# Fast Integration of Poisson Distributions for Dead Sensor Marginalization


Zoë Bilodeau [1], Christina Peters [2,3], Christopher D. Tunnell [3,4]

[1]Skidmore College [2]Department of Computer and Information Sciences, University of Delaware
[3]Department of Physics and Astronomy, Rice University [4]Department of Computer Science, Rice University


## The XENONnT Experiment

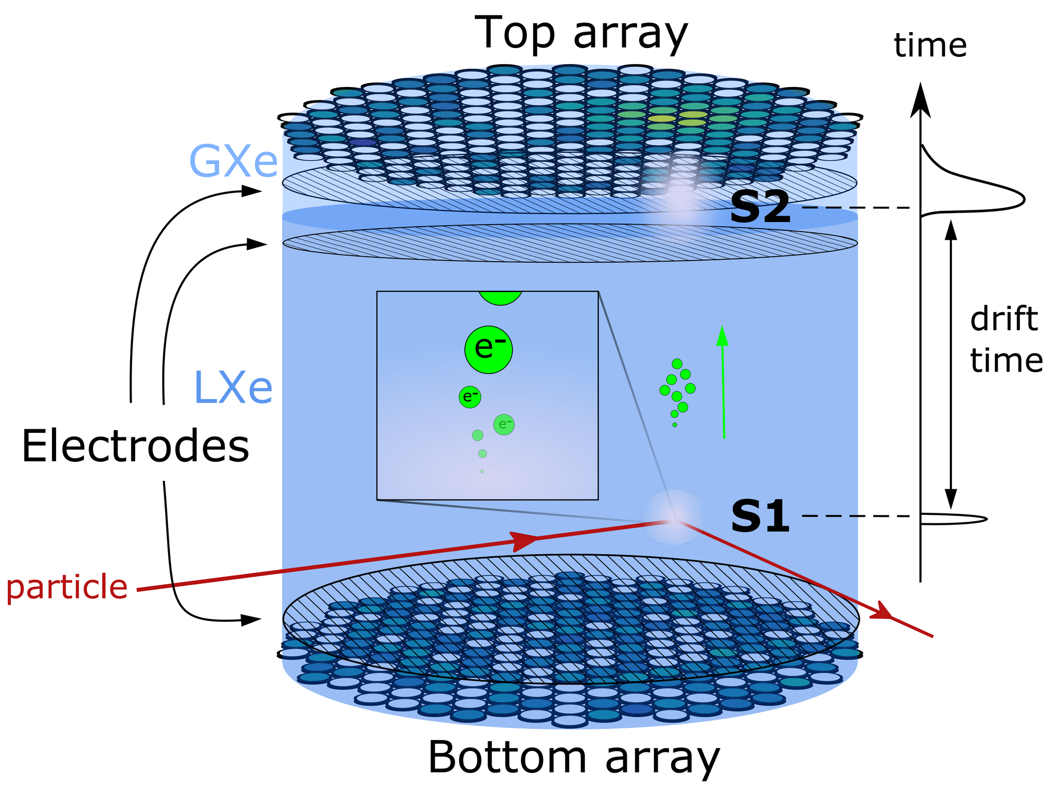


Figure 1. A schematic of the working principle of a dual-phase liquid xenon TPC detector. Credit: [1].

- XENONnT (xenonexperiment.org) is an experiment that uses a dual-phase time projection chamber designed to detect dark matter particles
- The detector, as shown in Figure 1, is filled with liquid and gaseous xenon, which interacts with particles passing through the detector.
- Sensor readings and be used to determine the particle type and position of interaction, and are designed with the intention of identifying dark matter.

## The Problem: Broken Sensors

- Malfunctioning and deactivated sensors leave gaps in data.
- Goal: estimate the number of photons broken sensors would have detected.

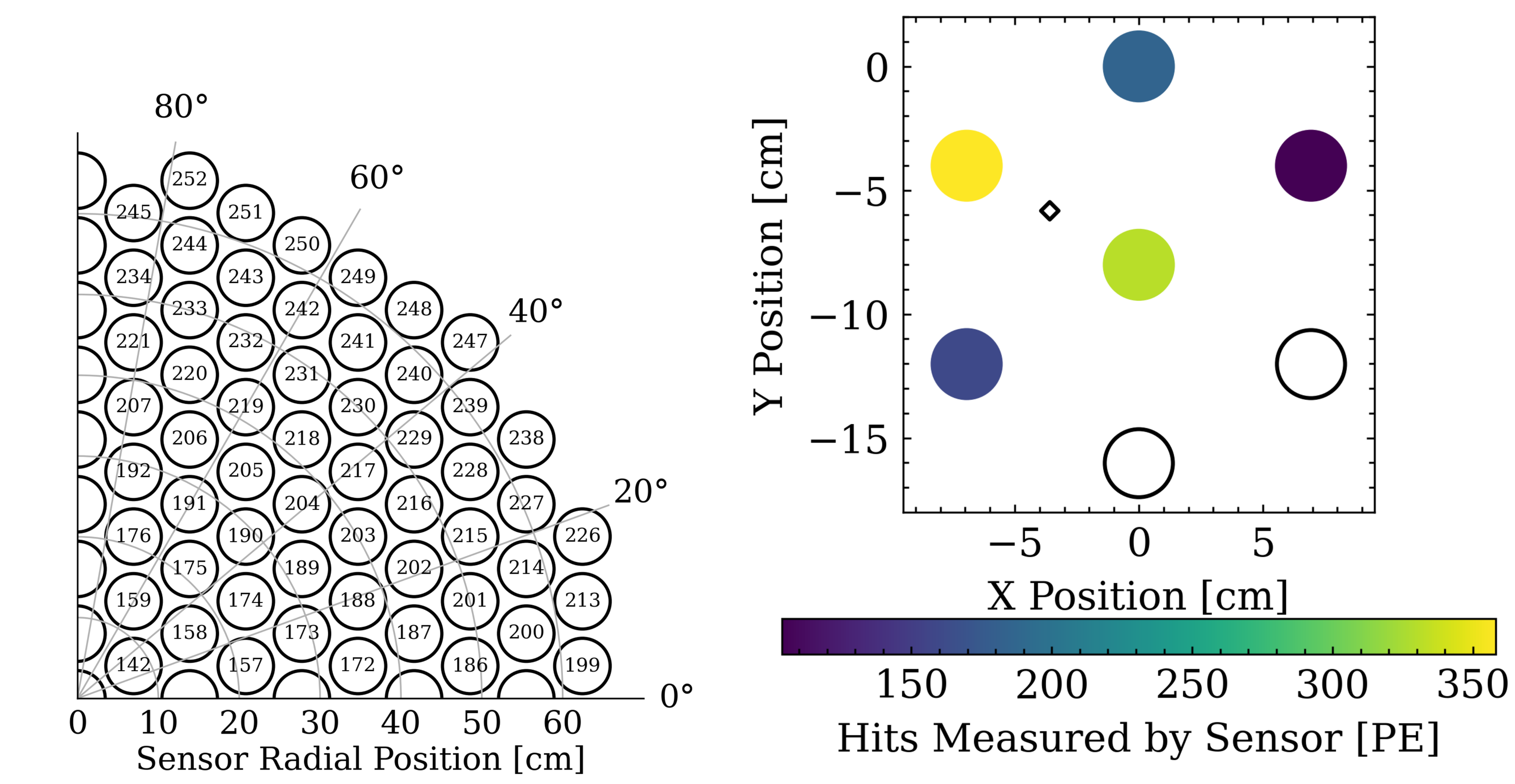


Figure 2. Left: Sensor positions in XENONnT. Credit: [4]. Right: Example simulated hit pattern with broken sensors, with interaction position indicated by black diamond.

- Gaps in data increase uncertainty of infered particle type and position of interaction.
- Goal: complete the Bayesian Network [6, 3] for position reconstruction from [4] shown in Figure 3 to account for dependencies between sensors.
- Goal: Allow positional reconstruction algorithms to run without special cases or retraining to account for broken sensors.

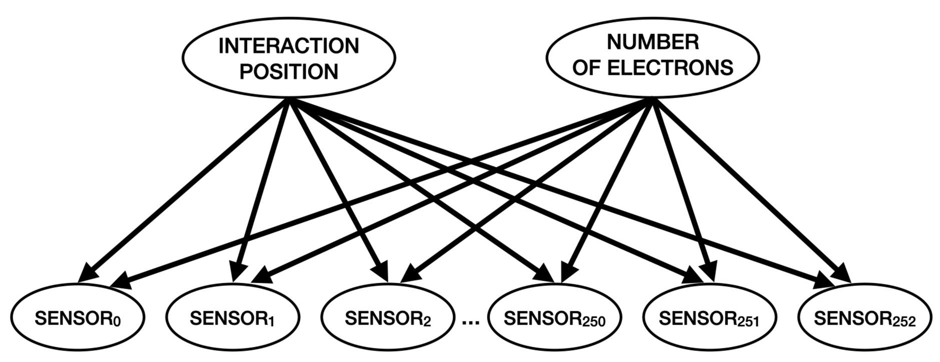


Figure 3. Structure of a Bayesian network for position reconstruction. Credit: [4].

## Proposed Solution: Estimate Missing Data

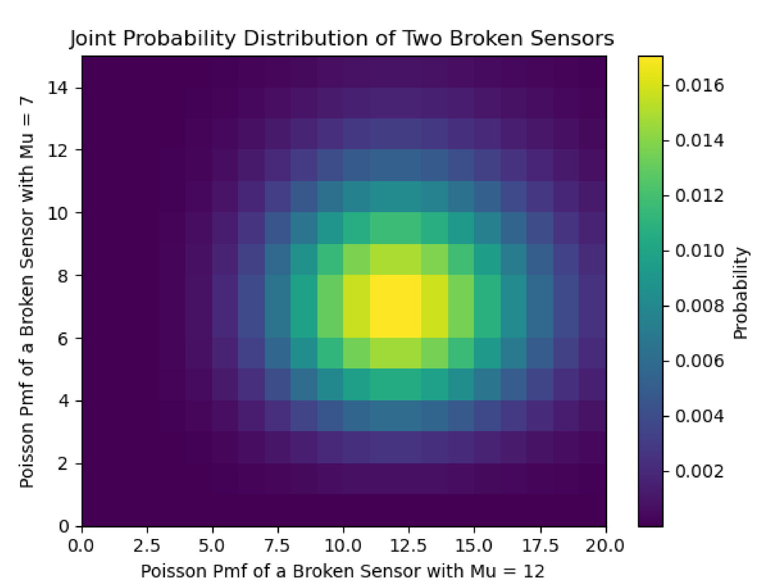


Figure 4. Bivariate Poisson distribution.

**Calculating Correlations:** Calculate **multivariate Poisson distribution** for groups of 7 adjacent sensors in the top array of sensors.
**Bivariate Poisson Distribution:** Consider correlations between two sensors using Equation 1, where $k$ is a number of photons a sensor may have detected and $\mu$ is the mean number of photons likely detected in the range of possible $k$ values. An example distribution is shown in Figure 4.
**Using the Distribution:** Calculate the joint probability distribution over both interaction position and the number of photons detected by the sensors to allow for inference of interaction position.

$$P(k_0, k_1|\mu_0, \mu_1, \mu_{0,1}) = e^{(-\mu_0 - \mu_1 - \mu_{0,1})}\frac{\mu_0^{k_0}}{k_0!}\frac{\mu_1^{k_1}}{k_1!}\sum_{k=0}^{min(k_0,k_1)}\frac{k_1!}{k!(k_1-k)!}\frac{k_2!}{k!(k_2-k)!}k!\left(\frac{\mu_{0,1}}{\mu_0\mu_1}\right)^k \quad (1)$$

**Problems Calculating the Distribution:**

- **Factorial overflow** occurs at k = 120, problematic as the number of photons detected by a sensor can be as large as ≈ 410.
- **Unreasonable storage requirements** distribution shape can be as large as $410^7$, not including the interaction position dimension.

## Strategies for Calculating Multivariate Poisson

**Addressing overflow / underflow:**

- Split up the multivariate calculations into two parts:
- First part: **univariate Poisson distribution** for each sensor

$$P(k_0|\mu_0) = e^{-\mu_0}\frac{\mu_0{}^{k_0}}{k_0!} \quad (2)$$

- Second part: **the summation term**

$$\sum_{k=0}^{min(k_0,k_1)}\frac{k_1!}{k!(k_1-k)!}\frac{k_2!}{k!(k_2-k)!}k!\left(\frac{\mu_{0,1}}{\mu_0\mu_1}\right)^k \quad (3)$$

- **log-probability**: calculating the log-probability distribution decreases the likelihood of underflow and loss of precision, as the probabilities can be added together instead of multiplied.
- Avoid overflow in factorial using **Ramanujan's log-factorial approximation** [5]:

$$x! \approx x\ln(x) - x + \frac{\ln(x(1+4x(1+2x)))}{6} + \frac{\ln(\pi)}{2} \quad (4)$$

**Reducing storage requirements:** To lessen the storage required for these calculations, we used the Python package **Zarr** [2] This package compresses Numpy arrays in chunks which are only uncompressed when they are to be used.

## Initial Results

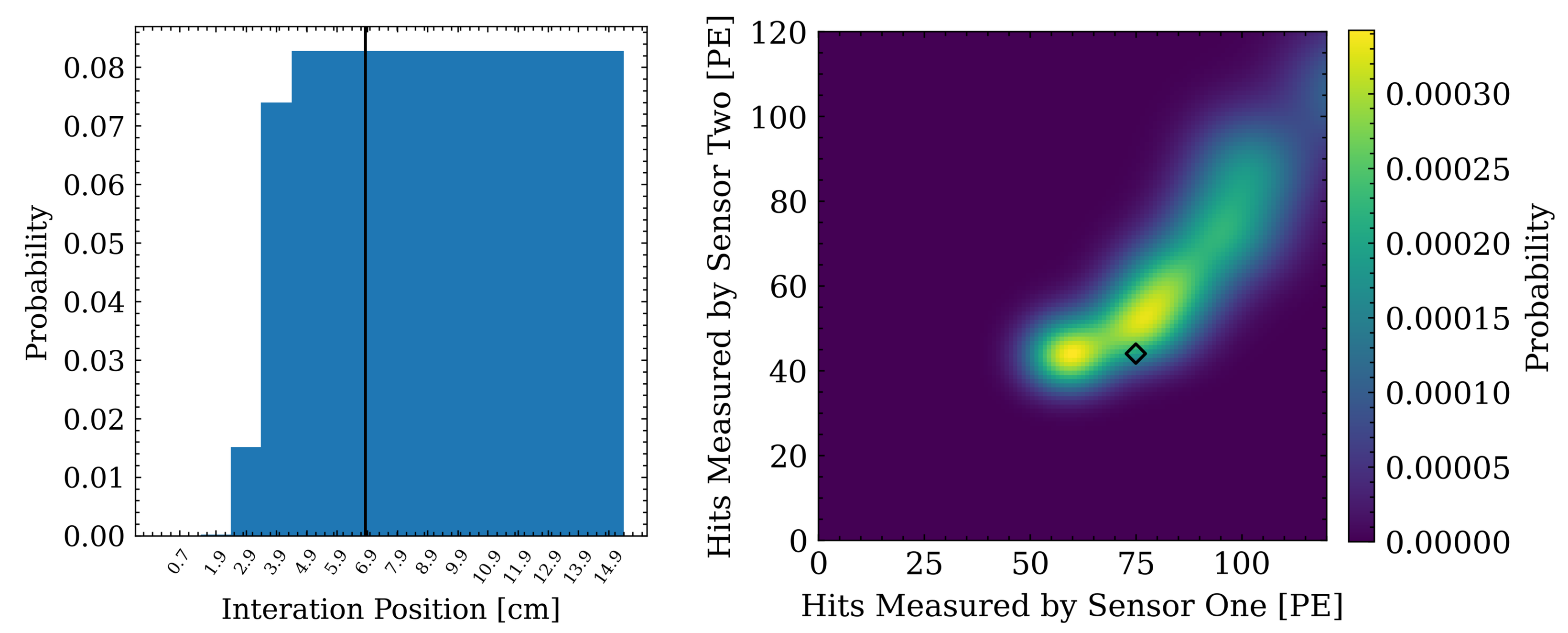


Figure 5. Left: Probability distribution over interaction position and Right: joint probability distribution over photons detected by the broken sensors, both for the example simulated hit pattern shown in Figure 2 [Right]. The vertical black line and black diamond indicate the true values.

**Discussion:**

- Probabilities nearly center around true interaction positions
- Distributions that range over 200 flatten, likely underflow

## Looking Ahead

- Address underflow in large probability distributions
- Proper benchmark testing for run-time and space requirements
- Improvements depending on benchmark results
- Lookup table for common data to improve speed
- Potentially further work to optimize Zarr compression

## Acknowledgements


Thank you to IRIS-HEP (iris-hep.org) for the introduction to this project and for funding my work through the summer and to the Skidmore College Opportunity Fund for supporting this project.


## References


[1] S. Liang, A. Higuera, C. Peters, V. Roy, W. U. Bajwa, H. Shatkay, and C. D. Tunnell. Domain-informed neural networks for interaction localization within astroparticle experiments. *Front. Artif. Intell.*, 2022.

[2] Alistair Miles, John Kirkham, Martin Durant, James Bourbeau, Tarik Onalan, Joe Hamman, Zain Patel, shikharsg, Matthew Rocklin, raphael dussin, Vincent Schut, Elliott Sales de Andrade, Ryan Abernathey, Charles Noyes, sbalmer, pyup.io bot, Tommy Tran, Stephan Saalfeld, Justin Swaney, Josh Moore, Joe Jevnik, Jerome Kelleher, Jan Funke, George Sakkis, Chris Barnes, and Anderson Banihirwe. zarr-developers/zarr-python: v2.4.0, January 2020.

[3] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan-Kaufmann, 1988.

[4] C. Peters, A. Higuera, S. Liang, V. Roy, W. U. Bajwa, H. Shatkay, and C. D. Tunnell. A Method for Quantifying Position Reconstruction Uncertainty in Astroparticle Physics using Bayesian Networks. *arXiv e-prints*, page arXiv:2205.10305, May 2022.

[5] S. Ramanujan. *The Lost Notebook and other Unpublished Papers.* Springer, 1987.

[6] T. Verma and J. Pearl. Causal networks: Semantics and expressiveness, in proceedings. In *4th Workshop on Uncertainty in Artificial Intelligence*, page 352, 1988.


## Where you can find us:


zbilodea@skidmore.edu