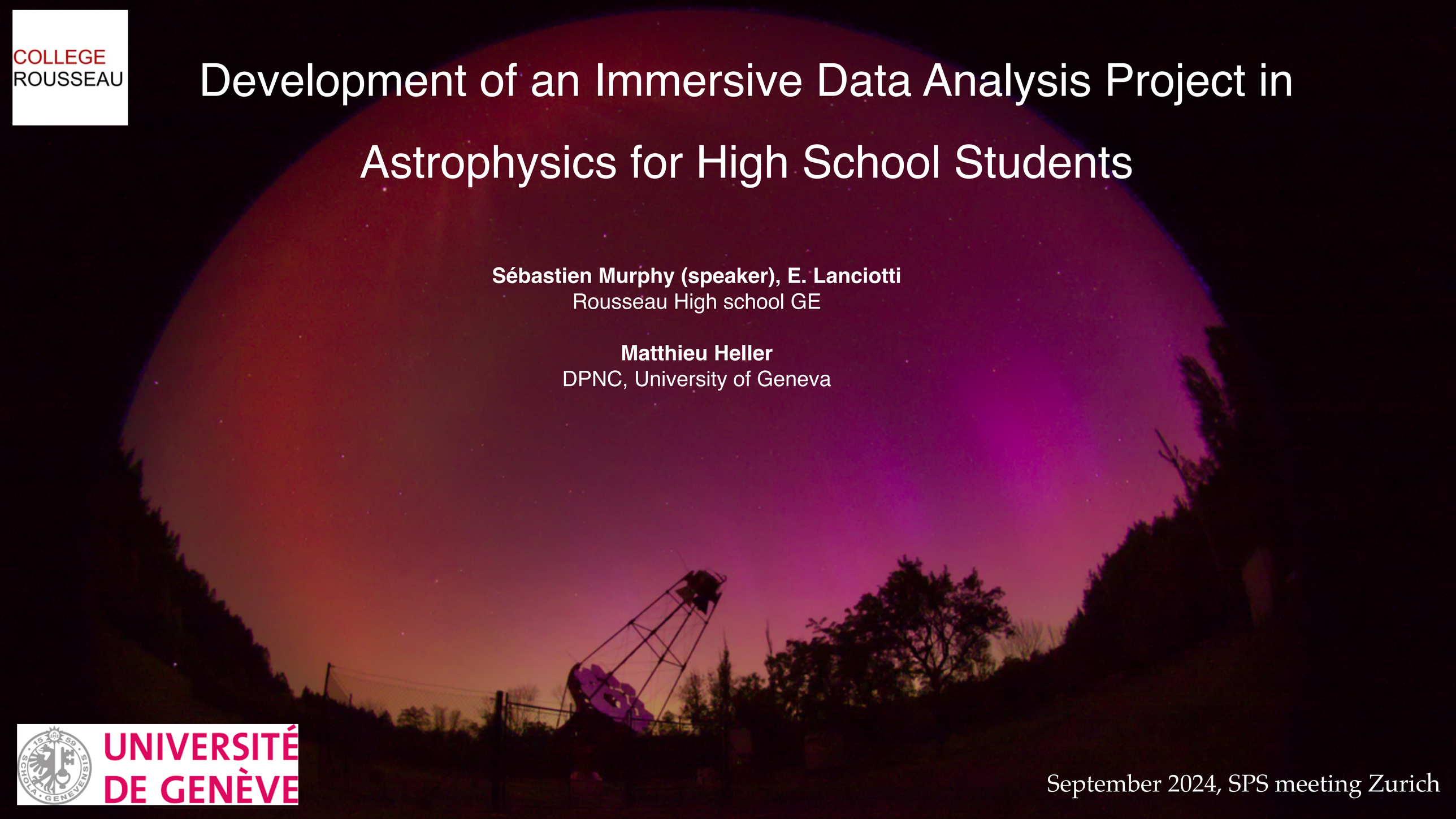
COLLEGE ROUSSEAU

Development of an Immersive Data Analysis Project in Astrophysics for High School Students

Sébastien Murphy (speaker), E. Lanciotti
Rousseau High school GE

Matthieu Heller
DPNC, University of Geneva



UNIVERSITÉ DE GENÈVE

September 2024, SPS meeting Zurich

INTRODUCTION

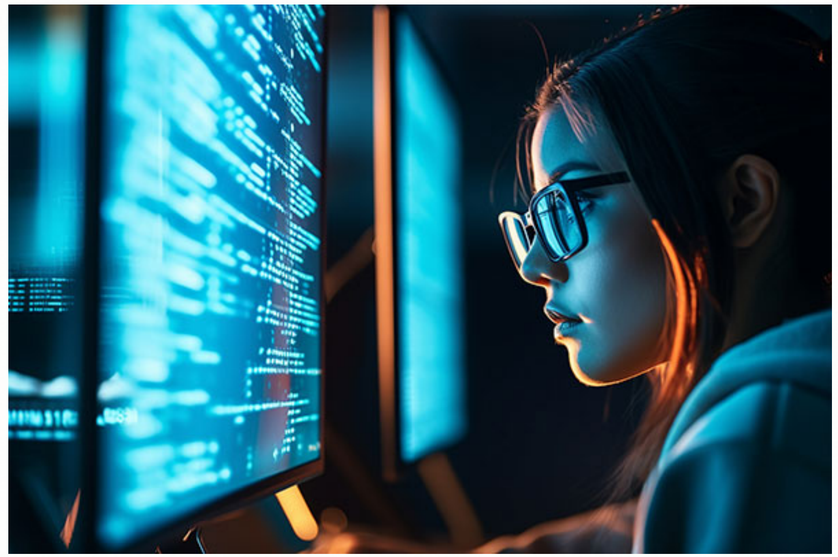
A **one-week astrophysics data analysis project** for 50 third-year students (17-18 year old) at Rousseau High School in Geneva

INTRODUCTION

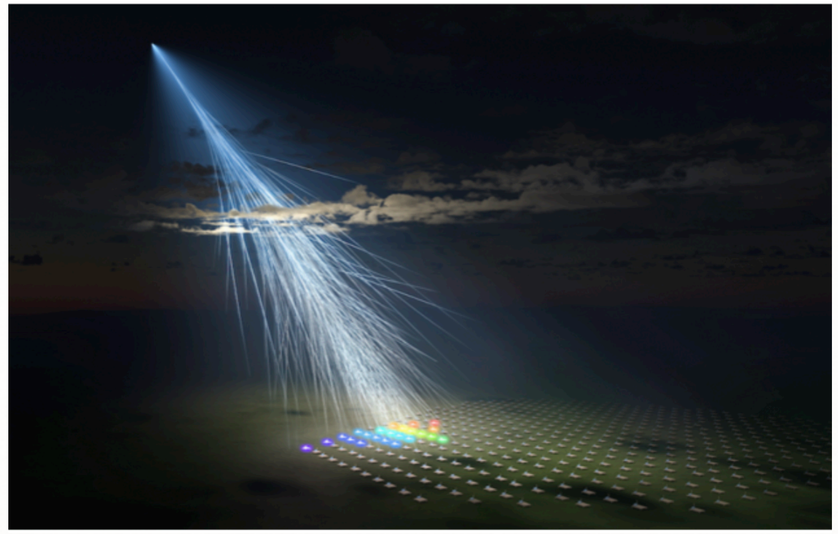
Students work in **teams of 4** to solve a problem.

- Full immersion : no other classes during the week.
- Assessed at the end of the week (scientific poster + oral exam)

INTRODUCTION



- Worked on **simulated data** from the **Single-Mirror Small Size Telescope, SST-1M** (see poster from B. Lacave), a former prototype for the **Cerenkov Telescope Array Observatory (CTAO)**.
- **Goal:** classify signal (gamma rays) from background (other cosmic rays)
- **Tools:** Python with Numpy and Matplotlib




Rationale
for the project

MOTIVATION FOR THE PROJECT

➡ Programming is **perceived as a difficult subject** by students (ages 16-18).
Partly because it's very abstract

- How to make programming more motivating and engaging?

MOTIVATION FOR THE PROJECT

➡ Programming is **perceived as a difficult subject** by students (ages 16-18).
Partly because it's very abstract

- How to make programming more motivating and engaging?

Make programming as concrete as possible.
Work on projects that are as close to real-life as possible, using an authentic project.

MOTIVATION FOR THE PROJECT

Students take on the role of physicists by working with simulated data from an existing experiment, engaging in project-based learning.

Numerous studies highlight the main benefits of using "authentic material" in teaching :

active student learning [1]

development of higher-order cognitive skills [2]

student engagement and motivation [3]

improved self-regulation capacity and reflective practice [4,5].

autonomy, collaboration between peers

References in backup

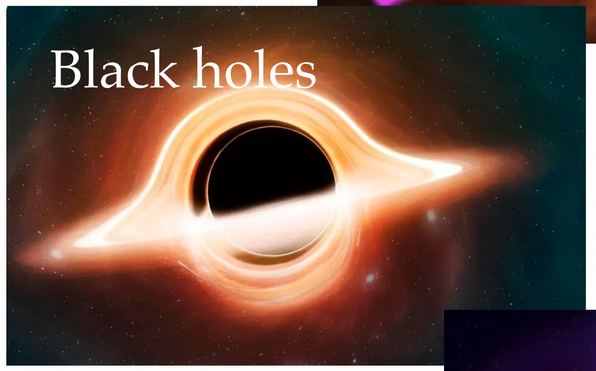
MOTIVATION FOR THE PROJECT

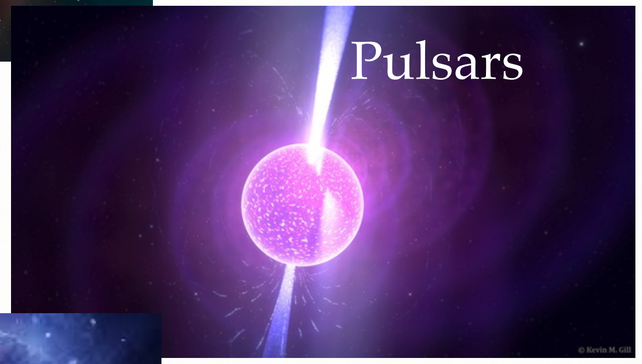
The topic is captivating

✓ We will explore radiation emitted by the **hottest and most extreme objects in the universe.**

✓ Unveiling the universe in "**colors (wavelengths) invisible to the human eye.**"

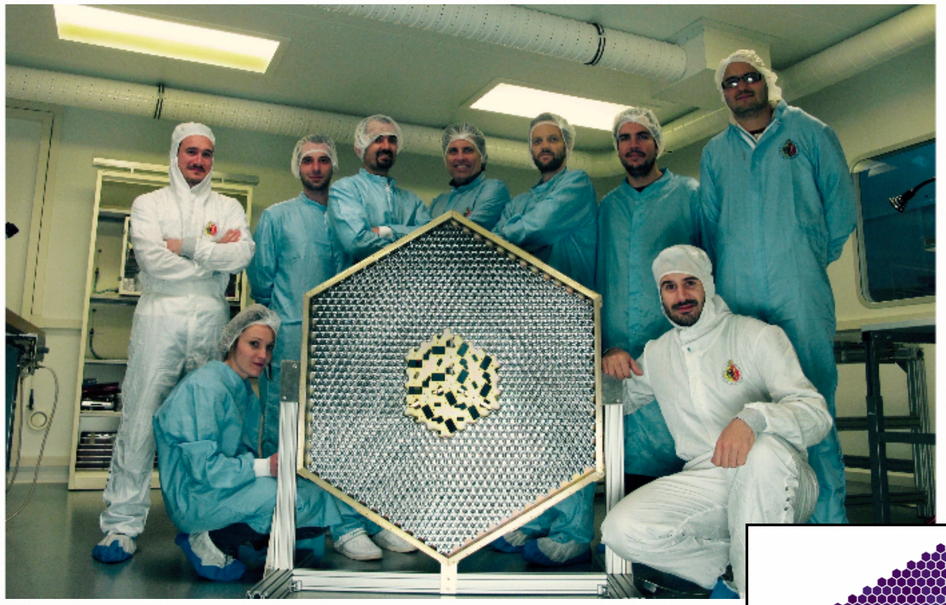
$\lambda < 10^{-11}$ **m**, $E >$ **GeV**





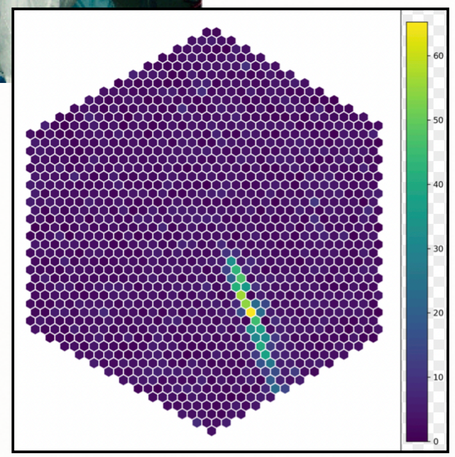
MOTIVATION FOR THE PROJECT

We happen to have a large team of physicists lead by Prof. T. Montaruli with MER M. Heller working on CTAO prototype **Small Size Telescope just a few km away from Collège Rousseau.**



UNIVERSITÉ DE GENÈVE

SST-1M
Single-Mirror
Small Size Telescope

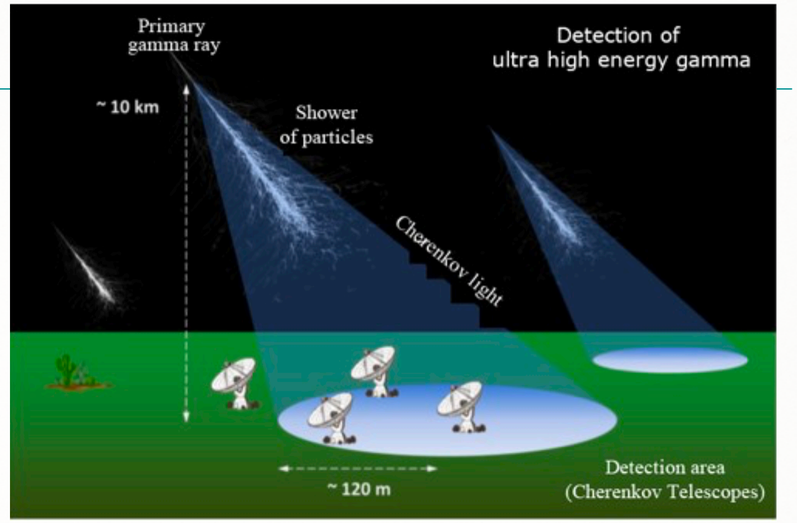


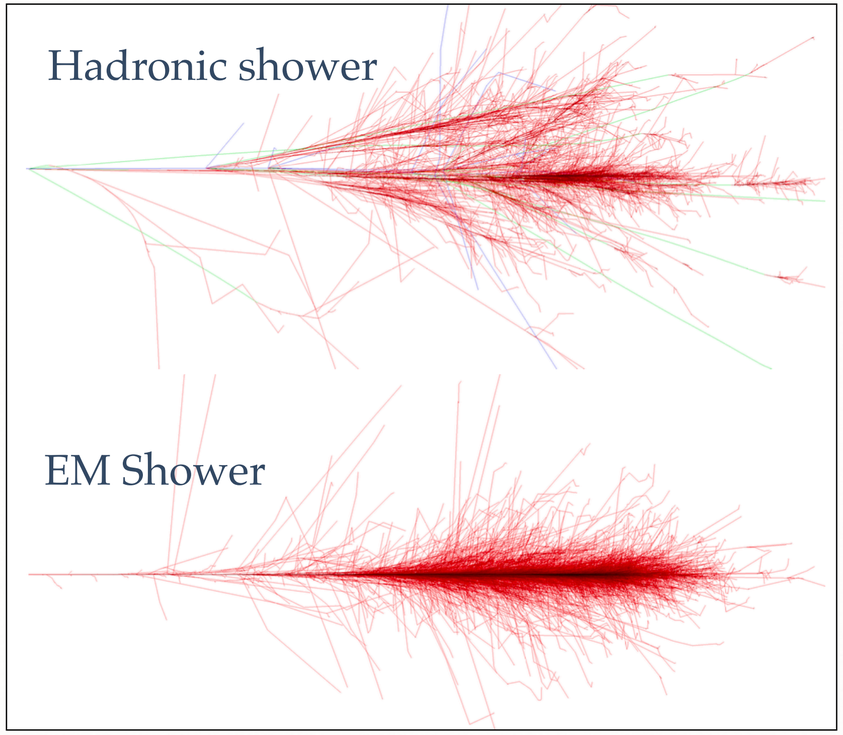
SST-1M Telescope installed at the Ondrejov Observatory in Czech Republic

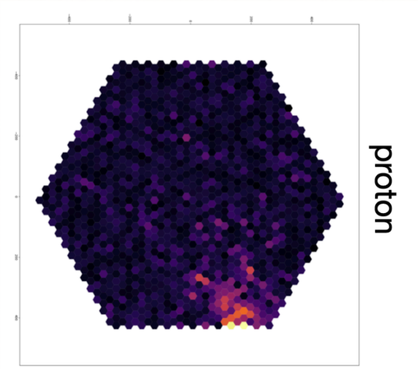
The Project

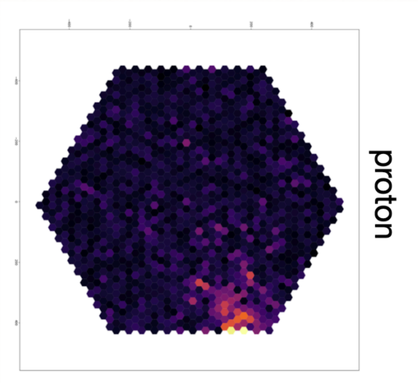
THE PROJECT

- Lead students to discover how to differentiate signal (gamma-rays) from background (cosmic rays) events when observed with Imaging Atmospheric Cherenkov Telescopes
- Hadronic showers are much more spread out and "disordered" compared to EM showers.







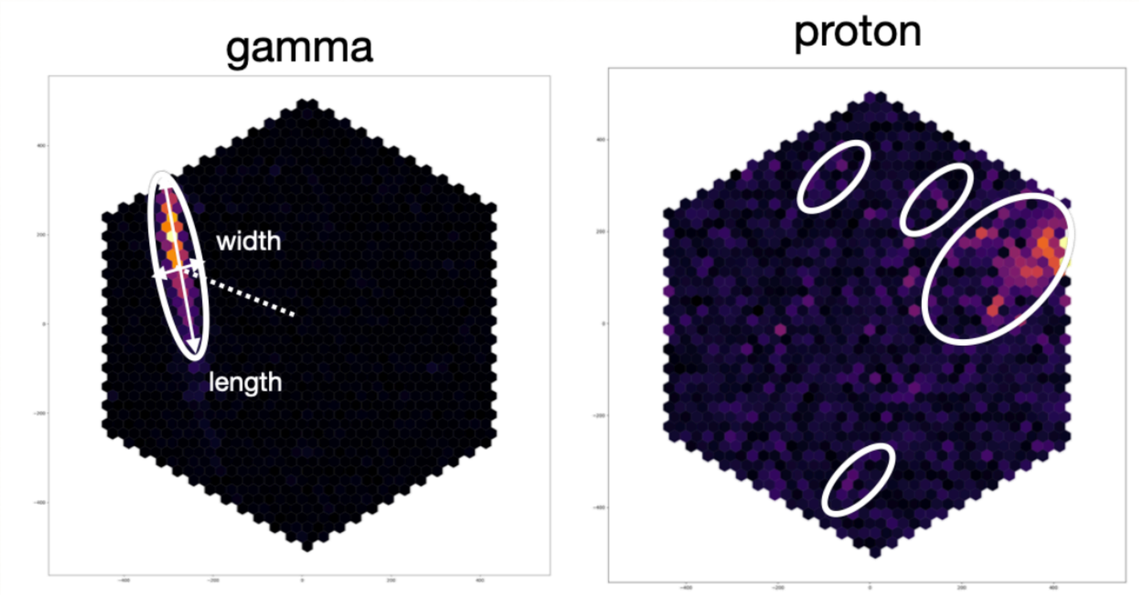


THE PROJECT

- Algorithm applied to each event to distinguish between protons and gamma rays. The image is modelled using ellipses.

➡ **Hillas parameters.**

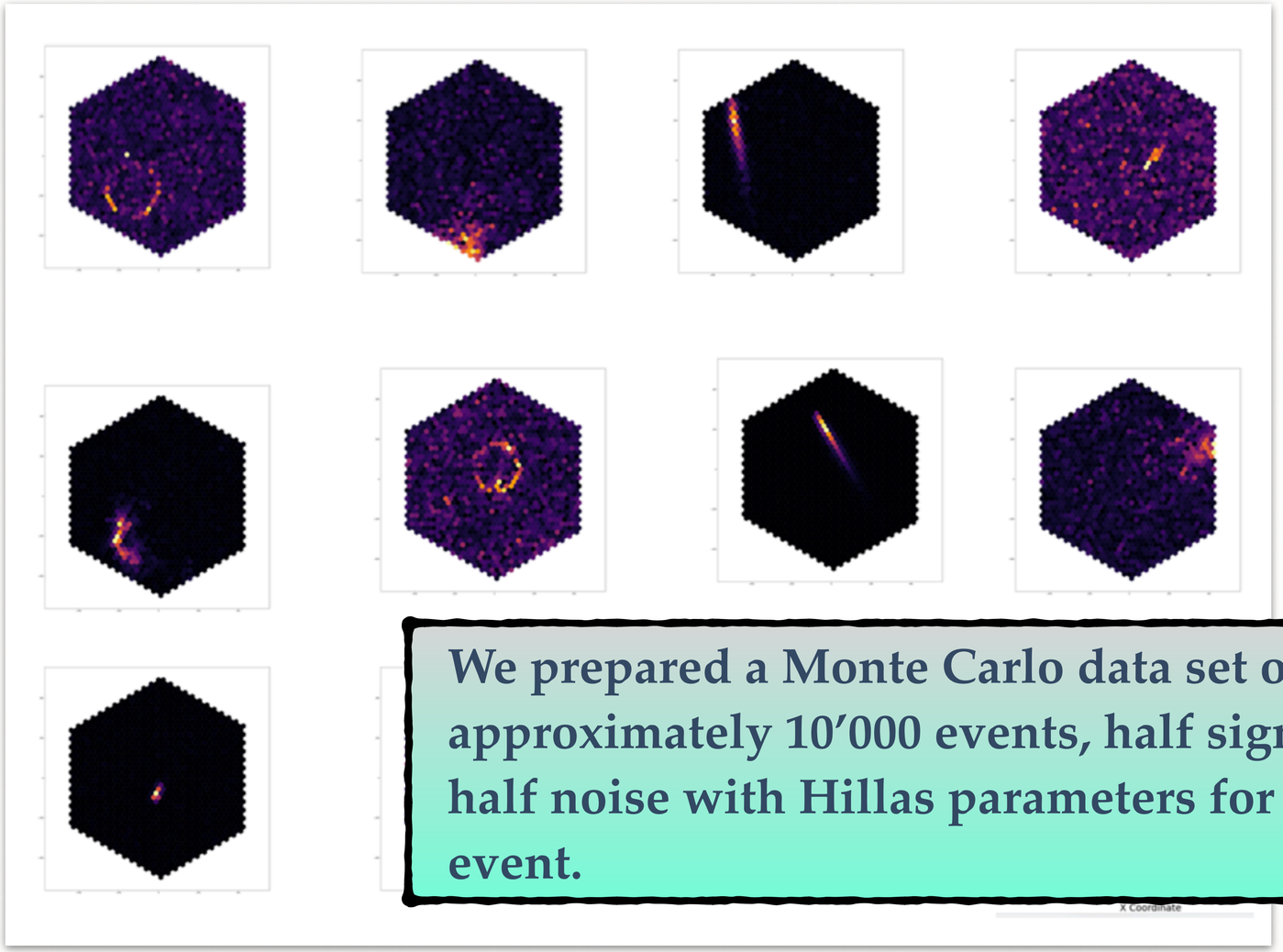
- Based on the parameters of these ellipses (**width, length, number of "islands",** etc.), we can retrieve an indication of the nature of the event.



For instance: proton events tend to have more "islands" and "rounder elipses"

THE PROJECT

Mission for the students: differentiate protons from gamma rays and estimate the effectiveness of their selection.

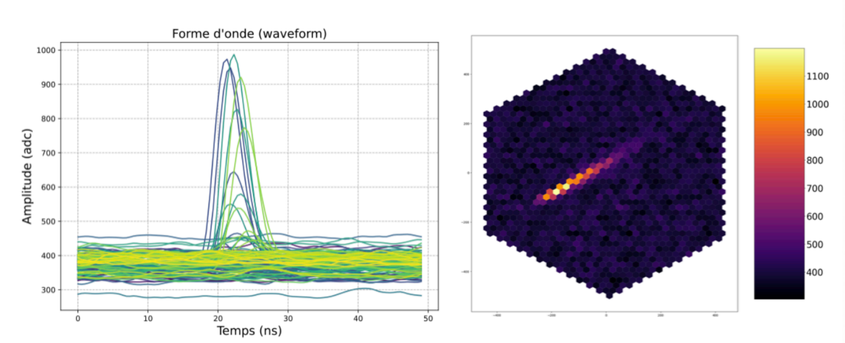


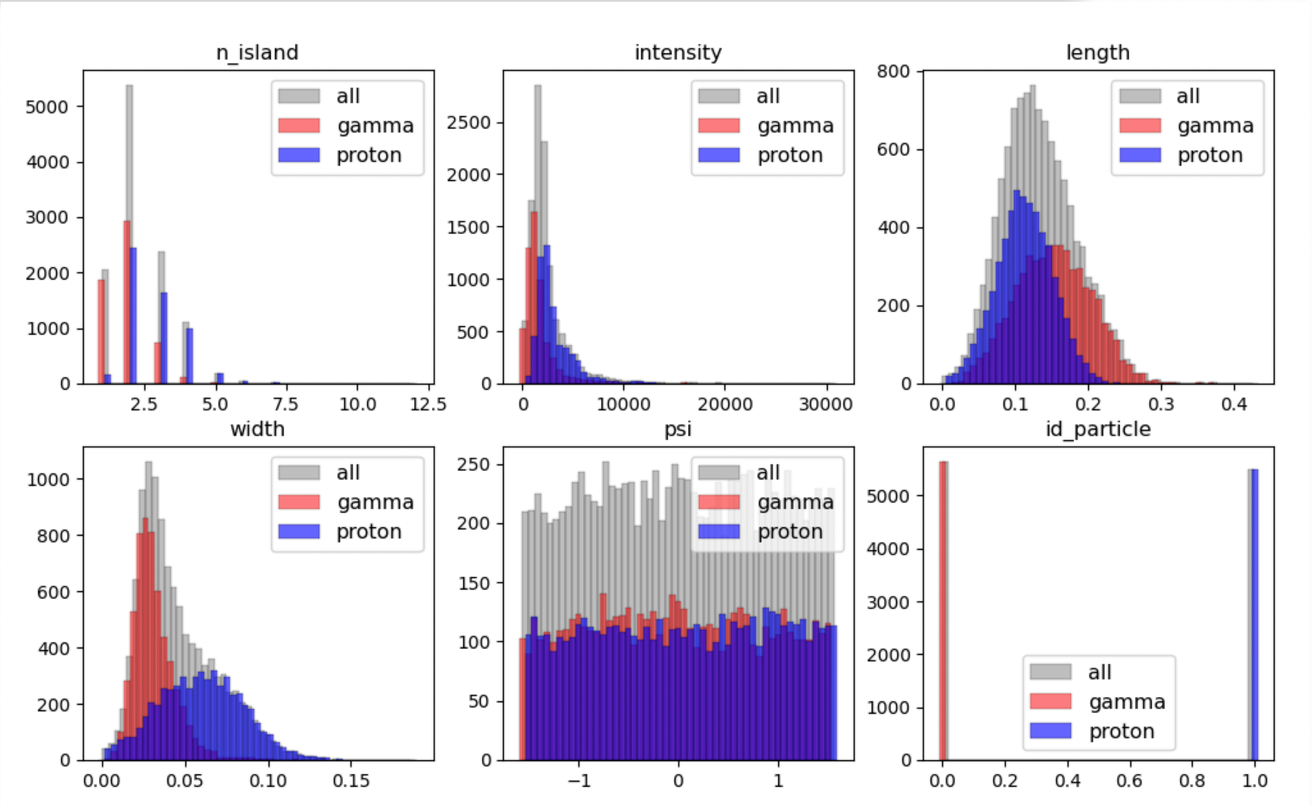
We prepared a Monte Carlo data set of approximately 10’000 events, half signal and half noise with Hillas parameters for each event.

THE PROJECT

Portion of CSV file with Hillas parameters

id_event	E [GeV]	imp_x [m]	imp_y [m]	n_island	intensité	length [m]	phi [rad]	psi [rad]	width [m]	id_particle
1.00	339.95	0.48	544.32	2.00	377.59	0.23	-1.44	1.37	0.01	0.00
2.00	256.45	447.06	408.24	2.00	257.42	0.19	2.92	0.61	0.02	0.00
3.00	343.15	535.02	-50.66	1.00	394.18	0.22	2.85	-0.12	0.01	0.00
4.00	1218.24	74.30	-4.41	6.00	1539.82	0.57	0.37	-1.17	0.08	1.00





plot the distributions, understand them, make the right cuts, optimise them, calculate efficiency and purity of sample

THE PROJECT : MAKE IT ACCESSIBLE

The programming and data analysis must be engaging:

- Avoid time lost on technical issues (to prevent demotivation), meaning no installation of libraries or specific environments.
- No object oriented, classes, etc.

From CTAPipe …

```
In [3]: class Camera:
            def __init__(self, geo_file, pix_size):
                self.pix_size = pix_size
                self.geo_file = geo_file
                self.hexa_side = pix_size/(2.*np.cos(np.
                self.camera_dict = {'pix_id': [], 'pix_x

            def read_geometry(self):
                self.camera_dict['pix_id'], self.camera_

            def compute_vertices(self):

                off_x = self.pix_size / 2.
                off_y = off_x*2/np.sqrt(3.)
                for i in range(len(self.camera_dict['pix
                                   dict['pix_x'][i]
                                   dict['pix_y'][i]
                                   vert_x'].append([c
                                   vert_y'].append([c
```

Camera object creation

We create a camera object.

Then we execute the main functions to fill the dicti

```
In [5]: cta_camera = Camera('camera_geometry.
        cta_camera.read_geometry()
        cta_camera.compute_vertices()
        cta_camera.add_patches()
```

To simplified CSVs, which they can read with Python and Numpy

ShowEventInLoop.py

```
1  #importer les fonctions
2  import csv
3  import matplotlib
4  matplotlib.use('TkAgg')
5  import matplotlib.pyplot as plt
6  import numpy as np
7  import os
8  from matplotlib.collections import PatchCollection
9  from matplotlib.patches import RegularPolygon
10 import glob
11 from matplotlib.colors import LogNorm
12 import random
```

Shell

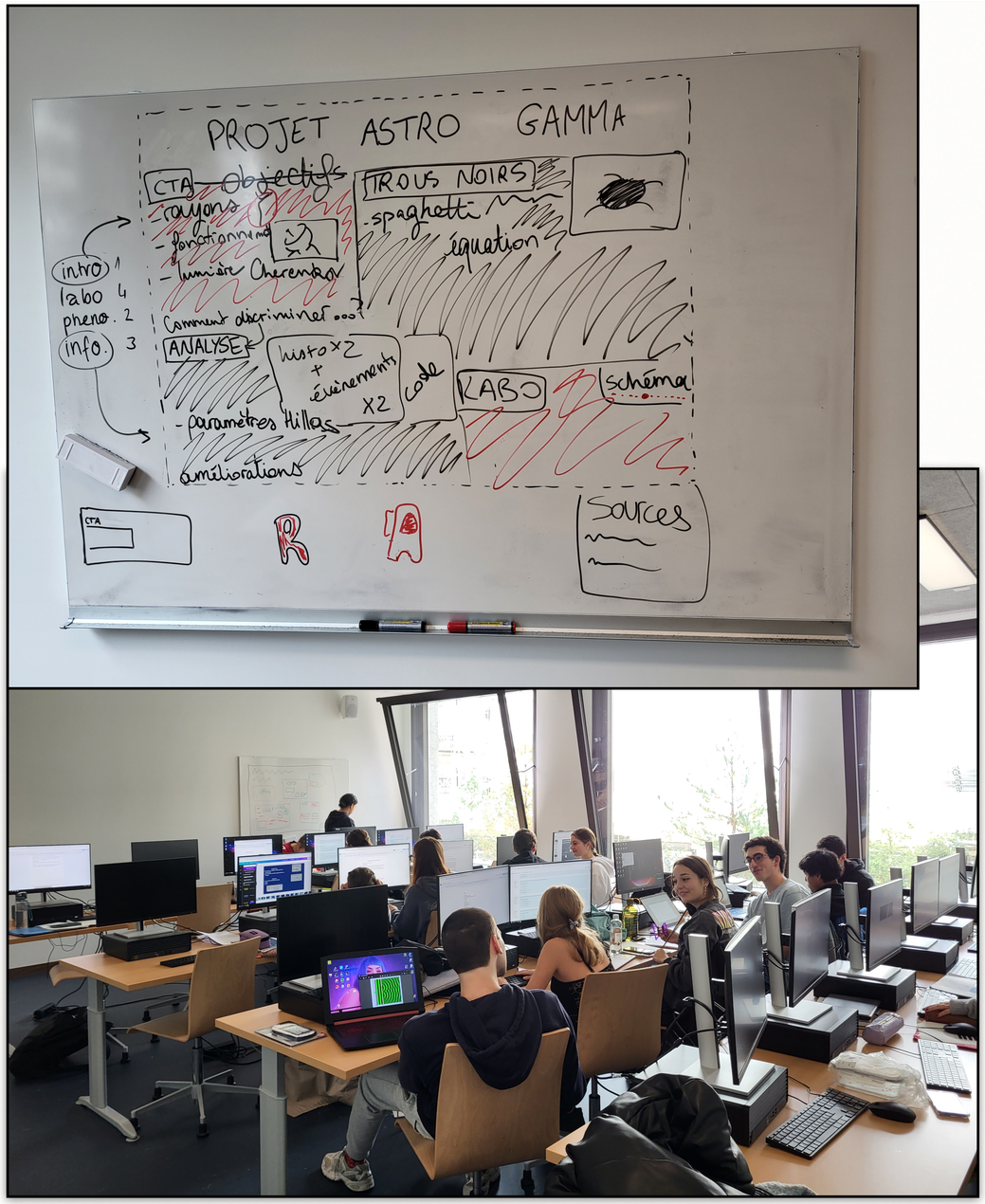
```
Python 3.10.11 (/Applications/Thonny.app/Contents/Frameworks/Python.framework/Versions/3.10/bin/python3.10)
>>>
```

```
#On ouvre le fichier CSV Hillas. Attention il doit être dans le même repertoire que ce fic
#hillas_file="Jour2-Hillas-New.csv"
hillas_file="Hillas-Para.csv"
print("opening:", hillas_file)
if  not  os.path.exists(hillas_file):
    print("le fichier: ", hillas_file," n'existe pas")
    exit(0)


header = np.loadtxt(hillas_file, dtype=str, delimiter=',',skiprows=0,usecols=range(0,10))
data = np.loadtxt(hillas_file, delimiter=',', skiprows=1, usecols=range(0,10))
```

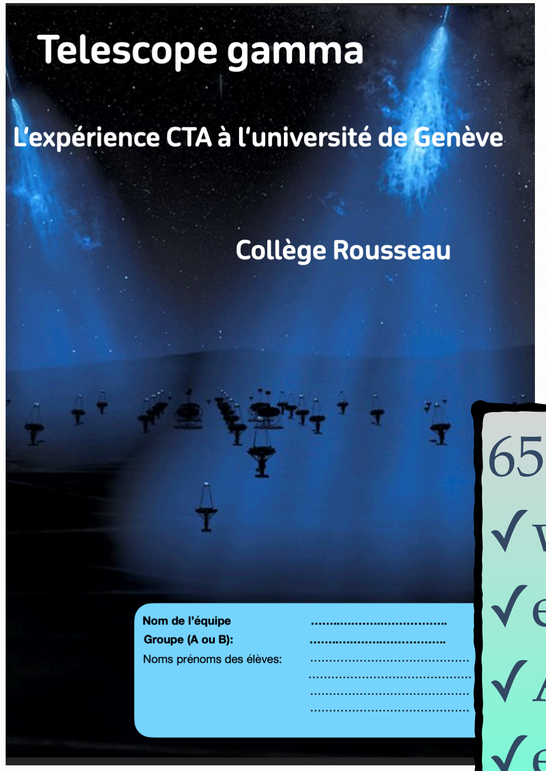
THE PROJECT: HOW IT UNFOLDED

- This project first took place from October 16 to 21, 2023.
- Will be held again this year October 14-18th 2024.



THE PROJECT: HOW IT UNFOLDED

- Day starts with a plenary where we present and focus on a physics topic
- Day wraps up with a meeting with the group leader (group leader changes every day)
- Rest the time, they work in autonomy (one lab on light interference during the week)



8 **PLANNING DE LA SEMAINE**

	Lundi		Mardi		Mercredi		Jeudi	
horaire de début	activité	salle	activité	salle	activité	salle	activité	salle
8:00	Plénière: *astrophysique et présentation de CTA (E. Lanciotti)*	salle de conférence	Apel Plénière : *que sont les rayons gammas?* *+ Introduction au protocole expérimental*	salle de conférence	Appel Plénière : *comment CTA détecte les rayons gamma ?* *S. Murphy*	salle de conférence	Plénière : *les photons et leur interaction avec la matière* *+ Introduction au protocole expérimental*	salle de conférence
8:30	Plénière : *aspects pratiques et évaluation (S. Murphy)*	salle de conférence						
9:00			**Groupe A.** Activité pratique sur interférence de la lumière. **durée 1:30** **Groupe B:** travail en autonomie en salle info	103 et 133	Travail en autonomie		**Groupe B.** Activité pratique : la spectroscopie. **Groupe A:** travail en autonomie en salle info	127,132 103, 133
9:30	Travail en autonomie	127,132 103, 133						
10:30			**Groupe B.** Activité pratique sur interférence de la lumière. **durée 1:30.**	103 et 133			**Groupe A.** Activité pratique : la spectroscopie. **Groupe B:** travail en autonomie en salle info	127,132 103, 133
							Travail en autonomie	127,132 103, 133
							15:10-15:55 Test écrit individuel	**127,132**
							Pour ceux qui le souhaitent: réunion des représentants de groupes avec le responsable de projet	103

65 page Activity Log book for each student :

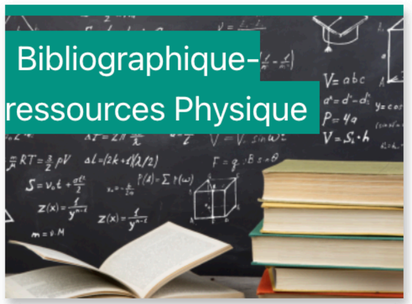
✓what they have to do each day

✓explanation of the physics

✓Assessment Grid

✓etc.

THE PROJECT: HOW IT UNFOLDED



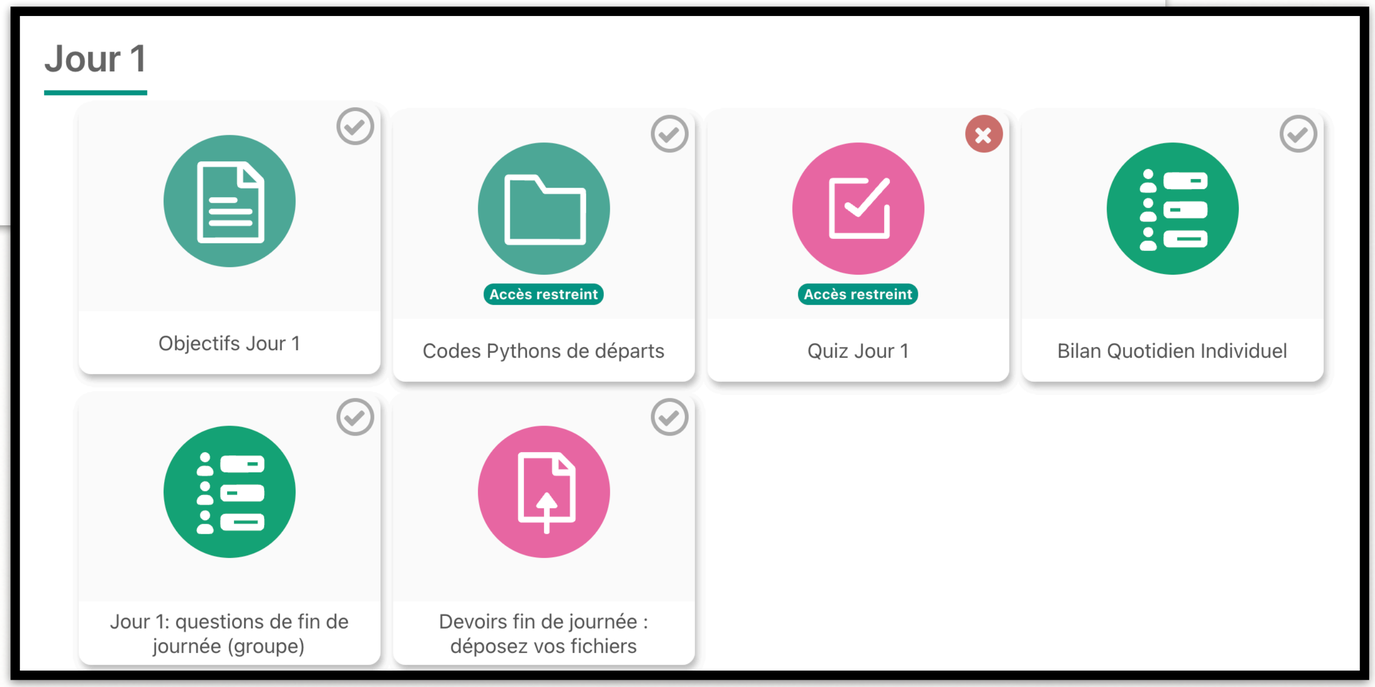
A Moodle page that includes:

✓starter codes

✓daily quizzes

✓Individual daily review and comments

✓daily submissions

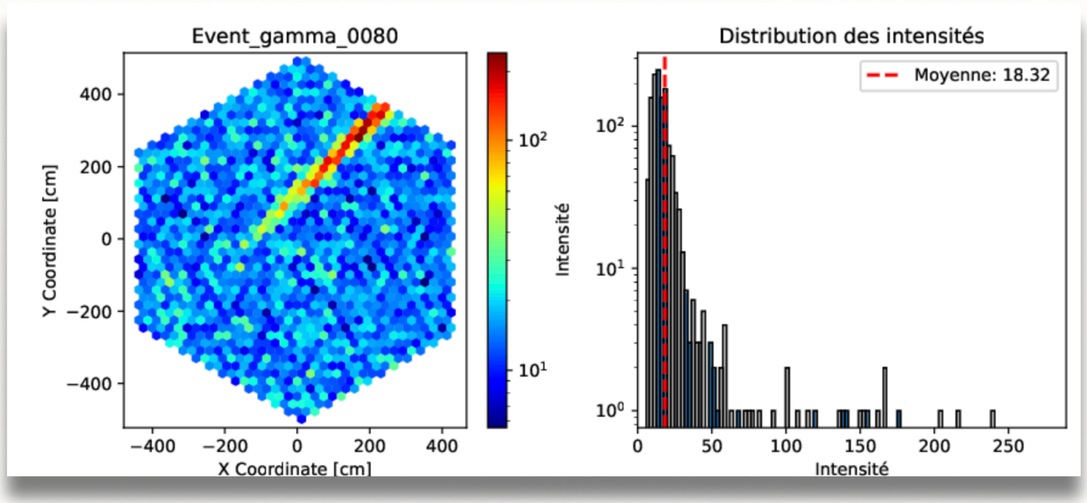


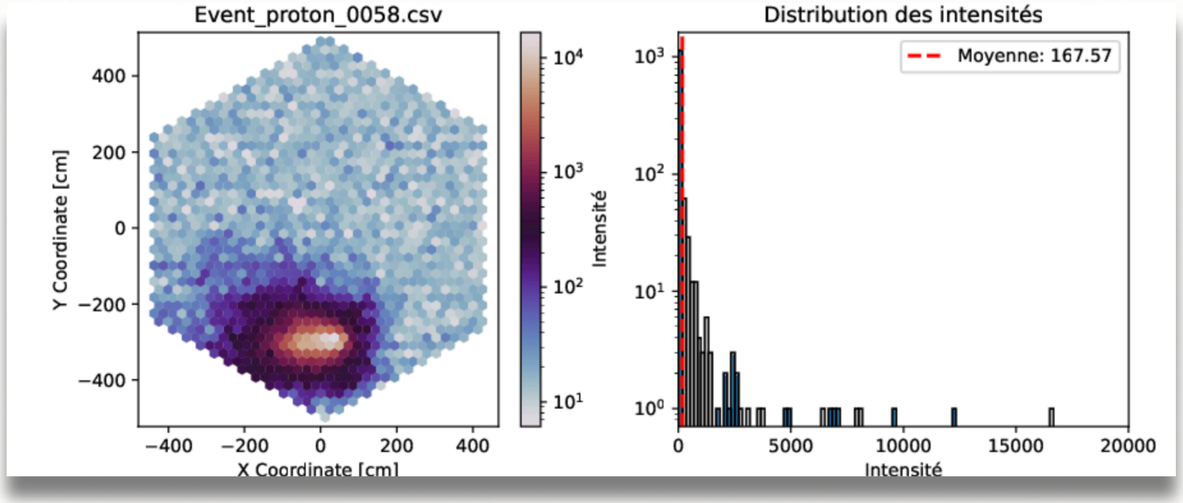
THE PROJECT: HOW IT UNFOLDED

- **They have four days:** Monday, Tuesday, Wednesday, Thursday, free to go where they want in the school.
- They have two hands-on physics experiments related to the topic (interference and spectroscopy).
- There is an **individual written assessment on Thursday**.
- They must submit a **scientific poster and give an oral presentation on Friday**. The poster includes:

  1. A summary of their analysis
  2. A physics topic of their choice related to CTA
  3. A summary of one of the two labs.

EXAMPLES OF STUDENT OUTPUT





Example of student answers on Moodle (translated from french). Question **how do gamma events from proton events ?**

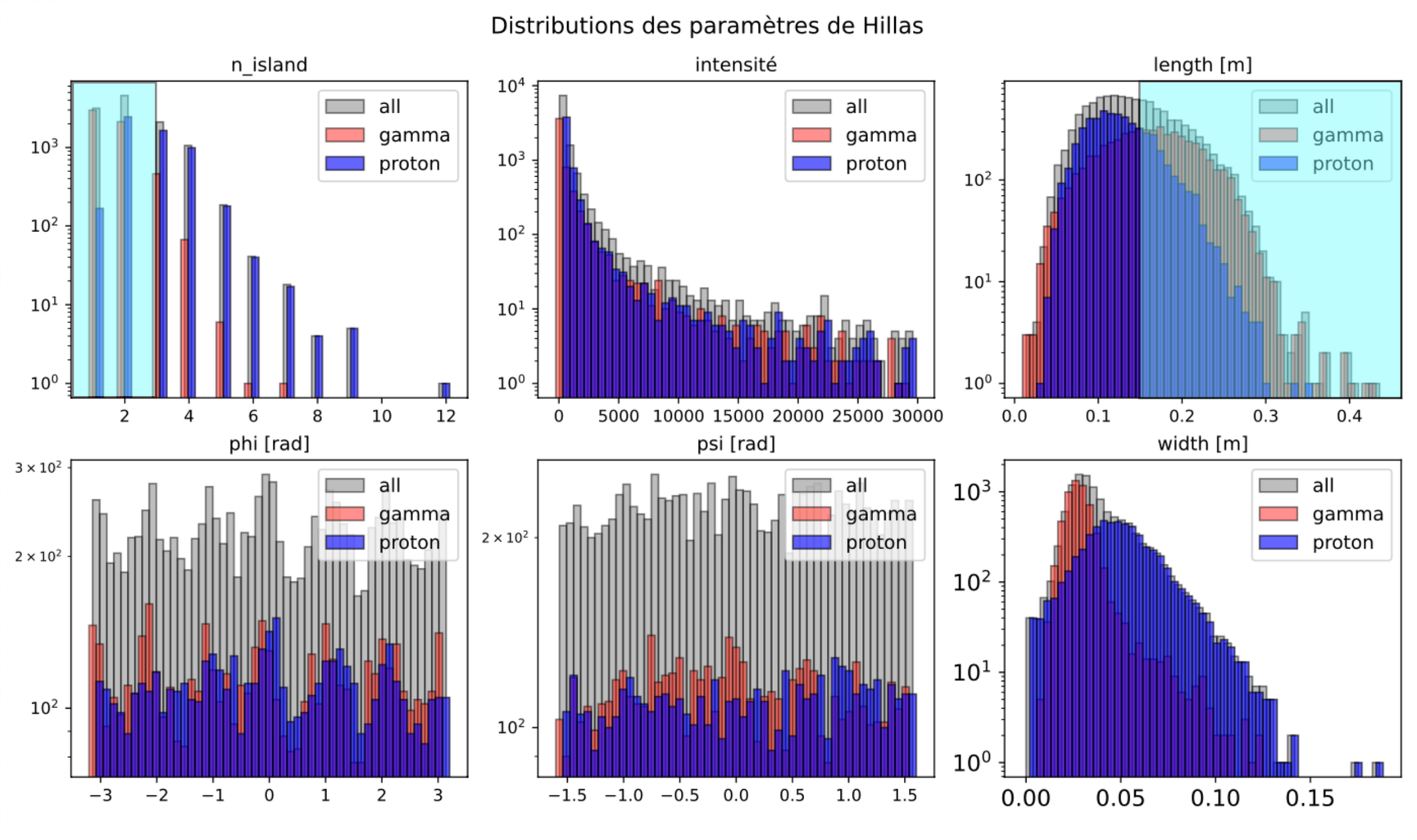
Gamma events form long, narrow streaks, while photon events create round or circular shapes.

Schematically, gamma events resemble comets (small streaks of light).

In general, the average intensity of gamma events is lower than the intensity of proton events.

EXAMPLES OF STUDENT OUTPUT

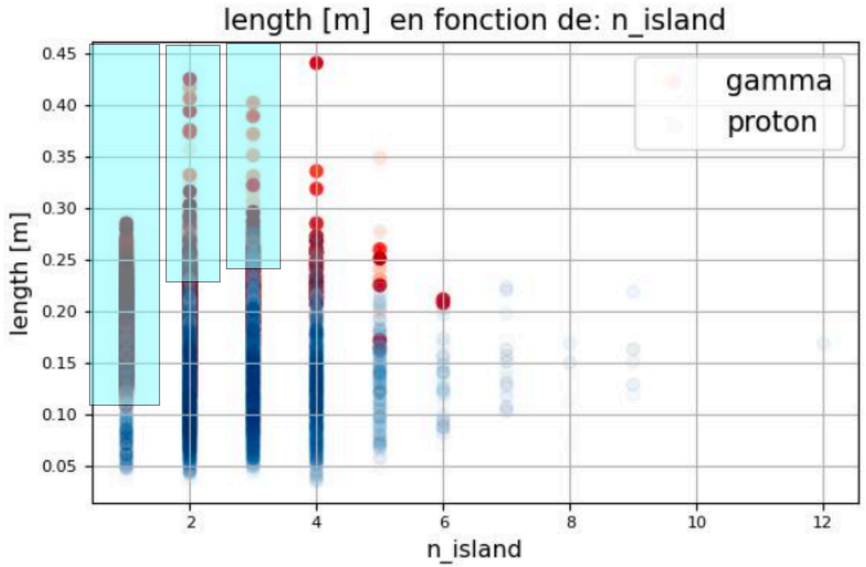
```
data_cut=data[(length > 0.15) & (n_island < 3)]
```



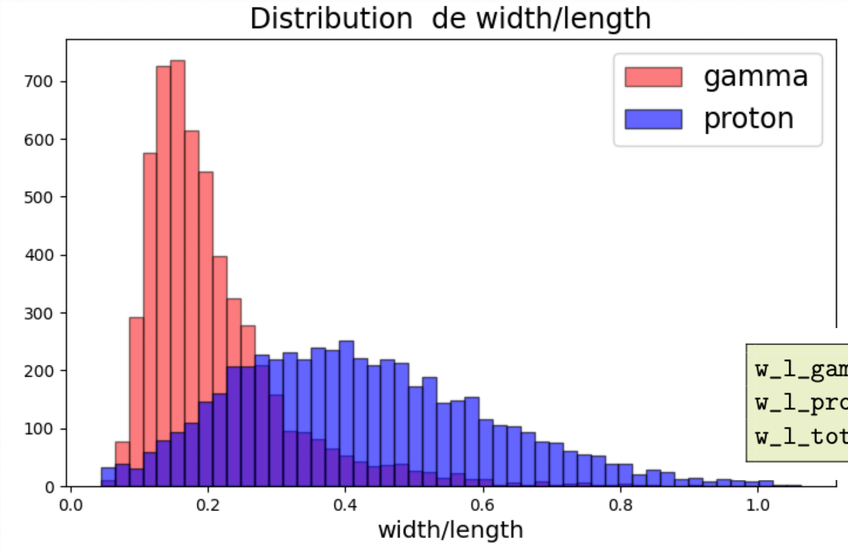
Très facile d'effectuer des coupures avec des masques

EXAMPLES OF STUDENT OUTPUT

```
if ((n_island == 1) & (length > 0.1)) | ((n_island == 2) & (length > 0.2)) | ((n_island ==
    3) & (length > 0.25)):
```



Looking at one parameter in relation to another and optimizing the cutoffs.



Creating new variables to improve discrimination.
Width/length is interesting because it gives the “roundness” of the ellipse

```
w_l_gamma=data_gamma[:,9]/data_gamma[:,6] # numpy array de w/l pour gamma
w_l_proton=data_proton[:,9]/data_proton[:,6] # numpy array de w/l pour protons
w_l_tot=data[:,9]/data[:,6]# numpy array de w/l pour tout
```

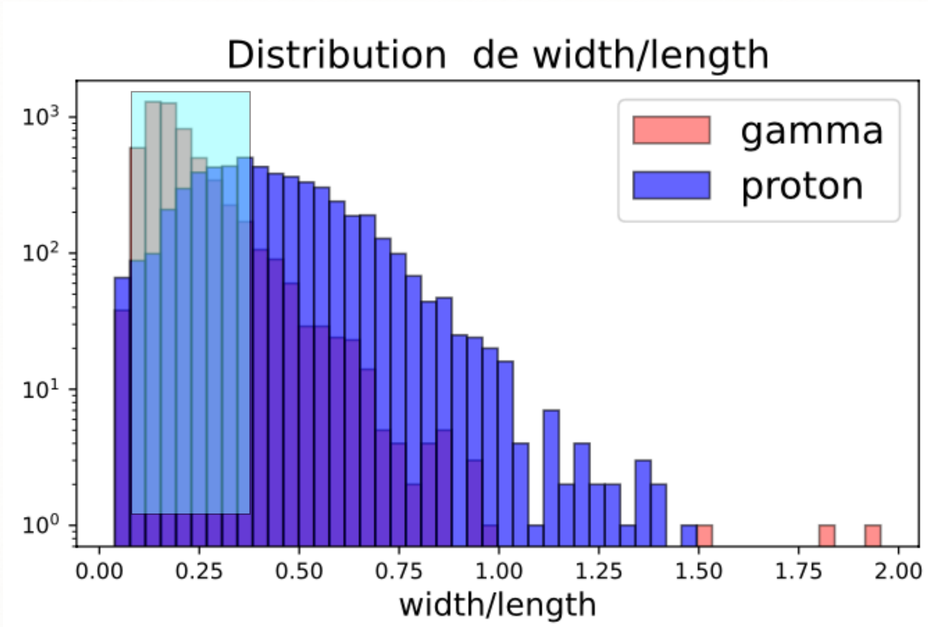
EXAMPLES OF STUDENT OUTPUT

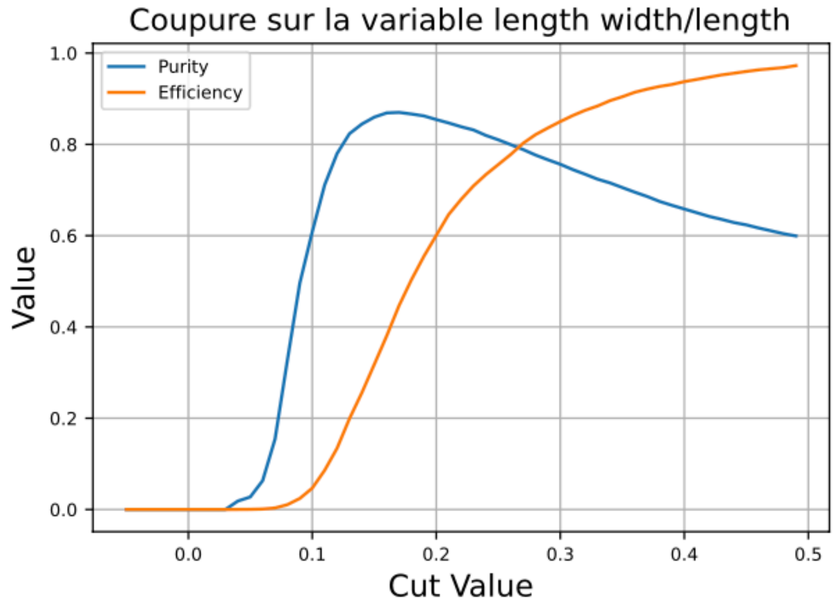
They have to calculate and submit the **purity and efficiency** of their selection.

$$P = \frac{N^{\gamma}_{sel}}{N^{\gamma+p}_{sel}} \qquad E = \frac{N^{\gamma}_{sel}}{N^{\gamma}_{TOT}}$$

What is the purity of our signal selection

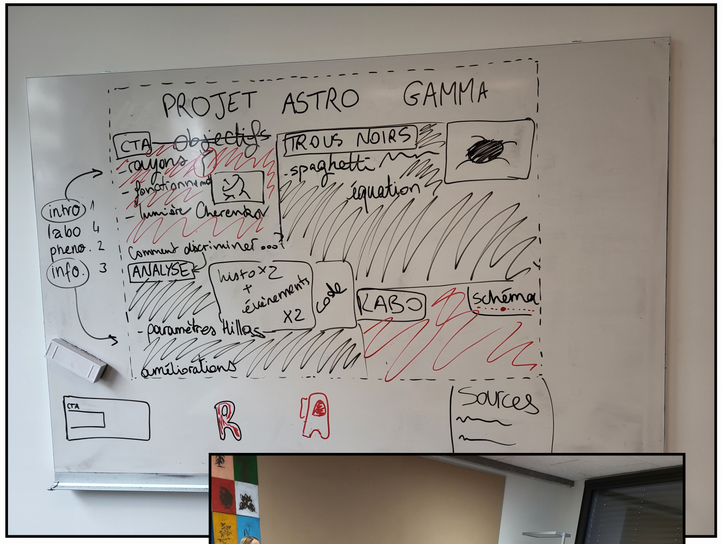
How we efficiently reject the background

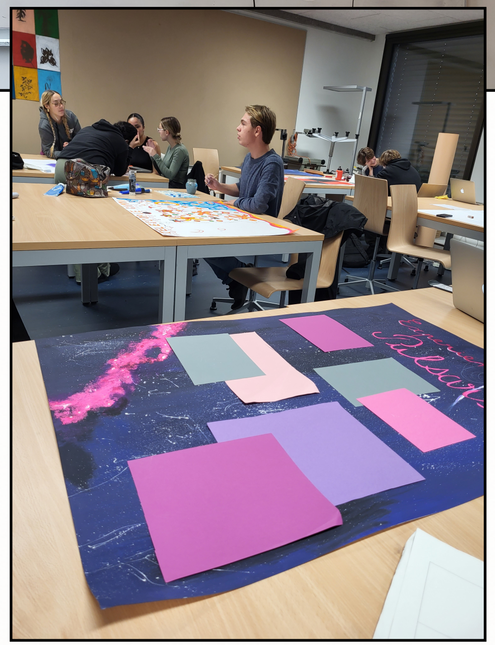


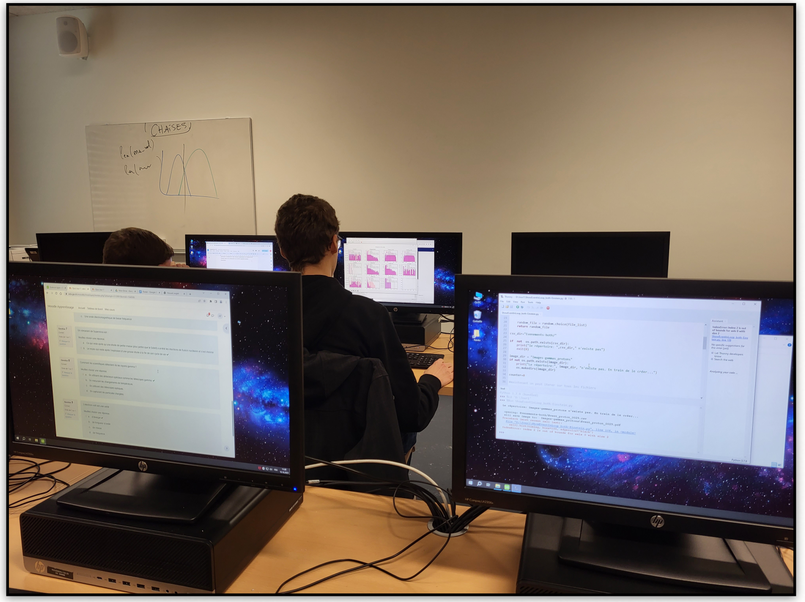


PICTURES


TELESCOPE GAMMA


PROJET ASTRO GAMMA
CTA
Objectifs
rayons
fonctionnement
lumière Cherenkov
TROUS NOIRS
spaghetti
équation
intro
labo
pheno.
info.
Comment discriminer ...
ANALYSE
histo x2
+
événement
X2
code
RABO
schéma
paramètres Hillas
améliorations
CTA
R
Sources




CHAISES

EXAMPLES OF STUDENT OUTPUT

A wide range of new skills acquired in computer science !

Equipe Meitner.

```
# Critères de sélection pour les coupures
# selection et affichage des particules qui ont maximum 3 iles et une largeur max de 0.05
data_select_1=data[(nb_island<=3) & (width<=0.05)]
nb_data_select_1=len(data_select_1)
print("Il y a ", nb_data_select_1,"particules qui ont maximum 3 iles et une largeur max de 0.05.")

# selection et affichage des particules GAMMA qui ont maximum 3 iles et une largeur max de 0.05
data_select_gamma_1=data[(nb_island<=3) & (width<=0.05) & (id_particle==0)]
nb_data_select_gamma_1=len(data_select_gamma_1)
print("Il y a ", nb_data_select_gamma_1,"particules GAMMA qui ont maximum 3 iles et une largeur max de 0.05.")

# selection et affichage des particules PROTONS qui ont maximum 3 iles et une largeur max de 0.05
data_select_proton_1=data[(nb_island<=3) & (width<=0.05) & (id_particle==1)]
nb_data_select_proton_1=len(data_select_proton_1)
print("Il y a ", nb_data_select_proton_1,"particules PROTONS qui ont maximum 3 iles et une largeur max de 0.05.")
```

Equipe Bell.

```
#Tous les events avec une longueur de plus que 1.160 sont des protons d'après la première figure
for i in range(len(data)):
    if length[i] > 1.16:
        data_removed.append(data[i,:])
        id_removed.append(data[i,0])

#Tous les events avec un n_island plus grand que 3 sont des protons avec quelques gammas d' après la première figure
for i in range(len(data)):
    if n_island[i] > 3:
        if data[i,0]  not in id_removed:
            data_removed.append(data[i,:])
            id_removed.append(data[i,0])

#Tous les events avec une width de plus de 0.0462 sont des protons avec un peu de gammas d' après la première figure
for i in range(len(data)):
    if width[i] > 0.0462:
        if data[i,0] not in id_removed:
            data_removed.append(data[i,:])
            id_removed.append(data[i,0])

```

Equipe Marie Curie.

```
#nb gamma avant coupure
hist, bin_edges, _ = plt.hist(width[id_particle == 0]/length[id_particle == 0], bins=600, color='red', alpha=0.3)
cutoff_index = np.searchsorted(bin_edges, 0.14)
gamma_before_cutoff = np.sum(hist[:cutoff_index])
print(f"Quantité de gamma avant la coupure: {gamma_before_cutoff}")
```

EXAMPLES OF STUDENT OUTPUT

A wide range of new skills acquired in computer science !

- Reading, understanding, and modifying a long code => trains **logic, attention to detail, and perseverance.**
- Using functions from external libraries => gives additional meaning to the concept and **utility of functions**.
- Use of Numpy with **Vectorized operations** on arrays, elementwise functions => functional approach, focusing on the “what” rather than the “how.”
- Boolean logic => **algorithmic thinking.**
- **Data visualisation, data science, and statistics** => essential for any future work.

Equipe Meitner.

```
# Critères de sélect
# selection et affic
data_select_1=data[(
nb_data_select_1=len
print("Il y a ", nb_
    0.05.")

# selection et affic
    0.05
data_select_gamma_1=
nb_data_select_gamma
print("Il y a ", nb_
    largeur max de 0

# selection et affic
     0.05
data_select_proton_1
nb_data_select_proto
print("Il y a ", nb_
    largeur max de 0
```

```
gamma_before_cutoff = np.sum(hist[:cutoff_index])
print(f"Quantité de gamma avant la coupure: {gamma_before_cutoff}")
```

EXAMPLES OF STUDENT OUTPUT

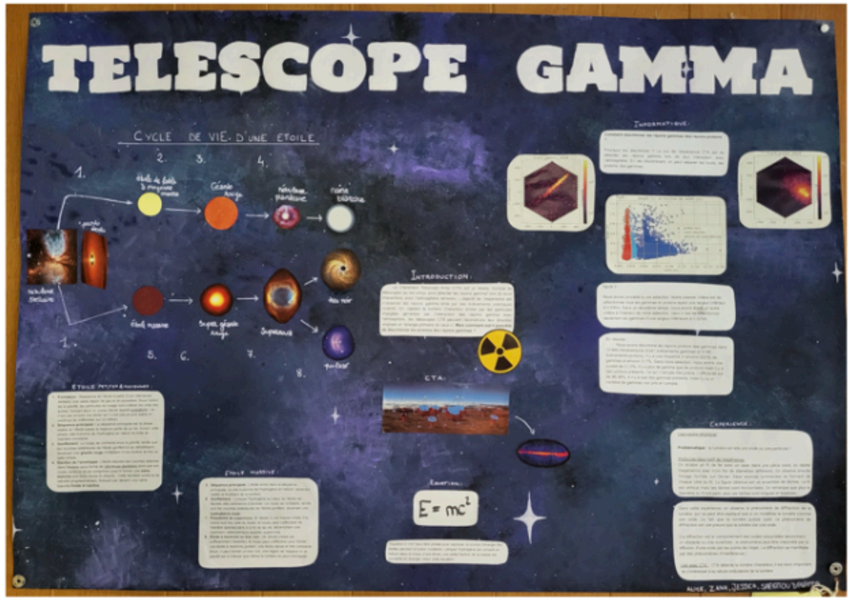
A wide range of new skills acquired in computer science !

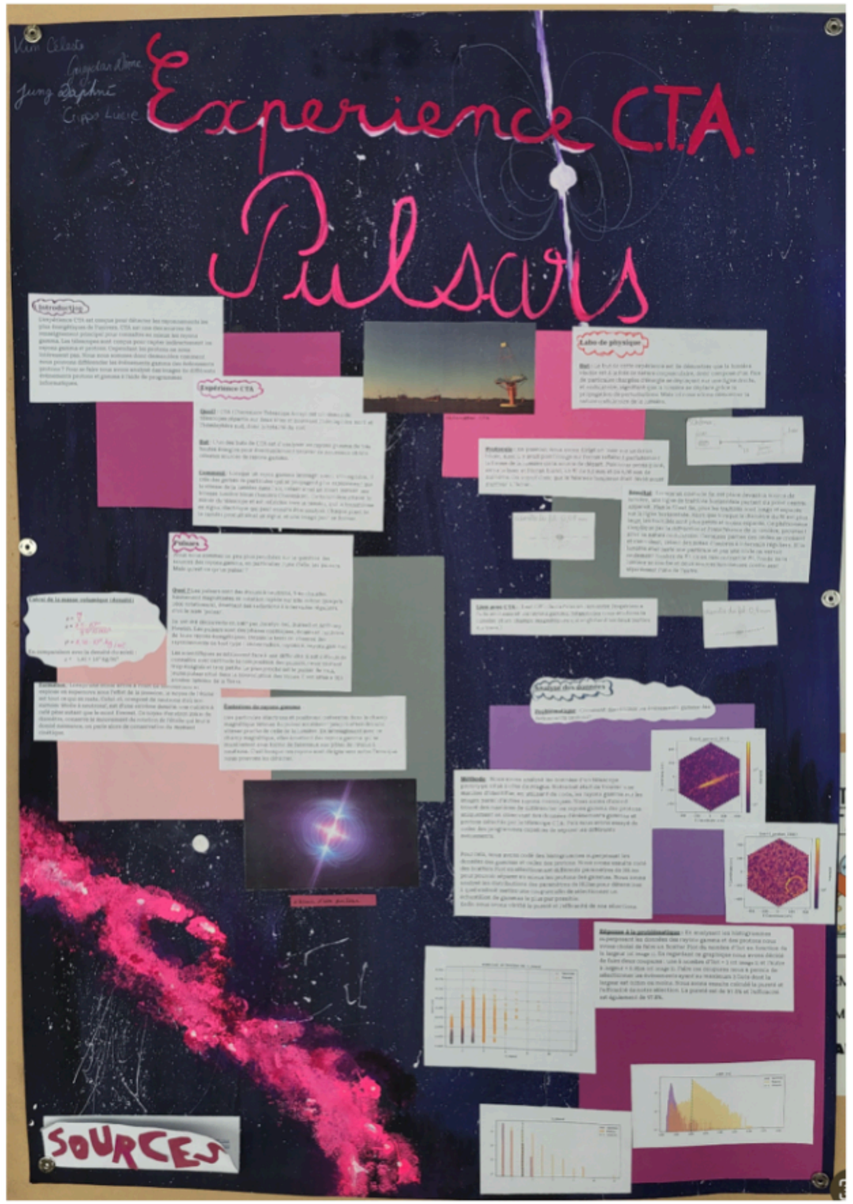
- They learnt all of this **while enjoying it and "having fun"**.
- A huge part of the job beforehand **was simplifying the framework so that the programming becomes engaging.**

"Immersion is when you learn without realising it."

Stephen Krashen - American linguist

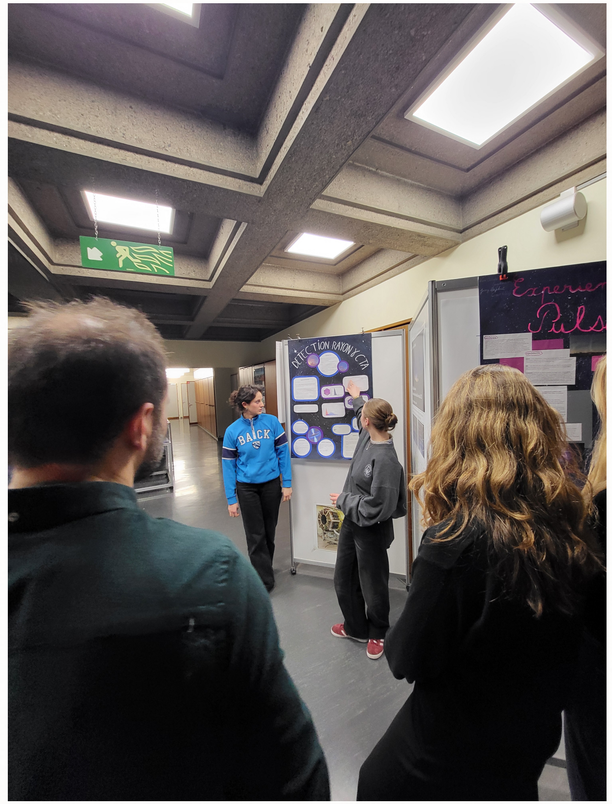
EXAMPLES OF STUDENT OUTPUT : POSTERS





THREE BEST POSTERS AND ANALYSES

- The three best posters were selected and displayed at UniGE.
- The corresponding teams visited the labs and clean rooms and had a discussions with the researchers.



THE PROJECT: STUDENT FEEDBACK

EXTENSIVE DATABASE OF COMMENTS, MOSTLY VERY POSITIVE

The freedom we are given is a **really enjoyable aspect** of the work (...)

The project is still very interesting and pushes us to discover many things and **finally see the usefulness of using Python**.

I find it quite hard, but not impossible because the teachers help a lot.

I felt like **we were united** (...) a sense of mutual support developed.

I really like physics, but computer science is hard.

I'm really interested in the topic, and I feel like I understand the physics side much better.

Given the success of previous edition, we are repeating the experience this year !

CONCLUSION

- This cross-disciplinary week was highly appreciated, **with high school students delivering analyses of university level**.

- The upcoming 2027 Maturité reform, which emphasises soft skills and cross-disciplinary learning, aligns well with the success of such immersive projects.

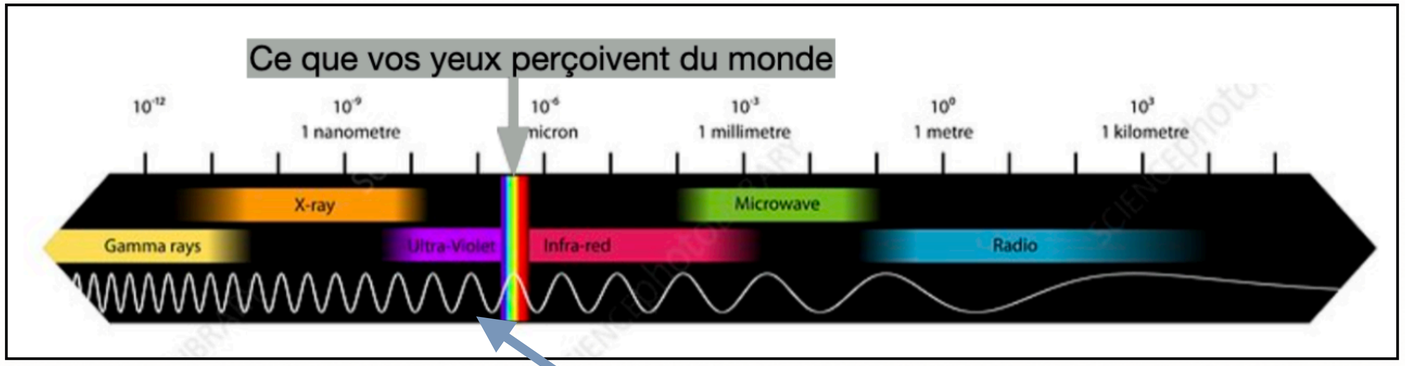
- This project highlights the **benefits of high school-university partnerships**. Strengthening these connections is essential for smoothing the transition to higher education.


Thanks for your attention

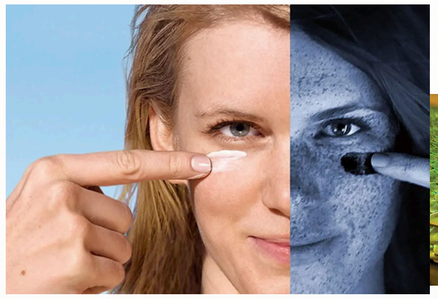
BACKUP

DIDACTIC TRANSPOSITION

Not "just programming"—focus on physics and explore new concepts in physics!

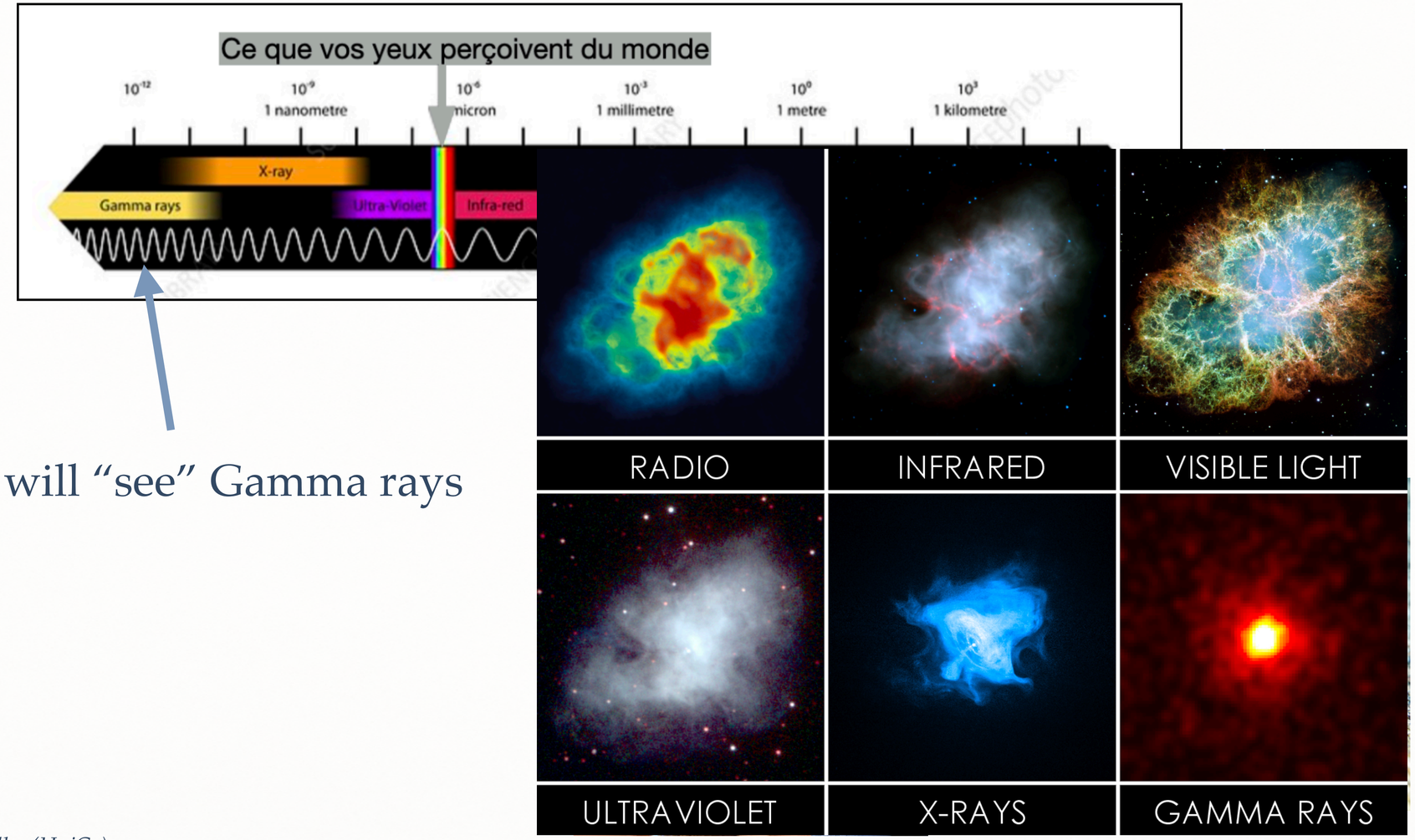


Some animals see UV



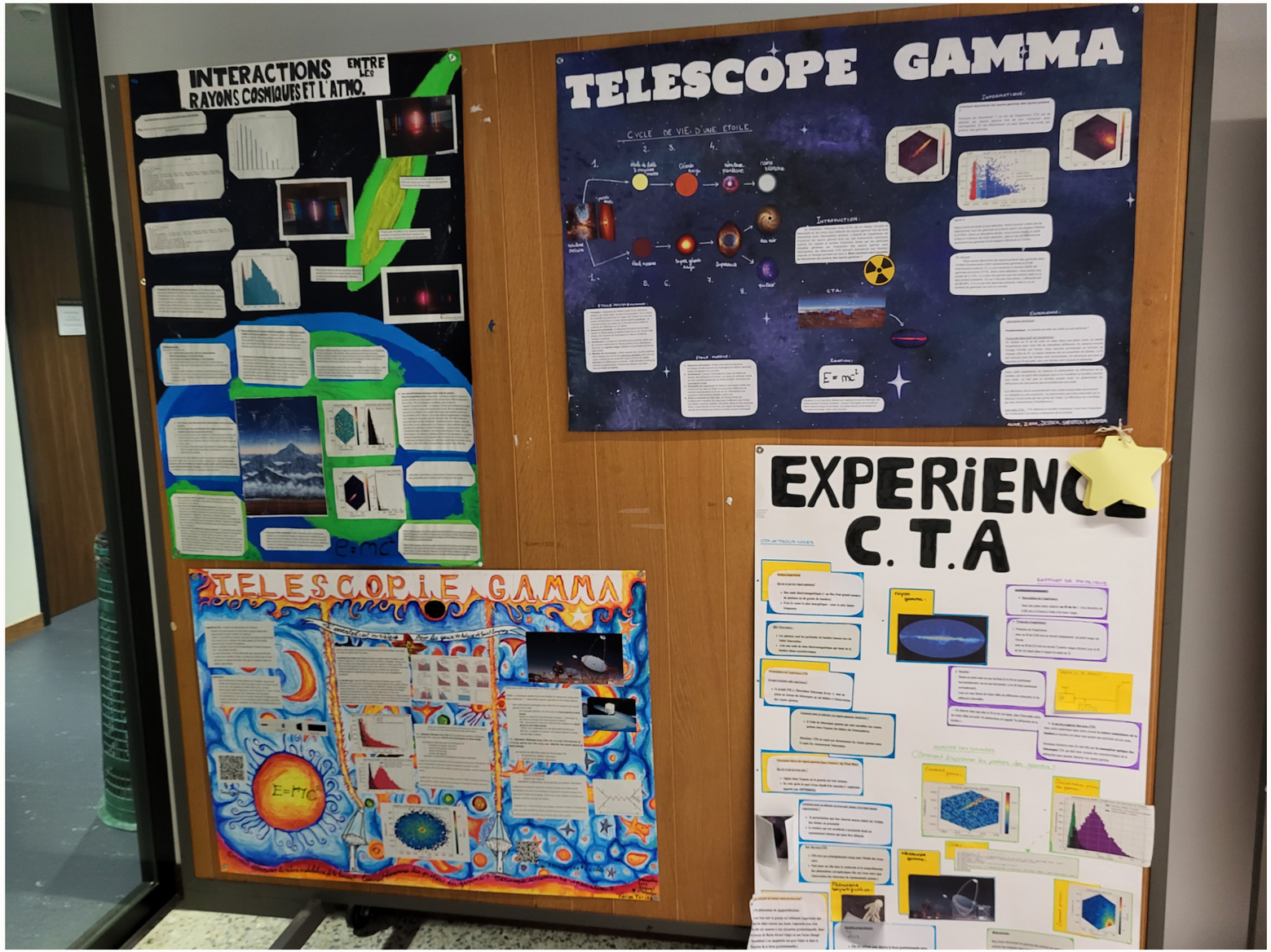
DIDACTIC TRANSPOSITION

Not "just programming"—focus on physics and explore new concepts in physics!



We will "see" Gamma rays

RESULTATS : posters élèves rendus



MOTIVATIONS : mot clés

Pendant une semaine les élèves vont effectuer un travail de chercheur en astrophysique.

Autonomie

Un projet, de l'aide mais c'est à eux de faire

Authenticité

Données d'une expérience existante.

Collaboration entre pairs

Indispensable de travailler et collaborer en équipe

PROJET TELESCOPE GAMMMA

Interdisciplinaire

Donner du sens aux apprentissages et mettre l'école en lien avec la réalité

Immersion

On apprend "sans s'en rendre compte"

Apprentissage par projet

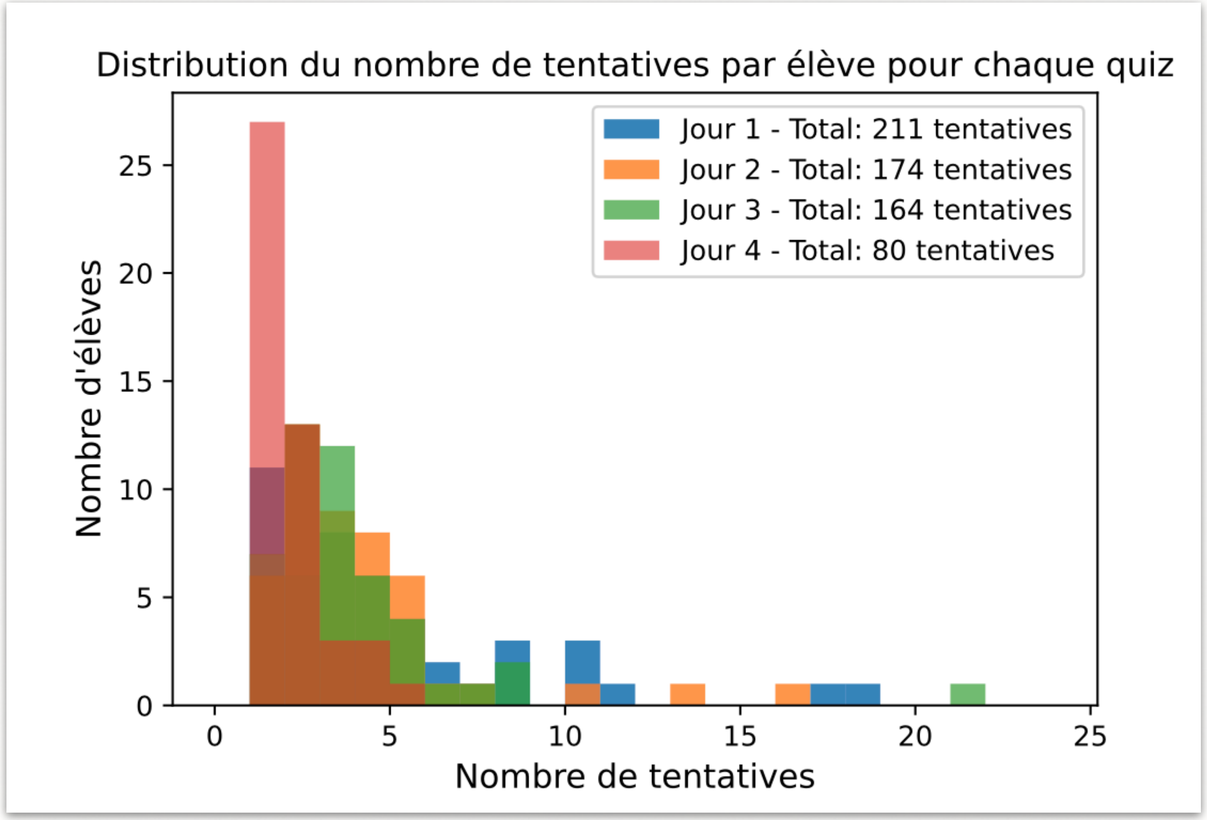
Longue durée

Thématique accrocheuse

RESULTATS : notes obtenues par équipes

Nom de l'équipe	Thème physique choisi	note totale sur la semaine (/6)
Oppenheimer	Interaction des rayons cosmiques dans l'atmosphère	4.4
Newton	Les trous noirs	5.5
Meitner	Les pulsars	6.0
Marie Curie	Créations des trous noirs	5.9
Hopper	Effet Cerenkov	4.4
Hamilton	Les supernovæ	5.3
Franklin	Cycle de vie d'une étoile	5.6
Feynman	Les pulsars	5.3
Einstein	Les supernovæ	5.7
Bell	Les pulsars	5.9
Ada	Tomber dans un trou noir	5.5
Turing	Détecteurs de rayons gamma terrestres et spatiaux	5.6

DEROULEMENT : tentatives aux quiz journaliers

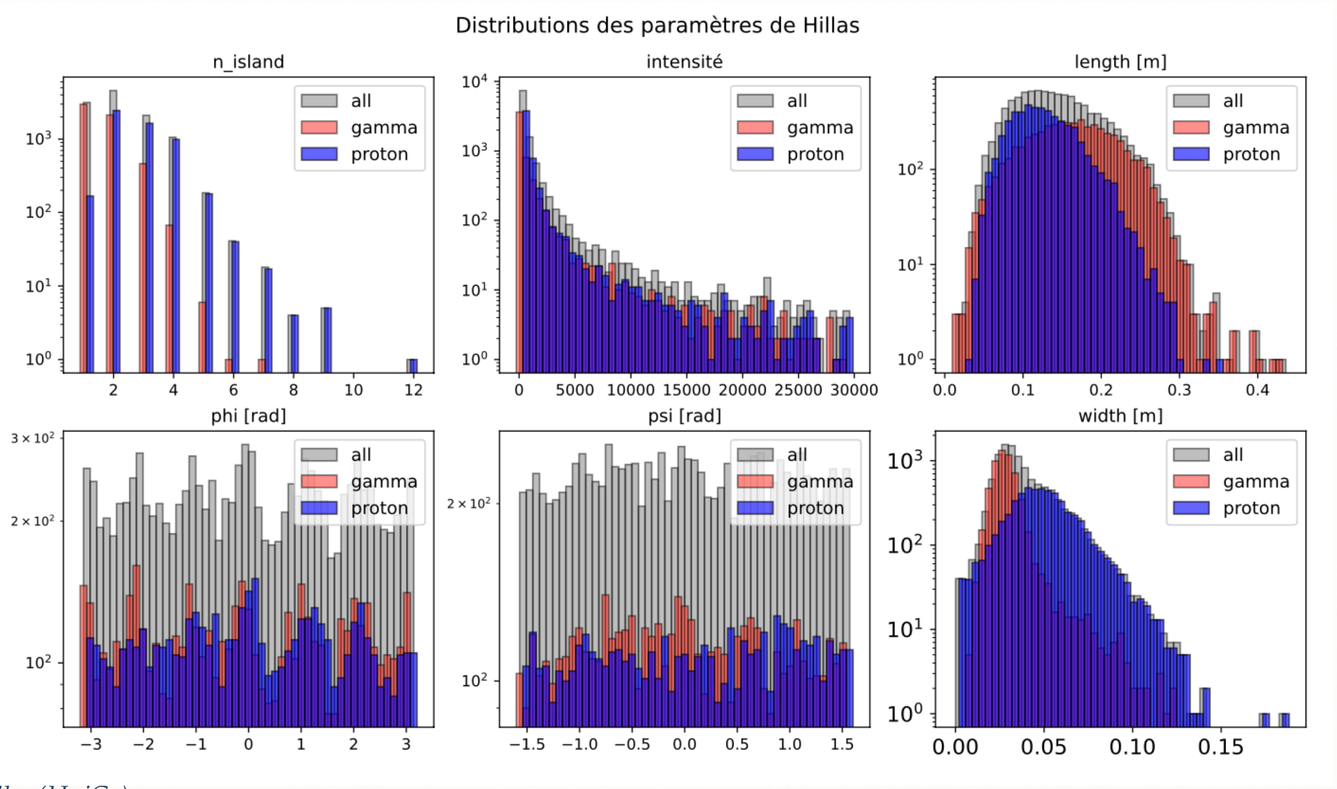


- 1er jour 211 tentatives (~4.5 tentatives par élève) quelques élèves à plus de 10 tentatives !
- Diminue ensuite.
- Tous les élèves ont obtenus les 85% du score chaque jour.

TRANSPOSITION DIDACTIQUE : ce qui est attendu des élèves

- **A partir du JOUR 2**: affichage et exploration des distributions des paramètres de Hillas programme AnaHillas.py

id_event	E [GeV]	imp_x [m]	imp_y [m]	n_island	intensité	length [m]	phi [rad]	psi [rad]	width [m]	id_particle
1.00	339.95	0.48	544.32	2.00	377.59	0.23	-1.44	1.37	0.01	0.00
2.00	256.45	447.06	408.24	2.00	257.42	0.19	2.92	0.61	0.02	0.00
3.00	343.15	535.02	-50.66	1.00	394.18	0.22	2.85	-0.12	0.01	0.00
4.00	1218.24	74.30	-4.41	6.00	1539.82	0.57	0.37	-1.17	0.08	1.00



La dernière colonne correspond au type de particule (0= proton, 1=gamma)