CYGNO Experiment

GRAN SASSO SCIENCE INSTITUTE
SCHOOL OF ADVANCED STUDIES
Scuola Universitaria Superiore

INFN
Istituto Nazionale di Fisica Nucleare

erc
European Research Council
Established by the European Commission

IN TIUM

INITIUM: Innovative Negative Ion TIme projection chamber for Underground dark Matter searches

Multivariate techniques for track reconstruction and particle identification

Annual Talk Astroparticle Physics Cycle -XXXV

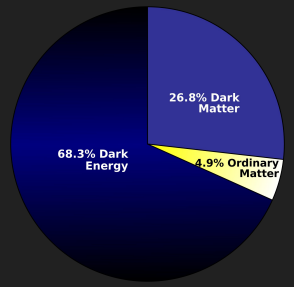
Candidate: Atul Prajapati
Supervisor: Prof. E. Baracchini

Date: 18/10/2021

Outline

1) Dark Matter introduction
2) CYGNO/INITIUM detector
3) Data Analysis and Simulation
4) Development of Discriminating variables
5) Nuclear recoil and Electron recoil discrimination using ANN
6) Future work

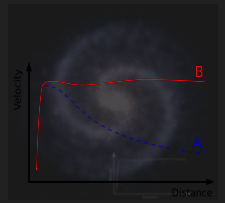
Dark Matter Introduction



Composition of the Universe

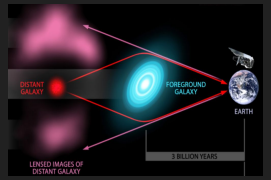
- Observations from galactic to cosmological scales indicate that primary source of gravity is from Non Baryonic Mass
- It is Neutral
- Must be stable or have lifetime more than age of the Universe
- Must be weakly interacting

1. Galaxy's rotation curve

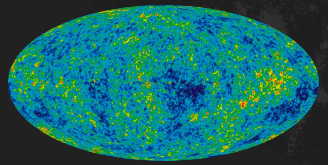


2. Motion of galaxies

These observations shows that there is more mass than measured.



3. Gravitational Lensing

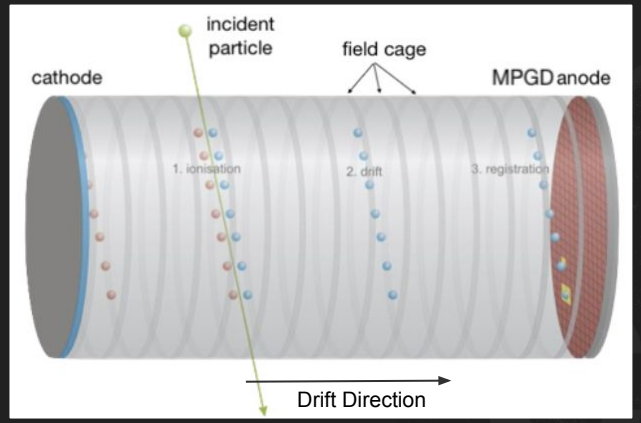


4. CMB measurements

WIMPs Detection Challenge

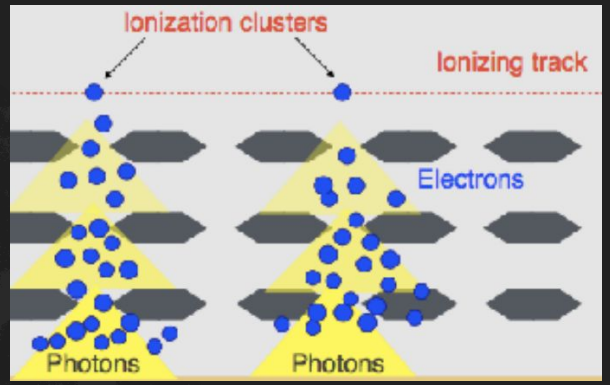
- Low Recoil Energy
  - Low energy threshold detector
- Low event rate
  - Large mass
  - Long term stability
- Background much higher than event rate
  - Background control (Cleanliness, shielding, ..)
  - Underground Operation (Reduces Muon induced neutron)
  - Electron/Nuclear recoil Discrimination

CYGNO/INITIUM Detector

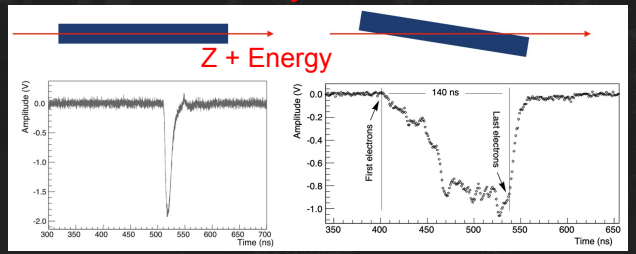


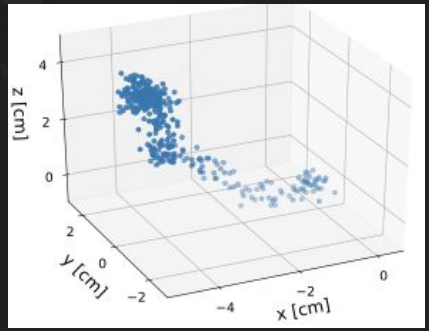
Amplification Region + Readout (sCMOS + PMT)

- Gaseous TPCs are inherently a 3D detector
- Tracking
- Head tail asymmetry
- dE/dX recognition
- Gas Flexibility



- CYGNO uses He:CF_4 gas mixture
- 3 GEM stack is used for charge amplification
- INITIUM is a part of CYGNO project which focuses on the development of TPCs with negative ion drift using SF_6 gas





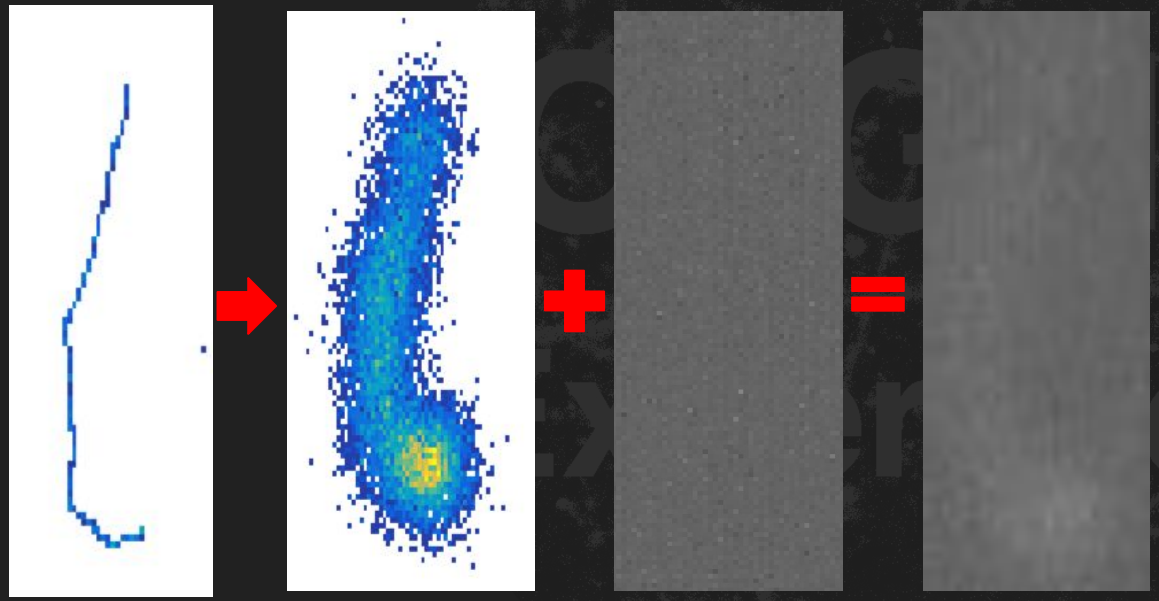
European Research Council
Established by the European Commission

INITIUM

https://iopscience.iop.org/article/10.1088/1742-6596/1468/1/012039
https://iopscience.iop.org/article/10.1088/1748-0221/15/07/C07036

Data analysis and simulation

Digitization



(a) Track w/o diffusion

(b) Track after diffusion

(c) Noise

(d) Digitised track

- Interaction of the particles with gas is simulated using either GEANT4 (for ER) or SRIM (for NR)
- These tracks are then projected to a 2D plane and detector effects are also added (diffusion)

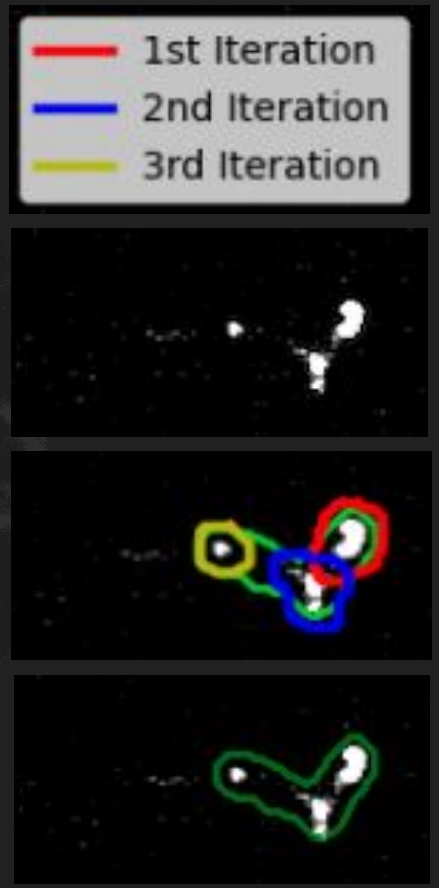
$$\sigma = \sqrt{\sigma_0^2 + B^2 Z}$$

- Actual camera noise is added to the diffused track to obtained the simulated image of the track

Reconstruction

- Images are reconstructed with a density based algorithm called IDBSCAN
- Outputs of IDBSCAN are fed to Superclustering algorithm which joins the different cluster obtained from IDBSCAN to form a supercluster around the complete track
- Two Superclustering algorithms that we used are GAC (Geodesic active contour) and Chan Vese
- Chan Vese works better than GAC in finding the supercluster for electron recoils.
- I worked on optimization of both these algorithms

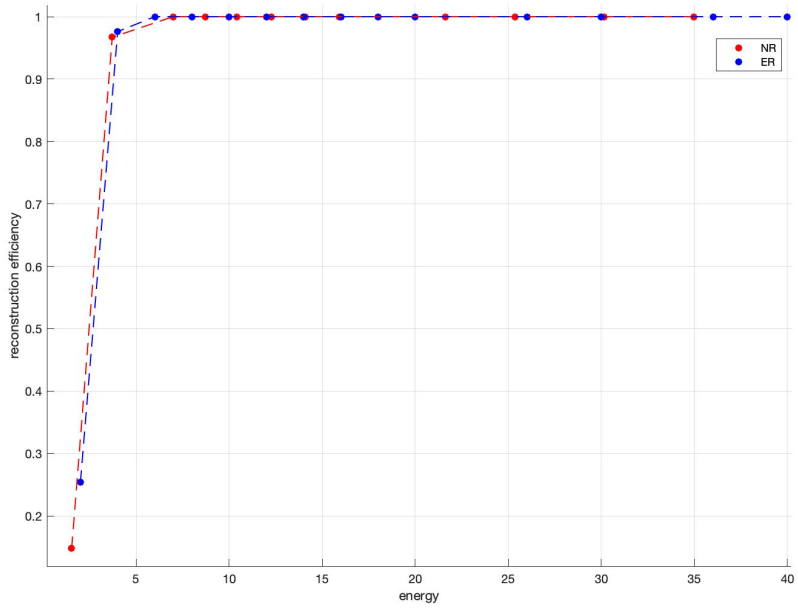
Iterations of IDBSCAN

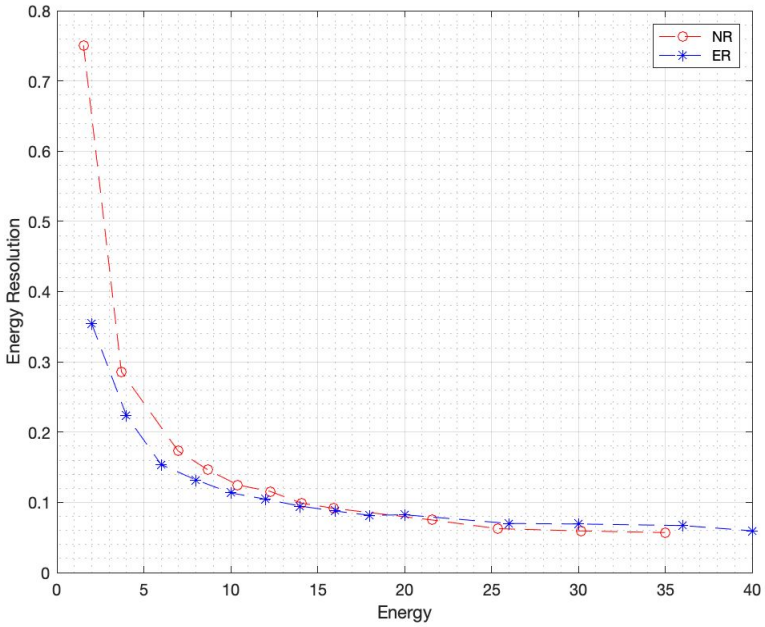


Reconstructed with GAC

https://inspirehep.net/literature/1805117

Reconstruction efficiency and Energy Resolution with CV





- Reconstruction efficiency is 100% at 6 keV and above for both ER and NR.
- Energy Resolution of the data at 6 keV is around 14% and for MC at 6 keV it is around 14%.

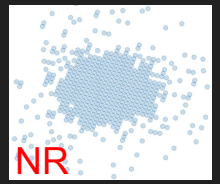
Discriminating Variable

Variables

Observables for recoil identification in gas TPCs
arXiv:2012.13649v1

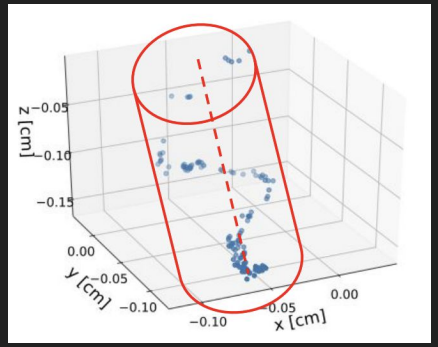
- Standard Deviation of Charge Distribution 2D(SDCD_2D):

$$SDCD = \sqrt{\frac{\sum_{i=1}^{N}(\mathbf{r_i} - \bar{\mathbf{r}})^2}{N}}.$$

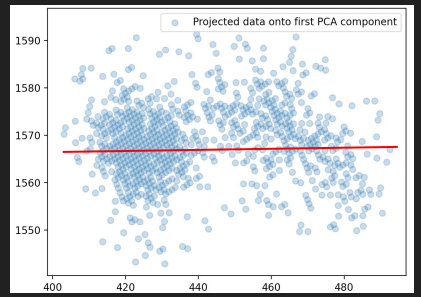


  - Electron recoils (ER) are longer, so the spread of charge is higher for ER when compared to Nuclear recoils (NR).
- Charge Uniformity 2D (ChargeUnif_2D):
  - For each point within the charge distribution, find the average distance to all other points.
  - ChargeUnif_2D is standard deviation of values computed in step 1.
  - Electron recoils tend to have charge distribution which is dense in some areas and sparse in other areas, while nuclear recoils are generally uniform.
- Maximum Density 2D (MaxDen_2D):
  - MaxDen is the value of most intense pixel from the image after rebinning it by a factor 2.
  - Electrons lose their energy at a slower rate than nuclei, this suggests that electron recoils are travel greater distance between interactions resulting in more sparse energy distribution.

Variables



- Cylindrical Thickness 2D (CylThick_2D):
  - For each charge, calculate the squared distance from the principal axis.
  - CylThick is the sum of all squared distances.
  - It is a measure of how much a recoil track deviates from the trajectory approximated by the principal axis.
  - Electrons experience far more scattering compared to nuclei, so principal axis approximates NR's trajectory much more accurately than it does for ER.
- Length Along Principal Axis 2D (LAPA_2D):
  - Project all the points in the charge distribution on to the principal axis.
  - LAPA is the difference between maximum and minimum projected value.
  - ER are longer compared to NRs, therefore projection is also longer.



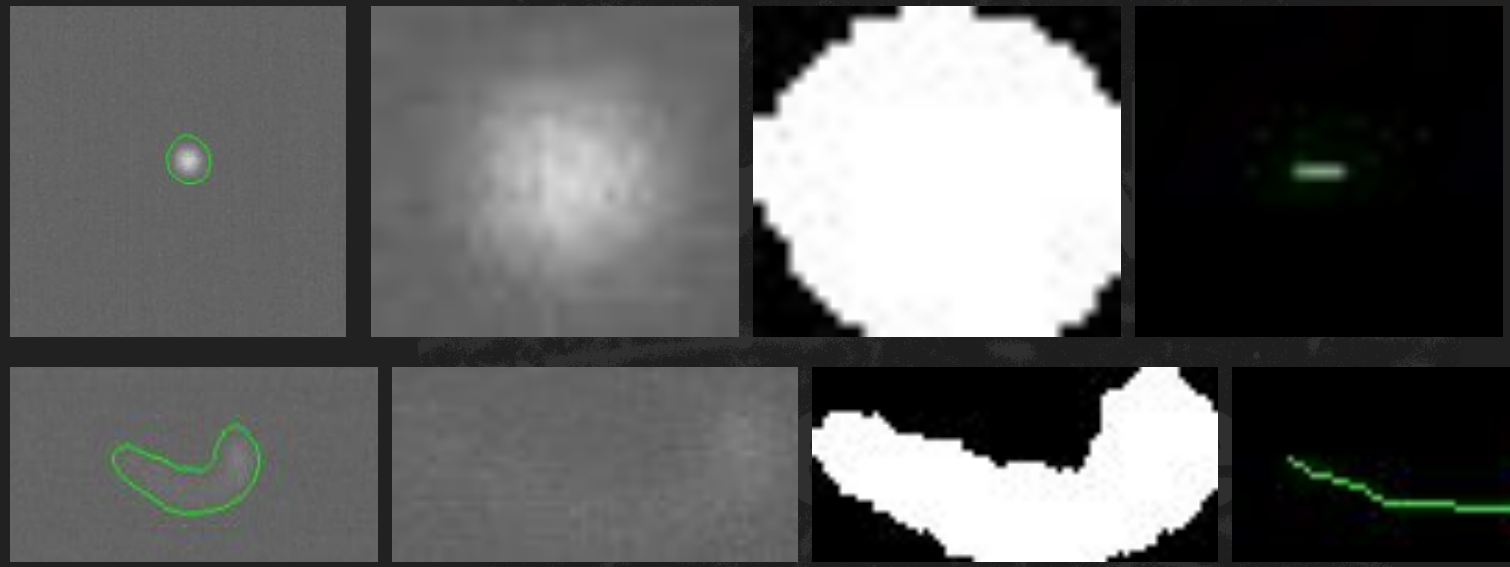
Variables

- eta:
  - MaxDen_2D divided by length (found by skeletonization)
- Light Density (delta):
  - Integral of the track divided by number of pixels in the track.
  - NR deposit higher energy over a short distance, therefore Light Density is higher for NR.
- Slimness:
  - Ratio of minor over major axis of the ellipse which bounds the track.
  - Electrons recoils suffer more scattering, so minor axis of the bounding ellipse is bigger when compared to NR which are generally straight.
- Skeleton length (thin_track):
  - Length in pixels found by thinning procedure.

GEM-based TPC with CCD Imaging for Directional Dark Matter Detection
arXiv:1510.02170v3

E Baracchini et. al., “Identification of low energy nuclear recoils in a gas TPC with optical readout”, arXiv:2007.12508v1

Length with Skeletonization



60 keV
He recoil

30 keV
electron
recoil

(a) Supercluster (b)Track (c)Thresholded track (d) Skeleton

Typically, directional detector use the diffused track length as length estimate, but that the REAL track length is more discriminating because is not affected by the diffusion.

Length comparison using different technique

	MC Length [mm]	Distance formula 2D	Skeleton (After Recon.)	Reconstructed length [mm]
10 keV NR	0.25	0.25	0.61	7.4
30 keV NR	0.6	0.6	0.76	8.8
300 keV NR	2.8	2.35	2.05	11.22
30 keV ER	6.68	8.76	7.07	13.48
60 keV ER	21.43	29.53	20.1	24.5
100 keV ER	53.14	71.19	28.88	22.8

Reconstructed length is the current official length used in our analysis.

MC Length: Length from MC simulation of track

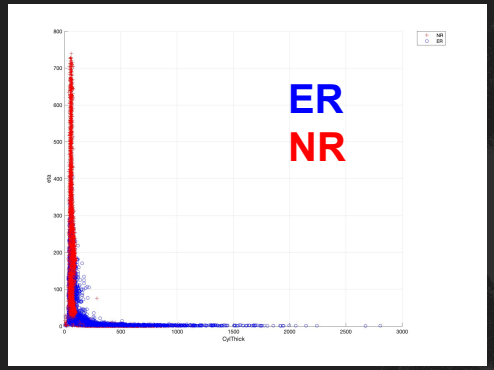
Skeleton: Length found by skeletonization

Distance formula : Length using 2 point distance formula

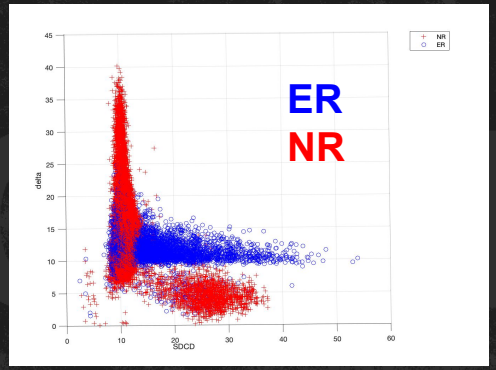
Reconstructed length: Major axis of the ellipse bounding the track

Some Variables in 2D and 3D

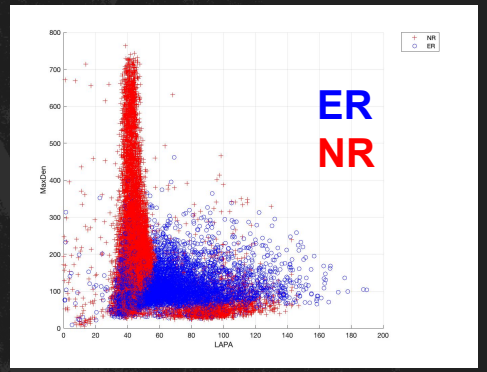
eta vs CylThick



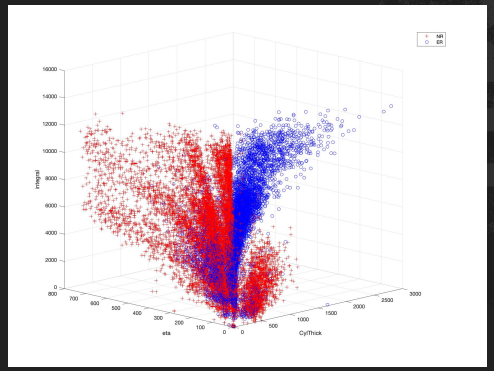
delta vs SDCD

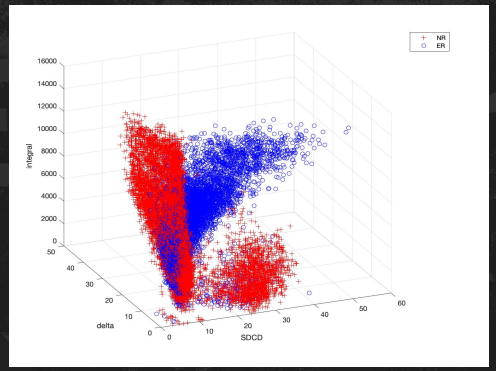


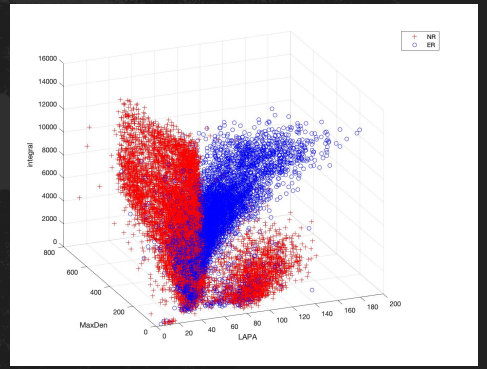
MaxDen vs LAPA



Variables in 2D





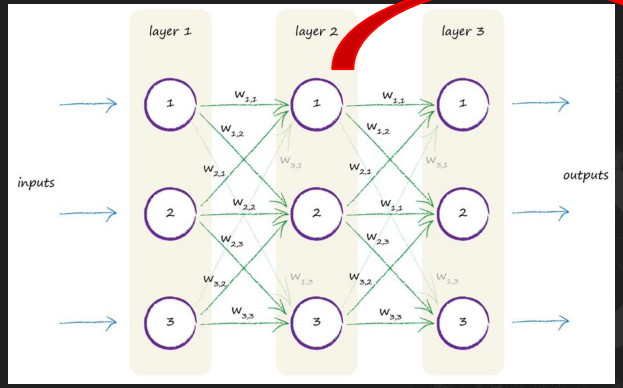


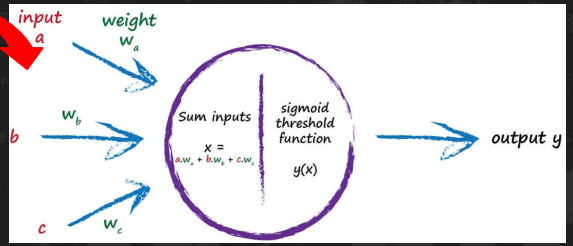
Variables in 3D

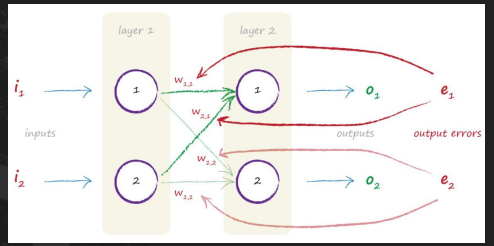
- In 3D, we can put a selection on energy and then other selection on the variables itself to increase the discrimination.

Discrimination of ER and NR using Neural Networks

Artificial Neural Networks

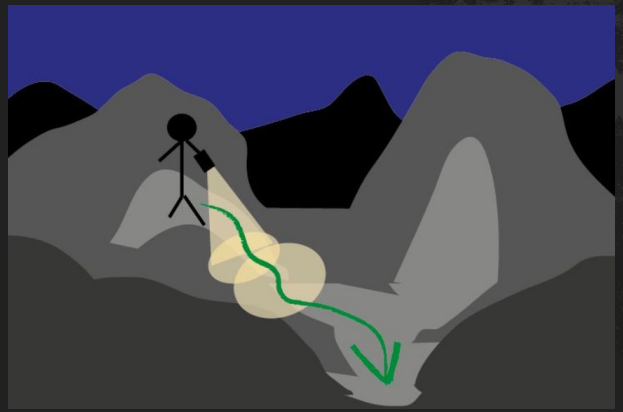


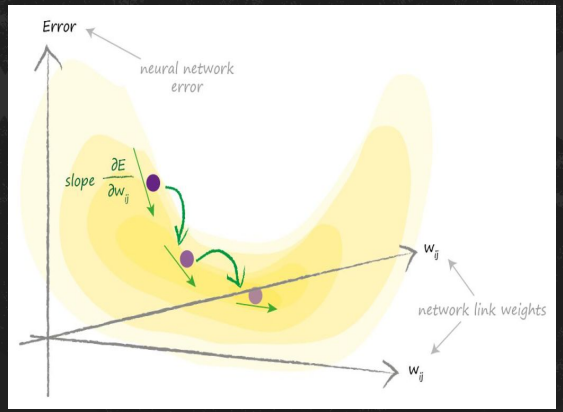




X = Σ(Weights * inputs)
Output = f(X)
Where, f is the activation function

$Error_O$ = Actual - Predicted
$Error_h$ = $Weights^T$ * $Error_O$
Where, O refers to output layer and h refers to hidden layers



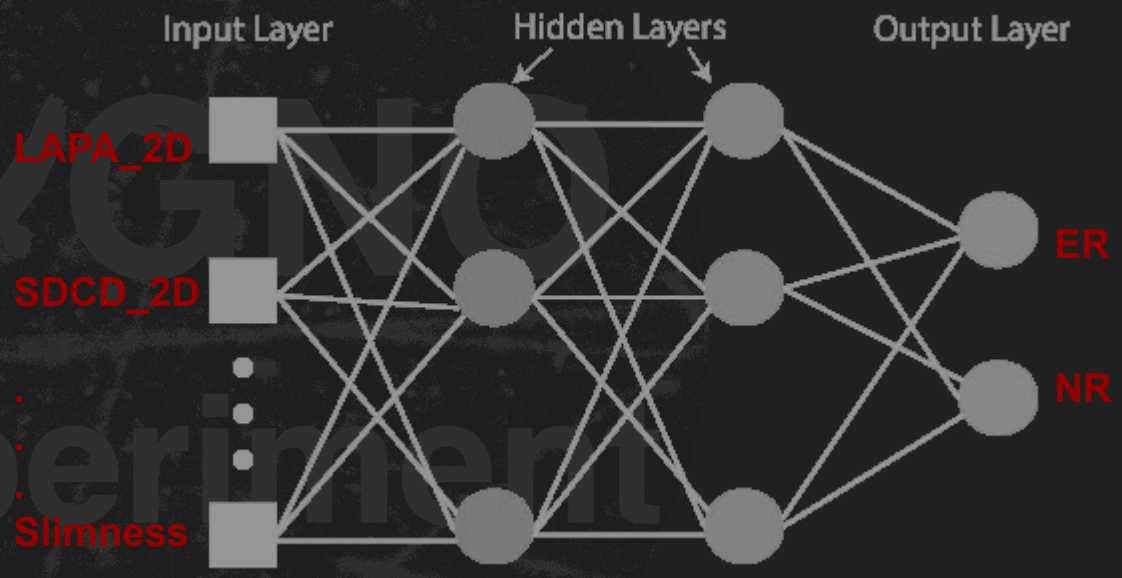


new Weights = old Weights - α . $\frac{\partial(Error)}{\partial(Weights)}$

Where, α is the step size (learning rate).

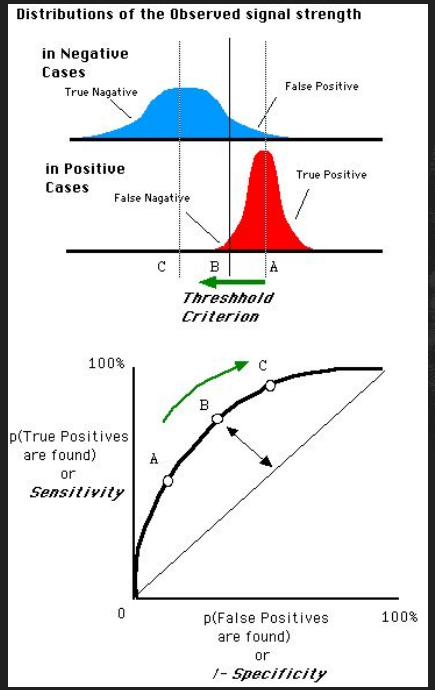
Feedforward Patternnet for our problem

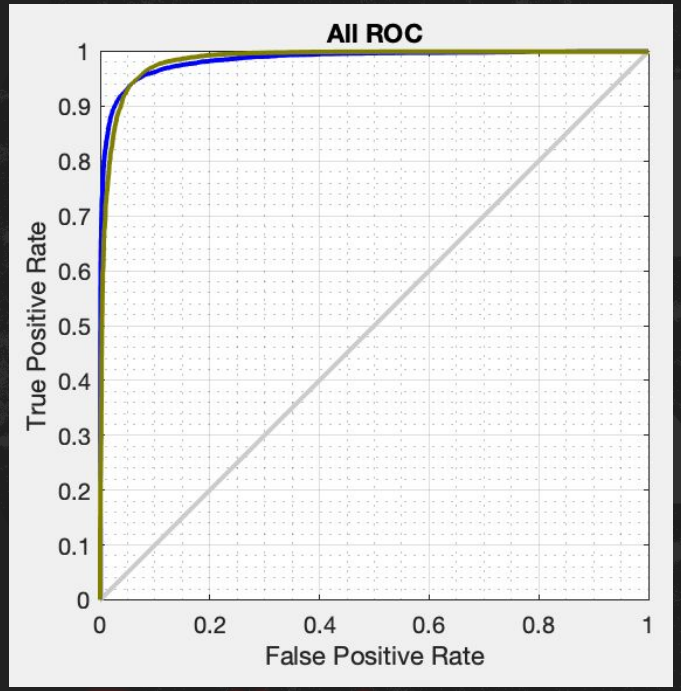
- Pattern net with 3 hidden layers of size [10,10,10] neurons were used.
- Data division [90:5:5]
- Training algorithm
  - Levenberg Marquardt
- Loss: MSE
- Inputs: LAPA_2D, skel_track, SDCD_2D, CylThick_2D, MaxDen_2D, eta, sc_size, sc_nhits, sc_integral, sc_length, sc_width,delta, slimness
- Output: Nuclear recoil and Electron Recoil class

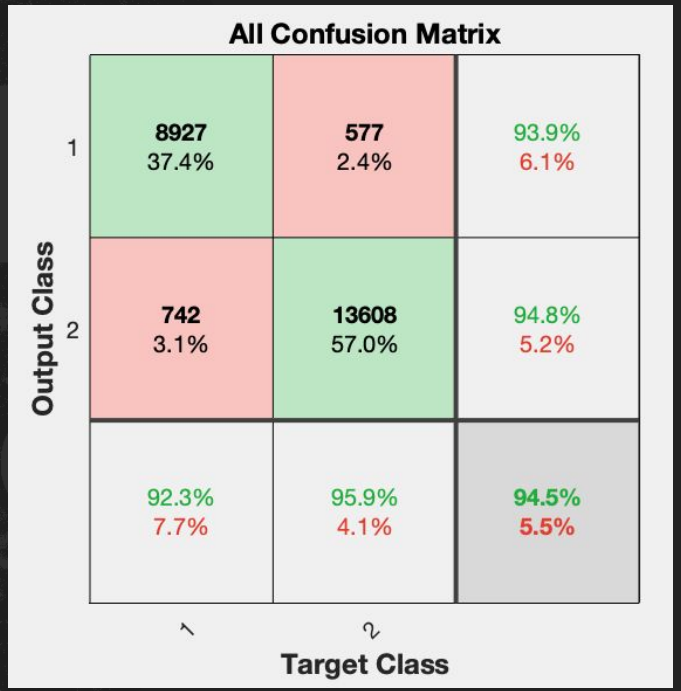


- Energy Range : 1-40 keV for both ER and NR

Result of DNN (ROC and CM)







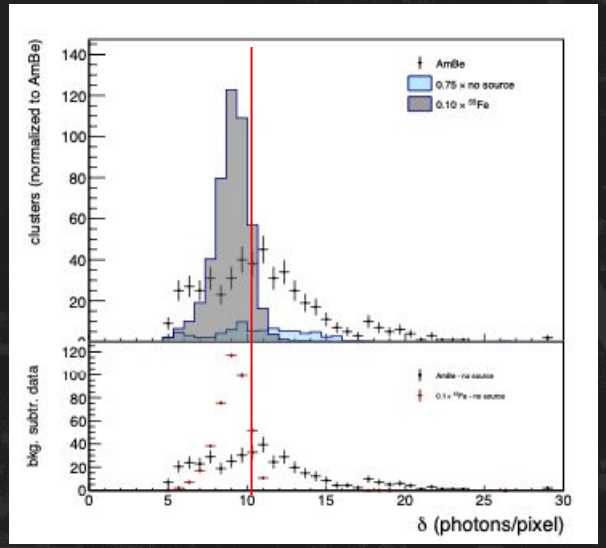
Preliminary

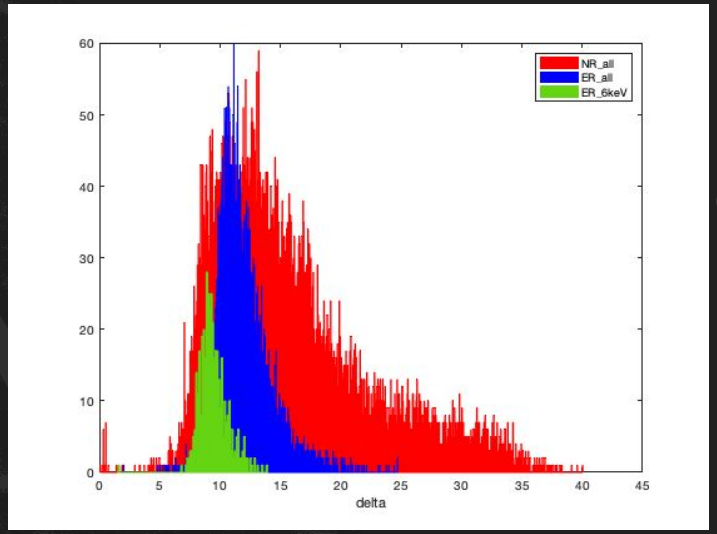
1 - ER class
2 - NR class

Predicted = Output
True = Target

Result of DNN

- Figure (a) is distribution of delta variable for NR (Red: 1-40 keV), ER (Blue: 1-40 keV), ER (Green: 6 keV).
- Table (b) is the signal and background efficiency with different thresholds (data without any preselection).





(a)

working point	Signal efficiency[1-40] keV			Background efficiency [6] keV		
	ε_S^{presel}	ε_S^{δ}	ε_S^{total}	ε_B^{presel}	ε_B^{δ}	ε_B^{total}
WP_{50}	0.98	0.51	0.50	0.70	0.050	0.035
WP_{40}	0.98	0.41	0.40	0.70	0.012	0.008

Threshold	ε_s^{total} [1-40] keV	ε_B^{total} [1-40] keV
0.995	0.50	0.005
0.998	0.40	0.001

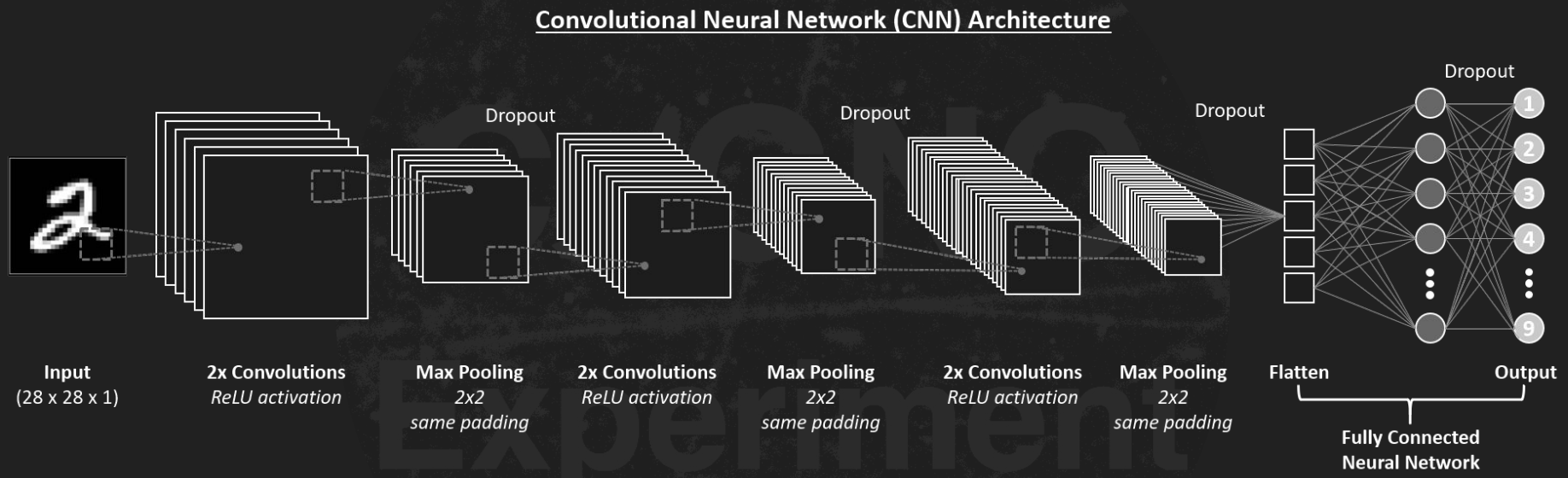
(b)

E Baracchini et. al., “Identification of low energy nuclear recoils in a gas TPC with optical readout”, arXiv:2007.12508v1

Preliminary

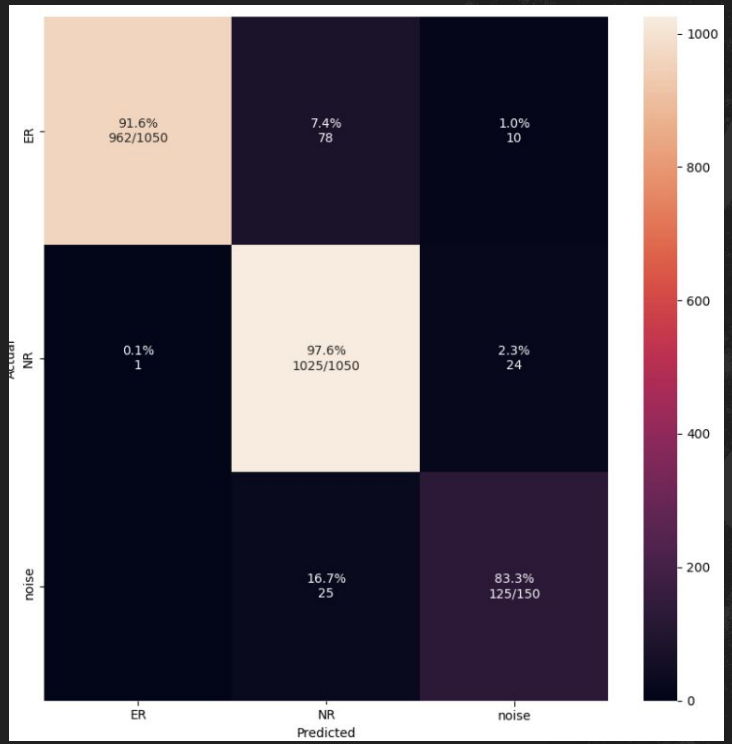
Preliminary

Convolutional Neural Networks



- Layers: 3 Convolutional layers
- Kernel size: 3
- Activation : ReLU
- Optimizer : Adam
- Loss function: Cross Entropy
- Input : Images (2304x2304x1)
- Energy range: 1-100 keV
- Output : NR and ER class

Result of CNN with input as simulated images



- Results of the classification is very similar to the results of the DNN with discriminating variables.
- ER classification -> 92.3% with DNN
  ->91.6% with CNN
- NR classification -> 95.9% with DNN
  -> 97.6% with CNN

Preliminary

- This work was done with Gustavo Viera Lopez from Computer Science.

Future Plans

Future Plans

- Include a larger sample for training (~ 5000 points at each energy sample)
- Adding these discriminating variables to a CNN
- Building a CNN for reconstruction of the track , and computing the variables like energy, length and so on.
- Combining the signal from PMT with CNN to obtain 3D track.
- Results of these studies will be further used to evaluate the actual experimental sensitivity including the expected background.

List of publications:

[1] F. D. Amaro, E. Baracchini et. al., *Directional Dark Matter Searches with CYGNO*, MDPI, July 2021, DOI: 10.3390/particles4030029

List of Schools/Conferences:

[1] School on Underground Physics, SOUP 2021, 28 June - 2 July (Online)

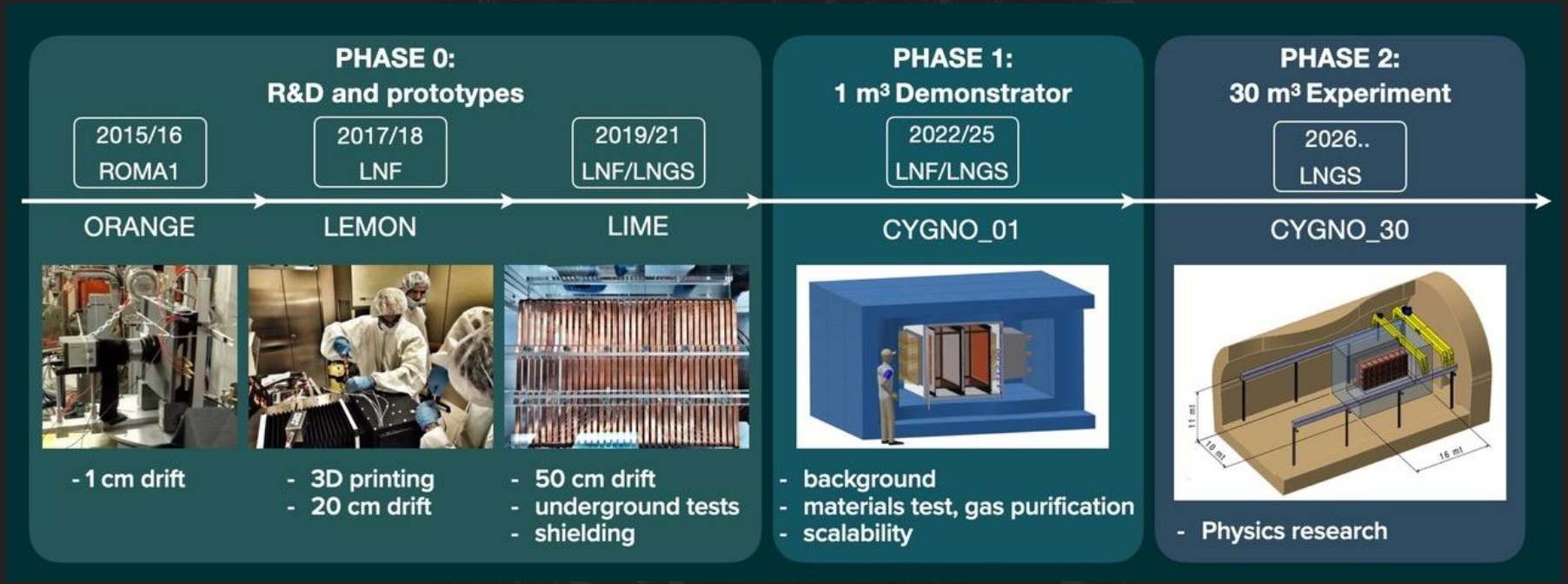
[2] 4th International School on Deep Learning, DeepLearn 2021 Summer, Las Palmas de Gran Canaria, Spain, 26 July - 30 July (In person)

[3] Quantum Sensors for Fundamental Physics, QSFP 2021, 6 September -17 September (Online)

[4] Advanced Computing and Analysis Techniques in Physics Research (ACAT) 2021, 29 November - 3 December 2021, Submitted abstract “Reconstruction and Particle Identification with CYGNO Experiment”(on behalf of CYGNO collaboration)

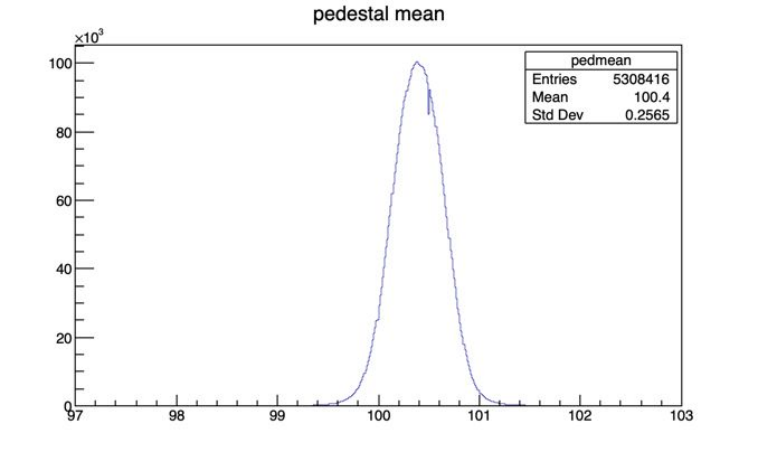
Backup Slides

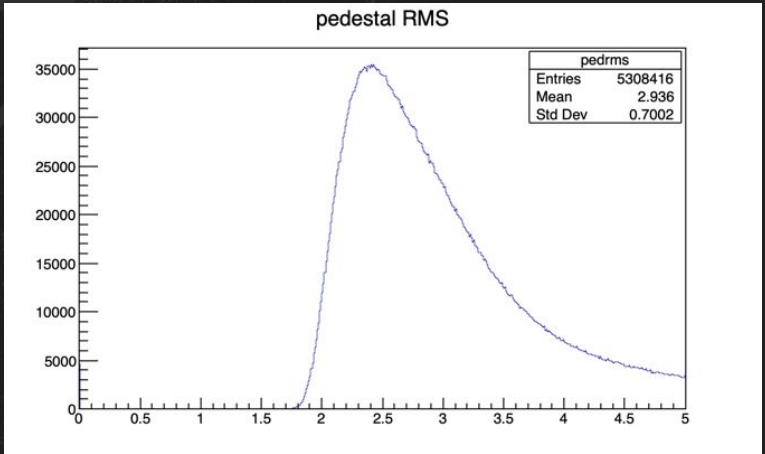
Timeline of the project



Noise Simulation

- ECDF (empirical cumulative distribution function) map for each pixel by reading the same pixels from all the images
- Then desired number of noise images can be generated with these ecdf maps
- These images can be used for digitization of the simulated tracks and mean and rms of the noise distribution is used for noise suppression during the process of reconstruction of the images.

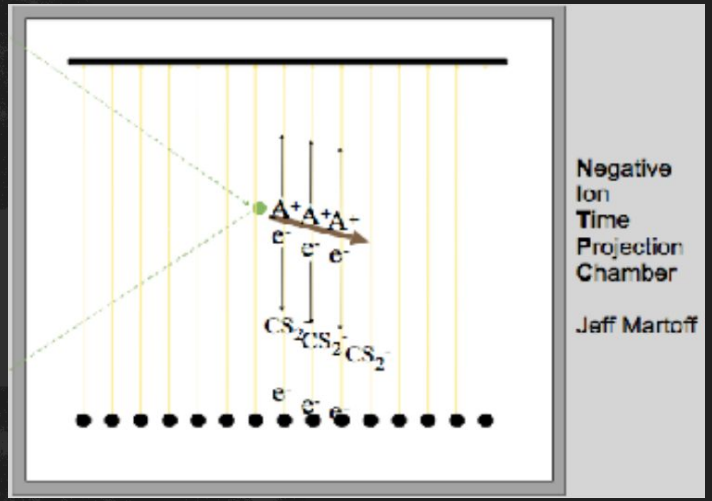




INITIUM

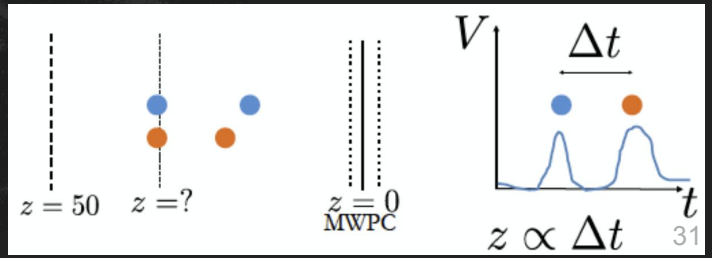
Reduced Diffusion = Improved Tracking Capability

- Electronegative dopant is introduced in the gas mixture (like SF_6)
- Primary ionization electrons captured by electronegative gas molecules at O(100) μm
- Anions drift to the anode acting as the effective charge carrier instead of the electrons and reducing both longitudinal and transverse diffusion to thermal limit
- Presence of multiple charge carriers with different mass give rise to difference in time signal, because anions with different mass drift with different velocities



Start Together

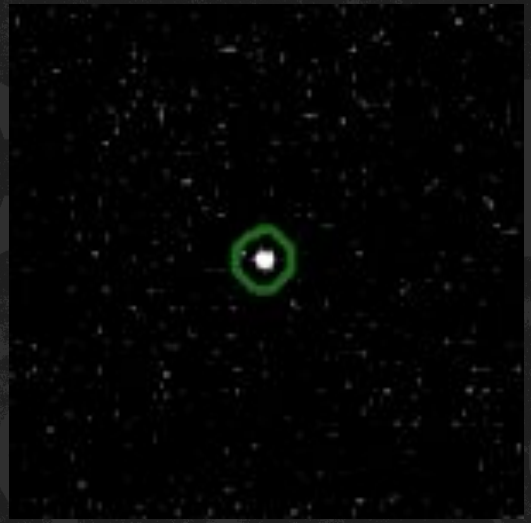
But drift at different velocity



https://doi.org/10.1088/1748-0221/13/04/P04022

CV vs GAC

(a) Chan vese



(b) GAC

- 6 keV He NR track reconstructed with Chan vese (a) and GAC (b) algorithm.
- Chan vese includes a lot of noise around the actual track, thus biasing the energy estimate.

- GAC algorithm looks for the local features (like change in gradient locally).
- For ER, energy deposition is not continuous, so GAC fails to reconstruct the complete track.