

---

# A deep ensemble approach to X-ray polarimetry

---

**A.L. Peirson\***, **R.W. Romani**  
Kavli Institute for Particle Astrophysics and Cosmology  
Stanford University  
alpv95@stanford.edu

## Abstract

X-ray polarimetry will soon open a new window on the high energy universe with the launch of NASA’s Imaging X-ray Polarimetry Explorer (IXPE). Polarimeters are currently limited by their track reconstruction algorithms, which typically use linear estimators and do not consider individual event quality. We present a modern deep learning method for maximizing the sensitivity of X-ray telescopic observations with imaging polarimeters, with a focus on the gas pixel detectors (GPDs) to be flown on IXPE. We use a weighted maximum likelihood combination of predictions from a deep ensemble of ResNets, trained on Monte Carlo event simulations. We derive and apply the optimal event weighting for maximizing the polarization signal-to-noise ratio (SNR) in track reconstruction algorithms. For typical power-law source spectra, our method improves on the current state of the art, providing a  $\sim 40\%$  decrease in required exposure times for a given SNR.

## 1 Introduction

Measuring X-ray polarization, the degree of order in X-ray electric field oscillations, has been a major goal in astrophysics for the last 50 years. X-ray polarization measurements offer rich opportunities to probe the magnetic field topology and emission physics of high energy astrophysical sources, such as accreting black holes and astrophysical jets [1, 2]. The recent development of photoelectron tracking detectors [3] has greatly improved the prospects of doing so. The gas pixel detector (GPD) [4] has brought soft X-ray polarimetry (1-10 keV) to the PolarLight CubeSat test [5], the scheduled NASA IXPE mission [6], and the potential Chinese mission, eXTP [7].

Imaging X-ray polarization telescopes feed GPDs which directly image charge tracks formed from photoelectrons scattered by incoming X-ray photons. IXPE [2, 8, planned for launch December 9th 2021] will use three co-aligned X-ray telescopes, whose focal planes are imaged by GPDs with hexagonal pixels. IXPE’s sensitivity is limited by the track analysis algorithm used to recover source polarization, spatial structure and energy, given a measured set of electron track images. In the 1 – 10 keV range, the cross-section for photoelectron emission is proportional to  $\cos^2(\theta)$ , where  $\theta$  is the angle between the normal incidence X-ray’s electric vector position angle (EVPA) and the azimuthal emission direction of the photoelectron. By measuring a large number of individual photoelectron emission angles  $\theta$ , one can recover the above distribution to extract the source polarization parameters: polarization fraction ( $0 \leq p_0 \leq 1$ ) and electric vector position angle (EVPA,  $-\pi/2 \leq \phi < \pi/2$ ). In practice, the recovery of photoelectron emission angles from track images is imperfect. Track images are noisy due to Coulomb scattering and diffusion, and, especially for low energies, are often barely resolved. Emission angle estimates are highly heteroskedastic.

The current track reconstruction method for GPDs is a moment analysis described by [3]. Impressive accuracies are achieved from a simple weighted combination of track moments. However, simple moments cannot capture all of the image information, especially for long high energy tracks, and so a

---

\*<https://www.alpeirson.com>

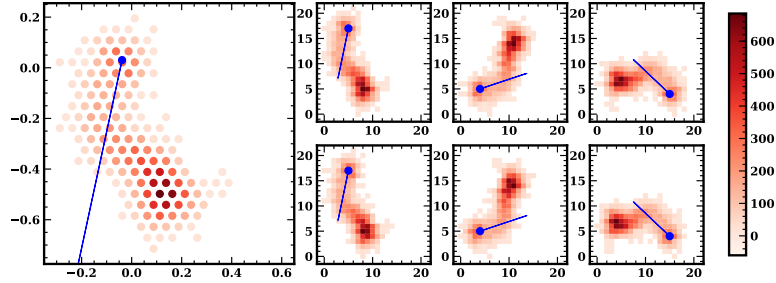


Figure 1: Example square conversions of a 6.4 keV hexagonal track (left panel). The six panels to the right show shifts along the  $120^\circ$  GPD axes; shifting odd rows (upper) or even rows (lower). For each hexagonal track, the ResNet-18s are fed column-wise pairs of square conversions, along with the initial photoelectron direction (blue line) as a label.

more sophisticated image analysis scheme may lead to improved track angle recovery. The problem of recovering polarization parameters from a dataset of (IXPE) electron track images has recently been announced as an open problem in the machine learning community [9].

In this work, we aim to both improve individual event estimates and explicitly model their uncertainty to increase final polarization sensitivity. We demonstrate a two step method: (1) Use a deep ensemble of ResNets [10] to predict electron track angles  $\theta$  and their uncertainties. (2) Combine predicted angles and their uncertainties in a weighted estimator of the polarization parameters  $p_0, \phi$  that maximizes the SNR. Our empirical findings indicate a substantial improvement over the current state-of-the-art track reconstruction [3, 11]. While the results shown here are specific to IXPE’s GPDs, the methods are general, and can be applied to other imaging detector geometries. Extended results, including spatial and energy resolution are discussed in [12, 13].

## 2 Step I: deep ensemble

To extract the angles from individual tracks we use a supervised deep learning technique known as deep ensembles [14]. Deep ensembles are made up of an ensemble of individual neural network models, each trained independently on the same data set to predict the desired output features. Deep ensembles provide estimates of the predictive uncertainty by exploiting different random initializations of the same model at the start of training which leads to widely different prediction functions [15]. In terms of performance and scalability, deep ensembles remain the current state of the art in uncertainty quantification [16]. In our case, we have an image to feature regression problem. Convolutional neural networks have been designed with a spatial inductive bias appropriate for image regression problems. So our deep ensemble will be made up of  $M$  individual ResNet-18s [10].

To make the hexagonal track images admissible inputs to standard ResNet architectures, we first convert the hexagonal images to square image arrays by shifting every other column and rescaling the distance between points, as described in [17]. Since there are two possible shifts (odd and even rows), we apply both and stack the two shifted images, similar to color channels in *rgb* images. We do this to more closely approximate spatial equivariance of the ResNet convolution kernels in the hexagonal space. At test time, we apply the deep ensemble to the same track 3 times, each time rotated by  $120^\circ$  in hexagonal space. We find this reduces all relevant prediction bias on  $\theta$  (and later  $p_0, \phi$ ) introduced when converting from hexagonal to square coordinates. Fig. 1 summarizes this process. An alternative solution is to use a model with native hexagonal convolutions, such as [17]. In practice, we found it more effective and expedient to leverage existing square convolution architectures.

For part (1) of our method, we need to predict emission angle  $\theta$  and its associated uncertainty for each track. Instead of the Gaussian negative log-likelihood (NLL) used in [14], we use the von Mises (VM) distribution NLL as our loss function, the maximum entropy distribution for circular data with specified expectation value. This more appropriately reflects the distribution of our estimators  $\hat{\theta}$ , which are clearly periodic. The VM distribution is parameterized by concentration parameter  $\kappa$ ; for large  $\kappa$  the VM converges to a Gaussian with variance  $\sigma^2 = 1/\kappa$  and for small  $\kappa$  it approaches a uniform distribution. We parameterize the true angle  $\theta$  as a 2D vector  $\mathbf{v} = (\cos 2\theta, \sin 2\theta)$  to capture

the periodicity. Only  $-\pi/2 \leq \hat{\theta} < \pi/2$ , as opposed to  $-\pi \leq \hat{\theta} < \pi$ , is required for polarization estimation since  $-\pi/2 \leq \phi < \pi/2$ ; see eq. 2). Each of the ResNet-18 models in our ensemble computes the loss as

$$\text{Loss}(\hat{\mathbf{v}}, \hat{\kappa}|\mathbf{v}) \propto -\hat{\kappa}(\hat{\mathbf{v}} \cdot \mathbf{v}) + \log I_0(\hat{\kappa}). \quad (1)$$

where  $I_0$  is the modified Bessel function of the first kind. Each ResNet-18 in the ensemble ( $j = 1 : M$ ) outputs the 3-vector  $(\hat{\mathbf{v}}_{ij}, \hat{\kappa}_{ij})$  for track  $i$ . Then  $\hat{\theta}_{ij} = \arctan(\frac{\hat{v}_{ij2}}{\hat{v}_{ij1}})/2$  and  $\hat{\kappa}_{ij}$  defines the predicted aleatoric uncertainty. The final track angle prediction  $\hat{\theta}_i$  is the circular average over the ensemble predictions  $j = 1 : M$ . The epistemic uncertainty on  $\hat{\theta}_i$  is also assumed to follow a VM distribution, with concentration parameter  $\kappa_i^e$ ; we estimate it using the appropriate maximum likelihood estimator for  $\hat{\kappa}_i^e$  given the independent sample set  $\{\theta_{ij}\}_{j=1}^M$ , see appendix. The total predictive error on each track angle  $\hat{\theta}_i$  is given by summing the aleatoric and epistemic variances:  $1/\hat{\kappa}_i^{\text{tot}} = (1/M \sum_j 1/\hat{\kappa}_{ij}) + 1/\hat{\kappa}_i^e$

**Data.** Our initial dataset consists of 3.5 million GEANT4 [18] Monte Carlo simulated tracks, where each track is labelled with its 2D emission angle vector, an example of which is shown in fig. 1. The track energies uniformly span 1.0 – 10.0 keV, IXPE’s most sensitive range, and are unpolarized (uniform track angle distribution). Since we don’t know the true event energy, we want a model that can make predictions for tracks of all energies. We have confirmed the simulated track data matches real flight detector data to extremely high precision, and that our method is robust to any remaining covariate shift, maintaining the relative improvement shown in §4 for real detector data.

**Training.** We apply pixelwise normalization to the square track images. Each ResNet-18 model is trained with a Momentum Optimizer, and a step-wise learning rate beginning at  $1 \times 10^{-3}$  and ending at  $1 \times 10^{-5}$  on 2 NVIDIA GeForce RTX 2080 GPUs. ResNets with batch size 512, 1024, 2048 are all considered for selection in the final ensemble. We randomly select  $M = 10$  trained ResNets to compose the final ensemble, whose distributed results estimate the epistemic error.

### 3 Step II: polarization estimation

The basic problem is to estimate  $p_0$  and  $\phi$  from a set of measured track angles  $\{\hat{\theta}_i\}_{i=1}^N$ . As described in the introduction, true track angles exhibit a sinusoidal modulation with period  $\pi$

$$p(\theta|p_0, \phi) = \frac{1}{2\pi} (1 + p_0 \cos[2(\theta - \phi)]) \quad (2)$$

where  $0 \leq p_0 \leq 1$ ,  $-\pi/2 \leq \phi < \pi/2$  and  $-\pi \leq \theta < \pi$ . One could estimate  $(p_0, \phi)$  using the maximum likelihood estimator (MLE) for eq.2. Equivalently, [19] have shown that the minimum variance unbiased estimator for  $(p_0, \phi)$  is given by

$$\hat{p}_0 = \sqrt{\hat{Q}^2 + \hat{U}^2}, \quad (3)$$

$$\hat{\phi} = \frac{1}{2} \arctan \frac{\hat{U}}{\hat{Q}}. \quad (4)$$

where  $\hat{I} = N$ ,  $\hat{Q} = \frac{1}{I} \sum_{i=1}^N 2 \cos 2\hat{\theta}_i$ , and  $\hat{U} = \frac{1}{I} \sum_{i=1}^N 2 \sin 2\hat{\theta}_i$ . However, since  $\hat{\theta}_i$  are imperfectly recovered, we adjust eq. 2 to include the VM uncertainty on our estimates,  $\hat{\theta} = \theta + \text{VM}(0, \kappa)$ . Computing the convolution of eq. 2 with the VM, we find:

$$p(\hat{\theta}|p_0, \phi) = \frac{1}{2\pi} (1 + p_0 \frac{I_1(\kappa)}{I_0(\kappa)} \cos[2(\hat{\theta} - \phi)]) \quad (5)$$

In other words,  $\hat{\theta}$  follow the same distribution as  $\theta$  but smeared by a modulation factor  $\mu = I_1(\kappa)/I_0(\kappa)$ . Current analyses [11, 9] treat  $\mu$  as constant for all tracks and calculate it based on broadband calibration measurements: no connection to individual track predictions. In effect, they assume homoskedastic emission angle measurements. Here, we explicitly model the heteroskedasticity of our predictions  $\hat{\theta}$  by using a deep ensemble (§2) to estimate event uncertainties  $\hat{\kappa}$  and include them in the likelihood for final  $(p_0, \phi)$  predictions.

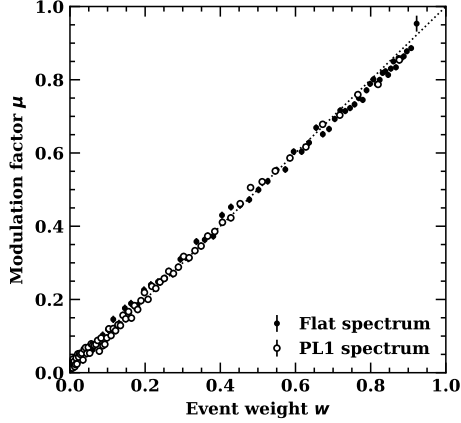


Figure 2: Measured modulation factor  $\mu$  as a function of predicted track weight  $w = I_1(\hat{\kappa})/I_0(\hat{\kappa})$  for large test data sets of 1 – 10 keV simulated events. Each  $\mu$  bin is calculated from 20,000 individual track events. Open circles represent a PL1 source, closed circles a flat spectrum.

Table 1: Sensitivity analysis for  $10^5$  2-8 keV photons with a  $dN/dE \sim E^{-1}$  spectrum.

Method	MDP <sub>99</sub> (%)
Moment Analysis	$4.45 \pm 0.02$
Deep ensemble	$4.21 \pm 0.02$
Deep ensemble w/ weights	<b><math>3.38 \pm 0.01</math></b>

Instead of explicitly maximizing a likelihood function based on eq.5 and our predictions  $\{\hat{\theta}_i, \hat{\kappa}_i\}_{i=1}^N$  to estimate  $(p_0, \phi)$ , we can define event weights

$$w_i = I_1(\hat{\kappa}_i)/I_0(\hat{\kappa}_i), \quad (6)$$

and use eq.3-4 by defining

$$\hat{I} = N_{\text{eff}} = \frac{(\sum_{i=1}^N w_i)^2}{\sum_{i=1}^N w_i^2}, \quad (7)$$

$$\hat{Q} = \frac{1}{\hat{I}} \sum_{i=1}^N 2w_i \cos 2\hat{\theta}_i, \quad (8)$$

$$\hat{U} = \frac{1}{\hat{I}} \sum_{i=1}^N 2w_i \sin 2\hat{\theta}_i. \quad (9)$$

As  $N$  approaches infinity, this converges to the MLE [19]. The posterior distribution  $p(p_0, \phi | \hat{p}_0, \hat{\phi})$  is given in appendix §A.2.

**Figure of merit.** We require a figure-of-merit to define the quality of polarization reconstruction using different methods. Aligned with prior work, we adopt the standard figure-of-merit used in X-ray polarimetry: minimum detectable polarization (MDP) [20]. MDP<sub>99</sub> is the polarization fraction that has a 1% probability of being exceeded by chance for an unpolarized ( $p_0 = 0$ ) source. This can be found by integrating the posterior distribution (see appendix):  $\text{MDP}_{99} = 4.29/(\mu\sqrt{N})$  for an unweighted polarization estimate. Here  $\mu$  is the empirical modulation factor for the entire set of track observations (i.e.  $\mu = \hat{p}_0$  for  $p_0 = 1$ ). For a weighted estimate  $N$  is replaced with  $N_{\text{eff}}$  (eq. 7) [19]. MDP<sub>99</sub> is effectively a (inverse) ratio of recovered signal  $\mu$  to noise  $\sim 1/\sqrt{N}$ . For a given detector, track reconstruction algorithms with lower MDP<sub>99</sub> are better. In the appendix, we prove our chosen weighting minimizes the MDP<sub>99</sub> (maximizes SNR).

## 4 Results and discussion

We evaluate our two step method and the standard moment analysis on a power law 1 dataset with  $N = 10^5$  tracks. Once emission angles are estimated, all methods use eqs.3-4 for polarization prediction. To further strengthen our analysis, we ablate using predicted uncertainties weights. Table 1 shows the results. Our weighted approach using deep ensemble predictions yields significant improvement over the standard analysis, with a  $1 - 3.38/4.45 \sim 24\%$  decrease in  $MDP_{99}$ . The equivalent decrease in IXPE's required exposure times to reach the same SNR is  $1 - (3.38/4.45)^2 \sim 40\%$ . Our method without any uncertainty weighting provides a significantly smaller improvement. The gain in sensitivity of our method comes in a small part from an improvement in  $\hat{\theta}$  accuracy, and in a larger part from properly modelling  $\hat{\theta}$  heteroskedasticity, using deep ensemble predicted uncertainties as appropriate event weights. In fig. 2 we assess the trustworthiness of our deep ensemble predicted VM uncertainties. We find the empirical modulation factor matches the deep ensemble predicted modulation factor (event weights) very closely for a selection of test spectra. In the appendix, we visualize our deep ensemble  $\hat{\theta}$  at multiple energies to assess any prediction bias.

Our results on real (non-simulated) detector data indicates similar exposure time reduction. A decrease in exposure time of 40% adds public value since NASA-funded projects like IXPE can observe significantly more sources, at better SNR, over the mission lifetime. We expect our approach of using deep ensemble predicted uncertainties to take into account heteroskedasticity could be applied to other problems in engineering and the physical sciences, for example in high energy particle physics: fitting Cauchy distributions to the frequencies of noisy (deep learning) measured decay states.

## Acknowledgments and Disclosure of Funding

We would like to thank Christopher Fifty for a thorough read-through of the manuscript and the referees for helpful and constructive comments. This work was supported in part by the NASA FINESST program (grant 80NSSC19K1407) and grant NNM17AA26C from the Marshall Space Flight Center.

## References

- [1] Krawczynski, H., Matt, G., Ingram, A.R., Taverna, R., Turolla, R. et al. Using X-Ray Polarimetry to Probe the Physics of Black Holes and Neutron Stars. 51:150, May 2019. Conference Name: Bulletin of the American Astronomical Society.
- [2] Weisskopf, M. An Overview of X-Ray Polarimetry of Astronomical Sources. *Galaxies*, 6:33, March 2018.
- [3] Bellazzini, R., Angelini, F., Baldini, L., Brez, A., Costa, E. et al. Novel gaseous x-ray polarimeter: data analysis and simulation. In *Polarimetry in Astronomy*, volume 4843, pages 383–393. International Society for Optics and Photonics, February 2003.
- [4] Bellazzini, R., Spandre, G., Minuti, M., Baldini, L., Brez, A. et al. A sealed Gas Pixel Detector for X-ray astronomy. *Nuclear Instruments and Methods in Physics Research A*, 579:853–858, September 2007.
- [5] Feng, H. and Bellazzini, R. The X-ray polarimetry window reopens. *Nature Astronomy*, 4(5):547–547, May 2020. Number: 5 Publisher: Nature Publishing Group.
- [6] Sgrò, C. and IXPE Team. The Imaging X-ray Polarimetry Explorer (IXPE). *Nuclear Instruments and Methods in Physics Research A*, 936:212–215, August 2019.
- [7] Zhang, S.N., Feroci, M., Santangelo, A., Dong, Y.W., Feng, H. et al. eXTP: Enhanced X-ray Timing and Polarization mission. In *Space Telescopes and Instrumentation 2016: Ultraviolet to Gamma Ray*, volume 9905, page 99051Q. International Society for Optics and Photonics, January 2017.

- [8] O’Dell, S.L., Baldini, L., Bellazzini, R., Costa, E., Elsner, R.F. et al. The Imaging X-ray Polarimetry Explorer (IXPE): technical overview. 0699:106991X, August 2018. Conference Name: Space Telescopes and Instrumentation 2018: Ultraviolet to Gamma Ray.
- [9] Moriakov, N., Samudre, A., Negro, M., Gieseke, F., Otten, S. and Hendriks, L. Inferring astrophysical X-ray polarization with deep learning. *arXiv:2005.08126 [astro-ph]*, May 2020. arXiv: 2005.08126.
- [10] He, K., Zhang, X., Ren, S. and Sun, J. Deep Residual Learning for Image Recognition. *arXiv:1512.03385 [cs]*, December 2015. arXiv: 1512.03385.
- [11] Kitaguchi, T., Black, K., Enoto, T., Hayato, A., Hill, J.E., Iwakiri, W.B., Kaaret, P., Mizuno, T. and Tamagawa, T. A convolutional neural network approach for reconstructing polarization information of photoelectric X-ray polarimeters. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 942:162389, October 2019. arXiv: 1907.06442.
- [12] Peirson, A.L., Romani, R.W., Marshall, H.L., Steiner, J.F. and Baldini, L. Deep ensemble analysis for Imaging X-ray Polarimetry. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 986:164740, January 2021.
- [13] Peirson, A.L. and Romani, R.W. Towards Optimal Signal Extraction for Imaging X-ray Polarimetry. *arXiv:2107.08289 [astro-ph]*, July 2021. arXiv: 2107.08289.
- [14] Lakshminarayanan, B., Pritzel, A. and Blundell, C. Simple and scalable predictive uncertainty estimation using deep ensembles. In *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS’17*, pages 6405–6416, Long Beach, California, USA, December 2017. Curran Associates Inc.
- [15] Fort, S., Hu, H. and Lakshminarayanan, B. Deep Ensembles: A Loss Landscape Perspective. *arXiv:1912.02757 [cs, stat]*, December 2019. arXiv: 1912.02757.
- [16] Hoffmann, L. and Elster, C. Deep Ensembles from a Bayesian Perspective. *arXiv:2105.13283 [cs, stat]*, May 2021. arXiv: 2105.13283.
- [17] Steppa, C. and Holch, T.L. HexagDLy - Processing hexagonally sampled data with CNNs in PyTorch. *SoftwareX*, 9:193–198, January 2019. arXiv: 1903.01814.
- [18] Agostinelli, S., Allison, J., Amako, K., Apostolakis, J., Araujo, H. et al. Geant4—a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250–303, July 2003.
- [19] Kislak, F., Clark, B., Beilicke, M. and Krawczynski, H. Analyzing the data from X-ray polarimeters with Stokes parameters. *Astroparticle Physics*, 68:45–51, August 2015.
- [20] Weisskopf, M.C., Elsner, R.F. and O’Dell, S.L. On understanding the figures of merit for detection and measurement of x-ray polarization. *arXiv:1006.3711 [astro-ph]*, page 77320E, July 2010. arXiv: 1006.3711.

## Checklist

1. For all authors...
  - (a) Do the main claims made in the abstract and introduction accurately reflect the paper’s contributions and scope? [\[Yes\]](#)
  - (b) Did you describe the limitations of your work? [\[Yes\]](#)
  - (c) Did you discuss any potential negative societal impacts of your work? [\[N/A\]](#)
  - (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [\[Yes\]](#)
2. If you are including theoretical results...

- (a) Did you state the full set of assumptions of all theoretical results? [Yes]
  - (b) Did you include complete proofs of all theoretical results? [Yes]
3. If you ran experiments...
- (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [No] These will be provided after IXPE’s launch, December 9th
  - (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes]
  - (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes]
  - (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
- (a) If your work uses existing assets, did you cite the creators? [Yes]
  - (b) Did you mention the license of the assets? [N/A]
  - (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]
  - (d) Did you discuss whether and how consent was obtained from people whose data you’re using/curating? [N/A]
  - (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
- (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
  - (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
  - (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

## A Appendix

### A.1 Epistemic error estimator

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  $M$  models  $\{\hat{\theta}_{ij}\}_{j=1}^M$ :

$$\bar{R}_i^2 = \left( \frac{1}{N} \sum_{j=1}^M \cos 2\hat{\theta}_{ij} \right)^2 + \left( \frac{1}{N} \sum_{j=1}^M \sin 2\hat{\theta}_{ij} \right)^2 \quad (10)$$

$$\frac{I_1(\hat{\kappa}_i^e)}{I_0(\hat{\kappa}_i^e)} = \bar{R}_i, \quad (11)$$

with the modified Bessel functions  $I_0$  and  $I_1$ .

## A.2 Posterior distribution on $p_0, \phi$

For a weighted polarization estimate with observations  $\{\hat{\theta}\}_{i=1}^N$  and weights  $\{w_i\}_{i=1}^N$  using eqs. 3–4, [19] derive the joint error distribution (posterior) for the polarization fraction and EVPA estimators as

$$p(p_0, \phi | \hat{p}_0, \hat{\phi}) = \frac{\sqrt{N_{\text{eff}} \hat{p}_0 \mu^2}}{2\pi\sigma} \times \exp \left[ -\frac{\mu^2}{4\sigma^2} \left\{ \hat{p}_0^2 + p_0^2 - 2\hat{p}_0 p_0 \cos(2(\hat{\phi} - \phi)) - \frac{\hat{p}_0^2 p_0^2 \mu^2}{2} \sin^2(2(\hat{\phi} - \phi)) \right\} \right], \quad (12)$$

where

$$\sigma = \sqrt{\frac{1}{N_{\text{eff}}} \left( 1 - \frac{\hat{p}_0^2 \mu^2}{2} \right)}. \quad (13)$$

and

$$N_{\text{eff}} = \frac{(\sum_{i=1}^N w_i)^2}{\sum_{i=1}^N w_i^2} \quad (14)$$

This assumes a uniform prior over  $(p_0, \phi)$ . Here,  $\mu$  is the empirical modulation factor for the entire set of track observations (i.e.  $\mu = \hat{p}_0$  for  $p_0 = 1$ ).  $\mu$  is effectively the instrument polarization response.

Confidence intervals and the  $\text{MDP}_{99}$  (§A.3, the 99% upper limit for  $p_0 = 0$ ) are derived from this posterior probability distribution. In cases where  $\mu$  and  $p_0$  are not close to 0, the Gaussian approximation for the marginalized errors below is sufficient

$$\sigma(p_0) \approx \sqrt{\frac{2 - \hat{p}_0^2 \mu^2}{(N_{\text{eff}} - 1)\mu^2}}, \quad (15)$$

$$\sigma(\phi) \approx \frac{1}{\hat{p}_0 \mu \sqrt{2(N_{\text{eff}} - 1)}}. \quad (16)$$

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

## A.3 Minimum detectable polarization ( $\text{MDP}_{99}$ )

The  $\text{MDP}_{99}$  is calculated for an unpolarized  $p_0 = 0$  source and is given by (using posterior eq. 12)

$$\int_0^{\text{MDP}_{99}} \int_{-\pi/2}^{\pi/2} p(p_0, \phi | \hat{p}_0, \hat{\phi}) d\phi dp_0 = 0.99 \quad (17)$$

$$\text{MDP}_{99} = \frac{4.29}{\mu \sqrt{N_{\text{eff}}}} \quad (18)$$

## A.4 Maximising the SNR

We define the signal-to-noise ratio (SNR)

$$\text{SNR} \propto 1/\text{MDP}_{99} \propto \mu \sqrt{N_{\text{eff}}}. \quad (19)$$

This is simply the inverse of the  $\text{MDP}_{99}$  (without constants); an optimal weighting scheme should maximize the SNR for a fixed number of events  $N$ . We can expand the SNR explicitly using our weighted estimators from §3,

$$\text{SNR} \propto \sqrt{\frac{\left( \sum_{i=1}^N 2w_i \cos 2\hat{\theta}_i \right)^2 + \left( \sum_{i=1}^N 2w_i \sin 2\hat{\theta}_i \right)^2}{\sum_{i=1}^N w_i^2}}. \quad (20)$$



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

$$\text{SNR}^2 \propto \frac{\sum_{i,j,i \neq j} w_i w_j \cos 2(\hat{\theta}_i - \hat{\theta}_j)}{\sum_k w_k^2}. \quad (21)$$

The estimators  $\hat{\theta}_i$  are random variables. The true values  $\theta_i$ , also random variables, follow the distribution eq. 2 (since they are perfectly known). Assuming the  $\hat{\theta}_i$  are unbiased estimators of  $\theta_i$  (approximately true for both moment analysis and ResNet-18s, §A.5) we can say

$$\hat{\theta}_i = \theta_i + \epsilon_i \quad (22)$$

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

$$\mathbb{E}[\epsilon_i] = 0, \text{Var}[\epsilon_i] = \sigma_i^2. \quad (23)$$

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

$$\hat{\theta}_i \sim \frac{1}{2\pi} (1 + \mu_i p_0 \cos[2(\hat{\theta}_i - \phi)]), \quad (24)$$

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 decrease the sinusoidal modulation by a factor  $\mu_i$  for the specific event  $i$ .

Knowing the distributions of  $\hat{\theta}_i$ , eq. 24, we take the expectation over  $\text{SNR}^2$  (dropping constant  $p_0$ )

$$\mathbb{E} [\text{SNR}^2] \propto \frac{\sum_{i,j,i \neq j} w_i w_j \mu_i \mu_j}{\sum_k w_k^2}. \quad (25)$$

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

$$w_i = \mu_i, \quad (26)$$

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

Under the assumption of von Mises distributed  $\hat{\theta}$  errors

$$\epsilon_i \sim \text{VM}(0, \kappa), \quad (27)$$

we have shown in §3 that

$$w_i = \mu_i = \frac{I_1(\kappa_i)}{I_0(\kappa_i)}. \quad (28)$$

These weights maximize the SNR.

## A.5 Recovered emission angles

We show the recovered  $\hat{\theta}$  distribution for our deep ensemble and the moment analysis at two example energies in fig.3. The unpolarized examples show negligible residual polarization in our method while the polarized examples suggest an increase in deep ensemble polarization signal ( $\mu$ ), especially at higher energies. As described in the main text, the improvement in the actual polarization estimator is even greater when weighting using event uncertainties. All plots suggest a lack of  $\hat{\theta}$  prediction bias in our method.

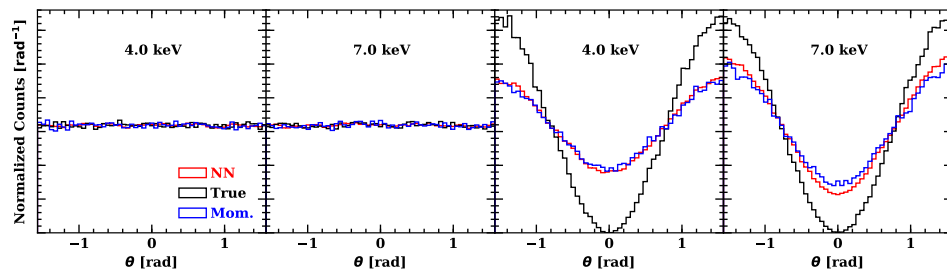


Figure 3: Track angle recovery for unpolarized,  $p_0 = 0$ , (left two panels) and 100% polarized,  $p_0 = 1$ , (right two panels) simulated data for 4.0 and 7.0 keV. The true photoelectron angle distribution is shown in black; standard moment analysis reconstruction is in blue and ResNet-18 deep ensemble in red.