Goodness-of-Fit Testing using Wasserstein Distance to evaluate model of Double Electron Capture background

Nevis Labs Columbia University REU
XENON Collaboration

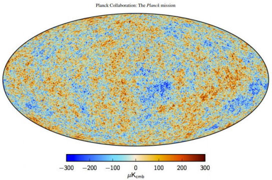
Eleni Mandarakas
eleni.mandarakas @gmail.com

Bucknell University

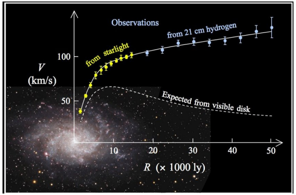
Mentors: Prof. Elena Aprile
Dr. Pueh Leng Tan

Introduction to XENON

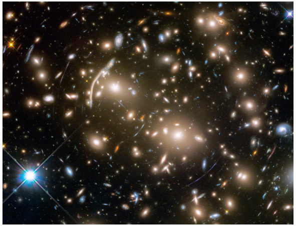
Dark Matter and Other Exotic Particles



Temperature Fluctuations in CMB [7]



Rotation Curves of Speeds of Galaxies [9]



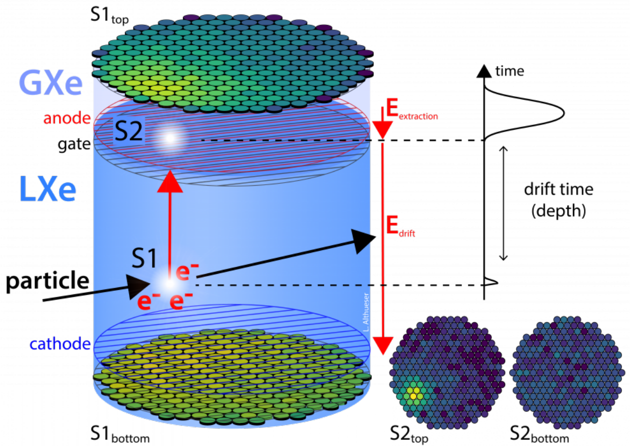
Gravitational Lensing [10]

XENON Collaboration

XENON Detector [4]

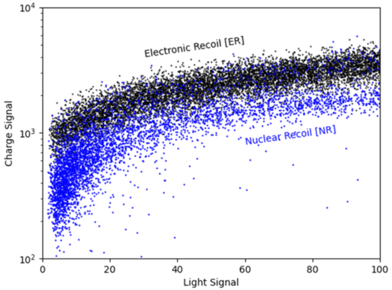
Laboratori Nazionali del Gran Sasso (LNGS) [8]

Dual Phase Xenon Time Projection Chamber (TPC)



[1]

ER and NR Discrimination



[12]

ER:

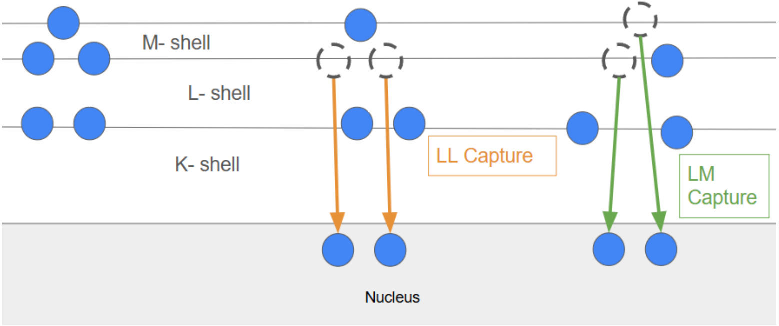
- ◇ Electronic Recoil
- ◇ Particle collides with electron
- ◇ Background
- ◇ low ionization density → low recombination rate → higher charge yield

NR:

- ◇ Nuclear Recoil
- ◇ Particle collides with nucleus / Xe atom
- ◇ ROI to find WIMPs
- ◇ high ionization density → high recombination rate → lower charge yield

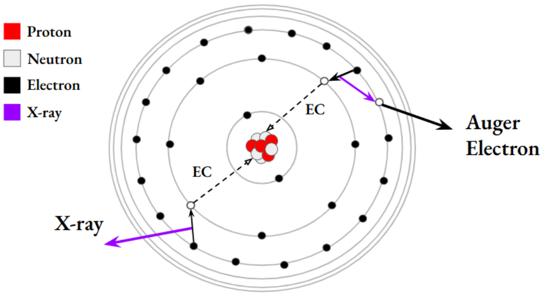
Double Electron Capture (DEC)

Double Electron Capture (DEC)

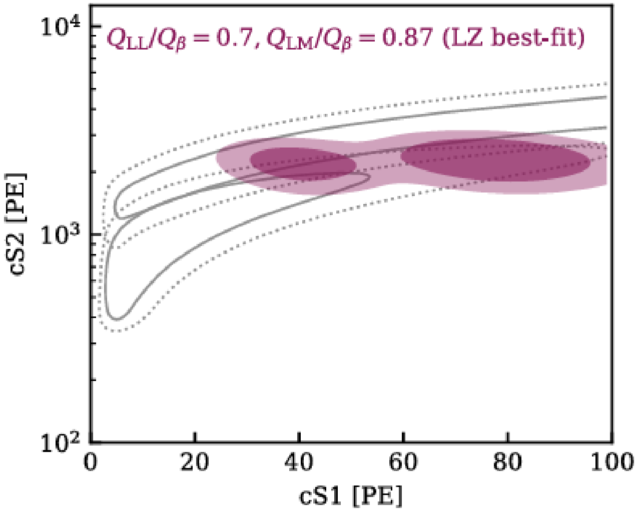


Double Electron Capture (DEC)

$$^{124}\mathrm{Xe} + 2e^- \rightarrow ^{124}\mathrm{Te} + 2\nu_e + \text{(X-rays and Auger electrons)}$$



Double Electron Capture (DEC)



[6]

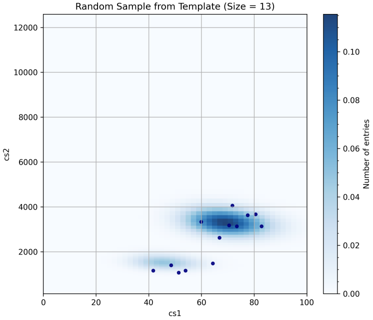
Hypothesis Testing

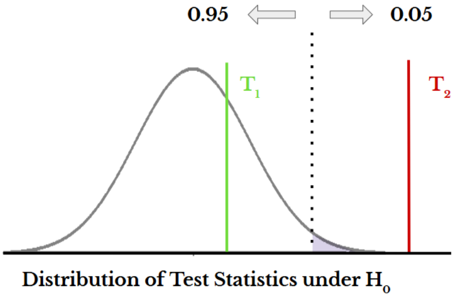
Hypothesis Testing

Goodness of Fit Test - One Tail

Goodness of Fit Test

- Calculate the test statistic - the difference between data and model.
- Compare statistic to test distribution under H_0 to see if it falls beyond the defined confidence interval (e.g. 95%).
- If beyond, reject the null.



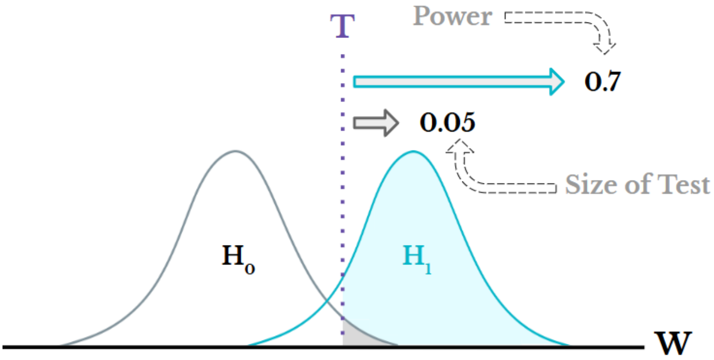


Hypothesis Testing

Power

Power Calculation

$$\text{Power} = \mathrm{P}_{H_1}(W > T)$$



Wasserstein Distance

Wasserstein Distance

General Definition

General Definition: Wasserstein Distance

The Wasserstein distance, also known as the Earth Mover's Distance (EMD), measures the minimum "work" required to transform one probability distribution into another, where "work" = (amount of mass moved) × (distance transported).

Formal Definition:

$$W_p(\mu,\nu) = \left(\inf_{\gamma \in \Gamma(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x-y\|^p \, d\gamma(x,y) \right)^{1/p}$$

Where:

- μ, ν: input distributions
- $\gamma \in \Gamma(\mu,\nu)$: set of all joint couplings with marginals μ, ν
- $d(x,y)$: metric distance between points x and y

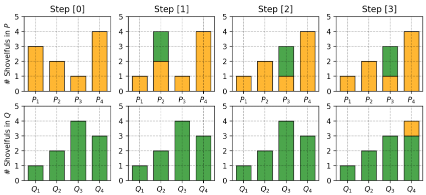


Figure 1. Discrete 1D example of EMD [11].

Wasserstein Distance

EMD Function

2-Wasserstein Distance (EMD) Formulation

Source: $\{x_i\}_{i=1}^{n} \subset \mathbb{R}^2$, $\mathbf{a} \in \mathbb{R}^n$
Target: $\{y_j\}_{j=1}^{m} \subset \mathbb{R}^2$, $\mathbf{b} \in \mathbb{R}^m$

$$\sum_{i=1}^{n} a_i = 1, \quad \sum_{j=1}^{m} b_j = 1, \quad a_i, b_j \geq 0$$

Cost matrix: $C_{ij} = \|x_i - y_j\|_2^2$

Transport plan: $\gamma \in \mathbb{R}^{n \times m}$

$$\sum_{j=1}^{m} \gamma_{ij} = a_i \quad \forall i \qquad \gamma_{ij} \geq 0 \quad \forall i, j$$
$$\sum_{i=1}^{n} \gamma_{ij} = b_j \quad \forall j$$

2-Wasserstein Distance:

$$W = \left(\sum_{i=1}^{n} \sum_{j=1}^{m} \gamma_{ij} C_{ij} \right)^{1/2}$$

Numerical Example: Sample Initialization

Source Distribution

Event Locations: $x = \left\{ \begin{matrix} (1,2), & (1,2), \\ (3,4), & (3,4) \end{matrix} \right\}$

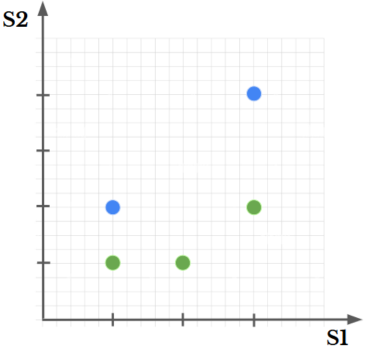
Locations: $x = \{(1,2),\ (3,4)\}$

Masses: $\mathbf{a} = [0.5,\ 0.5]$

Target Distribution

Locations: $y = \{(2,1),\ (3,2),\ (1,1)\}$

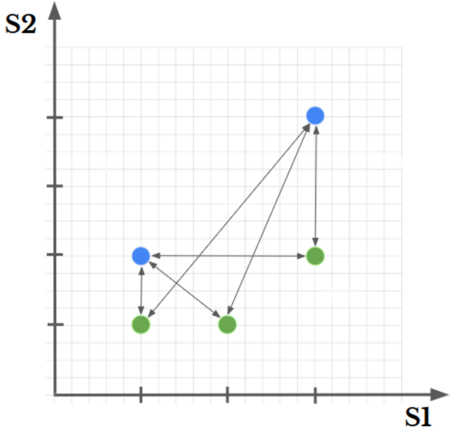
Masses: $\mathbf{b} = [0.25,\ 0.5,\ 0.25]$



Numerical Example: Cost Matrix $C_{ij} = \|x_i - y_j\|_2^2$

	$y_1 = (2,1)$	$y_2 = (3,2)$	$y_3 = (1,1)$
$x_1 = (1,2)$	2	4	1
$x_2 = (3,4)$	10	4	13

$$C = \begin{bmatrix} 2 & 4 & 1 \\ 10 & 4 & 13 \end{bmatrix}$$



Mass Marginals and Optimal Transport Plan

Mass marginals:

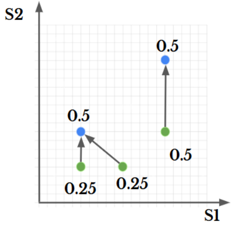
$$\mathbf{a} = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 0.25 \\ 0.5 \\ 0.25 \end{bmatrix}$$

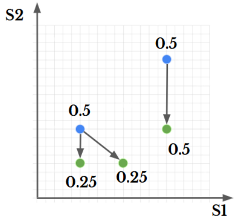
$$\sum_{i=1}^{n} a_i = 1, \quad \sum_{j=1}^{m} b_j = 1, \quad a_i, b_j \geq 0$$

Optimal transport plan:

$$\gamma = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} \end{bmatrix}$$

$$\begin{aligned} &\textstyle\sum_{j=1}^{3} \gamma_{ij} = a_i && \text{(row sums)} \\ &\textstyle\sum_{i=1}^{2} \gamma_{ij} = b_j && \text{(column sums)} \\ &\gamma_{ij} \geq 0 && \text{(non-negativity)} \end{aligned}$$





Solution: $\gamma = \begin{bmatrix} 0.25 & 0 & 0.25 \\ 0 & 0.5 & 0 \end{bmatrix}$

Computing the 2-Wasserstein Distance

$$C = \begin{bmatrix} 2 & 4 & 1 \\ 10 & 4 & 13 \end{bmatrix}, \quad \gamma = \begin{bmatrix} 0.25 & 0 & 0.25 \\ 0 & 0.5 & 0 \end{bmatrix}$$

Compute total cost:

$$\sum_{i=1}^{2} \sum_{j=1}^{2} \gamma_{ij} C_{ij} = 0.25 \cdot 2 + 0.25 \cdot 1 + 0.5 \cdot 4 = 2.75$$

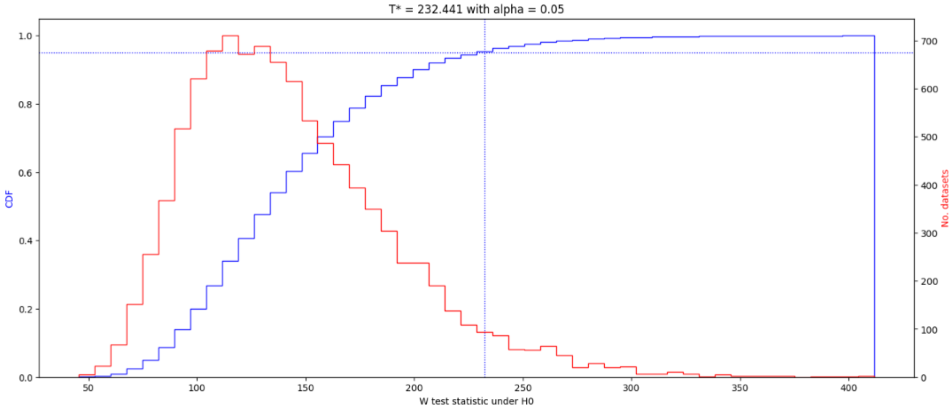
2-Wasserstein Distance:

$$W = \left(\sum_{i,j} \gamma_{ij} C_{ij} \right)^{1/2} = \sqrt{2.75} \approx 1.66$$

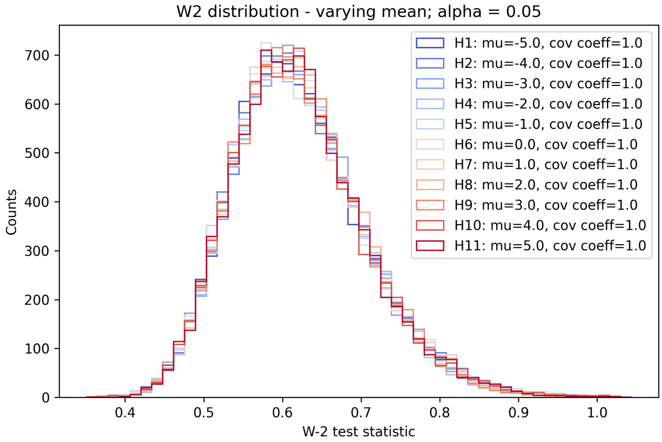
Wasserstein Distance

Wasserstein Distribution

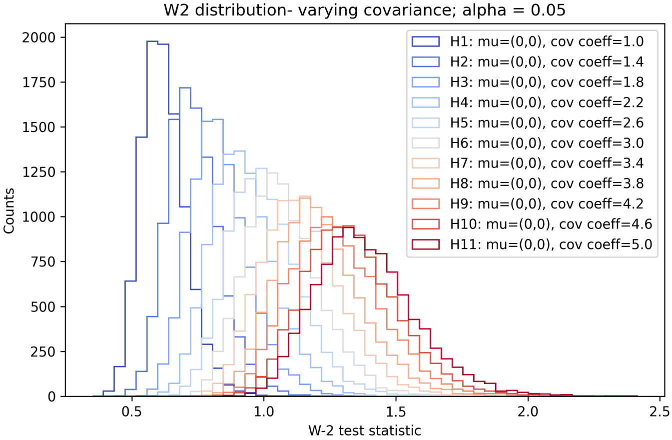
Visualizing a Wasserstein Distribution



Distribution Free? - Varying Mean



Distribution Free? - Varying Covariance



Conclusion

- Not Distribution Free
- Monte Carlo Random Sampling to approximate distribution for each H_0

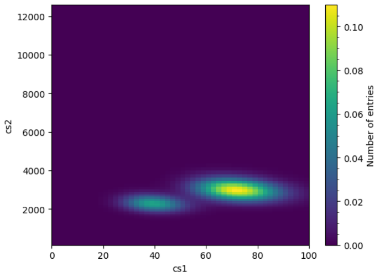
Methodology

Methodology

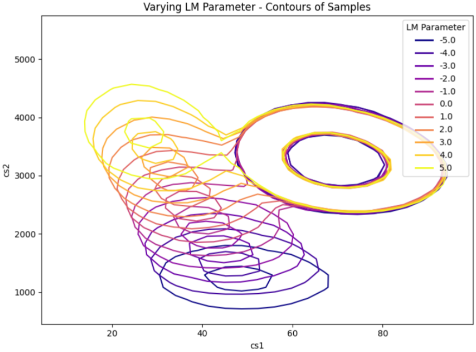
Templates

Template

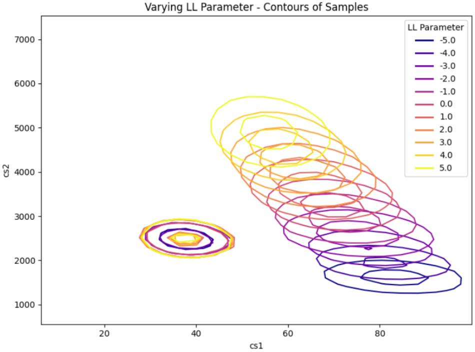
- 2D histogram
- vary LL and LM parameter by changing the light and charge yield
- 11 variations of LL Parameter/11 variations of LM Parameter - 121 possible templates



Vary LL and LM Parameter Individually



Varying LM Parameter - LL Parameter fixed at 0.0



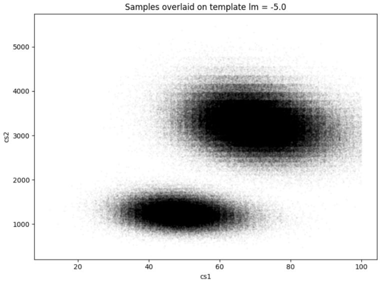
Varying LL Parameter - LM Parameter fixed at 0.0

Methodology

Sampling

Samples

- Sample size = 13 $\rightarrow$ expect to see approximately 13 DEC Events in the data set
- Template - use bin centers and "height" of bin to construct probability distribution
- Redraw sample and compare with template 10,000 times to create Wasserstein Distribution

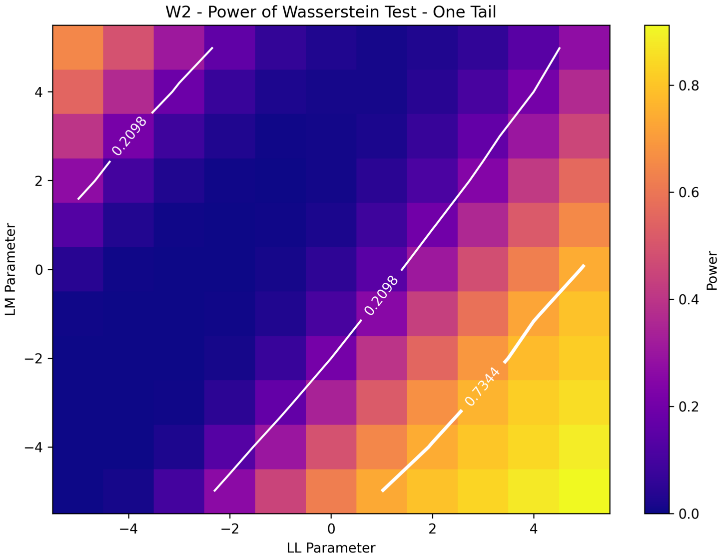


Conclusion

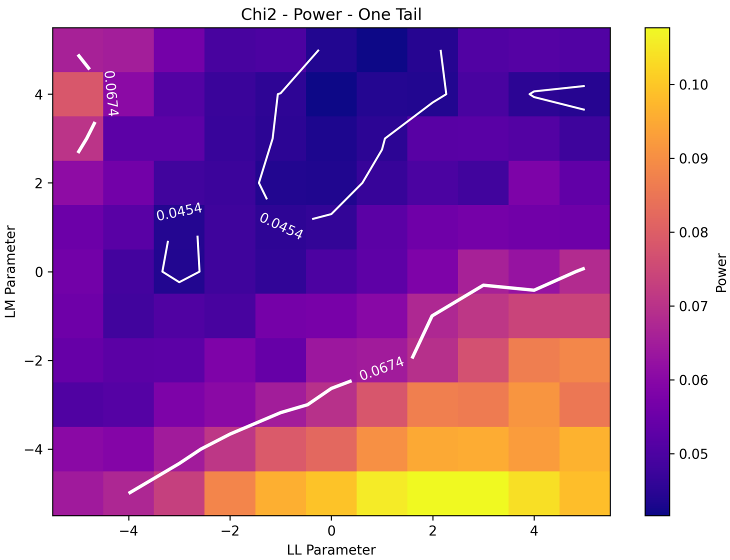
Conclusion

Results

Power Heatmap



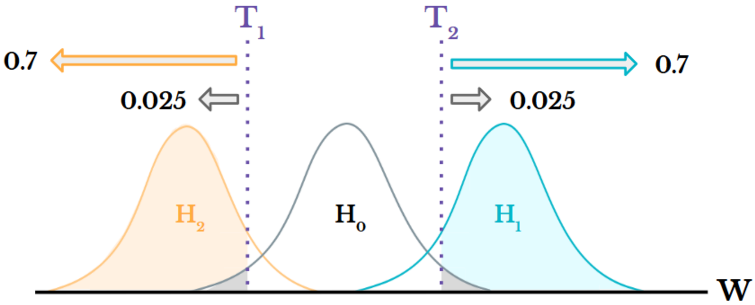
χ^2 Test Comparison



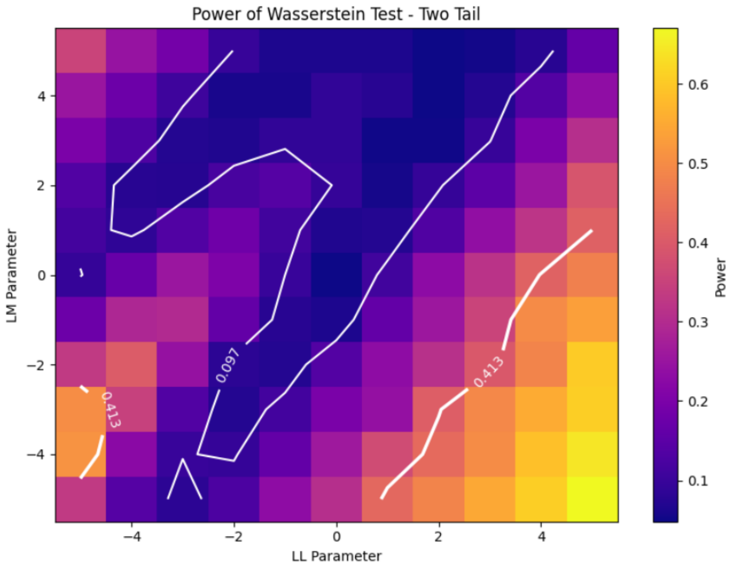
Conclusion

Future Work

Two-Tail Power



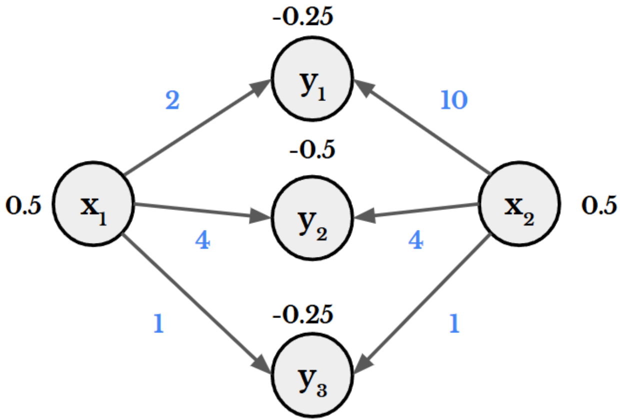
Two-Tail Power Heatmap



Thank you!

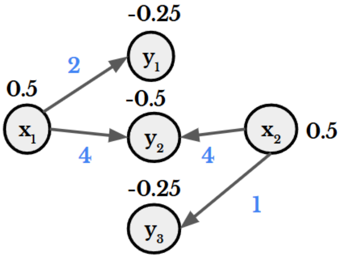
I am grateful to Dr. Elena Aprile for her constant support, Dr. Pueh Leng Tan for her mentorship, direction, and impactful discussions, and XENON Collaboration and Laboratori Nazionali del Gran Sasso for hosting and supporting the project. Thank you also to the Nevis Labs REU Program; Prof. Georgia Karagiori, Prof. Reshmi Mukherjee, Amy Garwood, Asia Latt, and Ashley L. Delphia. This material is based upon work supported by the National Science Foundation under Grant No. PHY-2349438.

Backup



Backup

Choose Basis:

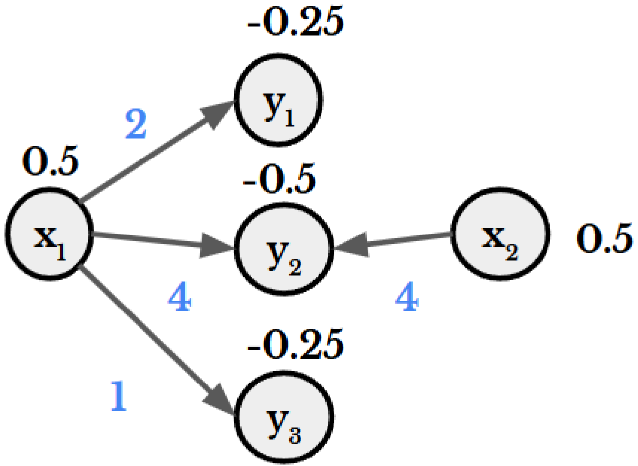


Cost - Dual Tail + Dual Head = 0

2 - 0 - 2 = 0
4 - 0 - 4 = 0
4 - 0 - 4 = 0
1 - 0 - 1 = 0

x_2y_1: 10 - 0 - 2 = 8
x_1y_3: 1 - 0 - 4 = -3

Backup



Lutz Althüser.
Schematic of the xenonnt liquid-xenon detector.
Image from XENON Collaboration, credited to Lutz Althüser, n.d.
Retrieved 2025-08-01.

E. Aprile et al.
Observation of two-neutrino double electron capture in ^{124}xe with xenon1t.
Nature, 568(7753):532–535, 2019.

E. Aprile and XENON Collaboration.
Wimp dark matter search using a 3.1 tonne$\times$year exposure of the xenonnt experiment.
arXiv preprint, 2025.

ETAP – Experimental Particle and Astroparticle Physics Group, Johannes Gutenberg University Mainz.
Xenon/darwin – ex-xenon research group overview.
https://www.etap.physik.uni-mainz.de/research-groups/ex-xenon/, n.d.
Accessed: 2025-08-01.

Marc Hallin, Gilles Mordant, and Johan Segers.
Multivariate goodness-of-fit tests based on wasserstein distance, 2020.

Robert Hammann.
Dec hypothesis test and wimp roi in xenonnt, 2024.
XENON1T Internal Note, accessible via XENON Wiki.

Daniel R. Herrera, U. São Paulo, and S.E. Jorás.
Full-sky map of the cmb temperature anisotropies observed by planck satellite.

https://www.researchgate.net/figure/
Full-sky-map-of-the-CMB-temperature-anisotropies-observed-by-Planck-satellite-Figure_
fig1_334447860, 2019.
Accessed: 2025-08-01.

Laboratori Nazionali del Gran Sasso.
Lngs overview.
https://www.lngs.infn.it/en/lngs-overview, n.d.
Accessed: 2025-08-01.

P.A. Murad.

László Szabados.
Symmetries in stellar, galactic and extragalactic astronomy.
In *Complex Symmetries*, pages 191–206. 2022.

Lilian Weng.
Gan by example (part 1), 2017.
Accessed: 2025-07-22.

XENON Collaboration.
Time projection chamber.
https://xenonexperiment.org/time-projection-chamber/, n.d.
Accessed: 2025-08-01.