



# Toward Optimal Signal Extraction for Imaging X-Ray Polarimetry

A. L. Peirson and Roger W. Romani

Kavli Institute for Particle Astrophysics and Cosmology, Dept. of Physics, Stanford University, Stanford, CA, 94305, USA

Received 2021 May 6; revised 2021 July 15; accepted 2021 July 16; published 2021 October 11

## Abstract

We describe an optimal signal extraction process for imaging X-ray polarimetry using an ensemble of deep neural networks. The initial photoelectron angle, used to recover the polarization, has errors following a von Mises distribution. This is complicated by events converting outside of the fiducial gas volume, whose tracks have little polarization sensitivity. We train a deep ensemble of convolutional neural networks to select against these events and to measure event angles and errors for the desired gas-conversion tracks. We show how the expected modulation amplitude from each event gives an optimal weighting to maximize signal-to-noise ratio of the recovered polarization. Applying this weighted maximum likelihood event analysis yields sensitivity (MDP<sub>99</sub>) improvements of  $\sim 10\%$  over earlier heuristic weighting schemes and mitigates the need to adjust said weighting for the source spectrum. We apply our new technique to a selection of astrophysical spectra, including complex extreme examples, and compare the polarization recovery to the current state of the art.

*Unified Astronomy Thesaurus concepts:* Polarimetry (1278); Convolutional neural networks (1938); Astrostatistics (1882); X-ray telescopes (1825); X-ray astronomy (1810)

## 1. Introduction

Imaging polarimetry in the classical soft X-ray band (1–10 keV) offers many novel probes of compact objects and nonthermal X-ray nebulae. The Imaging X-ray Polarimetric Explorer (IXPE; O’Dell et al. 2018; Weisskopf 2018) planned to launch in 2021 November, will provide our first polarization images of a variety of X-ray source classes. IXPE’s sensitivity is limited by the track analysis algorithm used to recover source polarization, spatial structure, and energy, given a measured set of photoelectron track images. In Peirson (2021, hereafter P21) we studied a set of simulated IXPE Gas Pixel Detector (GPD) events and demonstrated how convolutional neural networks (CNNs) track measurement can provide substantial improvements to the current state of the art in polarization recovery. In this work we improve on these results while simplifying the overall analysis pipeline from track reconstruction to polarization prediction. Further, we show that our analysis is near optimal. We demonstrate this analysis on several simulated astrophysical spectra: generic power laws, an intermediate synchrotron peak (ISP) blazar, and a bright accreting X-ray binary, comparing performance to IXPE’s default moment analysis scheme (Bellazzini et al. 2003). This range of spectra illustrates how our revised weighting performs well over a broad range of models, eliminating the need to recalibrate weights for each source spectrum. While the results shown here are specific to IXPE’s GPD, the methods are general, and can be applied to other imaging detector geometries.

In the 1–10 keV range, the cross section for photoelectron emission is proportional to  $\cos^2(\theta - \theta_0)$ , where  $\theta_0$  is the normal incidence X-ray’s electric vector position angle (EVPA) and  $\theta$  the azimuthal emission direction of the photoelectron. Specifically, the emission angles  $\theta$  follow the distribution

$$\theta \sim \frac{1}{2\pi}(1 + p_0\mu_{100}\cos[2(\theta - \theta_0)]), \quad (1)$$

where  $0 \leq p_0 \leq 1$  is the true polarization fraction of the source,  $-\pi/2 \leq \theta_0 < \pi/2$  is the true source EVPA (the expected direction where the  $\theta$  distribution peaks), and  $\mu_{100}$  is the

modulation factor. The modulation factor, determined by the interaction physics and polarimeter properties, is defined as the amplitude of the azimuthal modulation measured for a 100% polarized source (i.e., for  $p_0 = 1$ ), so  $0 \leq \mu_{100} \leq 1$ . The modulation factor  $\mu_{100}$  is a strong function of photon energy and of the track reconstruction algorithm. By measuring a data set of individual photoelectron emission angles  $\{\theta_i\}_{i=1}^N$ , one can recover the above distribution to extract the source polarization parameters ( $p_0, \theta_0$ ). For imaging X-ray polarimeters one also wishes to reconstruct the X-ray absorption (conversion) points and photon energy.

The sensitivity of a track reconstruction algorithm is typically measured by its minimum detectable polarization (MDP<sub>99</sub>) defined as

$$\text{MDP}_{99} \approx \frac{4.29}{\mu_{100}\sqrt{N}}, \quad (2)$$

where  $N$  represents the total number of observed events. This is the 99% confidence upper limit on polarization fraction  $p_0$  for a 0% polarized source. A lower MDP<sub>99</sub> is better. It is equivalent to the reciprocal of the signal-to-noise ratio (S/N), where we have an approximate Poisson counting noise,  $1/\sqrt{N}$ , and the recovered signal is represented by  $\mu_{100}$ . In an event-weighted analysis  $N \rightarrow N_{\text{eff}} < N_{\text{events}}$ . Therefore, the best track reconstruction algorithm will maximize  $\mu_{100}\sqrt{N_{\text{eff}}}$ .

The default track reconstruction method for the GPD is a moment analysis described by Bellazzini et al. (2003). Impressive accuracies for the absorption point and EVPA angle are achieved from a simple weighted combination of track moments. Track energy estimates are proportional to the total collected GPD charge. The track ellipticity also provides a rough proxy for track reconstruction quality. High ellipticity tracks typically have more accurate angle estimates.

We showed in P21 that an ensemble of CNNs trained on simulated photoelectron tracks can be used to predict not only point estimates of photoelectron angles  $\{\hat{\theta}_i\}$ , but also the total Gaussian uncertainty (standard deviation) on these estimates

$\{\hat{\theta}_i\}$ . We used these uncertainty estimates as weights in a maximum likelihood estimation (MLE) to predict source polarization parameters  $(p_0, \theta_0)$ , improving on moment-analysis accuracy. Our neural network (NN) approach also significantly decreased photon absorption point and energy resolution errors, estimating these simultaneously with  $\hat{\theta}_i, \hat{\sigma}_i$ .

Although our previous approach yielded improved polarimetry results, the weights were heuristic and some aspects complicated practical applications. For example, our event weighting scheme required tuning for individual spectra through a power-law weighting with index  $\lambda$ , designed to approximate the angle error distribution for a given source spectrum. Thus, the weights were suboptimal, especially over large spectral bands.

Here, we adopt a more general approach, studying a set of simulated IXPE event images to define event weights that give improved sensitivity and more uniform applicability. To improve event weight fidelity we incorporate an event classification scheme that helps separate good GPD events, converting in the gas, to those converting outside, but penetrating the active gas volume and triggering the detector. A retrained set of networks with a more realistic error distribution delivers good event measurements and error estimates for our (near-) optimal weighting scheme. We test this analysis against realistic astrophysical spectra, as might be expected in early IXPE observations.

## 2. Polarization Estimation

For clarity in the rest of the paper, we briefly summarize the steps required to go from an observed data set of photoelectron track events measured by a detector like IXPE to the final source polarization parameters  $(p_0, \theta_0)$  and their uncertainties. For additional details and evidence, the reader is referred to Kislat et al. (2015) and P21.

Imaging X-ray polarimeters measure photoelectron tracks, like those in Figure 1. In IXPE GPD events, photoelectron ionization tracks deposit charge onto a 2D hexagonal pixel detector array.

From an observed set of  $N$  tracks, one estimates a set of initial photoelectron directions  $\{\hat{\theta}_i\}_{i=1}^N$ . This can be done using, for example, the moment analysis (Bellazzini et al. 2003) or an NN (P21). The polarization fraction  $p_0$  and EVPA  $\theta_0$  can be reconstructed from the observations  $\{\hat{\theta}_i\}_{i=1}^N$  in a number of ways, for example, by binned fitting or numerical MLE for the distribution in Equation (1). We prefer and recommend the approach taken by Kislat et al. (2015). They rewrite Equation (1) as a function of linear Stokes parameters

$$\theta \sim \frac{1}{2\pi}(1 + Q \cos 2\theta + U \sin 2\theta), \quad (3)$$

and they define unbiased estimators for the Stokes parameters

$$\hat{I} = \sum_{i=1}^N 1 = N, \quad (4)$$

$$\hat{Q} = \frac{1}{\hat{I}} \sum_{i=1}^N 2 \cos 2\hat{\theta}_i, \quad (5)$$

$$\hat{U} = \frac{1}{\hat{I}} \sum_{i=1}^N 2 \sin 2\hat{\theta}_i, \quad (6)$$

from which the polarization fraction and EVPA estimates can be derived:

$$\hat{p}_0 = \frac{1}{\mu_{100}} \sqrt{\hat{Q}^2 + \hat{U}^2}, \quad (7)$$

$$\hat{\theta}_0 = \frac{1}{2} \arctan \frac{\hat{U}}{\hat{Q}}. \quad (8)$$

The advantage of this approach is that we have simple analytical expressions for the desired polarization parameters and their errors. It can be shown in the high  $N$  limit that  $\hat{Q}, \hat{U}$  are minimum variance estimators and recover the MLE result for  $\hat{p}_0, \hat{\theta}_0$ . Additionally, Stokes parameter estimation can be readily adjusted for event-weighted estimates by modifying Equations (4) and (6) to include event weights  $\{w_i\}_{i=1}^N$

$$\hat{I} = \sum_{i=1}^N w_i, \quad (9)$$

$$\hat{Q} = \frac{1}{\hat{I}} \sum_{i=1}^N 2w_i \cos 2\hat{\theta}_i, \quad (10)$$

$$\hat{U} = \frac{1}{\hat{I}} \sum_{i=1}^N 2w_i \sin 2\hat{\theta}_i, \quad (11)$$

and introducing

$$W_2 = \sum_{i=1}^N w_i^2. \quad (12)$$

The polarization fraction and EVPA are still estimated using Equations (7) and (8) but now the effective number of events is given by

$$N_{\text{eff}} = \sqrt{\hat{I}^2 / W_2}. \quad (13)$$

Kislat et al. (2015) derive the joint error distribution (posterior) for the polarization fraction and EVPA estimators as

$$P(\hat{p}_0, \hat{\theta}_0 | p_0, \theta_0) = \frac{\sqrt{N_{\text{eff}}} \hat{p}_0 \mu_{100}^2}{2\pi\sigma} \times \exp \left[ -\frac{\mu_{100}^2}{4\sigma^2} \{ \hat{p}_0^2 + p_0^2 - 2\hat{p}_0 p_0 \cos(2(\hat{\theta}_0 - \theta_0)) - \frac{\hat{p}_0^2 p_0^2 \mu_{100}^2}{2} \sin^2(2(\hat{\theta}_0 - \theta_0)) \} \right], \quad (14)$$

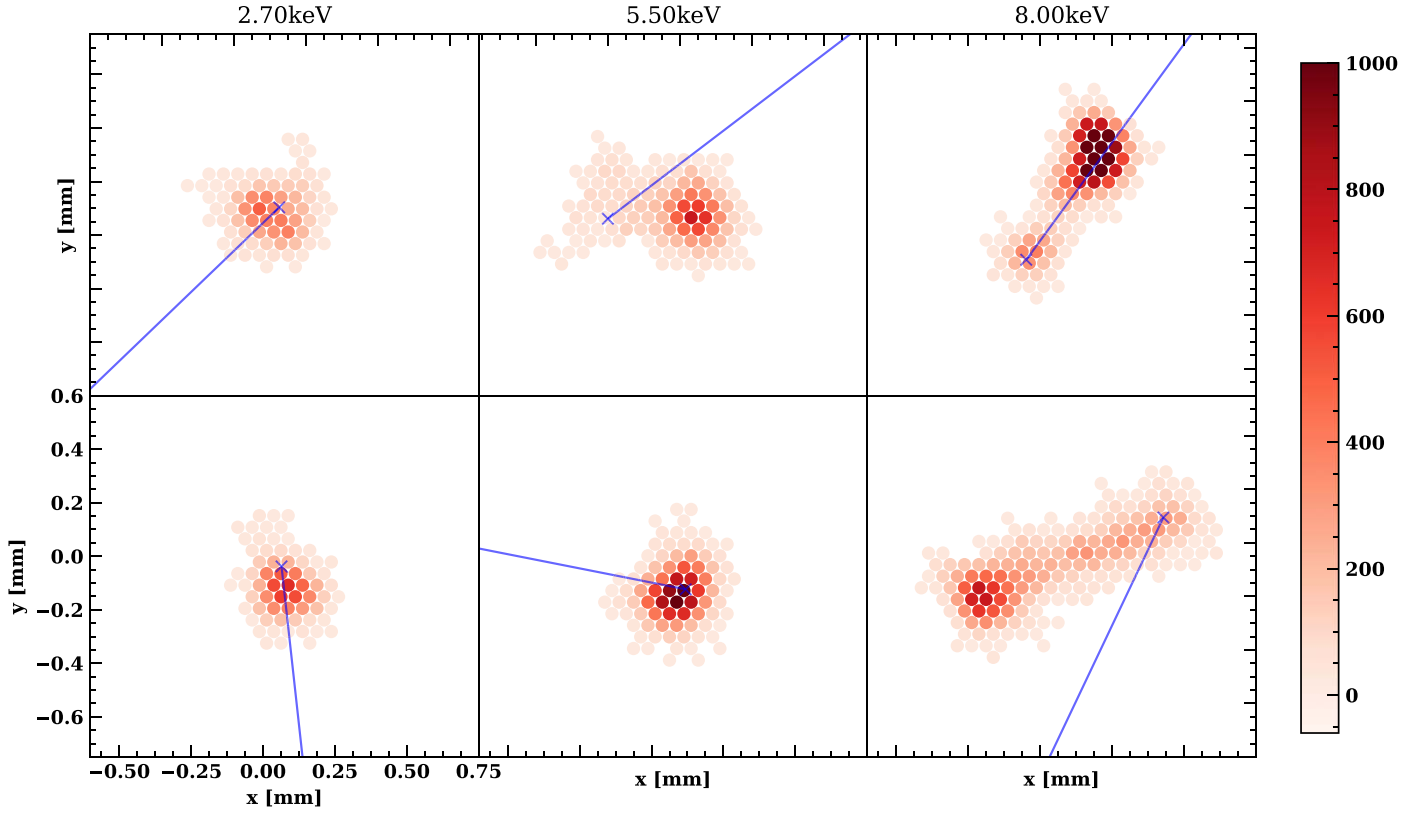
where

$$\sigma = \sqrt{\frac{1}{N_{\text{eff}}} \left( 1 - \frac{p_0^2 \mu_{100}^2}{2} \right)}. \quad (15)$$

Confidence intervals and the MDP<sub>99</sub> (Equation (2), the 99% upper limit for  $p_0 = 0$ ) are derived from this posterior probability distribution. In cases where  $\mu_{100}$  and  $p_0$  are not close to 0, the Gaussian approximation for the marginalized errors below is sufficient:

$$\sigma(\hat{p}_0) \approx \sqrt{\frac{2 - \hat{p}_0^2 \mu_{100}^2}{(N_{\text{eff}} - 1) \mu_{100}^2}}, \quad (16)$$

$$\sigma(\hat{\theta}_0) \approx \frac{1}{\hat{p}_0 \mu_{100} \sqrt{2(N_{\text{eff}} - 1)}}. \quad (17)$$



**Figure 1.** A selection of measured photoelectron track events from IXPE simulations. Pixel color density represents charge deposited, blue crosses show the photon  $x$ - $y$  conversion point and blue lines the initial photoelectron direction. Track morphologies vary widely and depend strongly on energy.

High  $\mu_{100}$  and high  $N_{\text{eff}}$  are both desirable to minimize the errors on recovered polarization parameters.

In this paper we present our results in  $(p_0, \theta_0)$ -space, as opposed to  $(Q, U)$ -space. We note in some cases it may be preferable to remain in  $(Q, U)$ -space since it significantly simplifies the posterior, Equation (14). On account of its simplicity and optimality, we use weighted Stokes parameter estimation throughout the remainder of this paper.

### 3. Optimal Event Weights

Photoelectron angle estimates  $\hat{\theta}_i$  from track images of X-ray polarimeters like IXPE are extremely heteroskedastic: the uncertainty on  $\hat{\theta}_i$  varies widely for different events  $i$ . This is mainly because track morphology is a strong function of energy. Low-energy tracks are short and circular making it very difficult to determine  $\hat{\theta}$ , while high energy tracks are long, making it easier to estimate  $\hat{\theta}$ . Track morphologies also vary widely at the same energy (see Figure (1)). Thus, an event-weighted estimate of polarization parameters, Equations (7) and (12), from an observation set  $\{\hat{\theta}_i\}$  would greatly improve both multiband and single-band polarization sensitivity, since some  $\hat{\theta}_i$  are much more accurate than others.

We showed in P21 that event weighting is very effective at improving sensitivity. There, we trained a deep ensemble of CNNs to predict  $(\hat{\theta}_i, \hat{\sigma}_i)$  pairs, where  $\hat{\sigma}_i$  is the total Gaussian uncertainty on each track estimate. We used  $w_i = \hat{\sigma}_i^{-\lambda}$  as event weights, where  $\lambda$  is a spectrum-dependent tuneable parameter. This improved  $\text{MDP}_{99}$  by more than 25% over the existing moment analysis. Although this prescription worked well, it is heuristic and by no means optimal.

Recently, Marshall (2021) showed that weighting events by their expected modulation factor  $\mu_{100}$  optimizes the expected S/N (minimizes the  $\text{MDP}_{99}$ ). This is illustrated in the following analysis. We define the S/N as

$$\text{S/N} \propto \mu_{100} \sqrt{N_{\text{eff}}}. \quad (18)$$

This is simply the inverse of the  $\text{MDP}_{99}$  (without constants); an optimal weighting scheme should maximize the S/N for a fixed number of events  $N$ . We can expand the S/N explicitly using our weighted estimators from Section 2, Equations (7)–(13),

$$\text{S/N} \propto \sqrt{\frac{\left(\sum_{i=1}^N 2w_i \cos 2\hat{\theta}_i\right)^2 + \left(\sum_{i=1}^N 2w_i \sin 2\hat{\theta}_i\right)^2}{\sum_{i=1}^N w_i^2}}. \quad (19)$$

Expanding, squaring, and dropping constants (which do not affect maximization) we obtain

$$\text{S/N}^2 \propto \frac{\sum_{i,j,i \neq j} w_i w_j \cos 2(\hat{\theta}_i - \hat{\theta}_j)}{\sum_k w_k^2}. \quad (20)$$

The estimators  $\hat{\theta}_i$  are random variables. The true values  $\theta_i$ , also random variables, follow the distribution Equation (1) with  $\mu_{100} = 1$  (since they are perfectly known). Assuming the  $\hat{\theta}_i$  are unbiased estimators of  $\theta_i$  (true for both moment analysis and NNs), we can say

$$\hat{\theta}_i = \theta_i + \epsilon_i, \quad (21)$$

where the measurement errors  $\epsilon_i$  are independent random variables drawn from the same family of distributions with

$$\mathbb{E}[\epsilon_i] = 0, \text{Var}[\epsilon_i] = \sigma_i^2. \quad (22)$$

The specific distribution of the measurement errors  $\epsilon_i$  will depend on the  $\hat{\theta}_i$  estimation method; however, since  $\theta_i$  are periodic,  $\epsilon_i$  should follow a periodic distribution. For any  $\epsilon_i$  distribution with the above properties, we can find the distribution for  $\hat{\theta}_i$  as the convolution of the  $\theta_i$ ,  $\epsilon_i$  distributions

$$\hat{\theta}_i \sim \frac{1}{2\pi} (1 + \mu_i p_0 \cos[2(\hat{\theta}_i - \theta_0)]), \quad (23)$$

where  $0 \leq \mu_i < 1$  and  $\mu_i(\sigma_i^2)$ . In other words, the distribution of estimators  $\hat{\theta}_i$  are the same as the distribution of the true values  $\theta_i$  but with a reduced modulation factor  $\mu_i$ . The measurement noise will blur the sinusoidal modulation signal by a factor  $\mu_i$  for the specific event  $i$ .

Knowing the distributions of  $\hat{\theta}_i$ , Equation (23), we take the expectation over  $S/N^2$  (dropping constant  $p_0$ )

$$\mathbb{E}[S/N^2] \propto \frac{\sum_{i,j,i \neq j} w_i w_j \mu_i \mu_j}{\sum_k w_k^2}. \quad (24)$$

Finally, maximizing this expression with respect to  $\{w_i\}$  in the large  $N$  limit we find

$$w_i = \mu_i, \quad (25)$$

i.e., the optimal weight for an event with observation  $\hat{\theta}_i$  is given by its expected  $\mu_{100}$ . Note for  $N \sim 20$  this already holds with high accuracy; useful polarization measurements typically have  $N > 1000$ .

Equation (25) makes intuitive sense, with each track weighted by its expected signal. Marshall (2021) adopted  $\mu_{100}(E)$  for the weight. This weight can be estimated from simulated or calibration data with 100% polarized photons at a range of energies. However, each energy bin requires many events for an accurate  $\mu$  estimate. Further, this weight represents only the mean response for events in a given energy bin, ignoring differences in the individual event quality, and so cannot provide optimal sensitivity. In practice, analysis of simulated data shows that this approach delivers only a 6%–7% improvement in  $\text{MDP}_{99}$  over an unweighted analysis. In P21 an individual event weight by a heuristic power law  $w_i = \hat{\sigma}_i^{-\lambda}$  already delivers  $\sim 25\%$  improvement in  $\text{MDP}_{99}$  over unweighted analysis; we expect optimal weights, reflecting the individual event  $\mu$ 's to deliver even better sensitivity.

### 3.1. NN Optimal Event Weights

Happily, we can train an ensemble of NNs to output individual event uncertainties, which can be converted to the optimal weights  $w_i = \mu_i$  if we know the distributions of the  $\theta_i$  estimation errors  $\epsilon_i$ . P21 assumed a simple Gaussian error distribution. This is inadequate when  $\epsilon_i$  is large (poorly measured tracks). Here, we take  $\epsilon_i$  to follow a von Mises distribution  $\text{VM}(0, \kappa_i)$ , the Gaussian on a periodic interval,

$$\epsilon_i \sim \frac{1}{2\pi I_0(\kappa_i)} \exp(\kappa_i \cos 2\epsilon_i). \quad (26)$$

Note that for large concentration  $\kappa_i$  (well-measured tracks) this reverts to a simple Gaussian with  $\sigma_i^2 \approx 1/\kappa_i$ .

We accordingly adjust our NN  $\theta_i$  loss function, minimizing the negative log-likelihood of the von Mises distribution to predict  $\hat{\theta}_i$ ,  $\hat{\kappa}_i$  pairs instead of the Gaussian uncertainty predictions  $\hat{\theta}_i$ ,  $\hat{\sigma}_i$  used in P21. The loss on input track image  $\mathbf{x}_i$  with true direction  $\theta_i$  is now

$$L(\theta_i|\mathbf{x}_i) = \log[I_0(\hat{\kappa}_i^a(\mathbf{x}_i))] - \hat{\kappa}_i^a(\mathbf{x}_i) \cos[2(\theta_i - \hat{\theta}_i(\mathbf{x}_i))]. \quad (27)$$

The predicted  $\hat{\kappa}_i^a$  are uncertainty parameters representing aleatoric uncertainty. Using a deep ensemble of NNs (Lakshminarayanan et al. 2017) we can additionally estimate the epistemic uncertainty arising from imperfect model specification. With the epistemic uncertainties assumed to follow  $\text{VM}(0, \kappa_i^e)$ ;  $\kappa_i^e$  can be estimated from the output of a deep ensemble with  $M$  NNs  $\{\hat{\theta}_{ij}\}_{j=1}^M$ :

$$\bar{R}_i^2 = \left( \frac{1}{N} \sum_{j=1}^M \cos 2\hat{\theta}_{ij} \right)^2 + \left( \frac{1}{N} \sum_{j=1}^M \sin 2\hat{\theta}_{ij} \right)^2, \quad (28)$$

$$\frac{I_1(\hat{\kappa}_i^e)}{I_0(\hat{\kappa}_i^e)} = \bar{R}_i, \quad (29)$$

with the Bessel functions  $I_0$  and  $I_1$ . The VM assumption may not be perfect. In that case, a more realistic epistemic error distribution can be recovered by bootstrap analysis of the ensemble predictions (P21); using this more realistic distribution can alter the weights and improve the MLE, but in practice the gains are small. Equation (29) gives an implicit prescription for the MLE for  $\kappa_i^e$ . We can convert to an equivalent Gaussian dispersion using the circular standard deviation

$$\sigma = \sqrt{-2 \log(\bar{R})}. \quad (30)$$

Combining  $\sigma(\hat{\kappa}_i^a)$  and  $\sigma(\hat{\kappa}_i^e)$  in quadrature yields the total  $\hat{\kappa}_i$ . The other loss function terms for track energies and absorption points are as in P21.

To convert the output  $\hat{\kappa}_i$  into optimal event weights, we need the distribution of  $\hat{\theta}_i = \theta_i + \epsilon_i$ . Convolving the distribution of  $\theta_i$  (Equation (1) with  $\mu_{100} = 1$ ) with  $\epsilon_i$  (Equation (26)) we find

$$\hat{\theta}_i \sim \frac{1}{2\pi} \left( 1 + \frac{I_1(\kappa_i)}{I_0(\kappa_i)} p_0 \cos[2(\hat{\theta}_i - \theta_0)] \right). \quad (31)$$

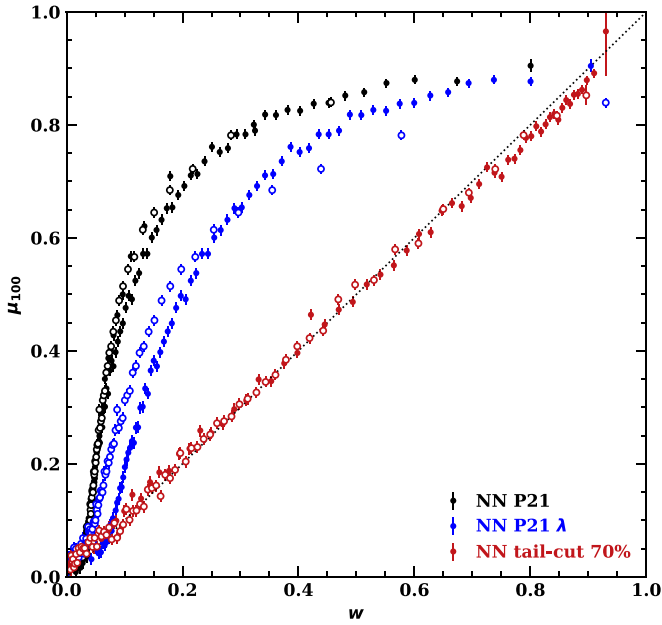
Comparing to Equation (23),

$$w_i = \mu_i = \frac{I_1(\hat{\kappa}_i)}{I_0(\hat{\kappa}_i)}. \quad (32)$$

Thus, Equation (32) can be used to transform our uncertainty parameters  $\hat{\kappa}_i$  into optimal event weights.

For a real detector, several effects mean that these weights will not be fully *optimal*. In detail, the error distribution is unlikely to follow a simple von Mises shape. In particular, for the GPD we find that events fall into several classes, with differing  $\epsilon_i$ . Further, with finite-sized training data sets and limited training time, our estimates of the  $\kappa_i$  will be imperfect. Nevertheless, we will show the approach achieves state-of-the-art results, while significantly simplifying previous NN analysis.

Truly optimal weights have  $w_i = \mu_i$  for any spectrum of events. However, the imperfections noted above will break this equivalence. These effects are often energy dependent, as illustrated in Figure 2, which shows the measured  $\mu_{100}$  as a function of weight computed from (Equation (32)) for two different event spectra. The spectrum independence of our



**Figure 2.** Measured  $\mu_{100}$  as a function of weight  $w$  for large test data sets of 1–10 keV simulated events. Each  $\mu_{100}$  bin contains 20,000 individual track events. Open circles represent a PL1 source convolved with IXPE’s effective area, closed circles a flat spectrum. For previous NN net analysis with untuned  $\lambda = 1$  weight (black) and with heuristic weight (blue), the weighting scheme is far from optimal and substantially dependent on the underlying spectrum. The new scheme of this paper (red) is much closer to optimal with minimal spectral dependence.

computation is imperfect, although appreciably better than for earlier schemes, such as the  $\hat{\sigma}_i^{-\lambda}$  scheme or the event ellipticity weighting employed for the moment analysis.

Note that we can post facto introduce a rescaling of the  $w_i$  to enforce linearity for any given source spectrum. This does indeed slightly improve MLE performance, at the expense of degraded S/N performance for other spectra. Thus, while it is acceptable to tune the  $w_i$  for a typical astronomical spectrum, we do not do that here as it is better to understand and model

the  $\epsilon_i$  to make the analysis as close to optimal as possible, using the  $w/\mu_{100}$  linearity as a check.

#### 4. Modulation Factor Recovery

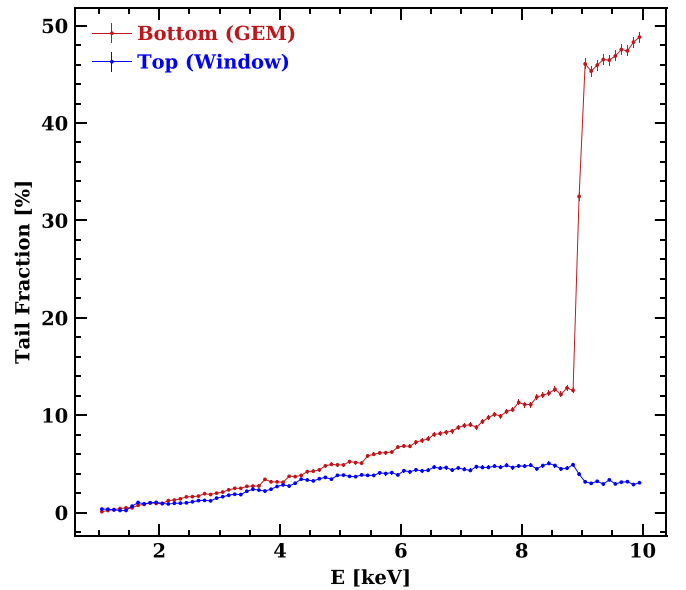
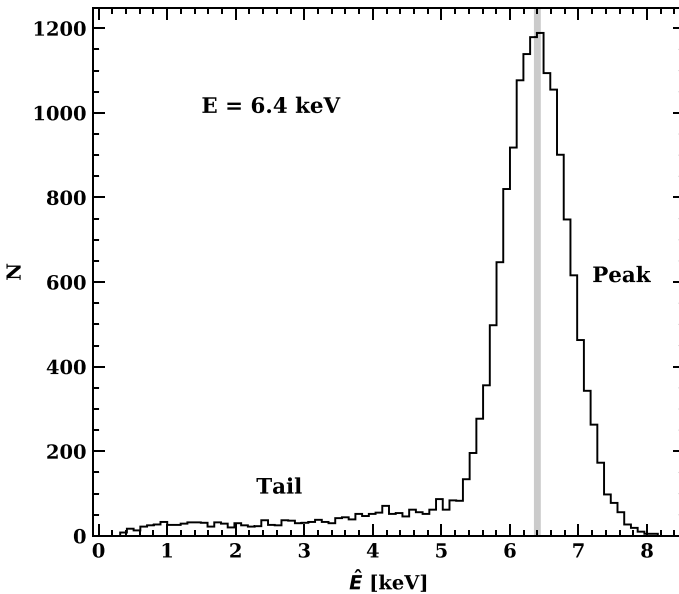
The spectral dependence in Figure 2 emphasizes that we must normalize our polarization sensitivity across a spectrum, through  $\mu_{100}$ , to recover true polarization amplitudes (Equation (7)). Since the sensitivity varies primarily with energy we usually compute  $\mu_{100}(E)$ ; this is complicated by the fact that the energy measured for each event  $\hat{E}_i$  is an imperfect reconstruction of the true energy  $E_i$ . For any arbitrary spectrum with  $N$  events and event weights  $w_i$

$$\mu_{100} = \frac{\sum_{i=0}^N w_i \mu_{100}(\hat{E}_i)}{\sum_{i=0}^N w_i} = \frac{\sum_{i=0}^N w_i^2}{\sum_{i=0}^N w_i}, \quad (33)$$

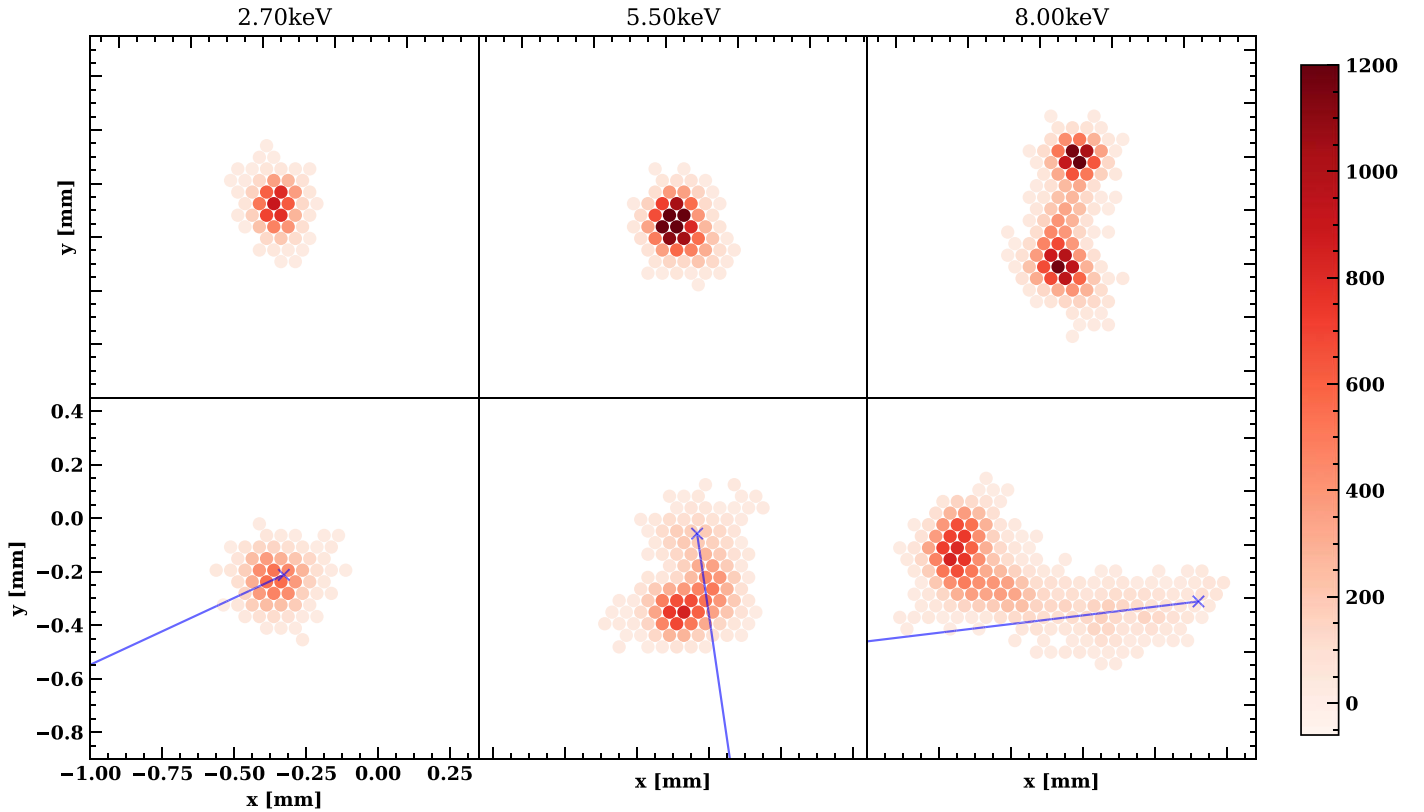
where the second equality holds when we have ideal values for the weights, avoiding all energy-dependent errors.

Alas, the basic NN trained on a collection of simulated events retains substantial energy dependence. In large part, this is due to multiple event classes: photon conversions occur in the Be window at the top of the gas cell as well as on the gas electron multiplier (GEM) at the bottom, with electrons emerging into the gas cell, triggering the detector and generating a track. Scattering near the conversion point in the solid ensures that these tracks have little to no correlation with the initial photoelectron direction. With decreased energy deposition in the gas, these events form a *tail* to the lower end of the energy point-spread function (PSF) (Figure 3). These tail tracks represent an increasing fraction of all events at high energy with a large increase at 8.9 keV representing the Cu absorption edge.

These tail events complicate spectral analysis, since a low recovered energy (low summed pixel count) event can be either a true low-energy photon or the window/GEM conversion of a high energy event. In P21, the NN recognized the tail events as poor reconstructions with little angle and energy information, assigning such events to energies in the middle of the range to



**Figure 3.** *Left:* recovered energy histogram for a 6.4 keV line source. A simple linear function of total charge deposited is used here for recovered energy, as in the standard moment analysis. The long–low-energy tail is produced by events converting in the window or GEM. *Right:* fraction of events that are tails as a function of energy. Red and blue traces show the Be window and GEM conversion, respectively. The jump in GEM conversions at 8.9 keV is prominent.



**Figure 4.** Example tail events (top row) and peak events (bottom row) for three different recovered energies. All plots follow the same spatial and color scale. These tail events convert in the GEM. Color denotes charge deposited in a given detector pixel. Note that tail events are more compact, short, and with high charge density at maximum.

minimize recovered energy dispersion. While this did decrease the energy rms error, as specified by the loss function, it produced a tail to the high side of low-energy events, as may be seen in the black histograms in the left two panels of Figure 6. Since most astrophysical spectra are expected to be power laws falling with energy, this high side tail from low-energy photons can significantly pollute the faint high energy bins with many low-weight, poorly measured low count events assigned a midrange energy.

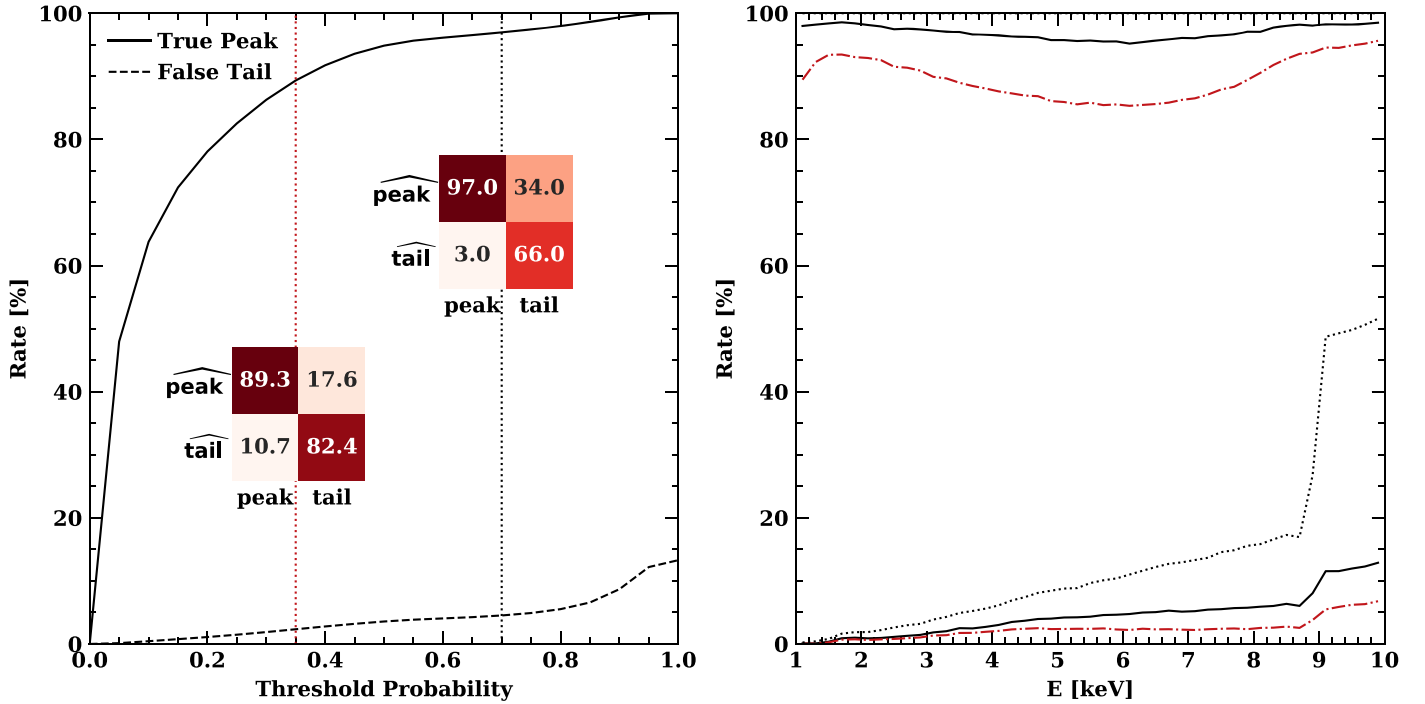
The tail events have dramatically lower polarization sensitivity. For example, while a flat spectrum of gas events produces a (weighted)  $\mu_{100} = 0.70$ , the GEM conversion events have  $\mu_{100} < 0.06$ . Thus, when analyzed together, tail events produce broad wings in the  $\epsilon_i$  distribution, departing from the peak tracks' von Mises dispersion and resulting in corrupted  $w_i$  estimates. The significant energy dependence of our weights' departure from  $\mu_{100}$  can be traced to the energy spectrum of tail events. In principle, with a very large training set an NN should be able to recognize different event classes, adjust the error distributions to match their varying uncertainties, and deliver robust  $\mu$  estimates. Limited training data and training time make this impractical.

The solution is to attempt to excise such events. The tail event tracks differ in morphology from those peak events converting in the gas (Figure 4). This figure compares events with the same recovered energy (largely determined through the summed pixel counts, the summed energy deposition in the gas). Tail events are typically due to photoelectrons ejected close to the GEM/window normal. Their tracks are thus more compact, with higher counts  $\text{pixel}^{-1}$ . Window events have larger drift diffusion than Cu GEM conversions. An NN can be

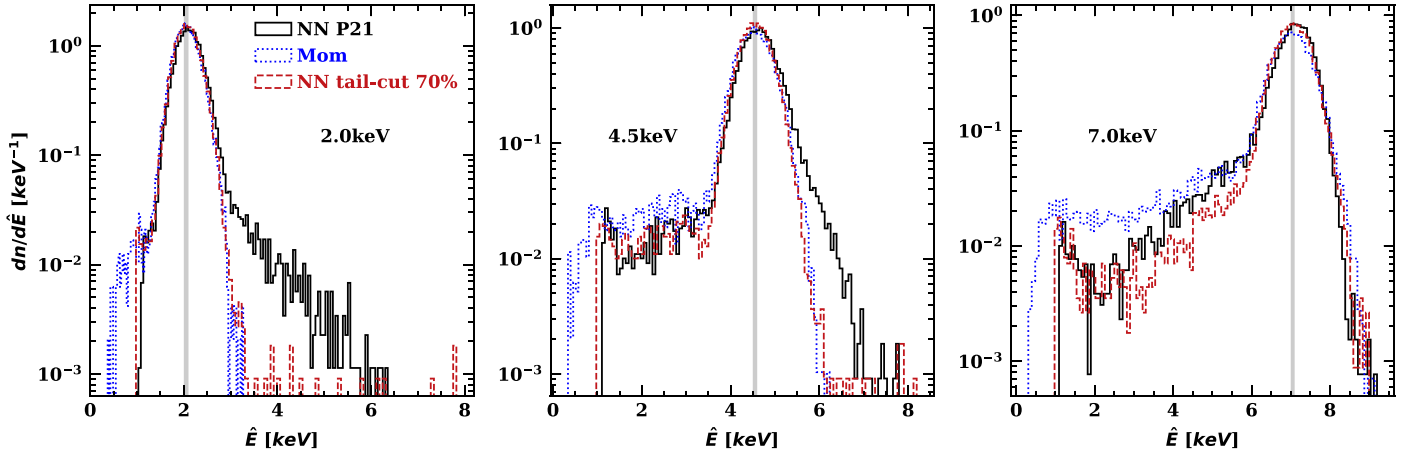
trained on simulated events to recognize these differences. We have thus trained a *preprocessing* CNN to do such basic event classification, assigning a tail probability for each event. This CNN shares a similar ResNet architecture (He et al. 2015) as the deep ensemble and is trained using a binary cross-entropy loss function. The output is a single scalar between 0 and 1 that represents the predicted event probability of peak or tail. Figure 5 shows how the fidelity of peak/tail separation varies with tail probability. GEM event identification is especially good, the remnant tail events are mostly window events blurred by diffusion. In the present analysis, we adopt a simple cut at 70% tail probability (right dotted line) and the inset gives the numerical percentages for peak and tail IDs, as well as the false-positive percentages. One might use this morphological classification index in a more sophisticated weighting scheme, for small additional gain.

## 5. Performance of Tail-suppressing NN Analysis

After suppression of the likely tail events, we have an event distribution more nearly dominated by gas/peak events. Thus, we train our event measurements NNs with a set of pure gas-conversion *peak* events and analyze post-suppression data sets as if they contain no tail events. Now, training with the same loss function, we achieve improved characterization of the remaining events. Figure 6 shows that the low-energy tail is suppressed by the cut, especially at high photon energies, while the undesirable high energy tail induced at low energies by training against the un-suppressed event set is avoided. The revised NN analysis slightly improves the energy resolution in the main peak, as well. This is better displayed in Figure 7,



**Figure 5.** *Left:* The solid curve shows the fraction of peak events retained as a function of the tail probability cut, while the dashed curve shows the fraction of the cut sample from the remaining tail events. Insets show the confusion matrices, normalized by column, for our adopted 70% cut and a 35% cut. *Right:* The top curves show the peak retention as a function of energy (black solid–70% cut, red dotted–dashed–35% cut). Below, the dotted curve shows the uncut fraction of the sample due to tail events, while the lower black and red curves show the residual tail pollution (70% and 35% cuts, respectively). Depending on how harmful tail events are to the desired measurements, different cut levels could be appropriate.



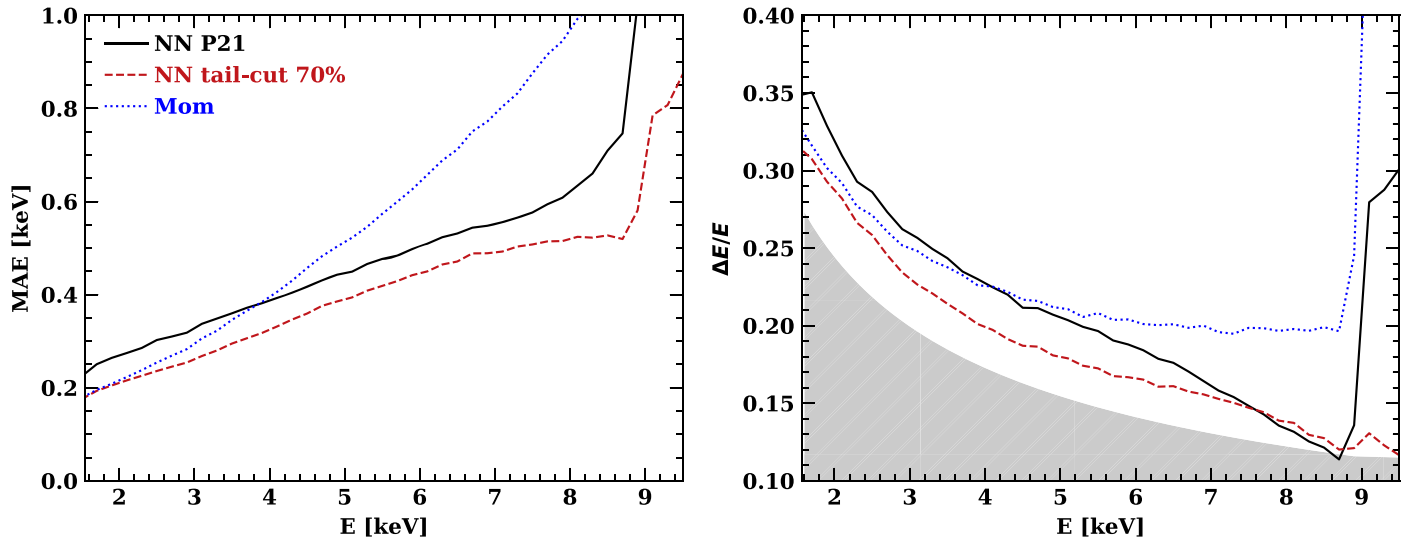
**Figure 6.** Response for three true energies. Note that while the **P21** analysis suppressed the large tail rate seen in the moment processing, events leaked to a high energy tail for medium to low true energies. Our morphological tail cut further suppresses the GEM/window tail, avoids the high energy leakage, and achieves comparable or better peak width.

where two measures of the main peak width are plotted, showing that the filtered NN analysis significantly improves recovery of the true event energy. While the IXPE spectral resolution will hardly be competitive with, e.g., CCD detector spectra of these bright sources, the improved resolution does aid in measuring polarization features associated with different spectral components, as illustrated below.

As expected, this means that the weight performance is also improved. Figure 2 shows that  $\mu_{100}(w)$  (red points) more closely approach the optimal weighting behavior. Following **P21** we can use  $MDP_{99}$  as a measure of the sensitivity. Table 1 illustrates, for two simple power-law spectra passed through IXPE’s energy response, how the

filtered,  $\mu$ -weighted NN analysis improves over both classic moment analysis and the **P21**  $\hat{\sigma}_i^{-\lambda}$  weights.

Note that the tail-suppressed spectral results are for a simple 70% tail probability cut on the event classification. Lower cuts of course improve the purity of peak identification at the cost of total event number (i.e., system effective area). With our present imperfect tail identification, we find that the peak events excluded by tail cuts rapidly increase  $MDP_{99}$ . In fact, for PL2 we find a minimum  $MDP_{99} = 3.73\%$  with a 95% confidence tail cut. This is  $< 0.6\%$  (relative)  $MDP_{99}$  improvement and the cut is too mild to improve the energy PSF. Accordingly, we adopt a 70% cut to significantly improve the spectral results (see Figure (6)) with minimal 2% (relative)



**Figure 7.** *Left:* mean absolute error in predicted energy as a function of true energy. *Right:* FWHM ( $3.46 \times$  median absolute deviation) of predicted energy distribution for a source at energy  $E$ . The gray band marks the limiting energy resolution in the purely exponential multiplication regime. NNs with a 70% tail cut perform better on every metric, more accurate per track and tighter resolution at all energies. All methods suffer from the large increase of tail events above 9 keV.

**Table 1**

Sensitivity Analysis for Two Power-law Spectra ( $dN/dE \sim E^{-N}$ ; PL2 for  $N = 2$ , PL1 for  $N = 1$ ) Each Normalized to Produce  $10^5$  2–8 keV Photons When Folded through IXPE’s Energy Response.

Spectrum	Method	MDP <sub>99</sub> (%)
PL2	Mom. w/ellipticity weights	$4.61 \pm 0.02$
	NN w/P21 $\lambda$ weights	$4.09 \pm 0.02$
	NN w/optimal weights	$3.75 \pm 0.02 \leftarrow$
	NN w/optimal weights 70%	$3.83 \pm 0.02$
PL1	Mom. w/ellipticity weights	$4.15 \pm 0.02$
	NN w/P21 $\lambda$ weights	$3.65 \pm 0.02$
	NN w/optimal weights	$3.38 \pm 0.01 \leftarrow$
	NN w/optimal weights 70%	$3.44 \pm 0.01$

**Note.** MDP<sub>99</sub> gives the sensitivities for the various weights and cuts; smaller MDP<sub>99</sub> is better. Mom. denotes moments analysis weighted by event ellipticity. NN denotes neural net analysis;  $\lambda$  is the approximate weight defined in P21.

MDP<sub>99</sub> loss. An experiment seeking only broadband polarization might impose weaker cuts. Conversely, as we improve the tail cut fidelity, stronger cuts enhance spectral performance with smaller MDP loss. Eventually, weighing with the classification NN result might be preferred, but combining spectral purity with polarization performance leads to a global performance metric whose spectrum-dependent value and error dispersion are not easily characterized. Thus, an *optimal* weight, including prefiltering factors, is difficult to construct. However, we find that our present MDP<sub>99</sub> performance is good and close to minimal for a range of spectral indices, so our construction is near-optimal for our present classification accuracy and  $\kappa$  resolution.

## 6. Astrophysical Spectra

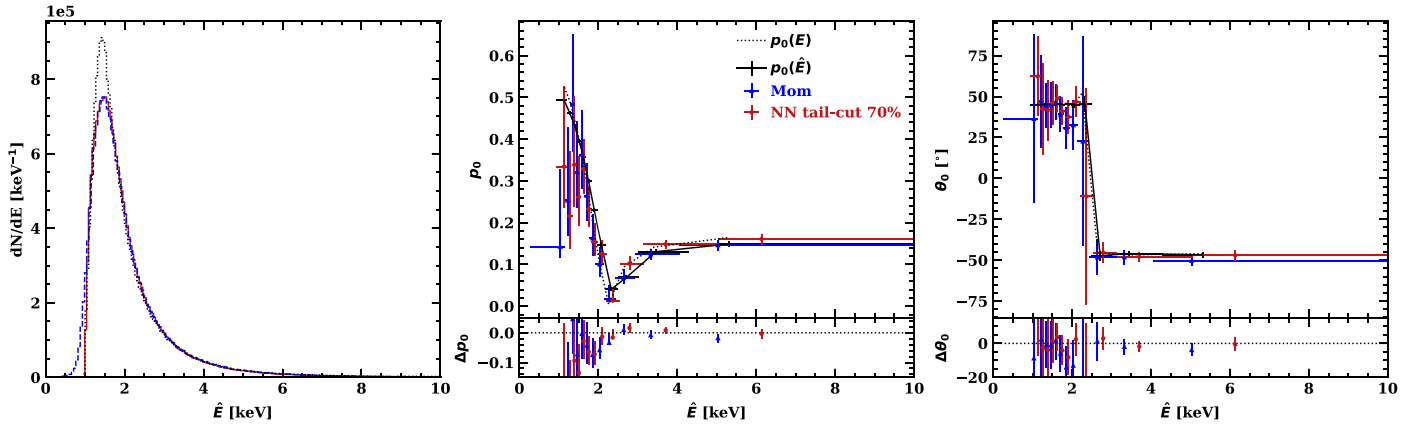
Most astrophysical spectra are moderate index power laws. When observing faint sources with a single emission process across the IXPE band (simple power law), optimal weighting analysis is especially valuable at combining signals from a wide energy range and a wide variety of weights to derive polarization detections or upper limits on weakly polarized

sources. To treat these spectra, we employ  $\mu_{100}(\hat{E})$ , computed for a PL1 spectrum, as in Figure 2. The traditional measure of polarization performance is the 99% upper limit of an  $f_{X(2-8 \text{ keV})} = 10^{-11} \text{ erg cm}^{-2} \text{ s}^{-1}$  power-law source. For connection with earlier results, we present here in Table 1 the performance for  $10^5$  photons collected for two power-law photon indices. We see that the heuristic P21  $\lambda$  weighting provided good sensitivity improvements over the traditional Moments method (even with ellipticity-derived weights). However, weight imperfection in both techniques restricted analysis to events with  $E > 2$  keV. Our near-optimal weighting scheme allows us to extend the analysis to the full 1–10 keV range, while improving the weights themselves; this provides additional MDP<sub>99</sub> decrease, with about half from the increased energy range. Cutting the worst of the tail events can provide a small additional boost in sensitivity. In practice, with our present event classification, dropping even the worst tail events (a 95% cut) provides only a very small  $< 1\%$  MDP<sub>99</sub> decrease. Instead, we find that improvements in energy resolution and spectrum recovery require substantially stronger tail exclusion. Thus, we adopt a 70% tail probability cut, which affords good spectral performance, but does entail a small MDP<sub>99</sub> increase.

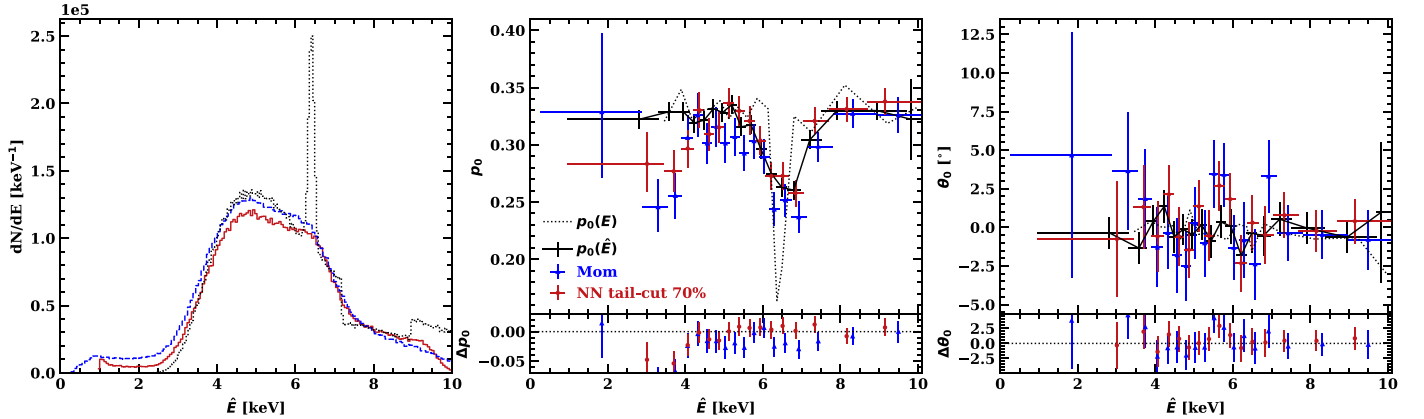
IXPE will also observe sources that are very soft and very hard or have strong cutoff and line features. Although the most sensitive analysis will always involve forward modeling polarized spectra through the IXPE response and constraining model parameters via fits in data space, our revised weighting scheme is sufficiently robust to the varying event energy distributions to allow good analysis of a range of astrophysical spectra. We illustrate with a few examples, representative of very bright sources with extreme spectral parameters that are reasonably handled by our analysis.

### 6.1. Soft, Broken Spectra–Intermediate Synchrotron Peak (ISP) Polarization

One source class of particular interest to IXPE are the ISP blazars. Blazars are jet-dominated active galactic nuclei energies. For ISPs the transition occurs at soft X-ray energies in the IXPE band. With an appropriate cutoff energy, the synchrotron component from a very few zones will dominate in



**Figure 8.** Simulation of an ISP with an exponential cutoff synchrotron spectrum passing to a weakly polarized Compton continuum at high energy. The dotted lines show the count spectrum,  $p_0$  and PA, folded through the IXPE effective area plotted as a function of the true photon energy. The solid lines in each panel show quantities plotted as a function of the recovered photon energy, the count spectra and polarization properties suffer blurring from the methods’ energy redistribution function. The residual plots at bottom show the smaller errors and modestly better recovery of the original polarization.



**Figure 9.** Simulation of a hard, absorbed HMXB similar to GX 301–2. As for Figure 8 the dotted lines show source properties as a function of the true photon energy, while the solid lines show the expected signal, and the various methods’ recovery of this signal as a function of the method-estimate recovered energy. Our NN analysis with tail cut decreases the tail counts at low-energy, and improves the isolation of the  $p_0$  dip associated with the unpolarized Fe K line at 6.4 keV (which is barely seen in the recovered count spectra). The residual plots at bottom show better recovery of PA/ $p_0$  with smaller errors.

the low-energy IXPE band. This exponentially cutoff synchrotron emission can be very steep and very highly polarized (Peirson & Romani 2019). At higher energies, the inverse Compton emission will have a faint hard spectrum and a low  $p_0$  at a different position angle (PA). In Figure 8, we show a model whose highest energy synchrotron component cuts off at 0.25 keV with Compton emission at an orthogonal PA dominating above  $\sim 3$  keV. We contrast the polarization recovery of the standard moments method and our new optimal weight analysis. The total number of counts generated for the spectrum between 1–10 keV is equivalent to an IXPE 10 day exposure of a  $10^{-10}$  erg cm $^{-2}$  s $^{-1}$  2–8 keV source.

For these very bright models IXPE should be able to make good polarization measurements independent of analysis method, and both moments analysis and optimal weighting are expected to give interesting polarization results. Although the optimal weighting scheme shows to greatest advantage when acting on a broad spectral band with a range of weights (as in Table 1), we do see significant gains in some aspects. First the PA/ $p_0$  error bars are, as expected, 20%–30% smaller than for moment analysis. The optimal weighting with cuts also minimizes the artificial low-energy tail to the moments’ polarization signal. The improved energy resolution of optimal weight analysis also better resolves the PA switch near 2 keV.

## 6.2. Hard, Absorbed Spectra–Bright Accretion Powered Binaries

Bright X-ray binaries are also interesting targets for IXPE. In Figure 9, we simulate a source whose spectrum resembles the pulse-average spectrum of the bright  $f_{2-10}$  keV  $\approx 10^{-9}$  erg cm $^{-2}$  s $^{-1}$  HMXB (high mass X-ray binary) GX 301–2. We take spectral parameters from Pravdo et al. (1995); the absorption in the accretion flow is large, leading to a hard spectrum, cut off sharply below 3 keV and we ignore the observed faint low-energy emission. The spectrum also features a strong Fe K $\alpha$  line at 6.41 keV and the absorption edge at 7.11 keV. We assume that the continuum is strongly (1/3) polarized. The Fe K $\alpha$  line is unpolarized except for weak polarization in a Compton shoulder of scattered flux centered at 6.3 keV (which, unfortunately, IXPE does not resolve). For this spectrum the total number of counts between 1 and 10 keV is equivalent to an IXPE 2.5 day exposure of a  $2 \times 10^{-9}$  erg cm $^{-2}$  s $^{-1}$  2–8 keV source. Both moments and NN have some difficulty recovering polarization below 4 keV. This is a result of the very large contamination of tail events at low energies, many arising from GEM conversions above 9 keV.

## 7. Conclusion

We have shown that an optimal weighting of track PAs (of known error distribution) recovers the source polarization with the highest possible sensitivity and resolution-limited independence from the underlying spectral form. In practice, for the IXPE GPD detector to approach this theoretical limit, we needed to extend our track analysis to separate events converting in the detector gas from the corrupted tracks of events converting in the detector walls. This, and our track geometry characterization, are accomplished by an ensemble of convolutional neural nets. Although track characterization is imperfect, we find that our analysis improves polarization sensitivity, approaching the *optimal* limit, while slightly improving energy resolution. In addition, the technique allows us to extend the IXPE sensitive band obtaining useful signals from  $\sim 1$  to  $> 10$  keV.

Applying our analysis to simulated power-law spectra we show 20%–25% sensitivity improvement over the standard (weighted moments) technique and  $\sim 8\%$ –9% improvement over our prior neural net analysis. The robustness of the measurements are illustrated by simulating spectra from especially soft (ISP blazar) and hard (absorbed neutron star binary) spectra from source classes planned for early IXPE observation. The *optimal* weighting shows modestly improved recovery of the model polarized spectrum at both extremes. Of course, these results need to be validated with real flight GPD data, including all the peculiarities of the as-built detectors. But in view of the improvements relative to existing methods, our analysis of simulated GPD events shows that the prospects for

improved imaging polarimetry with IXPE and similar systems are good.

We thank Luca Baldini and the anonymous referee for careful reading of the text, Alessandro Di Marco, Fabio Muleri, and Paolo Soffitta for advice on IXPE calibration, Herman Marshall for the discussion of polarization statistics, and Giorgio Matt for the discussion of source spectra. This work was supported in part by the NASA FINESST program (grant 80NSSC19K1407) and grant NNM17AA26C from the Marshall Space Flight Center.

## ORCID iDs

A. L. Peirson  <https://orcid.org/0000-0001-6292-1911>  
 Roger W. Romani  <https://orcid.org/0000-0001-6711-3286>

## References

- Bellazzini, R., Angelini, F., Baldini, L., et al. 2003, *Proc. SPIE*, **4843**, 383
- He, K., Zhang, X., Ren, S., & Sun, J. 2015, arXiv:1512.03385
- Kislat, F., Clark, B., Beilicke, M., & Krawczynski, H. 2015, *APh*, **68**, 45
- Lakshminarayanan, B., Pritzel, A., & Blundell, C. 2017, in Proc. 31st Int. Conf. on Neural Information Processing Systems, NIPS'17 (Red Hook, NY: Curran Associates Inc.), 6405
- Marshall, H. L. 2021, *ApJ*, **907**, 82
- O'Dell, S. L., Baldini, L., Bellazzini, R., et al. 2018, *Proc. SPIE*, **0699**, 106991X
- Peirson, A. L., & Romani, R. W. 2019, *ApJ*, **885**, 76
- Peirson, A. L., Romani, R. W., Marshall, H. L., Steiner, J. F., & Baldini, L. 2021, *NIMPA*, **986**, 164740
- Pravdo, S. H., Day, C. S. R., Angelini, L., et al. 1995, *ApJ*, **454**, 872
- Weisskopf, M. 2018, *Galax*, **6**, 33