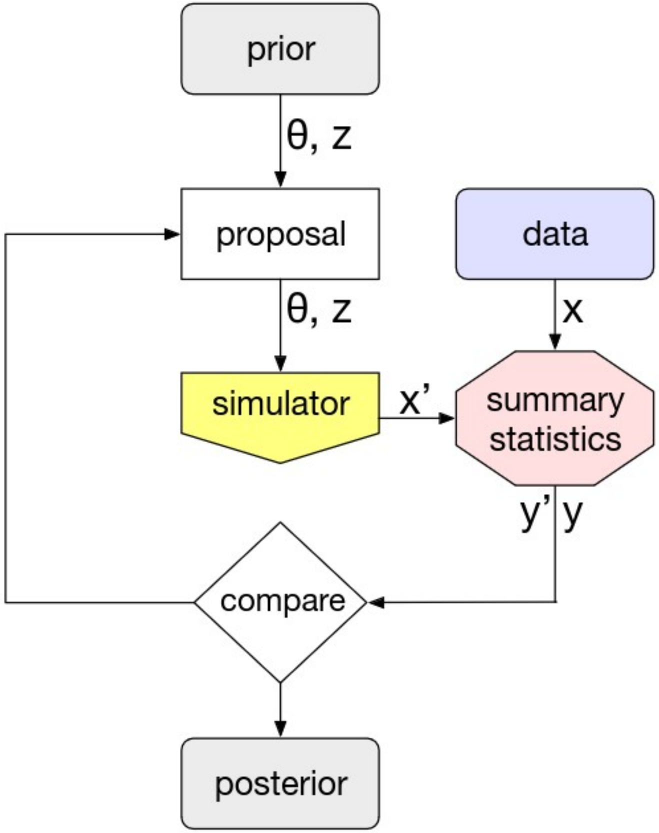
Implementing an Emissions Model for Particle Detectors with Probabilistic Programming

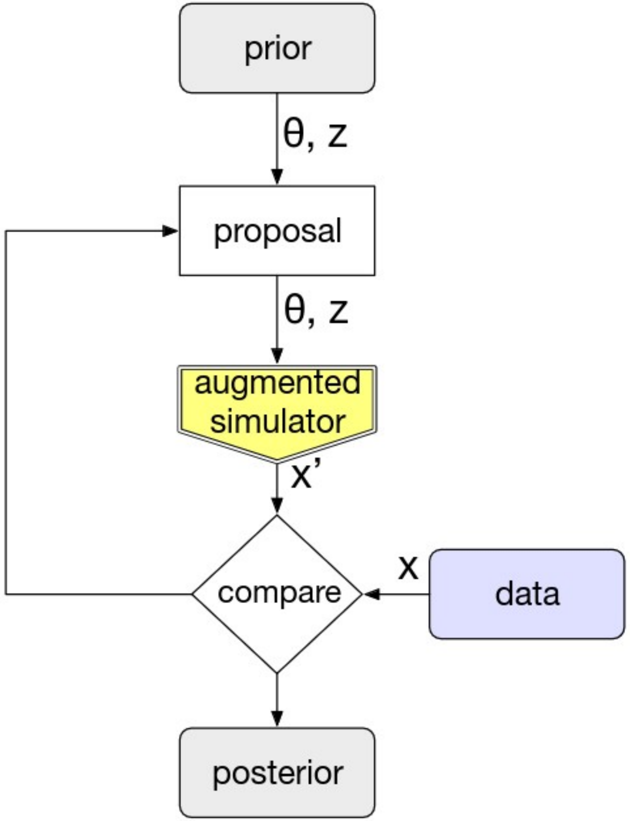
Qin, Juehang (jq8@rice.edu); Christopher Tunnell

RICE UNIVERSITY

ACAT 2024

The Probabilistic Programming Paradigm



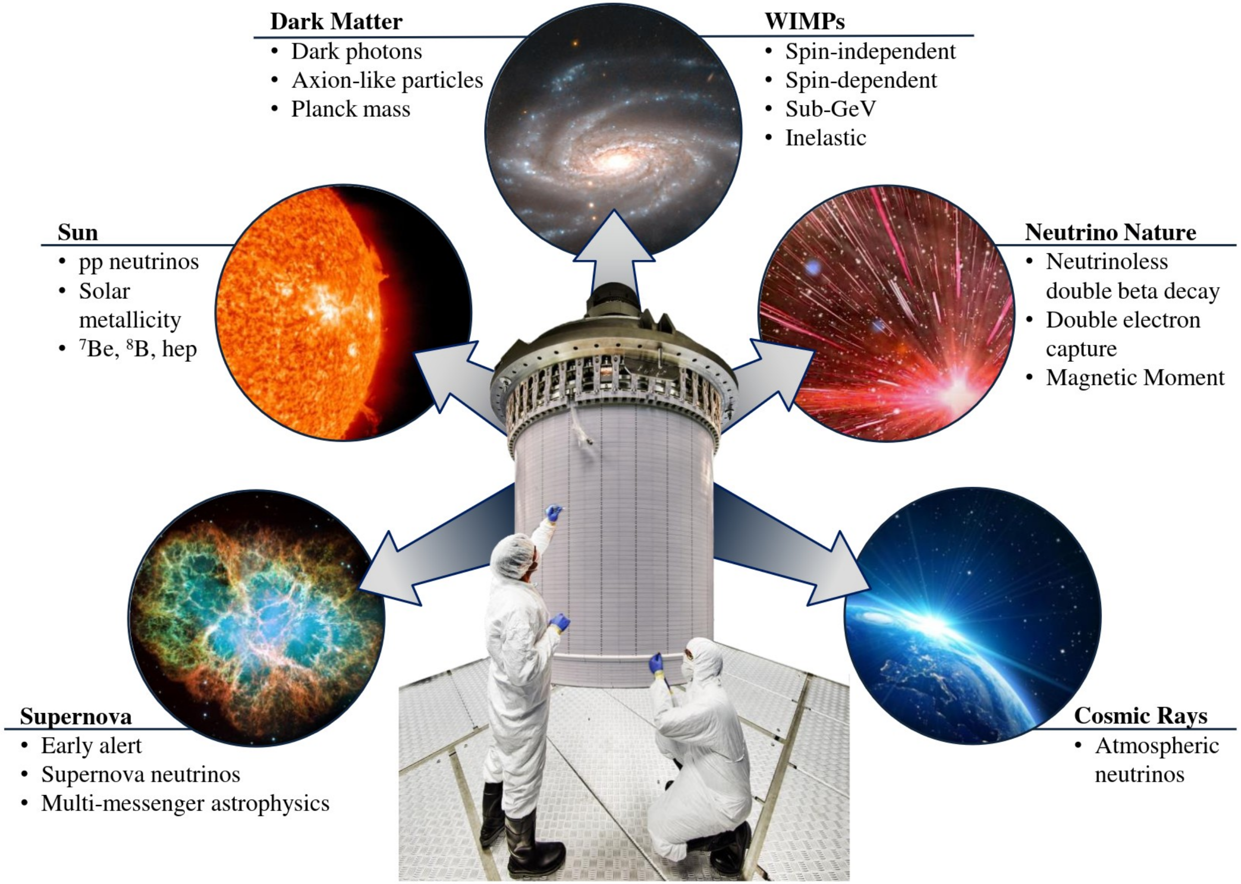


K. Cranmer et al. (2019), DOI:10.1073/pnas.1912789117

Benefits of Explicit Likelihoods

- Information loss occurs with summary statistics
- Template-based method has requirement to compute large sets of simulated events for inference to populate high-dimensional histograms
  - Unfeasible for high dimensional problems!
- Such methods should work in general for simulators with the correct properties; we focus here on liquid xenon TPCs for astroparticle physics

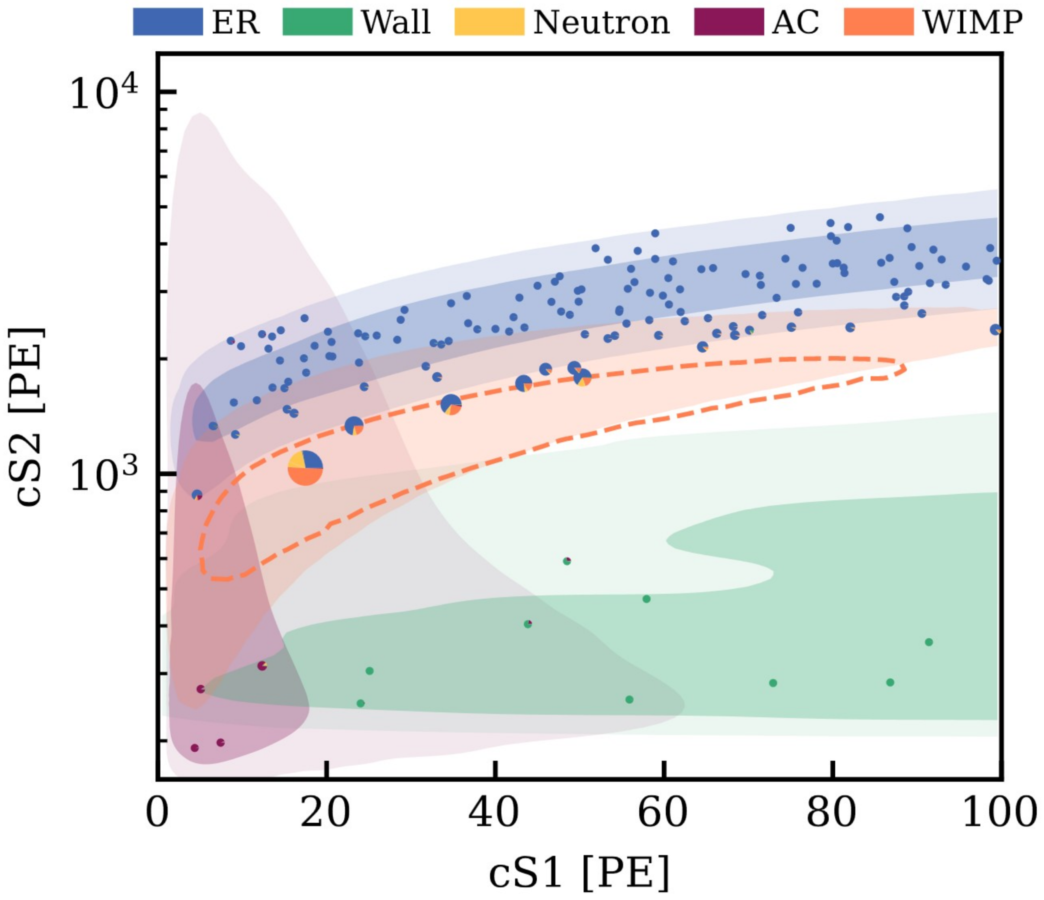
Not Just a Dark Matter Experiment



J. Aalbers et al. (2022), DOI: 10.1088/1361-6471/ac841a

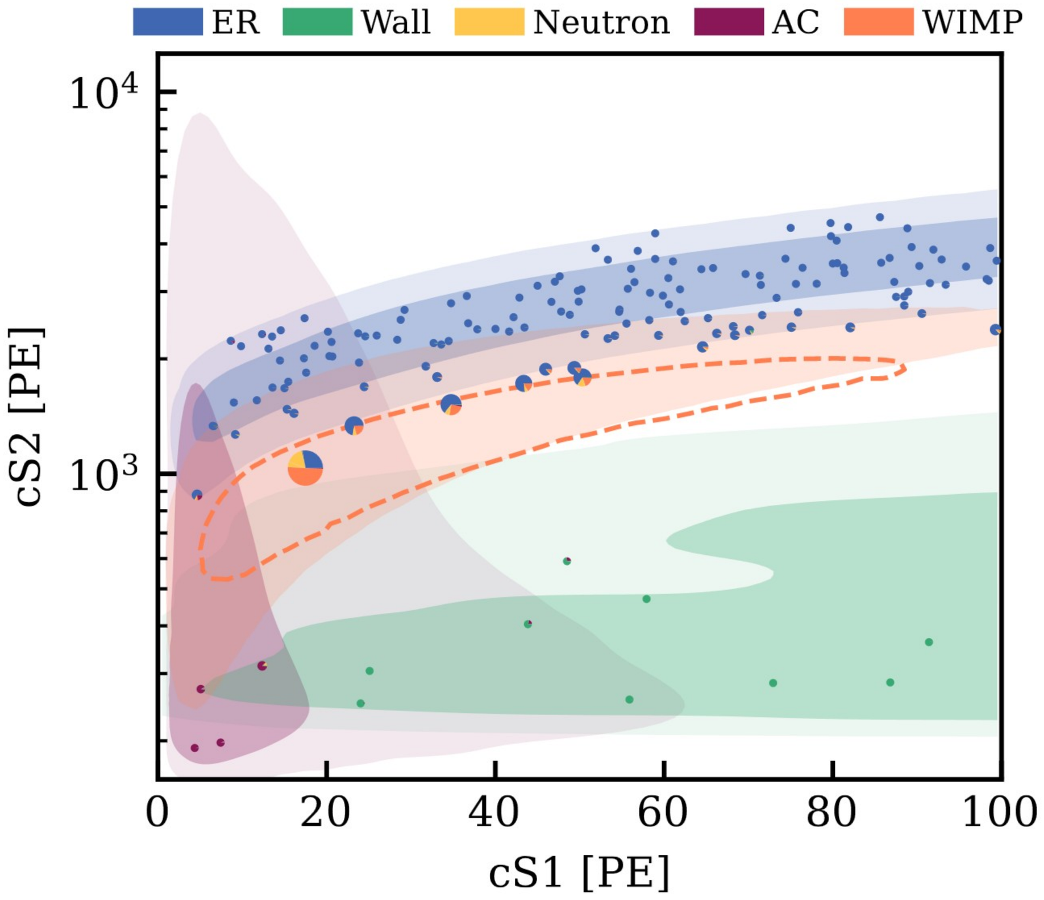
What the Data Looks Like in S1/S2 Space

XENON (2023), DOI: 10.1103/PhysRevLett.131.041003



- Example from the XENONnT WIMP analysis
- While we are counting discrete quanta, they are measured as continuous variables as the resolution is typically worse than 1 photon

Inference with Templates

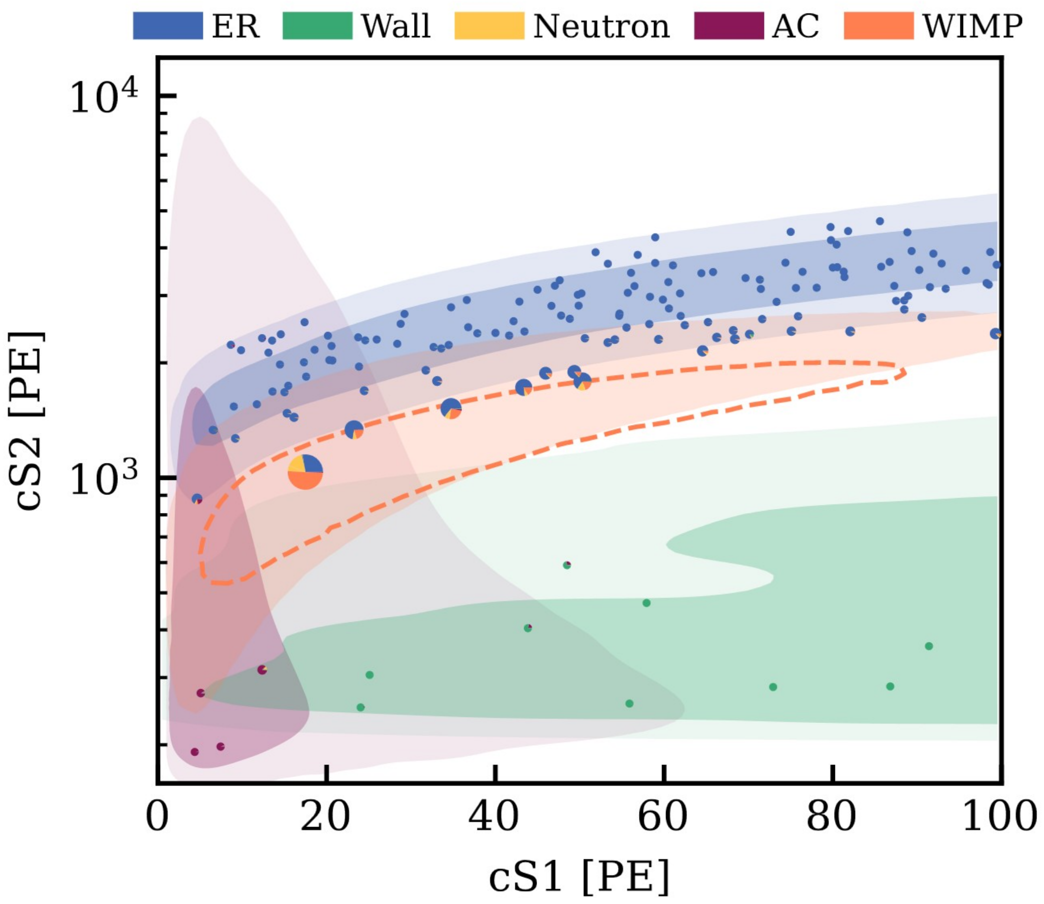


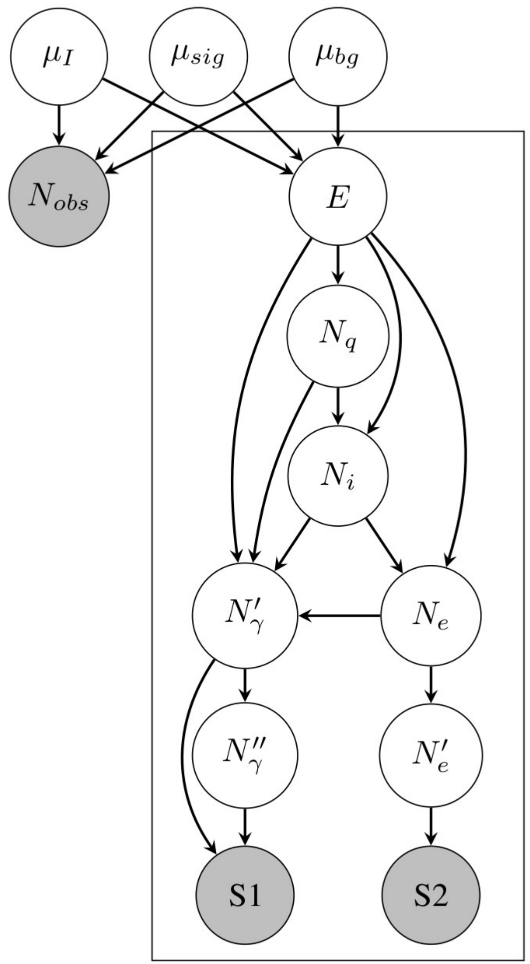
XENON (2023), DOI: 10.1103/PhysRevLett.131.041003

- Currently, inference is done using "templates", essentially histograms generated using events generators like NEST
- What if we can construct explicit likelihoods?
  - There exist prior work, see FlameNEST, arXiv:2204.13621, R.S. James et al.

What Does the Model Look Like?

XENON (2023), DOI: 10.1103/PhysRevLett.131.041003





Noble Element Simulation Technique

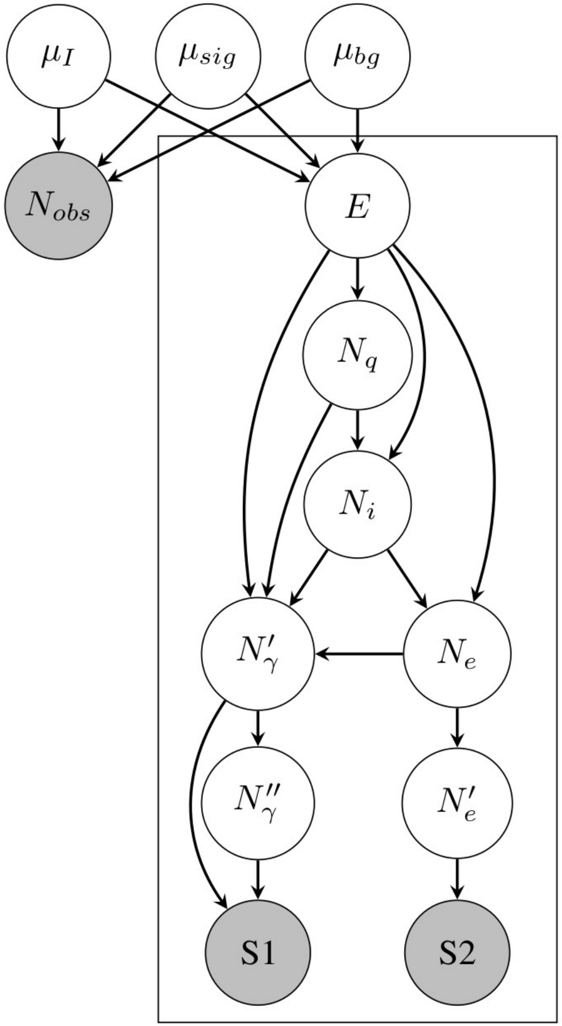
- The Noble Element Simulation Technique (NEST) is a toolkit for simulating quanta generation in liquid noble gas detectors
- In this work, I implement the NEST model for beta decays in liquid xenon in a probabilistic programming framework (numpyro)

M. Szydagis et al. (2022), arXiv:2211.10726

D. Phan et al. (2019), arXiv:1912.11554

NEST Beta Signal Model

- Parameters for the signal and background rates can be included directly
- This model has many discrete parameters!
  - Discrete parameters do not have gradients



Continuous Approximation

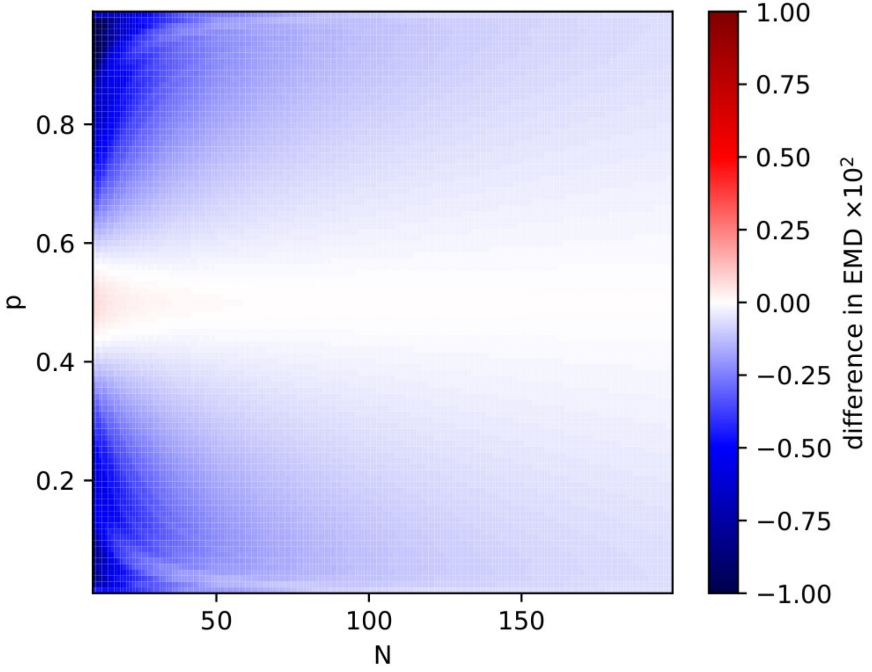
- We can simply replace the discrete random variables with continuous ones
  - This only works with large numbers
- Complementary to the approach of marginalising over discrete random variables (taken by FlameNEST, see arXiv:2204.13621, R.S. James et al.)

Continuous Binomial

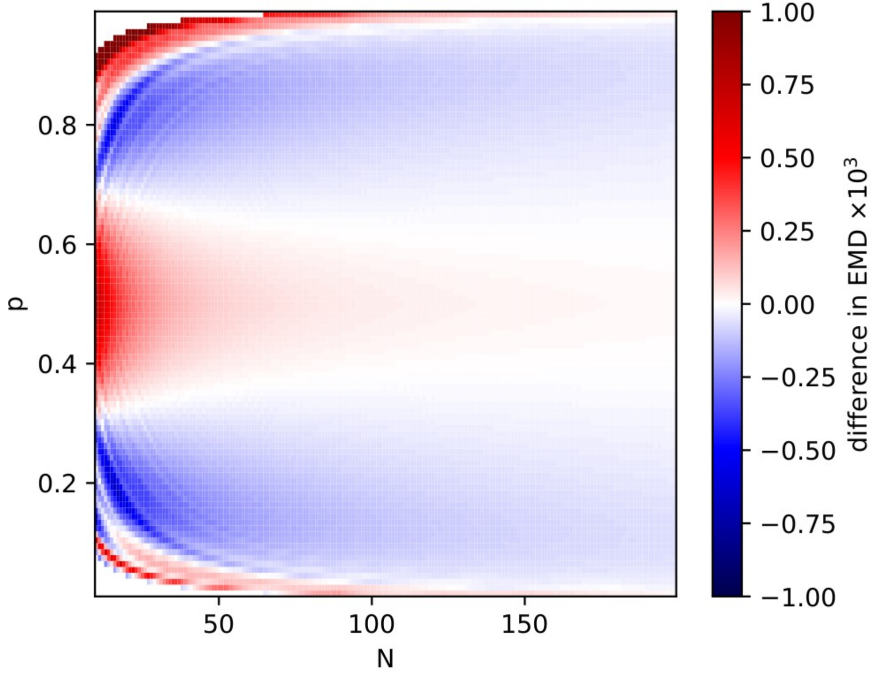
- Normal Approximation has no skew
- We can match a binomial distribution as a mixture of a Normal random variable and a Beta random variable multiplied by the number of trials
  - Enough degrees of freedom to match the first three moments: mean, variance, skew
- There is a very simple analytic solution for the mixture distribution weights!

$$w = \frac{\alpha' + \beta' + 2}{2\left(\alpha' + \beta' + 1\right)}$$

Performance of Continuous Binomial



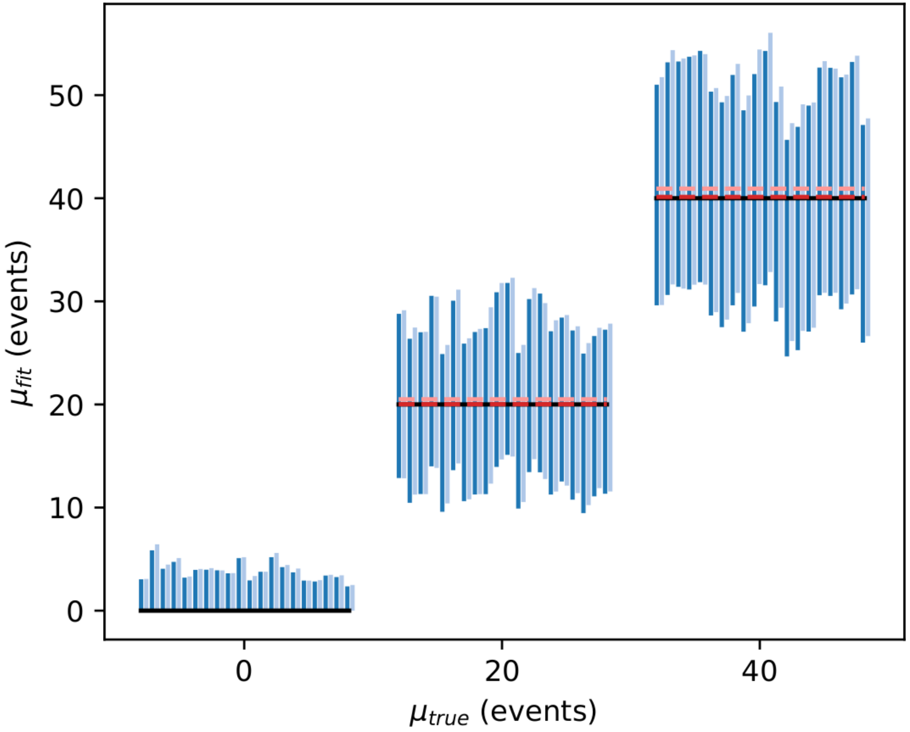
Compared with normal distribution



Compared with skew-normal

Note: blue means the mixture distribution performs better

Preliminary Performance for Line Fit



- Toy problem: flat background with a mono-energetic 200keV signal
- Repeated 20 times using a traditional spectral fit and the model constructed in this work.
- 90% credible intervals are 4.8 ± 0.9%, 1.6 ± 0.4%, and 0.5 ± 0.3% smaller
- Preaching to the choir here, but don't scoff at cheaply-obtained percent-level improvements!

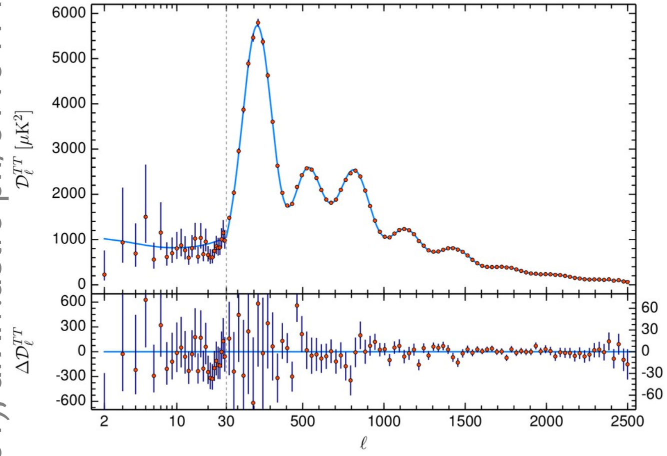
Further Work

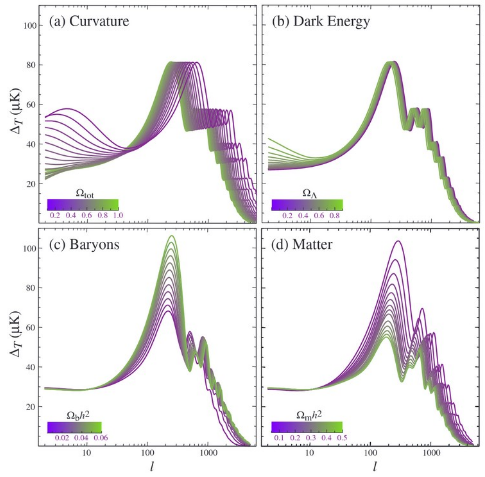
- Investigate use of this method to integrate multiple channels to estimate the same observable
  - S1 and pulse counting
- Investigate how this can be integrated into a frequentist analysis, which is traditional for particle physics
- Try out method on an actual analysis, instead of toy problems.
- This work also represents a differentiable detector model, application to design optimisation and end-to-end optimisation can be explored (see: Lukas Heinrich's plenary talk today!)
  - Note: not a full detector simulation yet!

Key Takeaways

- The signal model for liquid xenon detectors can be directly implemented in the probabilistic programming paradigm
  - Simulation-based inference in the purest form!
- Furthermore, with either a continuous approximation or a lot of compute and memory (marginalisation), this can be differentiable
- We can use this to enable Bayesian inference with gradients, and potentially also profile likelihoods using gradient descent! (the latter is future work)

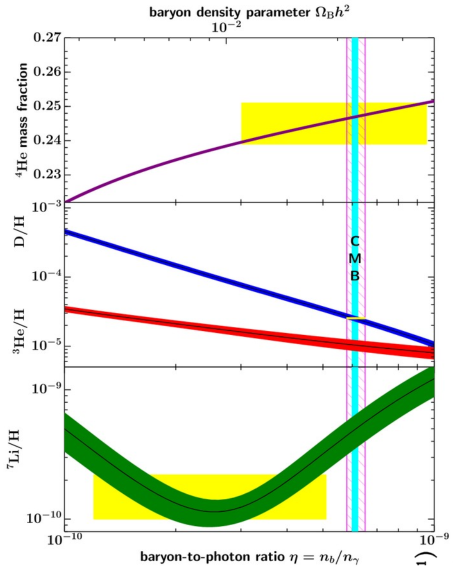
There is a lot of evidence for dark matter





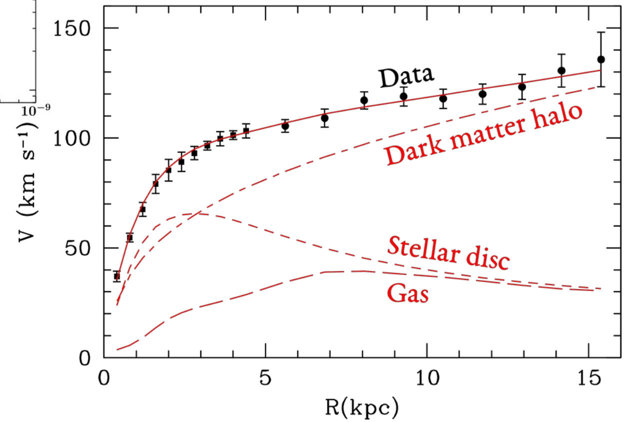
Planck (2018), arXiv:1807.06209

W. Hu, S. Dodelson (2001), arXiv:astro-ph/0110414



https://apod.nasa.gov/apod/ap170115.html

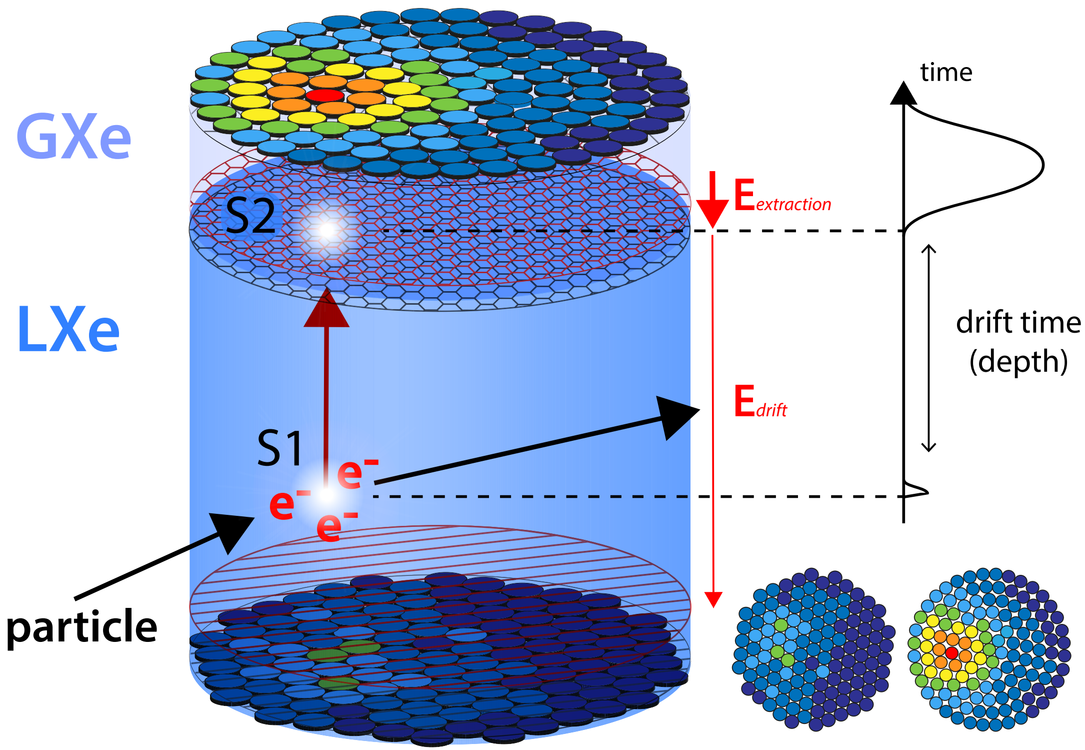
Particle Data Group (2022)



E. Corbelli, P. Salucci (2000), arXiv:astro-ph/9909252

Liquid Xenon TPCs

Image credit: Lutz Althüser, XENON Collaboration



- Scintillation (S1) occurs first, followed by the ionisation signal (S2).
- Time difference gives z-coordinate
- S2 PMT hitpattern allows for x-y position reconstruction.
- Extremely information dense!

Introduction

Model Implementation

Results

Further Work