Applications of Machine Learning methods to Dark Matter search

The CYGNO use-case

G. Mazzitelli on behalf of CYGNO/INITUM collaboration

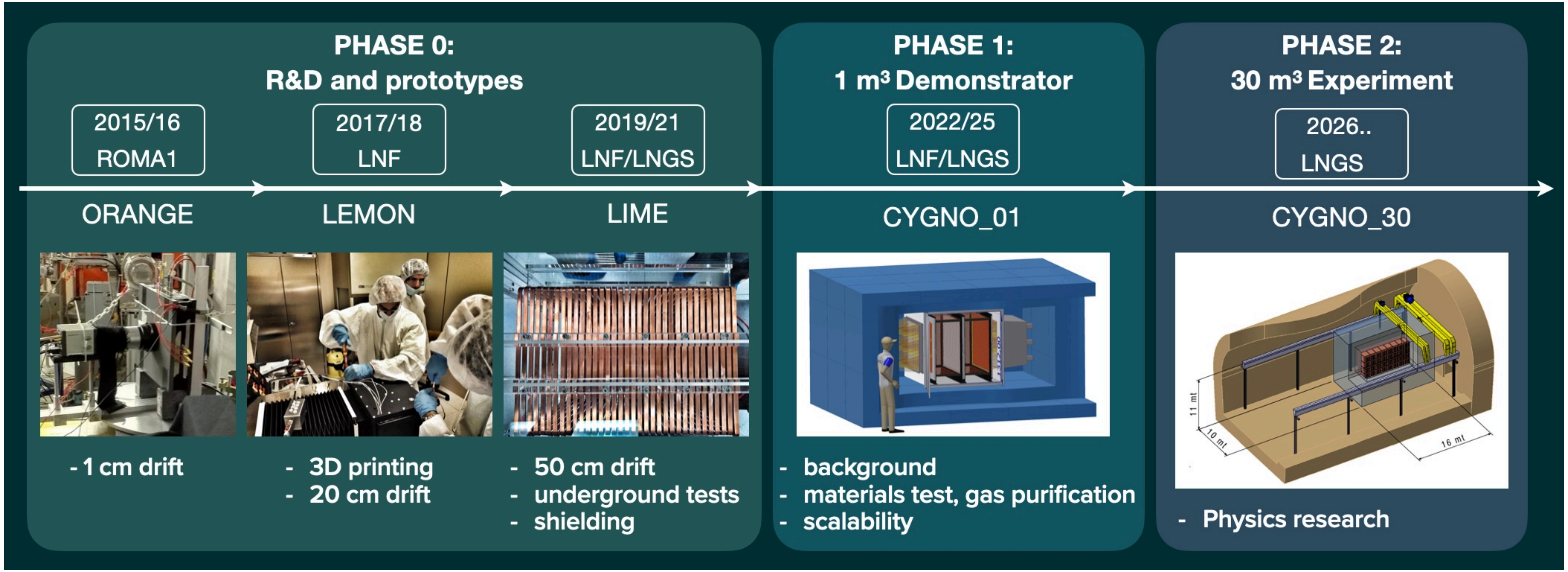
AI@INFN, 2–3 May 2022 - Hotel Europa, Bologna

AI@INFN - Artificial Intelligence at INFN

CYGNO a large TPC for dark matter and neutrino study

objective

exploiting the progress in **commercial scientific Active Pixel Sensors** (APS) based on CMOS technology to realise a large **gaseous Time Projection Chamber** (TPC) for **Dark Matter and Solar Neutrino search**.

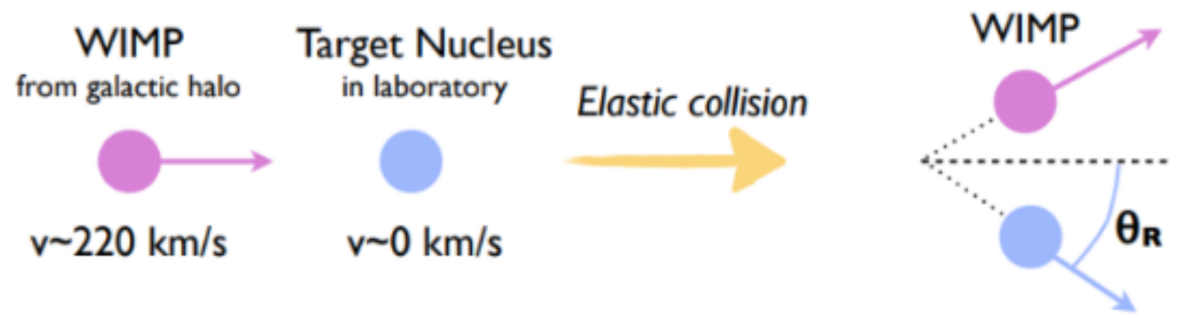


low-energy (1–100 keV) nuclear recoils
10^5 beta/gamme rejection

nuclear recoil threshold

why light mass gaseous based TPC?

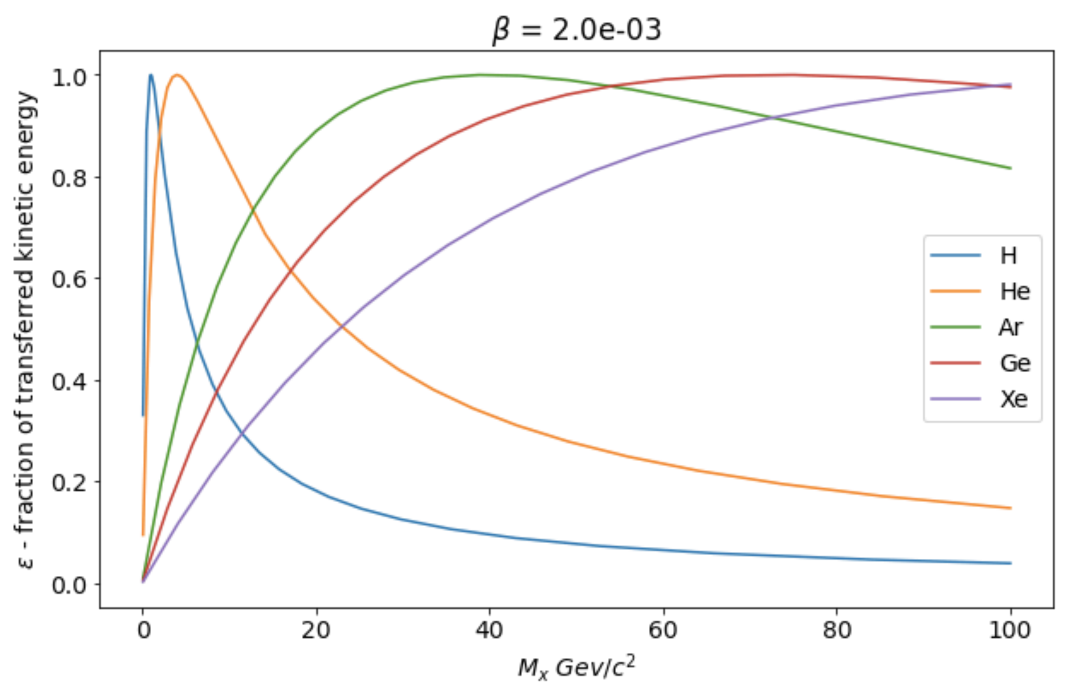
gassous
low (O 100 eV) threshold
just some ideas to increase sensitivity and scalability

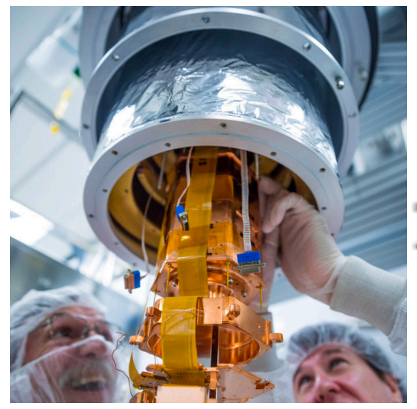


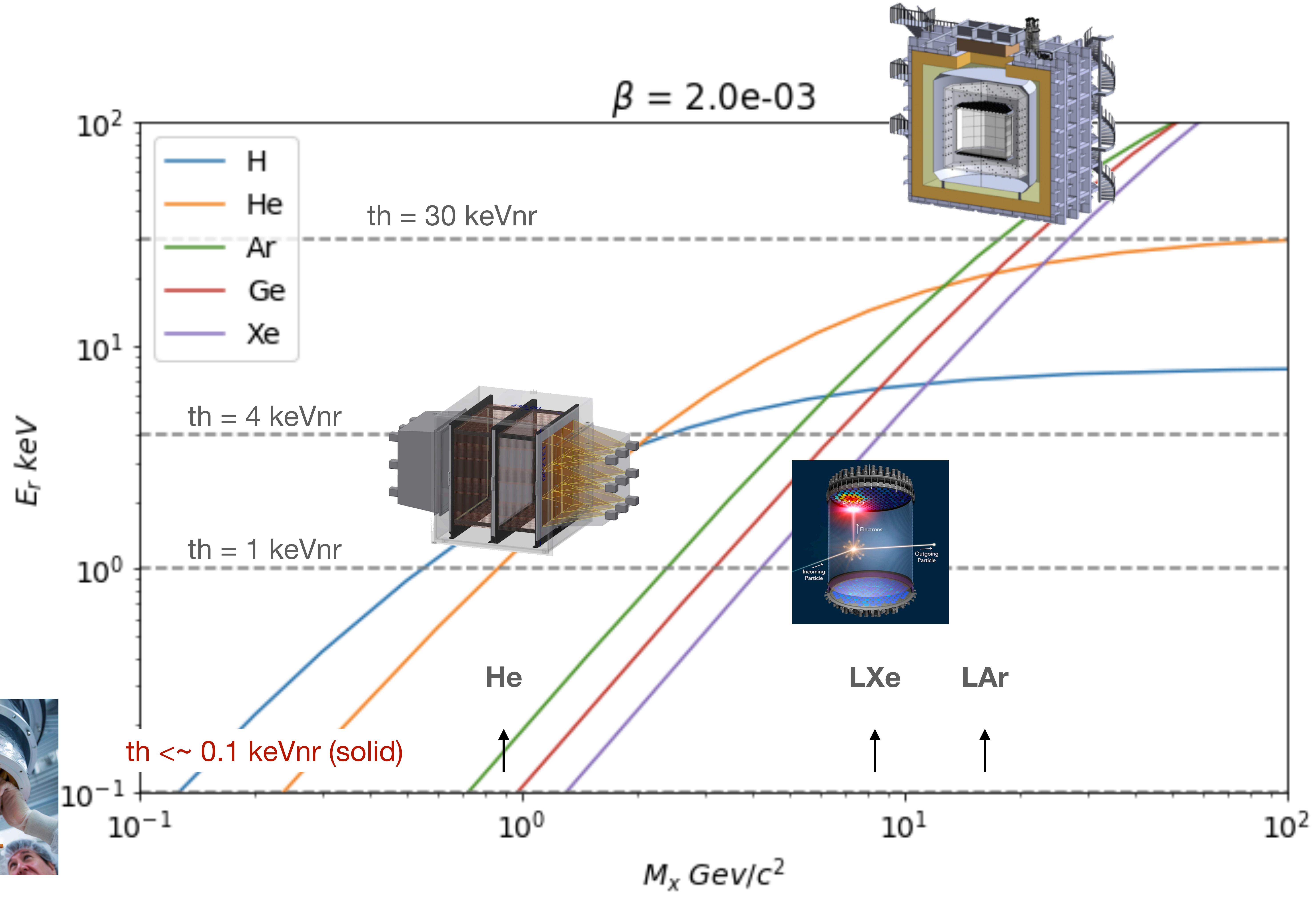
Nuclear recoils (partially) retain the incoming WIMP direction

$$\epsilon = \frac{4 \times \rho}{(\rho + 1)^2} \qquad \rho = M_w/M_T$$

$$E_{nr} = \epsilon \times \frac{1}{2} M_w v^2$$

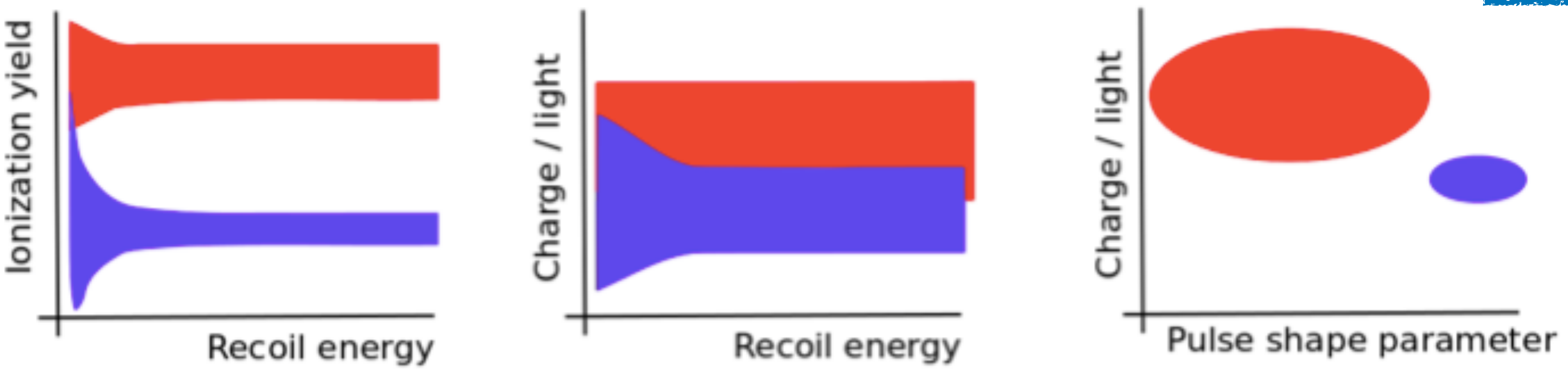






discrimination

liquid, cryogenic
medium (O 1000 eV) threshold
high sensibility and scalability

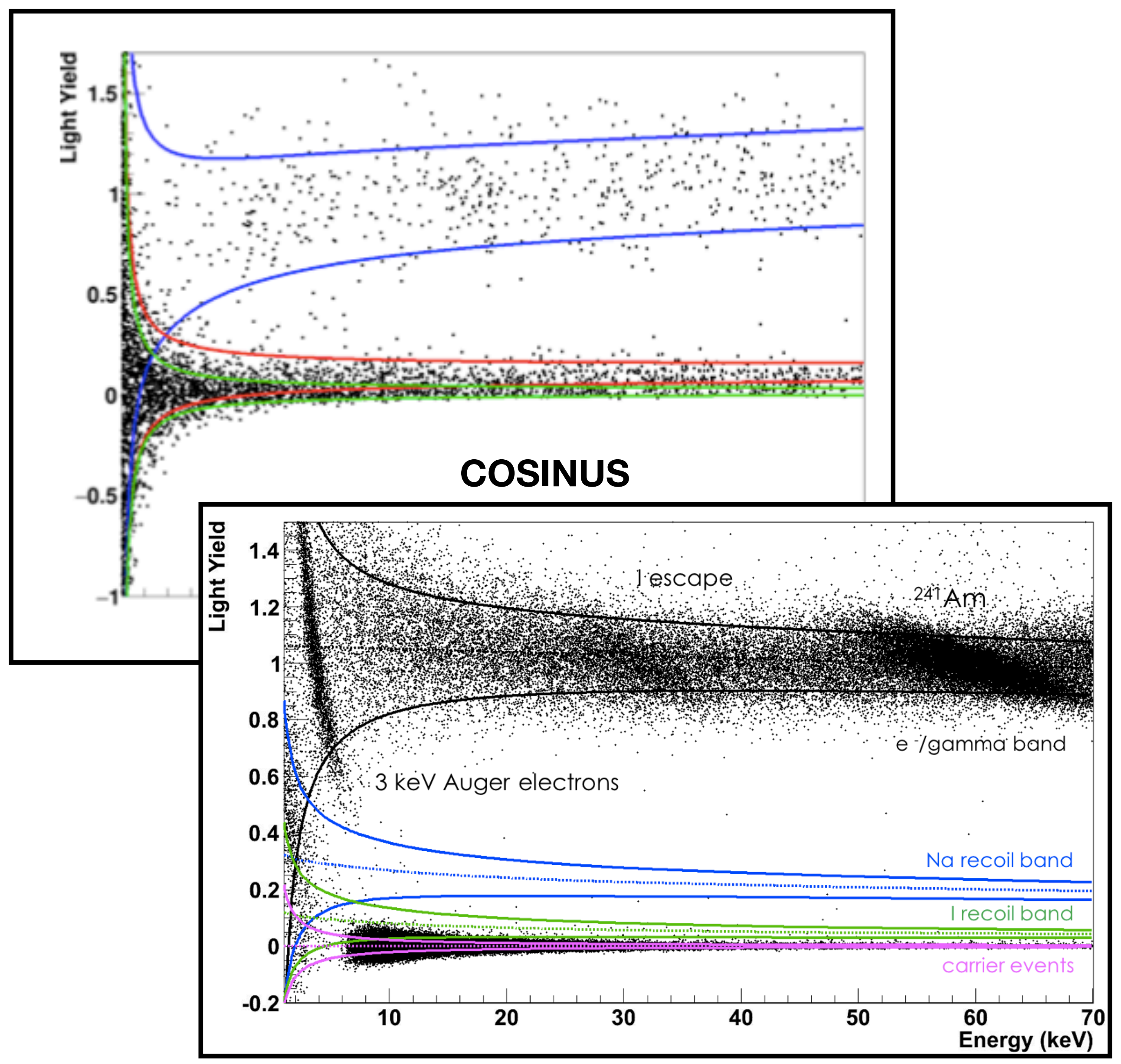


solid/cryogenics

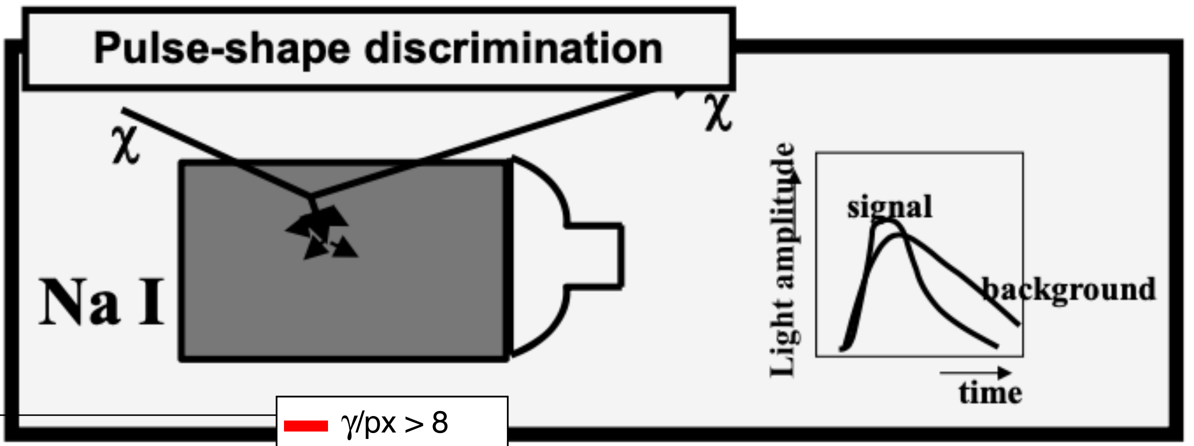
gass/liquid (Xe/Ar)

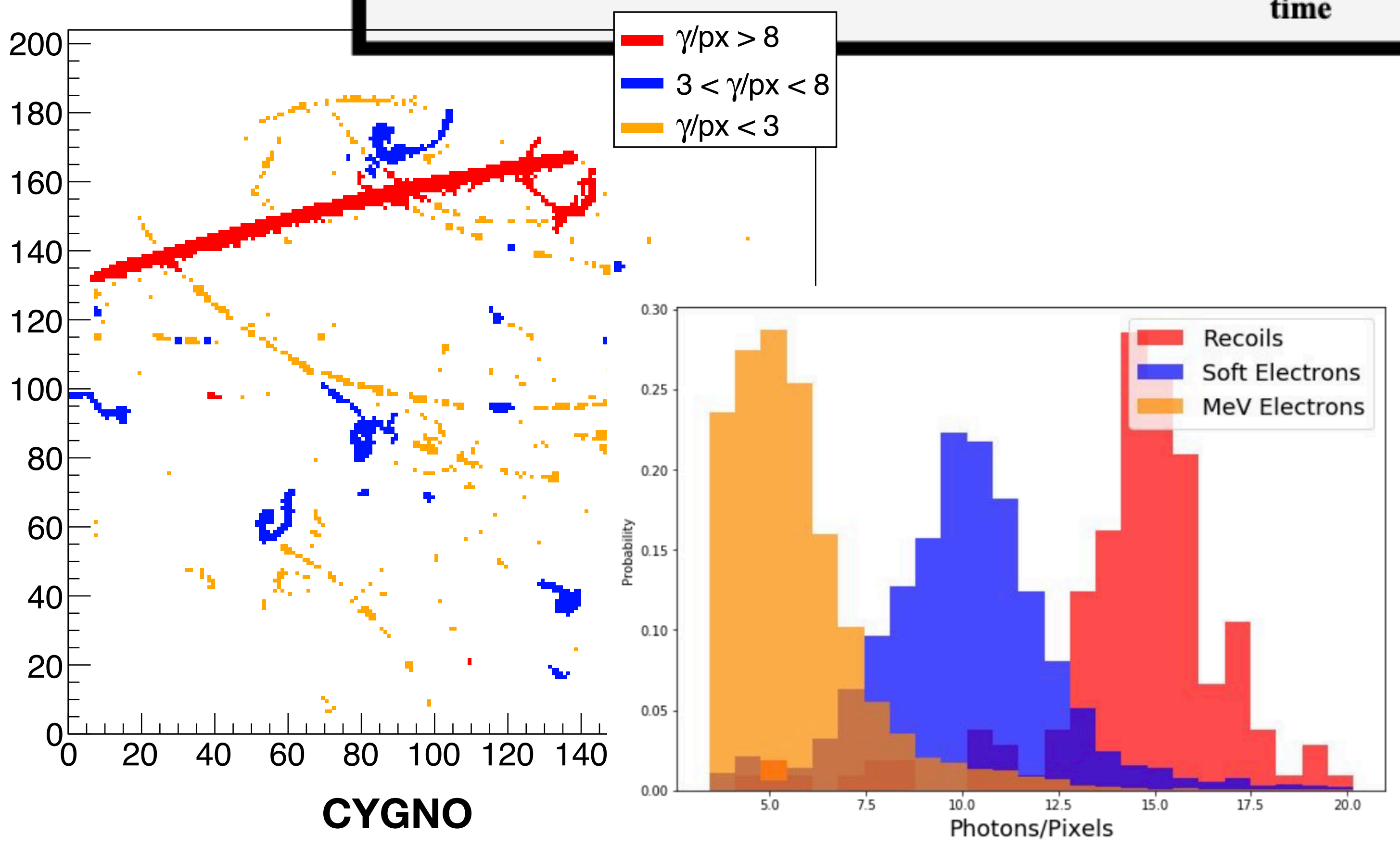
scintillator/liquid (Ar)

CREEST

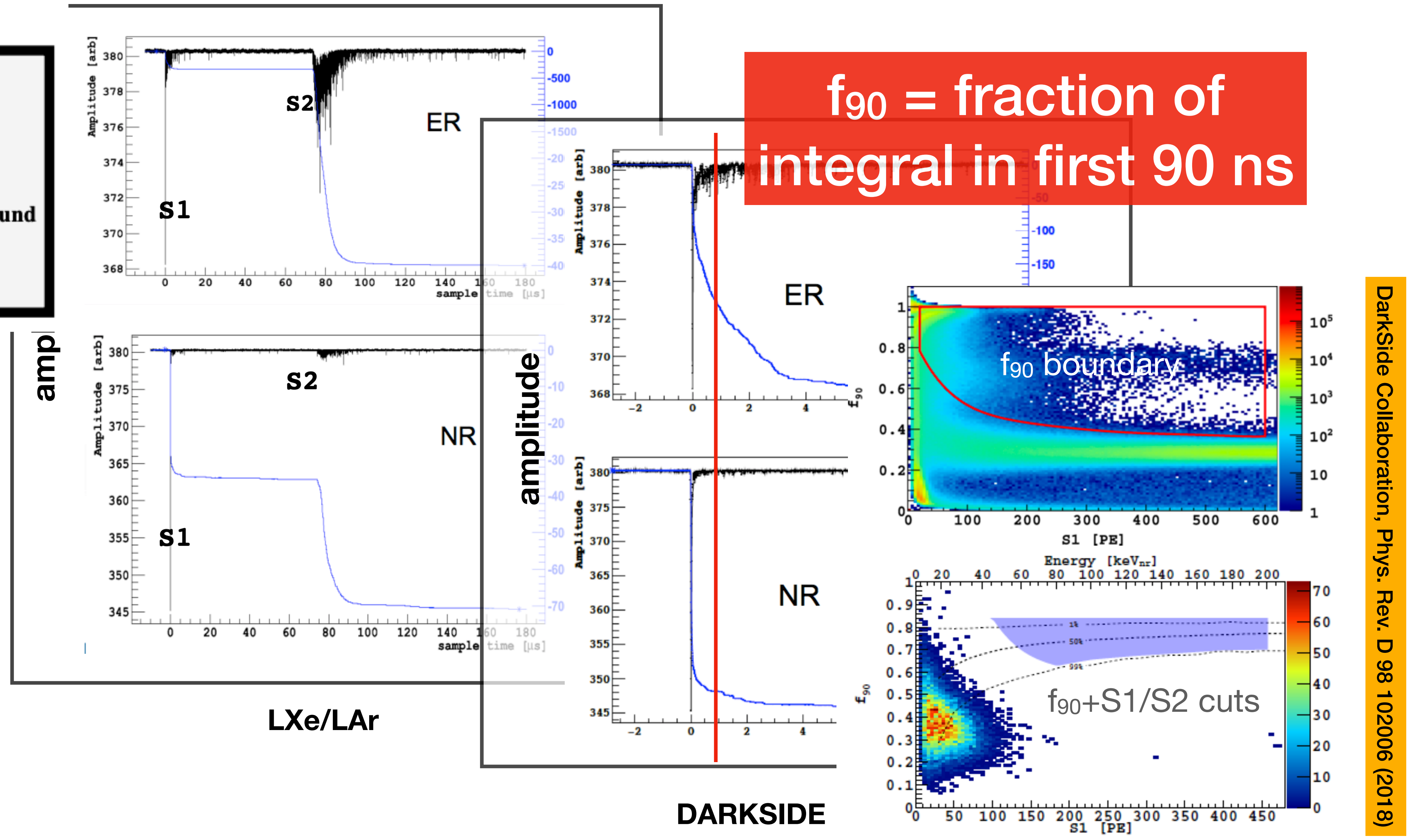


NaI detectors





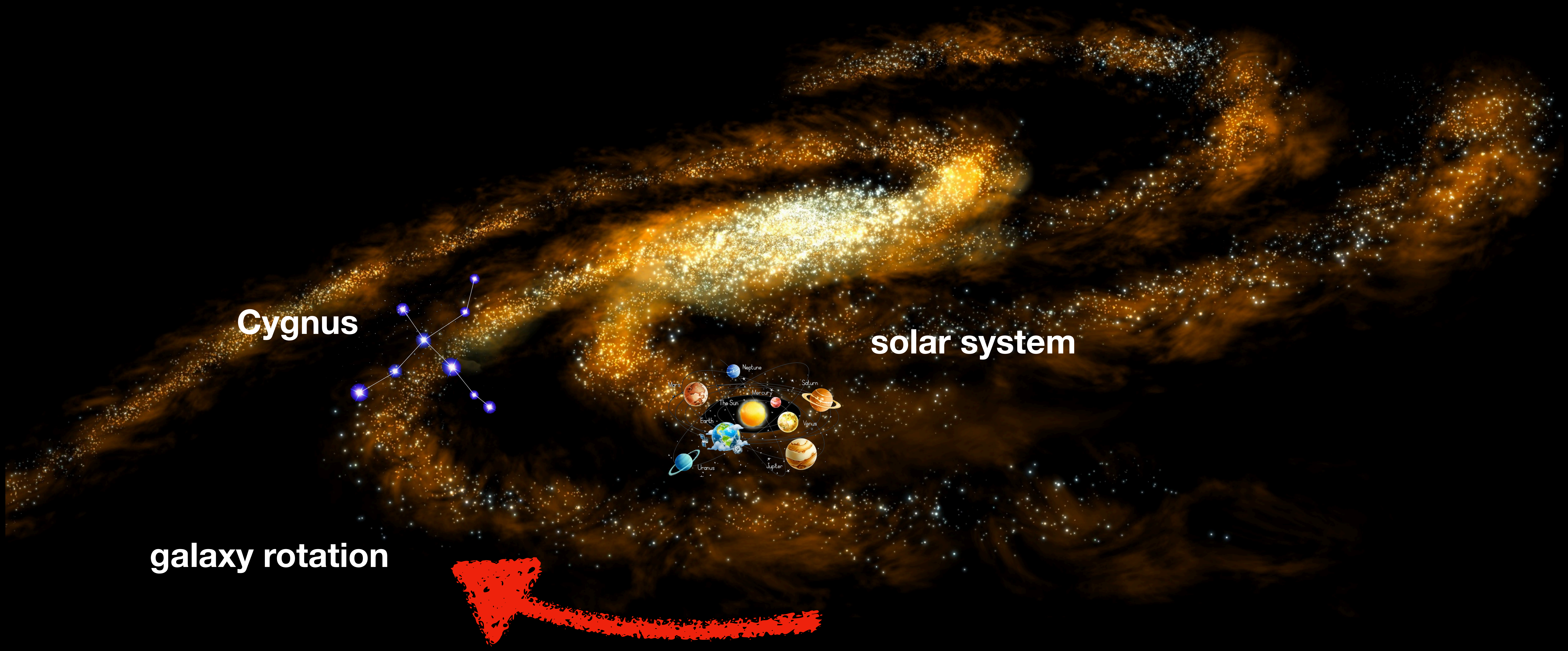
CYGNO



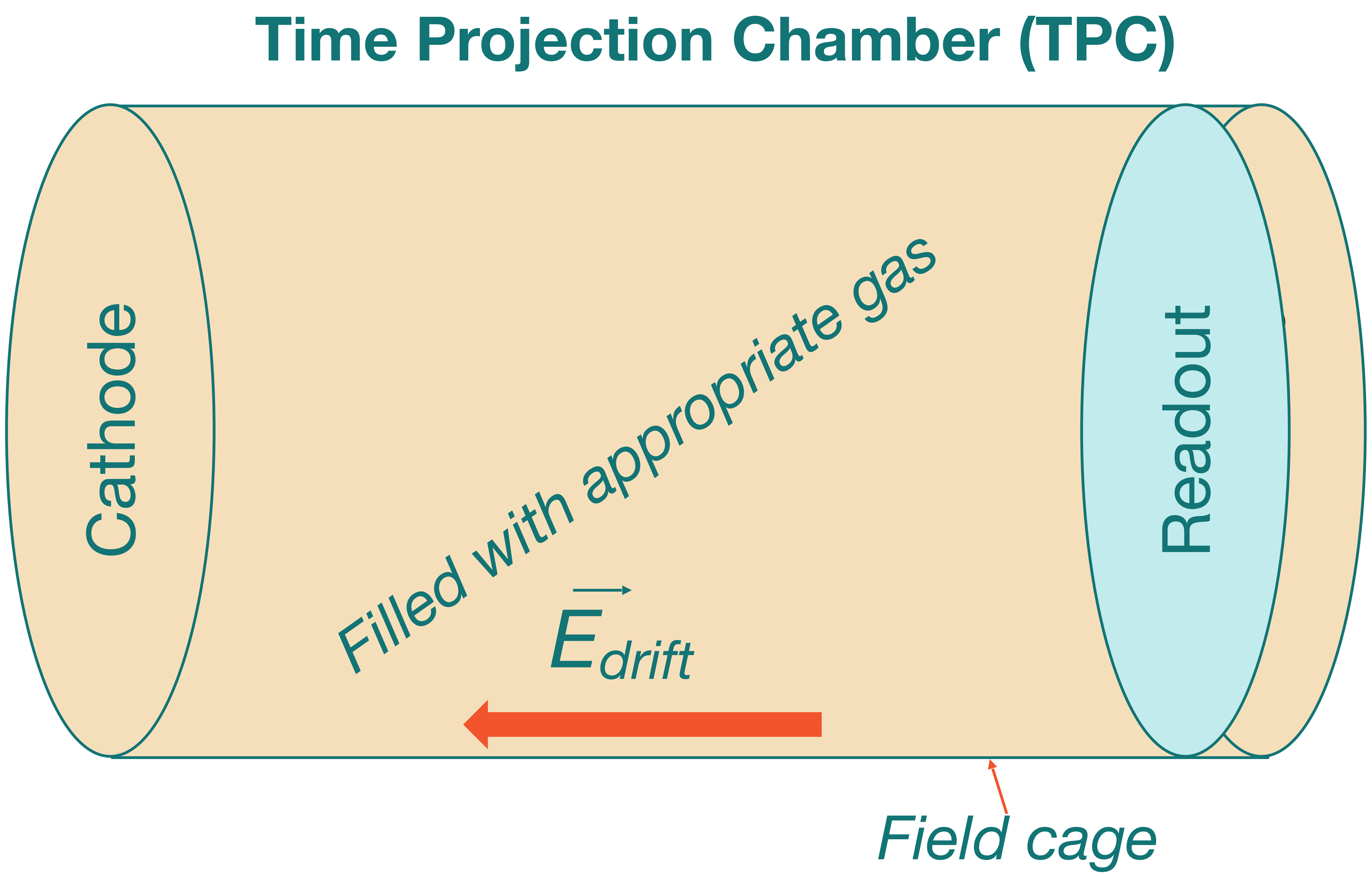
LXe/LAr

DARKSIDE

the dark matter when living on the earth

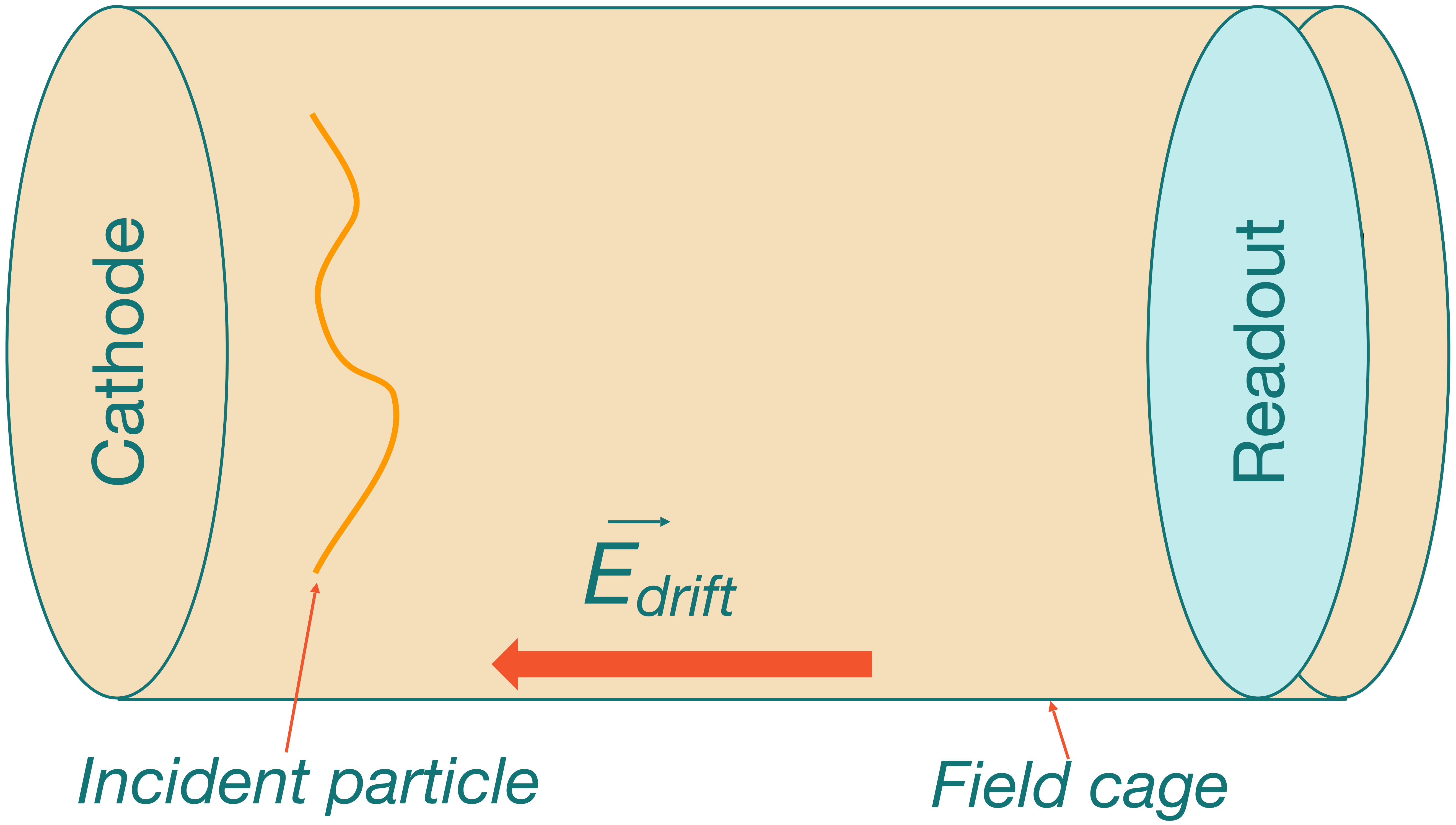


TPC Detector in a nutshell



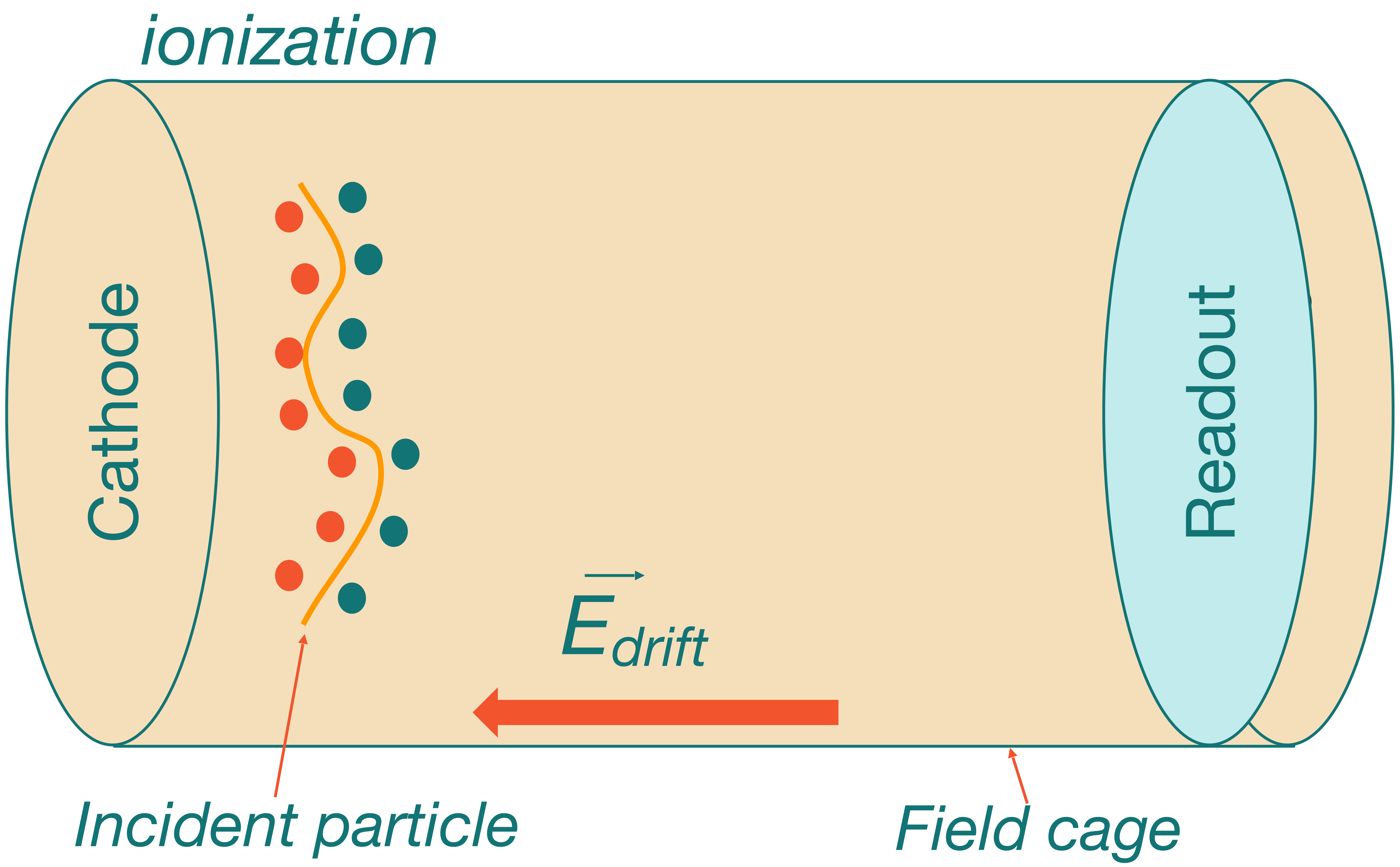
A TCP is costituite by a vessel filled with gas or liquid (Ar, Xe, etc) where an appropriate field is applied (typically kV/cm)

TPC Detector in a nutshell



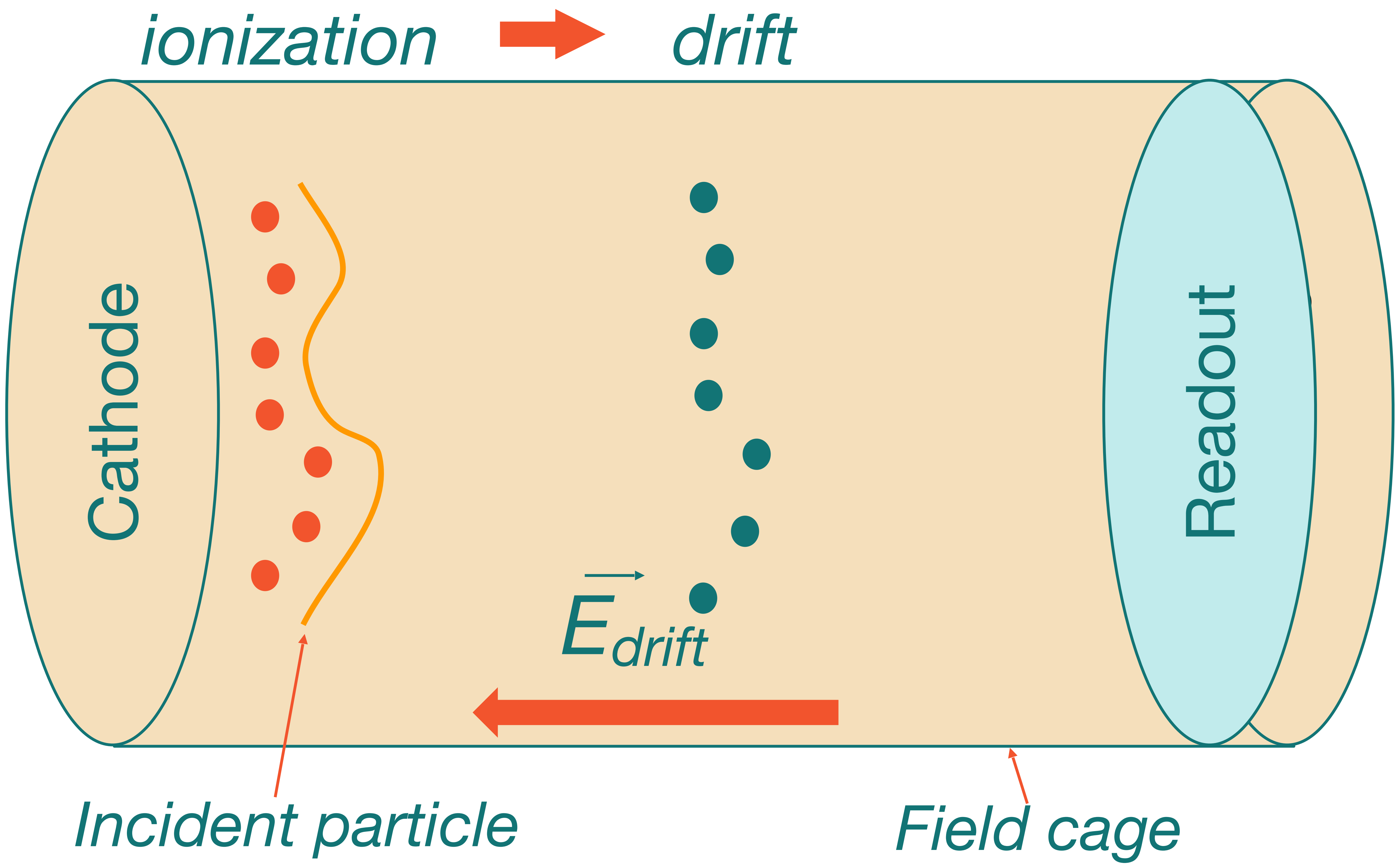
when a charged particle pass true the gas, have a well known probability to ionise the gas and …

TPC Detector in a nutshell



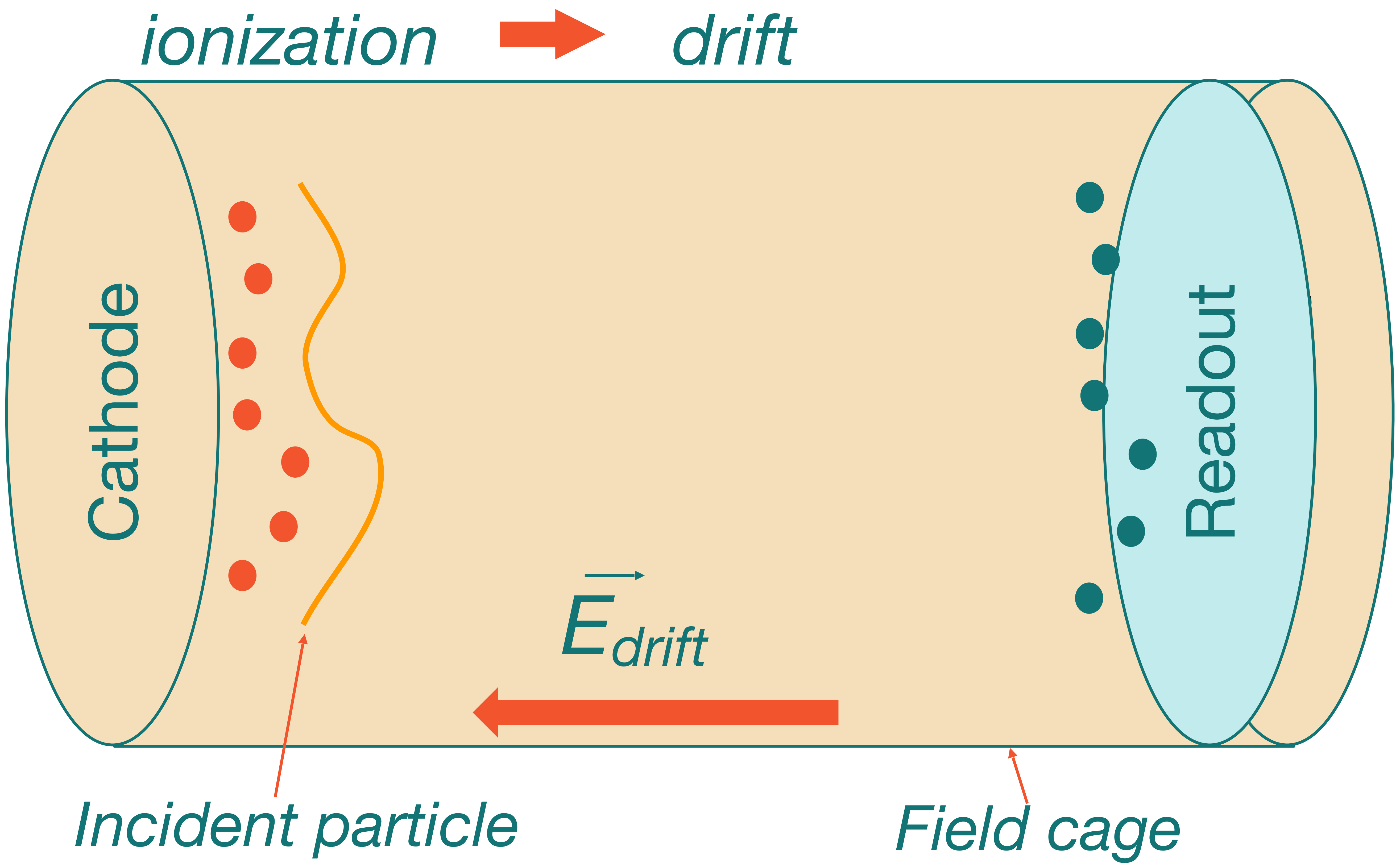
… produce free ions and electrons that …

TPC Detector in a nutshell



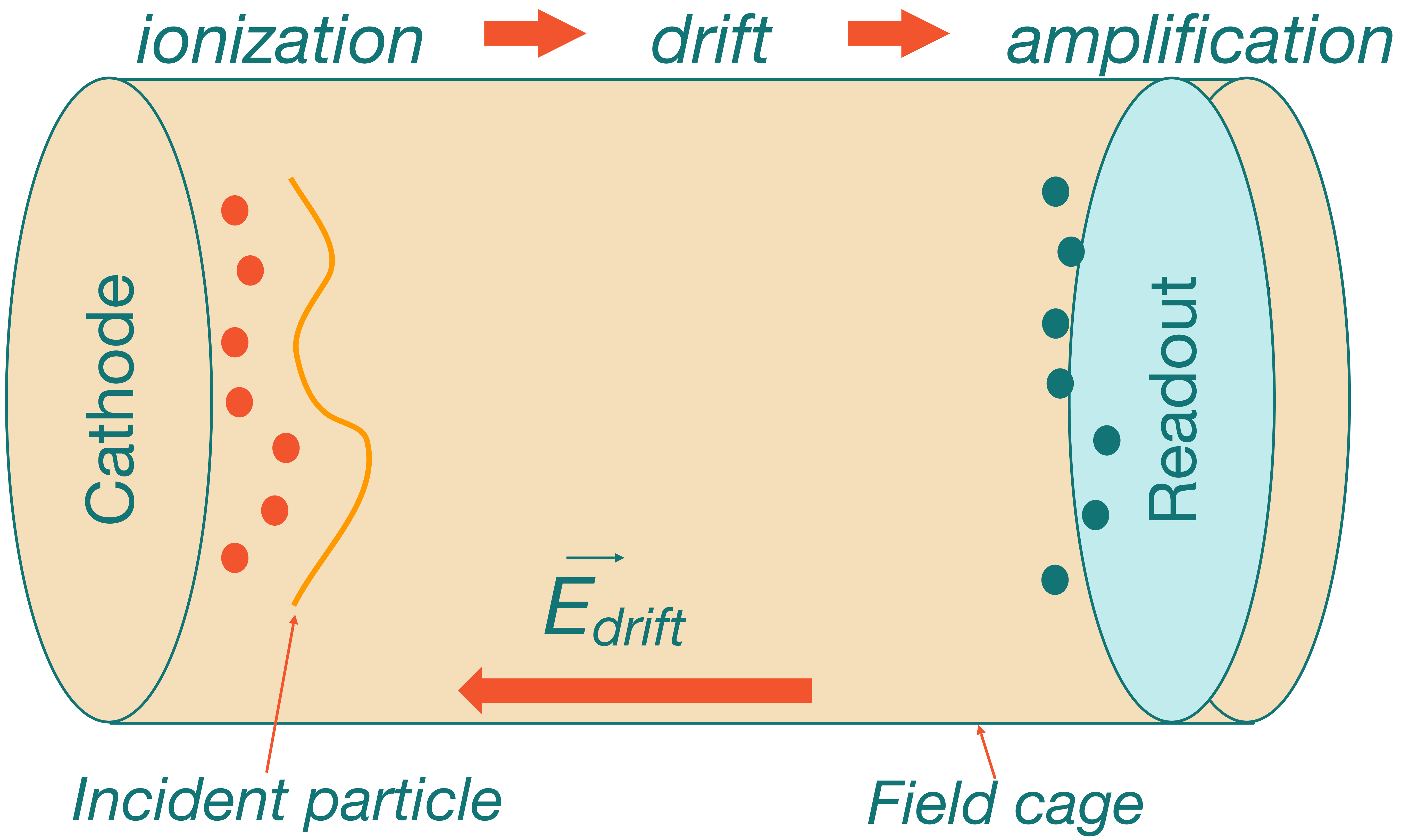
… start to drift in the direction of the anode and the cathode where …

TPC Detector in a nutshell



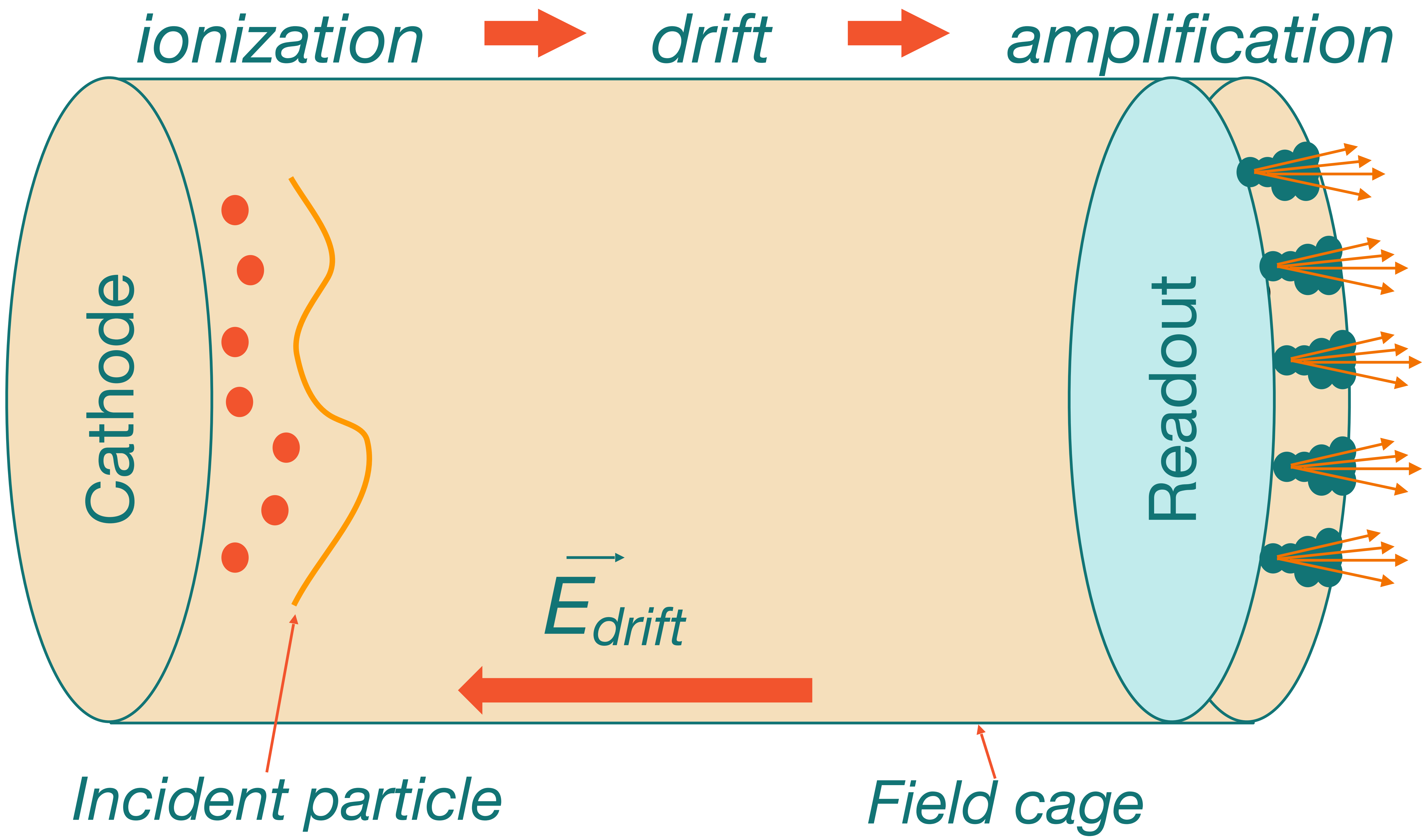
… a readout device is placed.

TPC Detector in a nutshell



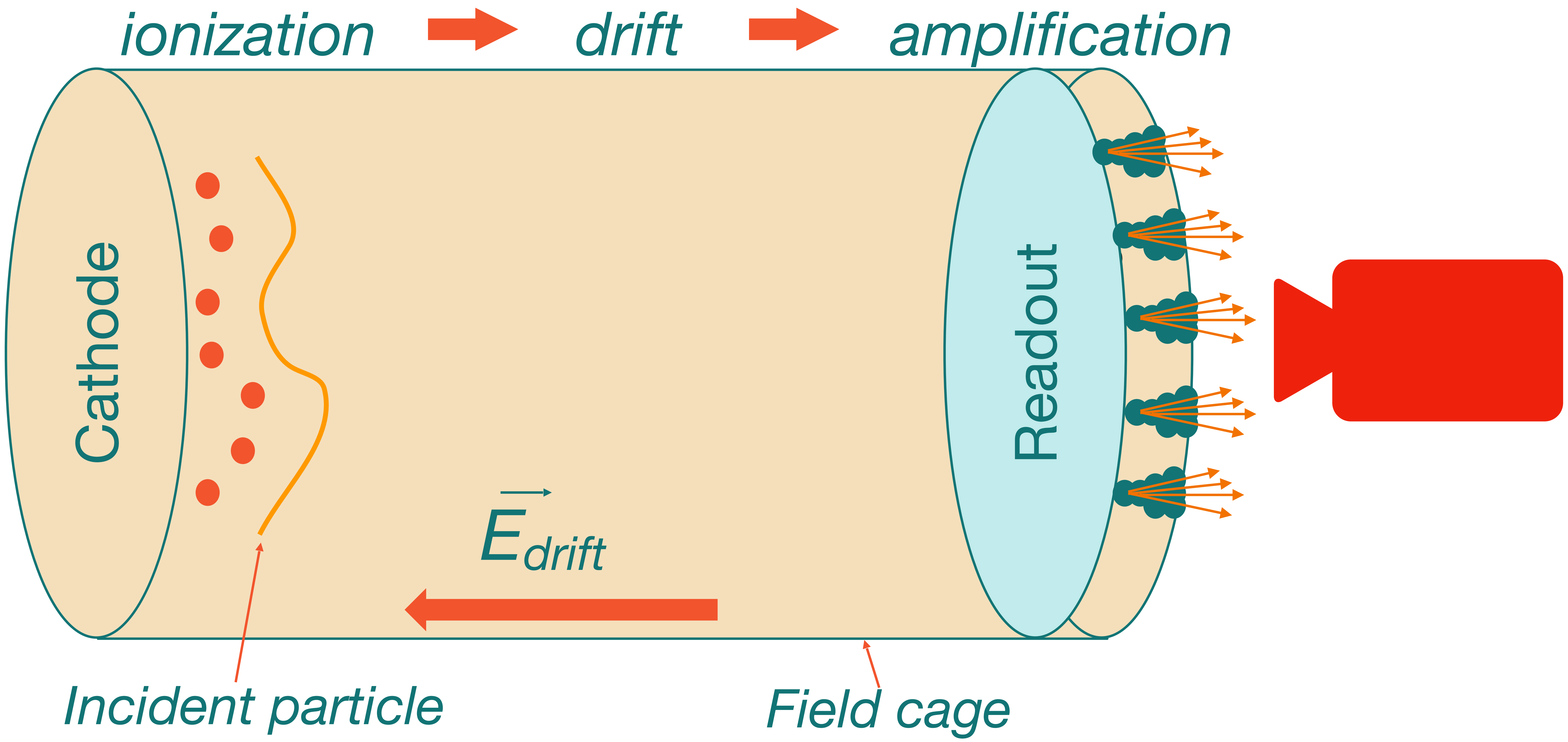
in gas TCP an amplification process by means of triple **Gas Electron Multiplier** (GEM) and produce an avalanche of electrons …

TPC Detector in a nutshell



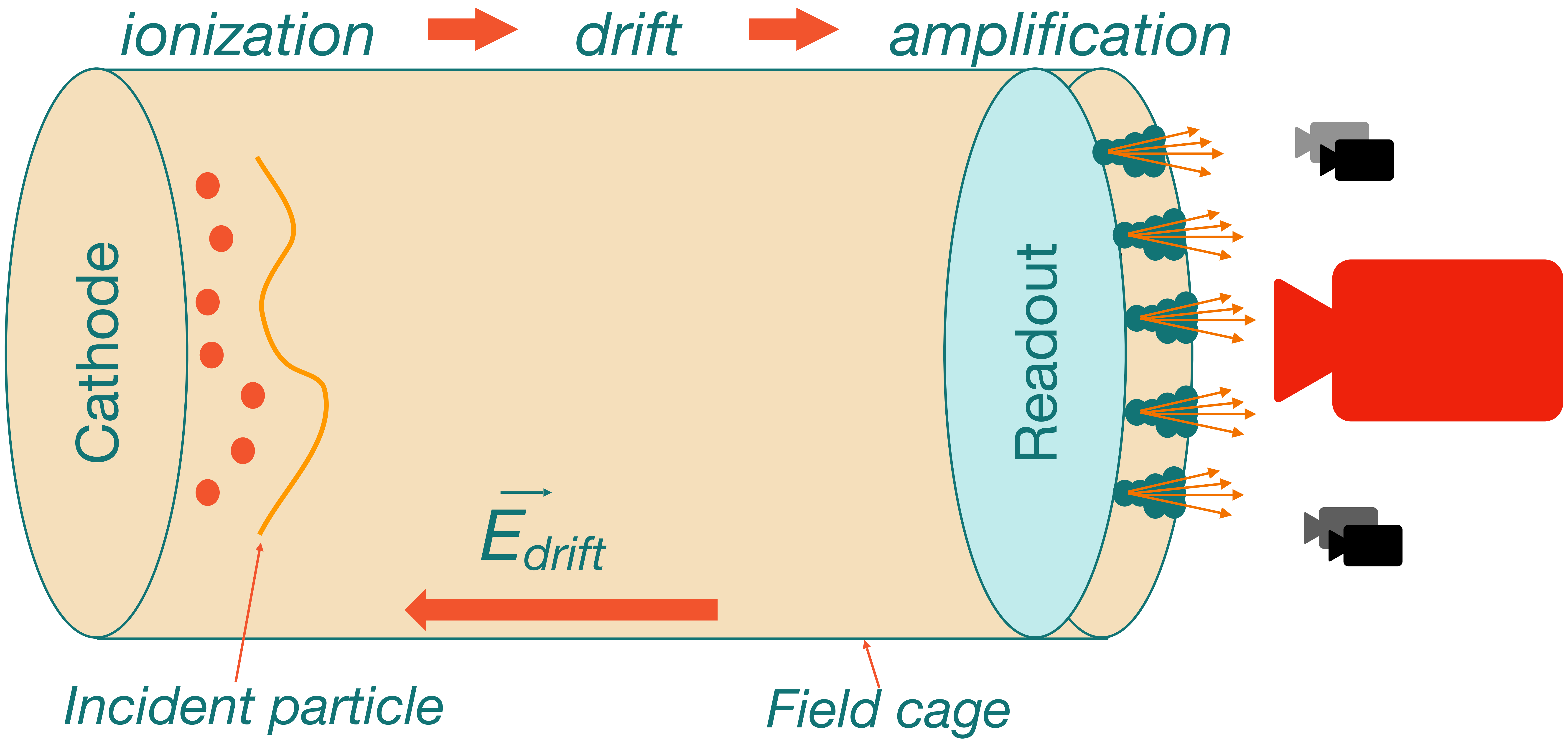
… that generate photons with an efficiency ~ 7-8% in HeCF4 gas mixture.

optical readout in a nutshell



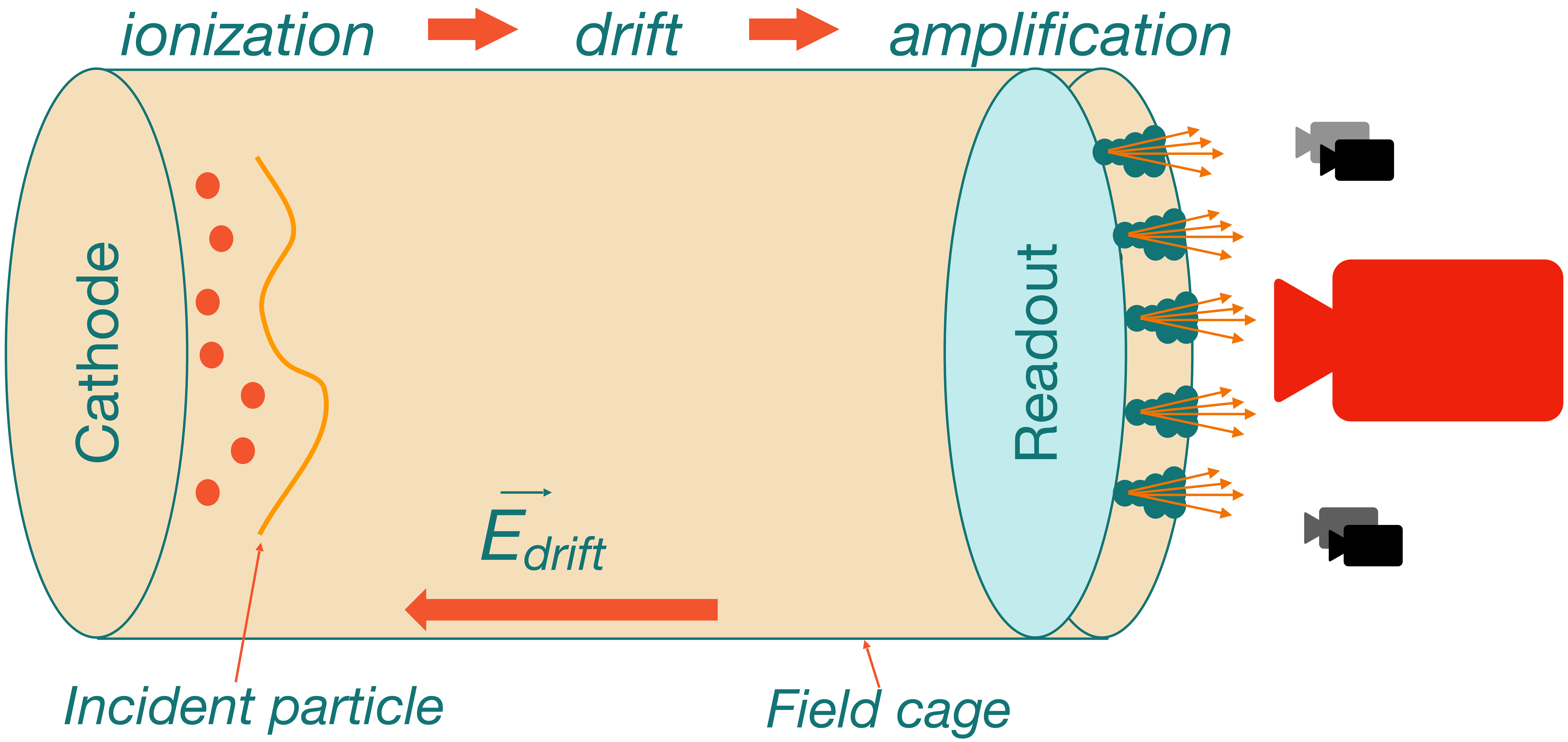
an **sCMOS** camera 2304×2304 resolution, 0.7 electrons rms, equipped with standard optics collects the photons produced

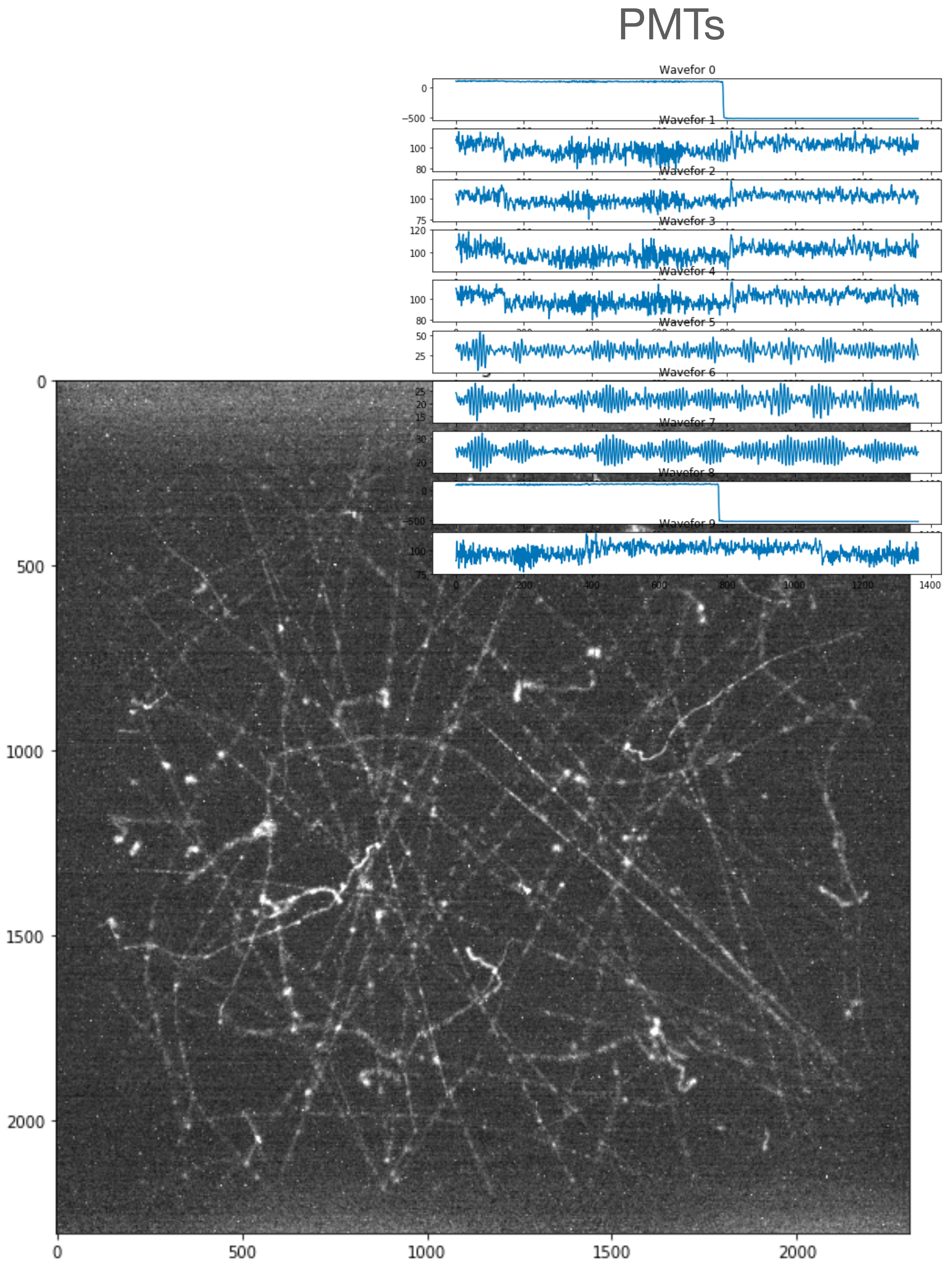
optical readout in a nutshell



4 **PMTs** symmetrically placed around the camera to detect the time shape longitudinal evolution

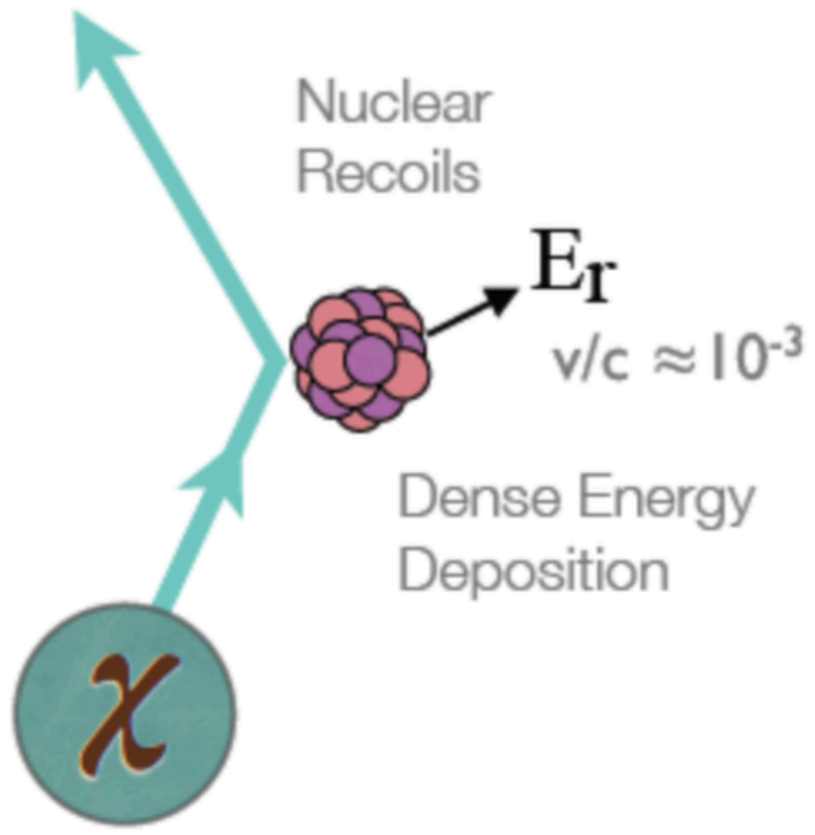
optical readout in a nutshell

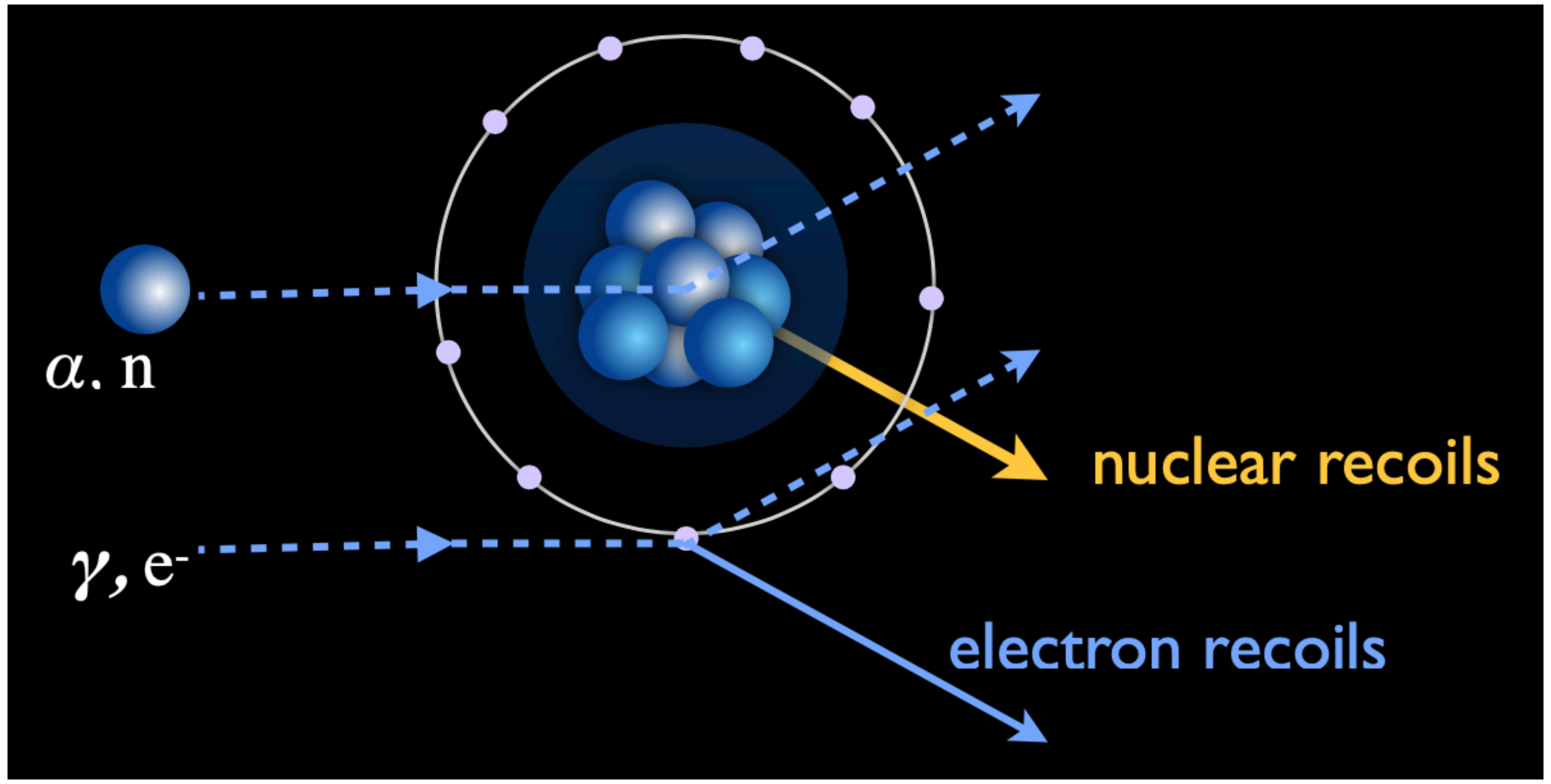




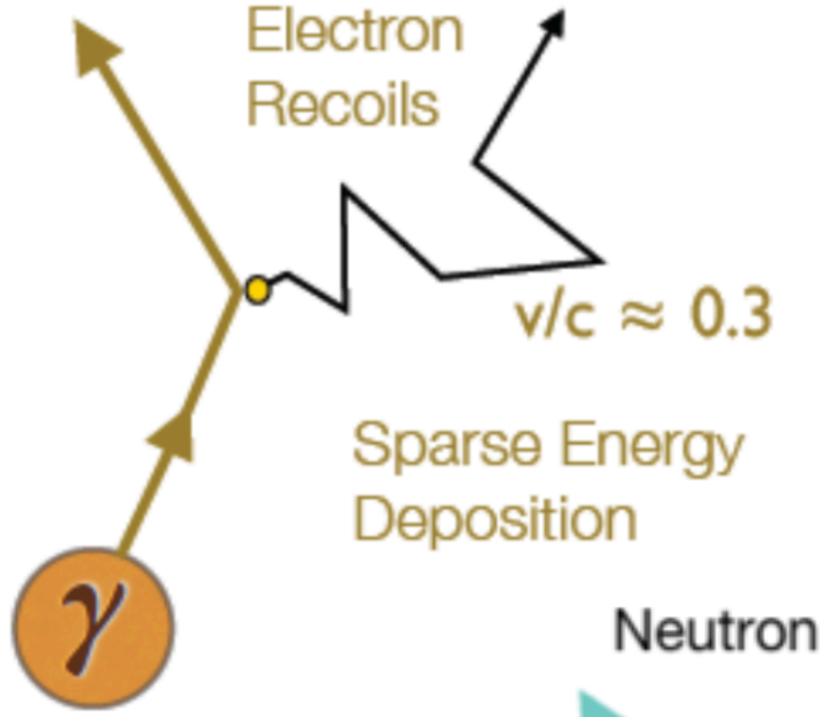
cosmic and radioactivity at see level,
in 500 ms image over 30*30 cm area

underground signal and background

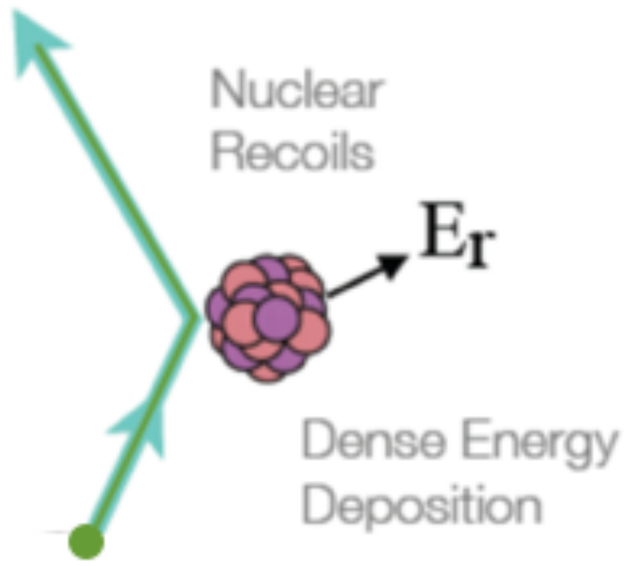




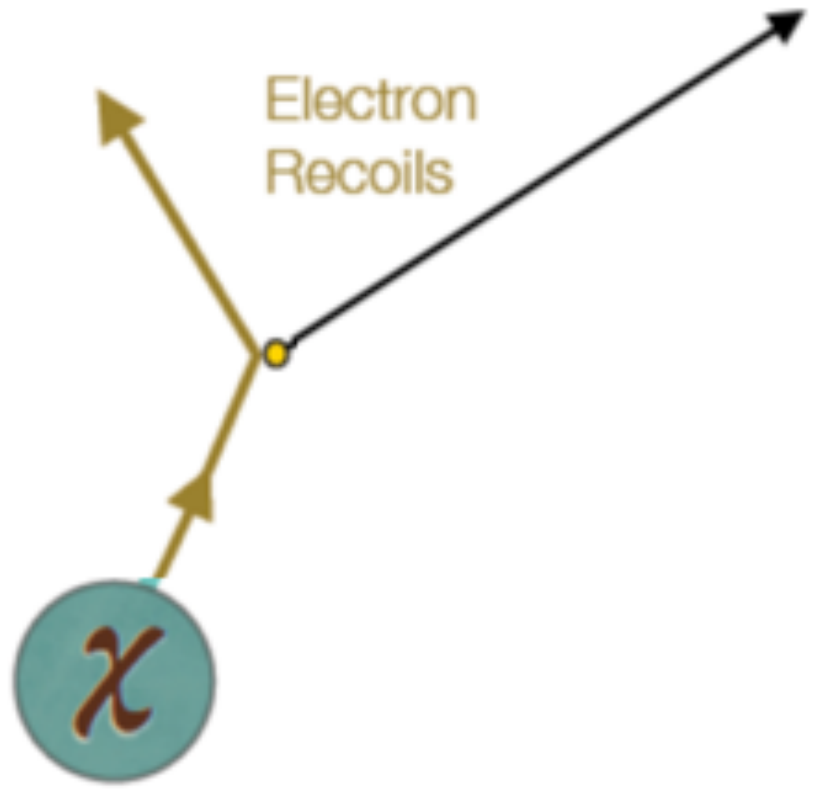
Background



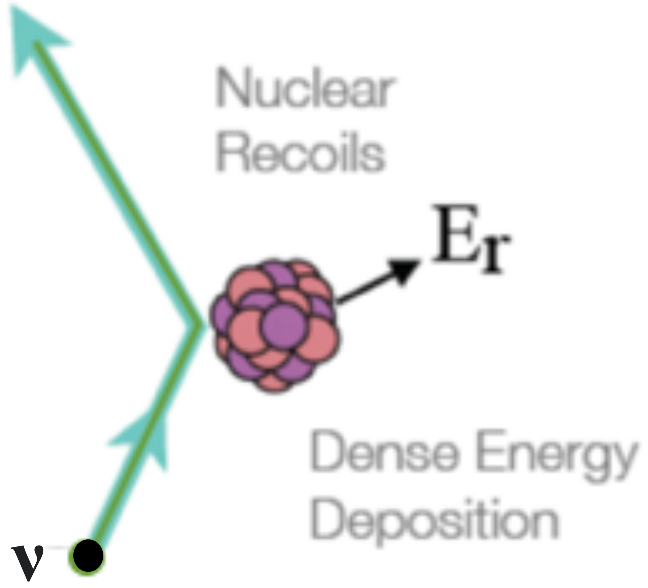
Neutron background

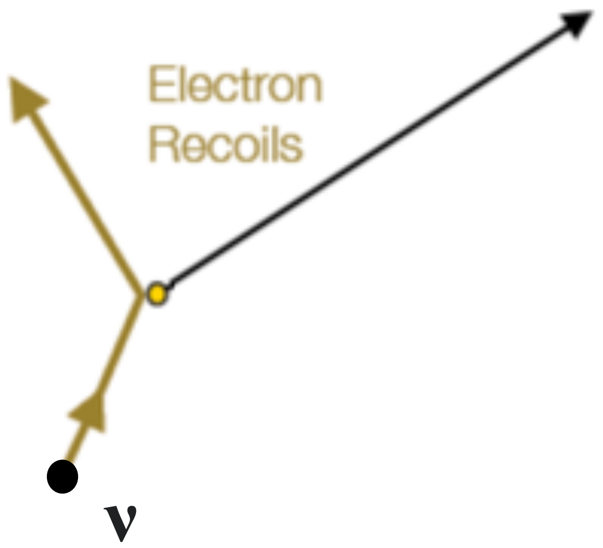


or alpha …

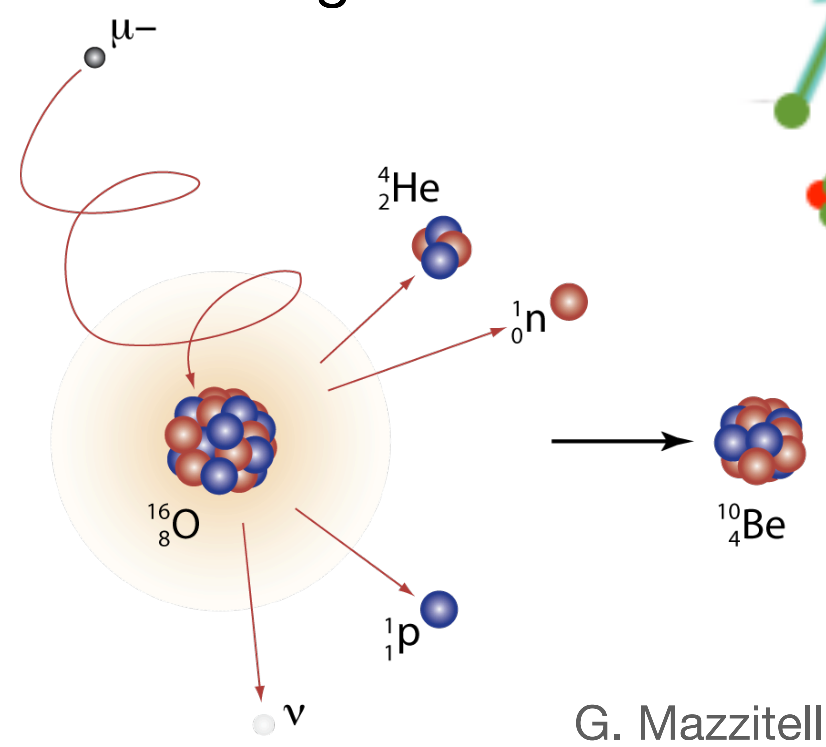


Signal <—> Background



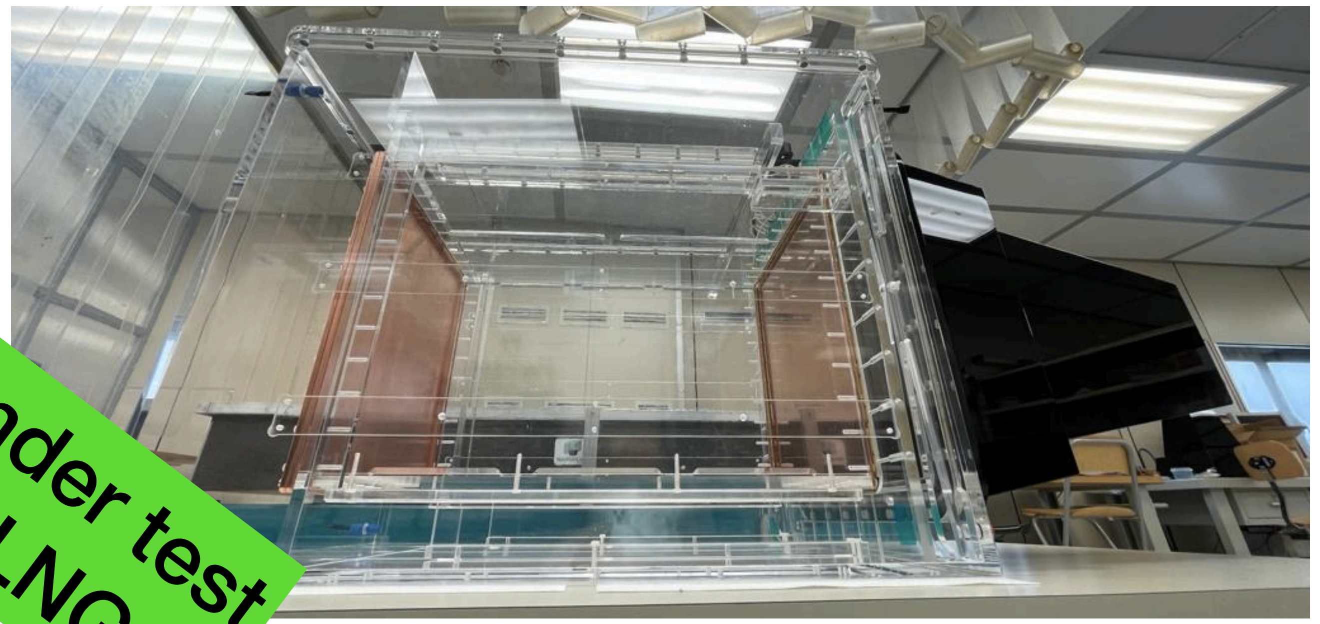


muons background



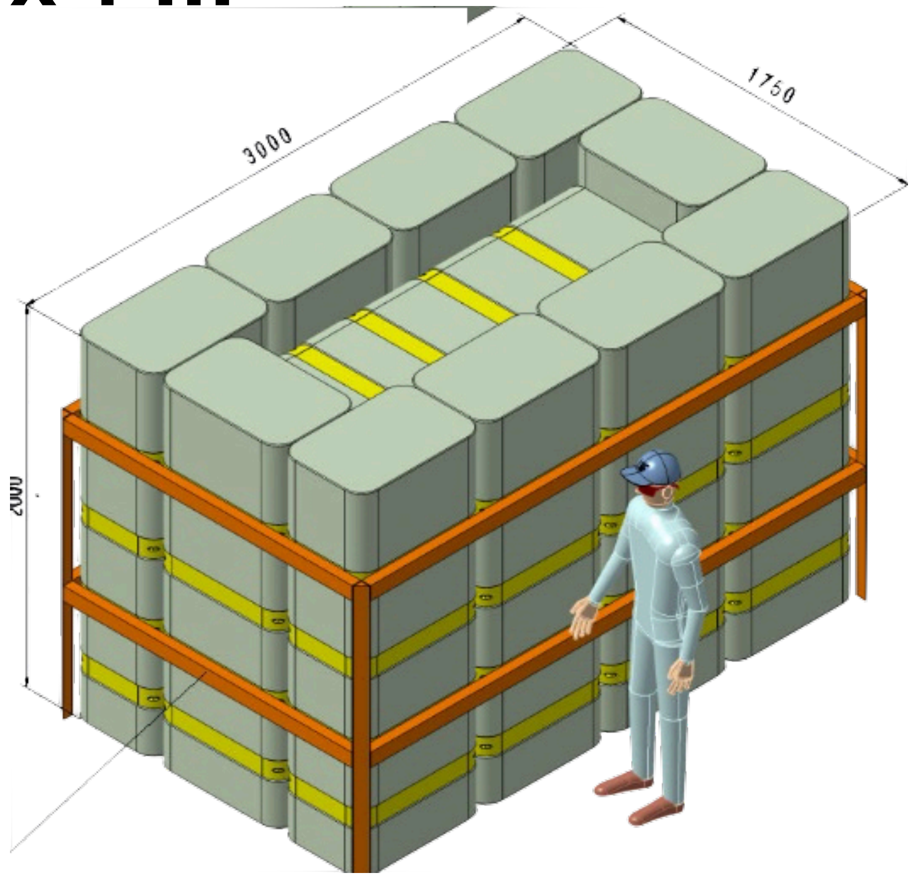
CYGNO (tech goal) to CYGNO 30 (physic goal)

the bet



under test at LNGS

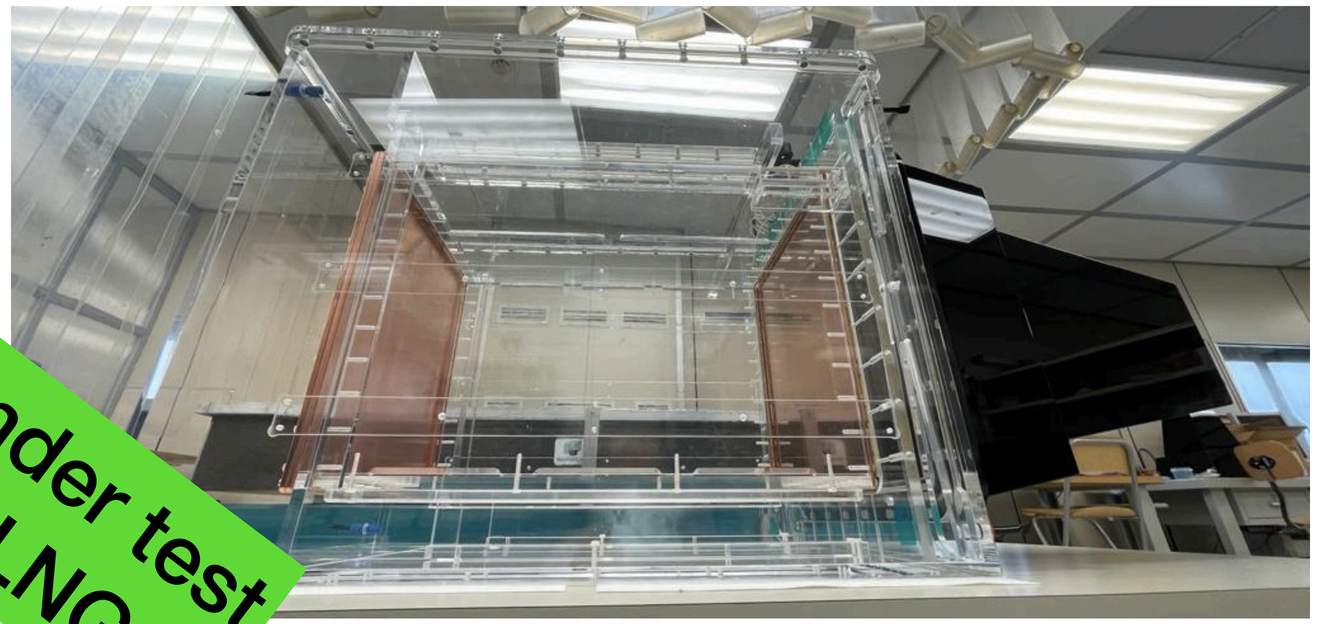
x 1 ...



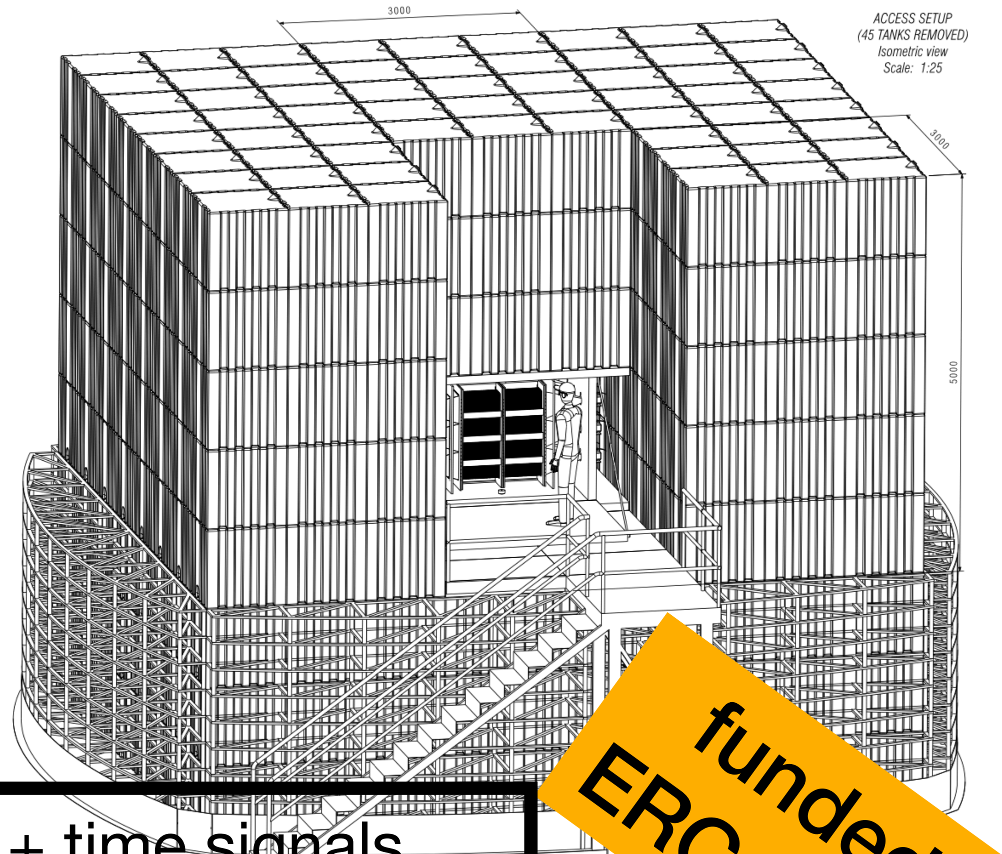
5*10 litres, 1 camera
2.5 MB/event 0.2—>0.01 Hz

CYGNO (tech goal) to CYGNO 30 (physic goal)

the bet

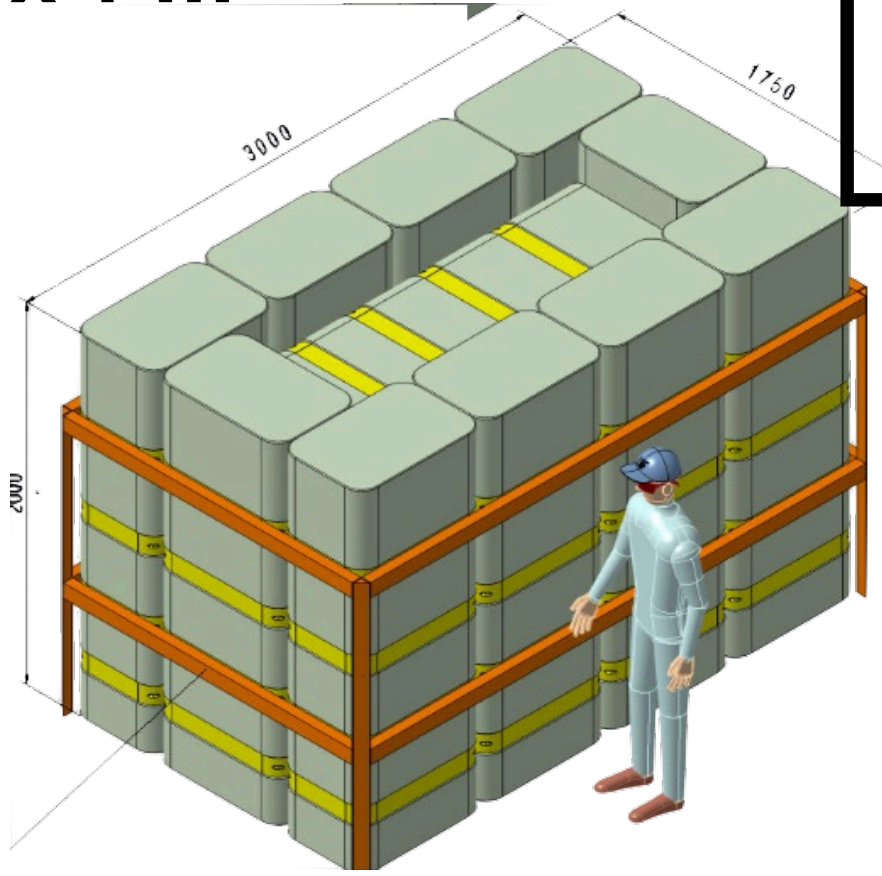


under test at LNGS

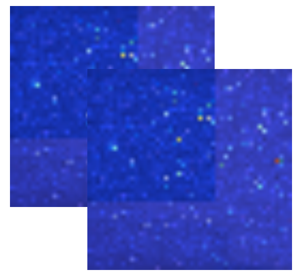


funded by ERC-INITIUM

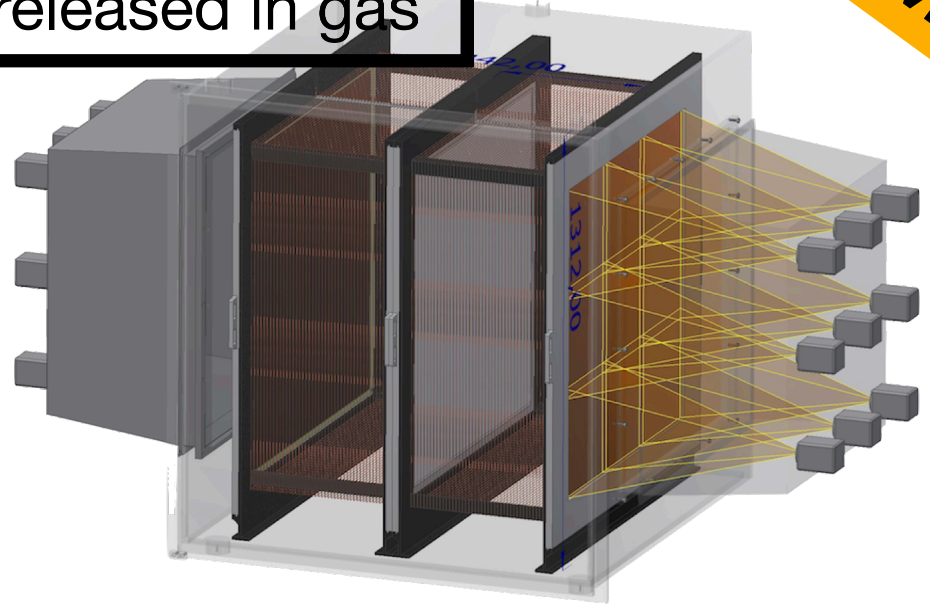
x 1 ...



10^7 readout channels + time signals
18 cameras monitoring 330*330 mm
each with **150 µm resolution** and a
sensitivity of ~ **1 ph / 2 eV** released in gas



x 9 ...



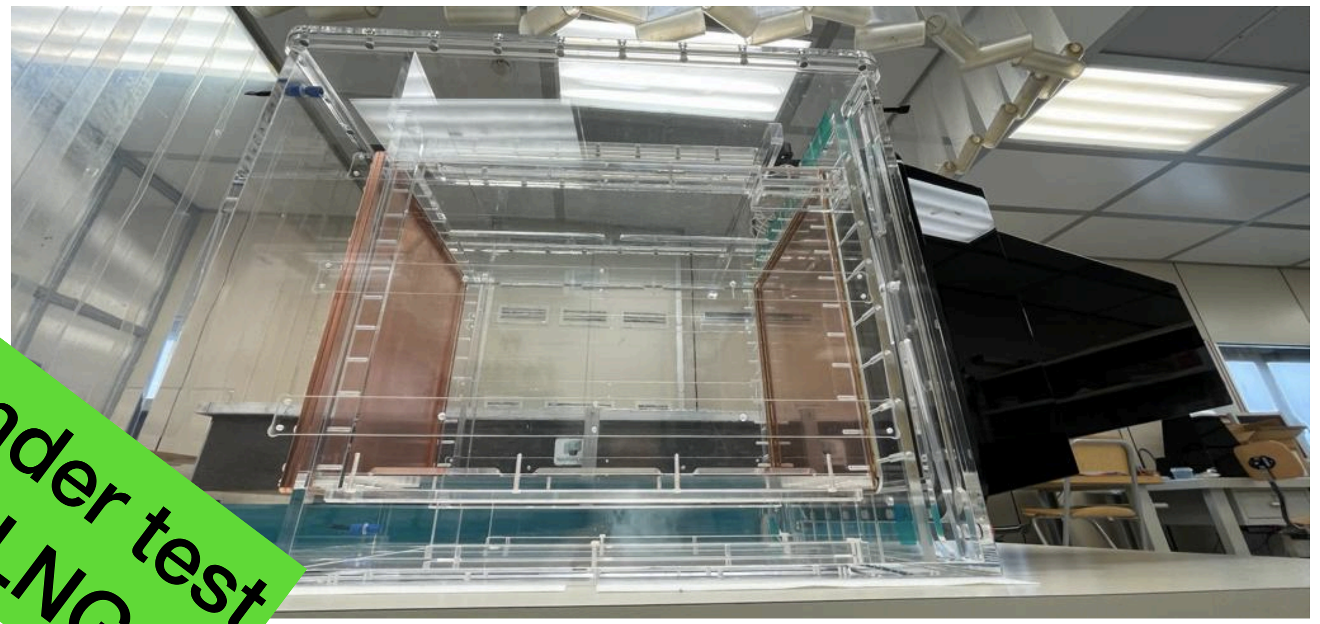
x 9 ...

5*10 litres, 1 camera
2.5 MB/event 0.2—>0.01 Hz

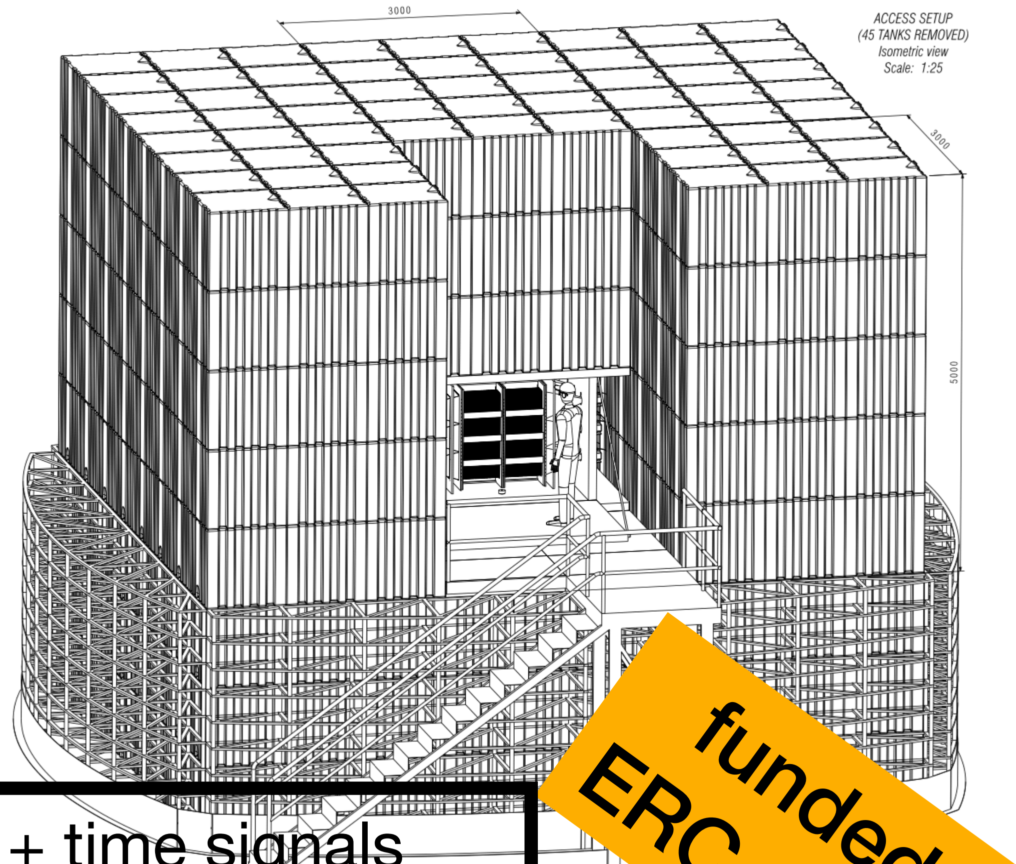
1*10^3 litres, 18 cameras
45 MB/event (Hz ?)

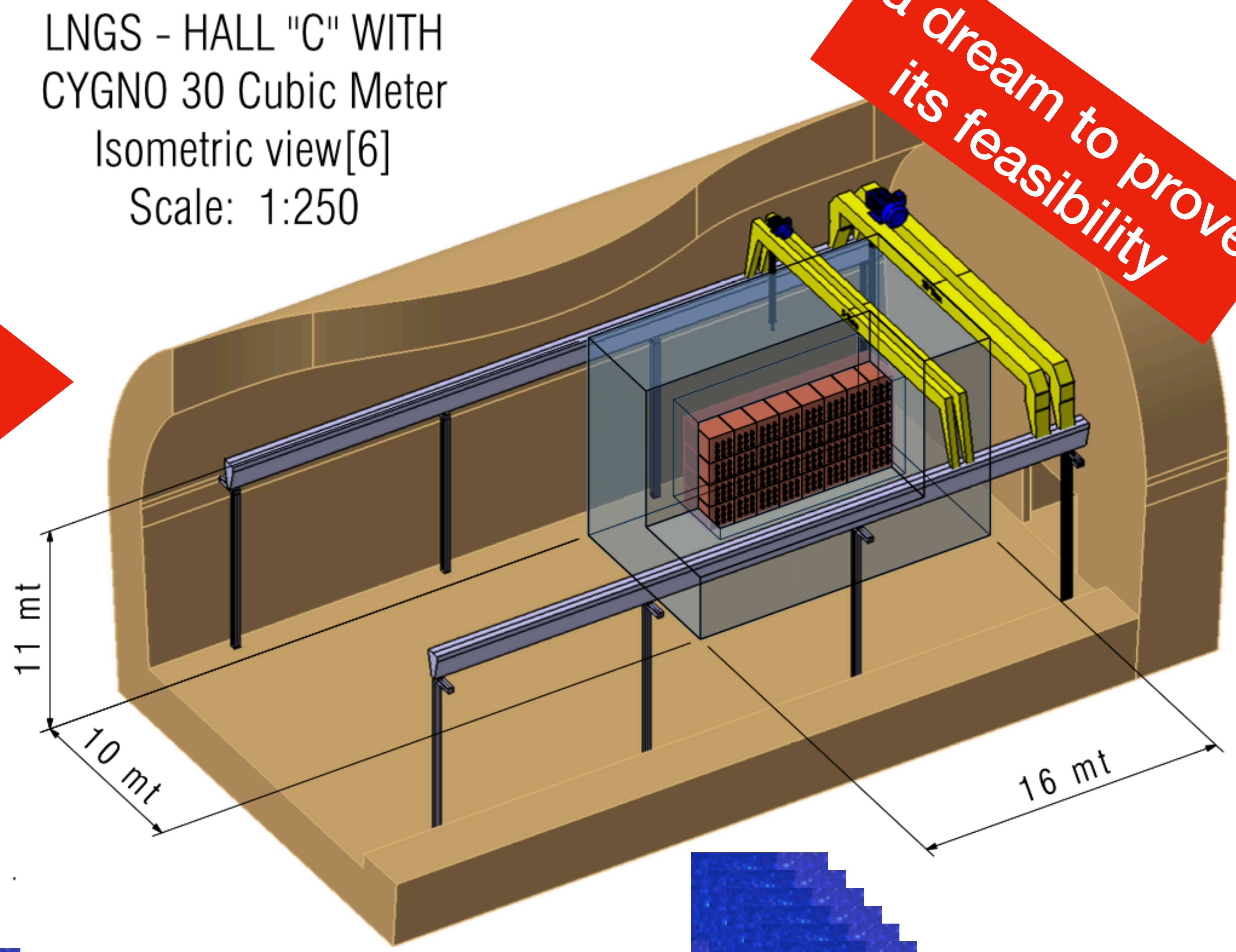
CYGNO (tech goal) to CYGNO 30 (physic goal)

the bet



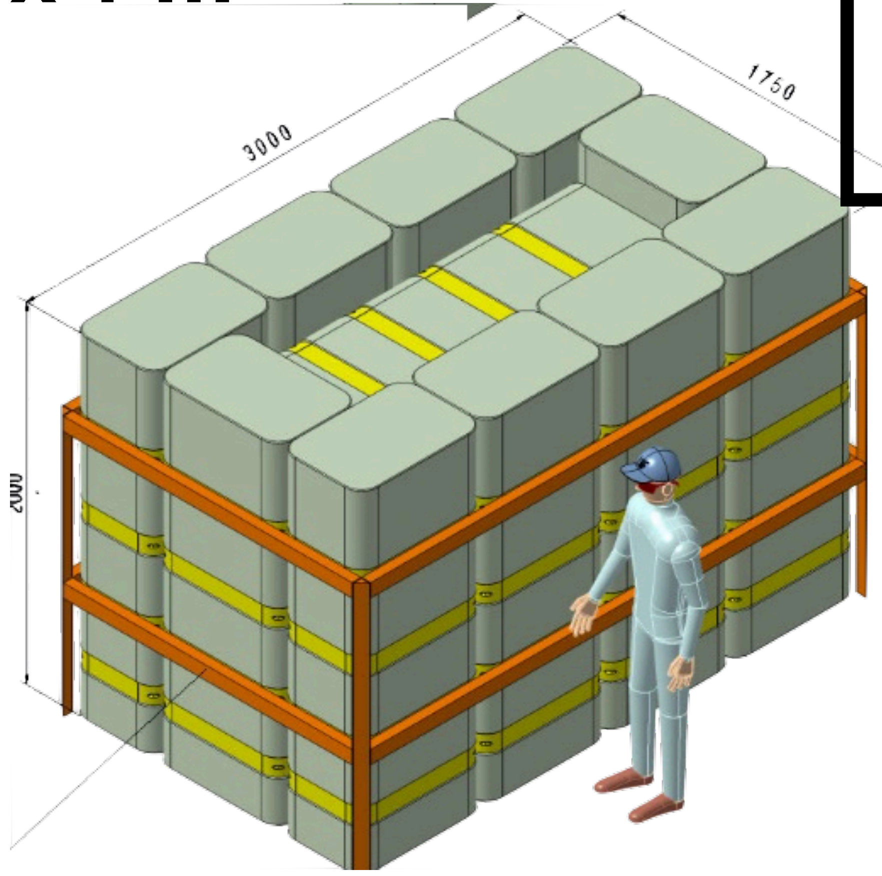
under test at LNGS





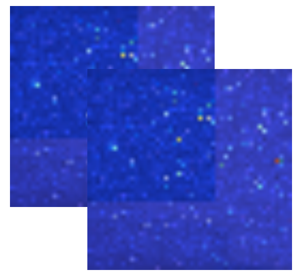
a dream to prove its feasibility

x 1 ...

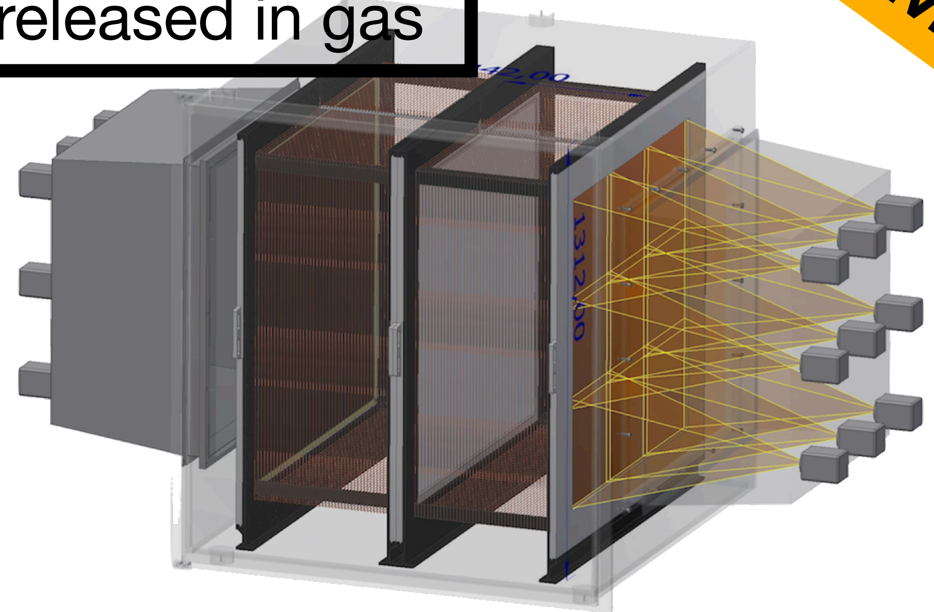


10^7 **readout** channels + time signals
18 cameras monitoring 330*330 mm
each with **150 µm resolution** and a
sensitivity of ~ **1 ph / 2 eV** released in gas

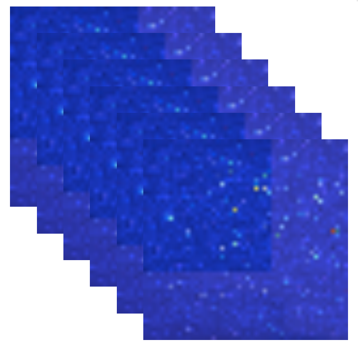
funded by ERC-INITIUM



x 9 ...



x 9 ...



x N ...

5*10 litres, 1 camera
2.5 MB/event 0.2—>0.01 Hz

1*10^3 litres, 18 cameras
45 MB/event (Hz ?)

3*10^4 litres, 540 cameras
1.3 GB/event (Hz ?)

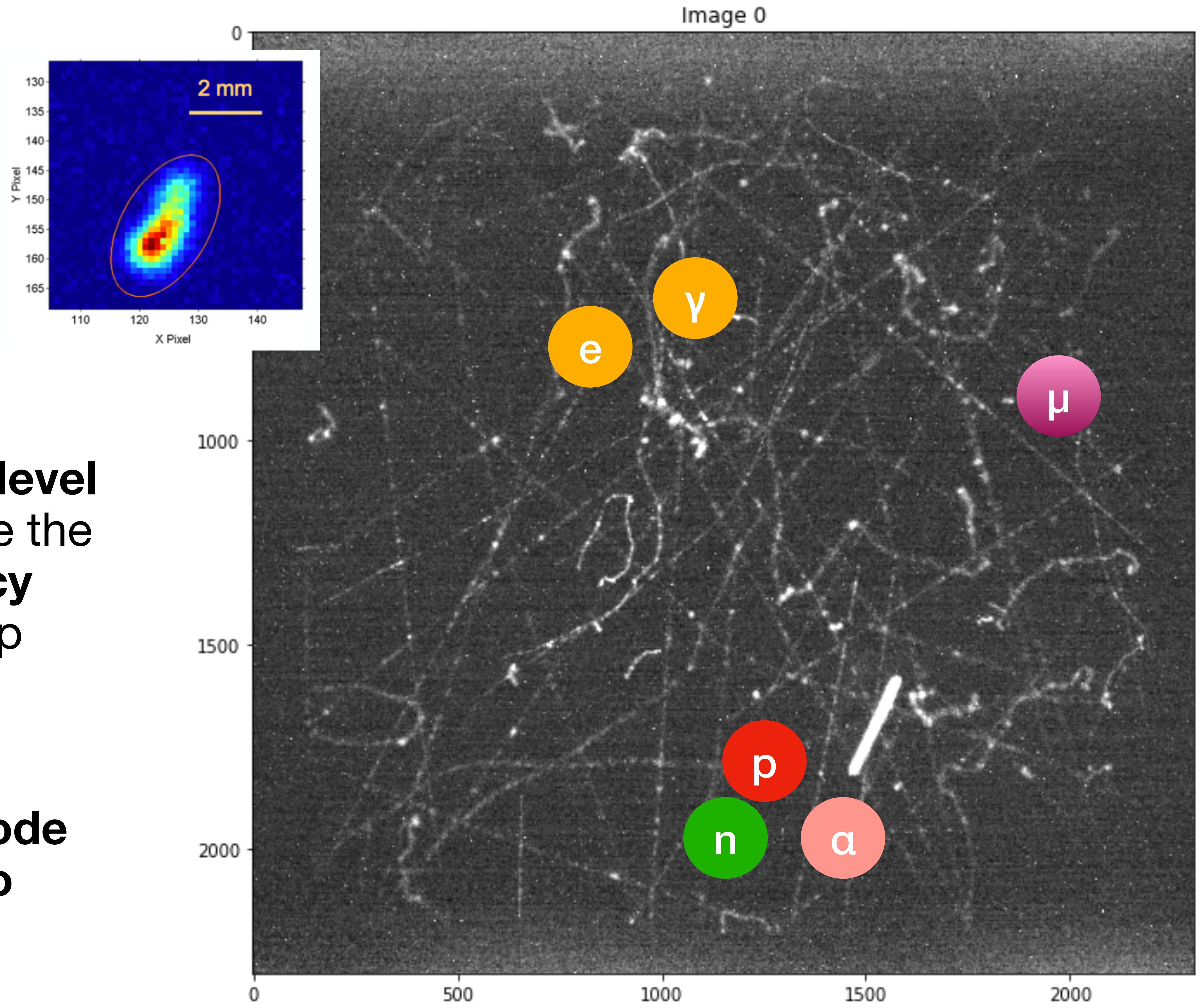
particle identification “gym”

nuclear recoil, neutrino vs beta/gamma discrimination

- exploiting dE/dx ionisation power (density)
- exploiting dE/dx ionisation profile vs path (shape, head-tail, snaking, etc)
- exploiting directionality
- exploiting time shape profile

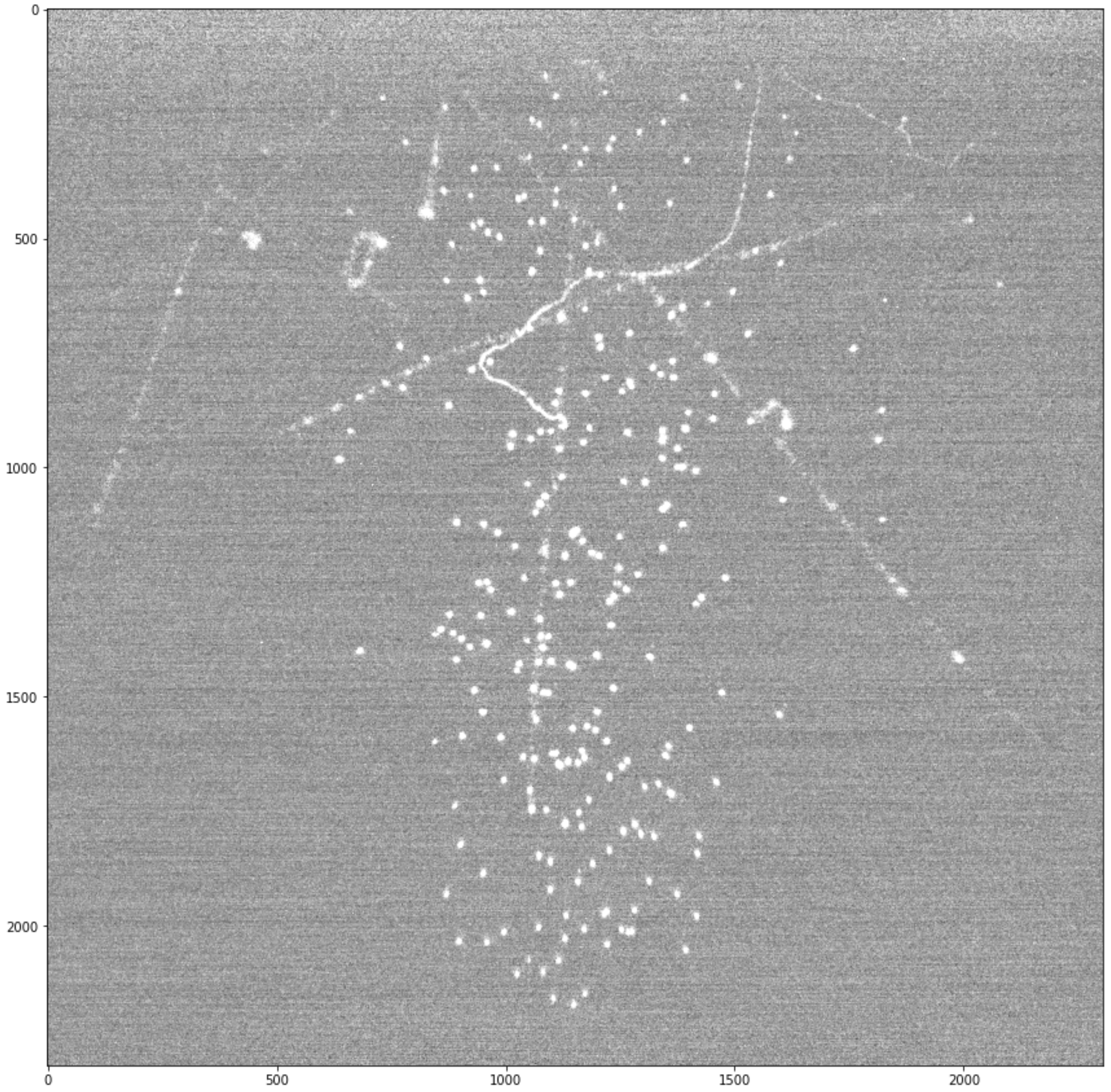
up to now we are **training our software** in the “**sea level gym**” where natural radioactivity and cosmic rays are the main issue to deal with, generating strong **occupancy and pileup** in the data. This has forcing us to develop code aimed mainly at removing **background** (not expected underground) than to identify the signal

LIME, just installed underground, will allow to **test code** in the real environment **and validate the Montecarlo** simulation



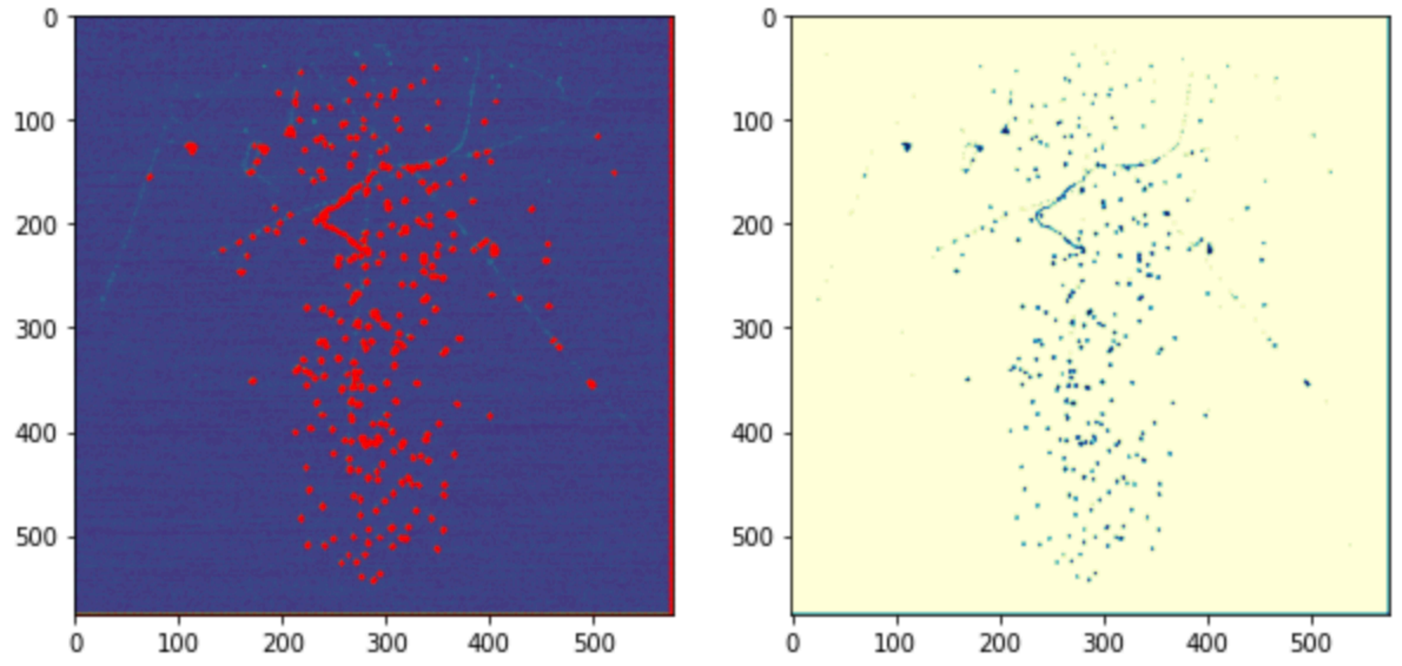
clustering

unsupervised learning NCC, KNN, HC to DBSCAN

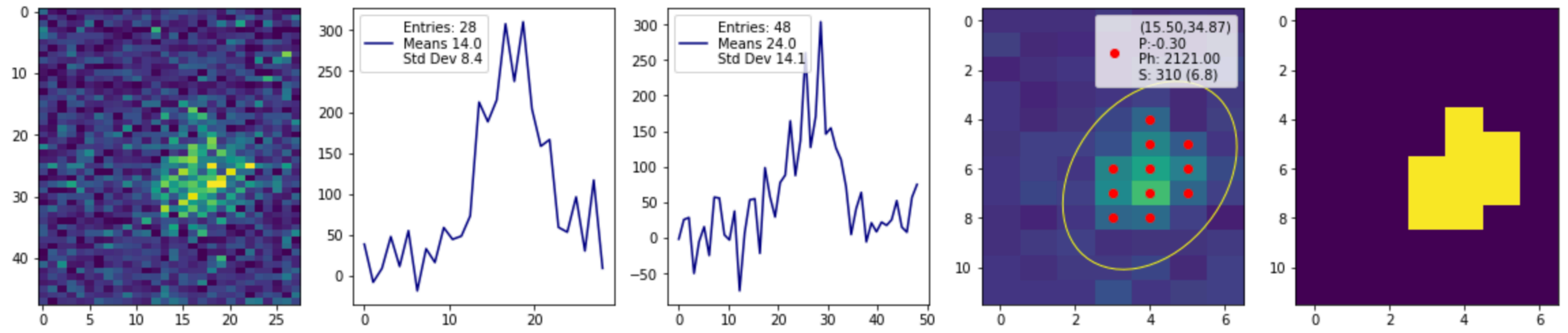


^{55}Fe source, in 50 ms image over 30*30 cm area

```
DEBUG: number of points, clusters: 4865 255
Elapsed time 10 events: 49.4
['iTr: 0.00', 'cluster_lable: 255.00', 'pixels: 310.00', 'photons: 2121.00', 'ph_pixels: 6.84', 'x0start: 1148.00', 'y0start: 2160.00', 'x0end: 1148.
00', 'y0end: 2176.00', 'width: 15.50', 'height: 34.87', 'pearson: -0.30']
```



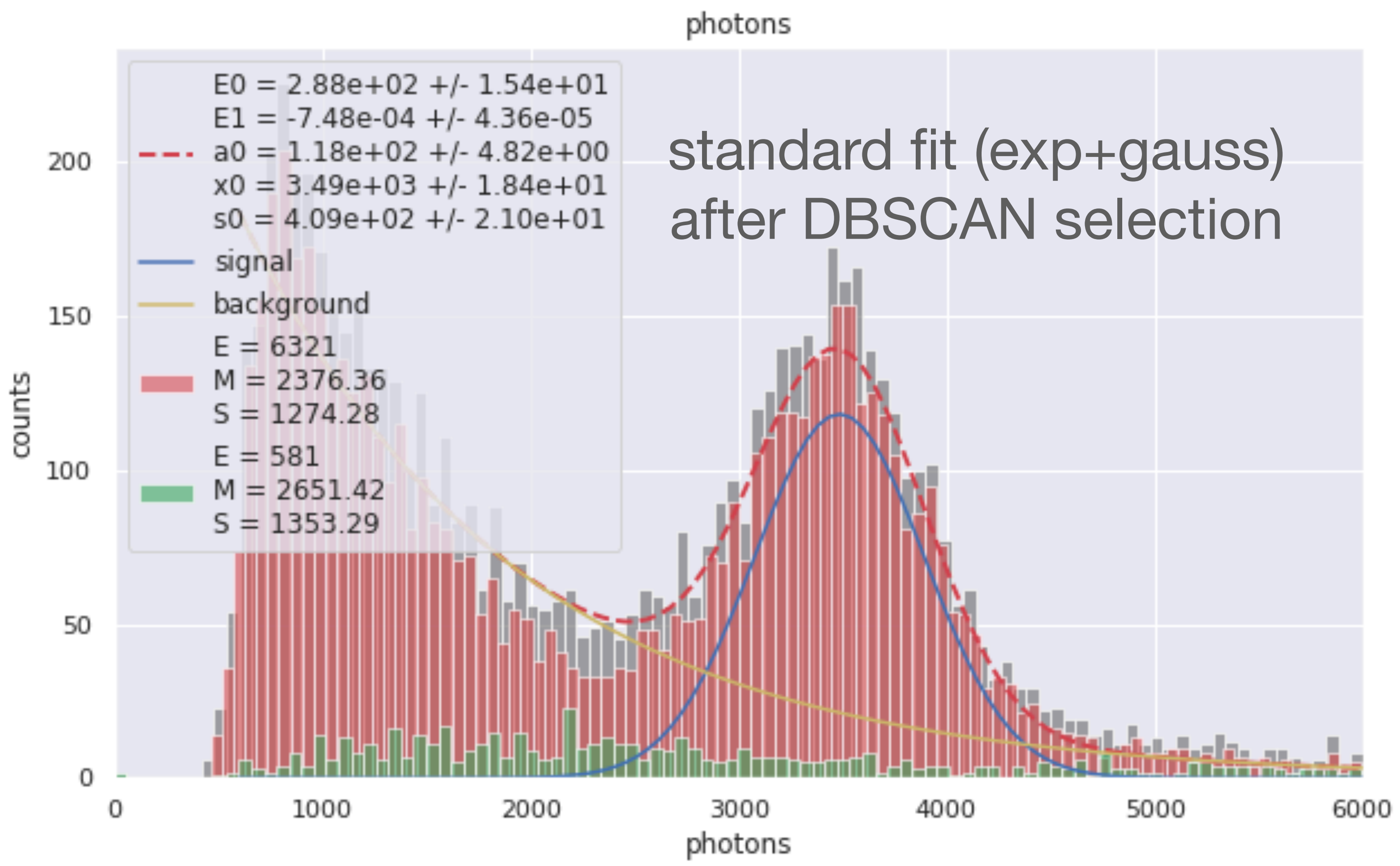
^{55}Fe source generate almost round spot due to the interactions of photos of **5.9 keV**, easily identifiable by cluster **shape/density** very useful for detector energy calibration. This signal up to know have to be select among **environmental and sensor background** and

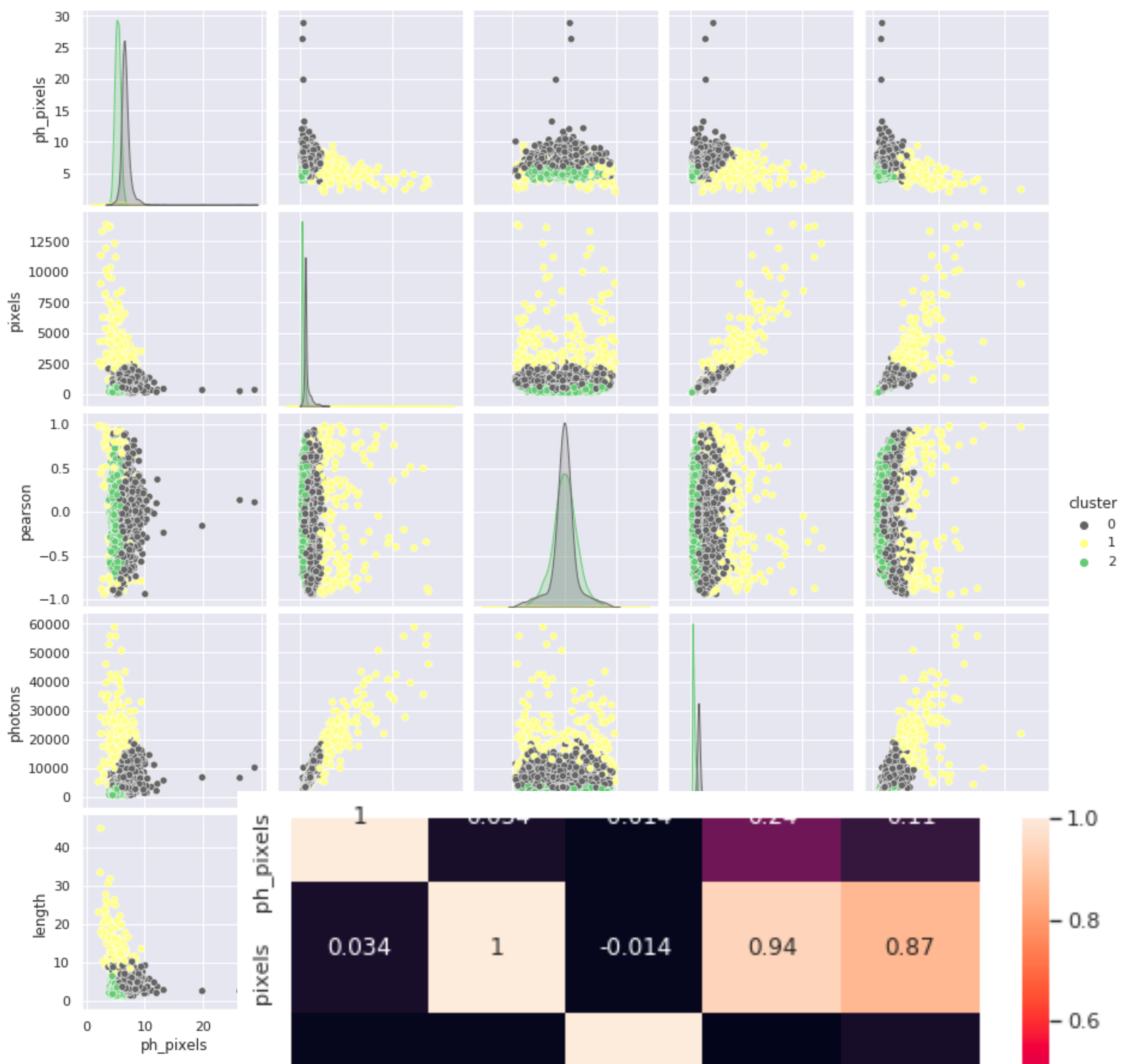


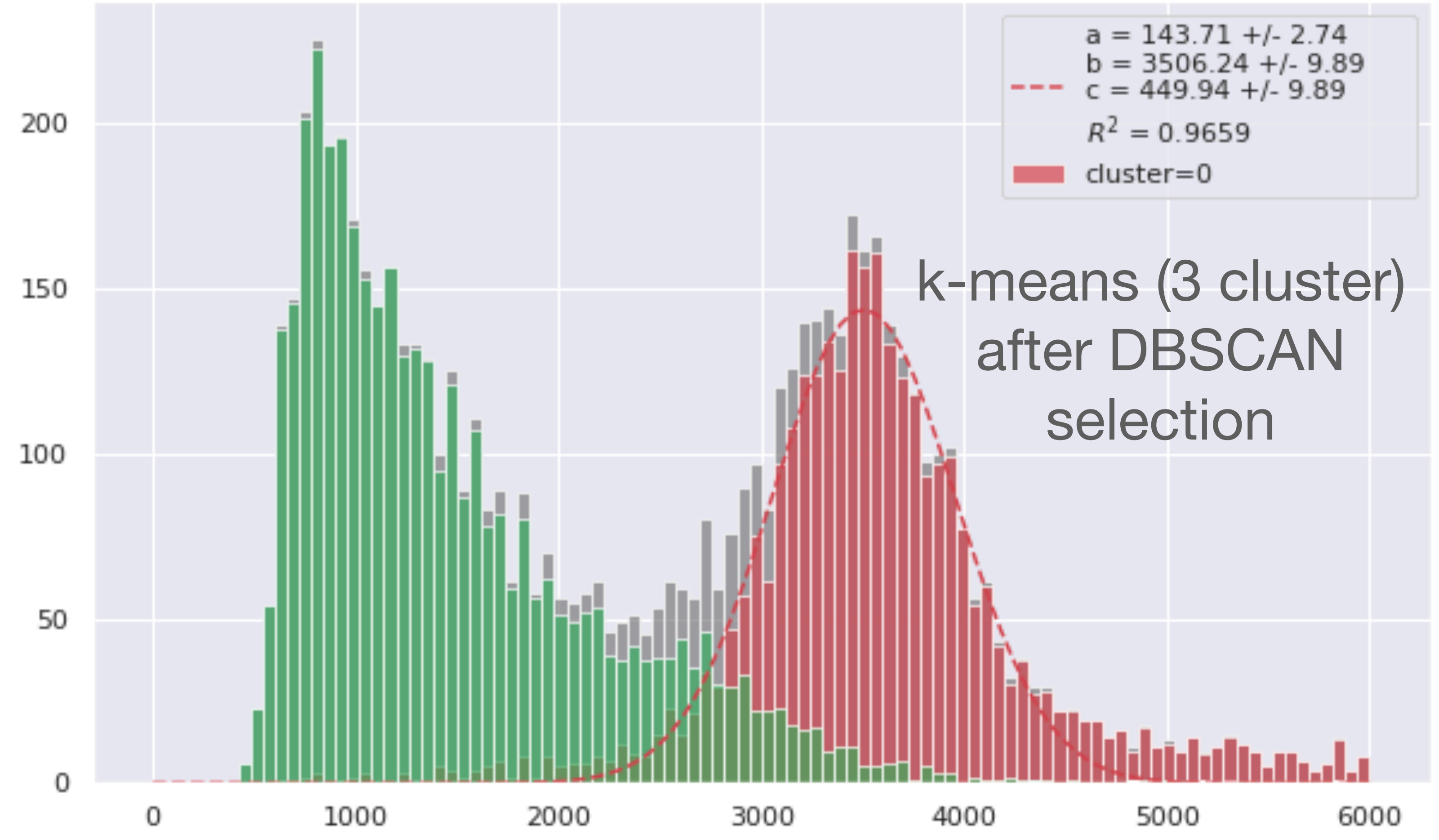
we start using NCC (Nearest Neighbors Clustering), K-Nearest Neighbors (KNN), Hierarchical clustering to arrive to Density Based Spatial Clustering able to detect closer in noise environments (DBSCAN)

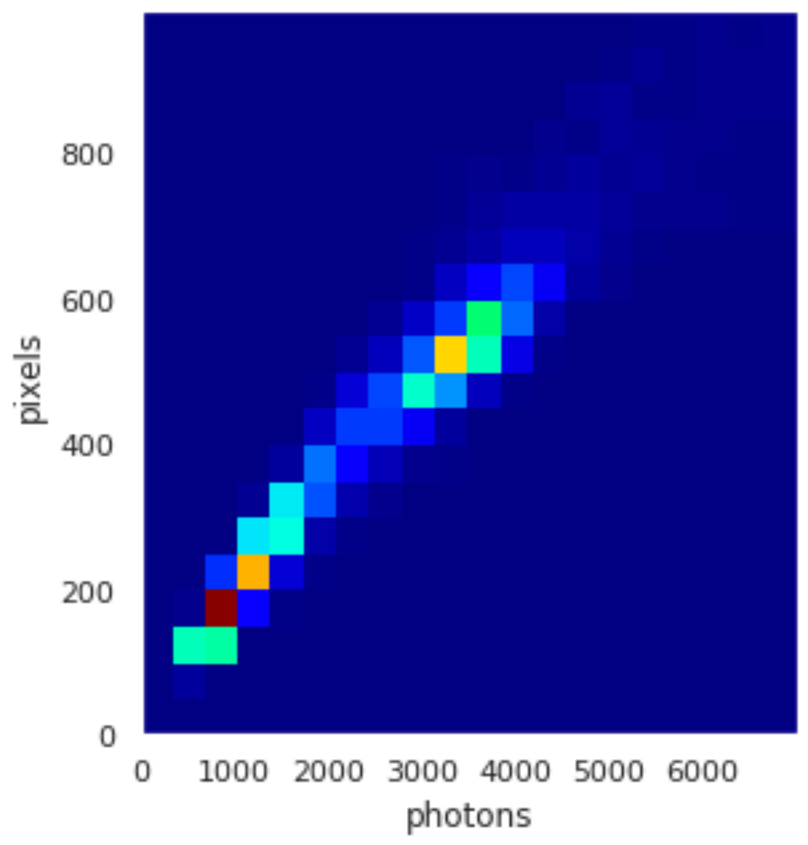
DBSCAN+K-means

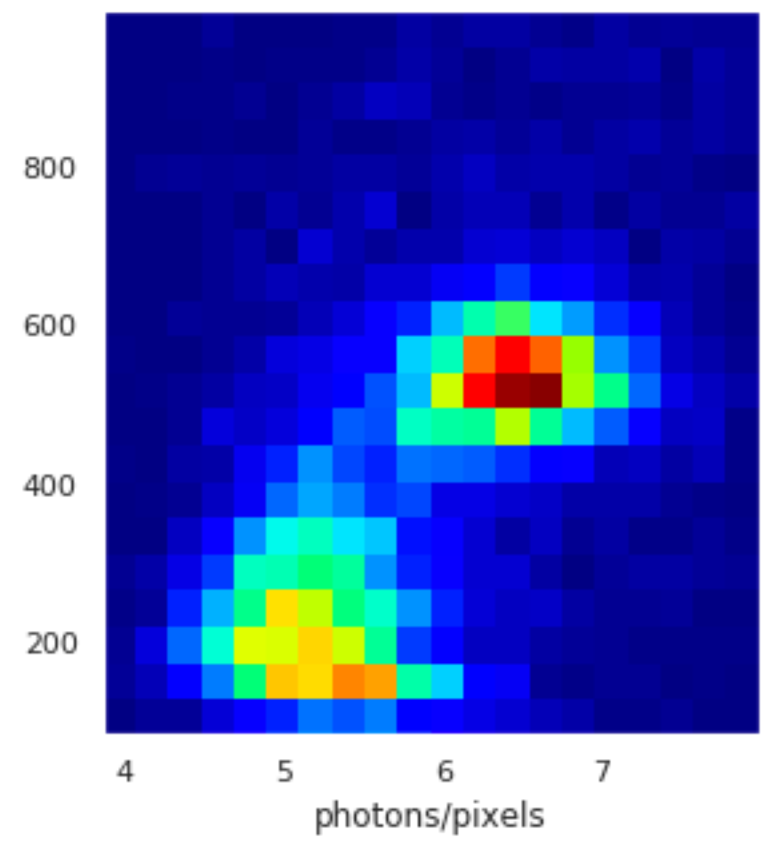
^{55}Fe source example detector resolution estimation











reconstruction flow chart

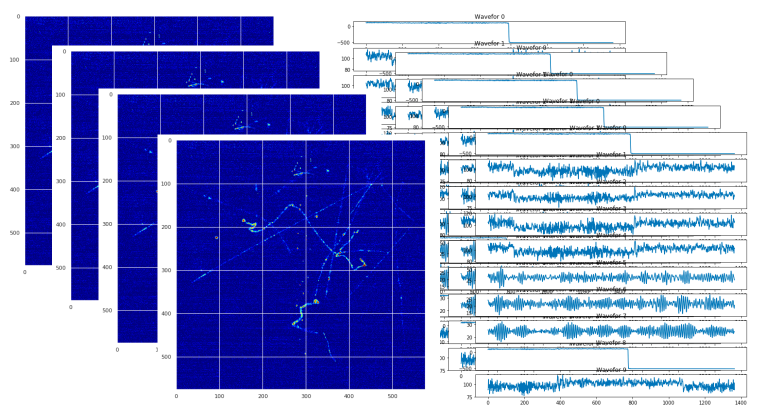
intensity DBSCAN

data preparation
zero suppression
noise filters
optical corrections

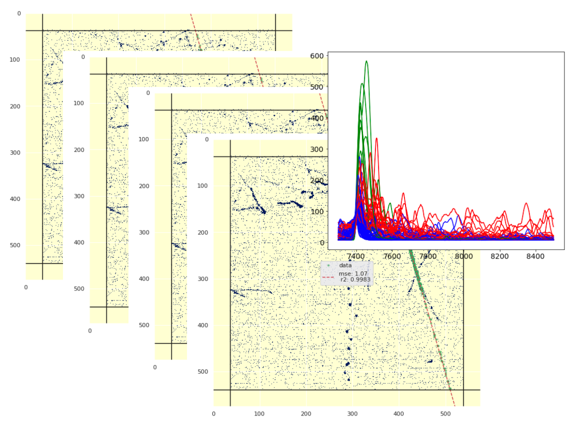
reconstruction
clustering &
superclustering
PMT info evaluation

sCMOS images

PMT waveforms

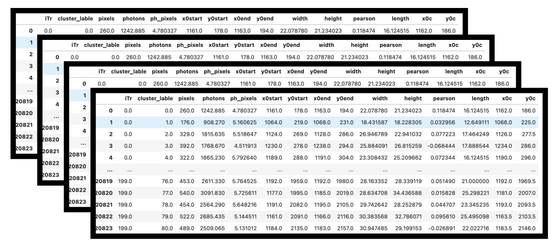


RAW events
calibration source, radioactivity,
cosmic rays, Dark Matter wind…



data “hits”

EVENTs data sets

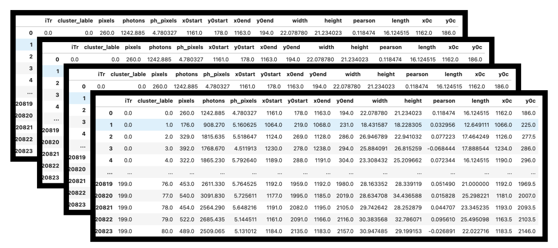


TTree clusters features
TTree PMTs features

E. Di Marco et al. @ INFN.it

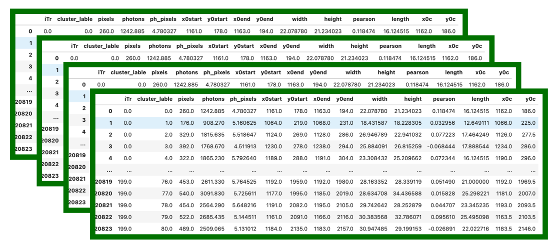
noise filter: median filter
optical correction: vignetting, optical distortion
superclustering: Geodesic Active Contour (GAC)

multi-EVENTs data sets



post-processing
calibrations
high-level analysis

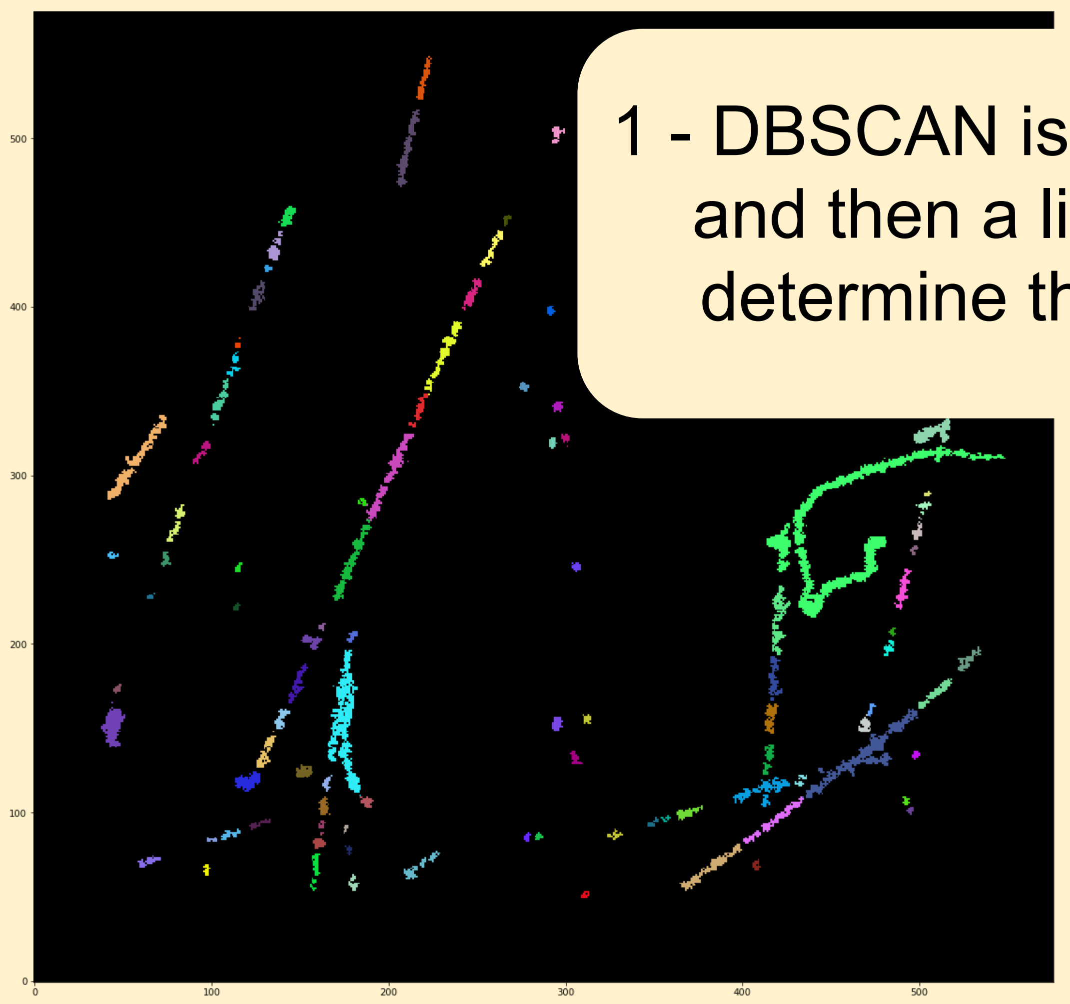
calibrated data sets



user analysis
physical outcomes

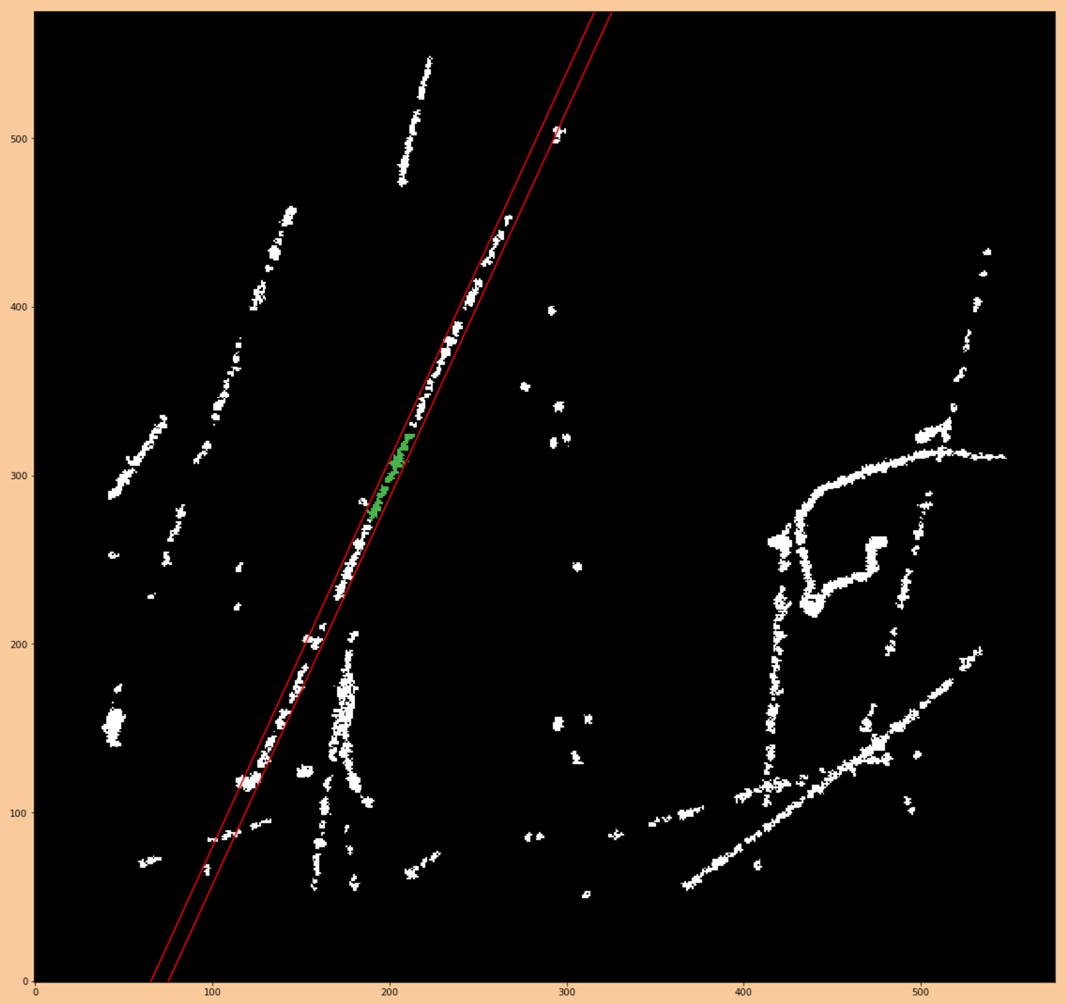
iDDBSCAN

i-ntensity Directional DBSCAN to identify cosmic rays

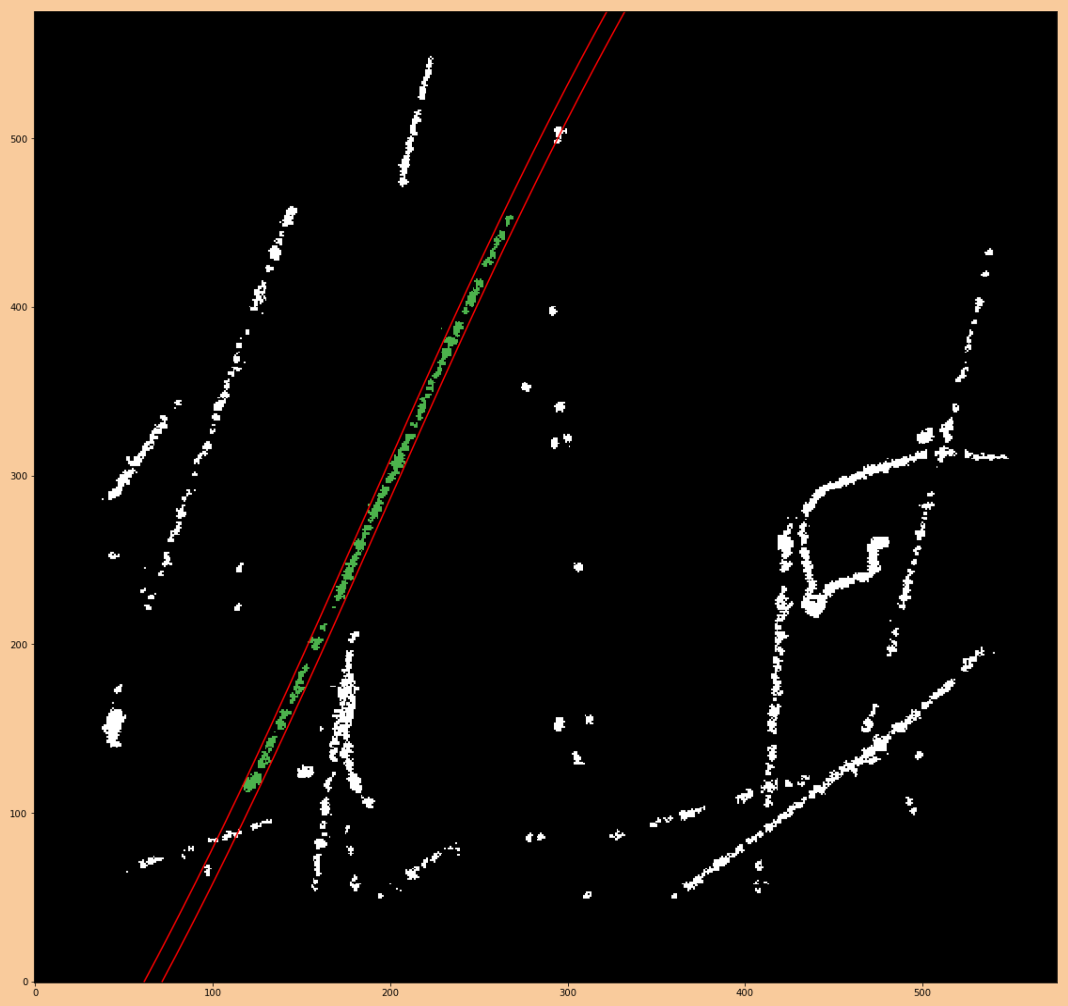


1 - DBSCAN is applied to detect the seeds and then a linear-RANSAC is used to determine the unidirectional clusters.

2 - The clustering restarts from the unidirectional clusters and a polynomial-RANSAC is used to fit them.



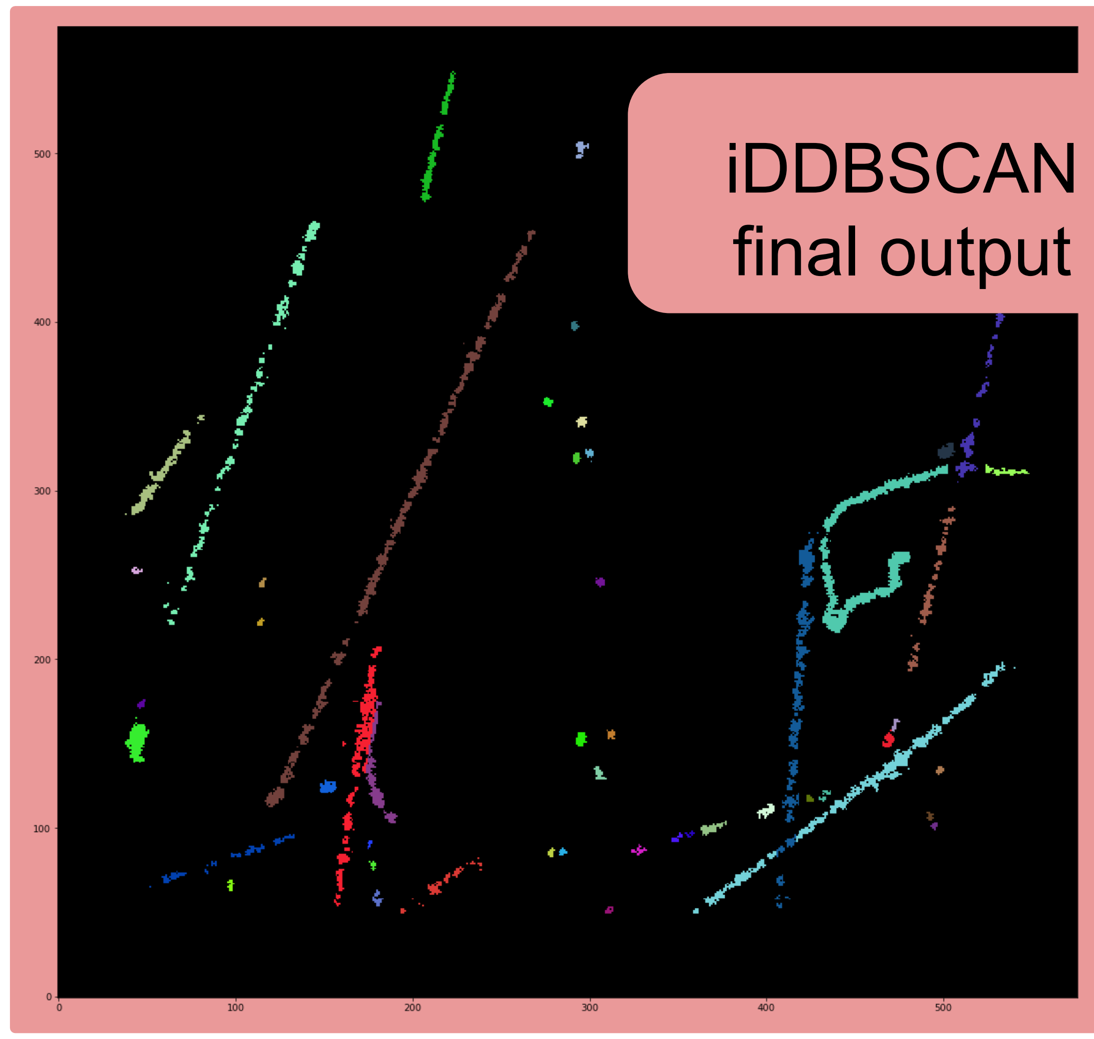
3 - Nearby points are added to the cluster based on the fit.



4 - The fit is updated whenever new points are added to the cluster.

5 - The procedure (2, 3, 4) is repeated on all unidirectional seeds.

6 - The remaining data is clustered using the DBSCAN logic.

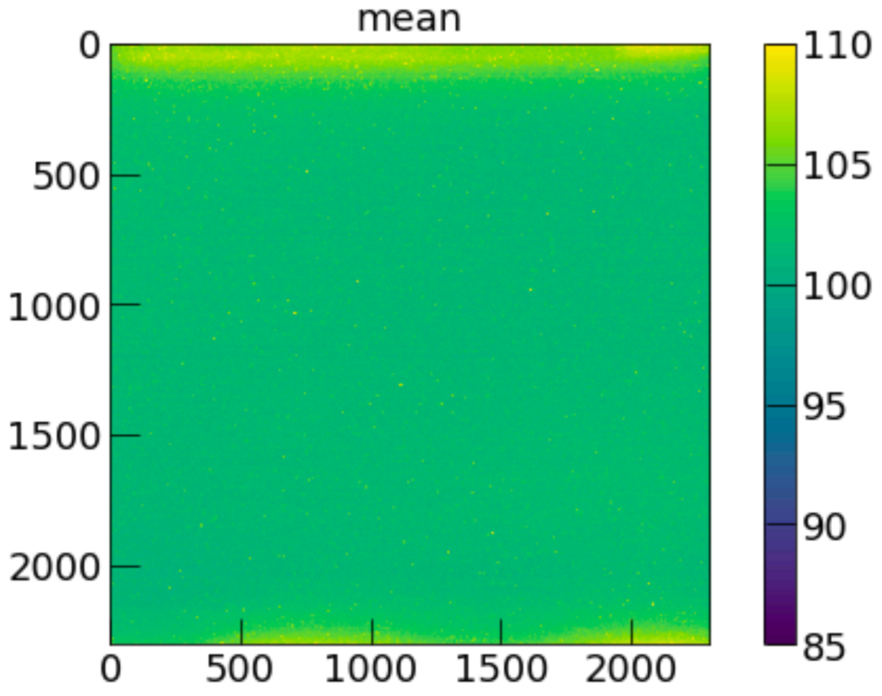


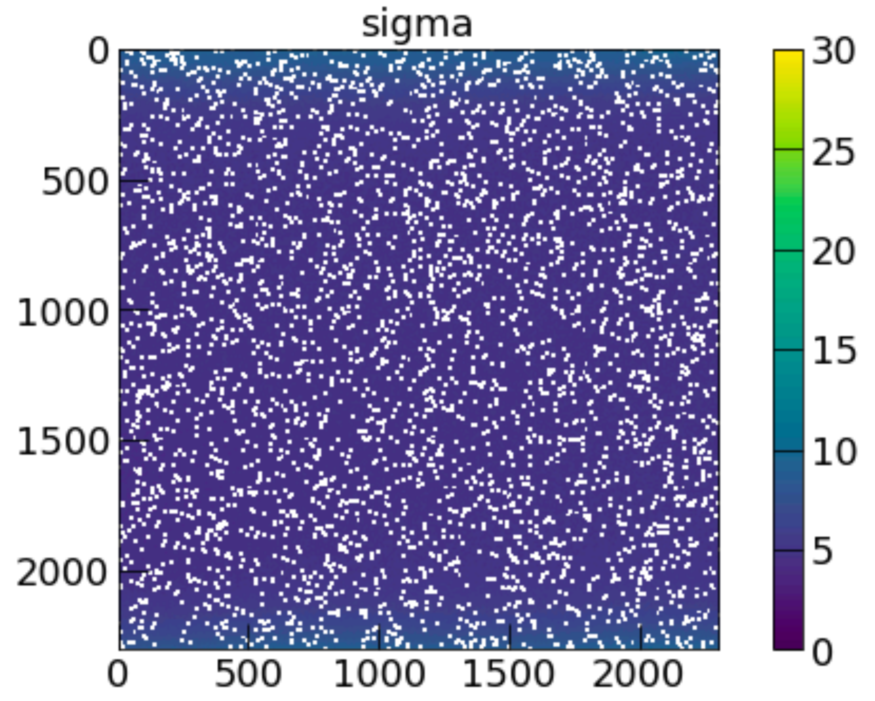
iDDBSCAN final output

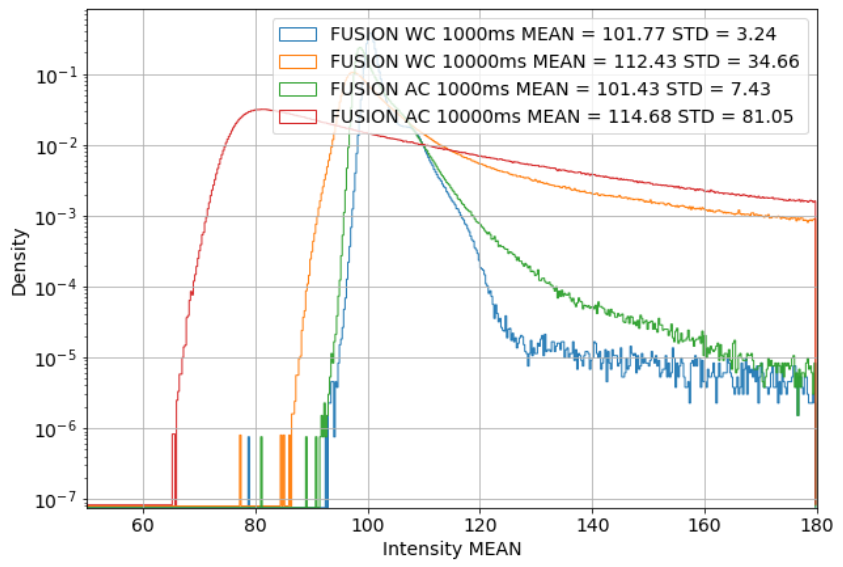
I. Pains et all. @ ufjf.br

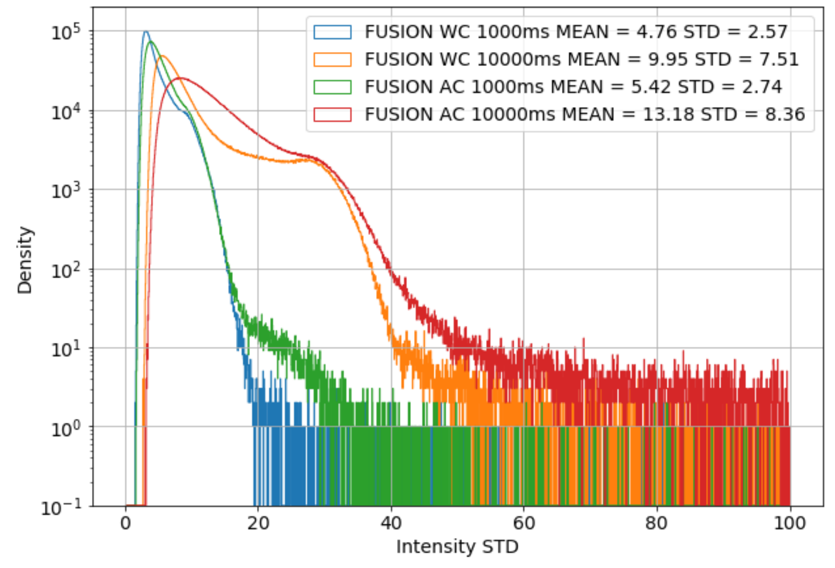
Fully Convolutional Network for pixel-wise selection

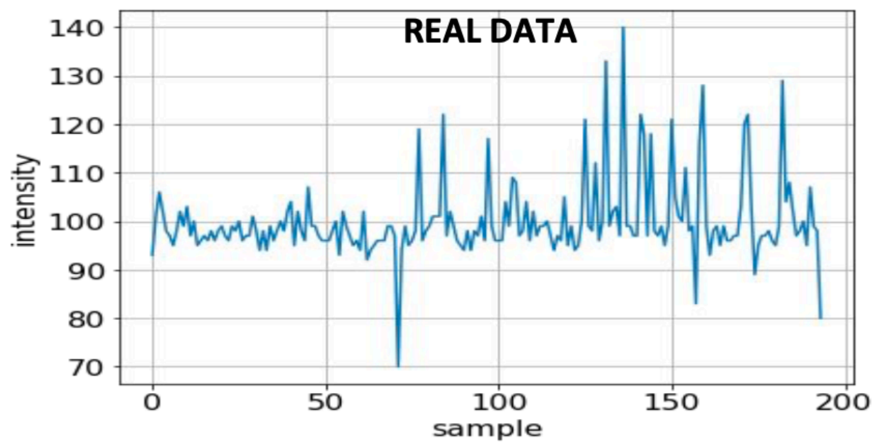
U-Net Fully Convolutional Network

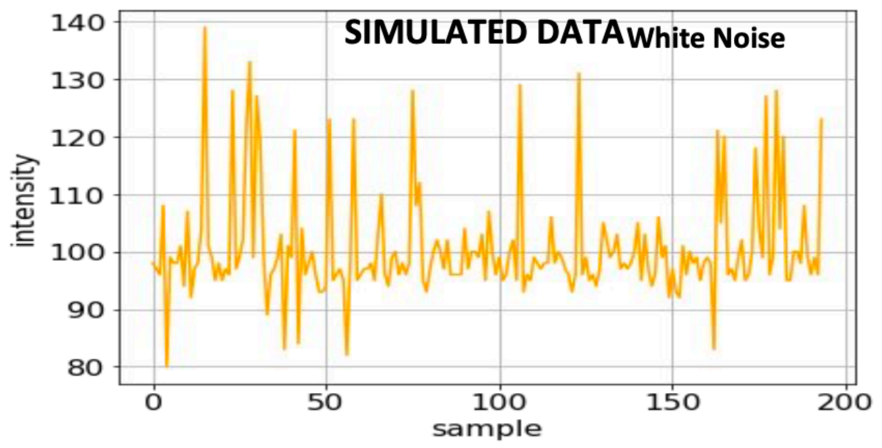






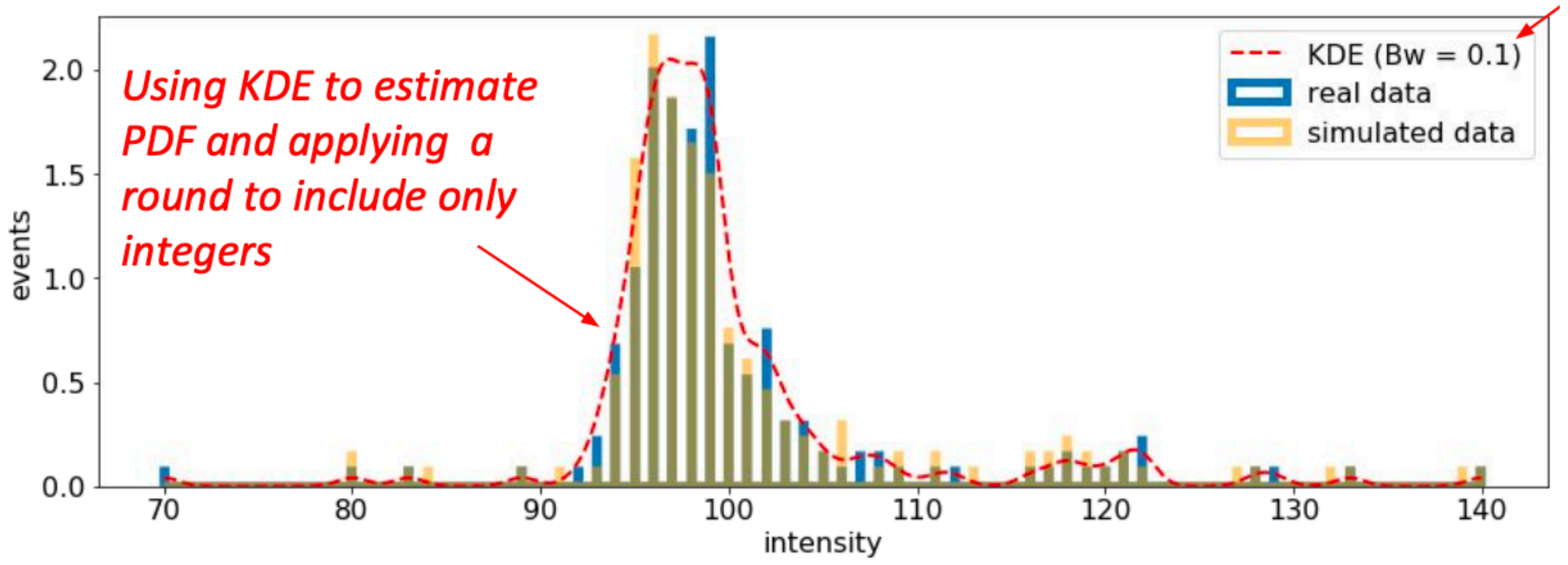






- handle with the sensor noise is not trivial
- Gaussian **thermal** noise (exposure)
- tail due to **telegraph** noise (random discrete fluctuations or switching events as a function of time) typical of CMOS devices
- in the real life GEM discharge, FC discharge + physics

noise is simulated considering the PDF function of each pixels of the cameras tested

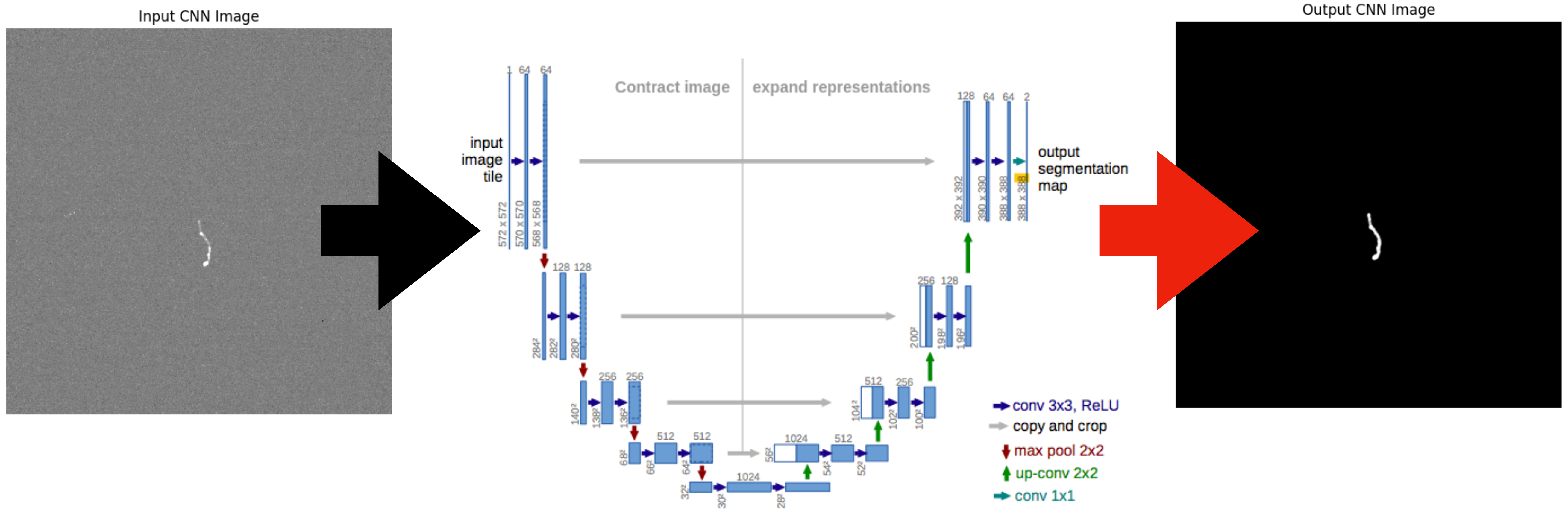


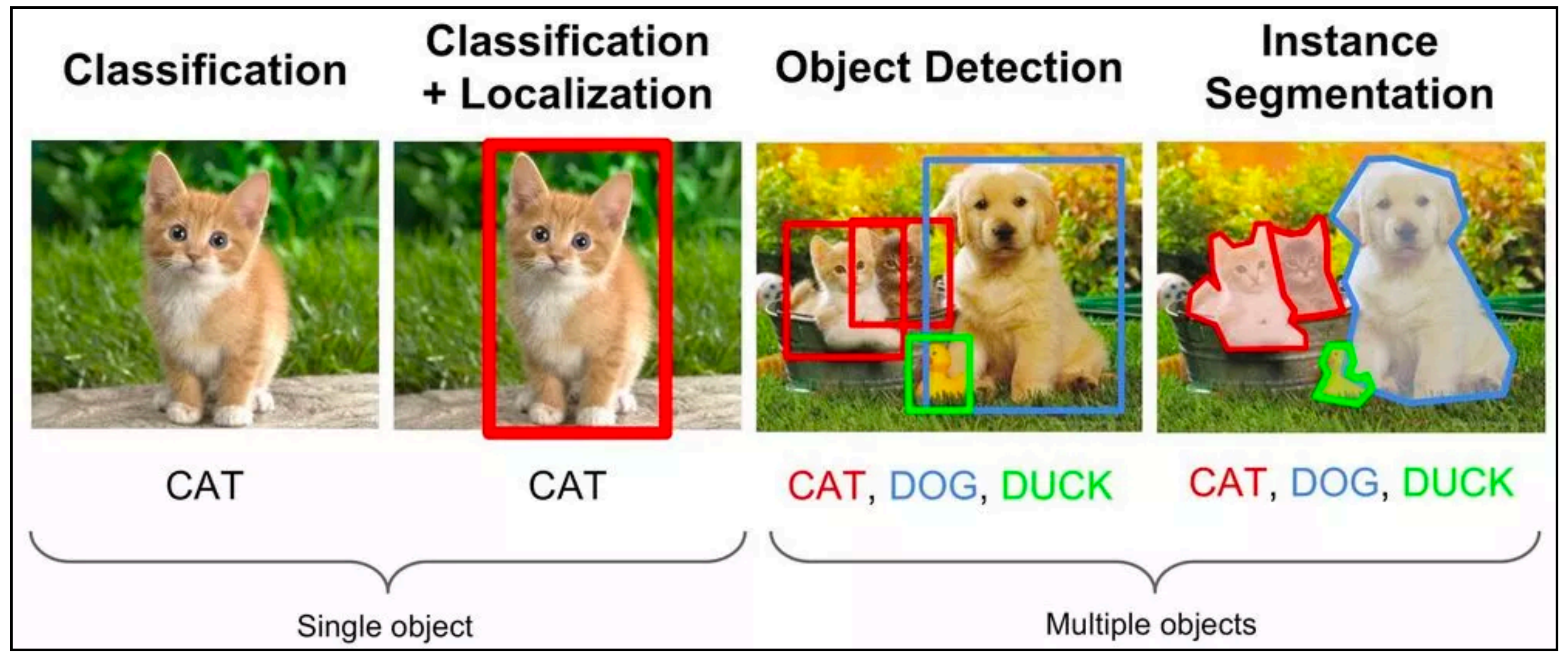
G .Lopes et al. @ .ufjf.br

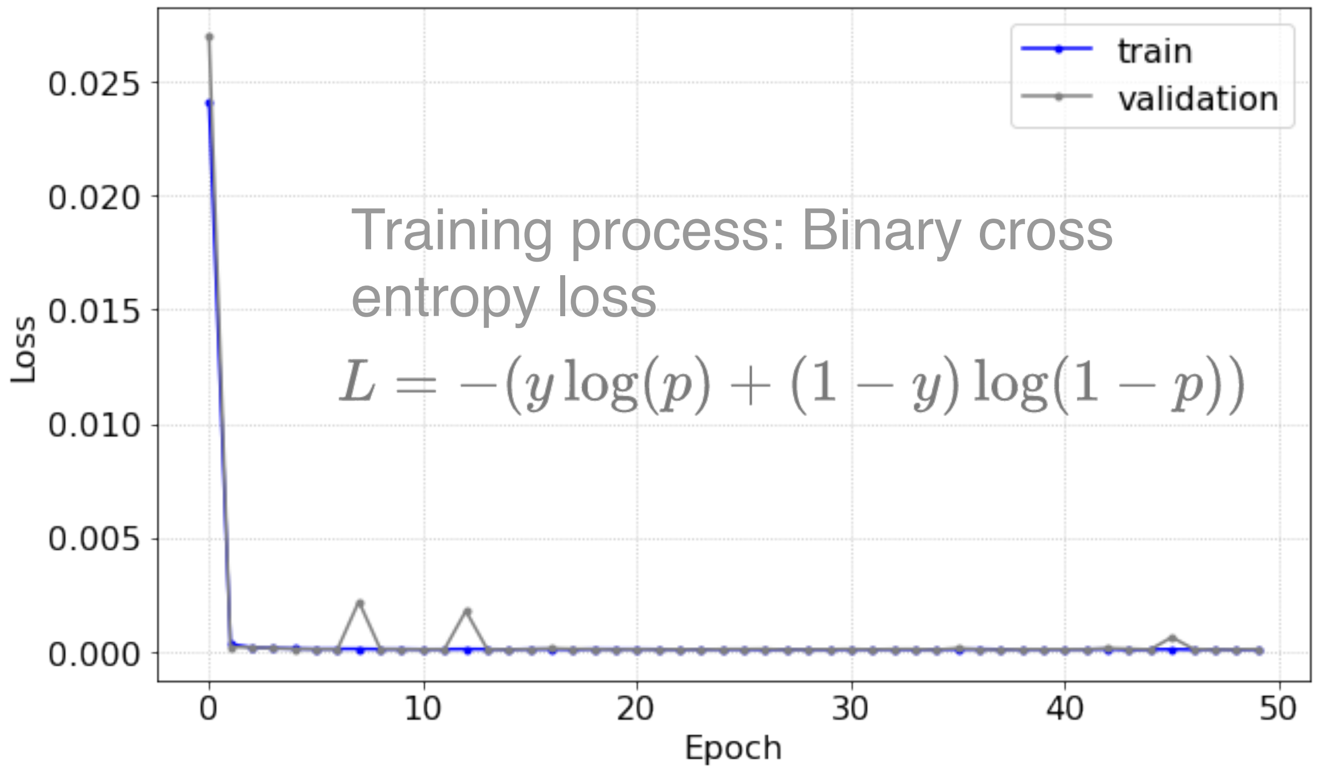
Fully Convolutional Network for pixel-wise selection

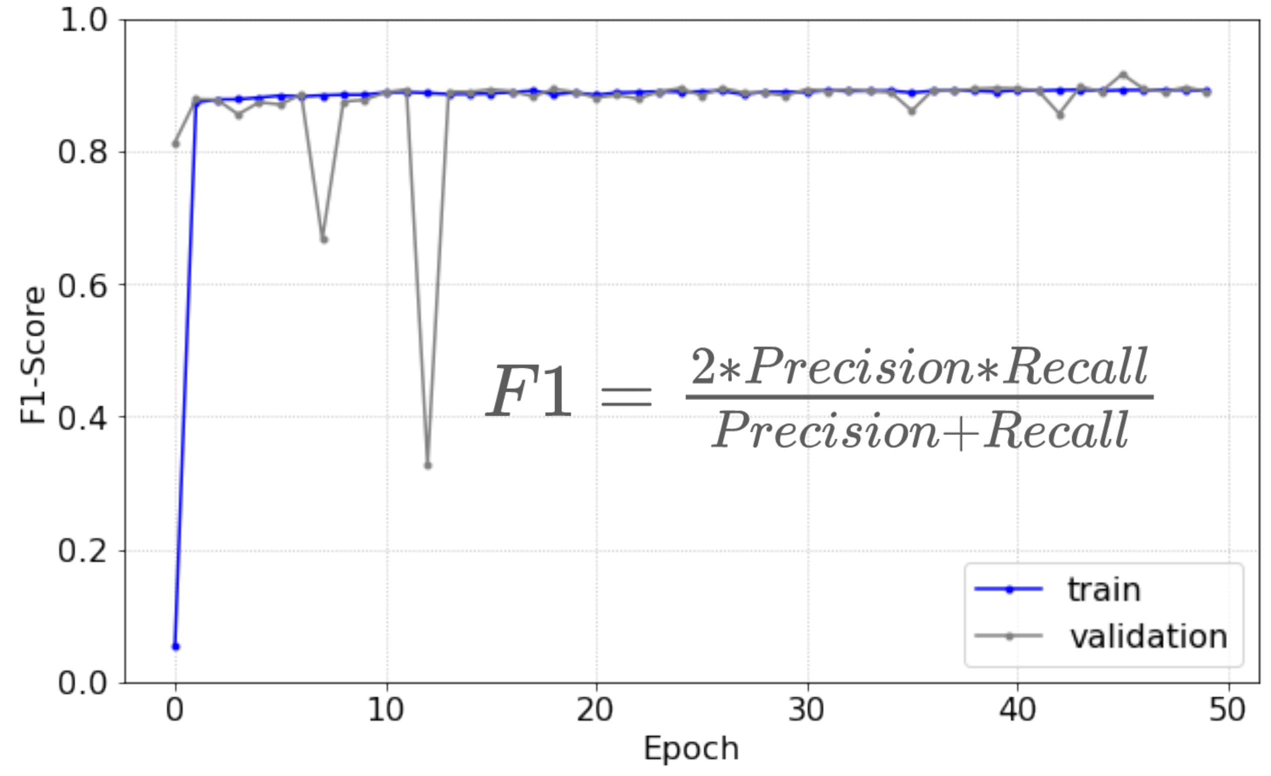
U-Net Fully Convolutional Network

- objective: increase signal pixels sent to reconstruction code, decreasing the number of background pixels, reducing the number of processed elements
- noise simulated by considering the density probability function of each pixel on camera.
- signal simulated with GEANT4 (for ER) or SRIM (for NR) 1-60keV
- each pixel is classified as signal or background
- 70% of samples were using for training







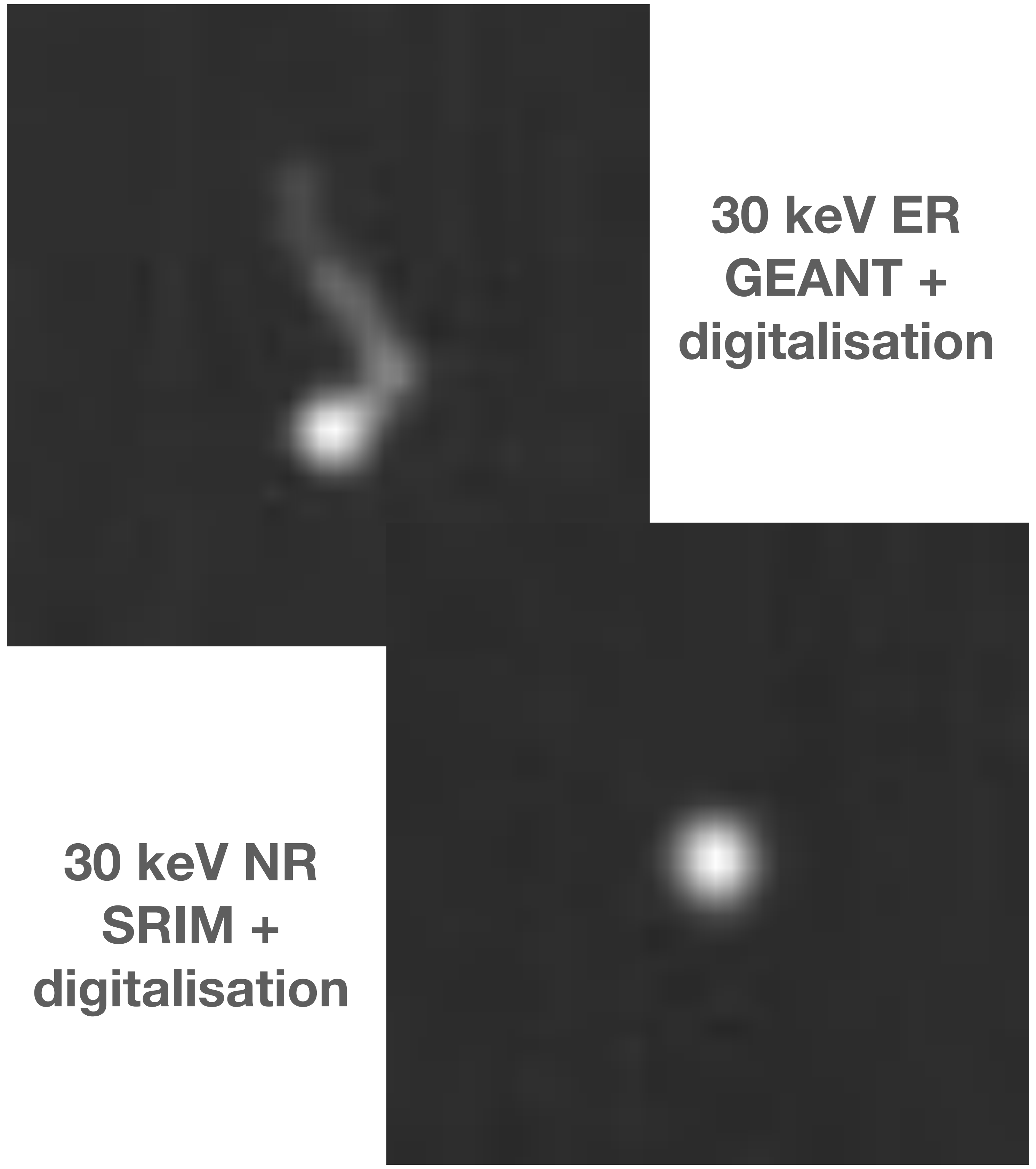


G .Lopes et al. @ .ufjf.br

deep learning models comparison

Deep Neural Network, Random Forest Classifier, Gradient Boosted Classifier

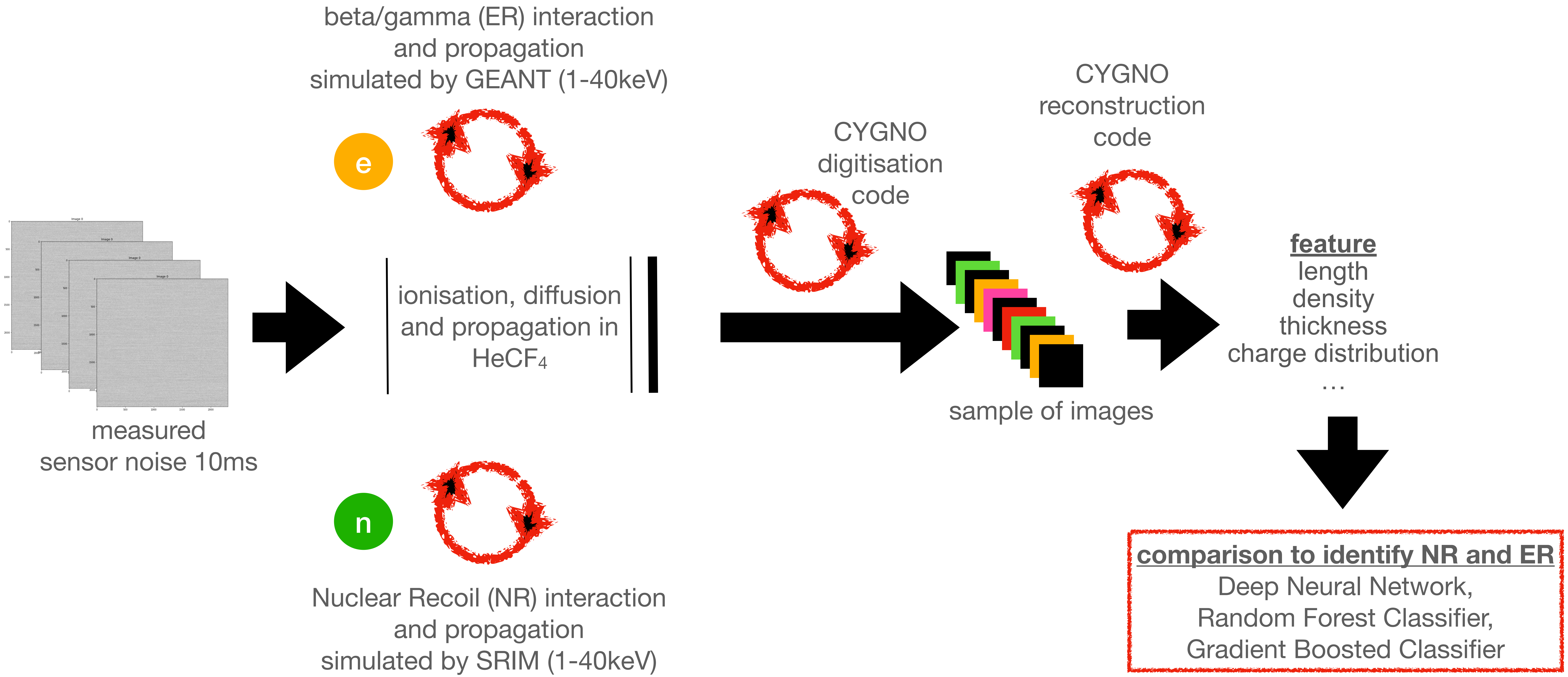
- **interaction** of the particles with gas is simulated using either GEANT4 (for ER) or SRIM (for NR)
- **detector/readout** effects are added to these track i.e. diffusion, camera noise, effective ionisation, gain fluctuations and geometrical acceptance etc.
- **digitised** images are reconstructed with a density based algorithm to find the cluster around the track.
- **topological** informations of reconstructed track can be used as discriminating variables (features)
- **features**: Length Along Principle Axis (LAPA), Maximum Density (MaxDen), Cylindrical Thickness (CylThick), Standard Deviation of Charge Distribution (SDCD), etc.
- the features were used for **training the networks**.



A. Prajapati and Gustavo V. Lopez. et al. @ GSSI.it

deep learning models comparison

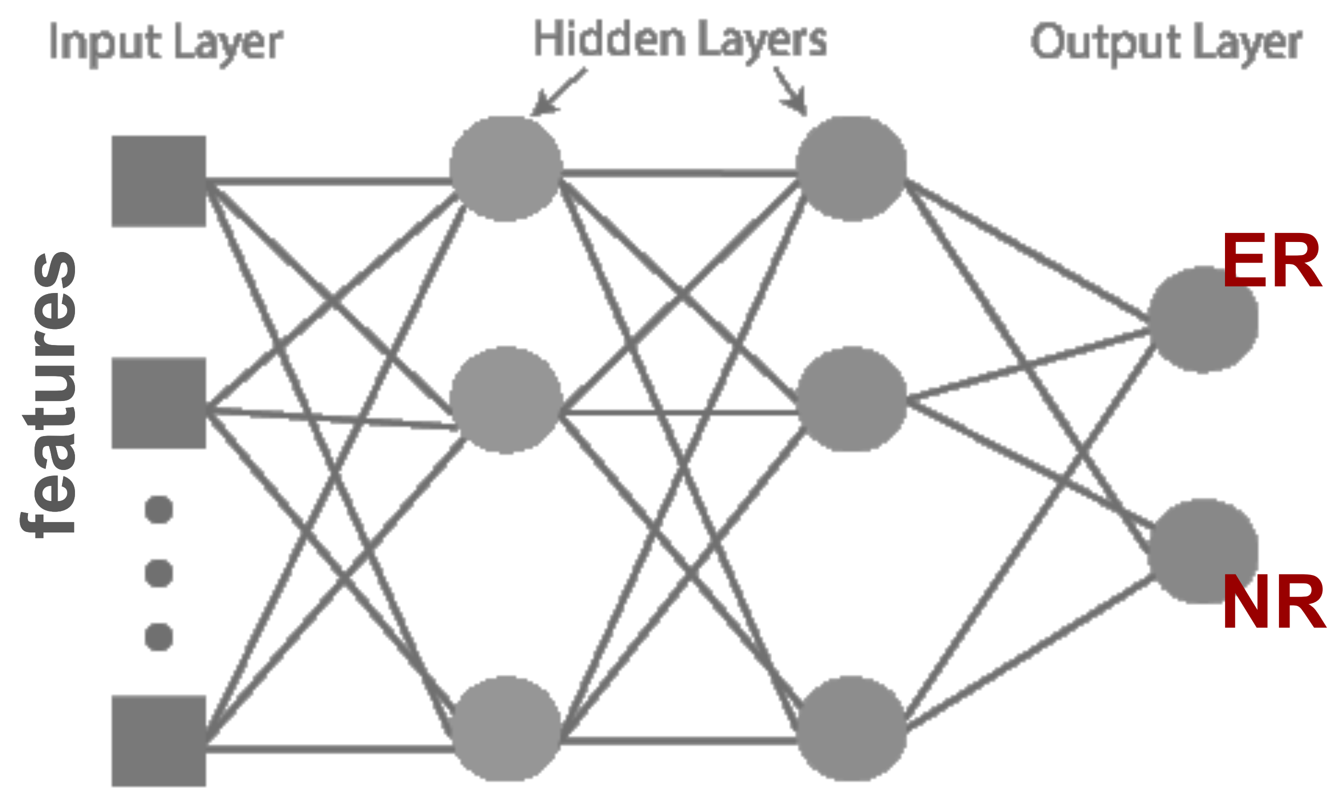
Deep Neural Network, Random Forest Classifier, Gradient Boosted Classifier



A. Prajapati and Gustavo V. Lopez. et al. @ GSSI.it

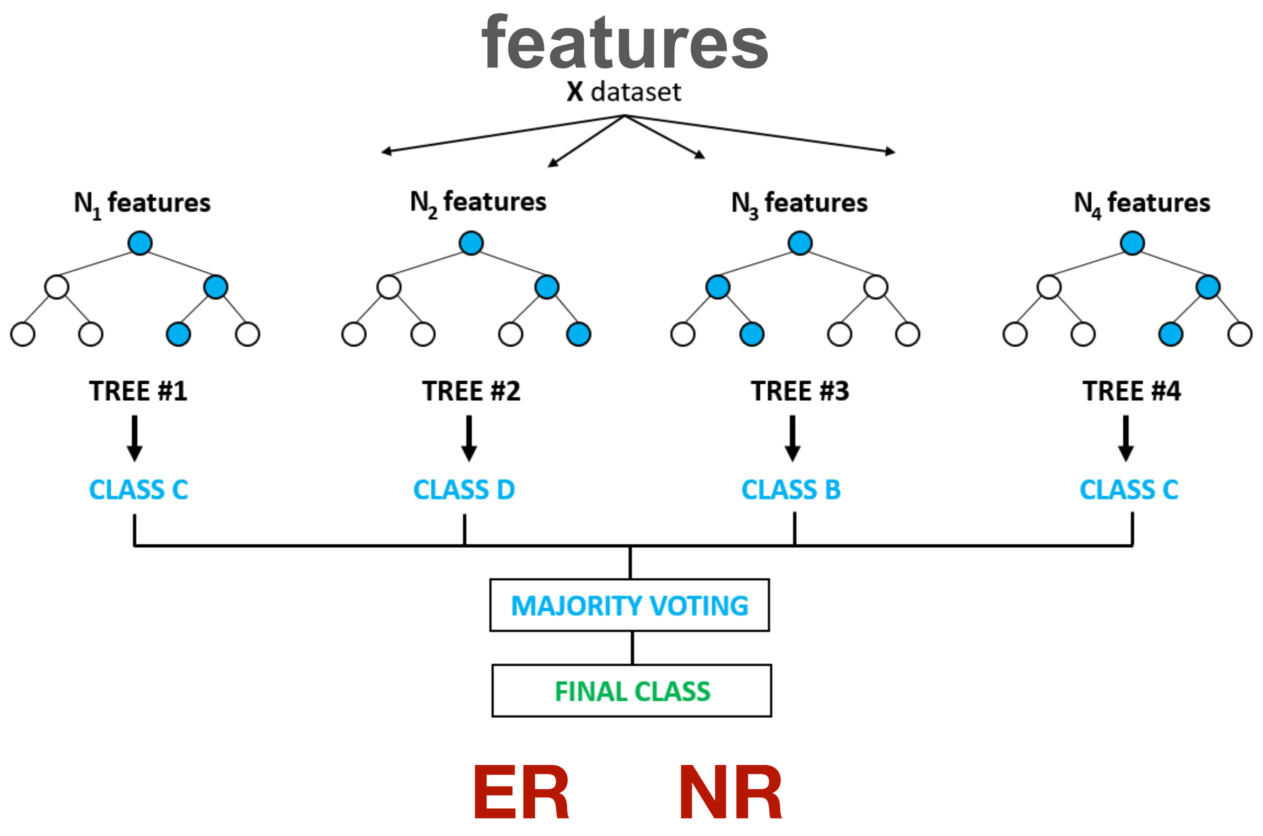
deep learning models compared

Deep Neural Network, Random Forest Classifier, Gradient Boosted Classifier



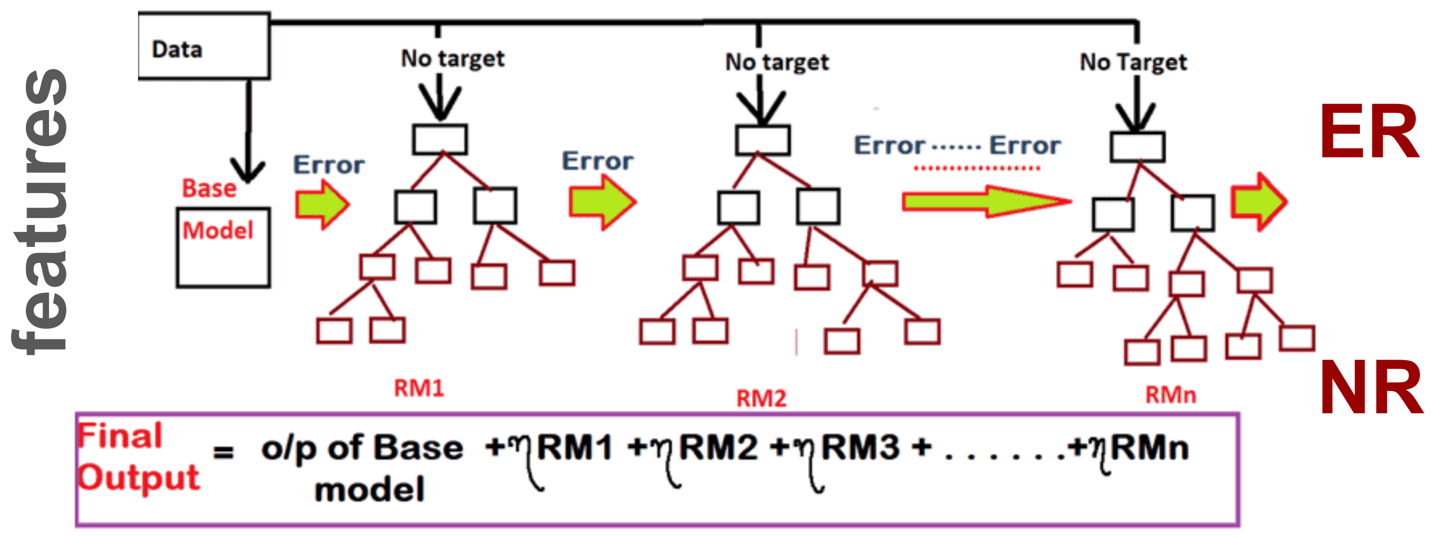
DNN

- the weights of the network are optimised iteratively
- result is the output of the last layer



RFC

- It can build each tree independently
- results are combined at the end of the process

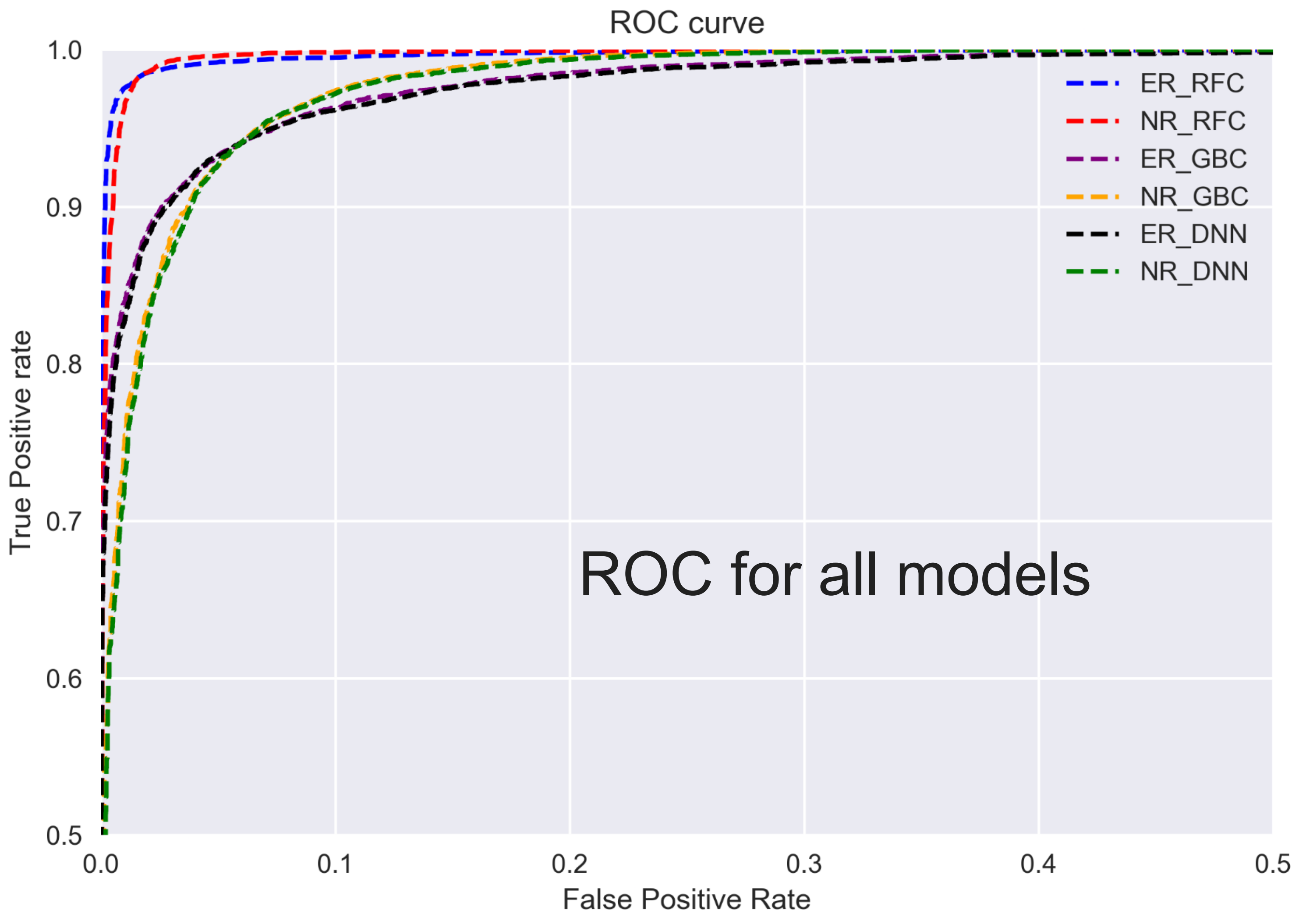


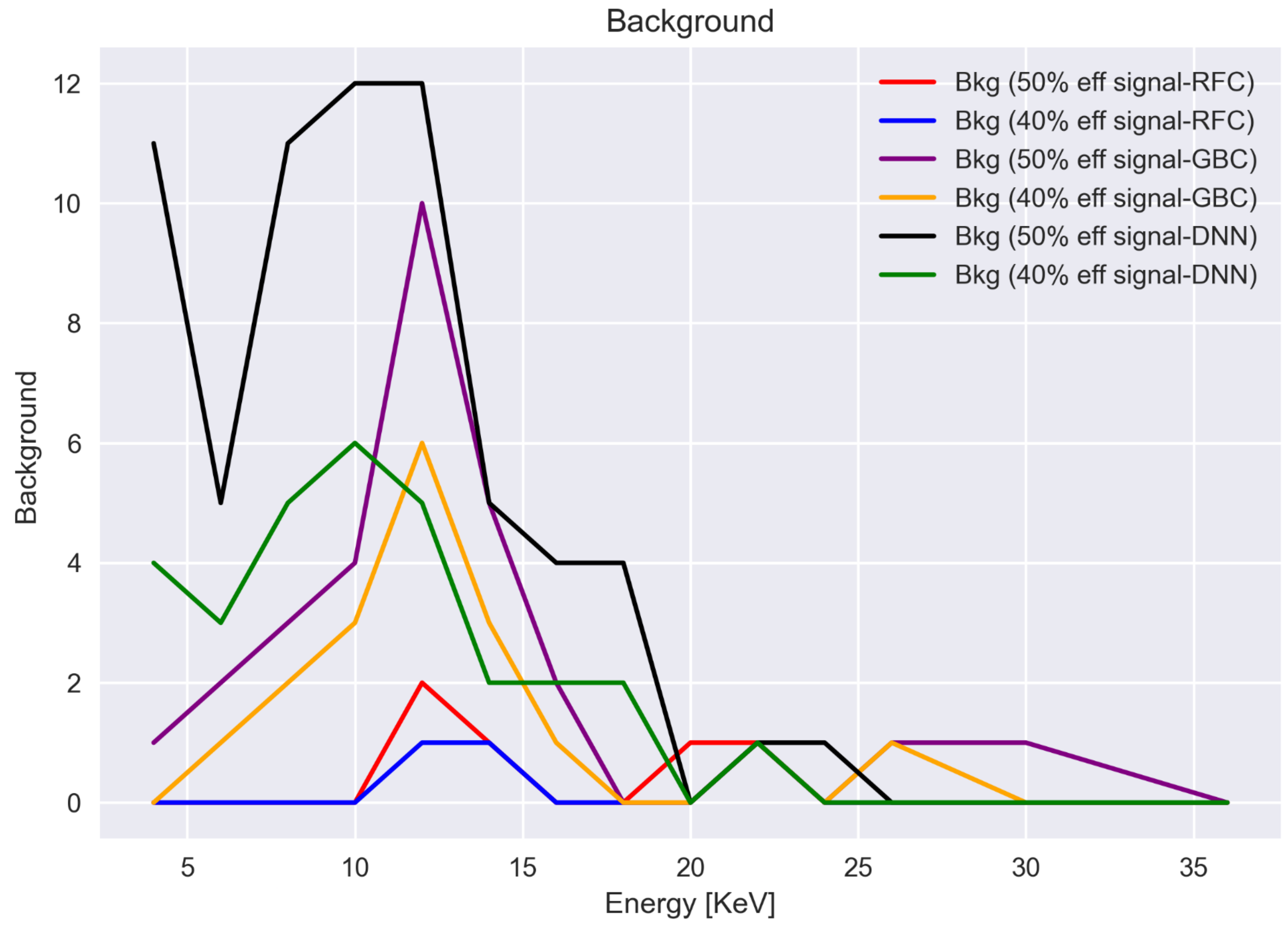
GBC

- It builds one tree at a time
- It combines results along the way

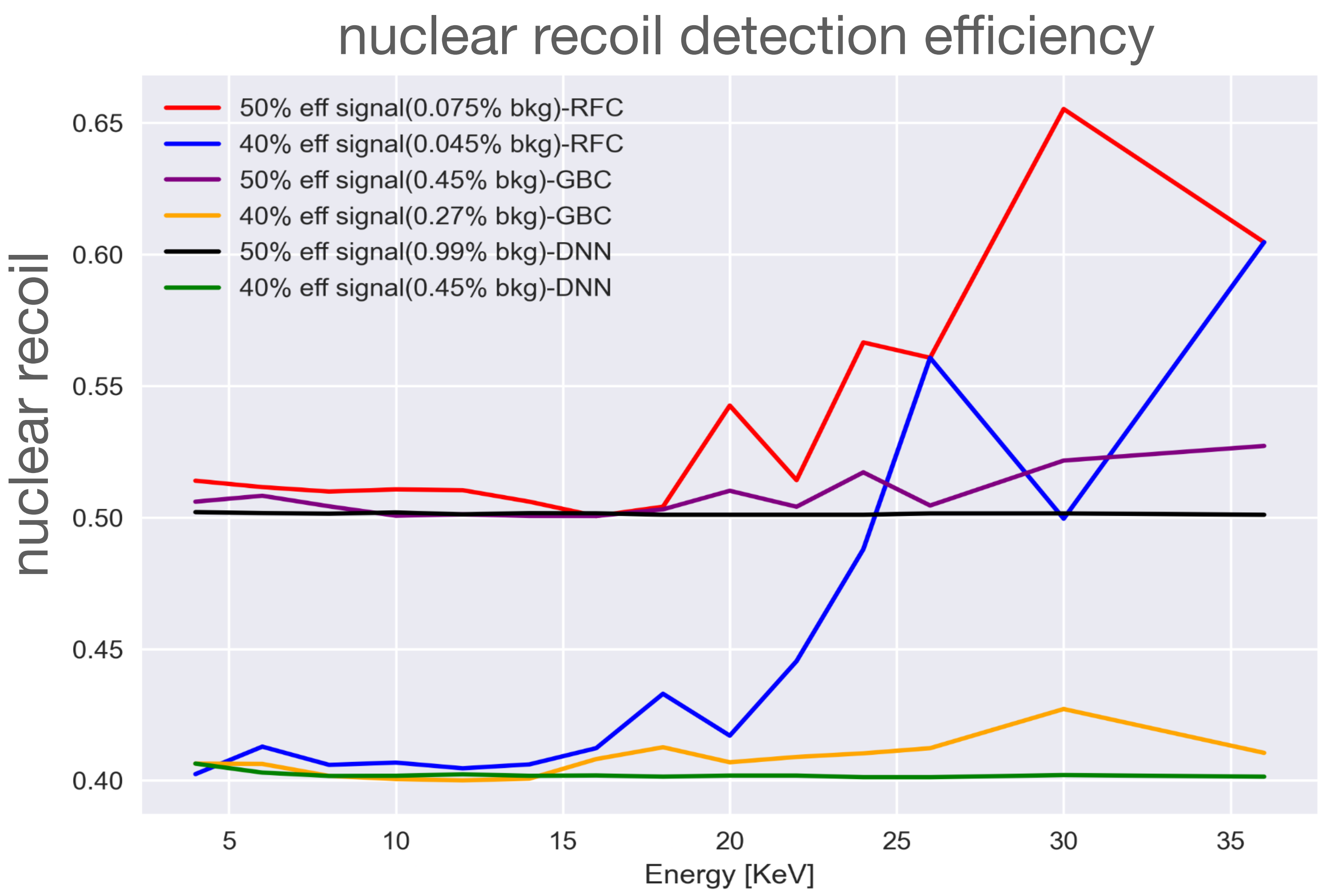
A. Prajapati and Gustavo V. Lopez. et al. @ GSSI.it

model comparison





background: ER classified as NR in each energy bin



Preliminary

models	NR det eff. [%]	background eff. [%]
RFC	50	0.075
	40	0.045
GBC	50	0.45
	40	0.27
DNN	50	0.99
	40	0.45
Traditional Approach	50	3.5
	40	0.8

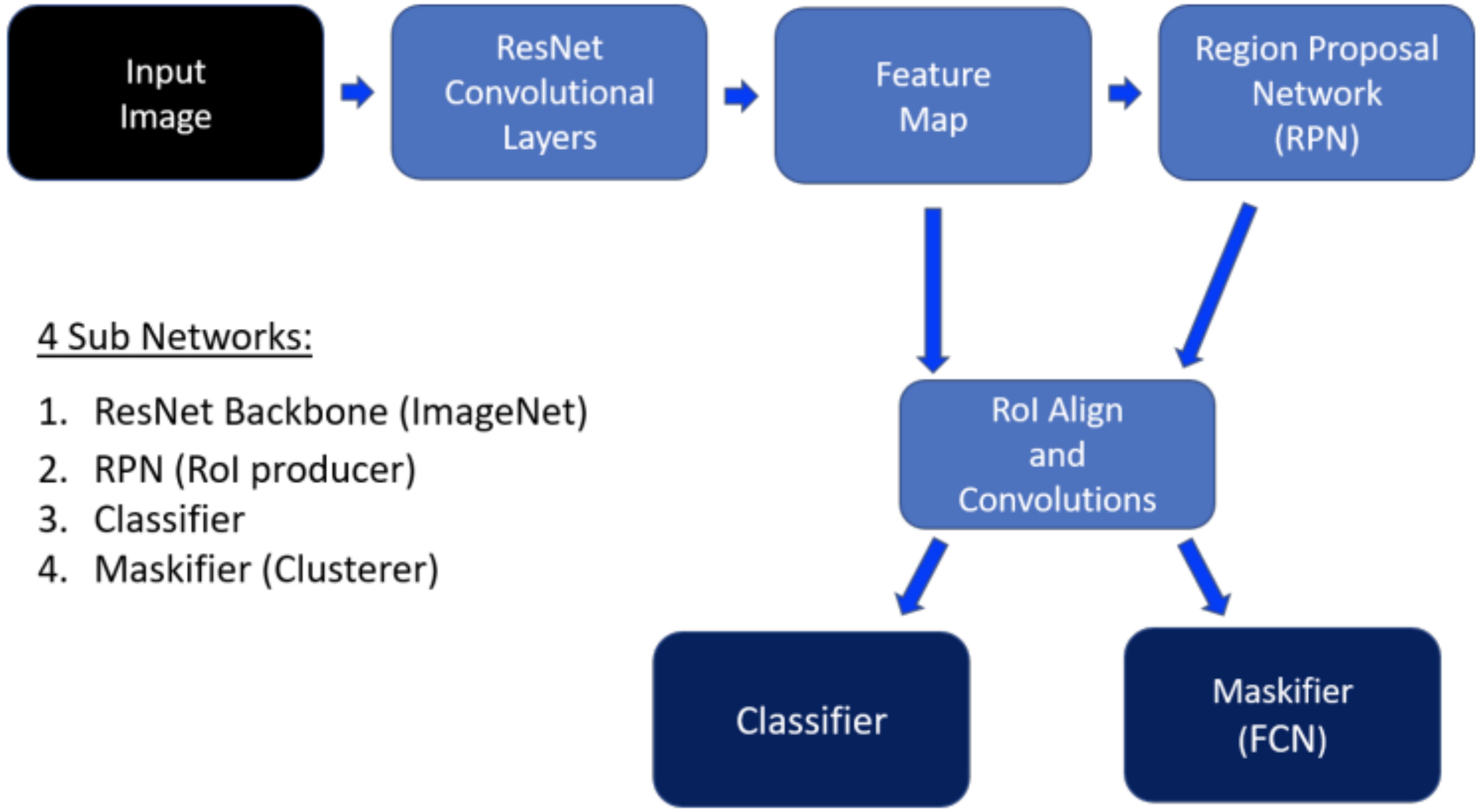
A. Prajapati and Gustavo V. Lopez. et al. @ GSSI.it

minimum NR efficiency in each energy bin is the signal efficiency and background (ER) efficiency is the overall background classified as signal in all energy bins. traditional approach: simple cut on the variable.

next steps 1/2

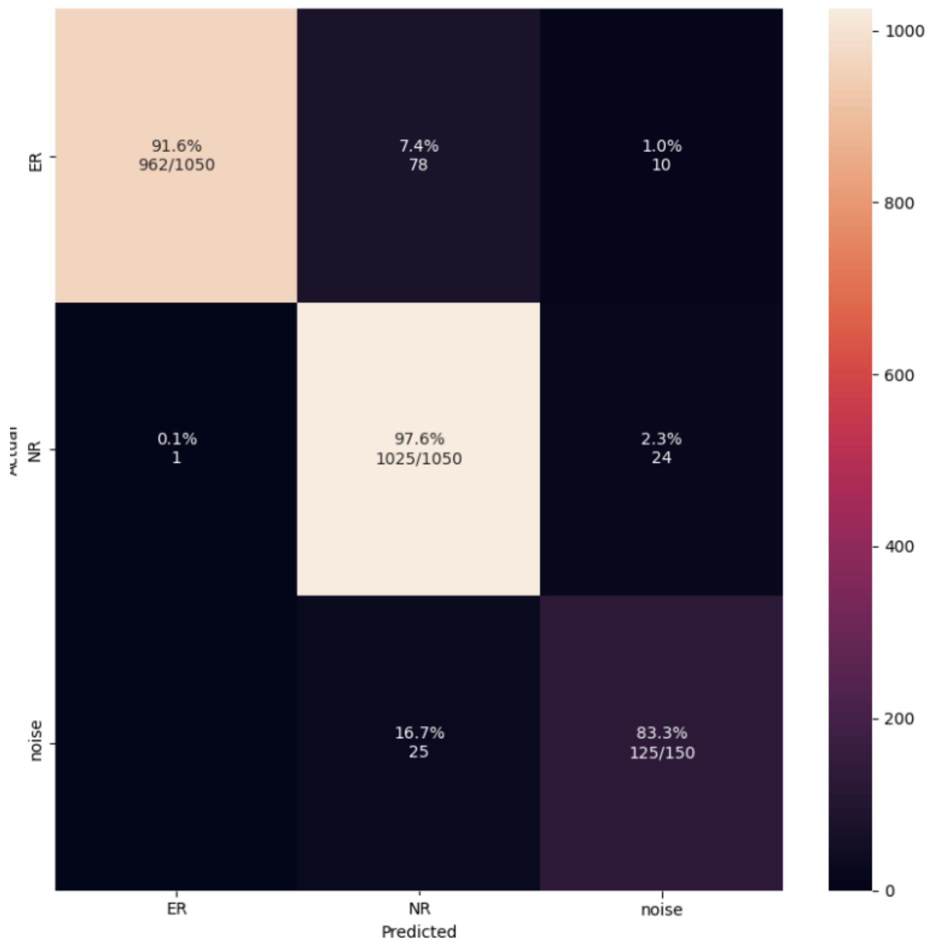
Convolutional Neural Network

- operate on real data with CNN for the **reconstruction, classification** and computing physical quantities associated to the track.
- A ResNet model and a CNN classifier is already been made and tested on simulated data.

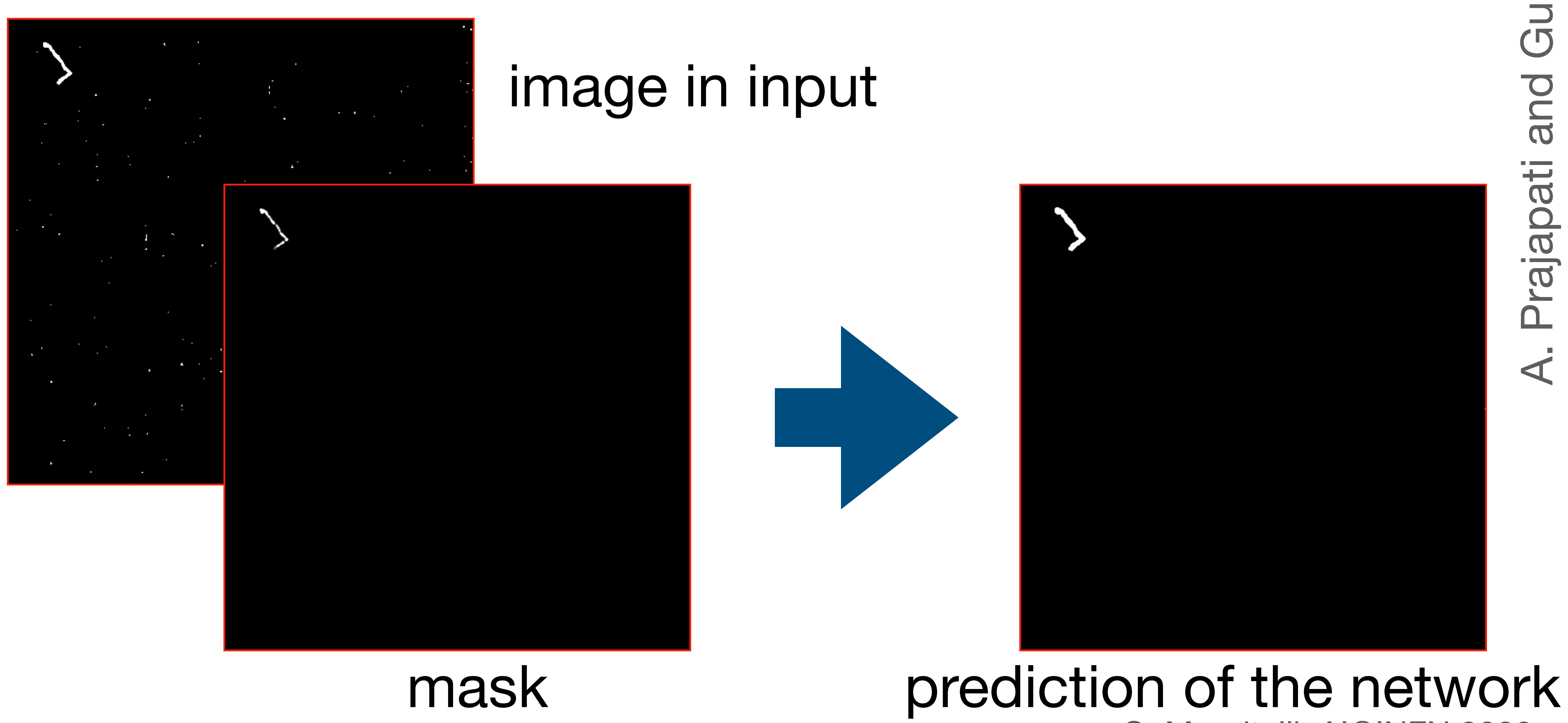


<u>results of CNN are very similar to the one of DNN</u>

<u>ResNet CNN model (60 keV ER)</u>



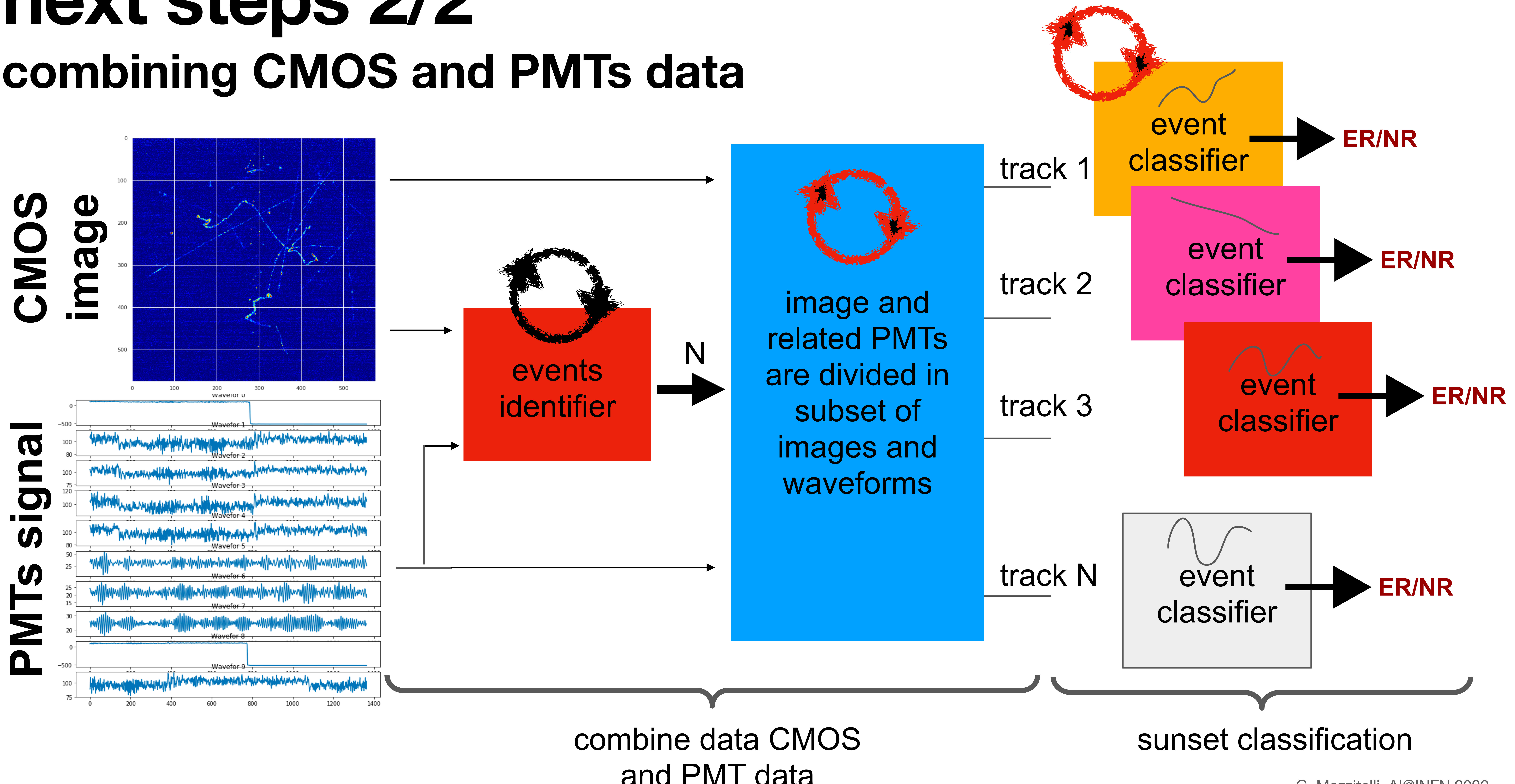
ER classification
92.3% DNN ->91.6% CNN
NR classification
95.9% DNN -> 97.6% CNN



A. Prajapati and Gustavo V. Lopez. et al. @ GSSI.it

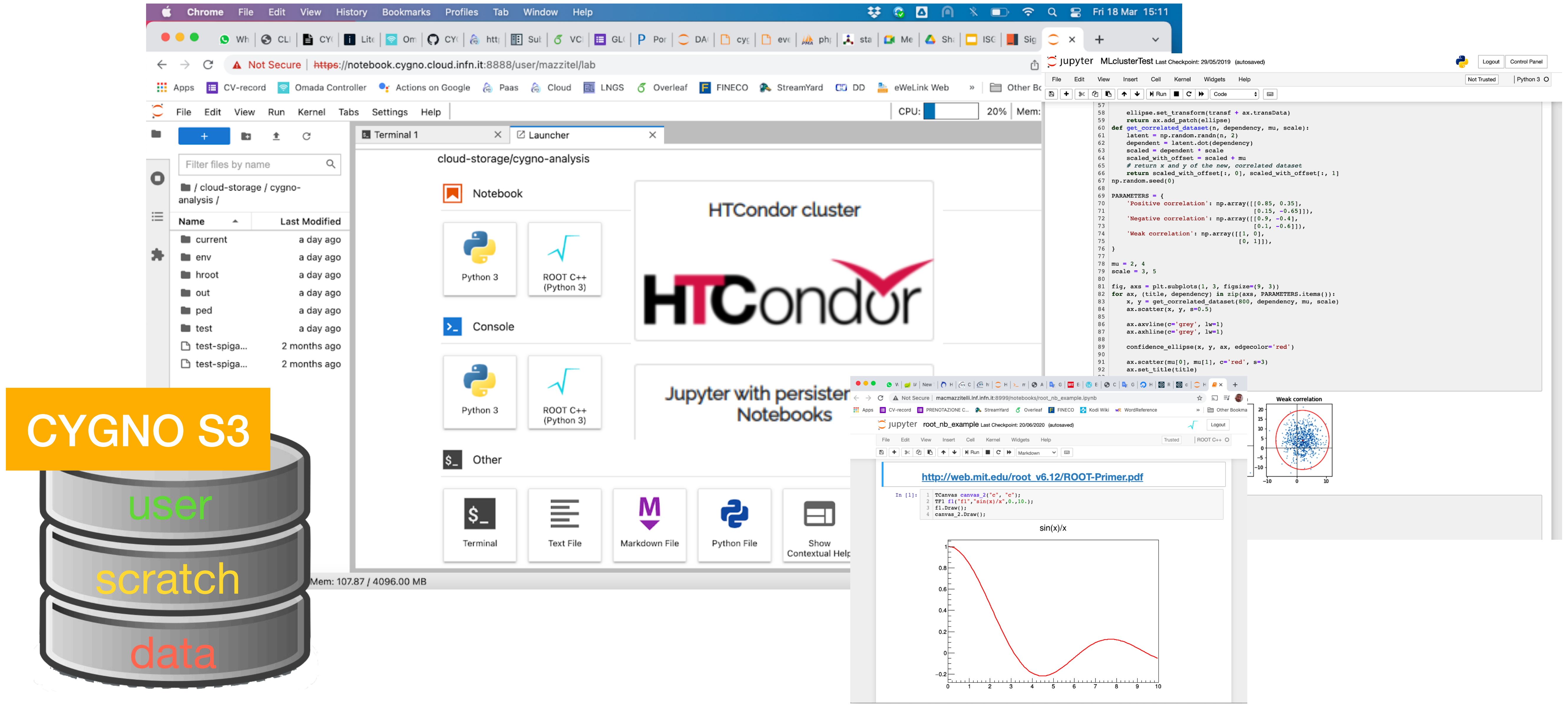
next steps 2/2

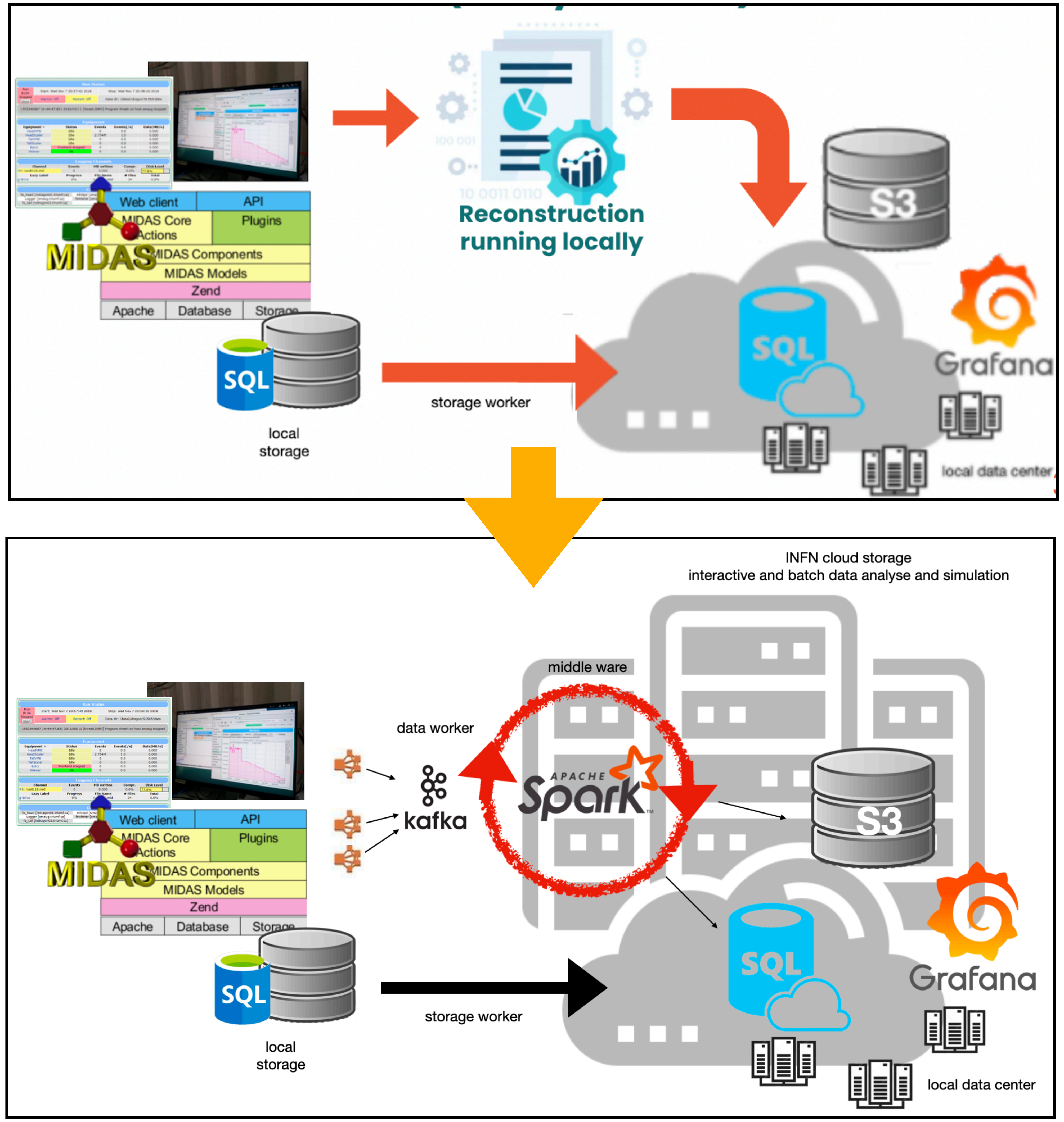
combining CMOS and PMTs data



infrastructure - schematic view

CLOUD INFN use-case





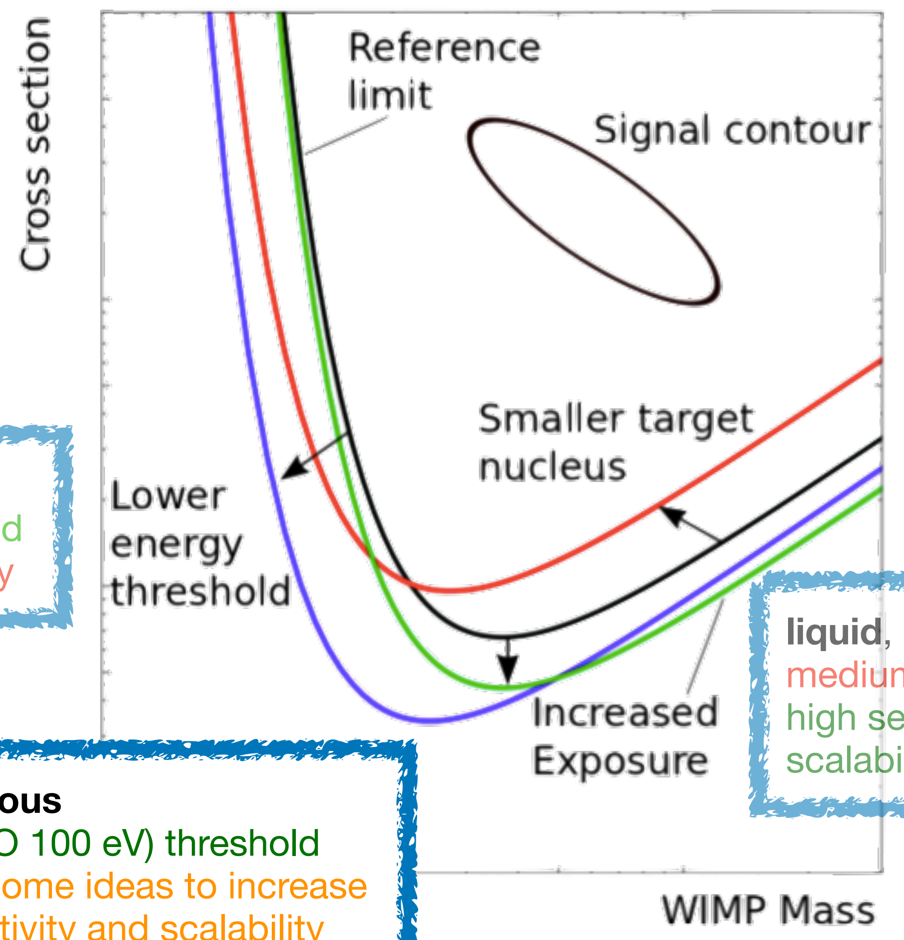
Conclusion

actually a work in progress…

- our experiment is probably an ideal use-case where **machine learning methods** can be **apply, tested and skills can grow up.**
- up to now we have played with our code in a “**gym**” where the environment is very different from the one where the detector is now installed and is going to start taking more realistic data
- this allow us to **develop code** able to deal with very **hard background condition** and trained us to start deal with unsupervised and supervise machine learning methods
- At the LNGS Laboratory the bet will be to look for **sub millimetric** tracks of NR respect to the most luckily **background** of the ER (beta/gamma) to find the ultimate operational **energy and directional threshold**

gas detector for DM & neutrino

“ionisation” detectors



solid, cryogenic
very low (O 10 eV) threshold
limited mass and scalability

gassous
low (O 100 eV) threshold
just some ideas to increase sensitivity and scalability

liquid, cryogenic
medium (O 1000 eV) threshold
high sensibility and scalability

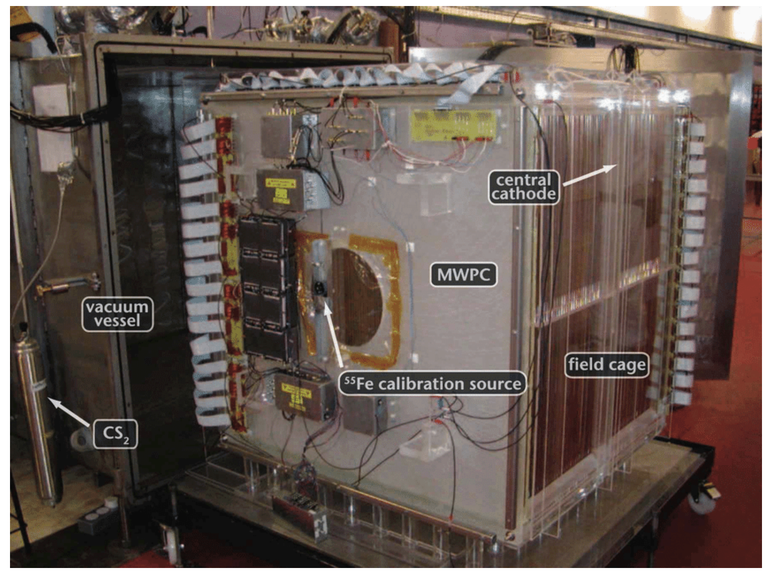
detector gas pressure

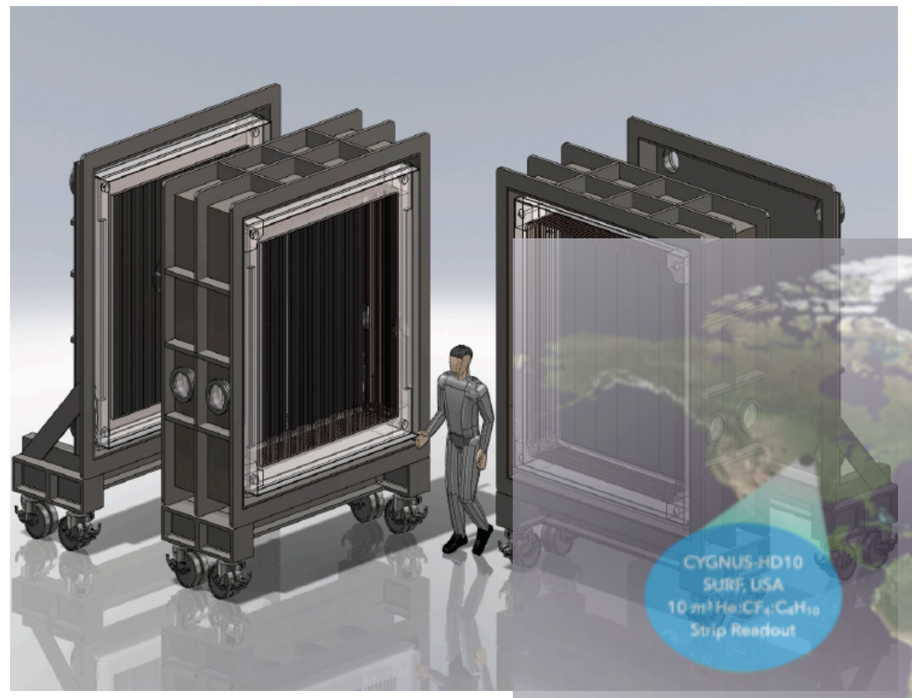
low pressure

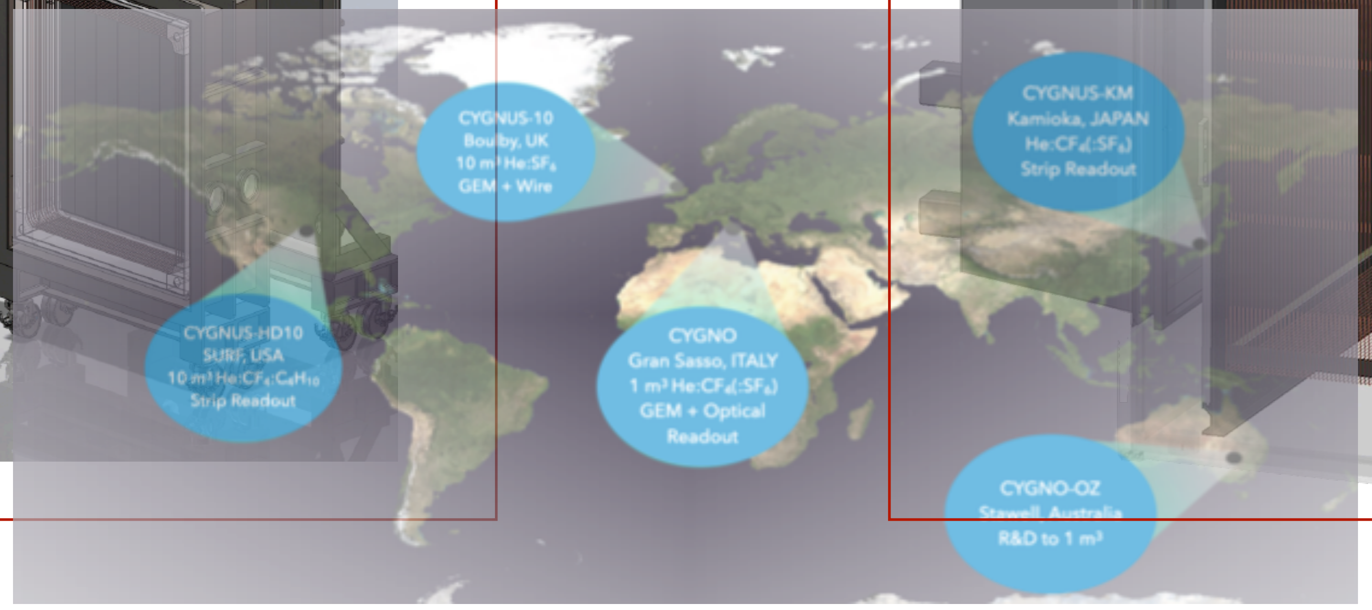
atmospheric pressure

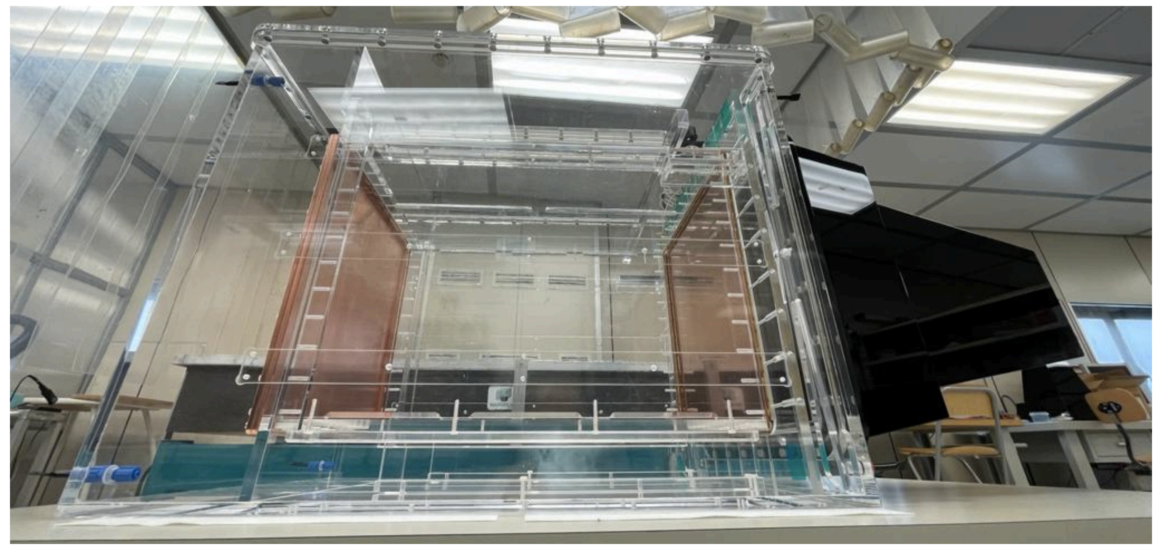
high pressure

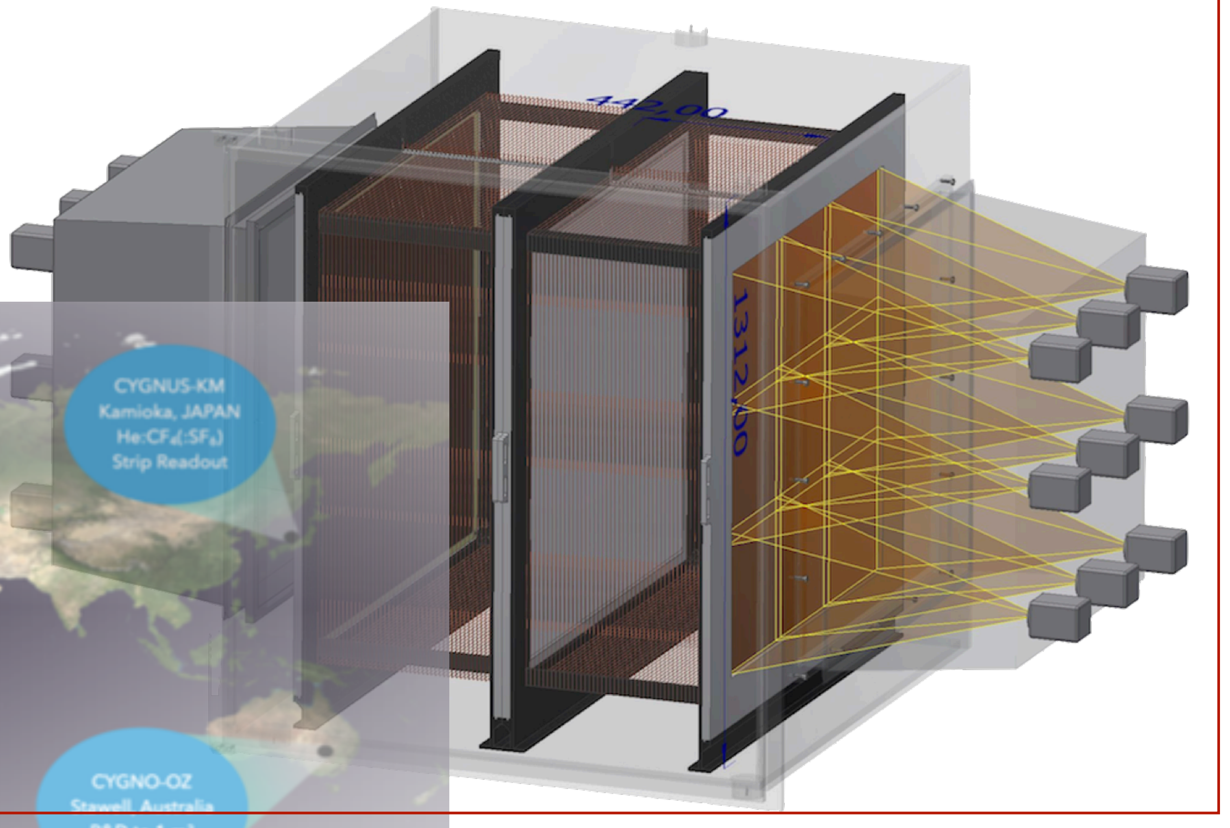
DRIFT
(negative Ion)
DMTCP
OPT with CCD



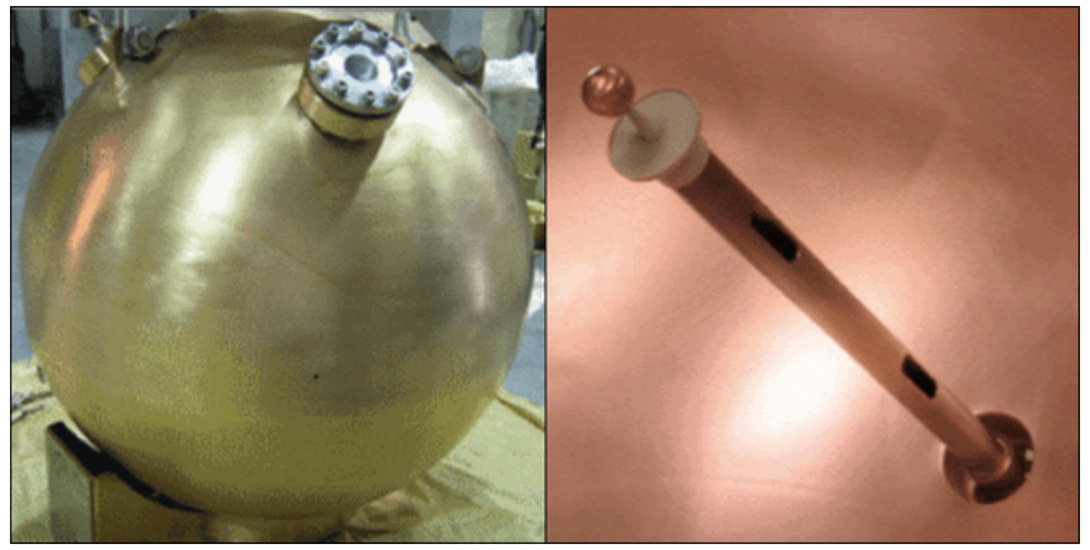




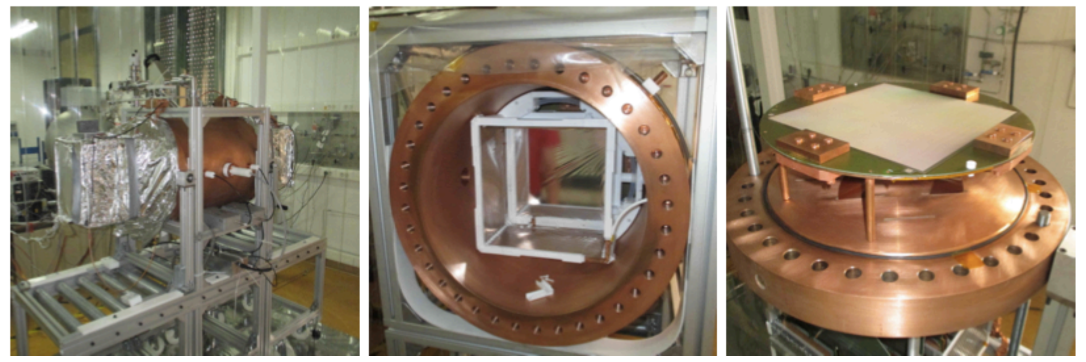




CYGNO/INITIUM



NEWS-G
(SNOLAB)
Ne:CH4



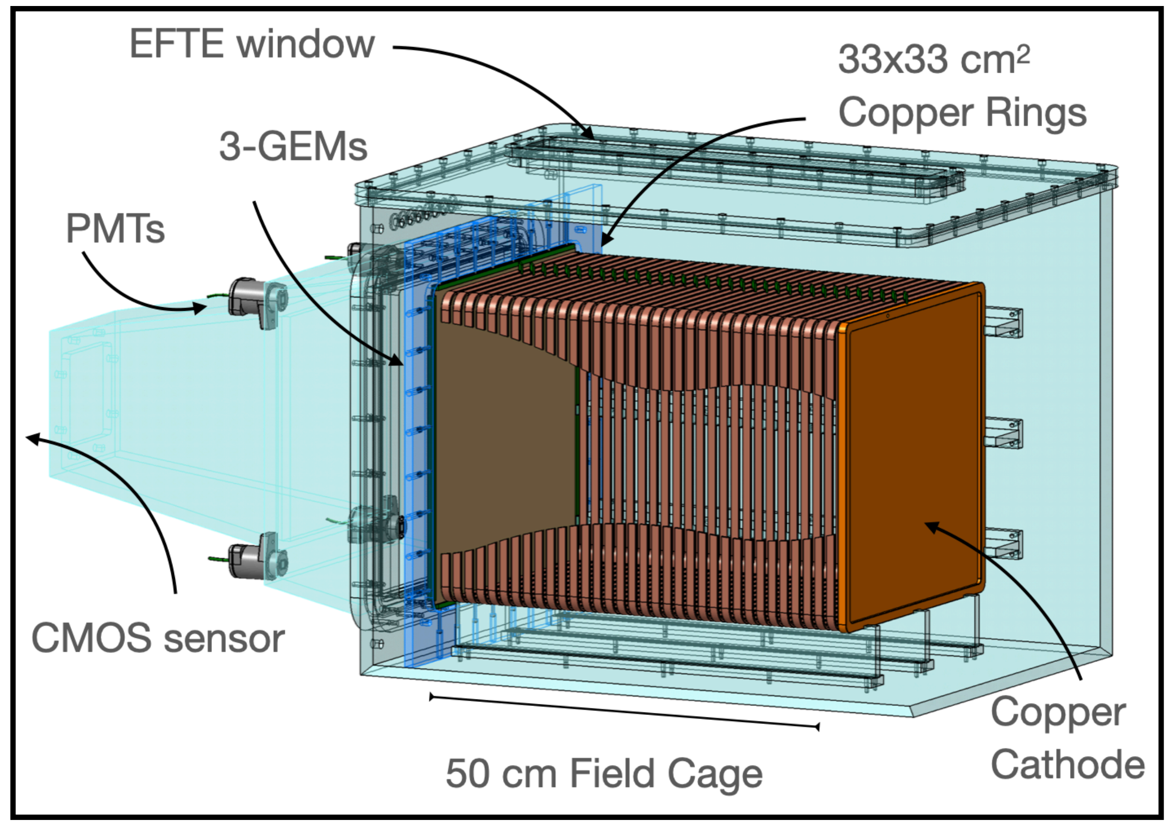
TREX-DM
(LSC)
Ne

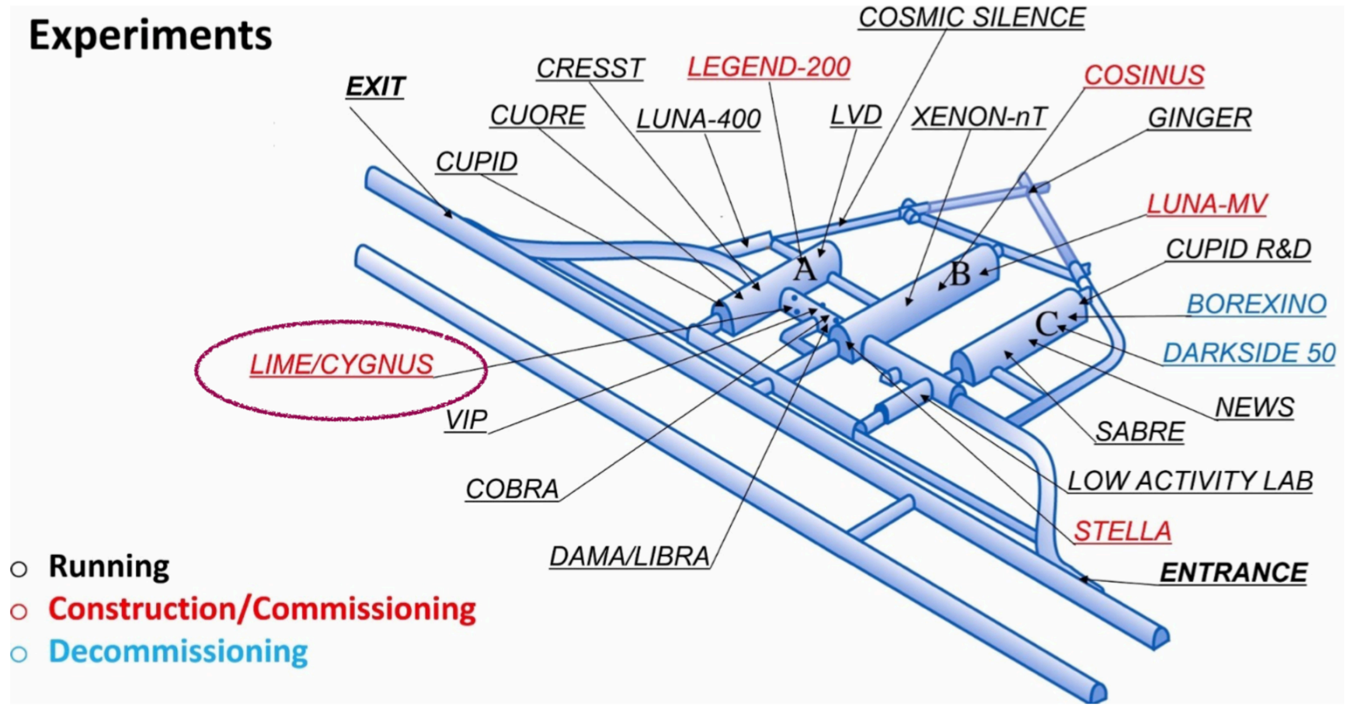
LIME prototype

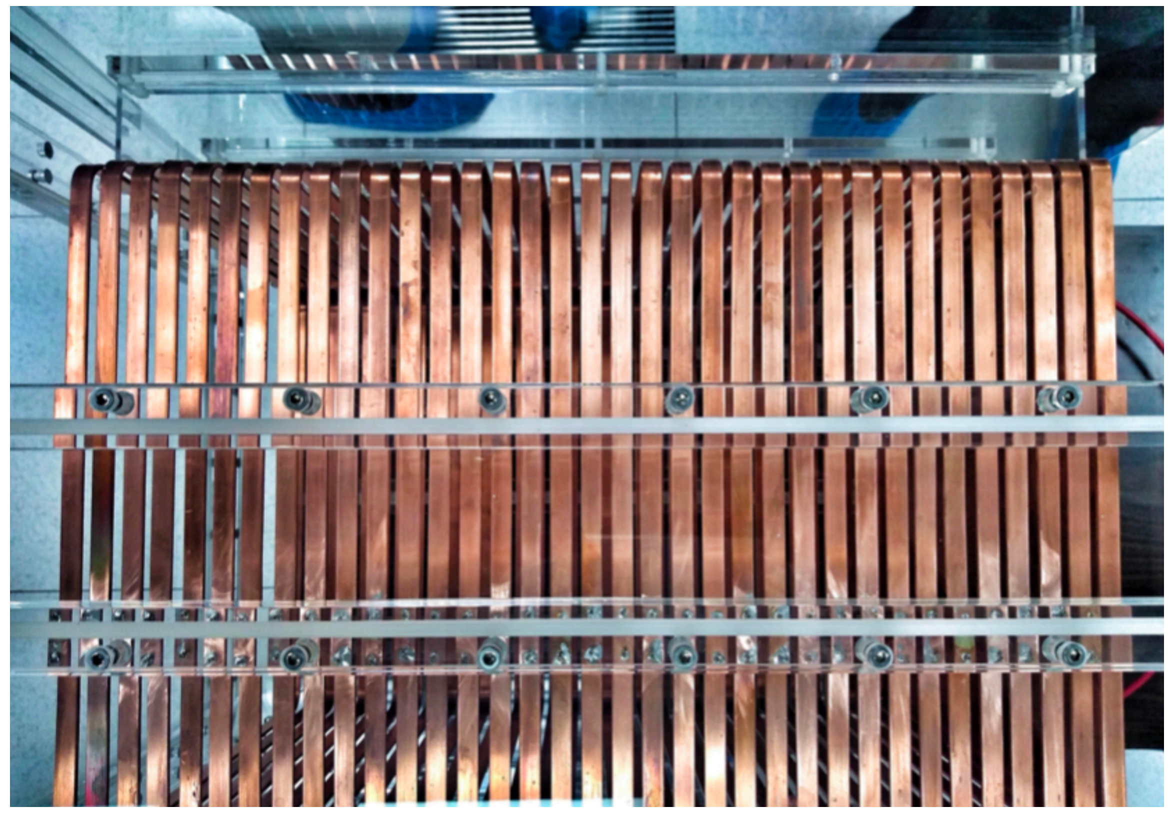
design

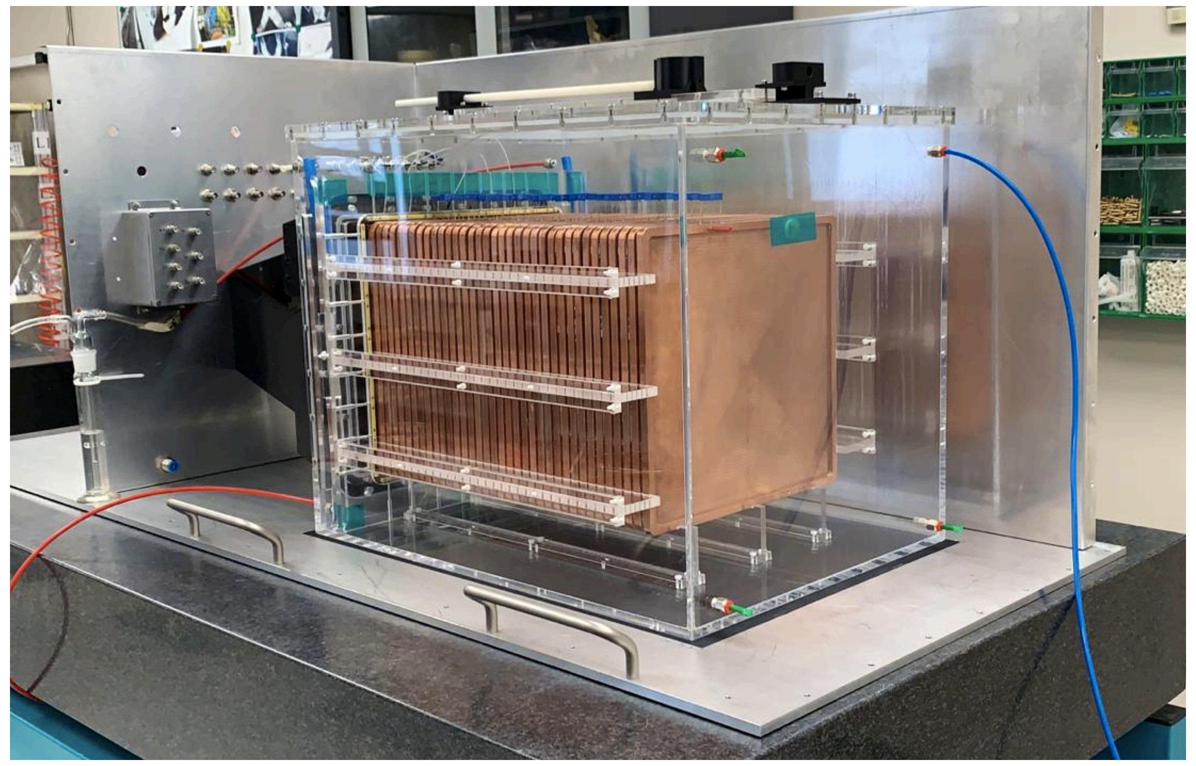
overground/underground first phase:

- **50 cm drift** made of **Cu ring** 33*33 cm
- **50kV Cu** cathode (up to 1kV/cm)
- triple **GEM** stack amplification stage
- low radioactivity PMMA **vessel**
- 50 litres sensitive volume with an **He/CF4** based mixture at **atmospheric pressure**
- a single sCMOS Active Pixel Sensors (APS) HAMAMATSU **camera** + Schneider commercial **optics**
- 4 **PMT** symmetrically placed around the sensor for **time** shape
- **Aluminum** faraday cage



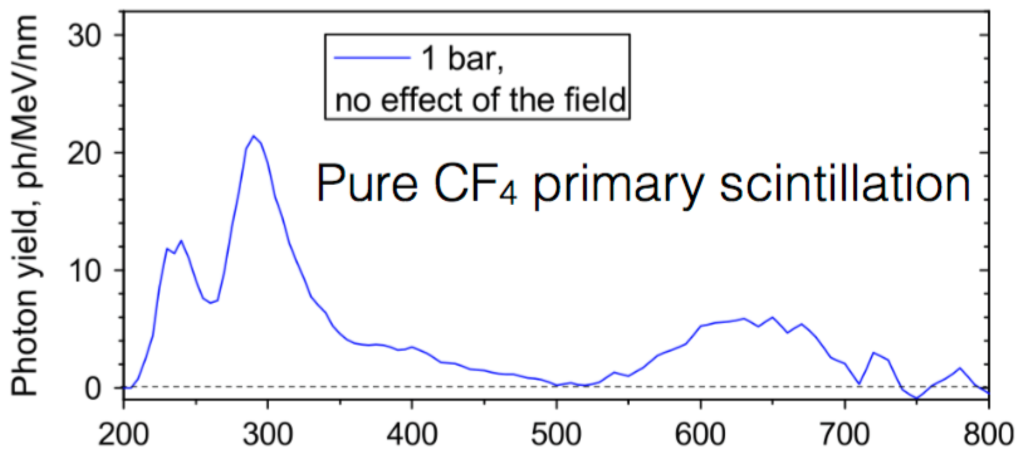


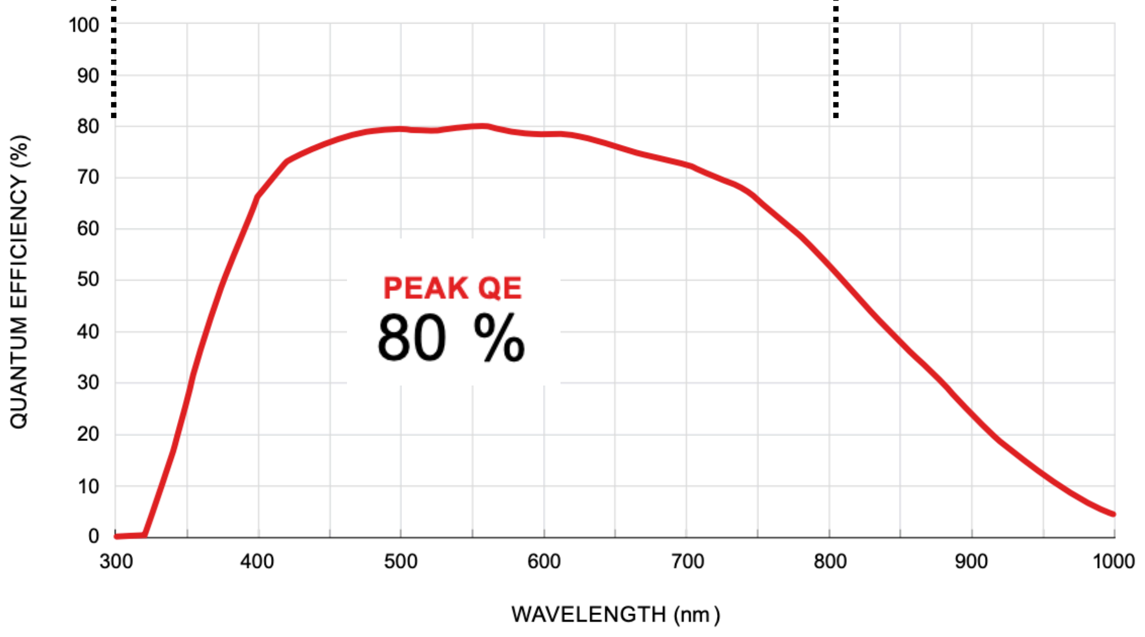




LIME prototype

tracking performance



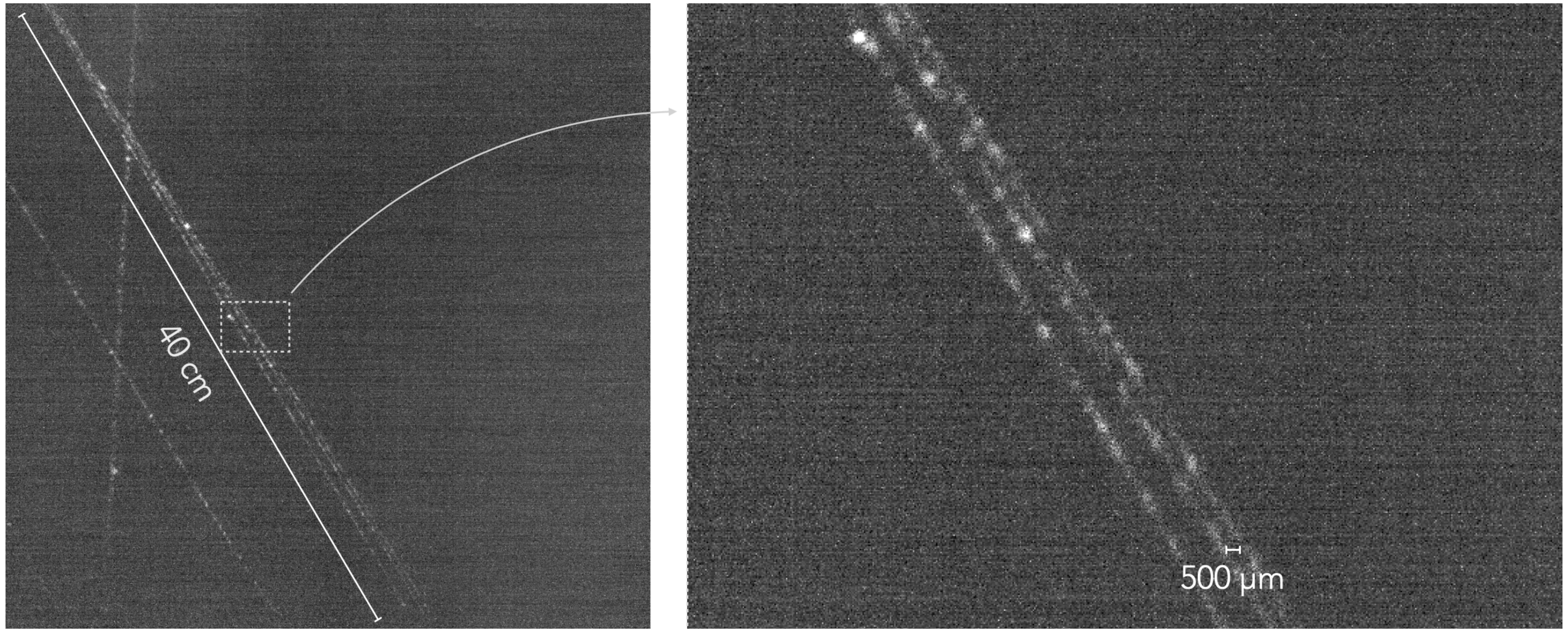


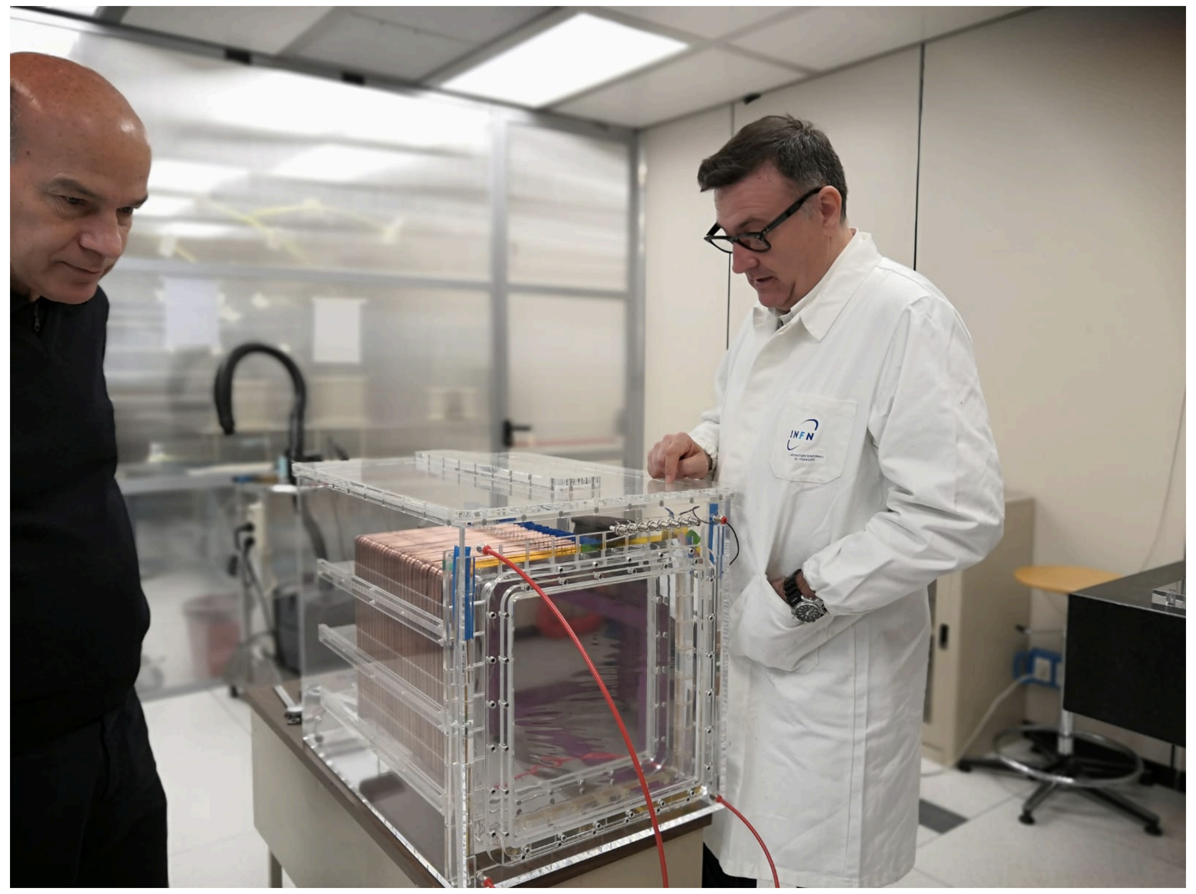
ORCA-Fusion

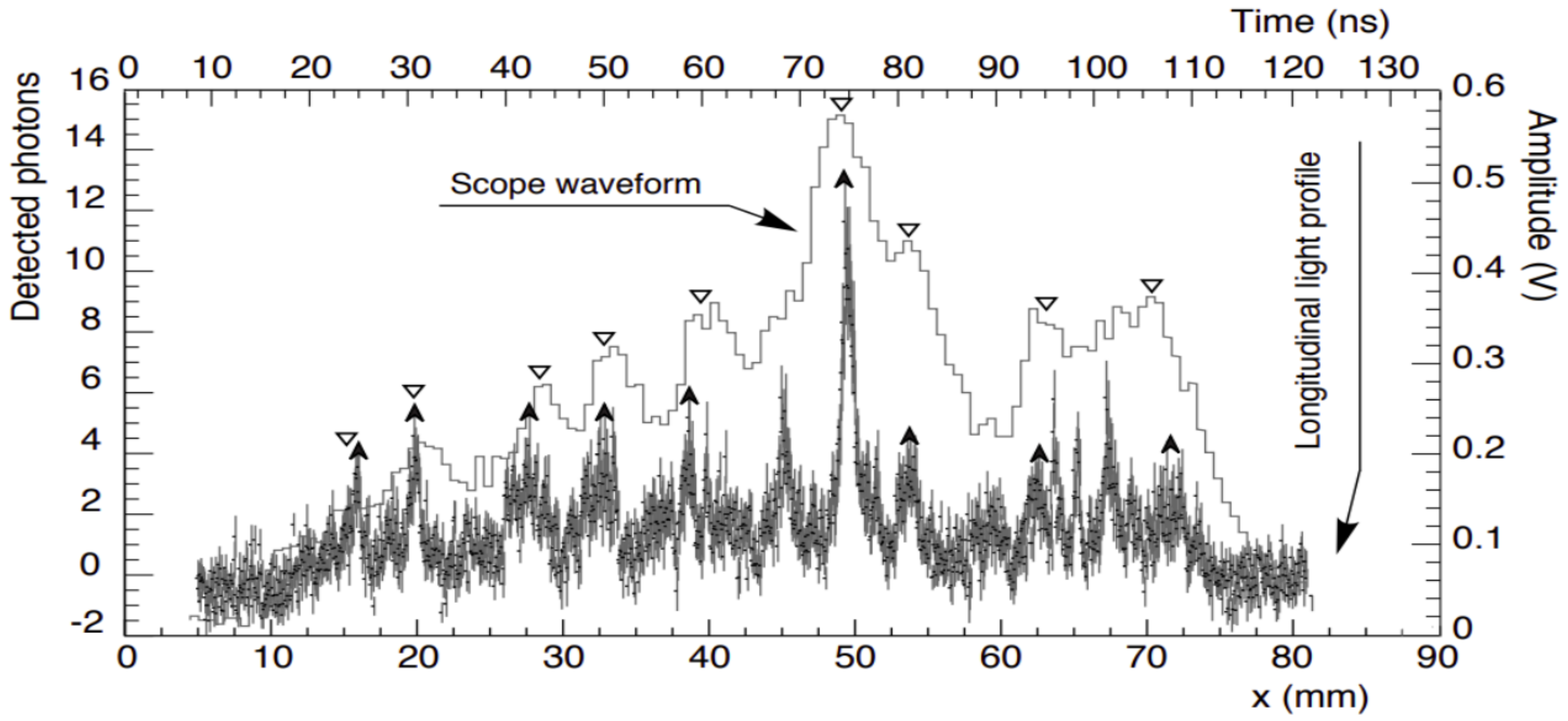
HIGH RESOLUTION
2304 × 2304
5.3 Megapixels

READOUT NOISE
0.7 electrons rms
Ultra-quiet Scan

Example of a few cosmic tracks in LIME (Long Imaging ModulE)





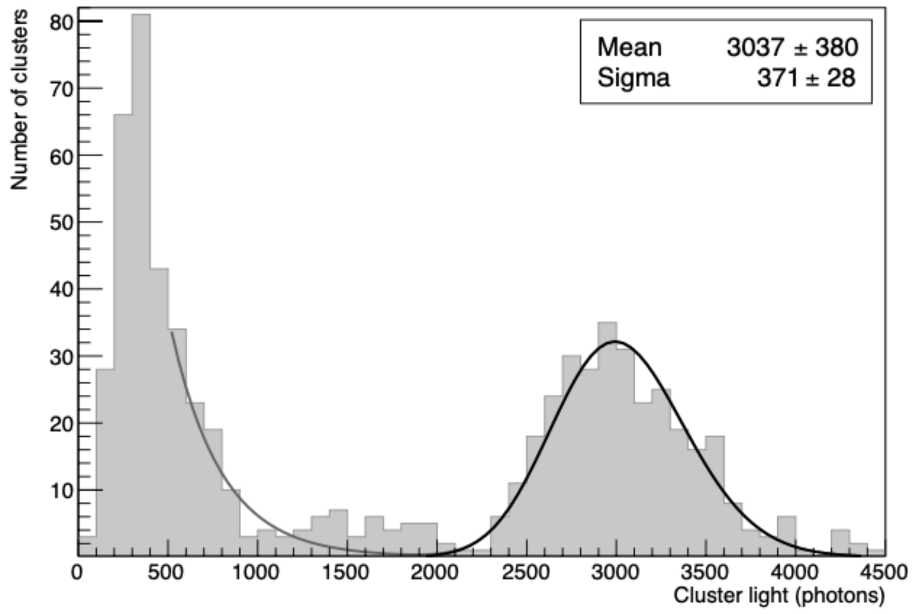


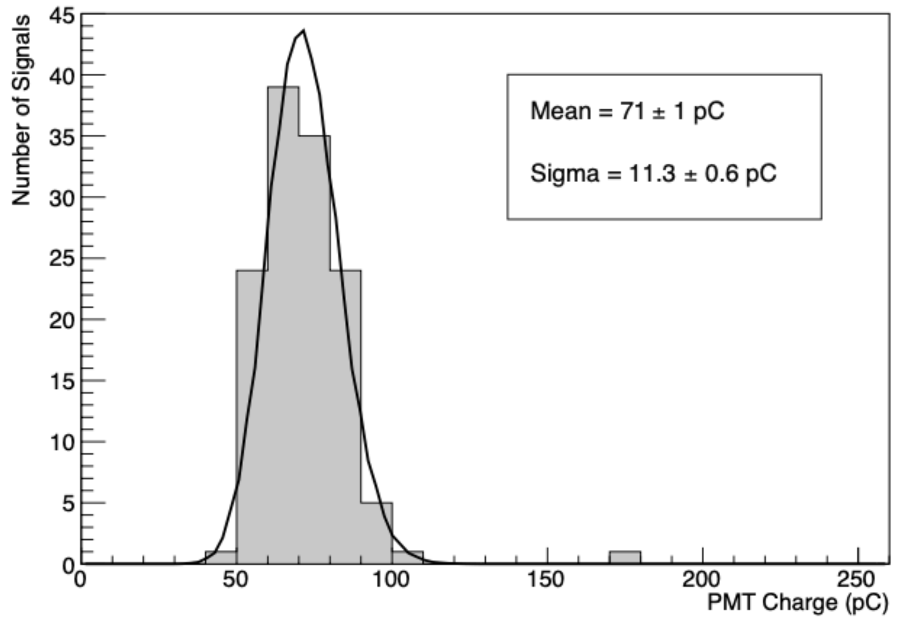
3D reco: PMT vs CMERA track clusters

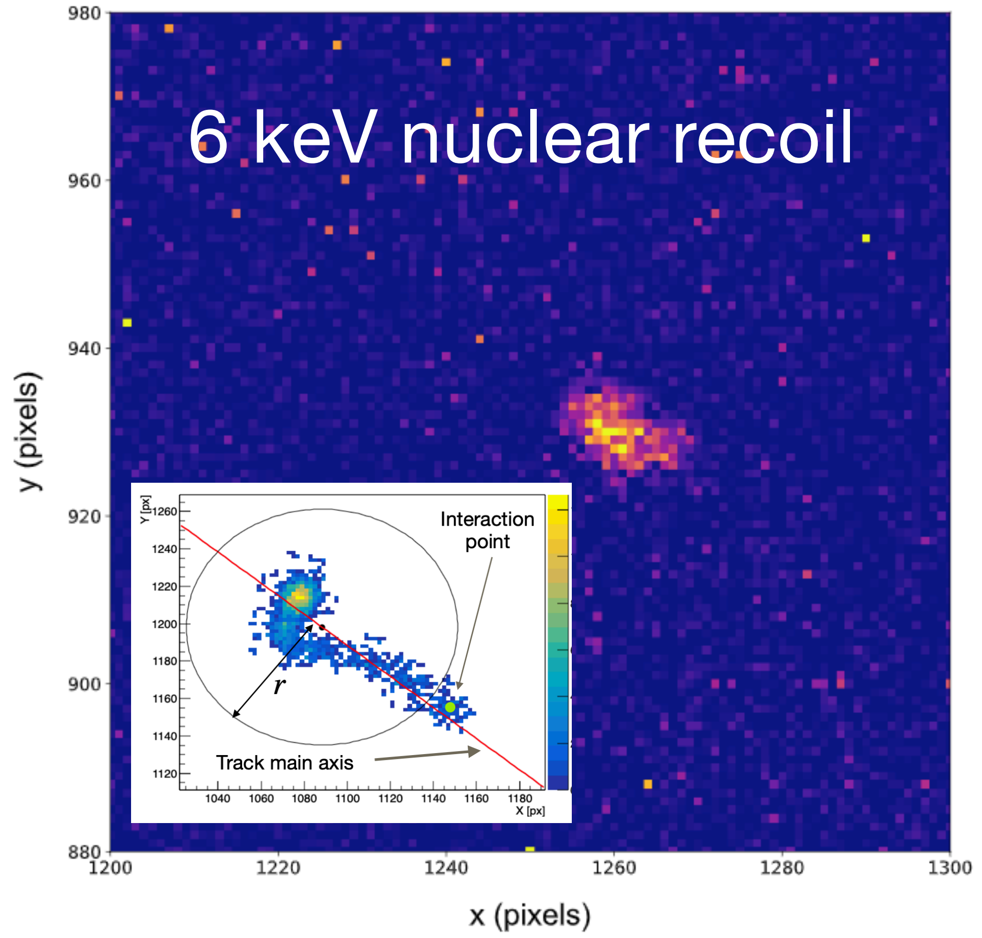
LIME prototype
overground performance

[The CYGNO Experiment. Instruments 2022, 6, 6.
https://doi.org/10.3390/instruments6010006]

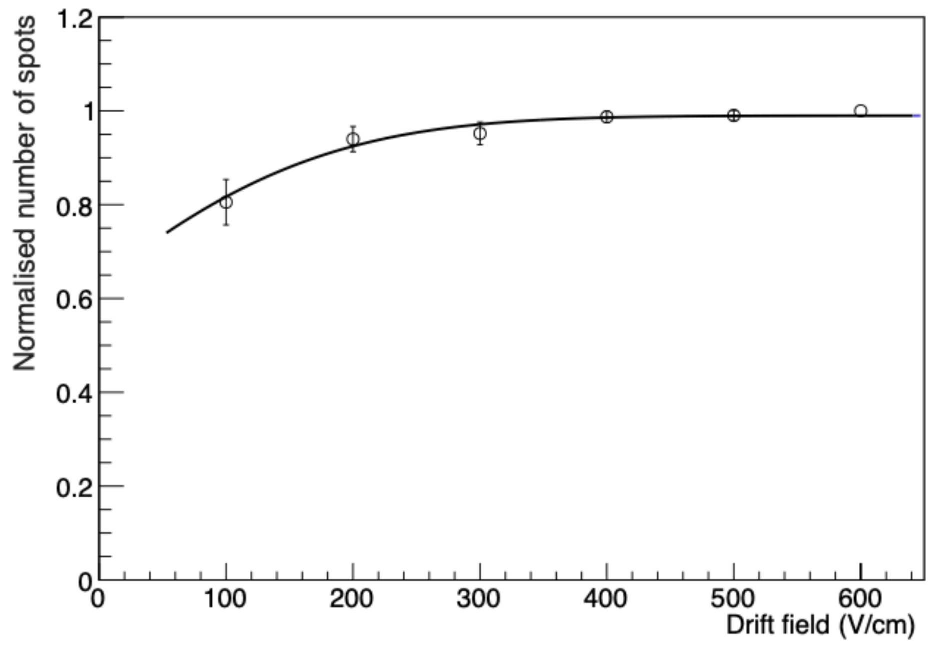
- distribution of the **light content** of the ^{55}Fe events reconstructed from the sCMOS images (left), and distribution of the **charge measured by the PMT** signals (right).

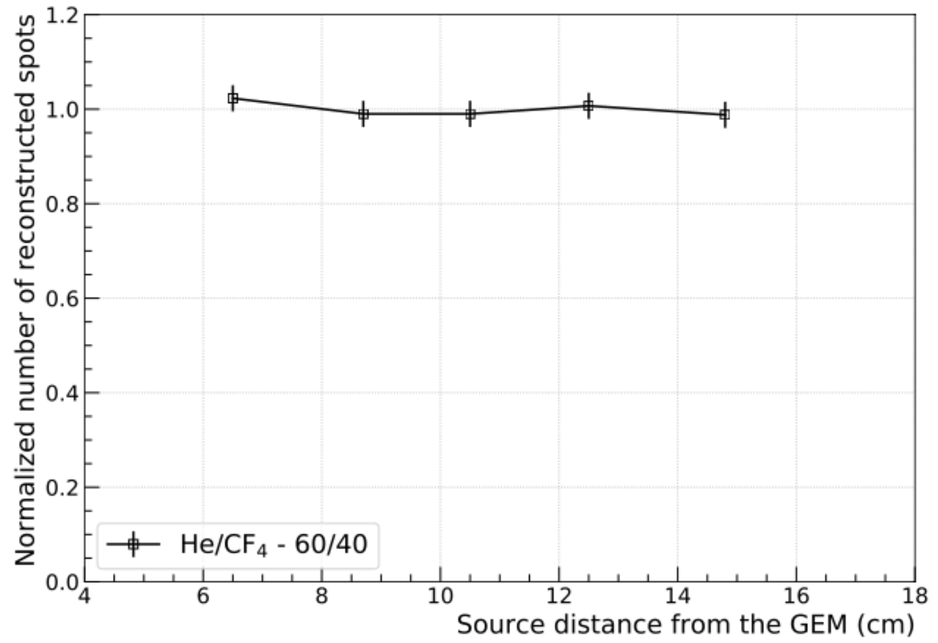




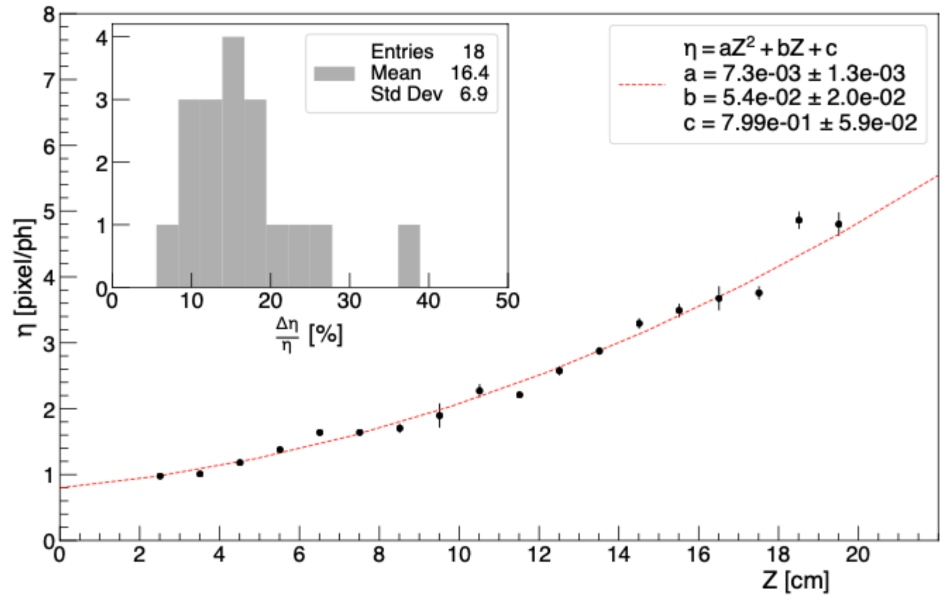


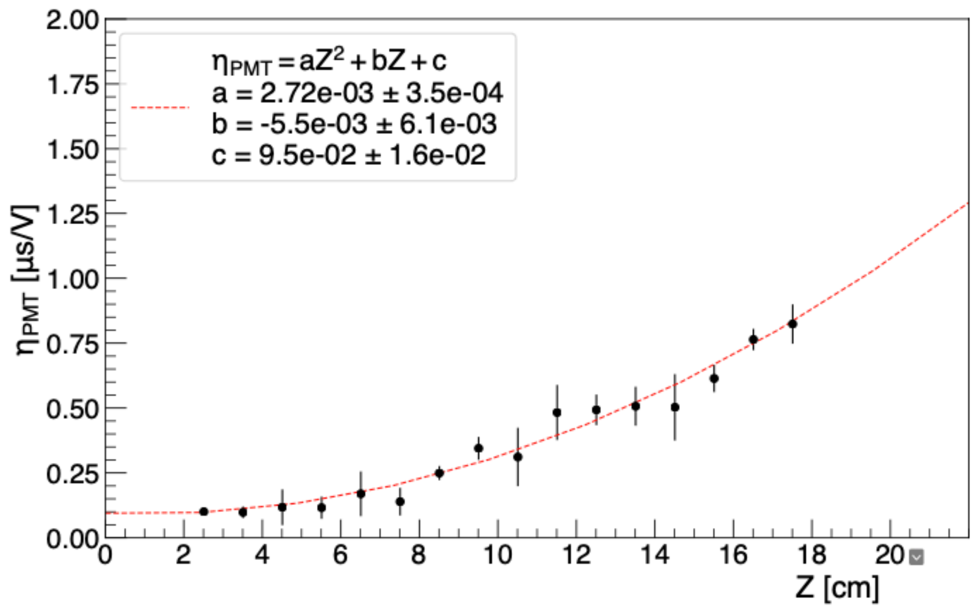
- behaviour of the normalised number of ^{55}Fe spots as a function of the drift **electric field** (left) and event depth in the **sensitive volume** (right)



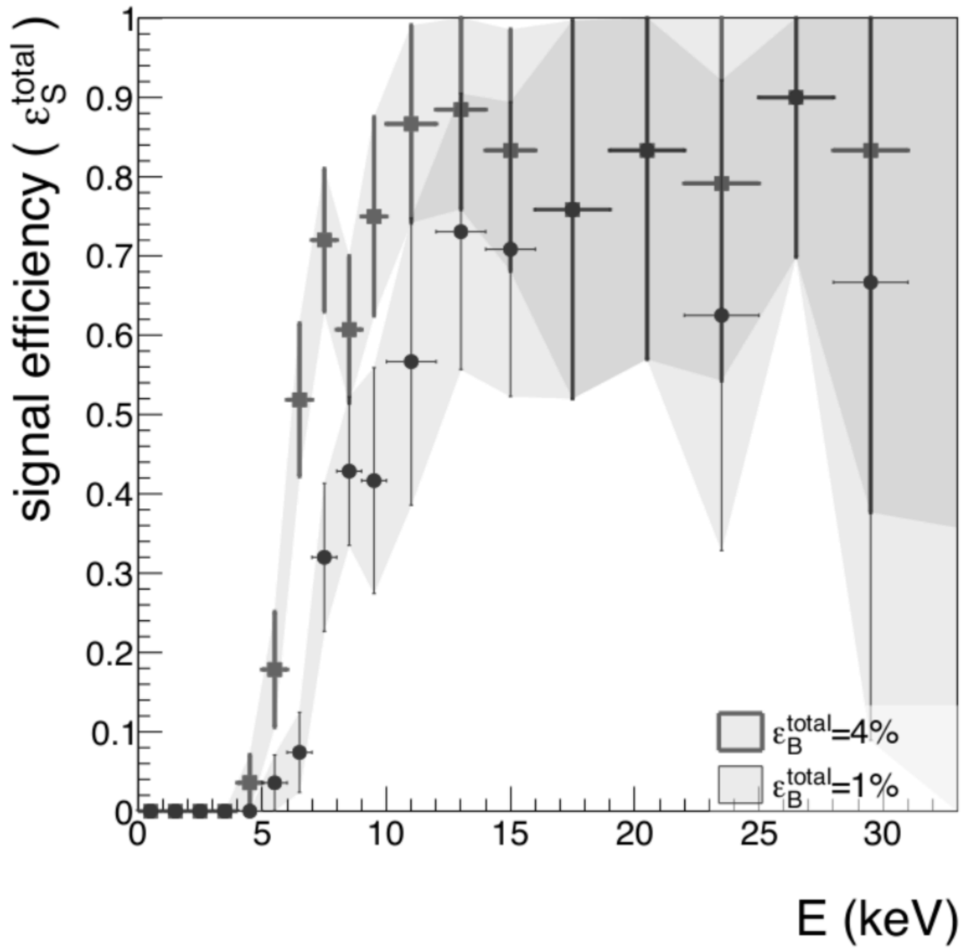


- dependence of η on the left and ηPMT on the right as a function of the track **distance** from the GEM





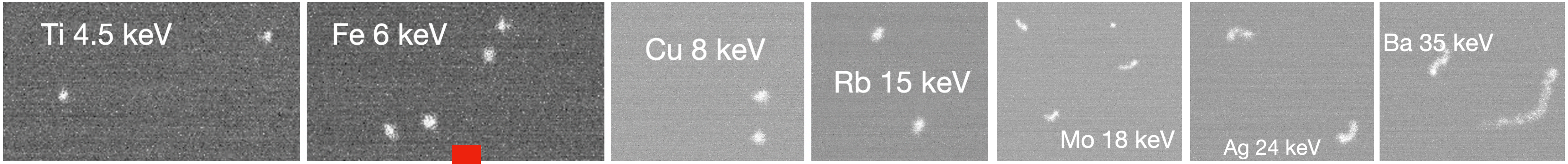
- **detection efficiency** for **nuclear recoils** (ε_{total}) as a function of their detected energy for electron recoils efficiency of 4% (squares) and 1% (circles).

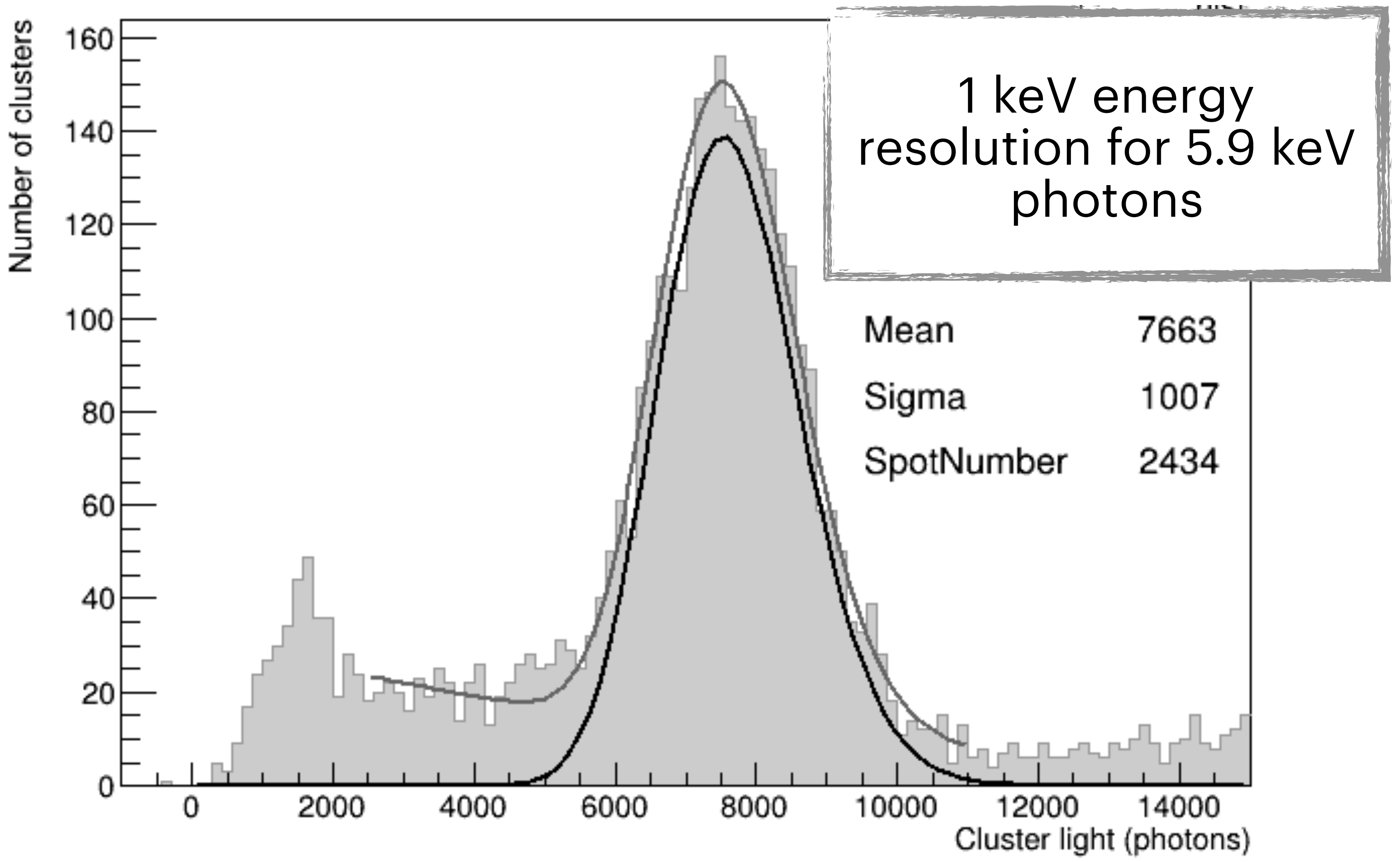


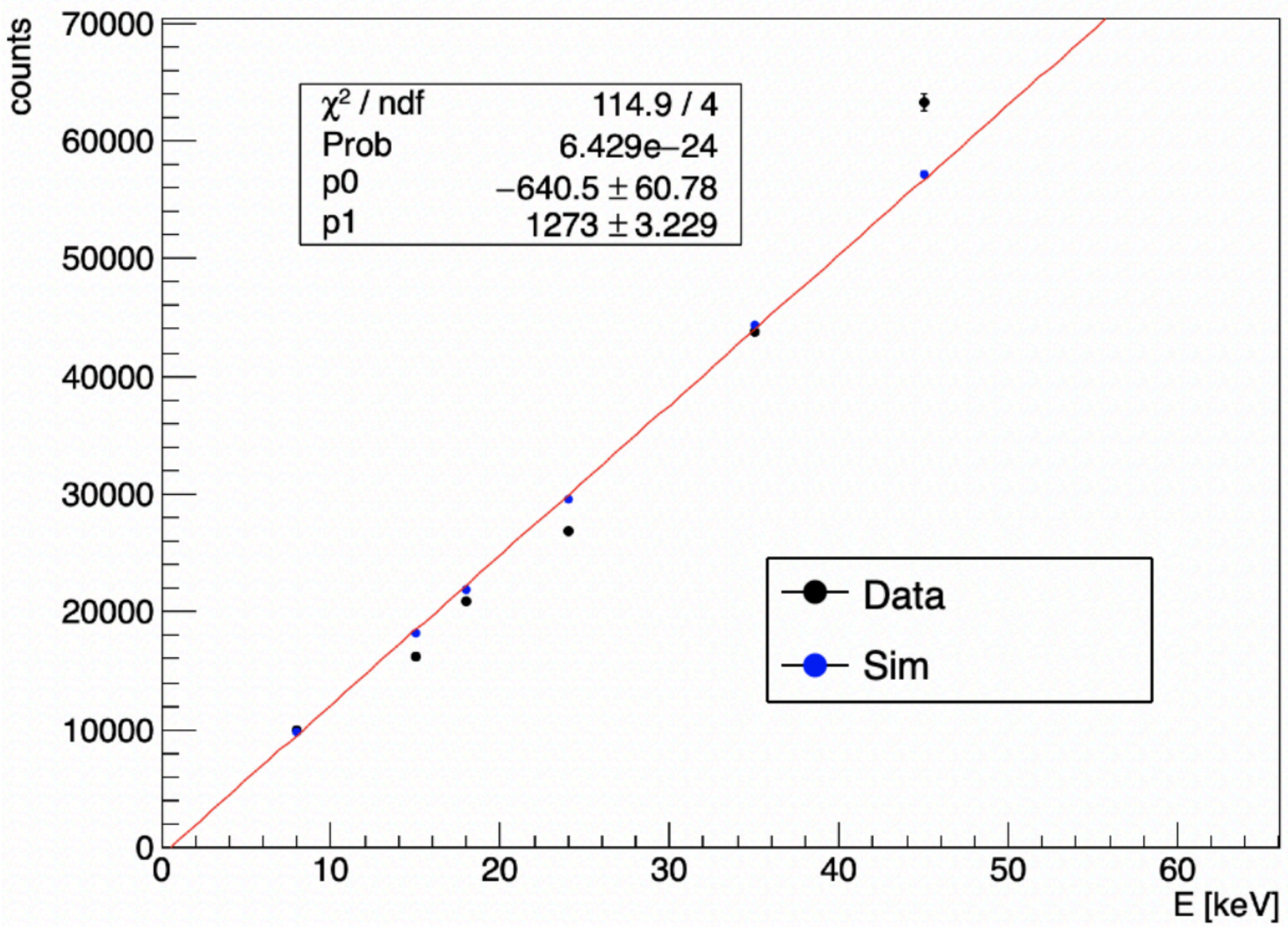
G. Mazzitelli, AI@INFN 2022

LIME prototype

beta/gamma performance/calibration







good linearity **response in the energy** range 4.5 keV - 45 keV