# Cosmic Frontier and machine learning

C. Tunnell (U. Chicago and Rice University)

July 4th, 2018

Caveat: all data either shown publicly or tweaked to make non-experiment specific

# ~~Cosmic Frontier and machine learning~~
# 12 minute on manifold learning

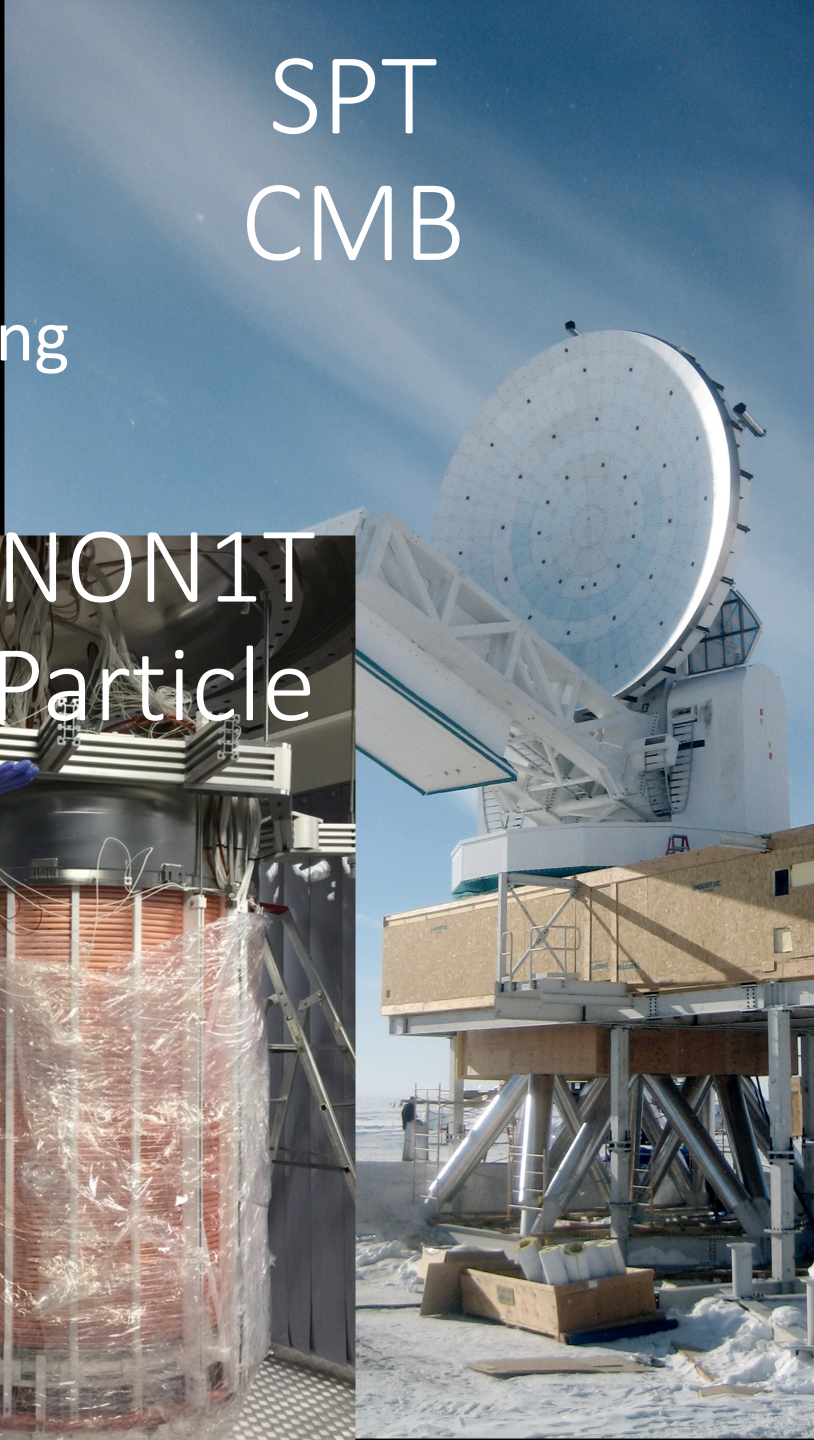
C. Tunnell (U. Chicago and Rice University)

July 4th, 2018

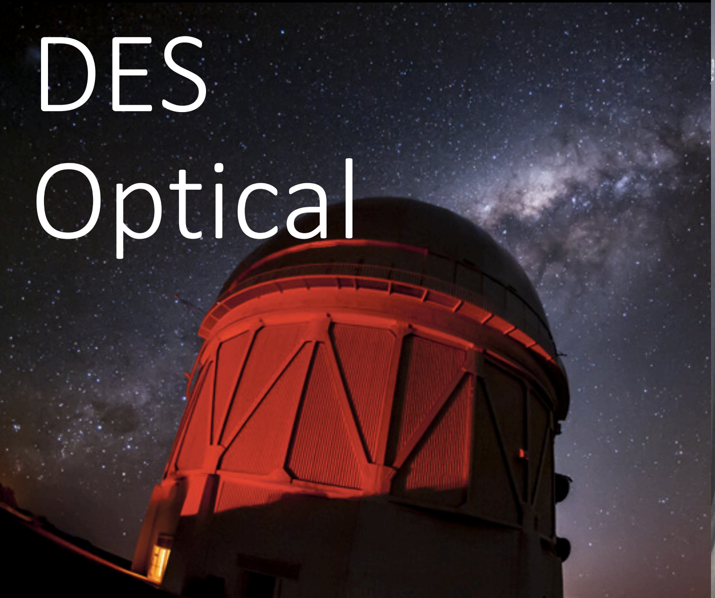
Caveat: all data either shown publicly or tweaked to make non-experiment specific

# Deep Skies Collaboration

Informal astrophysics group exploring useful ML applications in astrophysics







# Cutting through the hype

| Astronomy | Particle physics |
|---|---|
| Image data<br>(i.e. same as classifying cat/dog) | Tree-like data<br>(e.g. graph convolution applications limited) |
| Public data releases with docs | Limited data release, often requires membership |
| Toolchain more friendly to ML | ROOT |

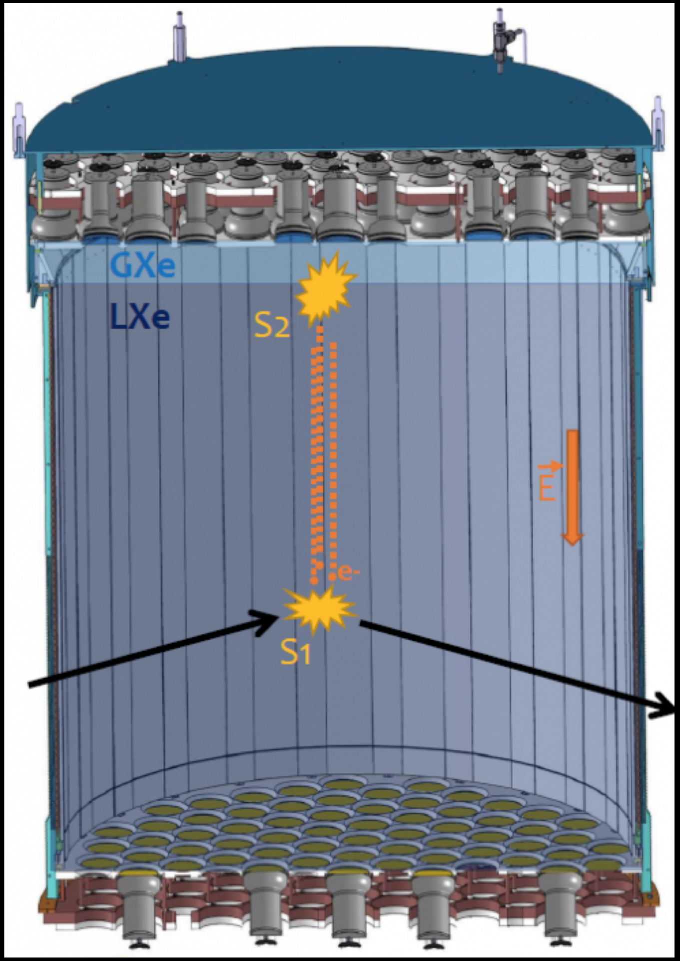
What new measurement does ML open?

Within direct-detection DM, limited impact.

I will discuss one neat example (which coincidentally first plenary said used widely elsewhere but not in HEP)


# What is XENON?

Liquid XENON dark matter detector instrumented with 248 photomultipliers and 10-ns flash ADCs. We make a world-leading new experiment every few years.




arXiv:1805.12562

# What is XENON?

Liquid XENON dark matter detector instrumented with 248 photomultipliers and 10-ns flash ADCs. We make a world-

We are current front runners … but

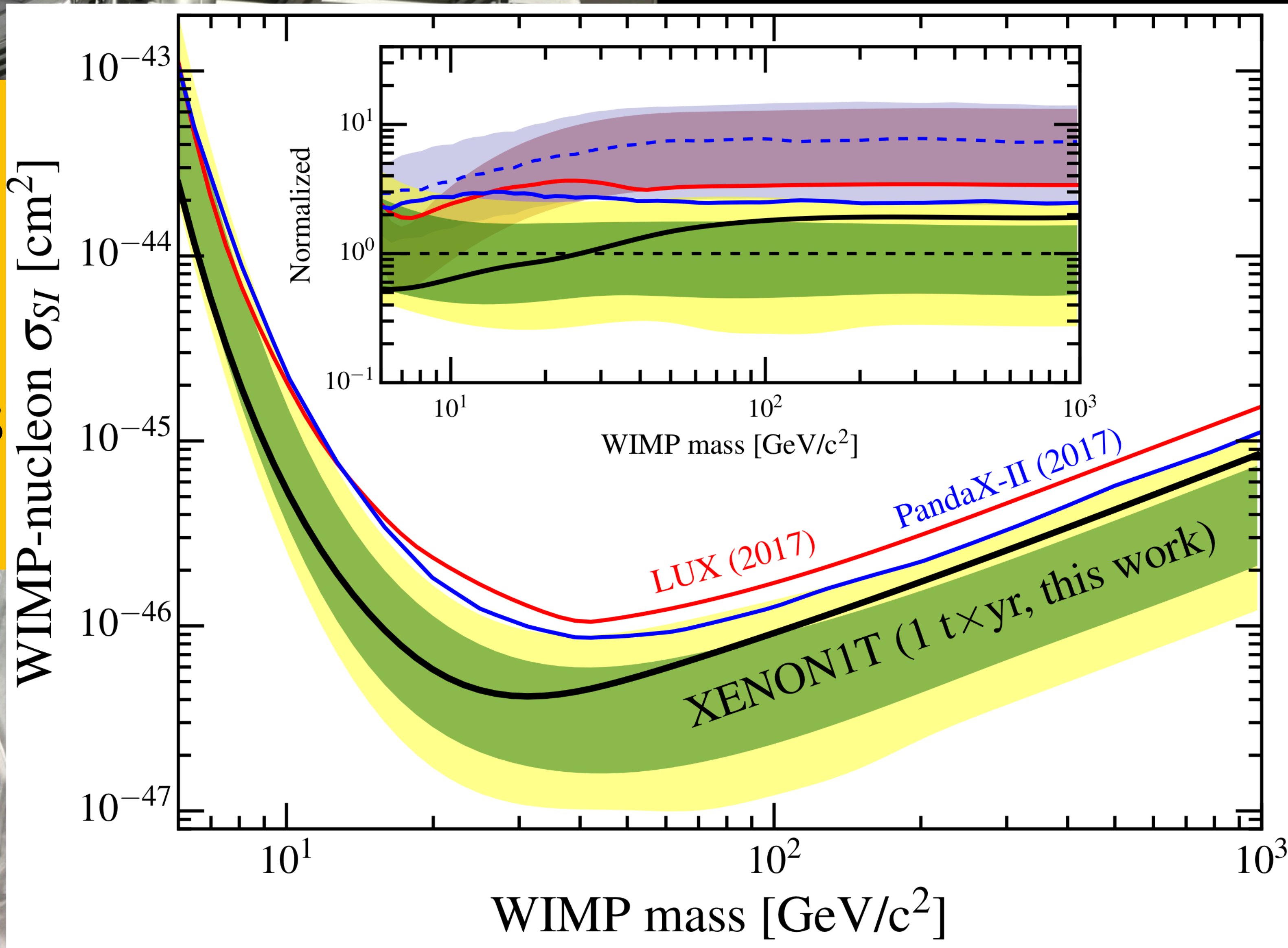
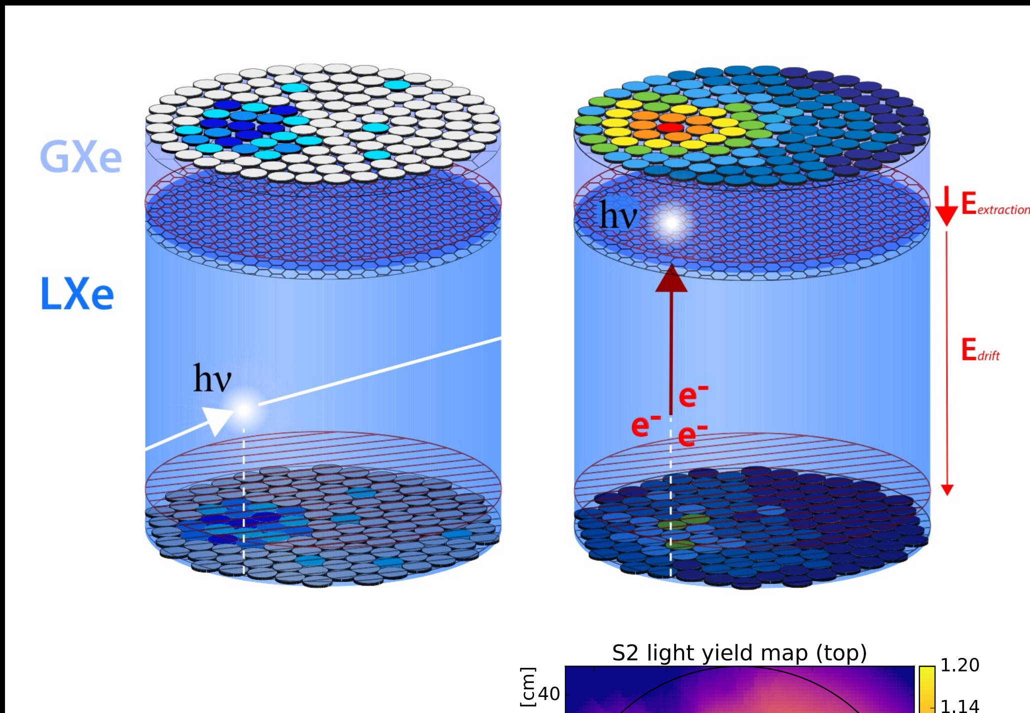
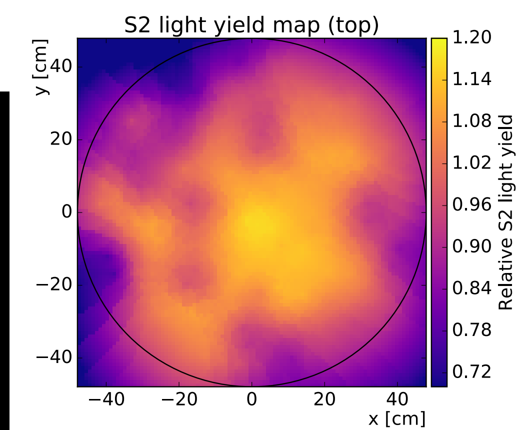


TABLE I: Best-fit expected event rates with 278.8 days live-time in the 1.3 t fiducial mass, 0.9 t reference mass, and 0.65 t core mass, for the full ($cS1$, $cS2_b$) ROI and, for illustration, in the NR signal reference region. The table lists each background (BG) component separately and in total, the observed data, and the expectation for a 200 GeV/c$^2$ WIMP prediction assuming the best-fit $\sigma_{SI} = 4.7 \times 10^{-47}$ cm$^2$.

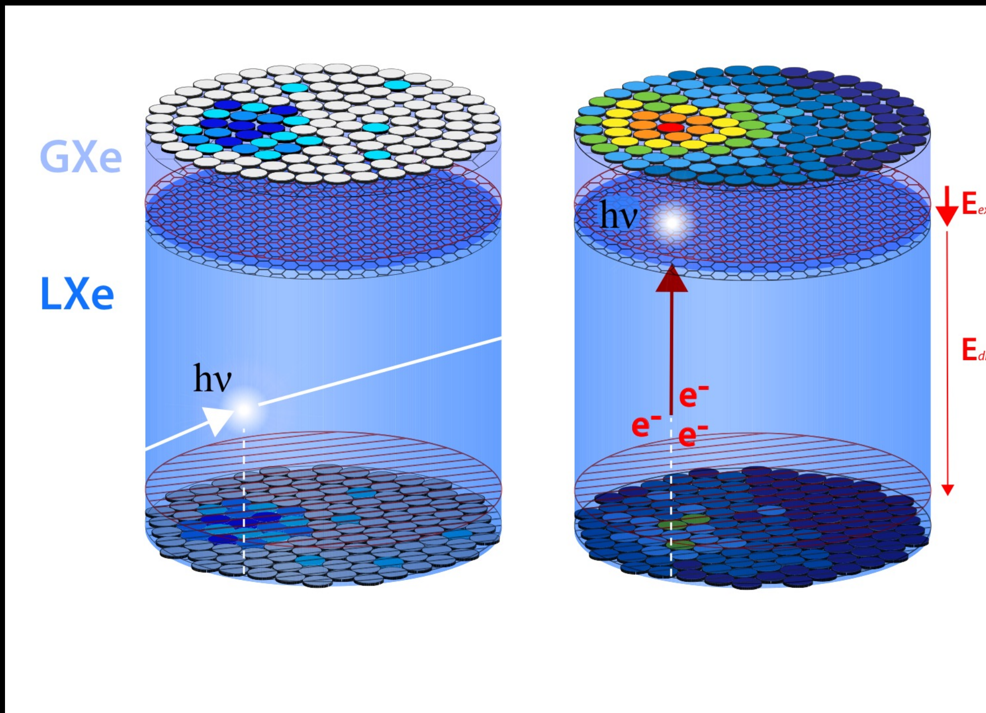
| Mass | 1.3 t | 1.3 t | 0.9 t | 0.65 t |
|---|---|---|---|---|
| ($cS1$, $cS2_b$) | Full | Reference | Reference | Reference |
| ER | 627±18 | 1.62±0.30 | 1.12±0.21 | 0.60±0.13 |
| neutron | 1.43±0.66 | 0.77±0.35 | 0.41±0.19 | 0.14±0.07 |
| CE$\nu$NS | 0.05±0.01 | 0.03±0.01 | 0.02 | 0.01 |
| AC | $0.47^{+0.27}_{-0.00}$ | $0.10^{+0.06}_{-0.00}$ | $0.06^{+0.03}_{-0.00}$ | $0.04^{+0.02}_{-0.00}$ |
| Surface | 106±8 | 4.84±0.40 | 0.02 | 0.01 |
| Total BG | 735±20 | 7.36±0.61 | 1.62±0.28 | 0.80±0.14 |
| WIMP$_{\text{best-fit}}$ | 3.56 | 1.70 | 1.16 | 0.83 |
| Data | 739 | 14 | 2 | 2 |

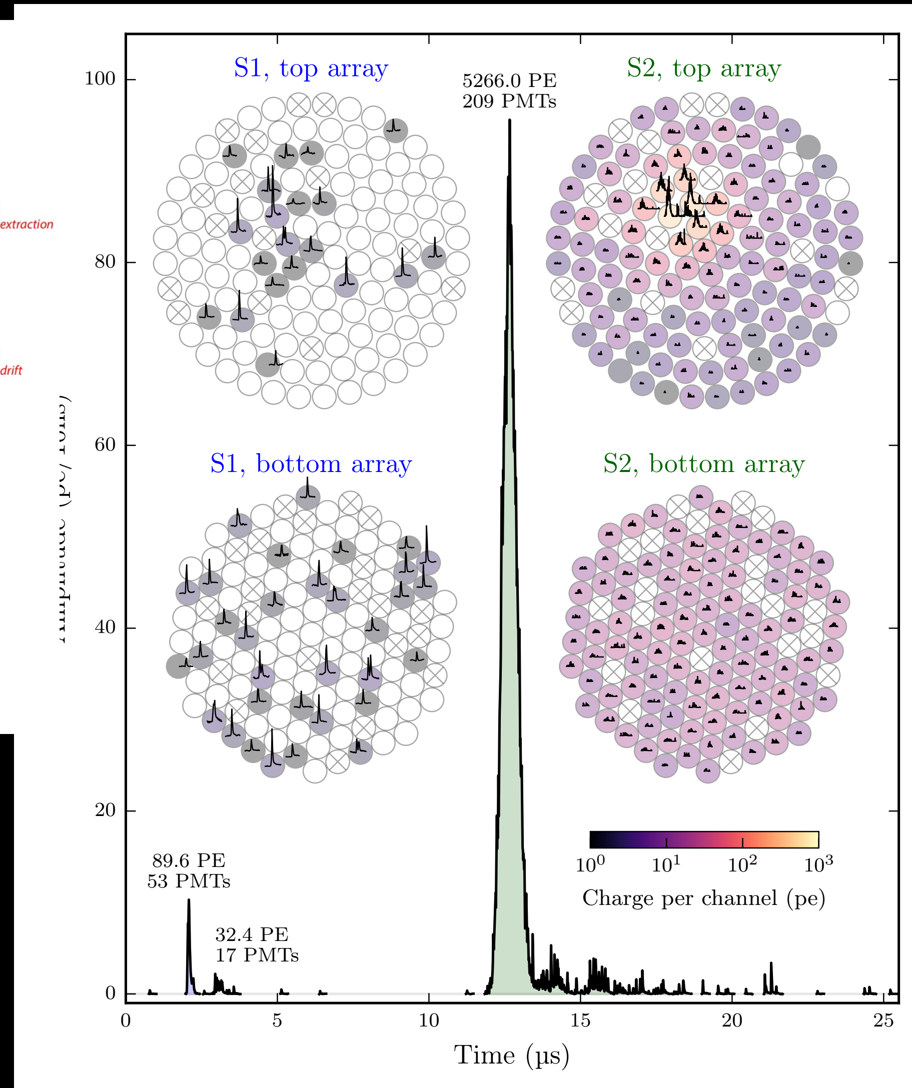



# THE CHALLENGE

1. The optics and response isn't uniform for S2s
    1. Hard to measure reflectivity Teflon ex situ
2. There is no internal calibration
3. When doing ML, including what we do **know** is not straightforward



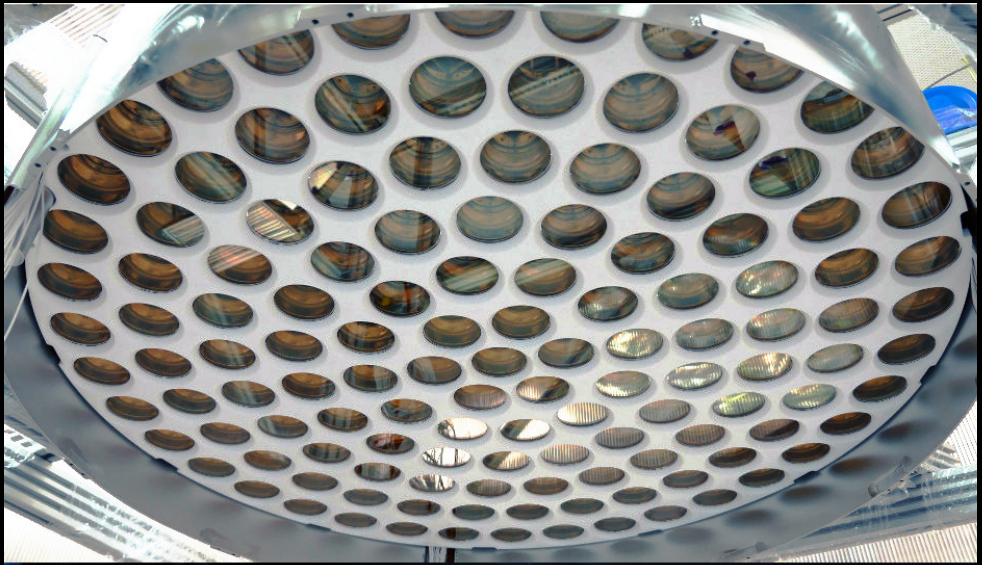

ML Challenge: Photosensors locations in 2D that aren't grid

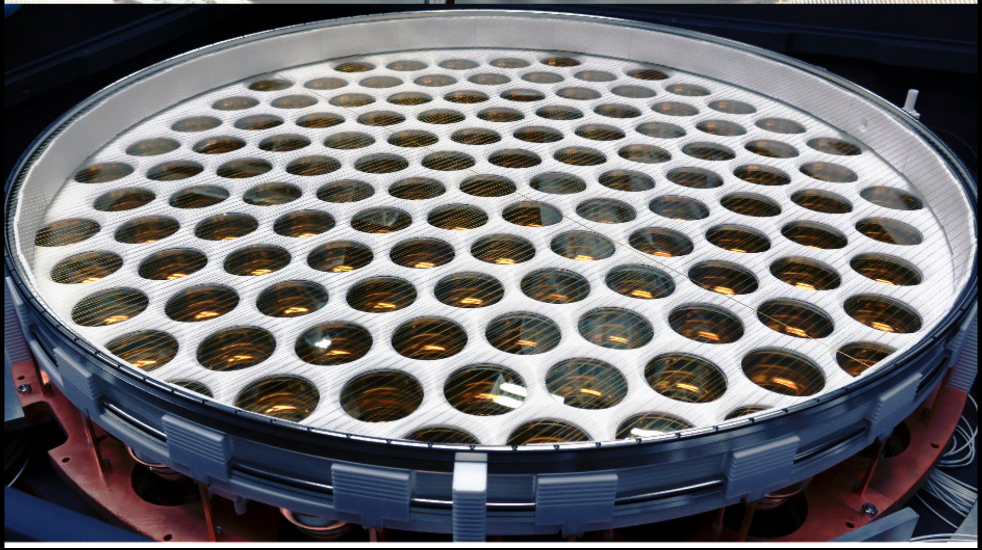
→ Sporadic spacing hard

  → Image not 2D but rather 248D

# Each photosensor represents input dimension

Looking for embedding
$\mathbb{R}^{128} \rightarrow \mathbb{R}^2$

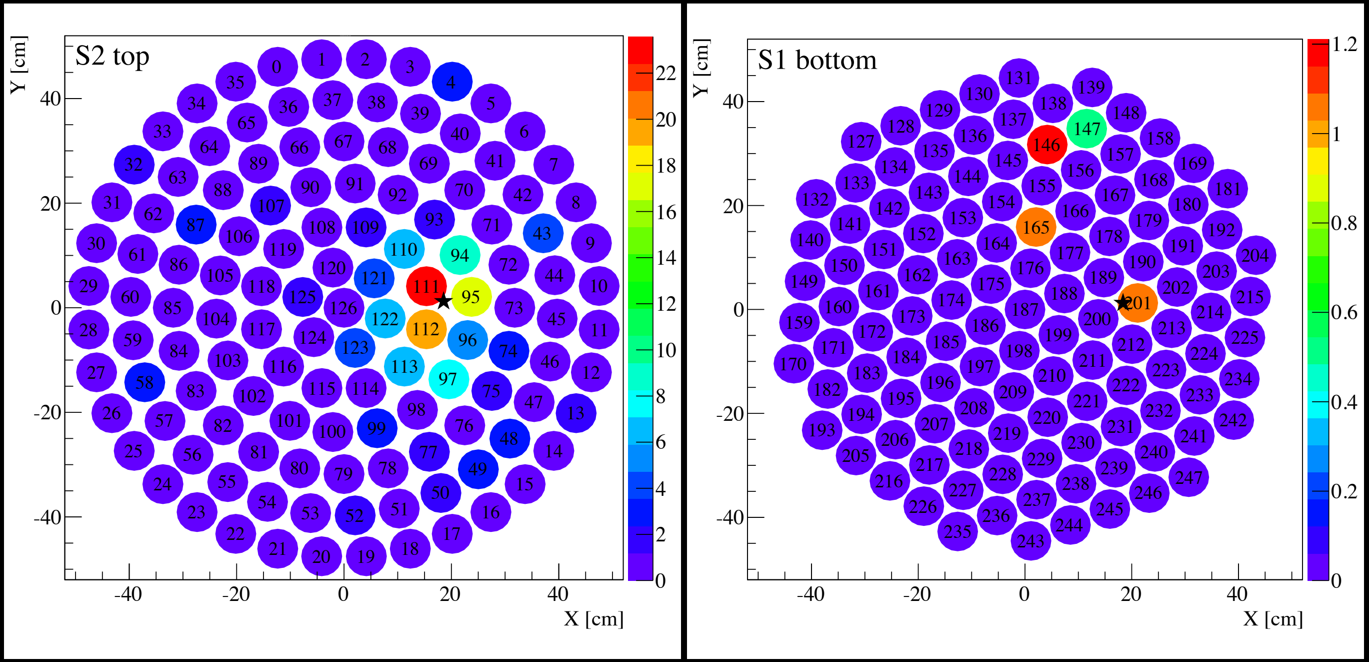




# Each photosensor represents input dimension

Looking for embedding
$\mathbb{R}^{128} \rightarrow \mathbb{R}^2$

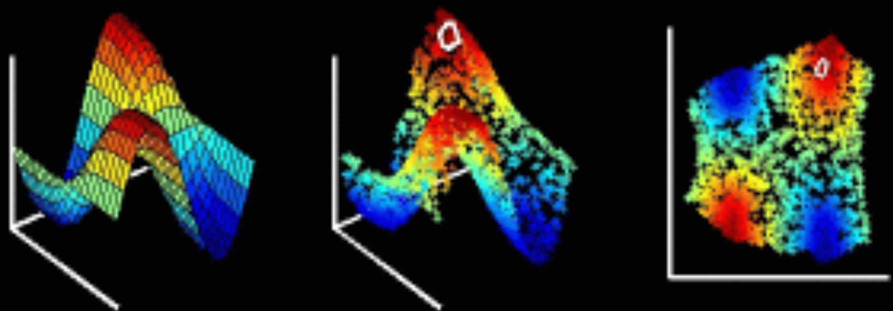
Key insight: sensors near in 2D see similar signals, so sensor dimensions correlated



# Dimensionality reduction

Often have higher dimensional data that we know has a low-dimensional representation

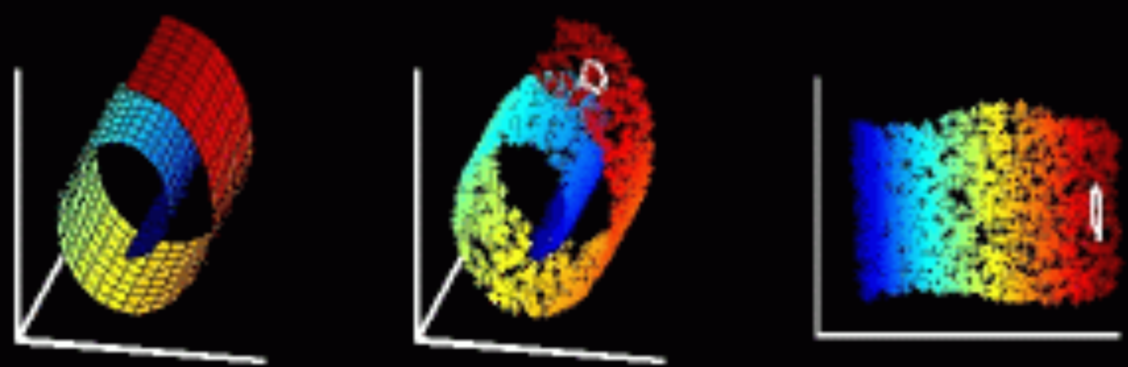
3D→2D: how to flatten? i.e. preserve neighbors



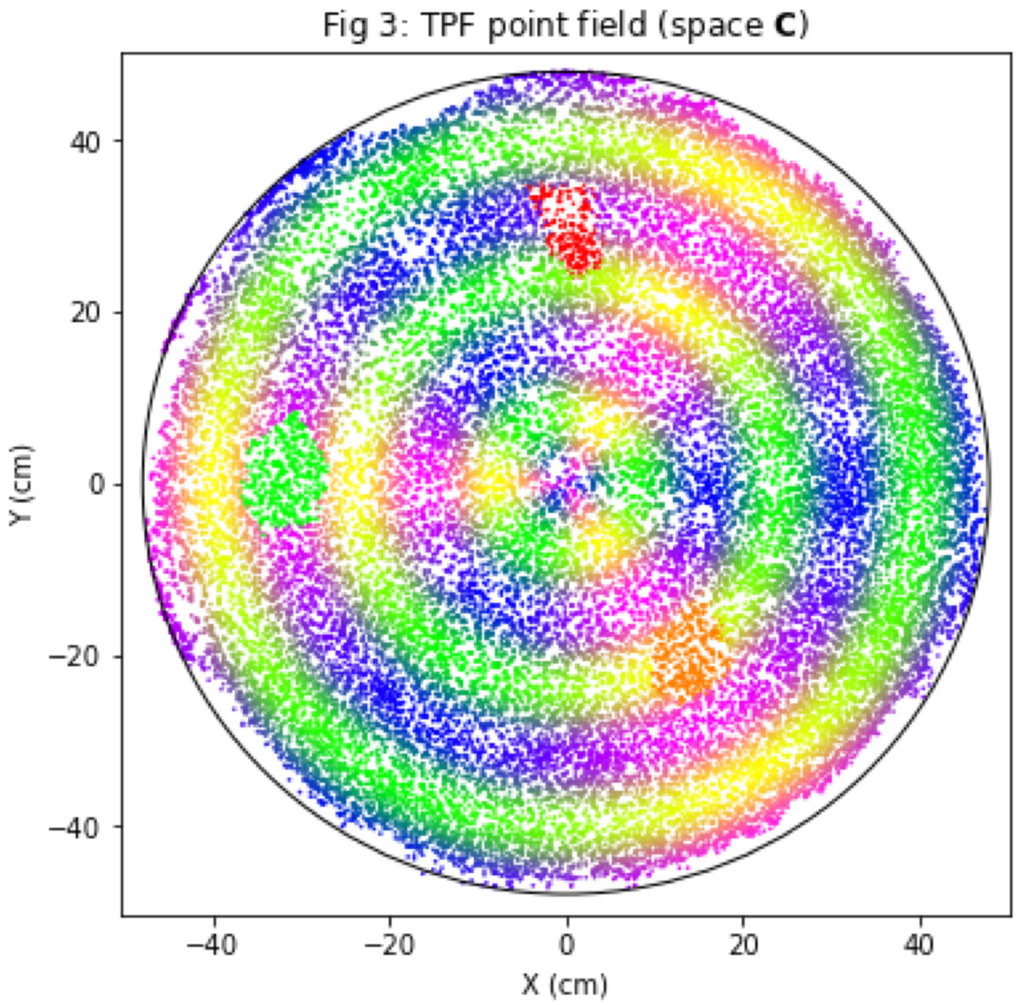
Twin peaks

Swiss roll

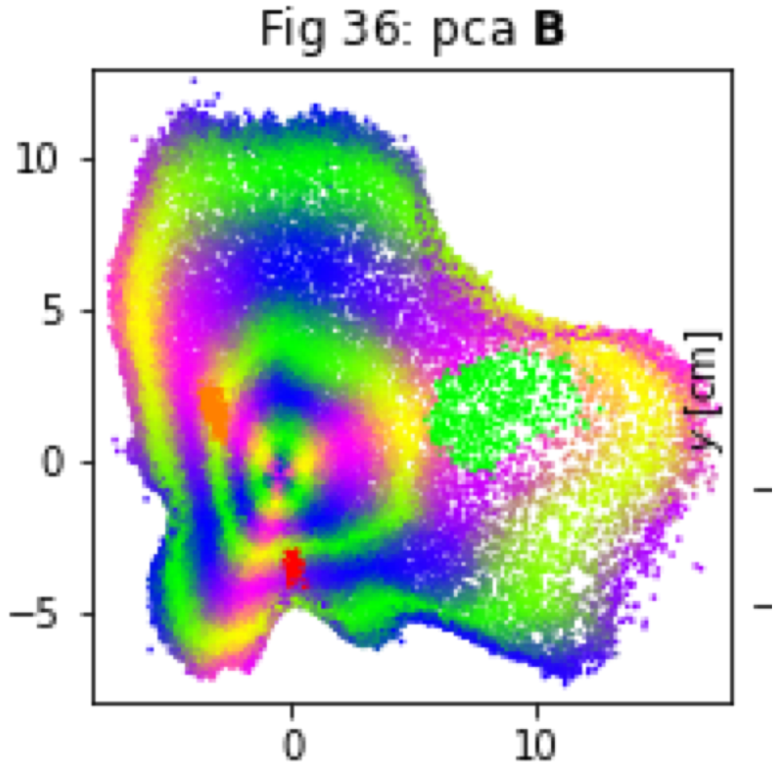
"Manifold learned"



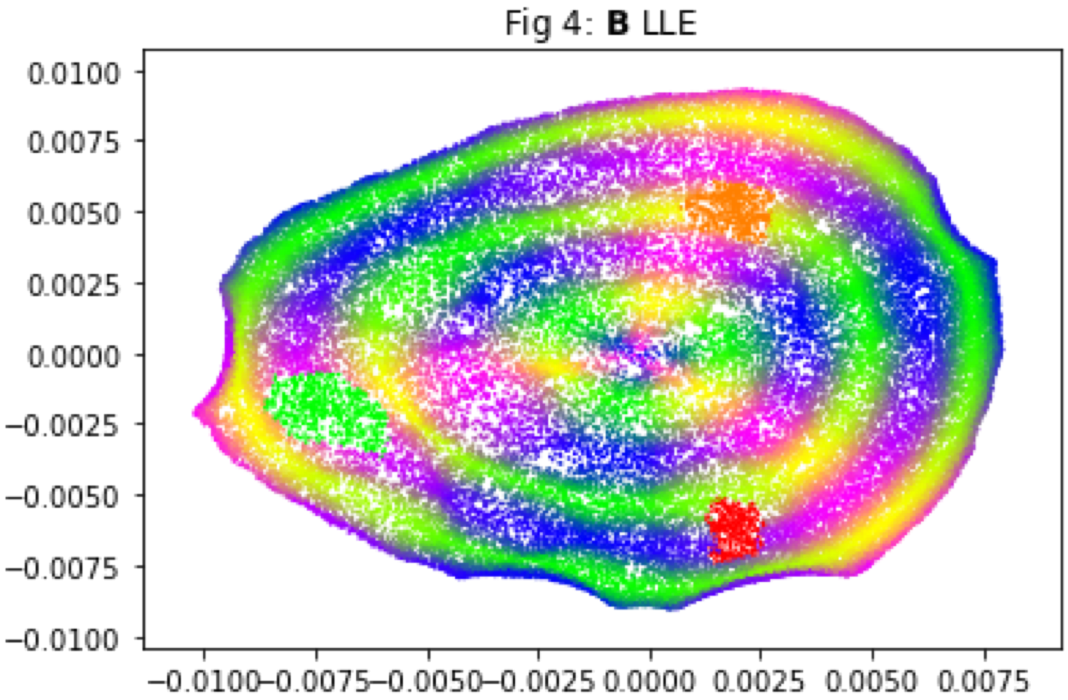
# Baseline



# Dimensionality reduction: Principal Component Analysis

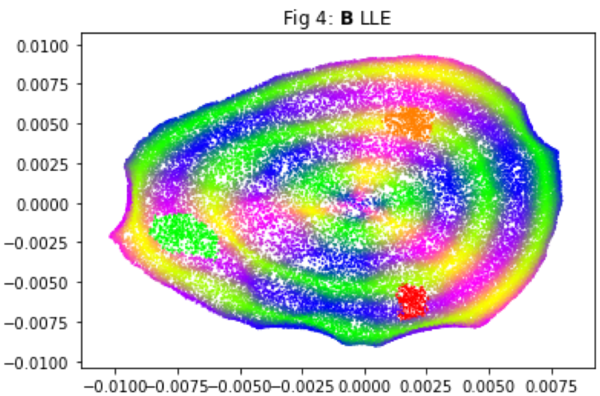


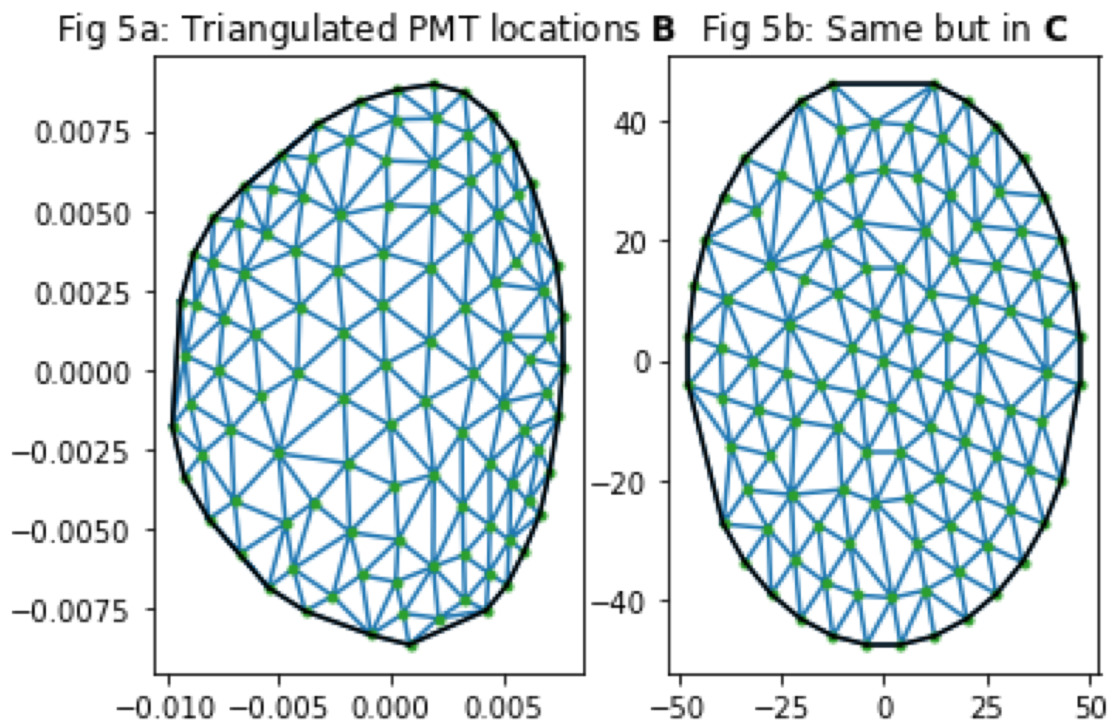
# Dimensionality reduction: Local linear embedding




Scale off…. now can use physics knowledge!!

# Stretching image




Triangulation interpolation from known positions in learned and lab space

# Applications / Conclusion

- DAQ: can find flipped channels
- Can “learn” optics/response of uncalibratable detector
- Can be used for detector alignment
- Only assumption neighborhoods preserved
  - P.S. ask me about spatial statistics and Ripley K functions if you ever want to check your manifold after learning on ‘uniform’ distribution