Istituto Nazionale di Fisica Nucleare
Laboratori Nazionali di Frascati

Status Of Reconstruction Code

G. Dho

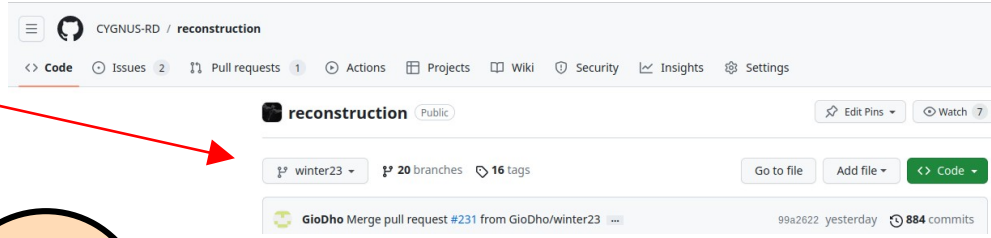
General Status

- The reconstruction code is technically working, but has some issues and needs development for the future requirements

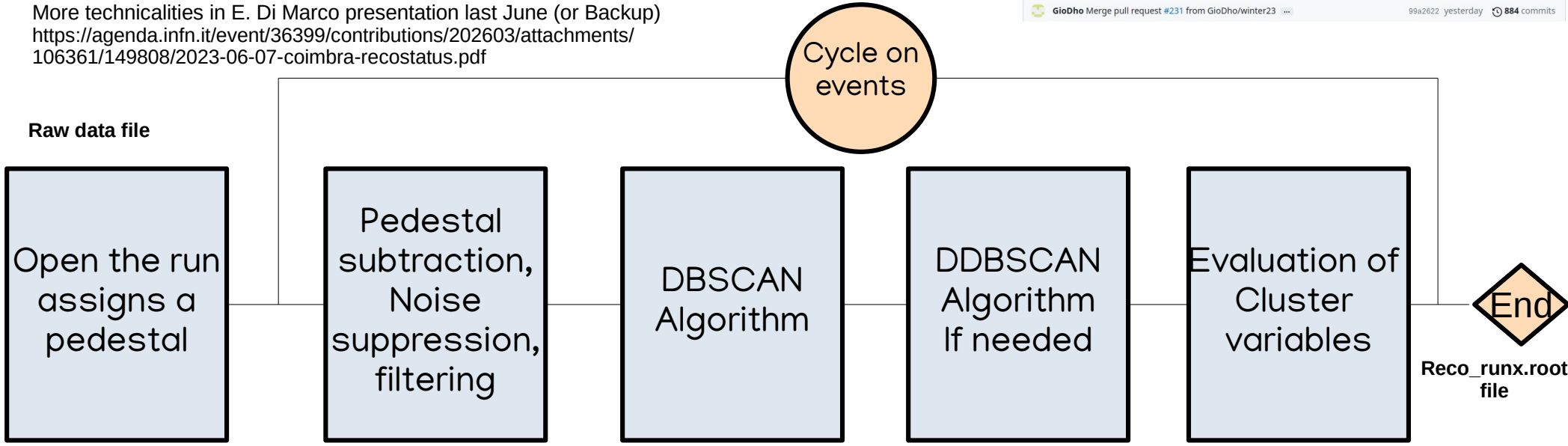
  - Latest branch
  - What to tackle now
  - What to tackle then
  - What to tackle in the future

Latest Branch

- The current more updated branch is **Winter23.**
- Some small new updates, but main structure is the same



More technicalities in E. Di Marco presentation last June (or Backup)
https://agenda.infn.it/event/36399/contributions/202603/attachments/106361/149808/2023-06-07-coimbra-recostatus.pdf



- The **goal** is to provide a single reconstruction code for all groups (MC, autoreco, analysis of LIME, MANGO etc… and PMT)

Latest Branch New Changes

- There have been some updates in the last month:
  - Now the code correctly runs on ROOT and midas data files (no h5 format yet)

Latest Branch New Changes

- There have been some updates in the last month:
  - Now the code correctly runs on ROOT and midas data files (no h5 format yet)
  - Environmental variables included (also not crashing for the “Renga-bugged” runs)

Latest Branch New Changes

- There have been some updates in the last month:
  - Now the code correctly runs on ROOT and midas data files (no h5 format yet)
  - Environmental variables included (also not crashing for the “Renga-bugged” runs)
  - Git hash saving corrected for future runs

Latest Branch New Changes

- There have been some updates in the last month:
  - Now the code correctly runs on ROOT and midas data files (no h5 format yet)
  - Environmental variables included (also not crashing for the “Renga-bugged” runs)
  - Git hash saving corrected for future runs
  - Pedestal and run search updated working both with .csv file from sql database and manual txt file with range for those who do not have a sql database (MC, MANGO)

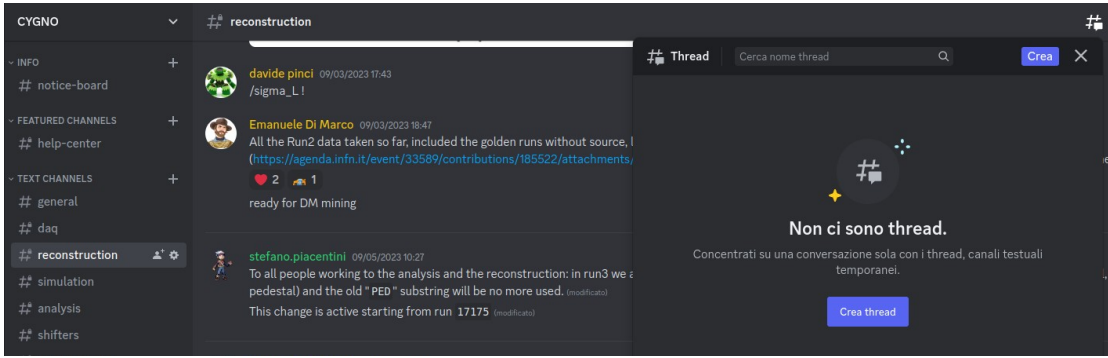
Latest Branch New Changes

- There have been some updates in the last month:
  - Now the code correctly runs on ROOT and midas data files (no h5 format yet)
  - Environmental variables included (also not crashing for the "Renga-bugged" runs)
  - Git hash saving corrected for future runs
  - Pedestal and run search updated working both with .csv file from sql database and manual txt file with range for those who do not have a sql database (MC, MANGO)
  - Speeding the debug mode for midas files (3 min → 30 s per image (if event<50))
- This means: minimal stability is available for all users (I know PMT people, I am soon working for you too), so **stop manually changing the code** for yourself in an untraceable way → we will lose track and have 100 versions.

Discord Bug Report

- But what if the code does not run or I find a bug?

Contact experts in the reconstruction channel on Discord



Discord threads on bug reports can be opened so that the bug can be fixed as soon as one finds it

- But what if a new development or piece of code wants to be added to main repository?

1) Fetch your fork, check for compatibility with latest version and open a PR

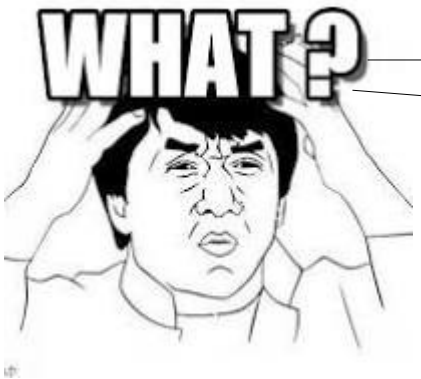
2) A check on the code will be performed

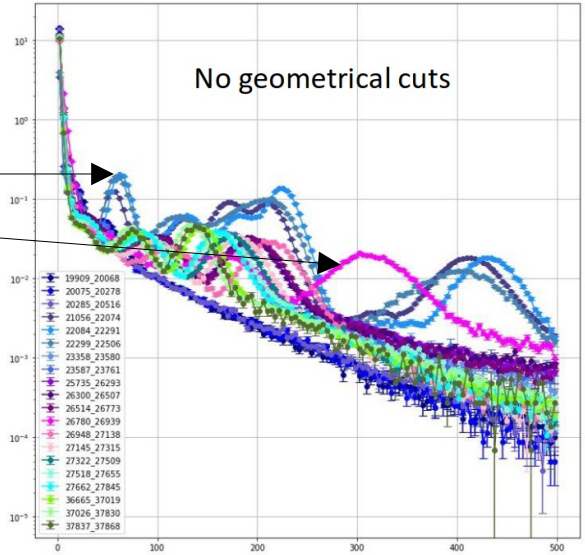
3) Some plots and tests on test runs will be required

Let's minimize errors

What to Tackle Now

- The main first order problem is to provide reliable data for LIME analysis and MC comparison





From Di Giambattista's simulation presentation

- A detailed history of the reconstruction parameters and conditions should be assessed

Might be impossible due to lack of some information

We plan to rerun on all Run3 a new more stable version of the code

What to Tackle Now II

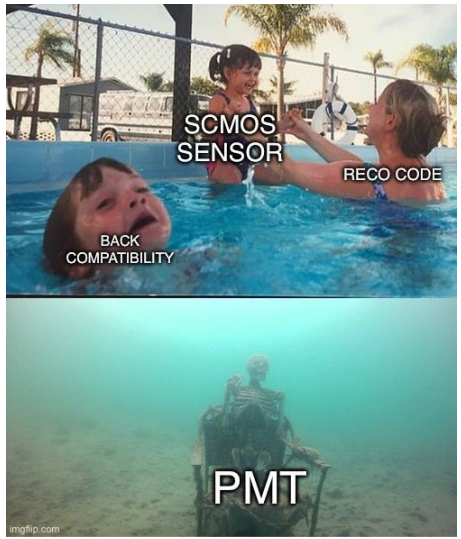
- Since we want to rerun on all data, we need a stable version which needs to have:

Saving of parameters	Correction of rotation	Decision on sensor border
Somewhat userfriendly and hardly hackable	In the end, we are a directional detector, we need to know the orientation of the detector	What to do with the camera noisy parts

What to Tackle Now II

- Since we want to rerun on all data, we need a stable version which needs to have:

Saving of parameters	Correction of rotation	Decision on sensor border	Inclusion of PMT variables
Somewhat userfriendly and hardly hackable	In the end, we are a directional detector, we need to know the orientation of the detector	What to do with the camera noisy parts	Do not forget about PMT



What to Tackle Now IIb

- Decision on borders:

I would add a flag on config file
to remove them or keep them

```
47
48 'scfullinfo'            : True,
49 'save_MC_data'          : False,
50
51 'tip'                   : '3D',
52 'saturation_corr'       : False,
53
54 # Superclusters parameters are hardcoded
55 'calibrate_clusters'    : False,
56
```

'Remove_border' : True,

What to Tackle Now IIb

- Decision on borders:

I would add a flag on config file to remove them or keep them

```
47
48 'scfullinfo'            : True,
49 'save_MC_data'          : False,
50
51 'tip'                   : '3D',
52 'saturation_corr'       : False,
53
54 # Superclusters parameters are hardcoded
55 'calibrate_clusters'    : False,
56
```

'Remove_border' : True,

- Save the parameters.. Still open question

Just githash:

- Ok for the version of the code,

But

- it would require to make new commits for every change of parameters
- No guarantees parameters will not be chaged

Save param inside recofile:

- Does not hurt. Hard to hack
- The self reco made by Giovanni could store them in Grafana

But

- Something could be harder to save (vignetting map)

At runtime create a file with all save in s3 and store link in sql table:

- Almost no possibility to hack it

But

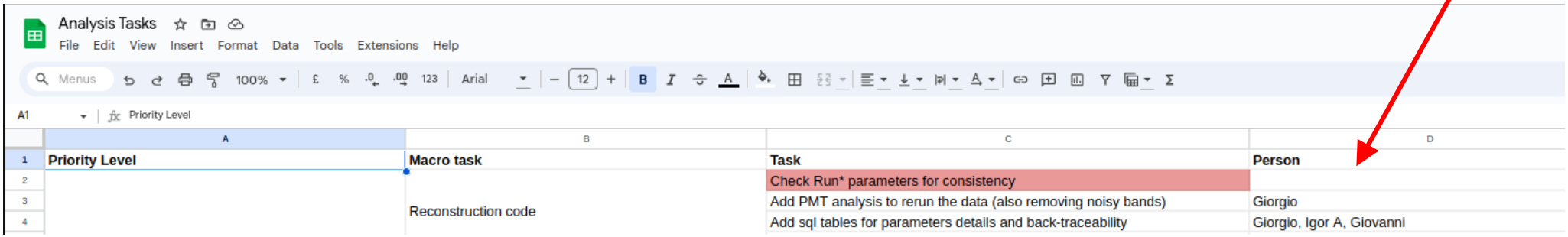
- Still to be determined how this file can be saved

What to Tackle Now III

- When it is time to rerun everything with all these features, it is important to check randomly runs along Run3 to check that the same parameters work for all sets

If not they need to be changed to correctly reconstruct (i.e. Ba, Eu) → Check for sc_integral consistency

- This is an important task



	A	B	C	D
1	**Priority Level**	**Macro task**	**Task**	**Person**
2			Check Run* parameters for consistency	
3		Reconstruction code	Add PMT analysis to rerun the data (also removing noisy bands)	Giorgio
4			Add sql tables for parameters details and back-traceability	Giorgio, Igor A, Giovanni

- Only then reconstruct all data (PMT will be incuded)

What to Tackle Then

- For the near future some updates are planned to improve the code and to move towards CYGNO-04:

Pedestal Reconstruction

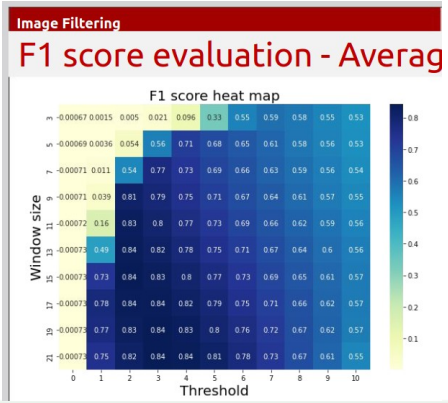
Automatic reconstruction of pedestal runs with itself (evaluate fake clusters) and with previous pedestal, to monitor the sensor evolution

Noise filtering techniques:

Already discussed in E. Di Marco presentation in June:

Convolutional method by I. Pains,

U-Net, optimised median filter (https://agenda.infn.it/event/32788/contributions/181084/attachments/97228/134158/Filtering%20CYGNO.pdf)



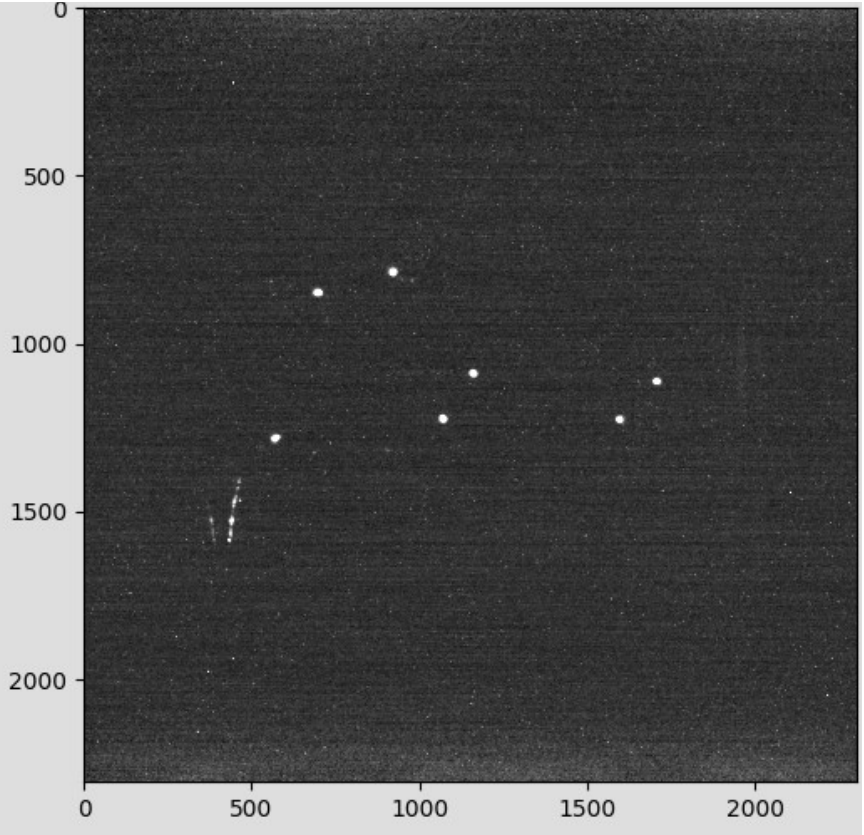
Adapt for QUEST:

Not only a matter of different counting and pixel numer

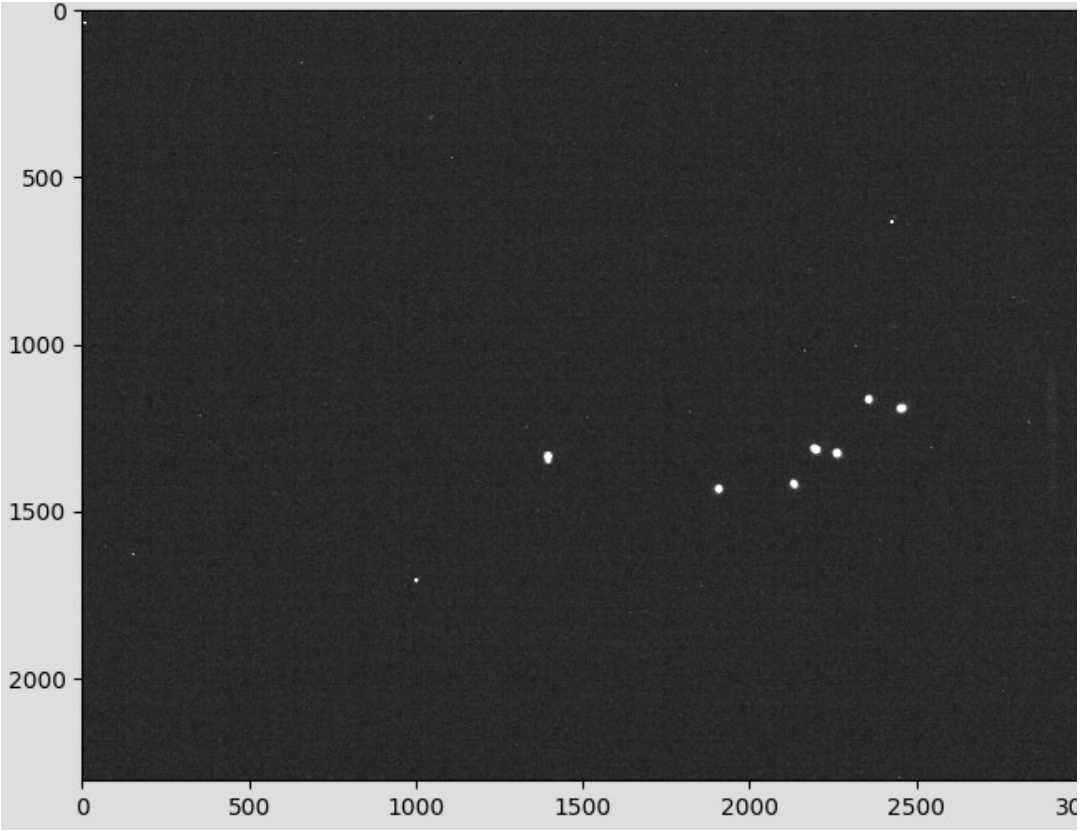
But can we improve the noise treatment?

(Fusion vs Quest)

Fusion



Quest



What to Tackle Then II

- Also for analysed data users, the idea is to provide tools to simplify life:

Provide opening data scripts

C++ and python scripts with all necessary tools and libraries to access all info contained in recoed files

middleware / conf / analyzer / dev / **reco-analyzer.py**

gmazzitelli reco example

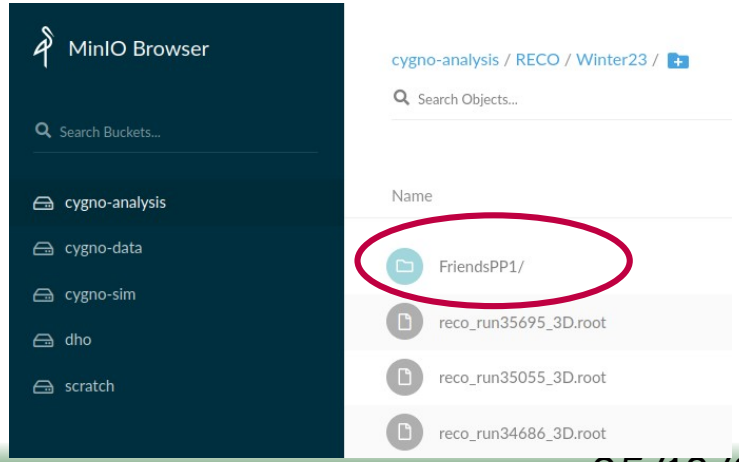
Code | Blame | Executable File · 95 lines (85 loc) · 3.52 KB | Raw

```
#!/usr/bin/env python3
#
# G. Mazzitelli 2022
# versione DAQ LNGS/LNF per midas reco-analyzer
#
import numpy as np
import uproot
import pandas as pd
import os


def GetLY(tf, verbose):
    df_A = tf['Events'].arrays(['sc_rms', 'sc_tgausssigma', 'sc_width', 'sc_length', 'sc_xmean',
                                'sc_ymean','sc_integral'], library = 'pd')

    if verbose: print (df_A)
    sel   = df_A[
        (df_A['sc_rms'] > 6) & (0.152 * df_A['sc_tgausssigma'] > 0.5) &
        (np.sqrt((df_A['sc_xmean']-2304/2)**2 + (df_A['sc_ymean']-2304/2)**2) < 800)  &
        (df_A['sc_integral'] > 30_000) & (df_A['sc_integral']<300_000)
    ].copy()

    p = np.array([7.51266058e-02, -1.32492111e+03])

    return p[0]*np.mean(sel['sc_integral'])+p[1], p[0]*np.std(sel['sc_integral']) / np.sqrt(len(sel))

def get_epoch(file_url):
    import requests
    from datetime import datetime
    r = requests.get(file_url)
    utc_time = datetime.strptime(r.headers['last-modified'], "%a, %d %b %Y %H:%M:%S %Z")
    epoch_time = (utc_time - datetime(1970, 1, 1)).total_seconds()
    return epoch_time

```

Automatise post-reco routines

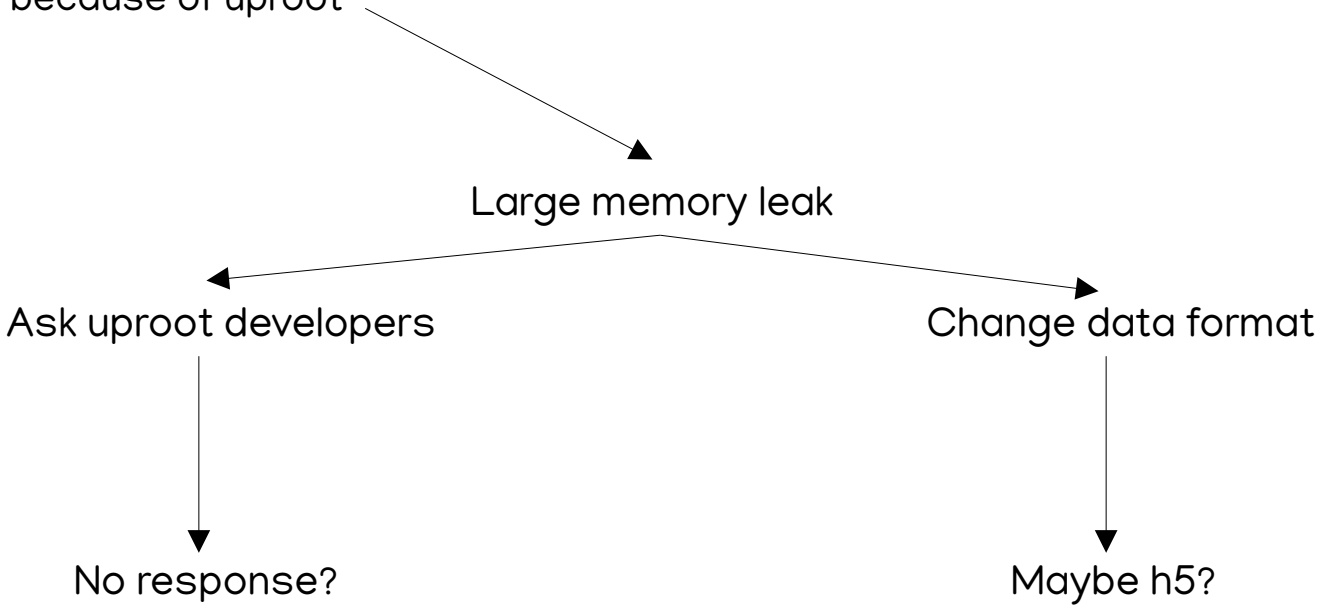
Important analyses which cannot be performed on run by run basis

- ^{55}Fe and Rita factor data correction
- Emanuele's Regression analysis And root friend files



What to Tackle Then III

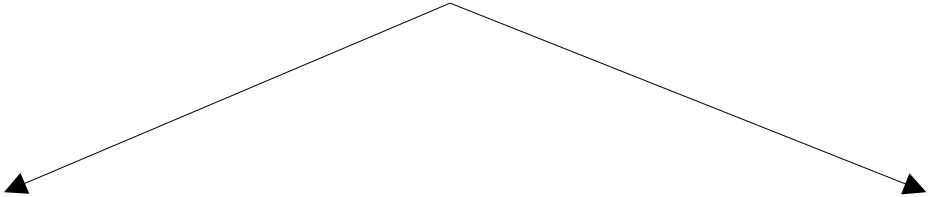
- For MC **big issue** because of uproot



PLEASE LET'S WORK TOGETHER

What to Tackle in the Future

- For CYGNO-04 the reconstruction structure needs some rework
- The data output of 4 QUEST cameras prevents from saving full resolution images
- The reconstruction code should be made modular



Reconstruction part
"first level trigger"

Strong connection with DAQ group

Analysis part of the events

Strong connection analysis group

What to Tackle in the Future II

Reconstruction part

1) Images from cameras and waveforms from PMT continuously taken.

2) Live coarse selection of only interesting pixels (exploit the work on GPU and trigger algorithms developed by Brasilian group)

3) Clustering algorithm applied to the selected pixels

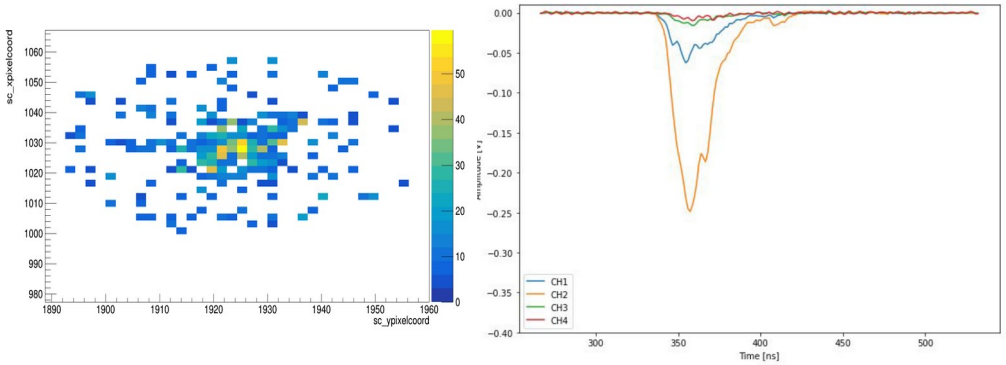
4) Bayesian fit (or something else) of PMT waveforms used to match PMT event with clusters

5) Output file contains **Event = track + corresponding waveforms**

No interaction needed by the analysis user who is no more involved in the track recognition part

https://agenda.infn.it/event/38032/contributions/213585/attachments/111926/159830/Trigger%20Proposal%20Status.pdf

May be written in CUDA and or C++



What to Tackle in the Future III

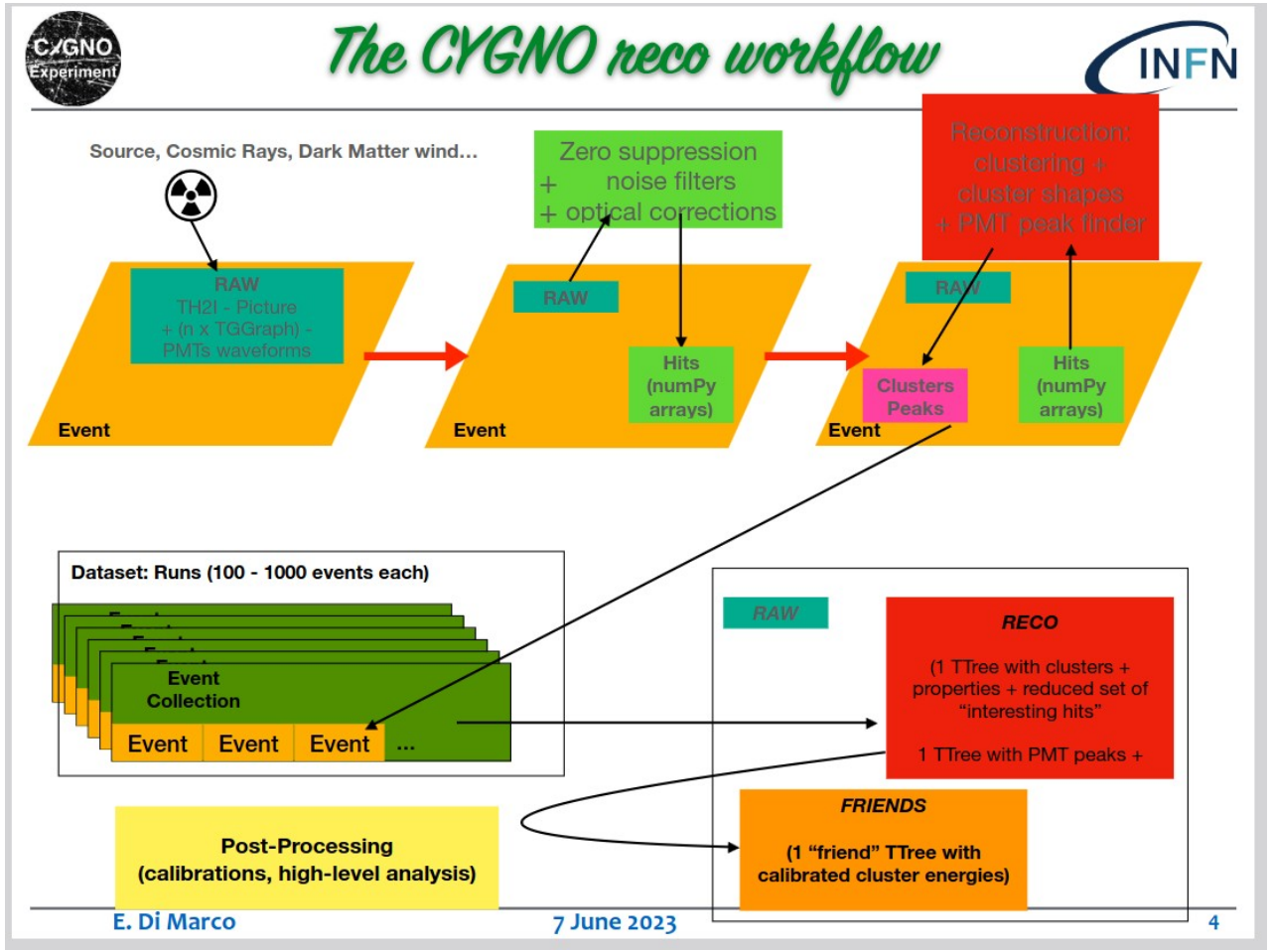
Analysis part

- The output of the reco part is accessible to everyone to perform any type of analysis desired
- Almost no analysis is performed in the first part (just integral, position and time of the cluster)
- An automatic analysis will be performed similar to the one made now
- More complex analyses (directionality) can start from the event generated by the reco without the need to reconstruct the raw original data (faster also for the analysis guy)

Conclusions

- Reconstruction code along with reconstruction flow is working fine
- Clearly there are issues as anyone could expect
- New version of the code is out which should be stable for all users
- Rereconstruction of Run3 is forseen
- More control will be put in new developments and Discord bug report threads is instantiated
- The devolpment for CYGNO-04 future data output requires strict work with DAQ personnel
    - Modular structure
    - First part: pure reconstruction exploiting Brasilian group software trigger
    - Second part of pure analysis on reduced data

Backup


The CYGNO reco workflow
INFN
Source, Cosmic Rays, Dark Matter wind...
RAW
TH2I - Picture
+ (n x TGGraph) -
PMTs waveforms
Event
Zero suppression
+ noise filters
+ optical corrections
RAW
Hits
(numPy
arrays)
Event
Reconstruction:
clustering +
cluster shapes
+ PMT peak finder
RAW
Clusters
Peaks
Hits
(numPy
arrays)
Event
Dataset: Runs (100 - 1000 events each)
Event
Collection
Event
Event
Event
...
RAW
RECO
(1 TTree with clusters +
properties + reduced set of
"interesting hits"
1 TTree with PMT peaks +
FRIENDS
(1 "friend" TTree with
calibrated cluster energies)
Post-Processing
(calibrations, high-level analysis)
E. Di Marco
7 June 2023
4