MOD in Figure 2. If the light intensity observed by a PMT is very low or the time constant for scintillation light is very long (hundreds of nanoseconds or more) compared to the duration of a single photoelectron pulse, then overlap of pulses is improbable. Photoelectrons can be counted by searching for peaks in the waveform, thus eliminating the impact of the PMT charge distribution on the counting procedure.

However, if multiple photoelectrons are likely to be detected by a PMT in a time much shorter than the single photoelectron pulse duration, such as from Cherenkov or fast scintillation light, peak finding is a poor photoelectron counting strategy. Peaks from different photoelectrons may not be clearly resolved in the waveform, resulting in a systematic bias toward under-counting. A simple and unbiased technique for photoelectron counting in this case would be to integrate the waveform and divide by the mean charge of a single photoelectron. Due to the broad charge distribution of most PMTs, this normalized integral charge procedure in a waveform with pulse pileup will have more variance than peak counting in a waveform without pileup. Fundamentally, information is lost in the pileup that is difficult to recover.

Realistic detectors typically span both of these extreme cases. Many experiments observe both Cherenkov or fast scintillation *and* slow scintillation light. Moreover, the light intensity observed by a given PMT can vary dramatically depending on the energy of the event and its location in the detector. To cover all these cases, we have designed a photoelectron counting method that combines both peak finding and normalized charge integration by using Bayes' Theorem to incorporate our external knowledge of how likely photoelectron pileup is for different events, PMTs, and times in the waveform. The method also outputs an estimated photoelectron production time for each pulse, which can be analyzed for particle identification purposes in many scintillators.

To make the discussion concrete, we focus the photoelectron identification procedure specifically on liquid ar-

<sup>1</sup>Certain commercial equipment, instruments, or materials are identified in this report in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

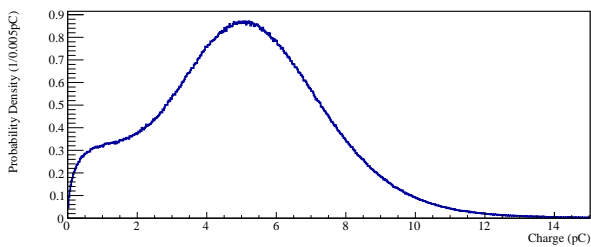


Figure 2: The single photoelectron charge distribution of the Hamamatsu Photonics R5912-02-MOD cryogenic 8” PMT at a gain of  $3 \times 10^7$ . As described in [7], the charge distribution is populated using only pulses which are adequate fits to a pulse shape model, so no assumptions about the shape of the ‘pedestal’ near zero are required.

gon scintillation light, primarily in the MiniCLEAN dark matter experiment, but the general approach can be easily adapted to the constraints of other experiments. In Section 3 we use a GPU-based fast Monte Carlo simulation of the MiniCLEAN detector to motivate the need for improvement in the canonical prompt-fraction particle identification technique. The Bayesian photoelectron counting method is described in Section 4, and improved particle identification test statistics are defined in Section 5. We then demonstrate the effectiveness of the Bayesian photoelectron counting method in Section 6 with gamma simulation and calibration data collected from the DEAP-1 detector underground at SNOLAB. In Section 7, we describe the MiniCLEAN dark matter experiment and motivate the need for improved particle identification specifically in large liquid argon dark matter detectors. Finally, Section 8 applies the Bayesian techniques to a complete Monte Carlo simulation of MiniCLEAN.

## 2. Single-Phase Dark Matter Detection with Liquid Argon

Laboratory searches for dark matter in the form of weakly-interacting massive particles (WIMPs) require sensitivity to nuclear recoils with tens of keV of kinetic energy. WIMP-induced recoils need to be distinguished from other sources of low energy particle tracks, such as fast neutrons and radioactive decay in detector materials. Dark matter experiments must simultaneously achieve very low levels of natural radioactivity and very high rejection factors for remaining sources of backgrounds.

Liquid argon offers a promising target for WIMP detection due to its combination of extremely low cost, moderate cryogenic requirements (similar to liquid nitrogen), straightforward purification techniques, and a reasonably high density and atomic number. Argon is a very efficient scintillator, producing approximately 40 UV photons per keV [8] of deposited energy from an electron track. Although very low levels of chemical impurities can be

achieved with liquid argon, atmospheric argon contains 1 Bq/kg of  $^{39}\text{Ar}$  [9], a beta decay isotope with an endpoint energy of 565 keV and half-life of 269 years. The  $^{39}\text{Ar}$  background must be mitigated either through the acquisition of argon depleted of  $^{39}\text{Ar}$ , through highly efficient identification and rejection of electronic recoils in the data, or both.

“Single-phase” liquid argon experiments search for nuclear recoil events using only scintillation light, in the absence of an electric field [10]. Prototype detectors, such as MicroCLEAN [1, 11] and DEAP-1 [2, 12], have demonstrated the performance characteristics of liquid argon scintillation detectors at the several kilogram scale. The larger scale dark matter experiments, MiniCLEAN [13, 14] and DEAP-3600 [15], are currently under construction with fiducial masses of 150 kg and 1000 kg, respectively. Given the exponential energy distribution of nuclear recoils from WIMPs, the ultimate sensitivity of these experiments will depend in part on how effectively  $^{39}\text{Ar}$  events can be rejected at energies of tens of keV or lower.

Argon scintillation is produced by the de-excitation of dimer states created by the ionization track of a particle [16]. There is both a short-lived singlet state,  $\tau_s$  with a lifetime of  $7.0 \pm 1.0$  ns, and a longer-lived triplet state,  $\tau_t$  with a lifetime of  $1600 \pm 100$  ns [17]. The relative amounts of singlet and triplet states produced by a charged particle depend on both the type of recoiling particle and the energy deposited, as measured in [1]. This, along with the large separation of the singlet and triplet time constants, offers excellent particle identification in liquid argon as described in [10]. Using a full simulation of the MiniCLEAN detector (see sections 6.2 and 8) with the above liquid argon scintillation characteristics, we have optimized<sup>2</sup> the integration bounds in Equation 1, and we take  $T_i = -28$  ns,  $\epsilon = 80$  ns, and  $T_f = 15360$  ns in the following sections.

The scintillation light yield of nuclear recoils in liquid argon at energies above 20 keV is  $25 \pm 1\%$  of the light yield for electronic recoils [11]. We refer to energies of all types of recoils in units of *electron equivalent* keV (keVee), where this 25% quenching factor has already been applied to energies for nuclear recoils. When referring to the full recoil energy, without quenching, we use units of keVr.

## 3. Sources of Background Leakage with $f_p$

In this section, we highlight the mechanisms in scintillator detectors which contribute to  $f_p$  leakage to motivate the potential improvement in a more sophisticated approach. In order to treat a wide variety of detector contributions to  $f_p$  leakage in a consistent manner, we take a predominantly Monte Carlo approach, assuming no functional form for any of the  $f_p$  distributions, except at the

<sup>2</sup>The  $f_p$  integral bound optimization minimizes the number of electronic recoils leaking into the 50% nuclear recoil acceptance region in MiniCLEAN Monte-Carlo.

stage of the initial statistical fluctuations in the number of states produced.

As shown in the following, realistic detector effects can degrade the achievable  $f_p$  discrimination between background and signal by many orders of magnitude relative to the fundamental limit, which is set by the statistical fluctuations in the scintillation states produced. For the case considered here, 12.5 keV apparent energy liquid argon scintillation events in MiniCLEAN, the  $f_p$  leakage fraction with 50% signal acceptance is increased from  $3.9 \times 10^{-11}$  with only fluctuations in the scintillation states produced to  $4.2 \times 10^{-6}$  once all realistic detector effects are included. In this section, we describe each of the factors that contribute to this lost rejection efficiency. In the following sections, we develop a new method that regains some of the lost rejection.

For the results in this section, we use a GPU-based fast Monte Carlo simulation to sample and fluctuate the properties of individual photoelectrons from thousands of events in parallel. This highly parallel approach allows us to reliably sample the tails of the  $f_p$  distributions over many orders of magnitude, without having to rely on extrapolation of a phenomenological model. Detector optics and full simulation of the data acquisition system are bypassed in this simplified model. In Sections 6 and 8, we will present results from our full detector simulations of DEAP-1 and MiniCLEAN, that use a more complete model including detector optics and DAQ.

We consider a general two-exponential scintillation model that is equivalent to, but parametrized slightly differently than, that presented in [1]. The assumed probability density function for the production times of argon scintillation photons is

$$P(t) = f \left( \frac{1}{\tau_s} e^{-t/\tau_s} \right) + (1 - f) \left( \frac{1}{\tau_l} e^{-t/\tau_l} \right), \quad (2)$$

where  $f$  is the fraction of states produced with the short-lived time constant<sup>3</sup>,  $\tau_s$ , and the remaining states are produced with the longer lived time constant,  $\tau_l$ . Generally all three of these parameters depend on the energy and type of the scintillating particle. For concreteness, we focus on scintillation in liquid argon detectors. However, this model and conclusions drawn from it are applicable to a wide range of scintillators.

For further concreteness in the fast Monte Carlo model, we assume the properties of the MiniCLEAN dark matter detector. The detector is described in detail in Section 7, but its relevant scintillation characteristics are summarized in Table 1. For the purposes of the fast Monte Carlo simulation, the MiniCLEAN detector is considered to be a 500 kg liquid argon target surrounded by 92 optical cassettes. Each optical cassette houses a Hamamatsu Photonics 8" R5912-02-MOD PMT and a 10 cm thick acrylic plug.

<sup>3</sup>Although  $f$  is sometimes called the ‘‘prompt-fraction,’’ it is not equal to the test statistic  $f_p$ .

Parameter	Value
Light yield	6 PE/keVee
Number of photoelectrons	75 PE
Nuclear quenching factor	0.25 keVee/keVr
$e^-$ singlet fraction ( $f$ )	0.308
Ar recoil singlet fraction ( $f$ )	0.678
Singlet time const. ( $\tau_s$ )	7.0 ns
Triplet time const. ( $\tau_l$ )	1600 ns
PMT gain	$3 \times 10^7$
PMT single PE mean charge	5 pC
PMT timing resolution	1.5 ns
PMT dark hit rate	1.5 kHz
Digitizer sample size	4 ns
Digitizer noise RMS	0.41 mV

Table 1: Basic scintillation and detector parameters for 12.5 keVee events in the MiniCLEAN detector.

The acrylic plug acts as a substrate for the TPB layer that converts the extreme UV argon scintillation light [18, 16] to a wavelength regime where the PMTs are sensitive.

Due to the presence of  $^{39}\text{Ar}$ , a beta decay isotope with 565 keVee endpoint, MiniCLEAN will observe  $10^{10}$   $^{39}\text{Ar}$  decays over all energies in a year of running with atmospheric argon, so we focus on this as the primary particle identification background to the signal of interest, WIMP induced nuclear recoils. For the discussion of the fast Monte Carlo simulation, we will further restrict the discussion to 75 PE (12.5 keVee) apparent energy events in MiniCLEAN which is meant as a feasible but aggressive energy threshold for a WIMP dark matter search. Refer to Section 7 for more details on the MiniCLEAN detector.

Leakage is defined to be the fraction of background events with an  $f_p$  test statistic greater than the median  $f_p$  test statistic for signal. This yields an approximate 50% signal acceptance probability. Except where noted, each leakage fraction is estimated using the very high statistics fast Monte Carlo simulation, including only the detector effects of interest. This allows trillions of events to be generated in several hours and avoids the need to extrapolate an analytic function to estimate very low leakage levels. Unless otherwise specified, all leakage fractions have been computed with sufficient statistics to reduce their statistical uncertainty to less than 20%. We will use the complete optical simulation to evaluate background leakage in Section 8.

In the remainder of this section, we identify various effects that increase the  $f_p$  leakage fraction. The simple estimate of leakage from binomial fluctuations at 75 PE (12.5 keVee) apparent energy would suggest background leakage as low as  $3.9 \times 10^{-11}$ . However, including the following realistic detector characteristics, we find that with the simple  $f_p$  test statistic one generally can only do as well as  $4.2 \times 10^{-6}$  at 12.5 keVee with 6 PE/keV detected scintillation light yield. Samples of the  $f_p$  distributions

obtained for electronic and nuclear recoils, including only those effects described below up to Section 3.4 are shown in Figure 5.

### 3.1. Binomial Fluctuations

If the detected photoelectrons could be perfectly identified as coming from the de-excitation of either the singlet or the triplet states, then the only uncertainty in particle identification would come from the Poisson fluctuations in the production and detection of photons from each dimer state. At a fixed number of photoelectrons, these Poisson fluctuations in the detection of the two states appear as binomial fluctuations in the fraction of photoelectrons coming from the singlet state. At this stage, with perfect identification of the scintillation states, we estimate  $f_p$  using  $f$  of Equation 2. With such a simple model,  $f_p$  is not a continuous distribution, so 50% acceptance of nuclear recoils can only be approximated. The numerically-estimated leakage purely from binomial fluctuations at 75 PE (12.5 keVee) is  $3.9 \times 10^{-11}$ , with a nuclear recoil acceptance of 46%. This represents a fundamental lower bound on the leakage, regardless of analysis technique, and can only be improved by increasing the overall light yield of the experiment.

### 3.2. Scintillation time profile

The originating dimer states that produce each photoelectron cannot be identified perfectly because both states generate photons with overlapping exponential time distributions. Argon has excellent particle discrimination properties because the time constants are well separated for the two dimer states, but the inherent ambiguity does increase the leakage to  $2.9 \times 10^{-10}$ .

### 3.3. $^{39}\text{Ar}$ Spectrum and Detector Energy Resolution

The overwhelming source of electron-like backgrounds in a large, well shielded liquid argon detector is the beta decay of  $^{39}\text{Ar}$ . The finite energy resolution of a real detector allows a distribution of  $^{39}\text{Ar}$  beta energies to be detected with a fixed number of photoelectrons. Figure 3 shows the true energy distribution for  $^{39}\text{Ar}$  events with an estimated energy of 12.5 keVee in the MiniCLEAN fast Monte Carlo simulation.

The true energy of recoils does not directly impact the statistical fluctuations in the fraction of detected singlet photons, which depends only on the number of photoelectrons. However, the scintillation properties of argon are energy-dependent, so the energy resolution does directly impact the distribution of singlet photons produced. Figure 3 shows the energy dependence of the singlet fraction as a function of energy [1] with the energy distribution at 75 PE overlaid. As a result, events with a fixed number of photoelectrons come from a range of expected singlet fractions. The leakage probability for  $^{39}\text{Ar}$  events with a given number of photoelectrons is dominated by those events

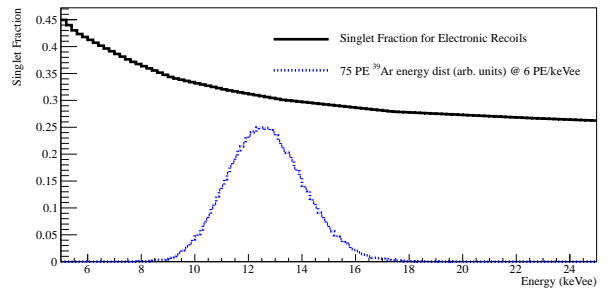


Figure 3: Energy dependence of the singlet fraction in electronic recoils [1] overlaid with the energy spectrum (in arbitrary units) of 75 PE events from  $^{39}\text{Ar}$ , assuming a detector detection efficiency of 6 PE/keVee.

whose *true* energy is lower than average, corresponding to singlet fractions closer to the 50% acceptance cut.

This energy-dependence of the scintillation light makes the leakage probability depend on both the number of photoelectrons in the event *and* the light yield of the detector. Improving the light yield of the detector and holding the energy threshold fixed, thus lowering the energy threshold in keVee, will reduce the background leakage due to the combined effect of higher photoelectron statistics and narrower energy resolution. Conversely, improving the light yield and holding the energy threshold fixed in number of photoelectrons will slightly increase the background leakage probability due to the energy dependence of the singlet fraction.

For 75 PE events in the MiniCLEAN detector with a light yield of 6 PE/keVee, the finite energy resolution of the detector increases the estimated leakage fraction to  $2.3 \times 10^{-9}$ .

### 3.4. PMT Charge and Time Response

Photomultiplier tubes can produce a range of different pulse sizes given a single photoelectron. This inherent charge resolution has a two-fold effect on background leakage. First, it introduces some randomness in the determination of the number of photoelectrons in an event, which must now be defined as the summed charge from all PMTs divided by the PMT gain. This effectively broadens the energy resolution of the detector, as shown in Figure 4. The second effect of the charge resolution is to introduce fluctuations in the apparent number of singlet and triplet states beyond the binomial fluctuations inherent in the underlying physics. Several photoelectrons can individually fluctuate to each have two or three times the mean charge, which broadens the tails of the  $f_p$  distribution for both nuclear recoils and electrons.

The time response of a PMT also influences  $f_p$  by smearing out the time structure of the scintillation light. The timing jitter of the R5912-02-MOD is approximately 1.5 ns, but measurements indicate that a photoelectron



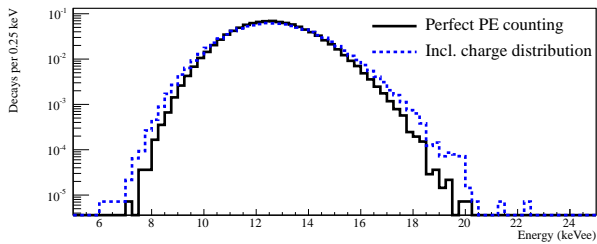


Figure 4: Energy distribution of  $^{39}\text{Ar}$  events falling in the 75 PE bin (12.5 keVee) before and after the R5912-02-MOD charge distribution is included.

also has a 4.7% chance of producing a “late” pulse, delayed by up to 60 ns [7]. In addition, the finite width of a PMT pulse, as shown in Figure 1, makes it possible for the prompt time window in the  $f_p$  estimation to only partially contain the charge from the pulse. The integral in the numerator of Equation 1 is relatively insensitive to these effects, although they do contribute in small part to the overall leakage.

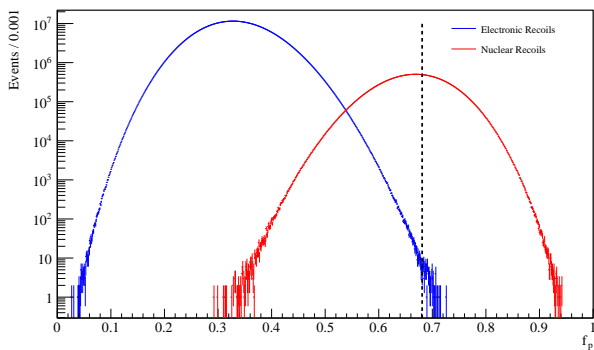


Figure 5: Sample  $f_p$  distributions for electronic (blue) and nuclear (red) recoils including the effects up to Section 3.4 (color version online). The dashed vertical line indicates 50% nuclear recoil acceptance.

To simulate these effects in the fast Monte Carlo simulation, we sampled charges from the full charge distribution, shown in Figure 4, and a simplified Gaussian timing distribution with a RMS of 1.5 ns that ignores the contributions of late pulsing. For speed reasons, the effect of PMT pulses partially extending beyond the prompt window was estimated by approximating the pulse shape with a Gaussian function. The  $f_p$  leakage of 75 PE electrons increased to  $7.1 \times 10^{-8}$ .

### 3.5. PMT Dark Current

Thermionic emission of individual electrons inside the PMT from the photocathode and dynode structure produces a “dark current” that can affect  $f_p$ . In the MiniCLEAN detector, the 92 PMTs each produce dark pulses

at a rate of approximately 1500 Hz at their operating temperature of 87K. Therefore, dark pulses will be uniformly distributed over the 16  $\mu\text{s}$  event window, with an average of 2.2 dark pulses per event. In order to compensate for this offset in the energy scale, we compute the leakage for 77 PE (instead of 75) when dark pulses are included. The general effect of dark pulses is to further broaden the energy resolution of the detector, which worsens the leakage for the same reasons discussed in Section 3.3. In addition, dark hits are most likely to occur outside the prompt time region which biases the  $f_p$  of both electronic and nuclear recoils to smaller values, but biases the nuclear recoils more, bringing the two distributions closer together. As a result, we find that the inclusion of a Poisson distribution of dark hits increases the  $^{39}\text{Ar}$  leakage fraction at 75 PE to  $1.2 \times 10^{-7}$ .

### 3.6. Digitization Noise

The sampling of the PMT signals introduces an additional source of noise into the calculation of the integrals in  $f_p$ . This noise can come from pickup of analog signals in the PMT cabling to the digitizers or intrinsic noise in the analog-to-digital converters. The precise impact of digitization noise on  $f_p$  depends on the amplitude and frequency composition. In the fast Monte Carlo simulation, we have introduced this noise as a per-sample Gaussian random variable with a RMS of 0.41 mV, consistent with measurements of the CAEN V1720 digitizers used by MiniCLEAN [14], DEAP-1 [12], and DEAP-3600 [15]. When zero suppression is enabled in the digitizer readout, the largest contribution of noise to  $f_p$  will occur when the number of samples is maximized by separation of photoelectron pulses in time and in different PMT channels. In the MiniCLEAN detector, with 92 PMTs, this separation of signals in time and PMT channel is quite probable for 75 PE events inside the fiducial volume of the detector. In this case, each photoelectron pulse is recorded in a separate block of 32 samples, which could contribute an additional Gaussian smearing per photoelectron of up to 3.7% of the mean single photoelectron charge at 5 pC gain. We find that this increases the leakage to  $4.2 \times 10^{-6}$  in the fast Monte Carlo simulation.

### 3.7. Leakage Summary

Table 2 shows the increasing leakage from an  $f_p$  cut as more detector effects are included in the fast Monte Carlo model. Clearly, energy resolution is a significant factor in the leakage due to the increased singlet fraction at lower event energies. The detector energy resolution comes from a mixture of photon counting statistics, which can only be improved by increasing the light yield of the detector, and the PMT and waveform digitizer response, which can be improved with better waveform analysis techniques, as shown in the next section.

Detector Effect	$f_p$ leakage
Binomial Fluctuations	$3.9 \times 10^{-11}$
Scintillation Time Profile	$2.9 \times 10^{-10}$
$^{39}\text{Ar}$ Distrib. + Energy Res.	$2.3 \times 10^{-9}$
PMT Response	$7.1 \times 10^{-8}$
PMT Dark Current	$1.2 \times 10^{-7}$
Digitization Noise	$4.2 \times 10^{-6}$

Table 2: Leakage of  $f_p$  for 12.5 keV apparent energy events only in a detector with 6 PE/keV (assuming 50% nuclear recoil acceptance) as different detector effects are added. Each row includes all of the effects above it, and the assumptions for each row are described in detail in the text. The optimal value of  $\epsilon$ , the length of the prompt window in  $f_p$ , is between 60 and 70 ns. The uncertainty for all leakages is 20%.

#### 4. Bayesian Photoelectron Counting

Several of the detector effects described in the previous section that increase the  $f_p$  leakage fraction can be mitigated with an analysis that can extract individual photoelectron times from the PMT waveforms. However, rather than solve a general deconvolution problem, we have instead developed an analysis that assumes the PMT waveform is produced by argon scintillation from a single recoiling particle. This prior can be incorporated into Bayes' Theorem in order to estimate the number of photoelectrons in the waveform, and then assign times to each of the photoelectrons based on the waveform shape.

The event analysis algorithm has 5 stages:

1. Time/voltage calibration of waveforms.
2. Identification of regions that contain pulses.
3. Bayesian identification of photoelectrons.
4. Reconstruction of the event position and energy.
5. Repeat of Bayesian identification of photoelectrons with improved priors.

These steps are explained in each of the following subsections.

##### 4.1. Waveform Calibration

With the PMT waveforms in a zero-suppressed voltage-time series, the absolute time offset of the waveforms in the event are adjusted so that the summed waveform reaches a maximum amplitude at  $t = 0$ . Due to the shortness of the singlet time constant, aligning based on the peak removes most of the jitter caused by the latching of the trigger, which can often be several digitizer samples or more.

The voltage of each PMT waveform is separately corrected to remove any baseline offset. For this first stage of analysis, the first 4 pre-samples from each block of samples in a given channel are averaged to estimate a constant baseline that is subtracted from all the blocks in the channel. After voltage calibration,  $f_p$  can also be calculated for

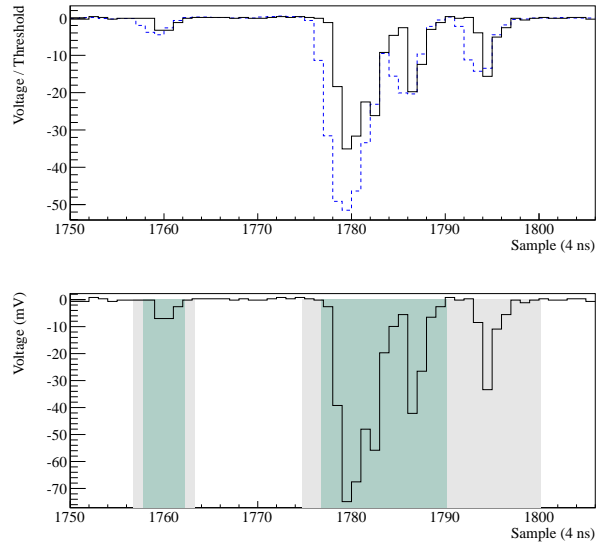


Figure 6: A typical voltage waveform from a single PMT in MiniCLEAN Monte Carlo simulation. The top panel shows the waveform normalized by 5 times the RMS of the electronics noise profile (black, solid) compared to the sliding integral value normalized by the corresponding threshold (blue, dashed). The sliding integration window enhances the right-skew PMT pulses relative to threshold while providing a filter for high frequency electronics noise. The bottom panel shows the pulse regions identified by the pulse finder. Green shaded regions are the regions where threshold is crossed, and the gray regions indicate a buffer region that extends the pulse boundaries. If threshold is crossed again within the buffer, the pulse boundary is further extended as in the right most pulse region.

the summed PMT waveform, as it is used in the Bayesian photoelectron identification stage.

##### 4.2. Pulse Finding

Single photoelectron pulses from the 8" R5912-02-MOD PMTs typically span 20 ns, as shown in Figure 1. We scan the calibrated waveforms for each PMT separately with a sliding 12 ns (3 sample) integration window and extract a *pulse region* whenever the integral exceeds 5 times the RMS of noise samples times the square root of the number of samples in the window. Once that detection threshold has been crossed, the boundaries of the pulse region are the times where the sliding window integral drops to below the RMS divided by the square root of the number of samples. Figure 6 shows the sliding window calculation applied to a pulse from 5 photoelectrons.

We find that the sliding window integration method is much more sensitive to low charge PMT pulses without significantly increasing the number of false positive detections. The width and right skew of the pulses make it difficult to efficiently discriminate low charge pulses from

background noise with a simple voltage threshold. An integration window, however, effectively filters out the high frequency noise while retaining the relatively low frequency PMT signal.

Once the samples corresponding to pulses have been identified, they are removed from the waveform, and the remaining baseline-only samples are used to estimate a local baseline for each block of samples in the original waveform. Unlike in the calibration stage, the baseline computed in this stage is allowed to vary with time, thus removing any lower frequency components, for example under- or over-shoot in PMT base electronics. The new baseline is subtracted from the waveform and the entire pulse finding procedure is repeated. After two passes, the results are stable and robust against low-frequency baseline variations.

#### 4.3. Bayesian Photoelectron Identification: First Pass

Once pulse regions have been extracted from each PMT waveform, the next task is to identify how many separate photoelectron pulses are contained in each pulse region, and what their arrival times were. As discussed in Section 1, two natural techniques to apply are normalized integral charge, where the integral of each pulse region is divided by the average single photoelectron charge, or peak counting, where the number of local maxima in the pulse region are counted after some filtering. Charge integration is a good strategy to apply at early times, where pileup will likely make multiple peaks indistinguishable, and a poor strategy at late times, where a pulse region will almost always contain a single photoelectron, but the variance in the single photoelectron charge distribution will create large fluctuations in the estimate. Conversely, peak counting is a poor strategy at early times, where it will generally undercount photoelectrons, but a good strategy at late times, where it is insensitive to charge fluctuations in single photoelectrons.

Bayes' Theorem provides a quantitative way to incorporate our knowledge of the time structure of argon scintillation light. For each pulse identified in each PMT waveform we would like to know the most probable number of photoelectrons in the pulse,  $n$ , given that the pulse spans time  $t_1$  to  $t_2$  (where  $t_1$  and  $t_2$  are defined by the pulse finding procedure in Section 4.2) and has an integrated charge of  $q$ . The probability mass function for  $n$  is

$$\begin{aligned} P_N(n|q, t_1, t_2) &= \frac{P_Q(q|n)P_N(n|t_1, t_2)}{P_Q(q|t_1, t_2)} \\ &= \frac{P_Q(q|n)P_N(n|t_1, t_2)}{\sum_{i=0}^{\infty} P_Q(q|i)P_N(i|t_1, t_2)}, \end{aligned} \quad (3)$$

where  $P_N$  is used to denote a probability mass function for the number of photoelectrons, and  $P_Q$  denotes a probability density function for integrated charge.

The function  $P_Q(q|n)$  is the  $n$ -photoelectron charge distribution for the PMT, which is assumed to have no

dependence on the time of the pulse. The charge distribution of noise,  $P_Q(q|n=0)$ , can be found by applying the pulse finding algorithm to a sample of waveforms with electronics noise only. The single photoelectron charge distribution,  $P_Q(q|n=1)$ , can be measured for each PMT using a triggered, low intensity light source, as in Figure 2. For  $n > 1$ ,  $P_Q(q|n)$  is the convolution of  $P_Q(q|n=1)$  with itself  $n$  times.

Our knowledge of the distribution of scintillation photons (and therefore the probability of pileup) enters into the equation via  $P_N(n|t_1, t_2)$ , the probability mass distribution for the number of photons in the given pulse. Suppose that events of a particular class produce  $\mu$  detected photoelectrons in the PMT on average with a photoelectron detection time probability density function (PDF) of  $S(t)$ . The probability mass function of observing  $n$  photoelectrons in the time interval  $[t_1, t_2]$  is then

$$P_N(n|t_1, t_2) = \sum_{j=0}^{\infty} \text{Pois}(j|\mu) \times \text{Bin}(n|j, I_j), \quad (4)$$

where  $\text{Pois}(j|\mu)$  is the Poisson probability of observing  $j$  photoelectrons in the entire waveform given the expected value  $\mu$ ,  $\text{Bin}(n|j, I_j)$  is the binomial probability of detecting  $n$  photoelectrons given  $j$  photoelectrons in the waveform, and the probability,  $I_j$ , of  $j$  photoelectrons falling in the time interval  $[t_1, t_2]$  is defined by

$$I_j = \int_{t_{j,1}}^{t_{j,2}} S(t) dt. \quad (5)$$

The PMT timing response (including dark hits, double pulsing, late pulsing, etc [7]) and the effects of detector optics are convolved with the scintillation time structure, Equation 2, to include realistic detector effects in the photoelectron detection time PDF,  $S(t)$ . Figure 7 shows sample time PDFs for nuclear and electronic recoils at energies of 5 and 25 keVee in the MiniCLEAN Monte Carlo simulation (described in detail in Section 7).

The integration bounds of Equation 5,  $t_{j,1}$  and  $t_{j,2}$ , depend on the hypothesized number of photoelectrons,  $j$ , due to finite width of the PMT pulses. For the hypothesis of a single photoelectron spanning the pulse bounds,  $t_1$  to  $t_2$ , the known pulse shape can be used to narrow the integration bounds in Equation 5 to the time interval in which the scintillation photon may have been produced,  $[t_{1,1}, t_{1,2}]$ . However, in the limit of many photoelectrons spanning the same time interval, the scintillation photons may have been produced over the entire range of the pulse,  $[t_1, t_2]$ . In practice, we compute the cumulative distribution function of the time PDF,  $S(t)$ , and determine the 1<sup>st</sup> and  $j^{\text{th}}$  ( $j+1$ )-quantiles,  $q_1$  and  $q_j$  respectively. Letting  $\tau$  represent the finite width of the digitization sampling, we then take  $t_{j,1} = q_1 - \tau/2$  and  $t_{j,2} = q_j + \tau/2$ . This gives a robust estimation of the correct integration boundaries for  $j$  photoelectrons without requiring exact knowledge of the PMT pulse shape with  $j$  photoelectrons piling up with



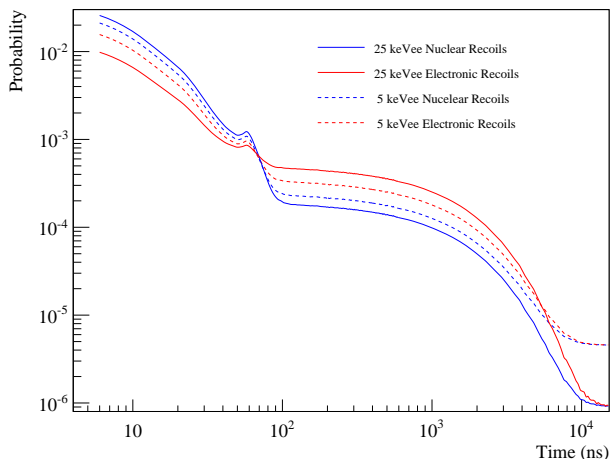


Figure 7: Photoelectron detection time PDFs for electronic and nuclear recoils at 5 keVee and 25 keVee observed energies from MiniCLEAN Monte-Carlo simulation (described in detail in Section 7). The peak near 60 ns is due to PMT double and late pulsing [7]. The energy dependence of the mean triplet fraction [1] is included in construction of the PDFs. The flat component in time, due to PMT dark hits, contributes a larger fraction to the signal at low energies resulting in convergence to a higher probability at late times. In the Bayesian photoelectron counting procedure, the PDFs are constructed by linear combination of the singlet and triplet components with a bootstrapped prior to using the  $f_p$  test statistic.

random time offsets. By inserting Equation 5 into Equation 4 and Equation 4 into Equation 3,  $P_N(n|q, t_1, t_2)$  can be evaluated.

The selection of  $\mu$  and  $S(t)$  for each PMT (see Figure 7) depends on our hypothesis of the particle type, position and energy, but estimating those quantities in turn requires the results of photoelectron counting. In order to bootstrap the process, we perform this first pass of the analysis assuming that  $\mu$  for each PMT is equal to the total integrated charge observed by the PMT divided by the average single photoelectron charge. The time distribution for photoelectrons is estimated by using the prompt-fraction test statistic,  $f_p$ , which is computed using all the PMT waveforms. Separate time PDFs for the singlet and triplet photons generated with the Monte Carlo simulation are linearly combined to create  $S(t)$  according to the singlet fraction calculated from  $f_p$ . A flat time distribution for PMT dark hits is then added into  $S(t)$  based on the relative magnitudes of the expected number of PMT dark hits and  $\mu$ . As described in Section 4.1,  $t = 0$  for each event is defined as the peak in the waveform summed over all PMTs in order to be consistent with the definition of  $S(0)$  as the maximum of the time PDF.

With these bootstrap priors, the most probable number

of photoelectrons in each pulse region can be estimated as the integer  $n$  which has maximal  $P_N(n|q, t_1, t_2)$  from Equation 3. In order to assign times to the photoelectrons, the cumulative distribution function (CDF) of the waveform in the region is used. For a pulse region with most probable  $n$  photoelectrons, we determine the  $n + 1$ -quantiles of the model CDF. Each photoelectron is assigned a unique detection time equal to one of these quantiles. Interpolation is used so that the quantiles can be between samples. This procedure is demonstrated in Figure 8 for a pulse region with 14 photoelectrons.

#### 4.4. Position and Energy Reconstruction

Given counts and times for photoelectrons detected by every PMT in the detector, more event properties can be reconstructed. With a radius of 43.7 cm, the MiniCLEAN detector is small enough that time is a weak handle on event position, so only the number of photoelectrons observed by each PMT is used by a maximum likelihood algorithm to estimate the event position and energy. The likelihood function contains a fast, Monte Carlo-derived optical model of the detector (which approximates the detailed optics of the optical cassettes) that can predict the expected number of photoelectrons for each PMT given a hypothesized position and energy. Once the fit has found the most likely event position and energy, a final set of expected numbers of photoelectron are calculated for each PMT and passed to the next stage of the analysis.

#### 4.5. Bayesian Photoelectron Identification: Second Pass

For second pass of the Bayesian photoelectron calculation, we repeat the calculation described in Section 4.3, but with an estimate of  $\mu$  for each PMT derived from the position and energy reconstruction stage. The estimate of the time PDF,  $S(t)$ , is left unchanged, except for the PMT dark hit contribution<sup>4</sup>. By replacing the crude Bayesian prior from the first pass with a much more accurate one based on event reconstruction, the second pass of the algorithm can reduce the energy bias and improve the energy resolution. In principle, additional iterations of position/energy reconstruction and Bayesian photoelectron counting could be performed, but we generally find no significant improvement after 2 passes.

## 5. Test Statistics for Particle Identification

With a set of detected photoelectron times for an event,  $\mathcal{T}$ , we can compute a variety of different test statistics. The simplest test statistic is a discrete version of  $f_p$ ,

$$r_p = \frac{|\{t | t \in \mathcal{T} \wedge T_i < t < \epsilon\}|}{|\{t | t \in \mathcal{T} \wedge T_i < t < T_f\}|}, \quad (6)$$

<sup>4</sup>In a larger detector than MiniCLEAN, a different  $S(t)$  could be computed for each PMT that includes the distribution of photon propagation times for an event at the reconstructed position.

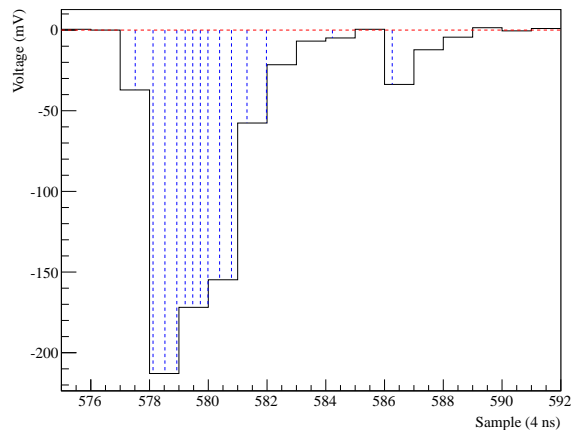
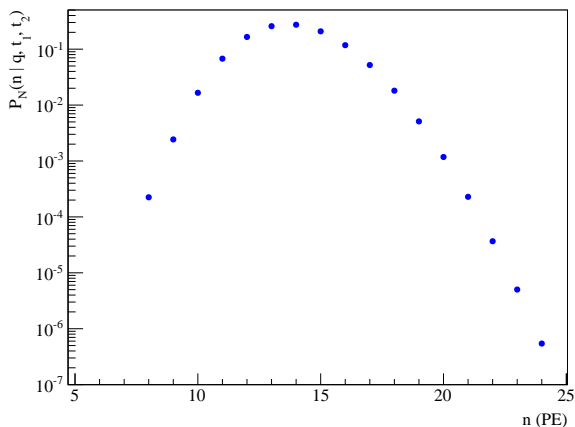


Figure 8: (Left) Sample distribution of  $P_N(n|q, t_1, t_2)$  for the pulse shown in the right panel. The Bayesian photoelectron counting algorithm assigns 14 photoelectrons to this pulse. (Right) The assigned times using the waveform shape are shown by the vertical blue dashed lines.

where  $\epsilon$  sets the prompt window, just as in the definition of  $f_p$ . The  $r_p$  test statistic removes some of the variance induced by the PMT charge distribution on  $f_p$ , but does not take full advantage of the scintillation time structure.

A more powerful approach to separating electronic and nuclear recoils is a likelihood ratio test statistic, as was also briefly described in [1]. By evaluating the likelihood of the observed photon detection times using time PDFs for a nuclear recoil hypothesis and an electronic recoil hypothesis, one can make better use of the full timing information from each photon. We define this second test statistic,  $l_r$ , as a normalized log-likelihood difference, comparing the nuclear recoil hypothesis to the electronic recoil hypothesis. Specifically, we define the test statistic to be:

$$l_r = \frac{1}{m} \sum_{t \in \mathcal{T}} (\log P_n(t|E) - \log P_e(t|E)), \quad (7)$$

where  $m = |\mathcal{T}|$ ,  $P_n$  is the time PDF for the nuclear recoil hypothesis given an event energy  $E$ , and  $P_e$  is the time PDF for the electronic recoil hypothesis given an event energy  $E$ . Sample distributions for  $P_n$  and  $P_e$  are shown in Figure 7 at 5 and 25 keVee. Due to the sign convention adopted in the definition of  $l_r$ , positive values are more nuclear recoil-like and negative values are more electronic recoil-like. The division by  $m$  is a convenience to keep the range of  $l_r$  similar for events with different numbers of photoelectrons. The time PDFs  $P_n$  and  $P_e$  are computed using separate Monte Carlo simulations of singlet and triplet scintillation photons, which are then linearly combined according to the measured energy (including the quenching factor where appropriate) and particle dependence of the singlet fraction.

Finally, it is always important to pair a likelihood ratio with a goodness-of-fit metric to reject events that do not conform to either hypothesis. Given the time PDFs  $P_n$  and  $P_e$  used to compute  $l_r$ , we can compute a Kolmogorov-

Smirnov test statistic for the nuclear recoil ( $K_n$ ) and electron recoil ( $K_e$ ) hypotheses, respectively.

## 6. Bayesian Photoelectron Counting in DEAP-1

To demonstrate the effectiveness of the Bayesian photoelectron counting technique, we apply it here to data collected by the DEAP-1 detector between November 28 and December 9, 2011. DEAP-1 is a cylindrical detector containing 7.6 kg of liquid argon that has been operated underground at SNOLAB since 2007. The detector has had various configurations as an R&D platform as described in [12], but the configuration in November-December 2011 is shown in Figure 9.

The central target volume is defined by a 28 cm long and 15 cm diameter acrylic cylinder that has been coated on the inside with TPB which converts 128 nm extreme UV light [18, 16] into 440 nm visible light [19] that can be transmitted through acrylic light guides which are coupled to PMTs through a layer of mineral oil. The TPB thickness on each acrylic end cap is 0.9  $\mu\text{m}$  and is coupled via a Kodial glass window and acrylic light guide to a high quantum efficiency (HQE) model R5912 8" PMT manufactured by Hamamatsu Photonics. The TPB coating on the inside of the acrylic barrel is 4-5  $\mu\text{m}$  thick, and the outside of the barrel is wrapped with a PTFE reflector.

In this configuration, DEAP-1 digitizes full 16  $\mu\text{s}$  waveforms with no zero-suppression using a CAEN V1720 250 MHz digitizer. Zero-suppression is applied in software to be consistent with the zero-length encoding assumed in the fast Monte Carlo simulation in Section 3 and the full MiniCLEAN simulation in Section 8.

### 6.1. DEAP-1 $^{22}\text{Na}$ Calibration Data

The DEAP-1 data set used for this analysis comes from a tagged, 10  $\mu\text{Ci}$   $^{22}\text{Na}$  radioactive source placed just outside the acrylic vacuum vessel, centered along the axis of

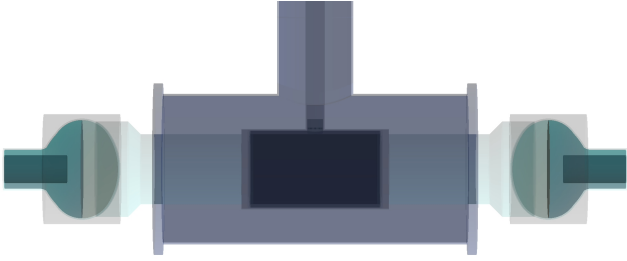


Figure 9: A rendering of the DEAP-1 detector from the GEANT4 simulation. The 7.6 kg liquid argon volume is coupled to acrylic light guides via TPB coated glass windows. Each light guide is coupled through a layer of mineral oil to Hamamatsu Photonics R5912 PMT.

the target volume. The dominant  $^{22}\text{Na}$  decay mode produces a positron that annihilates into back-to-back 511 keV gamma rays followed within ps by a 1274 keV de-excitation gamma ray. A small NaI crystal is placed behind the source to tag one of the 511 keV gamma rays in order to constrain the other 511 keV gamma to be within a  $9^\circ$  cone aimed directly at the center of the argon volume. The  $^{22}\text{Na}$  source geometry and trigger configuration used in this data is very similar to that described in [2], but the trigger only requires a coincidence between the two main PMTs facing the argon and the small PMT attached to the back NaI crystal. The 1274 keV gamma ray is untagged and uncorrelated in direction with the 511 keV gamma ray that enters the DEAP-1 target volume.

The light yield of the detector in this configuration was measured to be 4.5 PE/keV using the full energy peak from the 511 keV calibration gammas, and this is incorporated into the Bayesian counting procedure to include the energy dependence of the mean singlet fraction in liquid argon scintillation. From test bench measurements, for approximately 5% of photoelectrons, the HQE R5912 PMTs in DEAP-1 produce ionic after-pulses over timescales of several  $\mu\text{s}$  after the primary pulse. The time PDFs used in the Bayesian photoelectron counting, and in the calculation of  $l_r$ , include these effects with PDFs generated in the simulation described in the next section. With only 2 PMTs (rather than 92 as in the fast Monte Carlo simulation of MiniCLEAN in Section 3), there is significant pileup of photoelectron pulses which provides a good test of the Bayesian photoelectron counting transitioning from pileup of multiple photoelectrons in the prompt region to isolated pulses at late times as described in Section 4.3.

DEAP-1 is not capable of 3D position reconstruction, but can reconstruct the position ( $z$ ) of the event along the axis of the target cylinder. The  $z$  position is also used to cut background that occurs at the windows where the PMT light guides are coupled to the liquid argon target volume (see [12] for details). Only events reconstructing

within 10 cm of the  $z$  center of the detector are included in the analysis. As a data quality cut and to remove pileup of multiple recoils in the detector, we also apply a cut using the KS test statistics  $K_e$  and  $K_n$  described in Section 5.

The bottom panels of Figure 11 show the distributions of the three test statistics,  $f_p$ ,  $r_p$ , and  $l_r$ , as a function of the number of detected photoelectrons as output by the Bayesian photoelectron counting procedure. Events at low number of photoelectrons and high values of  $f_p$  which could leak into the nuclear recoil region are due in part to the effects described in Section 3. However, an additional class of high  $f_p$  events are also present in the region of 25 to 150 PE. The presence of this class of events is enhanced with the Bayesian photoelectron counting procedure and the  $r_p$  test statistic. Using the  $l_r$  statistic, which takes full advantage of the scintillation time profiles, these events are especially well identified at positive values of  $l_r$ . As described in the next section, simulations suggest that these events are due to the untagged 1274 keV gamma from the  $^{22}\text{Na}$  decays producing Cherenkov light in the large acrylic light guides in coincidence with the 511 keV gamma producing scintillation light in the liquid argon. The fast Cherenkov light produces an additional prompt component which the  $f_p$  test statistic is relatively insensitive to. The separation of these events from the scintillation only events highlights the improved particle identification capabilities of the Bayesian photoelectron counting technique and the  $l_r$  test statistic.

A one-dimensional slice of the distributions at 30 PE, or 6.67 PE/keVee, is shown in Figure 10. For the purposes of comparison, the  $l_r$  values have been linearly transformed to match the median values for electronic and nuclear recoils for  $f_p$ . This keeps a fixed 50% nuclear recoil acceptance for each test statistic. The reduction in the  $l_r$  tail relative to  $f_p$  on the nuclear recoil side shows that  $l_r$  is a superior background rejection tool for low energy events. Integrating above  $f_p$  of 0.7, the Bayesian photoelectron counting and the  $l_r$  test statistic reduce the leakage of 30 PE events into the nuclear recoil 50% acceptance region by a factor of  $7.8 \pm 0.3$  (stat).

As in [2], a well-motivated model can be constructed to extrapolate the tails of the  $f_p$  distribution at a particular energy. However, we lack such a model for the normalized likelihood ratio discriminant  $l_r$ . For this reason, we avoid comparison of the leakage for the particle identification parameters  $f_p$  and  $l_r$  except where there is sufficient statistics in the data to make a direct comparison.

## 6.2. DEAP-1 Simulation

To further test the Bayesian photoelectron counting technique and the test statistics  $f_p$ ,  $r_p$ , and  $l_r$ , we apply the complete procedure described in Section 4 to simulated DEAP-1  $^{22}\text{Na}$  calibration data in this section and to  $^{39}\text{Ar}$  events in MiniCLEAN in Section 8. The simulation and analysis framework used for both analyses, called *RAT*, was originally developed by the Braidwood collaboration [20] and is now maintained cooperatively

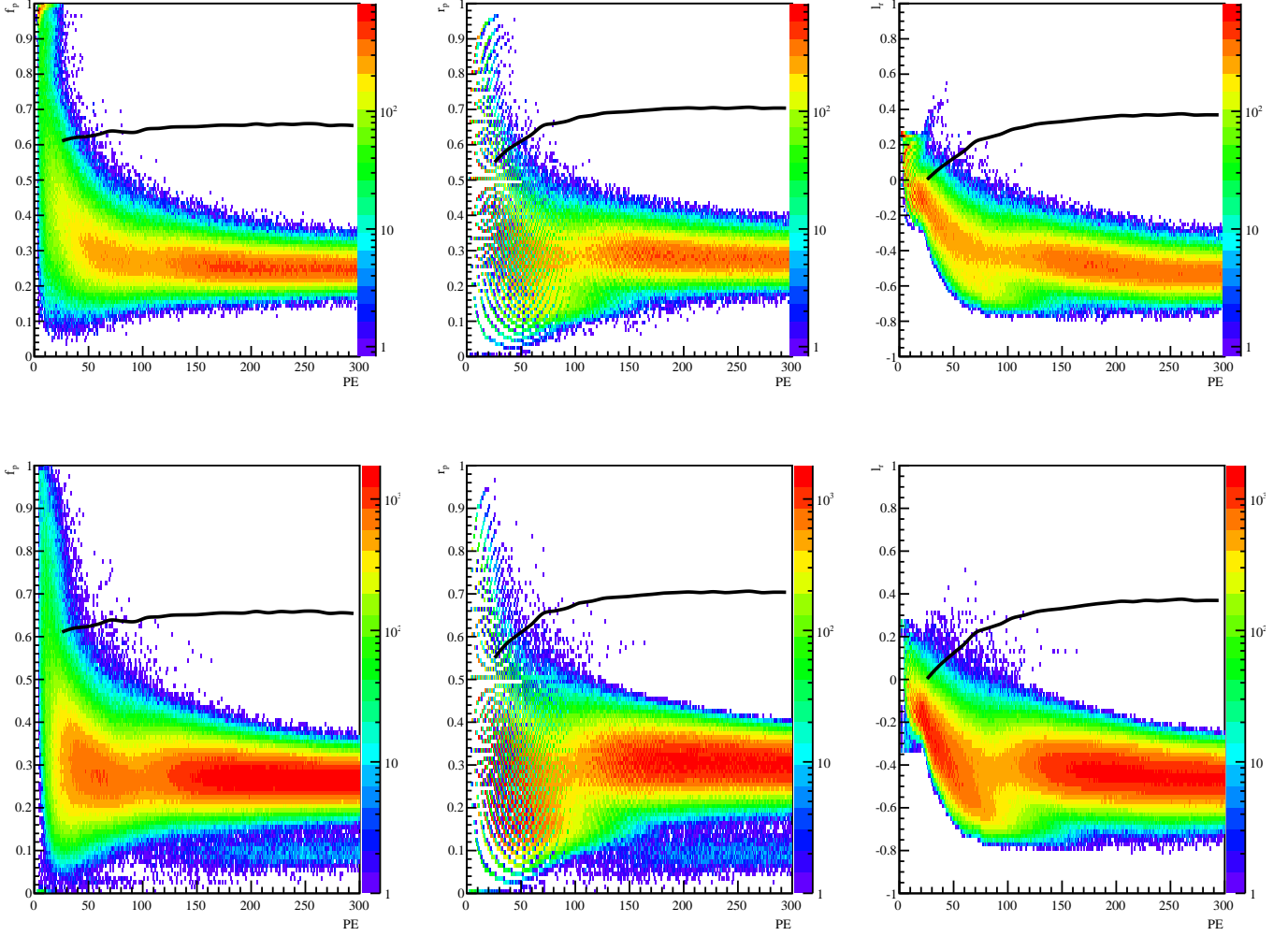


Figure 11: Distribution of  $f_p$  (left column),  $r_p$  (middle column), and  $l_r$  (right column) test statistics for simulated electronic recoils from a  $^{22}\text{Na}$  calibration source as a function of reconstructed number of photoelectrons in DEAP-1 simulation (top row) and data (bottom row). The solid black line in each panel represents the simulated mean profile of each test statistic for nuclear recoils in DEAP-1. The discrete nature of the  $r_p$  test statistic creates the structure which is apparent in the center panels. A population of events at positive  $l_r$  values below 100 PE due to pileup of Cherenkov light from the 1274 keV gamma with scintillation from the 511 keV gamma becomes increasingly apparent with the improved test statistic.

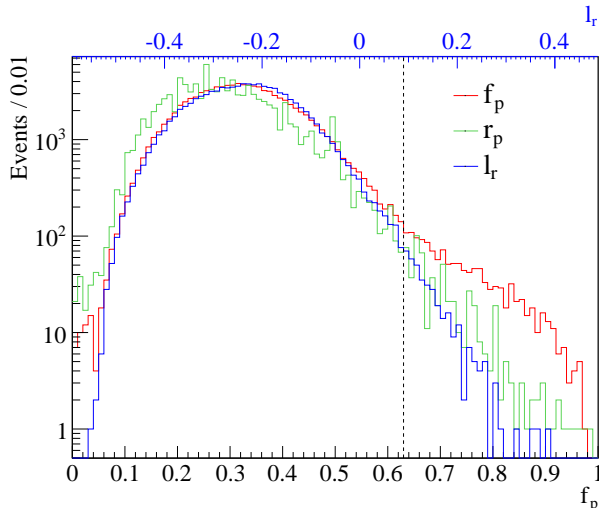


Figure 10: Distribution of  $f_p$ ,  $r_p$ , and  $l_r$  test statistics for electronic recoils for  $^{22}\text{Na}$  calibration events in DEAP-1 with 30 PE. The vertical dashed line indicates 50% nuclear recoil acceptance at 6.7 keVee. The  $l_r$  values have been linearly transformed such that the median values for the electron and nuclear recoil distributions match those for  $f_p$ . The shift in the  $r_p$  peak relative to  $f_p$  is due to the discrete nature of the test statistic.

by the DEAP/CLEAN collaborations and the SNO+ [21] collaboration. RAT brings together the electromagnetic and hadronic physics simulation provided by GEANT4, the data storage and processing tools provided by ROOT, and parts of the scintillation and PMT simulation from GLG4sim, an open source package released by the KamLAND collaboration.

RAT simulates the following detector effects:

- Propagation of primary particles, such as electrons, gamma rays, nuclear recoils, and neutrons through detector materials using GEANT4.
- Production of extreme UV (EUV) scintillation light by charged particles in the liquid argon. This includes the energy and particle dependence measured in [1].
- Propagation of individual EUV photons and wavelength-shifted photons through the detector with GEANT4, including both bulk and surface optical processes of liquid argon, TPB, glass, and metal surfaces.
- Detection of photons at PMTs and the production of realistic pulses including time, charge and shape variations, as well as pre-pulsing, late-pulsing, double-pulsing, and after-pulsing.
- Detector triggering, waveform digitization, zero suppression, and data packing for readout.

The output of the simulation stage is formatted identically to real detector data, allowing for the development of analysis algorithms that can directly operate on both simulated and real events.

A detailed geometry of the DEAP-1 detector in the configuration described above was used to generate events simulating the  $^{22}\text{Na}$  calibration data. The source configuration is modeled on the physical location of the source and trigger PMT with 511 keV gamma rays produced towards the center of the detector in a  $9^\circ$  cone coincident with a directionally uncorrelated 1274 keV gamma ray.

The distributions of the  $f_p$ ,  $r_p$ , and  $l_r$  test statistics for simulated  $^{22}\text{Na}$  calibration data is shown in Figure 11 along with the mean profile for simulated nuclear recoils for comparison. The simulated  $^{22}\text{Na}$  test statistic distributions broadly reproduce the basic shape of the distributions found in the data, although with limited statistics. In particular, the nuclear recoil-like events begin to emerge at positive values of  $l_r$  as in the calibration data.

## 7. The MiniCLEAN Detector

The MiniCLEAN detector [13, 14], shown in Figure 12, is a cryogenic scintillation detector capable of operating with liquid argon or liquid neon. The 500 kg central volume is surrounded by a spherical array of 92 light guides each viewed by a Hamamatsu Photonics cryogenic 8" R5912-02-MOD PMT. The conceptually simple, scalable design with maximal PMT coverage is similar to the approaches taken by other large dark matter and neutrino detectors (for example DEAP-3600 [15], SNO+ [21], KamLAND [22], etc). Unlike the smaller two-PMT DEAP-1, MiniCLEAN's 92 PMTs allow 3D position reconstruction. With a complete 3D vertex, MiniCLEAN is able to apply a strict fiducial volume cut, significantly reducing backgrounds from the TPB surfaces and PMTs. The large array of PMTs, relative to DEAP-1, with maximal coverage also reduces the pileup of scintillation photons and leads to a higher expected light yield in the MiniCLEAN design. Both effects contribute to improved particle identification as described in Section 3.

Each of MiniCLEAN's PMTs is fully immersed in the liquid and mounted in a cassette assembly that optically couples the PMT to a central target volume with a reflective tube capped by a 10 cm thick acrylic plug. The acrylic plug acts as a neutron shield and a supporting substrate for a  $2\ \mu\text{m}$  layer of TPB that directly faces the argon volume to convert the extreme UV scintillation photons to the visible. The TPB-coated acrylic blocks are suspended inside the detector by the cassette bodies to create a polyhedral approximation of a sphere at an approximate radius of 43.7 cm with 97% of the surface area covered with TPB. When filled with liquid argon, the total target mass is 500 kg, and after a nominal fiducial volume cut of 29.5 cm, the fiducial mass is 150 kg.

The data acquisition system is similar to the DEAP-1 design, employing 13 CAEN V1720 250 MHz, 12-bit digi-



tizers that compress 16  $\mu\text{s}$  waveforms into non-contiguous blocks of samples by applying zero-suppression. Each block of samples contains some number of pre- and post-samples that extend beyond the samples exceeding the zero-suppression threshold. In a typical configuration, the minimum size of a block containing at least one photoelectron pulse is 32 samples. The CAEN V1720 digitizers have been observed in the lab to have a noise RMS of 0.41 mV per sample when connected to a full length coaxial cable and PMT.

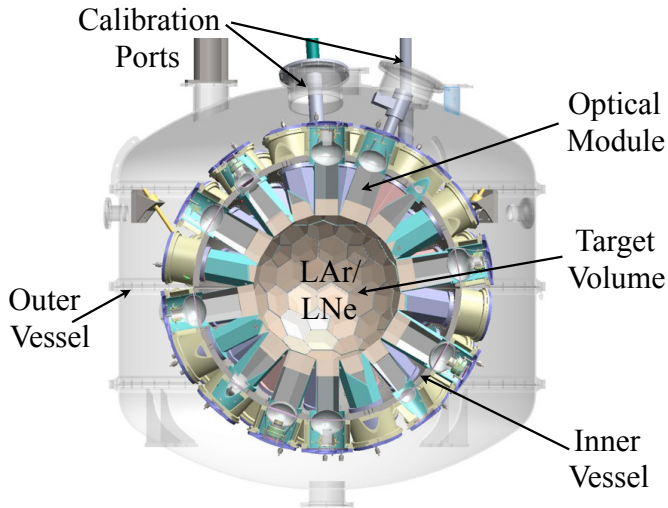


Figure 12: A 3D rendering of the MiniCLEAN detector. The optical modules have TPB-coated acrylic at the innermost radius and PMTs at the outermost radius.

## 8. Simulation of MiniCLEAN

A sample of 84 million  $^{39}\text{Ar}$  decays were simulated and processed through the entire analysis sequence as described in Section 4. The events have been limited to be within the fiducial radius, where the energy scale has no radial dependence. For 75-100 PE (12.5-16.7 keVee) events, Figure 13 shows the fractional error in the number of photoelectrons estimated using the Bayesian photoelectron counter on the first pass with bootstrapped priors compared to the second pass where the priors,  $\mu$  from Equation 4, are updated using position reconstruction. Although the updated Bayesian priors can induce small channel-to-channel correlations, the second pass of the algorithm is an overall improvement. The second pass of Bayesian photoelectron counting reduces the energy bias from 3.0% to 0.5% and improves the energy resolution from 4.1% to 3.9%.

Figure 14 compares the apparent event energy from normalized charge integration, where the number of photoelectrons is estimated from the total charge divided by the mean single photoelectron charge, to the number of photoelectrons estimated using the two-pass Bayesian photoelectron counter. Since events are restricted to a region

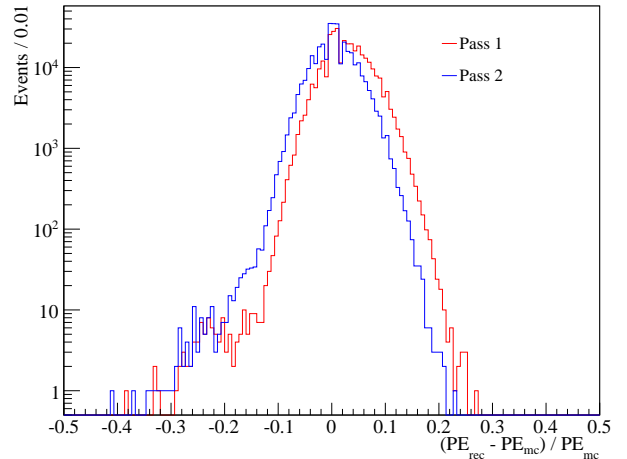


Figure 13: The fractional error in estimation of the number of photoelectrons in simulated electron events in the MiniCLEAN detector between 75 and 100 PE. The second pass of Bayesian photoelectron counting, which replaces the bootstrapped Bayesian prior with a prior derived from the reconstructed position and energy of the event, reduces the energy bias from 3.0% to 0.5%, reduces the energy resolution from 4.1% to 3.9%, and reduces the RMS of the  $l_r$  distribution by 5%.

where the energy scale has no radial dependence, the improvement in photoelectron counting resolution is a direct measure of the improvement of the energy resolution. The energy resolution is reduced from 4.3% with the charge integration method to 3.1% with the Bayesian photoelectron counting.

The improved energy resolution offered by the Bayesian photoelectron counting reduces the number of low energy background events entering the energy region of interest. Lowering of the threshold with reduced background in the presence of a WIMP-like signal can dramatically improve the sensitivity to signal which increases rapidly at low energies.

The 2D distributions of the  $f_p$ ,  $r_p$ , and  $l_r$  test statistics as a function of reconstructed number of photoelectrons are shown in Figure 15 for simulated events in the MiniCLEAN detector. The  $l_r$  test statistic significantly tightens up the distributions of electrons and nuclear recoils at very low numbers of photoelectrons, which is precisely where a WIMP signal will show increasing rate. We do not have enough CPU resources to fully simulate the number of events required to estimate the background leakage for each of these test statistics directly, as we did with the fast simulation in Section 3. However, we can linearly transform  $l_r$  into the same [0,1] range as  $f_p$  and  $r_p$  so the median  $l_r$  value for electrons is the same as the median  $f_p$ , and the same is true for nuclear recoil distributions. This allows the tails of the distributions to be compared

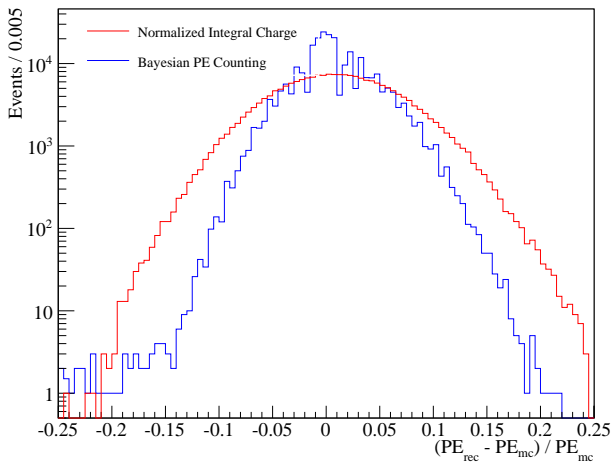


Figure 14: The ratio of estimated number of photoelectrons divided by the true number of photoelectrons for simulated  $^{39}\text{Ar}$  decays in the MiniCLEAN. Only events between 75 and 100 PE, and radius less than 295 mm are shown.

more easily, as is shown in Figure 16. Electrons with only 40 PE (6.67 keVee as in Figure 10 for DEAP-1 given the respective light yields) show a significantly smaller tail in the  $l_r$  distribution compared to  $f_p$  and  $r_p$ . Above  $f_p$  of 0.7, the  $l_r$  tail is reduced by a factor of  $7.2 \pm 0.2$  (stat) compared to the  $f_p$  distribution.

## 9. Conclusion

Many current and future neutrino and dark matter experiments depend on the detection of scintillation light with PMTs. The ability to both count and identify the times of individual photoelectrons in a PMT waveform directly impacts energy resolution and particle identification with timing. When scintillation light with both short and long time constants is present, simple strategies for photoelectron counting will alternately introduce biases or large variance in different situations. We have resolved this difficulty by showing how to apply a Bayesian prior assumption about the intensity and time structure of the light to the counting process. By iterating the procedure, a crude, bootstrapped prior can be refined with the addition of reconstructed observables, such as event energy and position.

Although the canonical prompt-fraction,  $f_p$ , particle identification parameter often offers excellent discrimination, realistic detector effects can degrade the discrimination power significantly with the extent depending on the scintillator and detector properties. Using the times output by the Bayesian photoelectron counting procedure, improved test statistics, such as  $r_p$  and  $l_r$ , can be constructed to better utilize the known time structure of the scintilla-

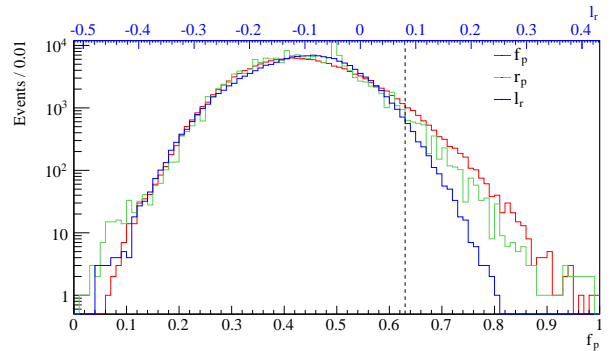


Figure 16: Distribution of  $f_p$ ,  $r_p$  and  $l_r$  test statistics for electronic recoils for simulated MiniCLEAN events at 40 PE. The vertical dashed line indicates 50% nuclear recoil acceptance at 6.67 keVee. The  $l_r$  values have been linearly transformed such that the median values for the electron and nuclear recoil distributions match those for  $f_p$ .

tion light. Although this is applicable to a wide range of scintillator detectors, here we have demonstrated the improved particle identification test statistics in the DEAP-1  $^{22}\text{Na}$  calibration data and MiniCLEAN  $^{39}\text{Ar}$  simulation. With improved energy resolution and improved particle identification test statistics, scintillator experiments looking for rare events can improve sensitivity to the signal of interest through the reduction of backgrounds and lowering of the energy threshold.

## Acknowledgments

This work has in part been supported by the United States Department of Energy, Office of High Energy Physics.

Support for DEAP-1 has been provided by the Canadian Foundation for Innovation and the Natural Sciences and Engineering Research Council. The High Performance Computing Virtual Laboratory (HPCVL) has provided us with CPU time, data storage, and support. We would also like to thank the SNOLAB staff for on-site support. The work of our co-op and summer students, including Christopher Stanford who operated DEAP-1 during the data-taking in this paper, is gratefully acknowledged.

Contributions by staff of NIST, an agency of the US government, are not subject to copyright in the US.

## 10. References

### References

- [1] W.H. Lippincott, et al. Phys. Rev. C **78** (2008) pp. 035801.
- [2] M.G. Boulay, et al. arXiv:0904.2930.
- [3] K Veshima, et al. NIM A **659** (2011) 161-168.
- [4] J. Angle, et al. arXiv:0910.3698.
- [5] D. Budjás et al. JINST **4** (2009) P10007.
- [6] S. Yoshida, et al. NIM A **622** (2010) 574-582.
- [7] T. Caldwell, S. Seibert, S. Jaditz. JINST **8** (2013) C09004.
- [8] M. Miyajima, et al. Phys. Rev. A **9** (1974) 1438.

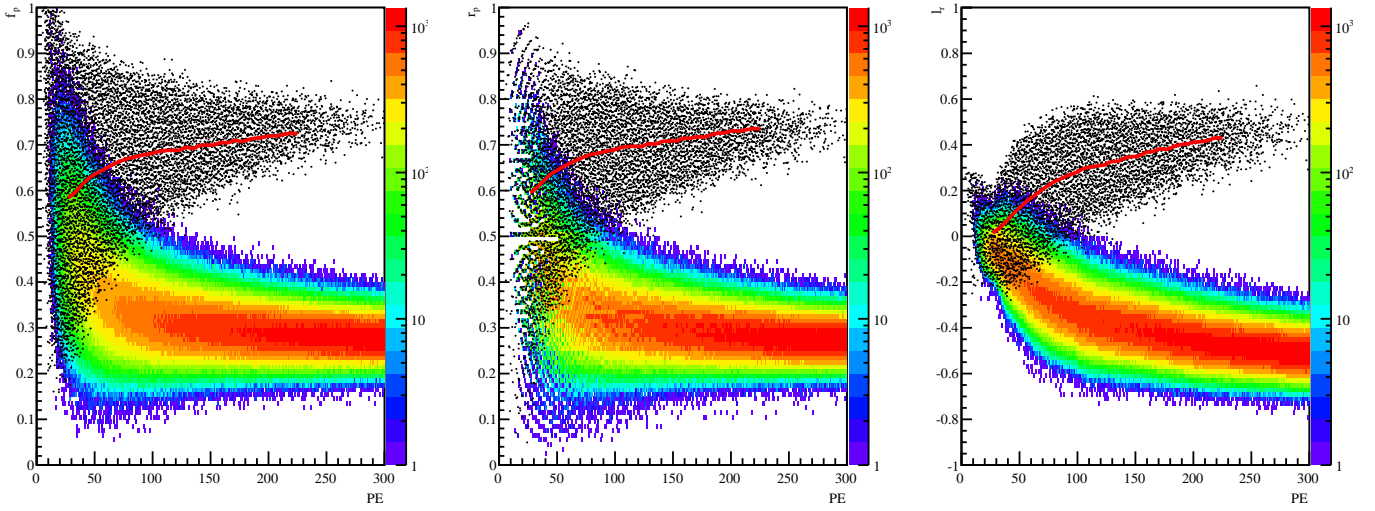


Figure 15: Distribution of  $f_p$  (left panel),  $r_p$  (center panel), and  $l_r$  (right panel) test statistics for simulated electronic (color scale) and nuclear (black points) recoils as a function of reconstructed number of photoelectrons in the MiniCLEAN detector. The solid red line in each panel indicates the mean profile for nuclear recoil.

- [9] P. Benetti, et al. NIM A **574** (2007) 83-88.
- [10] M.G. Boulay, A. Hime, Astroparticle Physics **25** (2006) 179-182.
- [11] D. Gastler, et al. Phys. Rev. C **85** (2012) 065811.
- [12] P.-A. Amaudruz, et al. arXiv:1211.0909.
- [13] A. Hime for the MiniCLEAN Collaboration. arXiv:1110.1005.
- [14] K. Rielage for the MiniCLEAN Collaboration. arXiv:1403.4842.
- [15] M.G. Boulay for the DEAP Collaboration. arXiv:1203.0604.
- [16] R. S. Muliken J. Chem. Phys. **52** (1970) 5170.
- [17] A. Hitachi, et al Phys. Rev. B **27** (1983) 5279.
- [18] A. Gedanken, et al. J Chem Phys. **57** (1972) 3456.
- [19] V. Gehman, et al. NIM A **654** (2011) 116-121.
- [20] T. Bolton. Nucl. Phys. Proc. Suppl. **149** (2005) 166-169.
- [21] M.C. Chen. Nucl. Phys. Proc. Suppl. **145** (2005) 65-68.
- [22] S. Abe, et al. Phys. Rev. C **84** (2011) 035804.