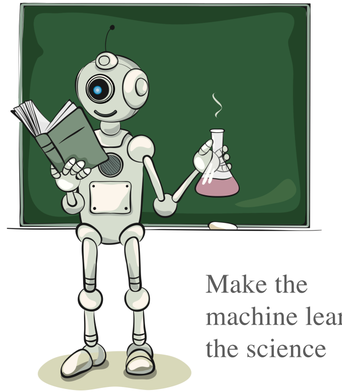
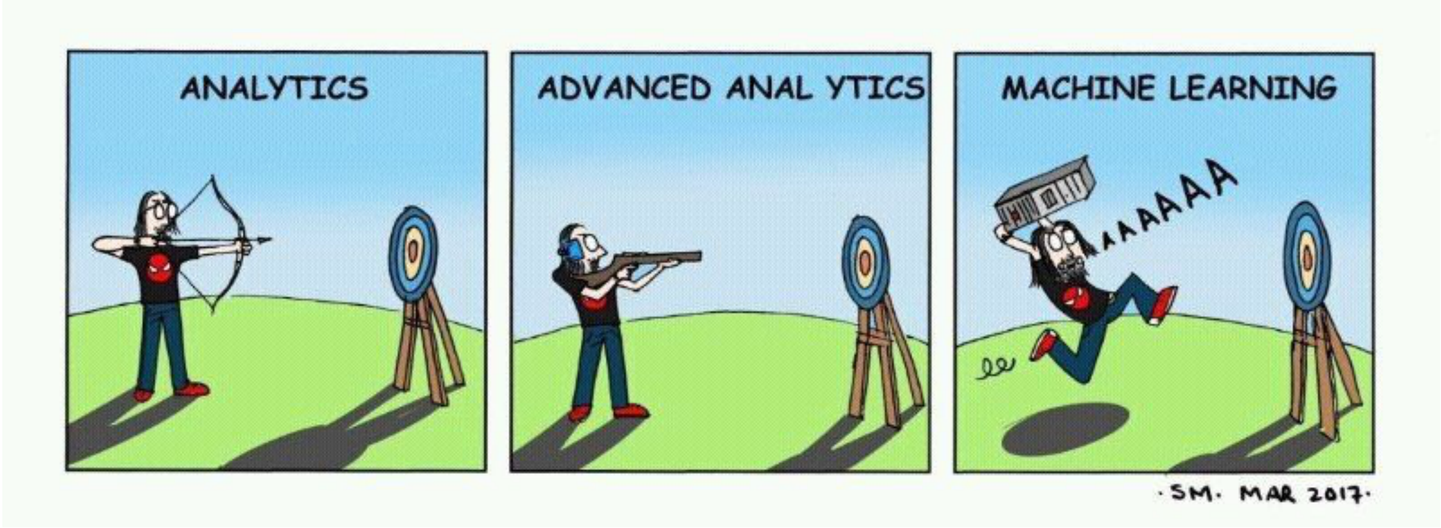


Make the
machine learn
the science

DIDACTS

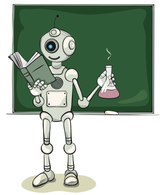
Data-Intensive Discovery Accelerated by Computational Techniques for Science

Aaron Higuera
On behalf of DIDACTS



How do we realize the full impact of machine learning in the physical sciences?

DIDACTS



DIDACTS *is a collaboration of physicist and machine learning experts with an overall goal of incorporate scientific knowledge into machine learning and data science methods in the context of scientific disciplines*

RICE

Prof. C. Tunnell
A. Higuera
S. Liang
A. Oranday

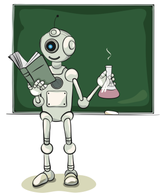
RUTGERS

Prof. W. Bajwa
V. Roy

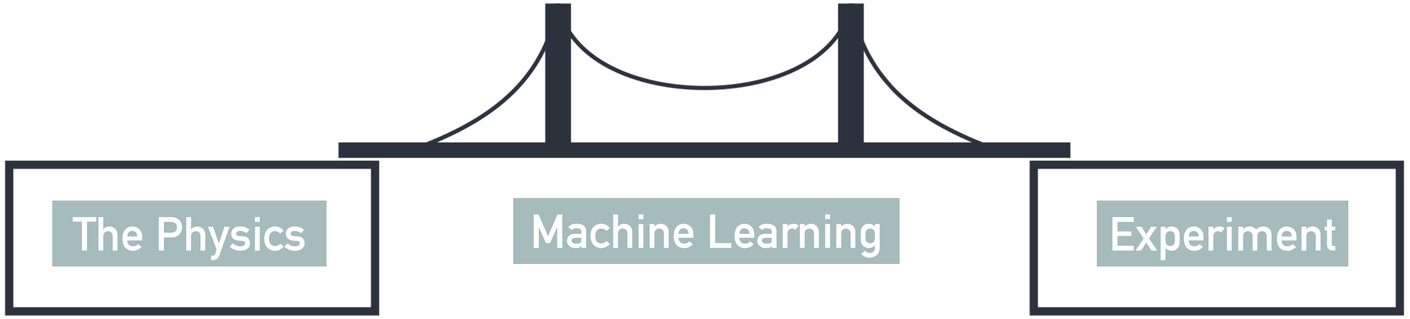
UNIVERSITY OF DELAWARE

Prof. H. Shatkay
C. Peters
M. Pergeorelies

DIDACTS

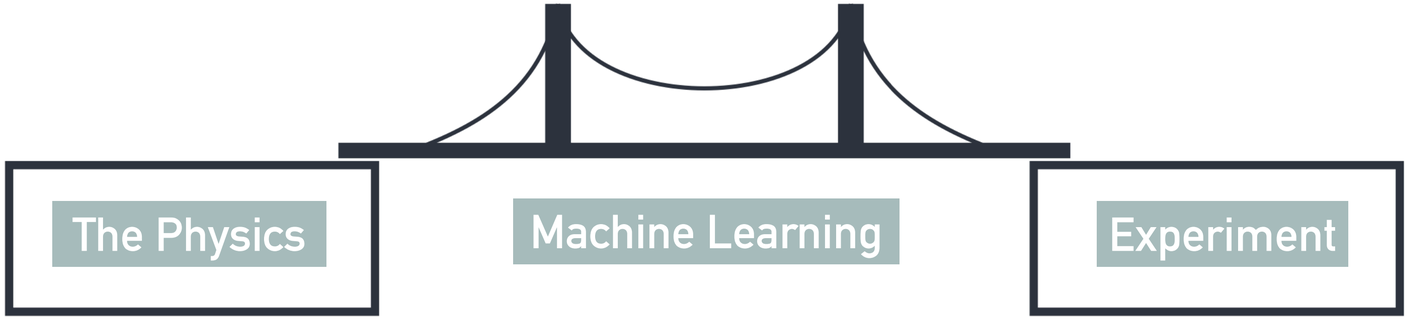


DIDACTS *is a collaboration of physicist and machine learning experts with an overall goal of incorporate scientific knowledge into machine learning and data science methods in the context of scientific disciplines*



- ***How do we realize the full impact of machine learning in the physical sciences?***
- ***How do we put our own knowledge in and get uncertainty out?***
- ***How we incorporate the physics we know into ML in such that it can uncover the physics we do not know?***

DIDACTS

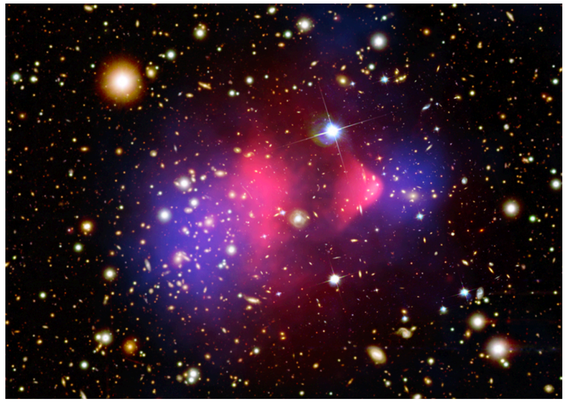


- ***How do we realize the full impact of machine learning in the physical sciences?***
- ***How do we put our own knowledge in and get uncertainty out?***
- ***How we incorporate the physics we know into ML in such that it can uncover the physics we do not know?***

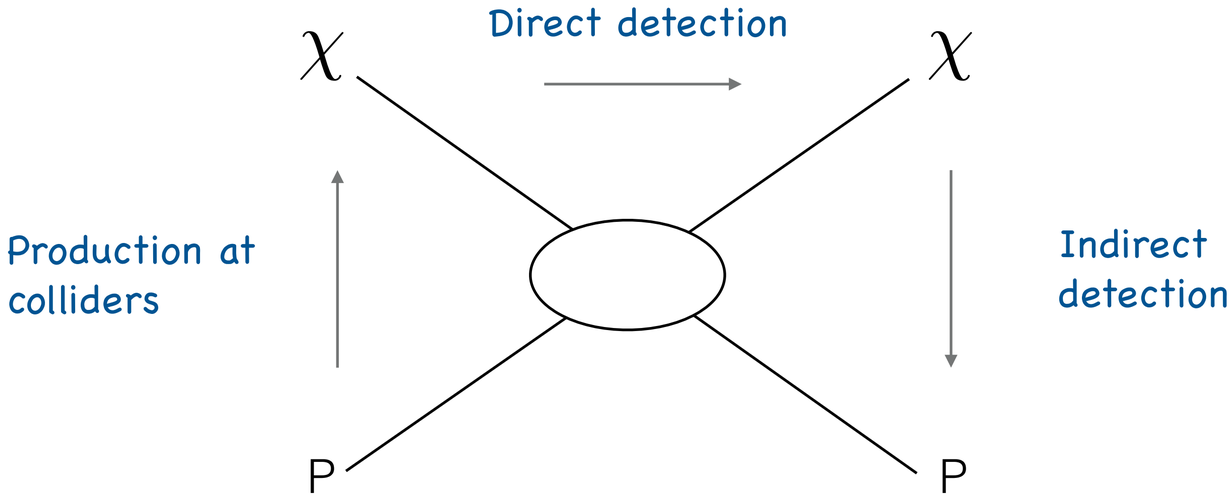
- **One example: Dark Matter**
  - **For every kilogram of ordinary matter, the Universe contains ~5.4 kg of Dark Matter**
  - **At DIDACTS we are looking into the challenging problem of Dark Matter direct detection using XENON1T experiment as test bed**

DARK MATTER

- 85% of the Universe consists of something that we do not understand: Dark Matter (DM)
- If DM would be a particle unlike any we have encountered in the SM, this would revolutionary our understanding of the universe
- How can we search for DM?

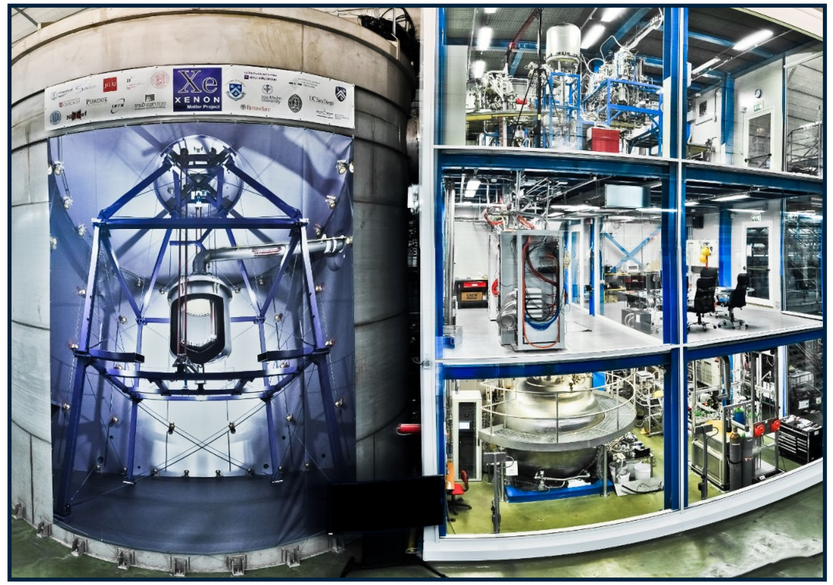


https://chandra.harvard.edu/photo/2006/1e0657/

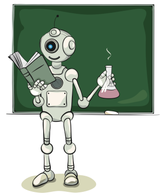


The XENON Experiment

- An international collaboration of ~160 scientist from 26 institutions
- Located in the LNGS in Italy (deep underground 3600 mwe)
- XENON1T detector is a LXe dual-phase TPC with 1.3 ton LXe of fiducial mass

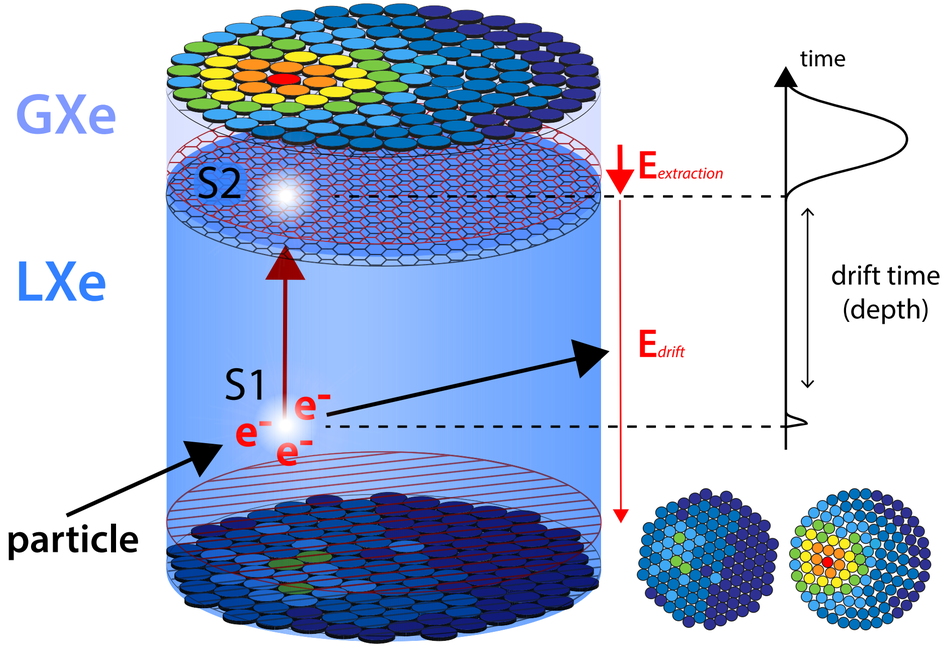


Dual-phase Time Projection Chamber

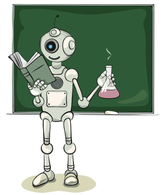


- Dual-phase Time Projection Chamber
- 0.125 kV/cm E-field w/ 150 cm drift length
- Prompt signal (S1): from scintillation (UV light)
- Delayed signal (S2): electrons are drifted and extracted in the gaseous phase producing second scintillation
- This detection system allows
  - Particle ID $(S2/S1)_{NR,WIMP} < (S2/S1)_{ER}$
  - Energy reconstruction
  - 3D position reconstruction (X, Y, Z)

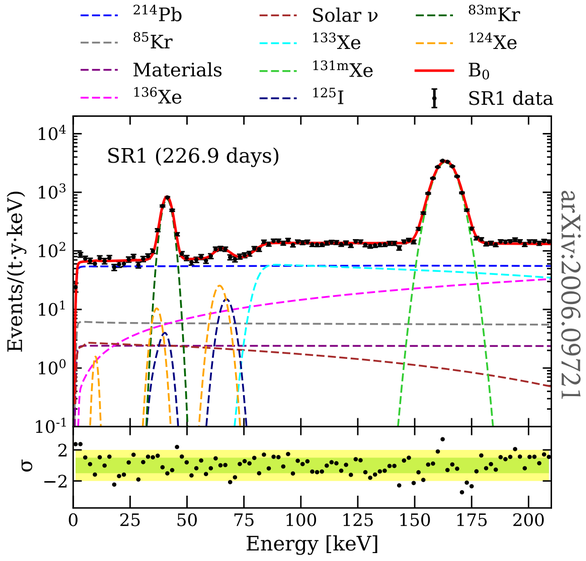
$$E = W\left(\frac{S1}{g_1} + \frac{S2}{g_2}\right)$$

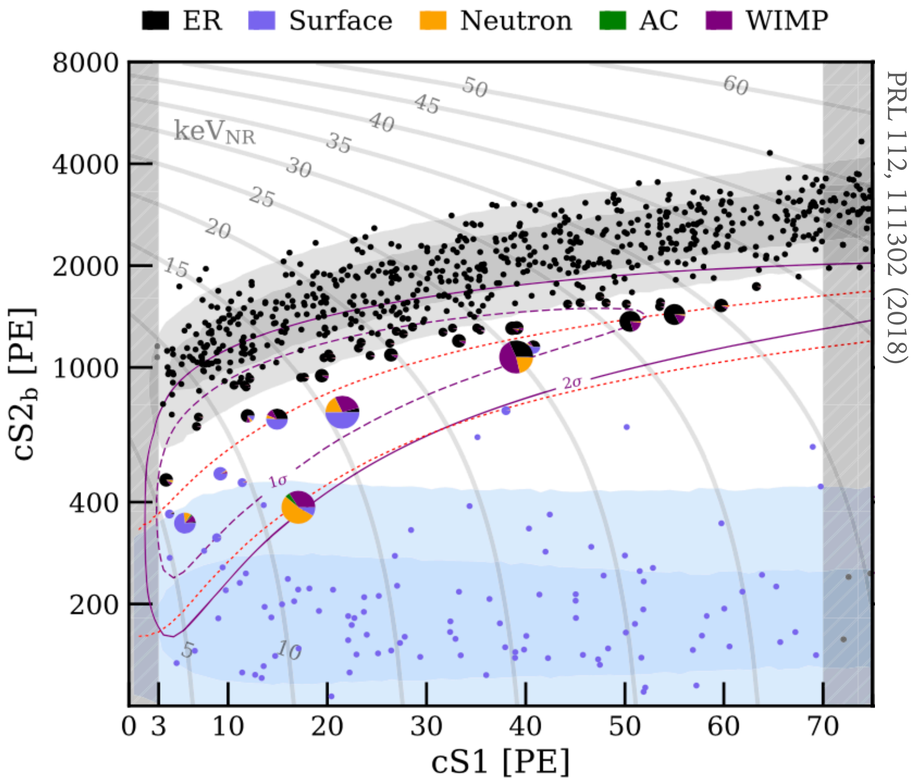


DM DIRECT DETECTION CHALLENGES

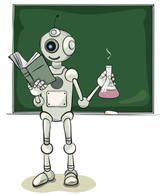


- Direct detection:
  - Need ultra-radio pure environment
  - Backgrounds are ER-like and NR-like
  - 10-50 Hz background rates vs 1 DM event (maybe)
  - Low energy searches [KeVs]

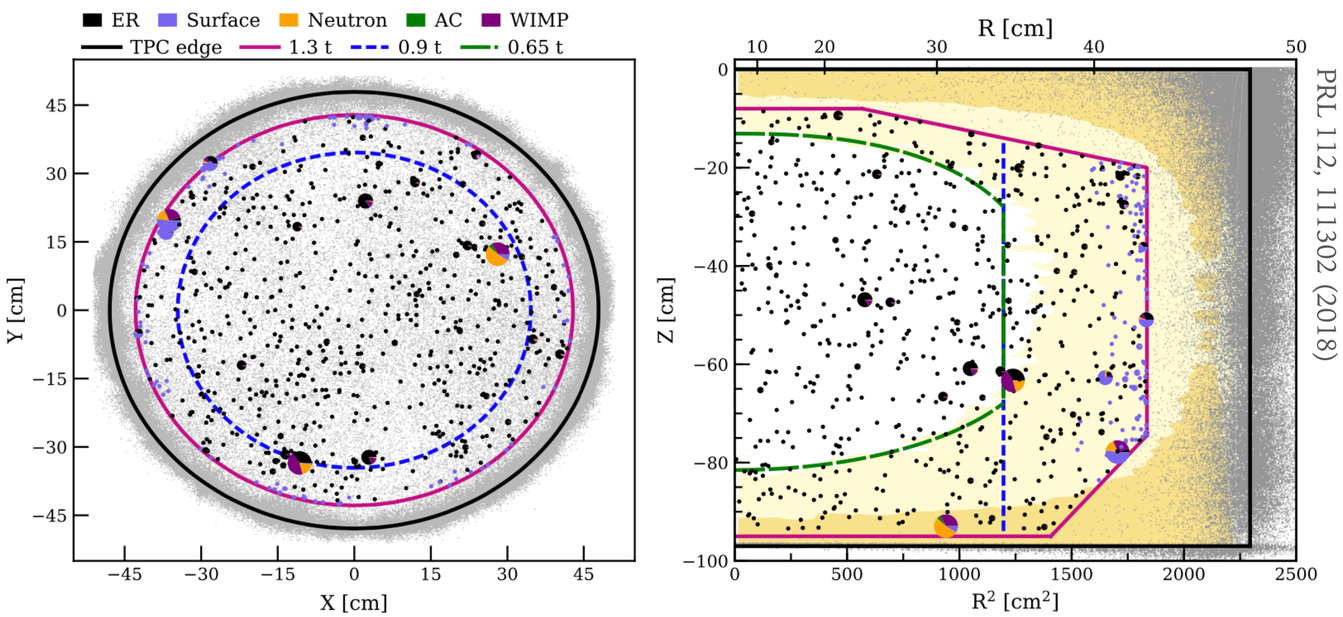




DM DIRECT DETECTION CHALLENGES



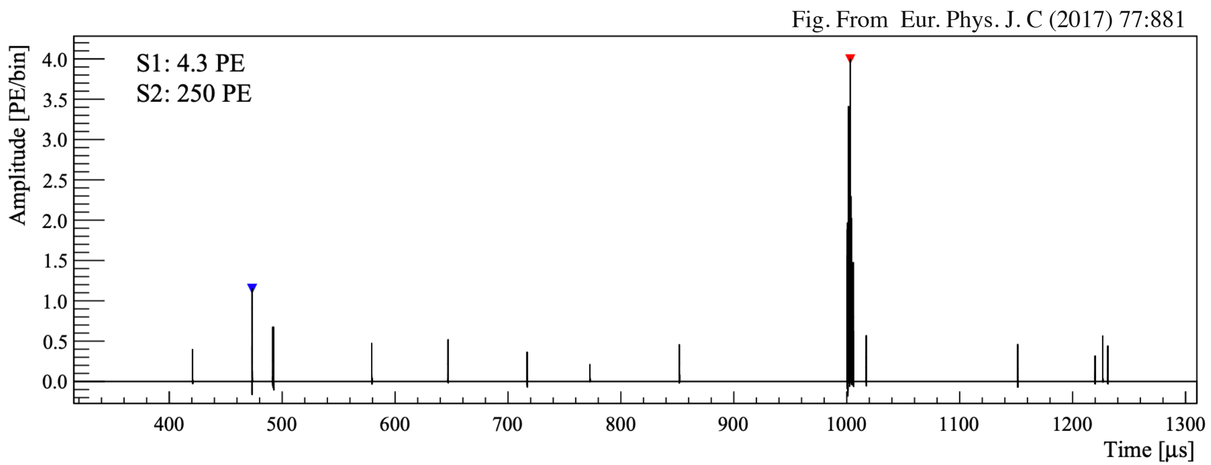
- Position reconstruction is a key player to reduce background
  - Fiducial volume

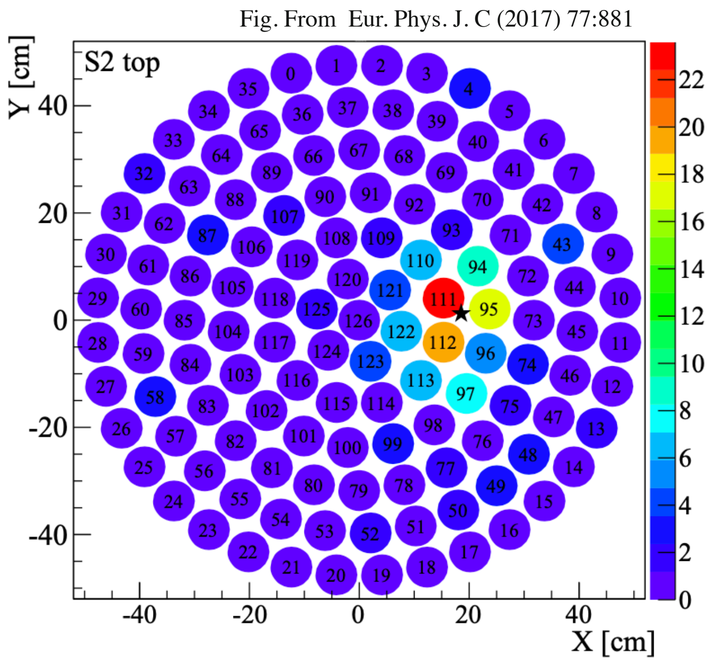


- What about 0νββ?
- LXe DM experiments aim for competitive sensitivity, via restricted fiducial volume to reduce backgrounds, and projected 1% energy resolution
- The sensitivity to the 0νββ process is proportional to both energy and position resolutions

Position Reconstruction

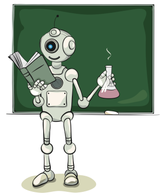
- The Z-position of the interaction is calculated from the drift time, it requires S1 and S2





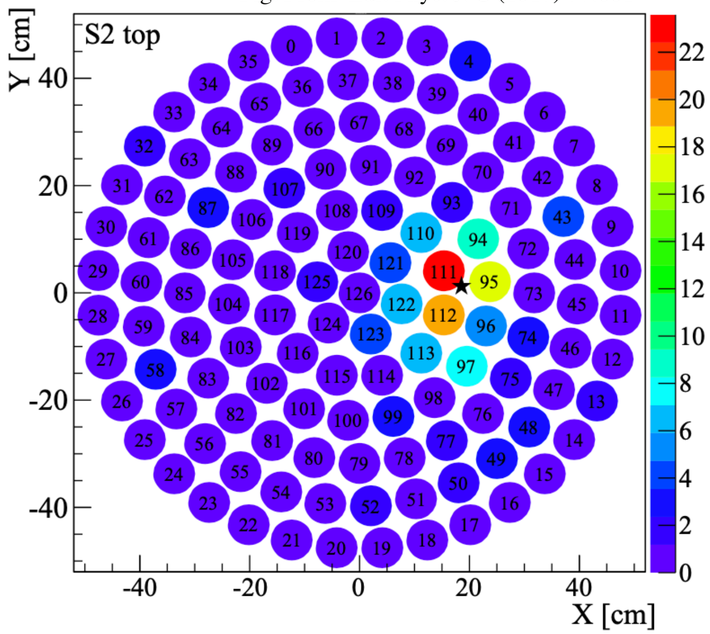
- X,Y position can be reconstructed from S2 (hit pattern)
  - Maximum PMT method
  - Maximum PMT charge-weighted sum
- The aforementioned methods can only give 127 unique positions (PMT positions) and are biased towards the center

Position Reconstruction

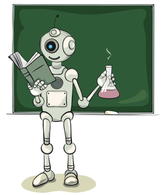


- The Z-position of the interaction is calculated from the drift time, it requires S1 and S2
- X,Y position can be reconstructed from S2 (hit pattern)
- “Classical ML approach” : Multi-layer Neural Network
  - Four layers,
  - 127 input layers (#PMTs)
  - two hidden layers and two output layers (X,Y)

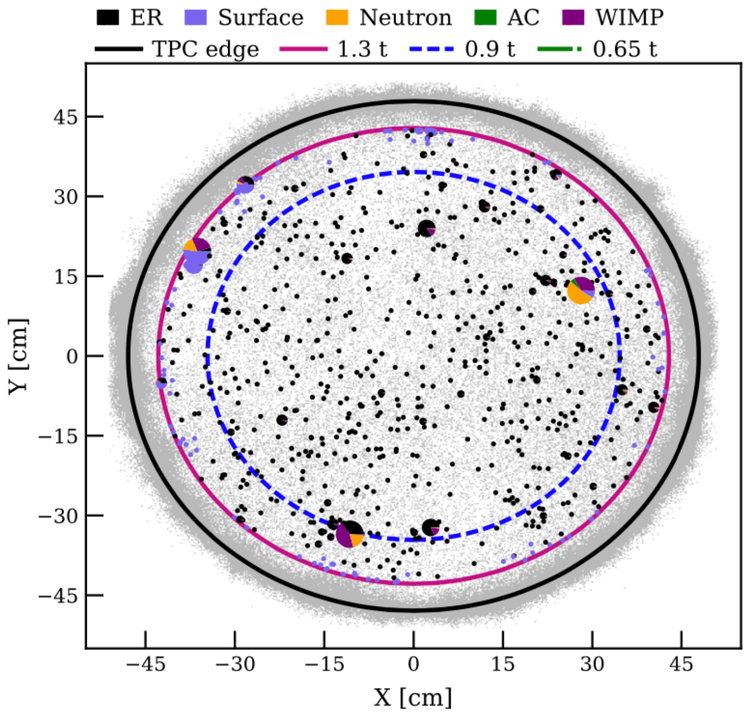
Fig. From Eur. Phys. J. C (2017) 77:881

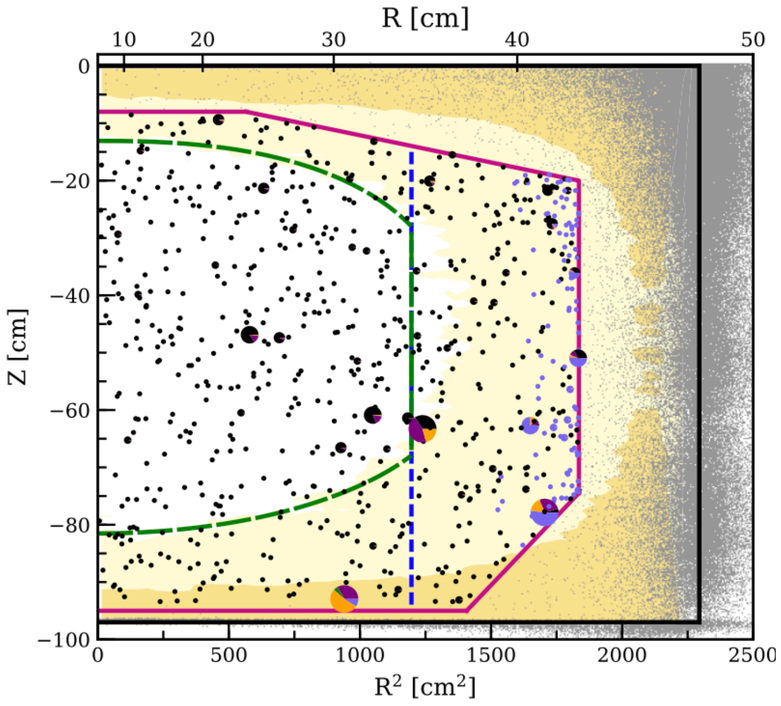


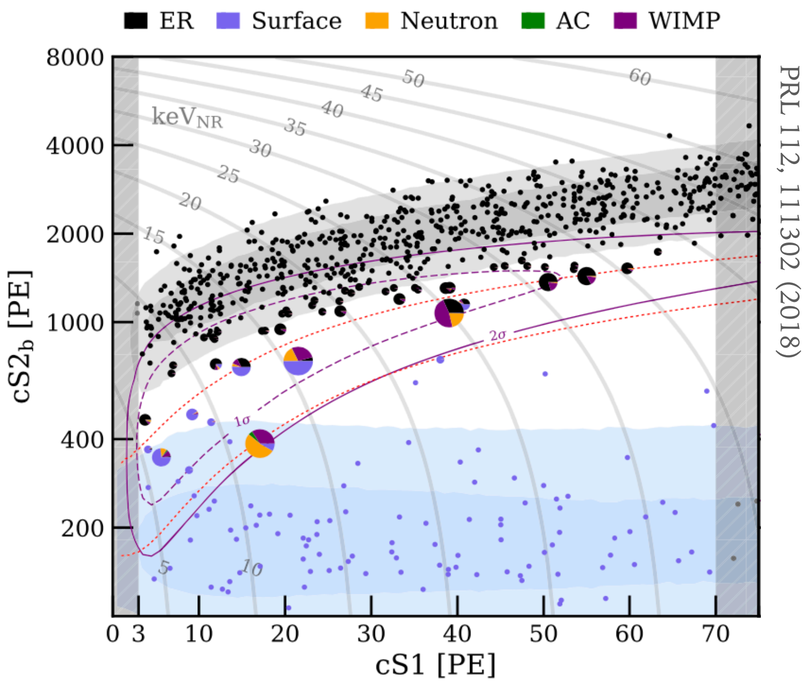
Position Reconstruction



- “Classical ML approach” Multi-layer Neural Network

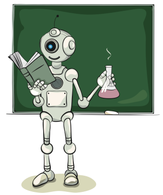






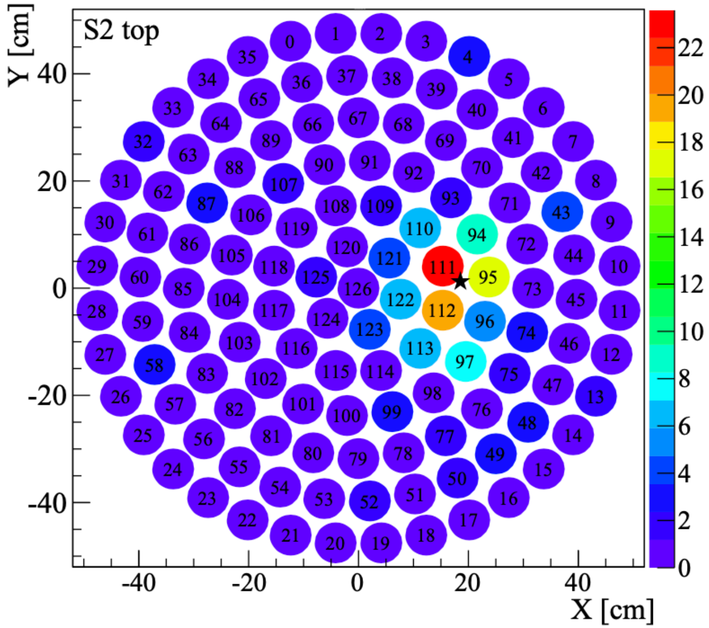
- **Any improvements on position reconstruction would be critical**

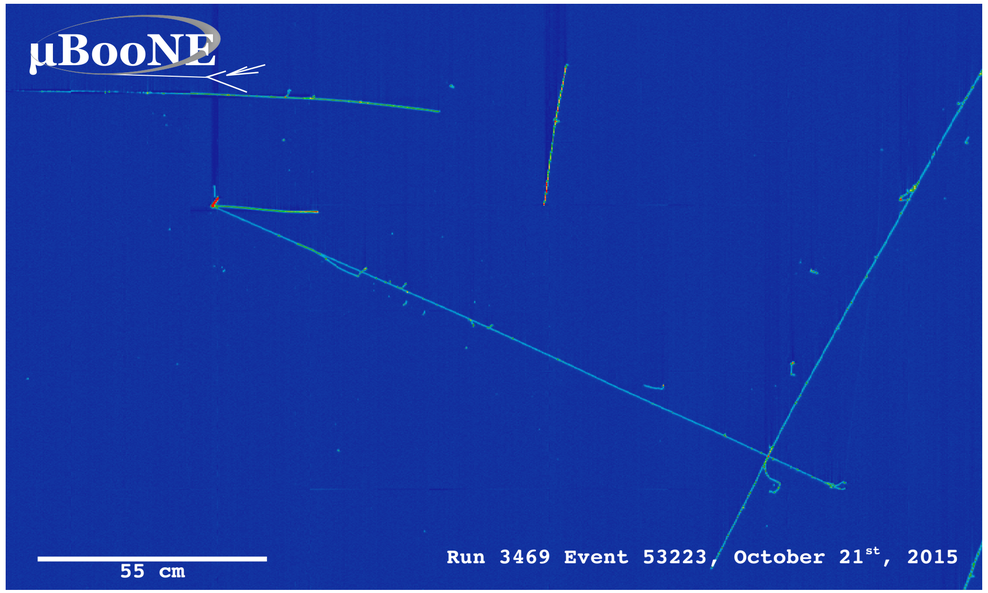
Position Reconstruction



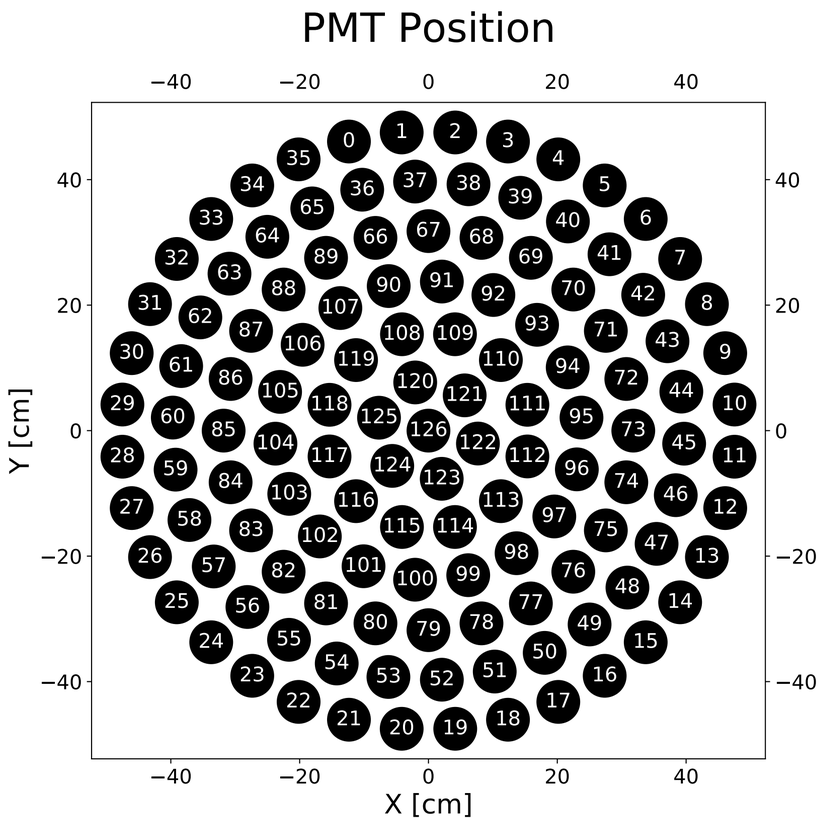
- Why no CNN?
  - No grid-structured data
  - Converting unstructured and sparse data can lead to loss of information

Fig. From Eur. Phys. J. C (2017) 77:881

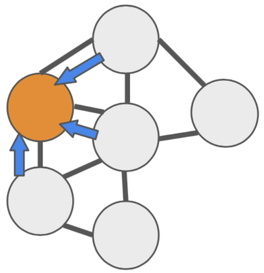




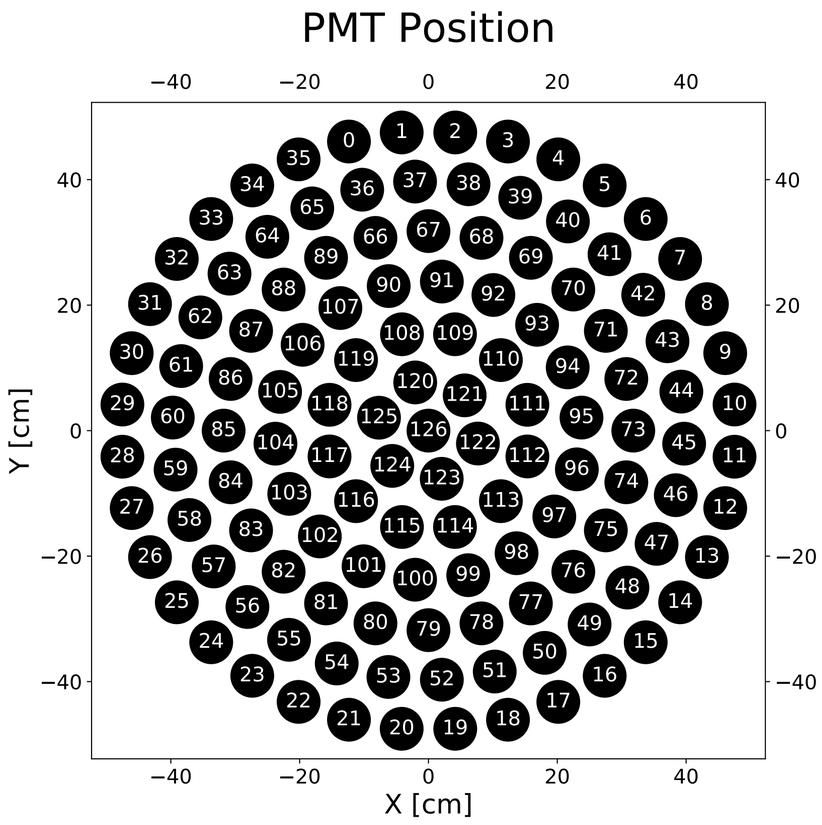
Graph Neural Networks



- Graphs
  - **Nodes** are quantized objects with some arbitrary features
  - **Edges** describe the relationship between nodes
- Data graph-structured

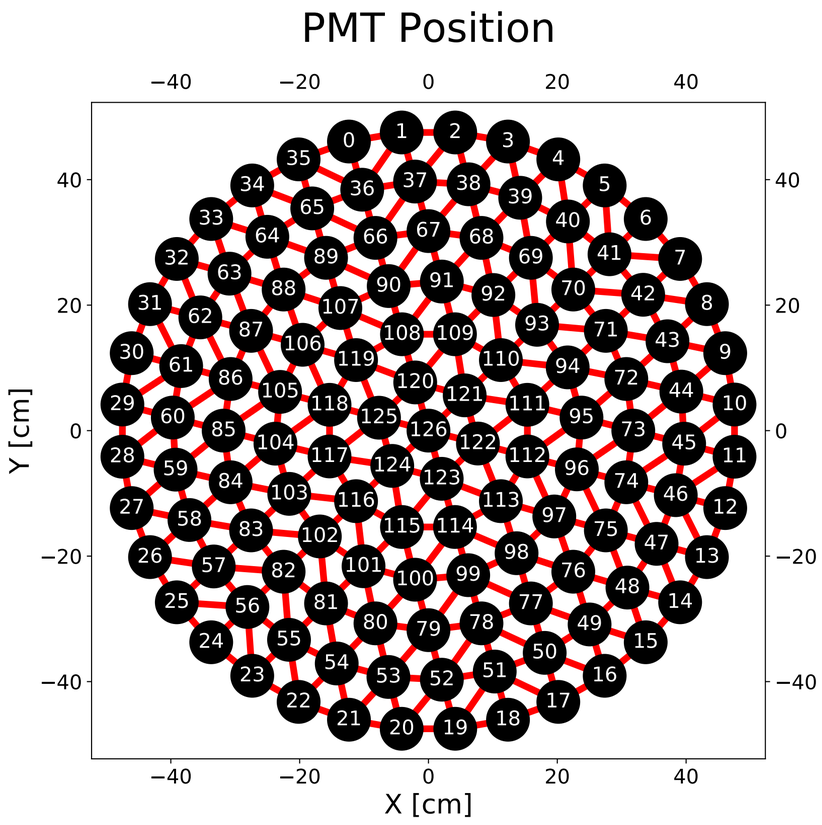


Graph Neural Networks



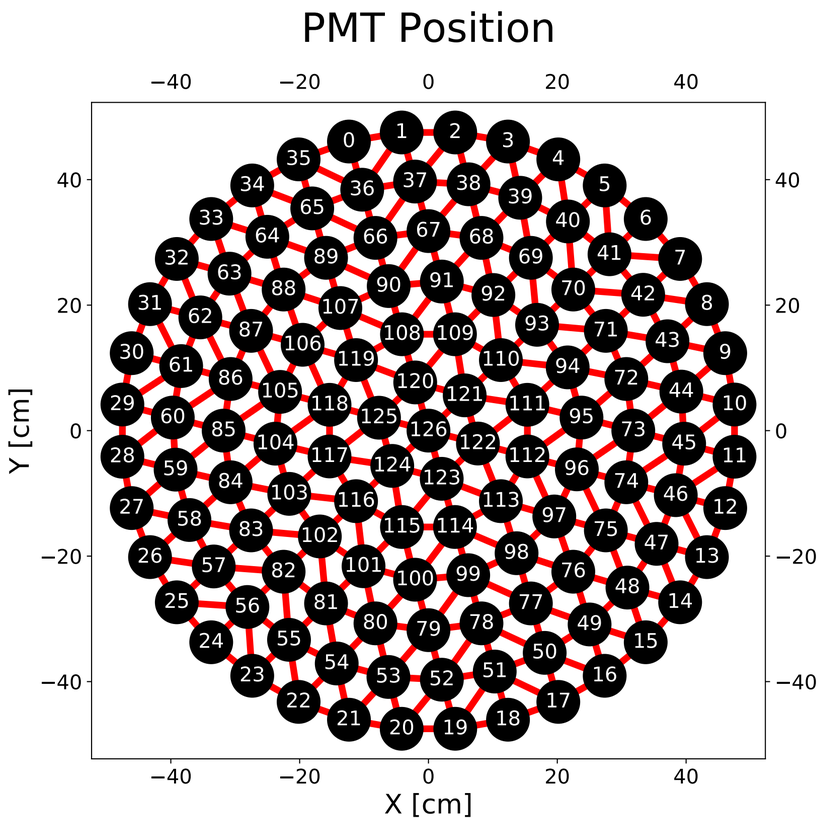
- Each node is a PMT
- Input features:
  - PMT location (X,Y)
  - S2 area [pe] (waveform area)
- Edges, describe the relationships between nodes
  - What is the best way to describe it?

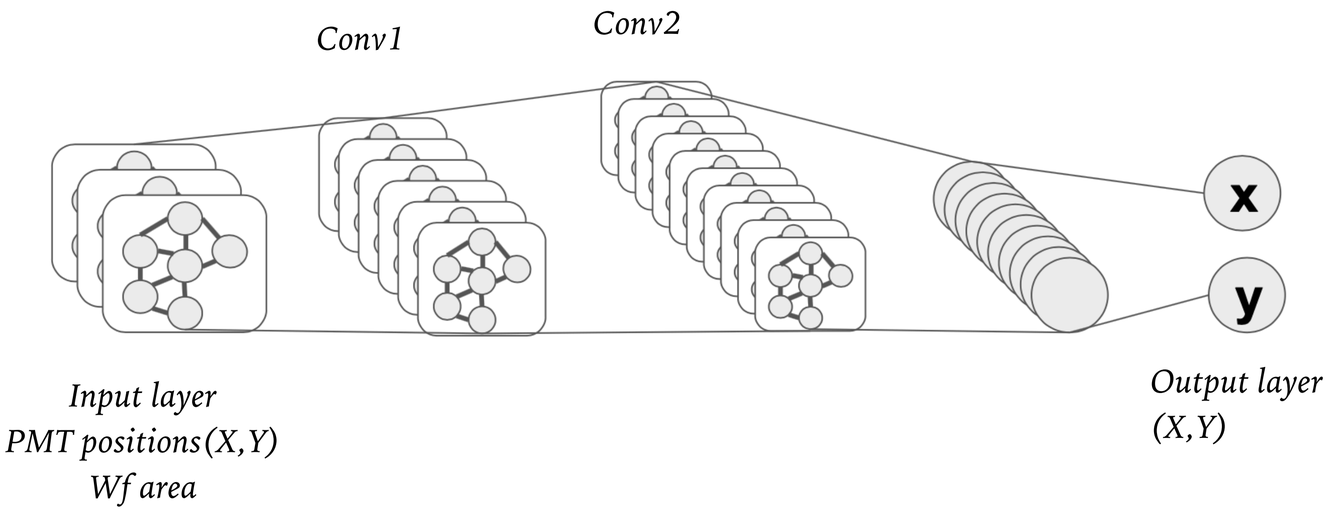
Graph Neural Networks



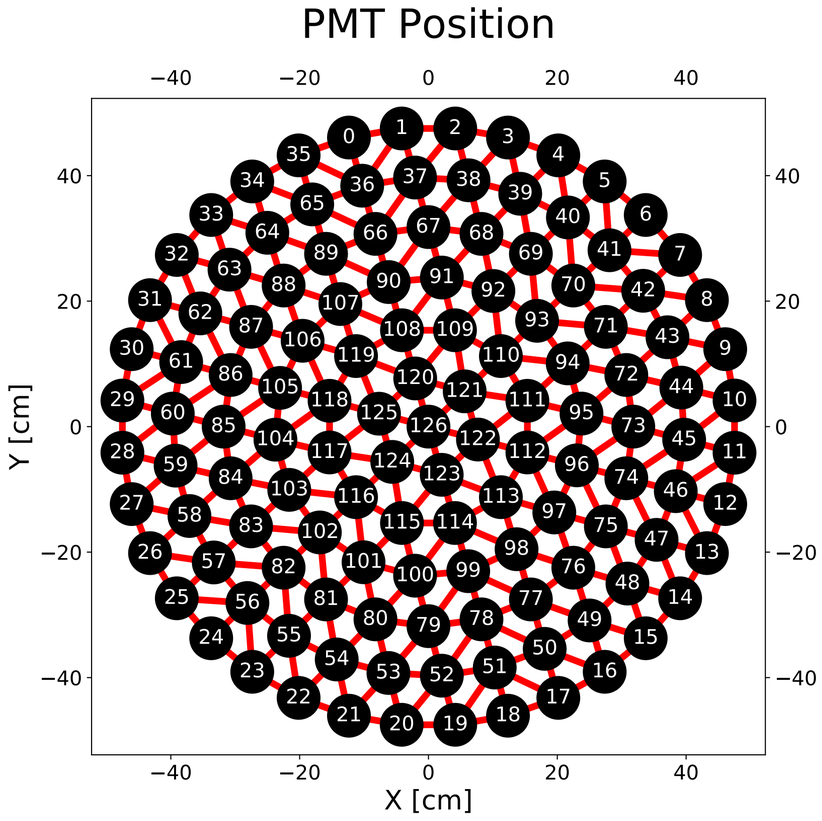
- Each node is a PMT
- Input features:
  - PMT location (X,Y)
  - S2 area [pe] (waveform area)
- Edges, describe the relationships between nodes
  - What is the best way to describe it?
  - Use Delaunay triangulation, this was chosen for being well connected without being definitely over-connected
- MC simulation: GEANT-based optical simulation

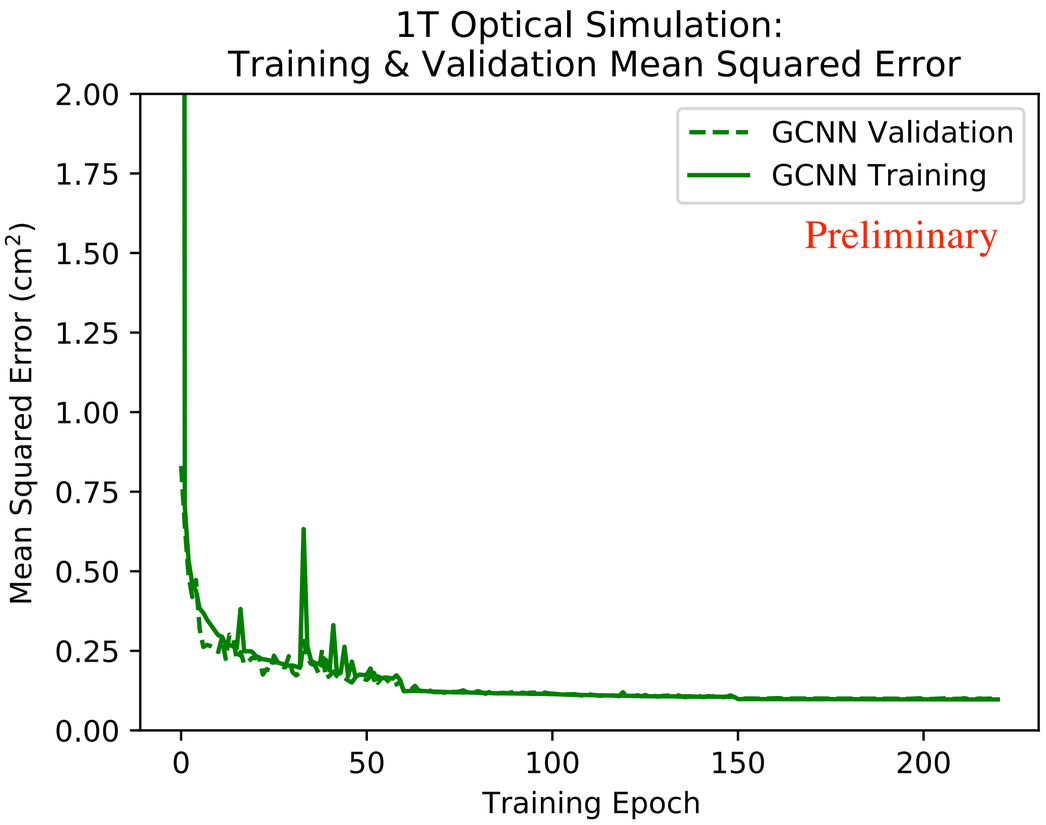
GCNN


PMT Position
Y [cm]
X [cm]

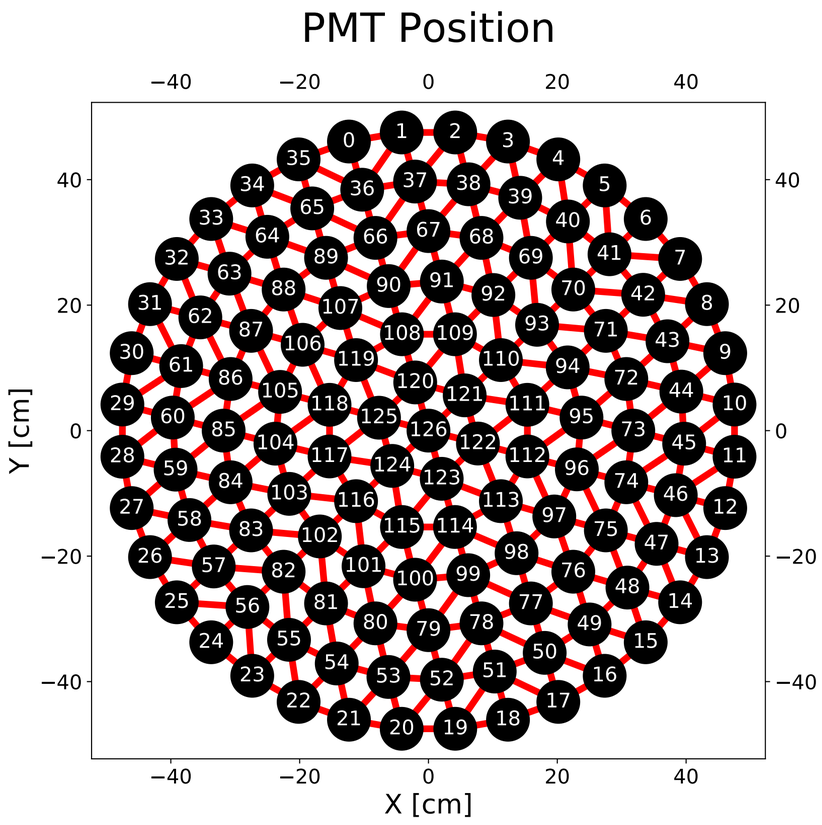

Conv1
Conv2
x
y
Input layer
PMT positions (X,Y)
Wf area
Output layer
(X,Y)

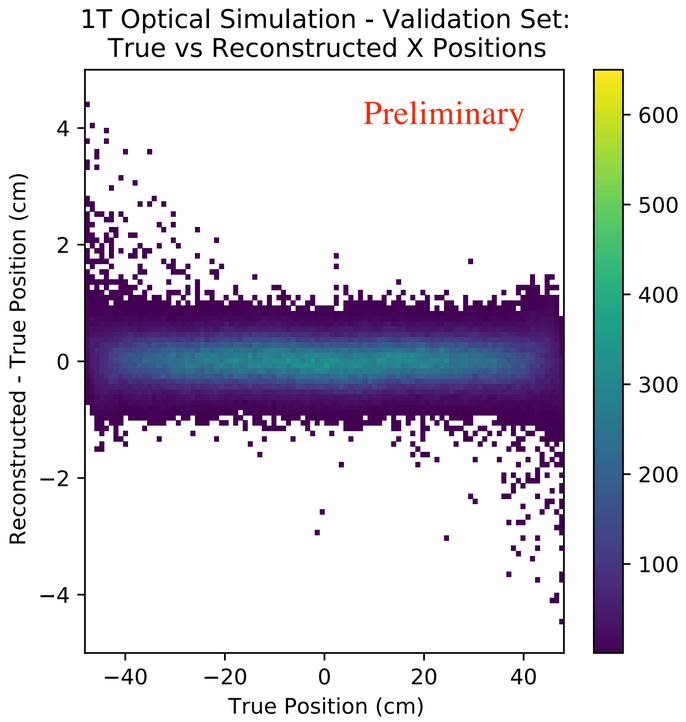
GCNN

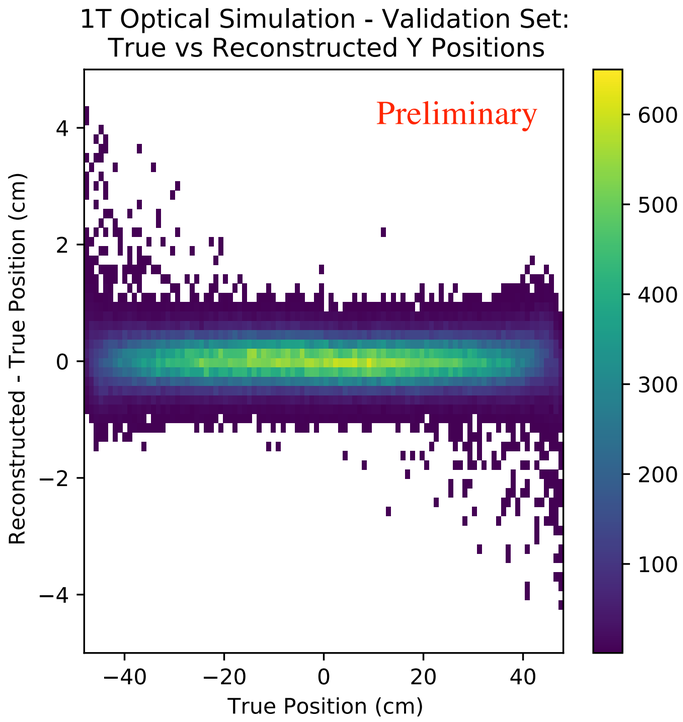

PMT Position
X [cm]
Y [cm]


1T Optical Simulation:
Training & Validation Mean Squared Error
GCNN Validation
GCNN Training
Preliminary
Mean Squared Error (cm^2)
Training Epoch

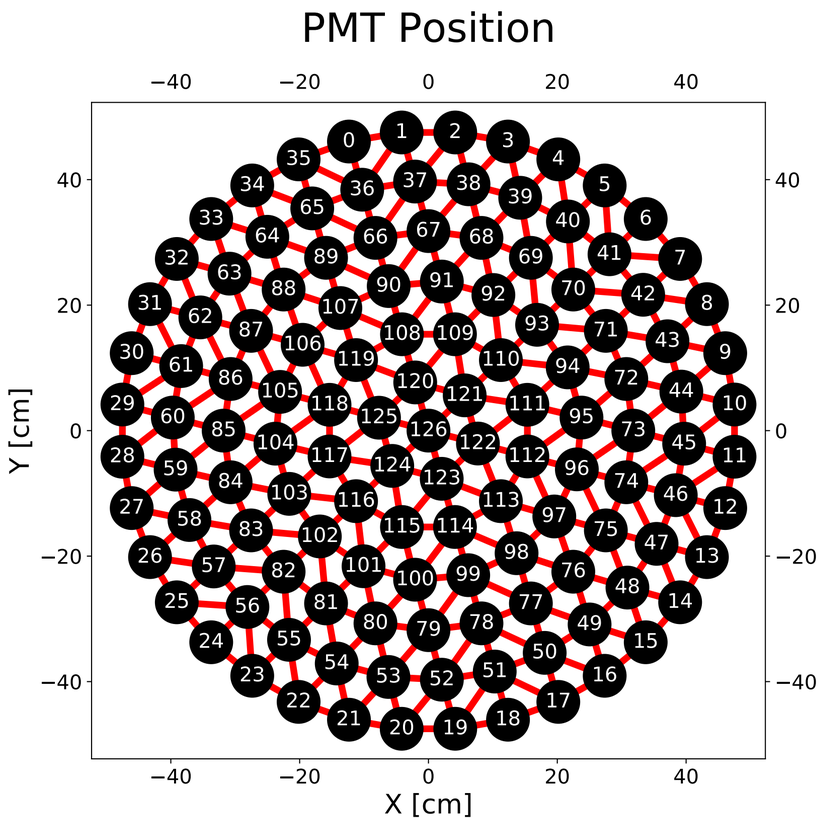
GCNN

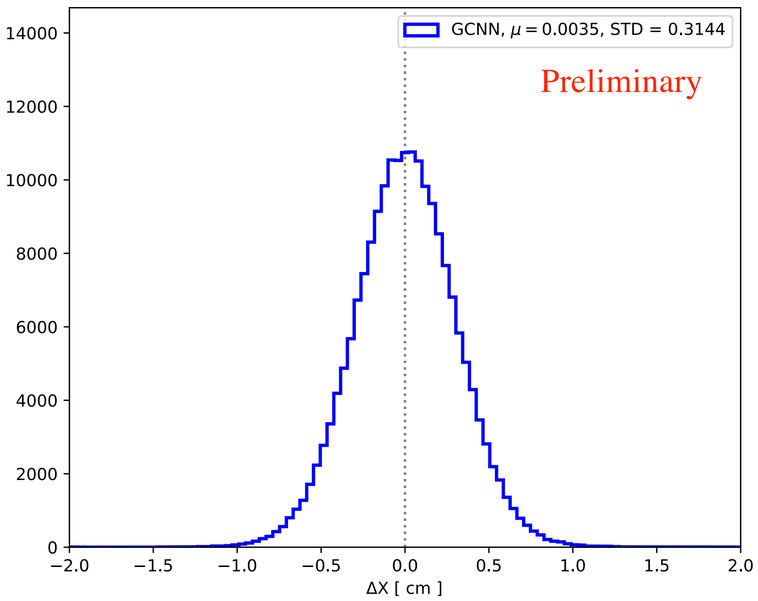

PMT Position
X [cm]
Y [cm]

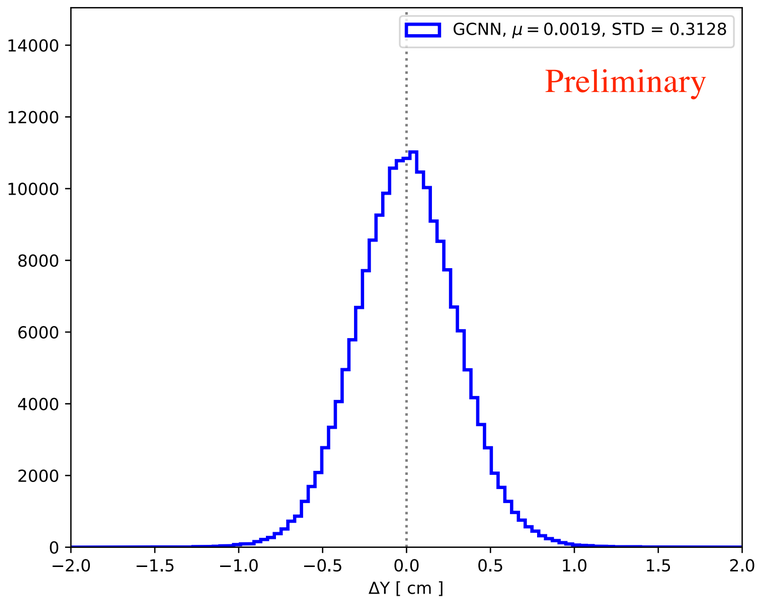

1T Optical Simulation - Validation Set:
True vs Reconstructed X Positions
Preliminary
Reconstructed - True Position (cm)
True Position (cm)


1T Optical Simulation - Validation Set:
True vs Reconstructed Y Positions
Preliminary
Reconstructed - True Position (cm)
True Position (cm)

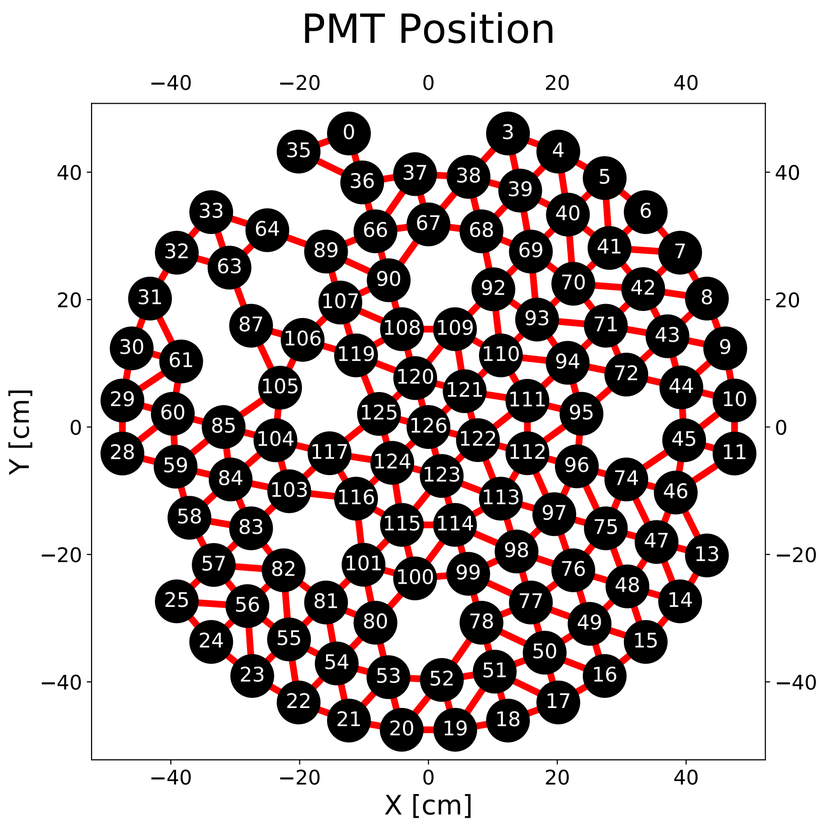
GCNN


PMT Position
X [cm]
Y [cm]


GCNN, $\mu = 0.0035$, STD = 0.3144
Preliminary
ΔX [cm]

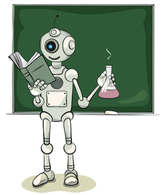

GCNN, $\mu = 0.0019$, STD = 0.3128
Preliminary
ΔY [cm]

GCNN

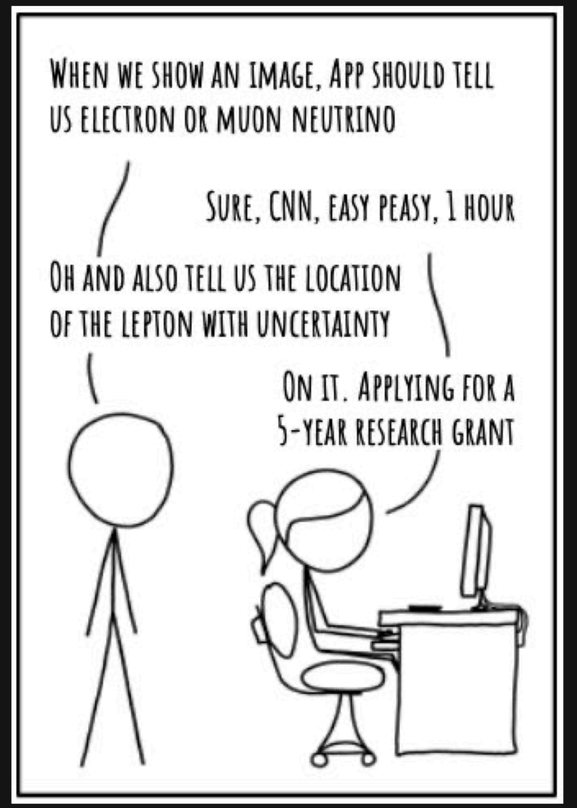


- Future work:
  - Dead PMTs (missing nodes)
  - Complete MC simulation: add detector effects (PMT gains, etc)
  - Graph optimization

Probabilistic Graphical Models

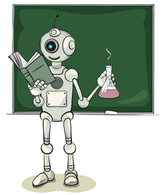


- ***How do we put our own knowledge in and get uncertainty out?***



Original image credit: xkcd

Probabilistic Graphical Models

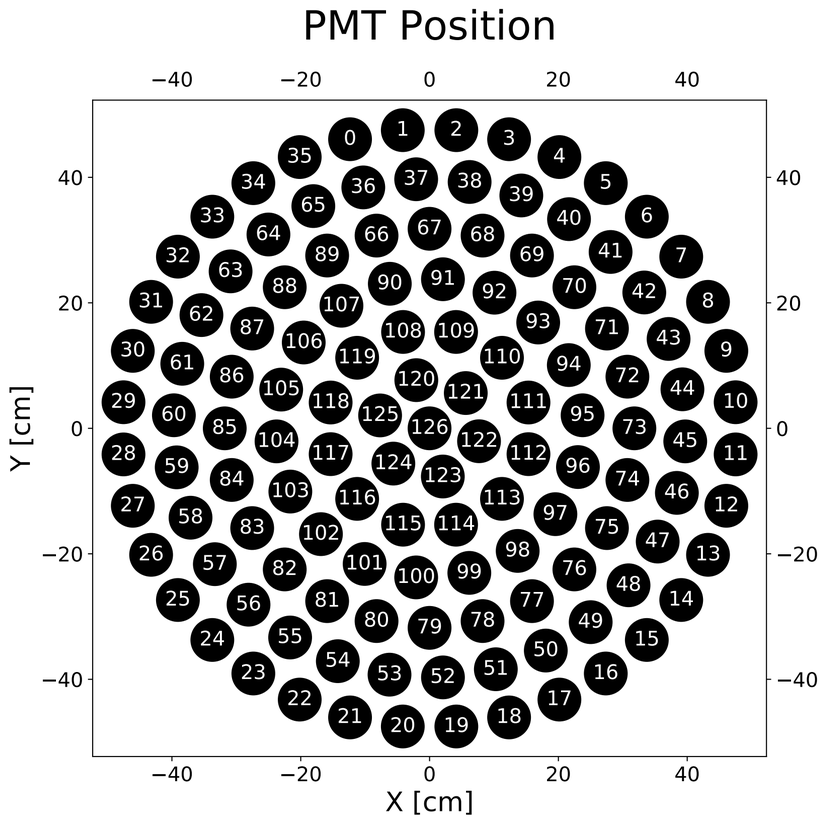


✦ ***How do we put our own knowledge in and get uncertainty out?***

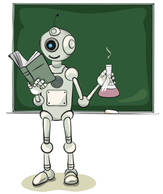
- Probabilistic Graphical Models
- Probabilistic: Encoding uncertainty via conditional probability

$$\Pr(X|Y = observations + knowledge)$$

- Graphical:
  - Nodes: random variables
  - Edges: dependencies
  - Missing edges: independencies
- A common example of PGM are Bayesian Networks:

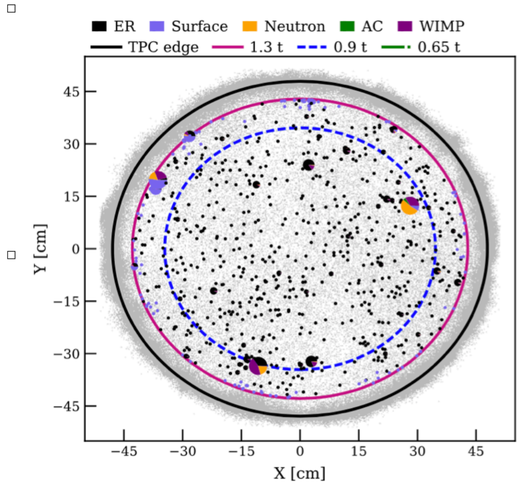


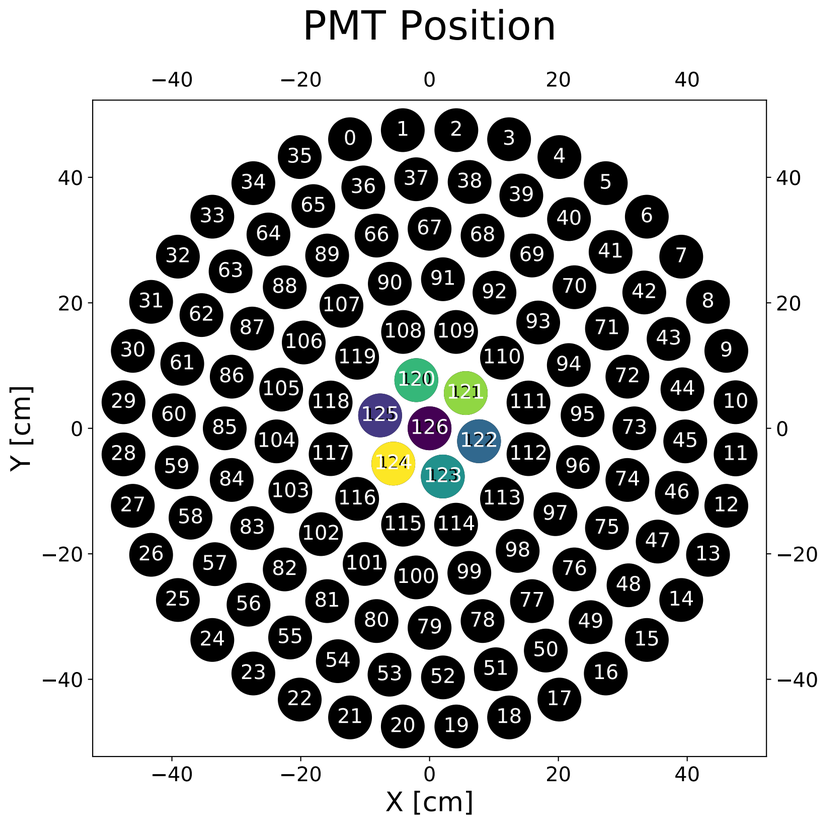
Probabilistic Graphical Models



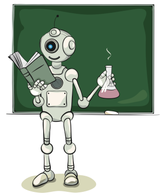
$$\Pr(X|Y = observations + knowledge)$$

- Having an algorithm that provides uncertainties in the reconstructed position is a plus
- Having an accurate position reconstruction with a quantified uncertainty will allow to maximize the use of detector volume fiducialization
- E.g. having a fiducial volume with 99% certainty

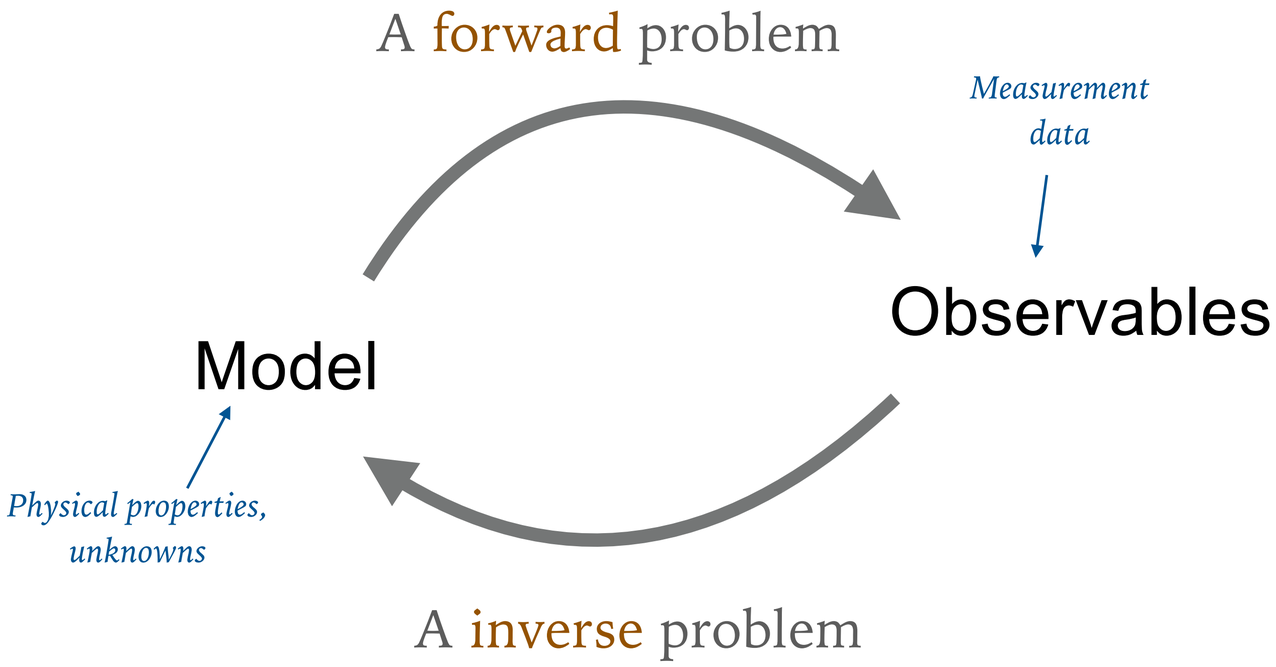




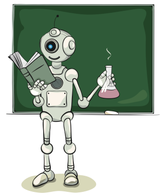
An Inverse Problem Approach



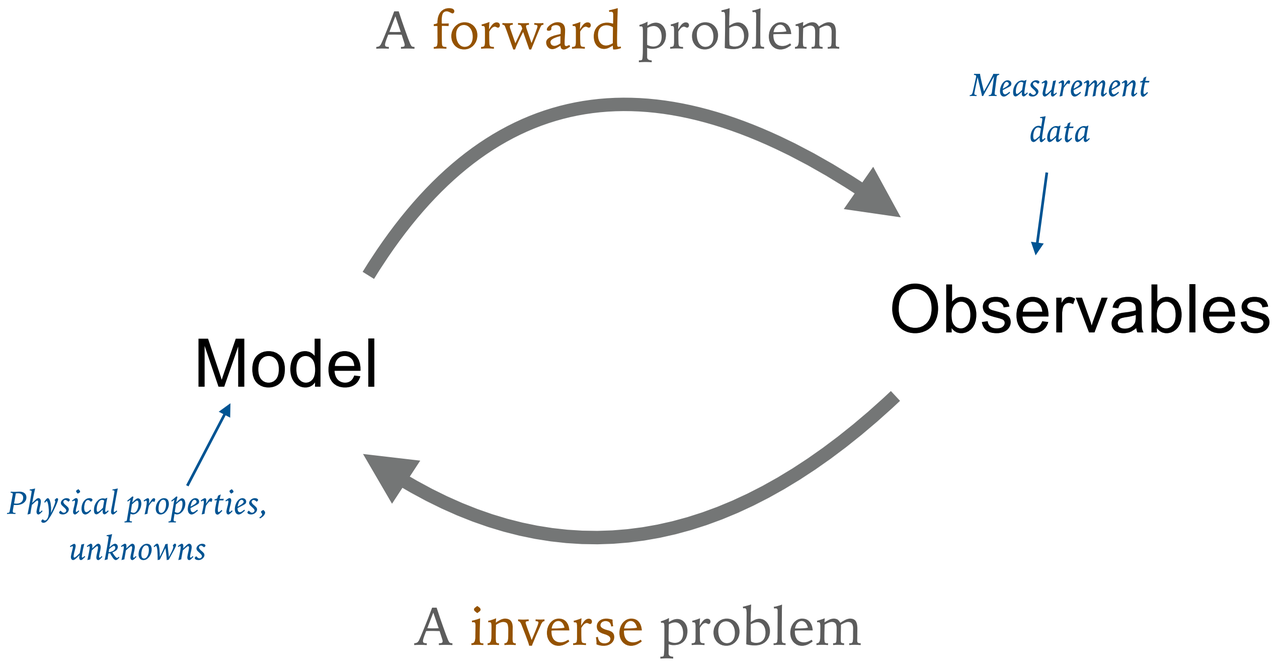
- The prediction of observations, given the values of the parameters that defined a model constitutes the “forward problem”
- The “Inverse problem” consists in using the results of actual observations to infer the values of the parameters characterizing the system (model)



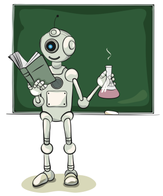
An Inverse Problem Approach



- Inverse problem does not required labeled training data
- For new physics searches e.g. DM searches, labels are only available in simulation and are model dependent
- In principle, by observing the data we can provide a estimation for the unknown parameters (position and energy) with an inverse problem formulation



Summary



- DIDACTS aims to test out multiple novel methods to enhance our knowledge
- DIDACTS is looking into the challenging problem of extreme rare searches using XENON1T as test bed
  - Position reconstruction is a key player to reduce background for extreme rare searches such as DM and 0vßß
  - Given graph-structured nature of XENON1T data DIDACTS is looking into:
    - Graph Neural Networks for position reconstruction
    - Can Probabilistic Graphical Models provide additional information(energy) with a quantified uncertainty?
    - An Inverse Problem approach for position and energy reconstruction
- More information about DIDACTS at https://didacts.org