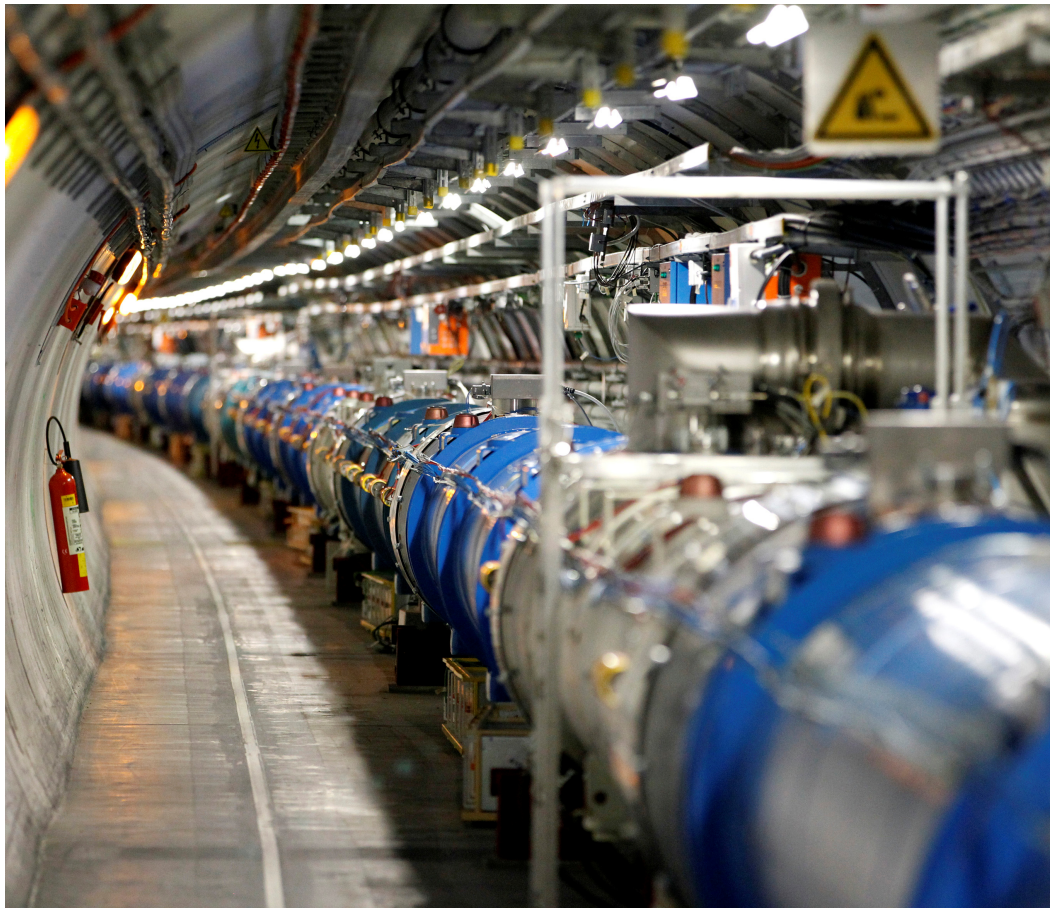
AN APPROACH TO TRACK RECONSTRUCTION FOR THE SCIFI TRACKER AT LHCb USING ANNs AND SPATIAL INDEXING

Gabriel Cammany Ruiz

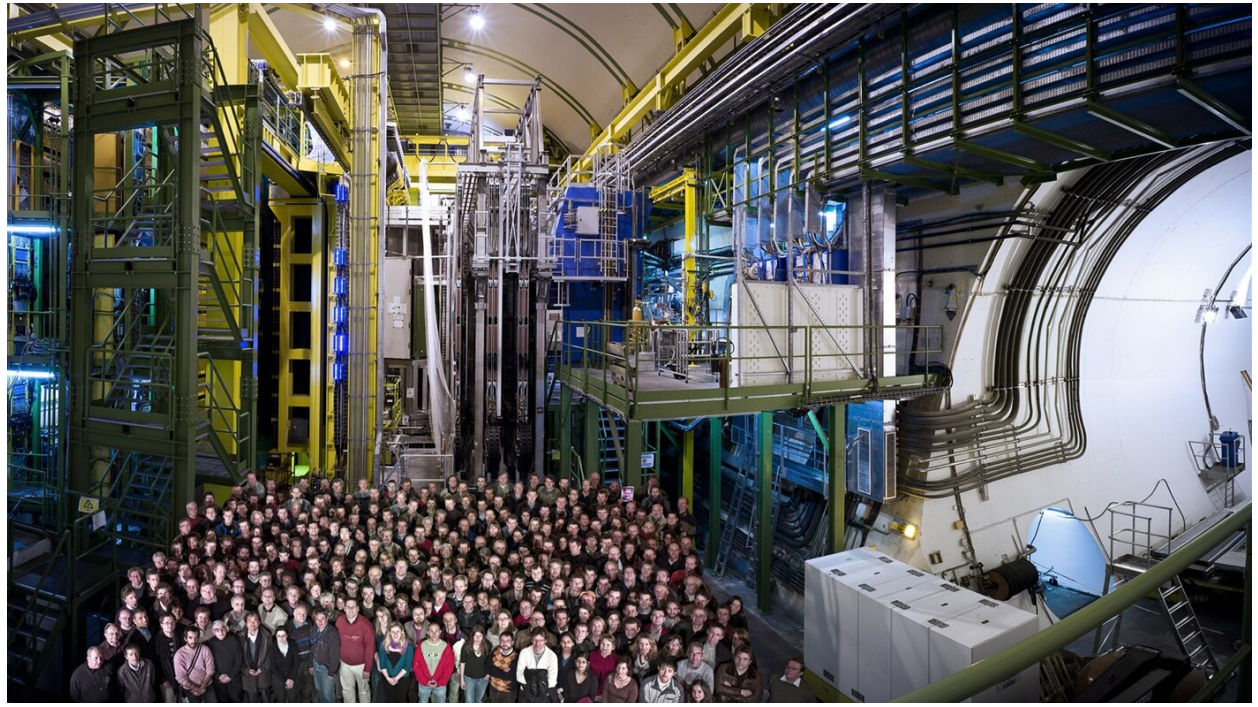
laSalle
Ramon Llull University

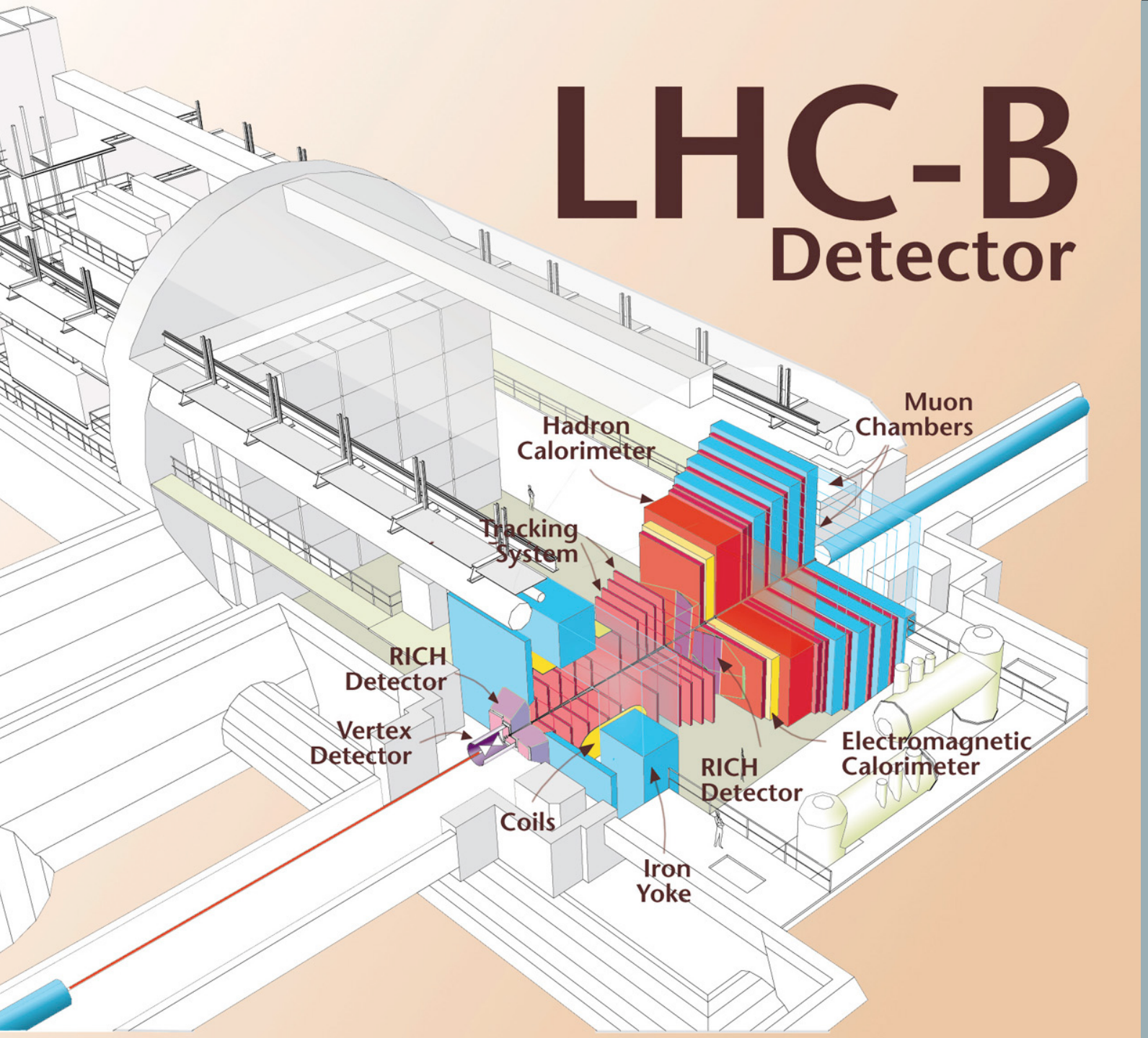
CERN AND LHC



- CERN – *Conseil Européen pour la Recherche Nucléaire.*
- The LHC is the largest and most powerful particle collider in the world.
- Tunnel 27 km in circumference located between the borders of France and Switzerland
- Main detectors are ATLAS, CMS, ALICE and LHCb.

LHCB DETECTOR





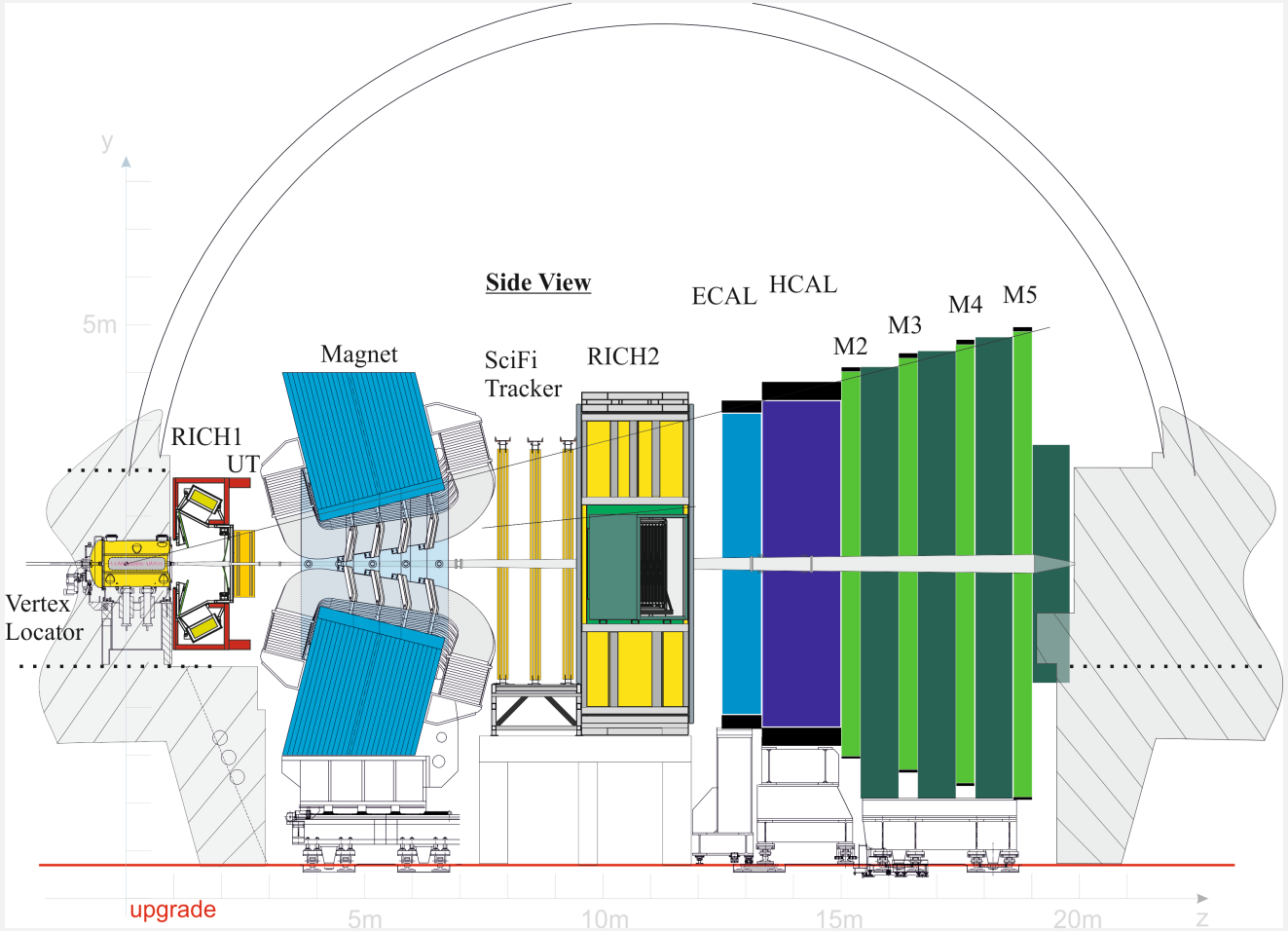
Explain the asymmetry between matter and antimatter observed in the Universe.

It uses a series of sub-detectors to detect the particles thrown forward by the collision in a direction 18 meters long.

An update will be made that will affect subdetectors and the trigger system.

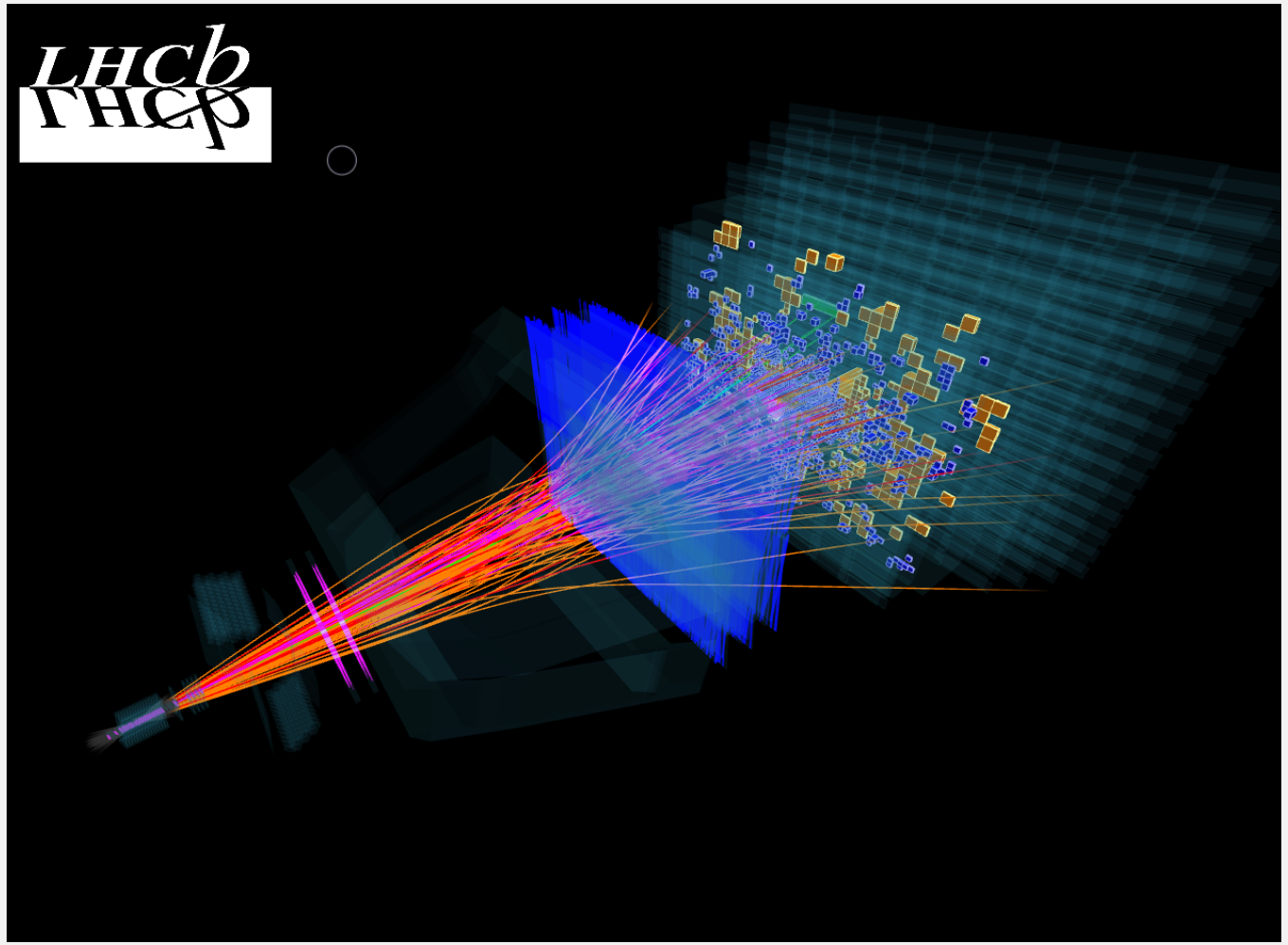
SISTEMES DEL LHCb

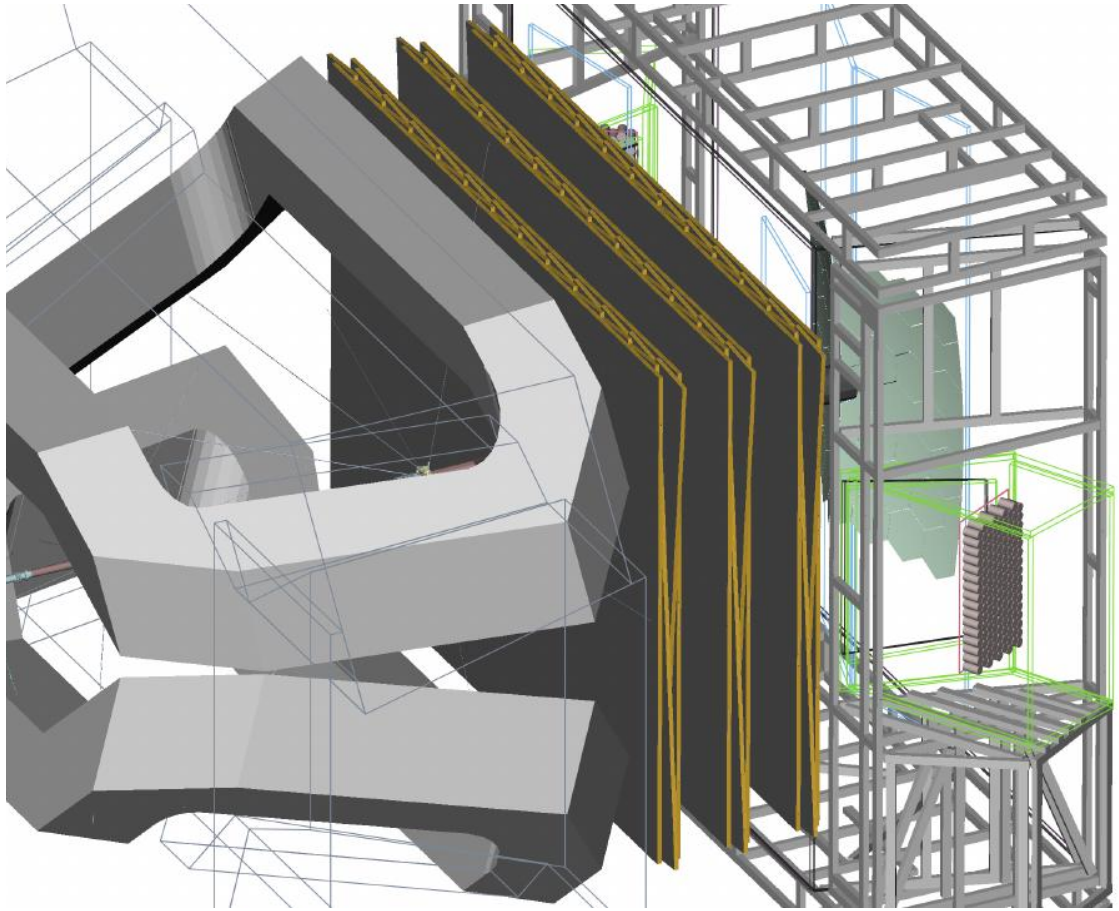
- **Tracking**
  - VELO + UT + Iman + SciFi
- **Identificació**
  - RICH1, 2
  - ECAL i HCAL
  - M1,2,3,4,5
- **Trigger**
  - HLT (High level trigger)



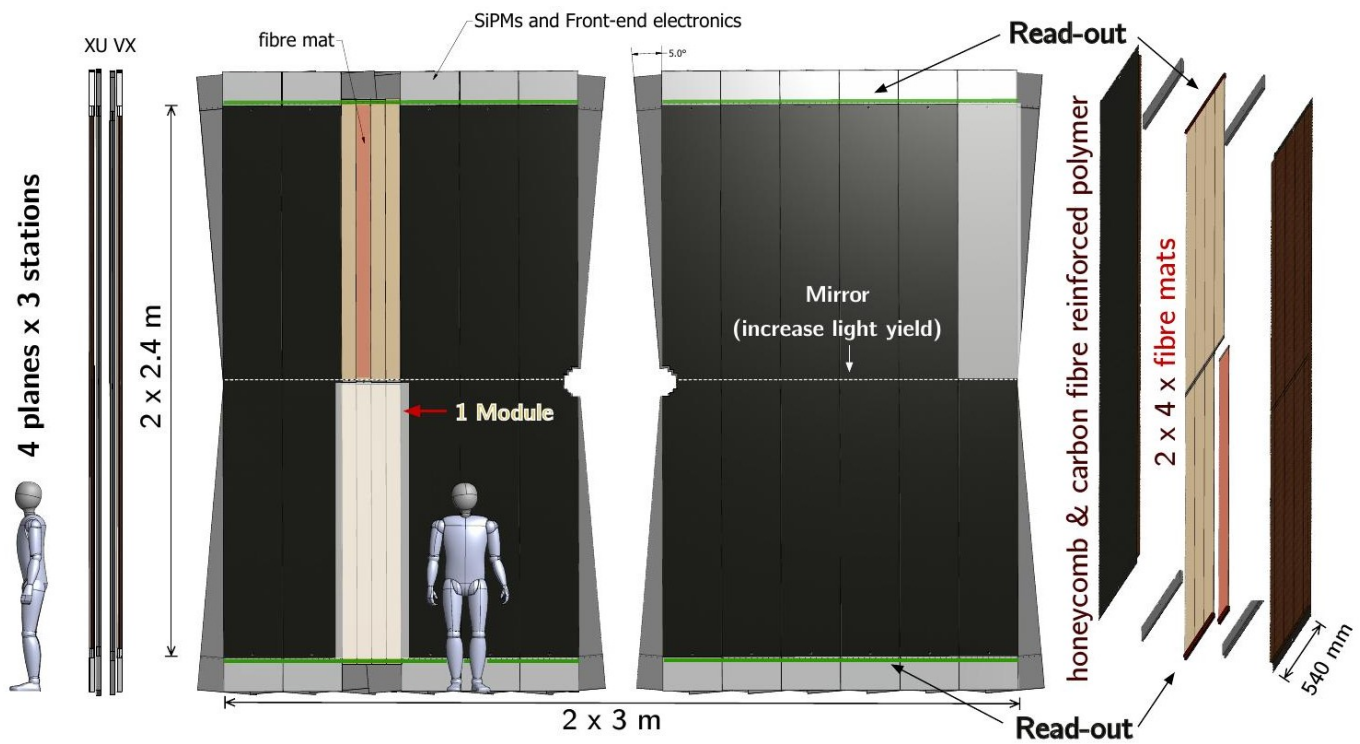
PARTICLE TRACKING SYSTEM

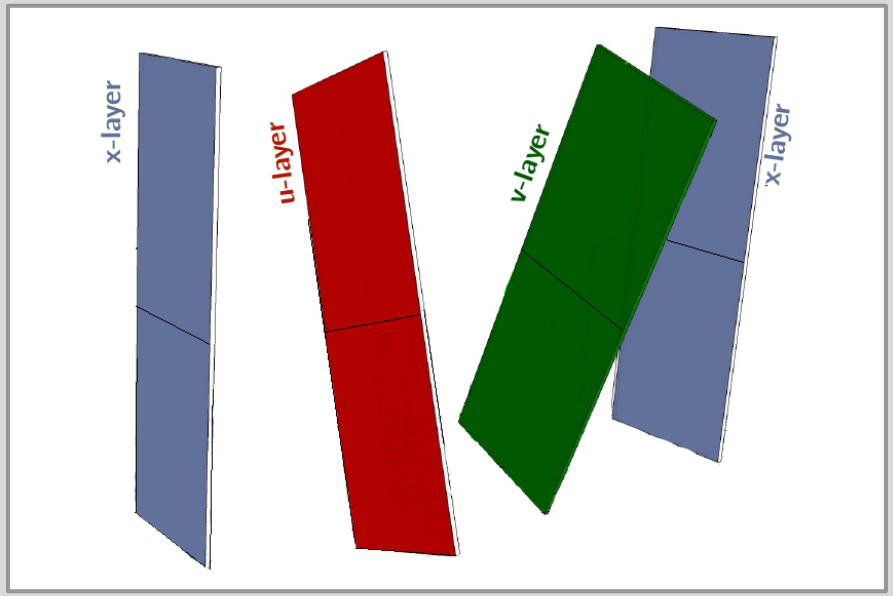
- Provide efficient reconstruction of charged particle trajectories.
- The reconstruction sequence consists of two parts.
  - Pattern recognition.
  - Validation of the results.

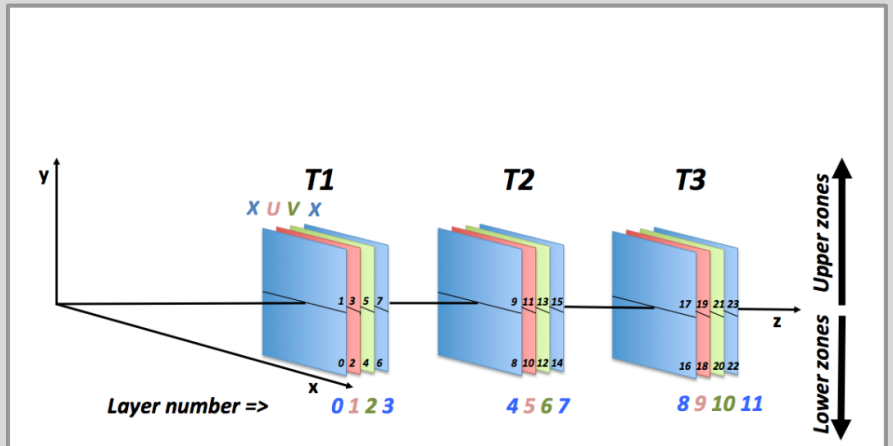


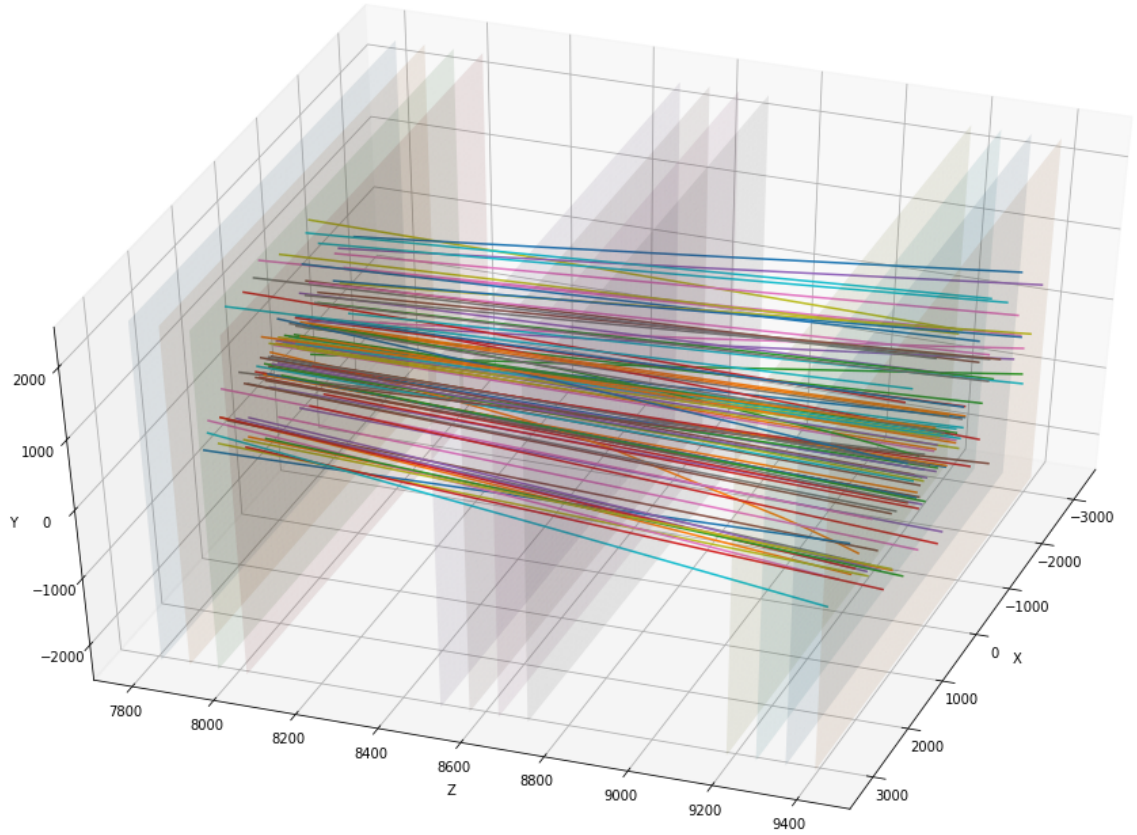


SCINTILLATING FIBRE TRACKER


4 planes x 3 stations
XU VX
2 x 2.4 m
fibre mat
SiPMs and Front-end electronics
1 Module
Mirror
(increase light yield)
Read-out
honeycomb & carbon fibre reinforced polymer
2 x 4 x fibre mats
540 mm
2 x 3 m
Read-out


x-layer
u-layer
v-layer
x-layer


y
T1
T2
T3
X U V X
1 3 5 7
0 2 4 6
9 11 13 15
8 10 12 14
17 19 21 23
16 18 20 22
z
x
Layer number =>
0 1 2 3
4 5 6 7
8 9 10 11
Upper zones
Lower zones

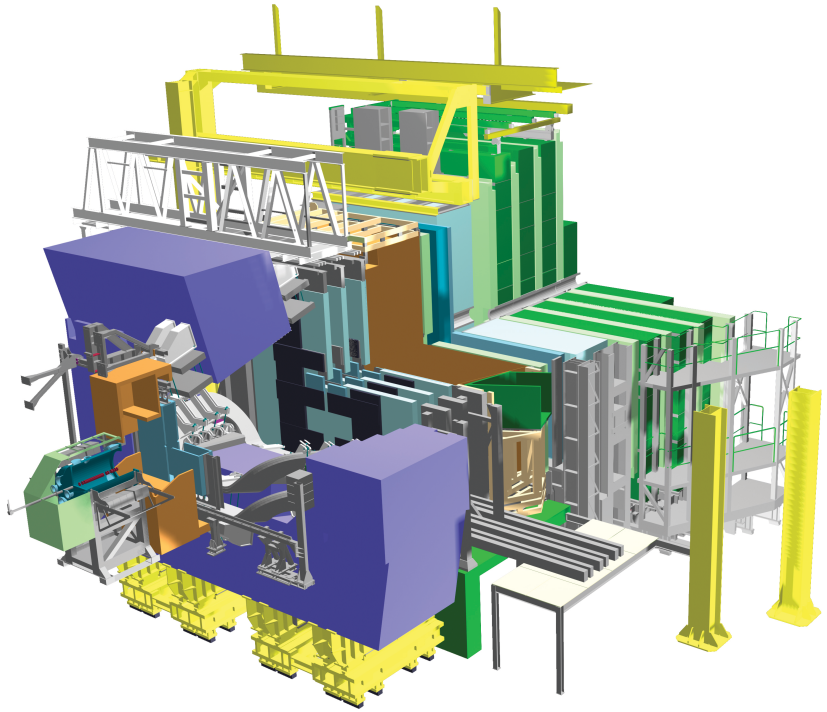

2000
1000
Y 0
−1000
−2000
7800
8000
8200
8400
8600
Z
8800
9000
9200
9400
−3000
−2000
−1000
0 X
1000
2000
3000

SCIFI DATA

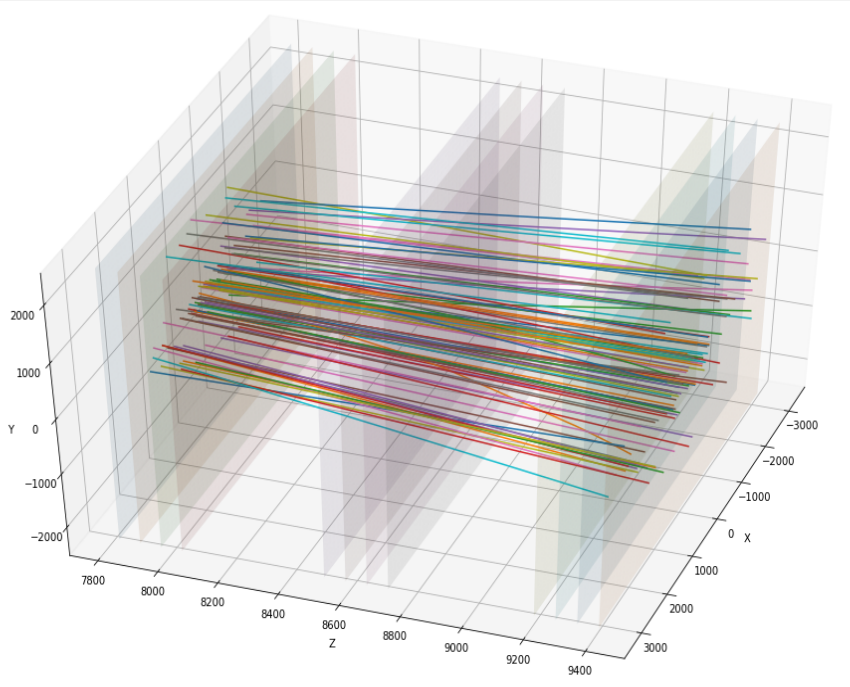
MONTE CARLO

The Monte Carlo method allows to simulate the real world detector and its response to a collision of certain particles.

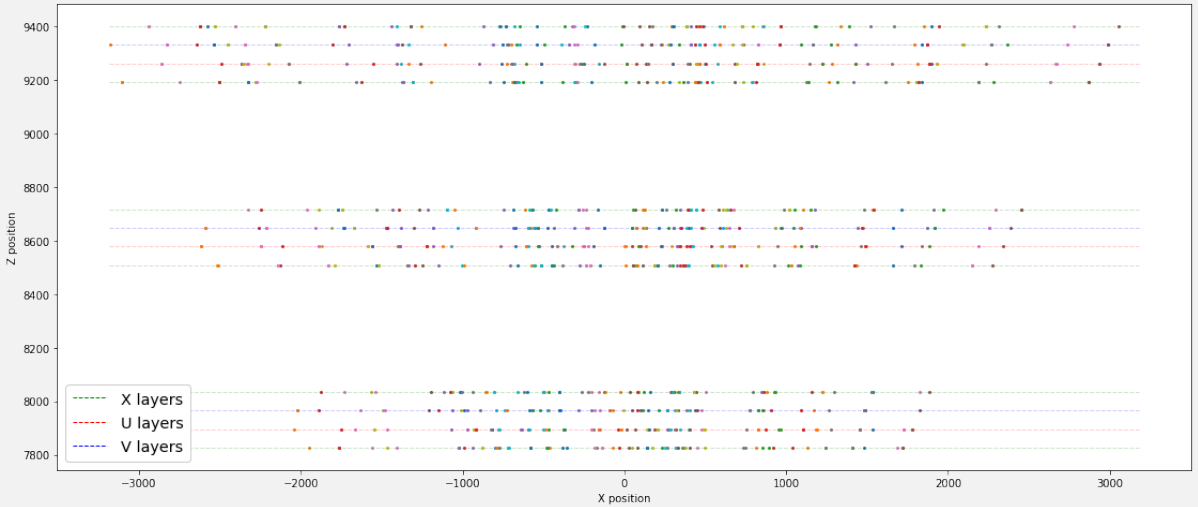
Accurate description, from its sensitive material to the glue that attaches each module for a particular T-station.



MCHITS



PRHITS

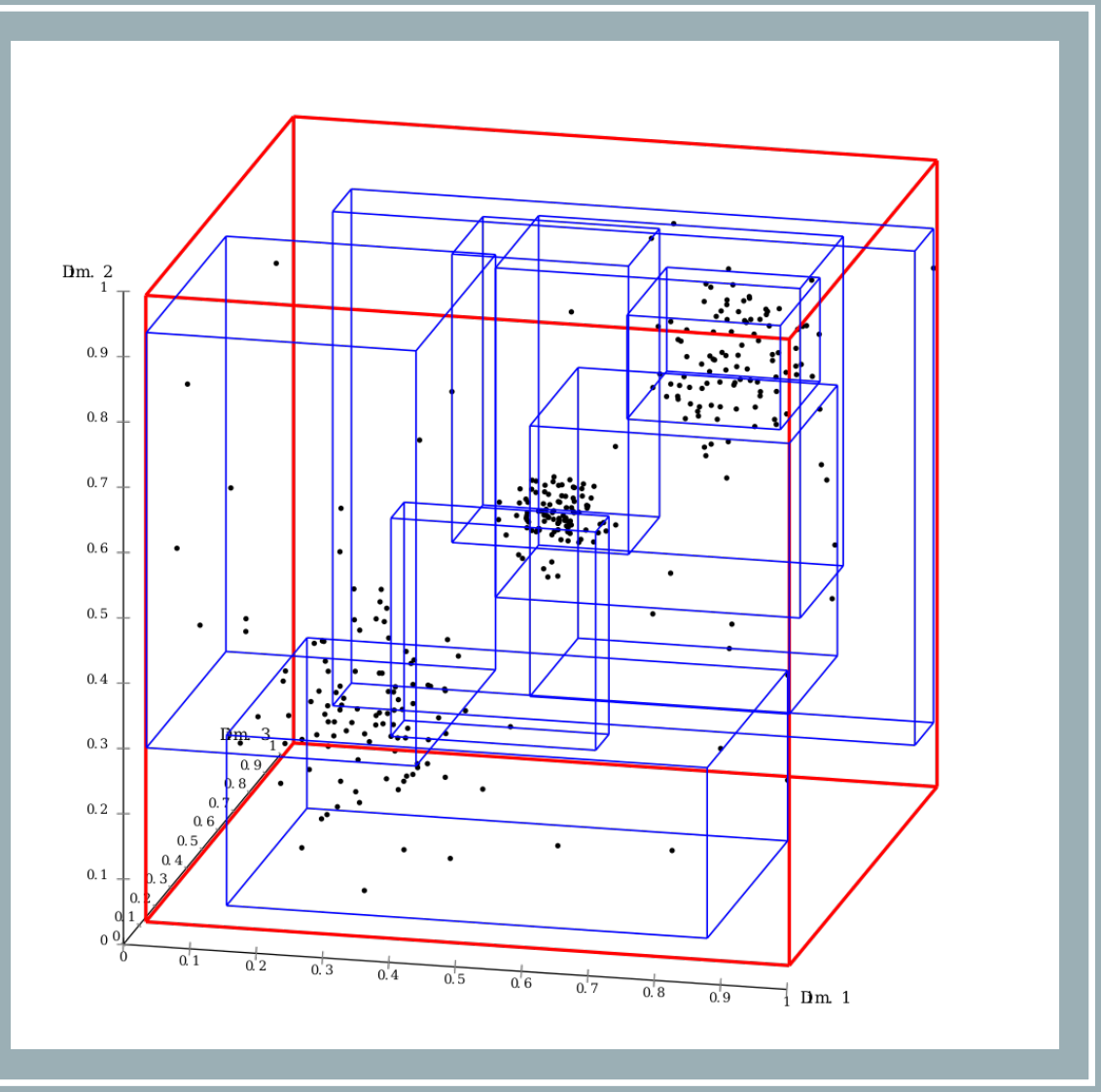


* *Hits* are three-dimensional points that represent the energy deposits of the particles in the subdetector.

ALGORITHM

CONCEPTS

- *Hit:* Interaction of a particle with a sensitive part of a detector and processed in order to be used by pattern recognition algorithms.
- *Seed:* Union of 2 to 4 hits from a station where each one belongs to a certain layer.
- *Track:* Set of seeds that represents the trajectory of a particle in the detector.
- *Index:* instance of an r-tree.
- *Ghost rate:* Ratio of seeds incorrectly categorized as valid but non-existent in reality.
- *Clone rate:* Ratio of duplicate seeds of the total reconstructed.
- *Event:* Proton-proton collision.



R-TREE

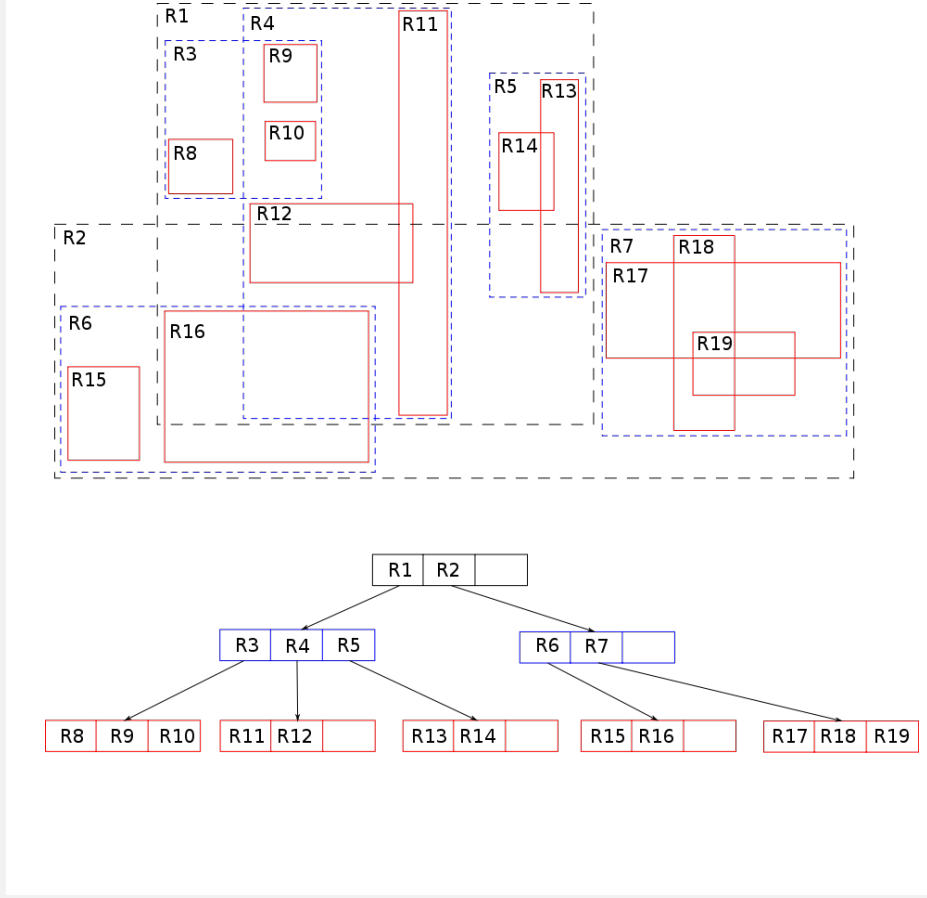
WHAT ARE THEY?

Tree-shaped data structures.

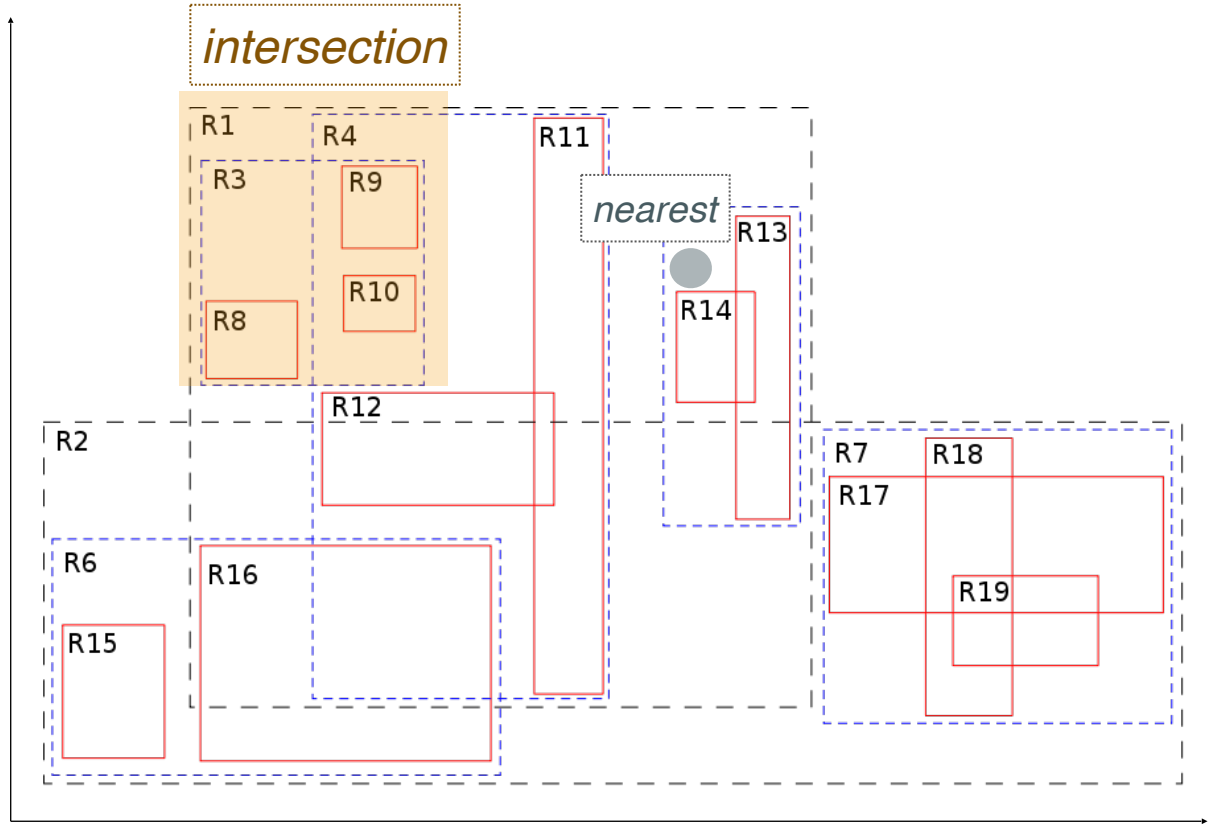
Index multidimensional information such as geographic coordinates.

Group nearby objects and represent them with their minimal bounding rectangle.

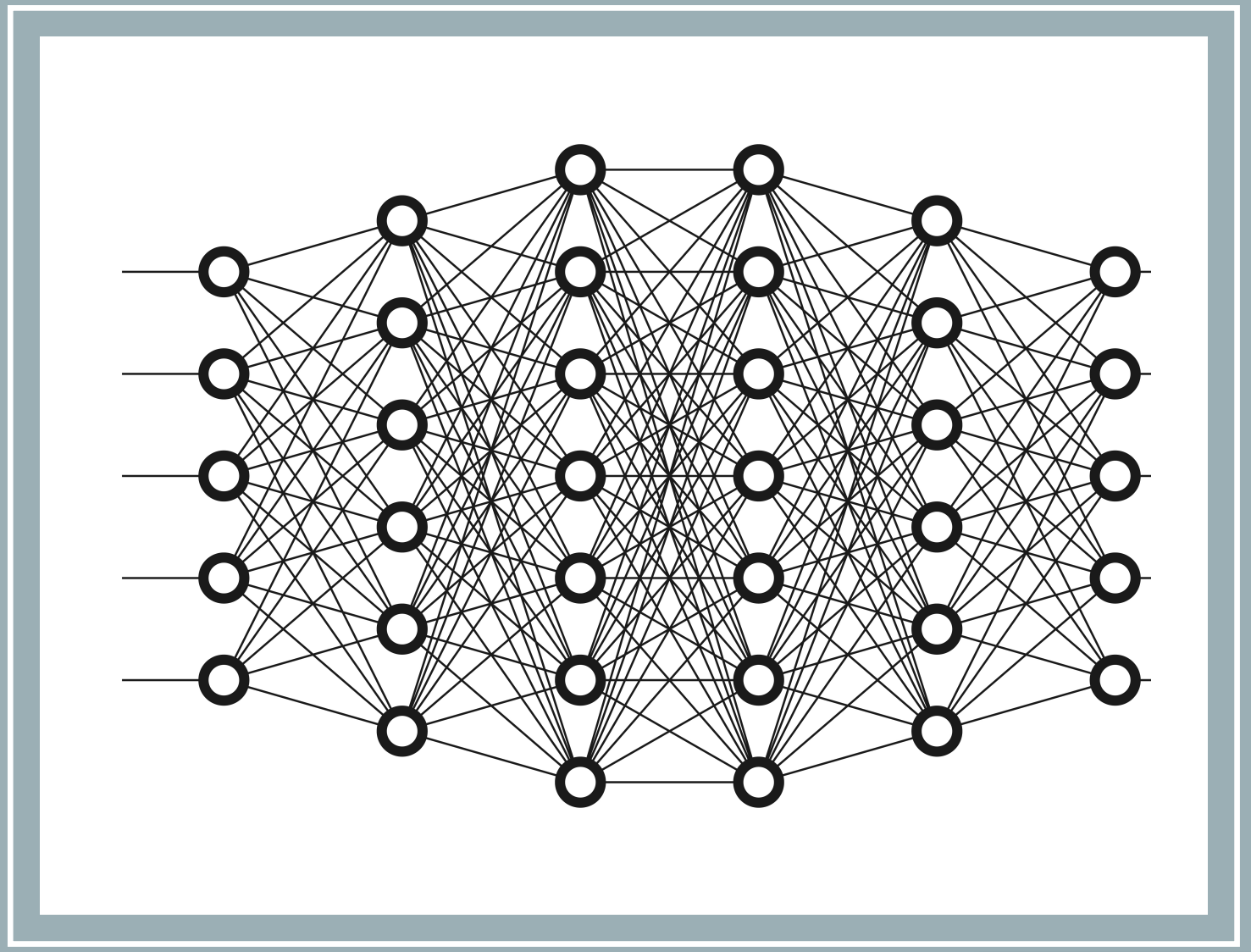
Search cost of $O(log_M(N))$ being M entries for each node and N the number of nodes.



OPERATIONS


intersection
nearest
R1
R4
R3
R9
R8
R10
R11
R13
R14
R12
R2
R7
R18
R17
R6
R16
R19
R15

ARTIFICIAL NEURAL NETWORKS



WHAT ARE THEY?

Computational systems that are inspired by biological neural networks.

They "learn" to perform tasks by considering examples, usually without being programmed with any specific task rules.

They automatically generate identifying characteristics of the learning material they process.

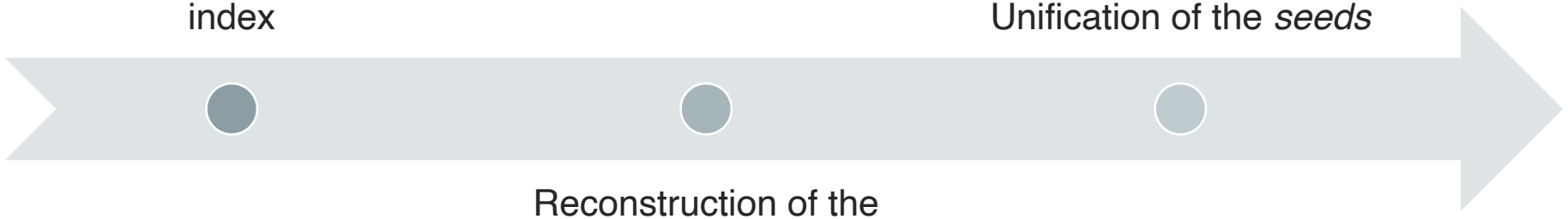
Two of the main applications of artificial neural networks are to predict or classify data.

GENERAL IDEA

All *hits* are added to an index

Reconstruction of the *seeds* in parallel in the three seasons

Unification of the *seeds*



GENERAL IDEA

INITIALISE *HITS_INDEX*

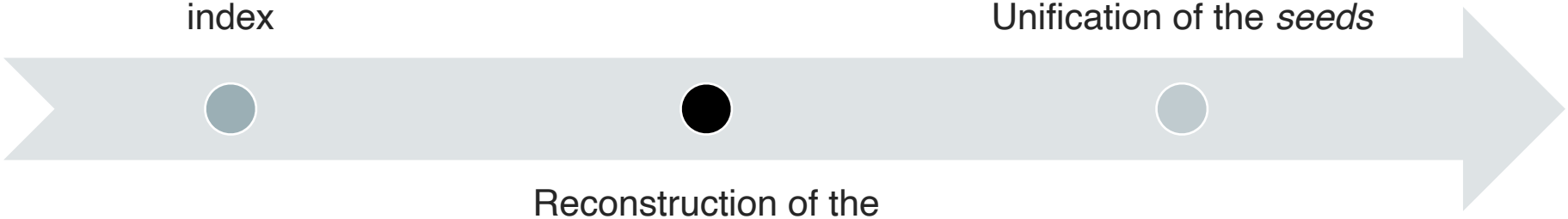
- Add all *hits* to *hits_index* with the following characteristics:
  - Coordinate system with 3 dimensions.
    1. X position in millimeters of the hit.
    2. Numeric code of the plan.
    3. Upper or lower area in the plan.
- Discard points that have a difference equal to or less than ± 1mm, in order to decrease the *clone* rate.

GENERAL IDEA

All *hits* are added to an index

Reconstruction of the *seeds* in parallel in the three seasons

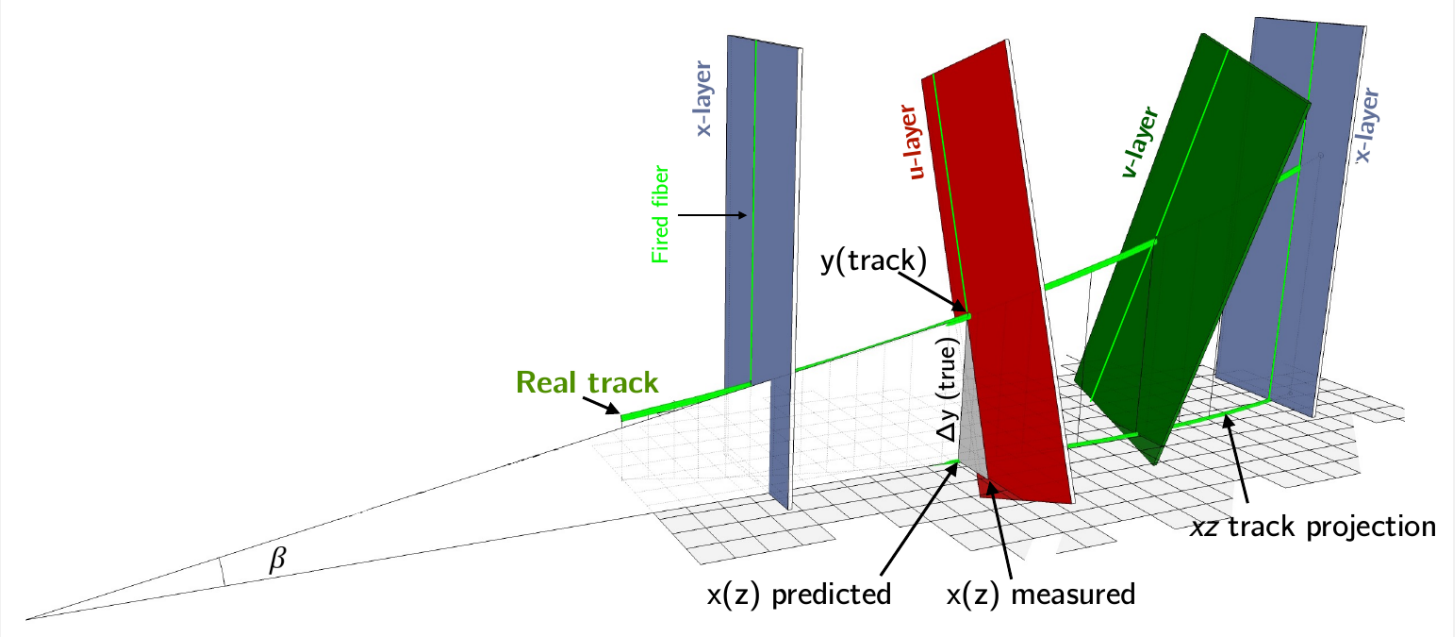
Unification of the *seeds*



RECONSTRUCTION OF *SEEDS* BY STATION

- Unify *hits* to form possible *seeds* in each season.
- Three threads of execution, one for each station.
- The *seeds* reconstructed by each are added to a new *index* with the following characteristics:
  - Coordinate system with 3 dimensions:
    1. X position in millimeters of the first *hit*.
    2. *Seed* height prediction.
    3. Station number (0 to 2).

HEIGHT PREDICTION



Projection of a line from position (0,0)

Computing beta angle:

$$\beta = \frac{X_{pred} - X_{real}}{\alpha_{stereo} * Z_{stereo}}$$

Calculate the possible *X* position of the particle in a given layer.

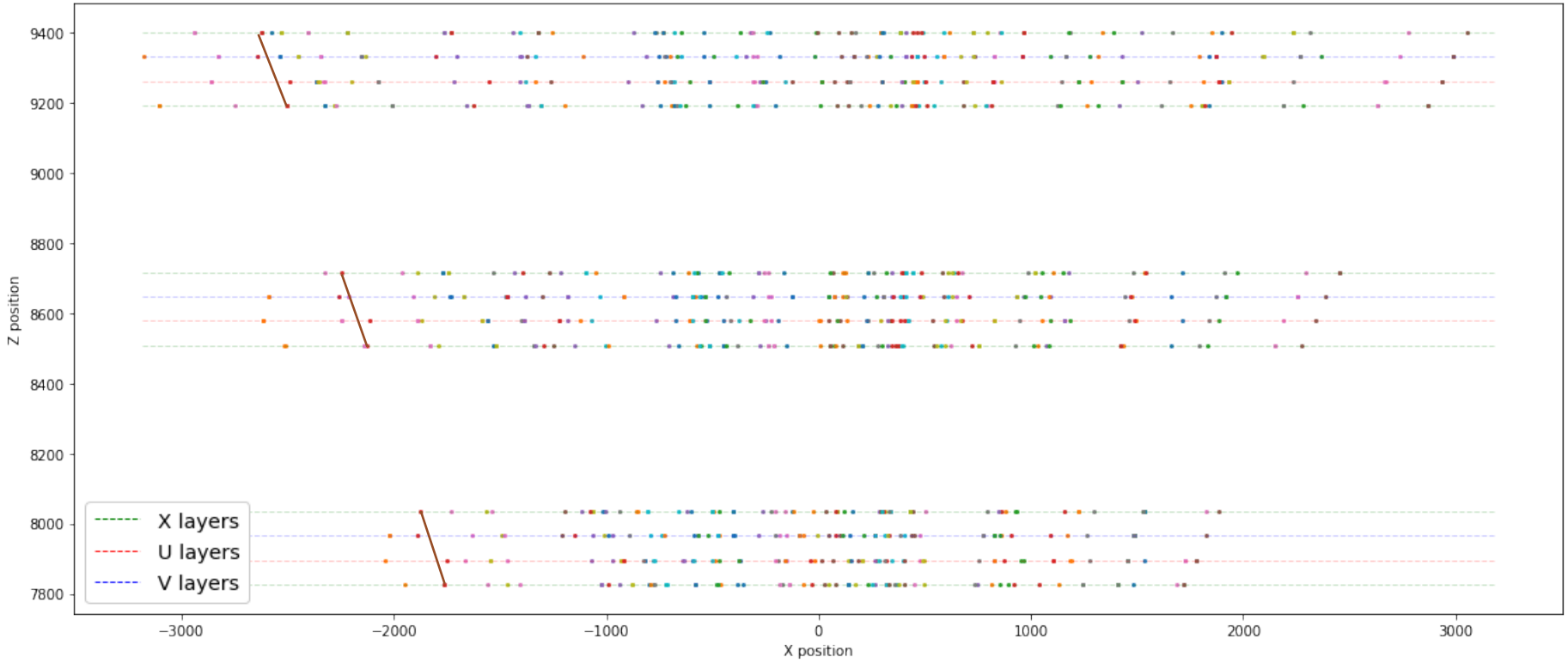
The height is finally obtained by:

$$Y_{pred} = \tan(\beta) * Z_{stereo}$$

RECONSTRUCTION OF *SEEDS* BY STATION

1. Select all *hits* from *hits_index* using the following coordinates: (X_{min}, X_{max}, X_0_layer, X_0_layer, 0, 1).
2. For each:
    1. Predict using the basic projection of a line the position of the particle in the following layers *U-V-X*.
    2. These values $PredX_U$, $PredX_V$ and $PredX_{X1}$ are used to access the *hits_index* to get the possible hits of the particle in the layers *U-V-*X_0:
        - Intersection at ($PredX_U - limit$, $PredX_U + limit$, $U_{planecode}$, $U_{planecode}$, Hit_{zone}, Hit_{zone}) with format (X_{min}, X_{max}, Y_{min}, Y_{max} Z_{min}, Z_{max})

RECONSTRUCTION OF *SEEDS* BY STATION

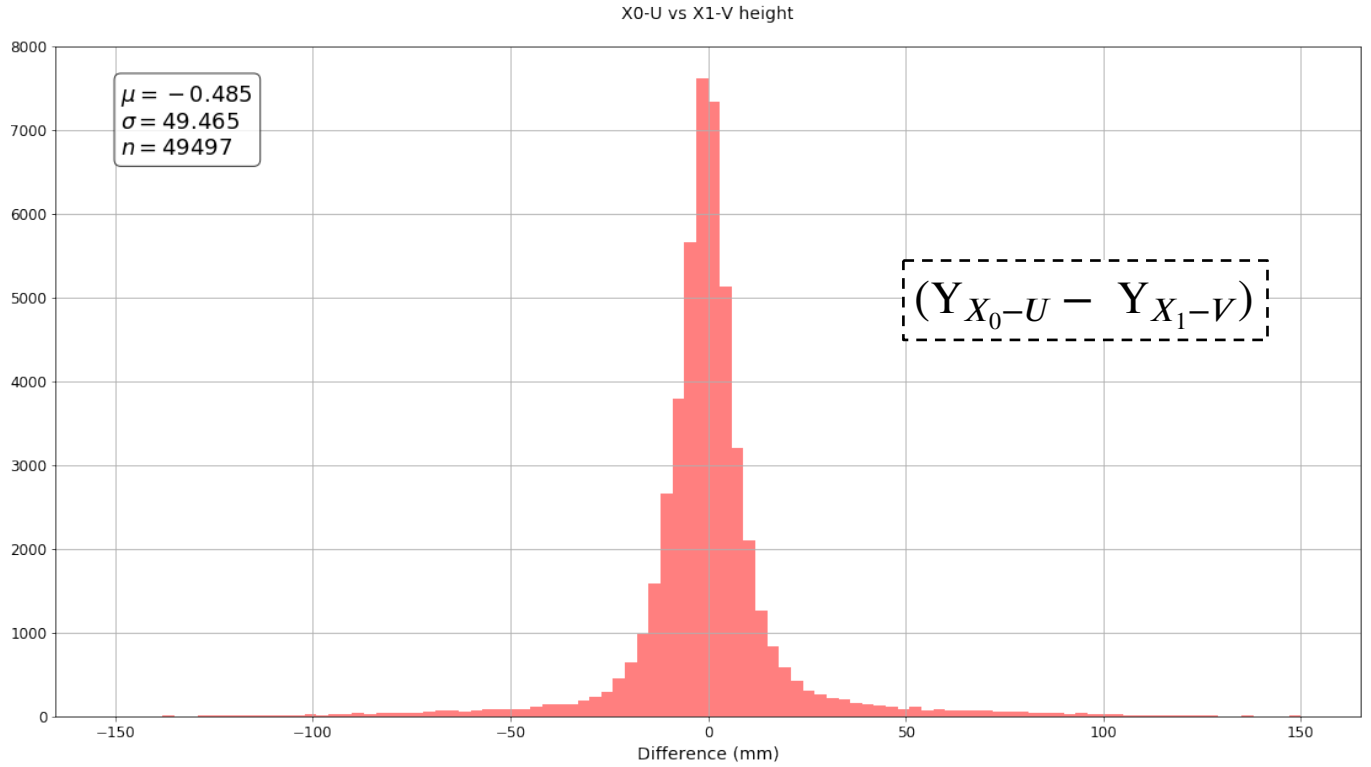


RECONSTRUCTION OF *SEEDS* BY STATION

RECONSTRUCTION OF *SEEDS* BY STATION

3. Once we've obtained *hits* for each layer:
   1. If none is returned, incomplete seeds are created.
   2. If there are more than 0 *hits* for each layer, the height of each possible combination is calculated and possible seeds are formed.
   3. All these that there is a difference between Y_{X_0-U} i Y_{X_1-V} greater than *50 mm* are discarded.
4. The *seed* with the best probability of a valid combination is selected using the ANN model called *seed_station_validation*.
5. If a probability greater than 20% is obtained, it is inserted in the index *seeds_index*.

RECONSTRUCTION OF *SEEDS* BY STATION



GENERAL IDEA

SEED UNIFICATIONS

1. Get all the *seeds* of the first station of the *rtree* called *index_seeds*.
2. For each:
    1. If 25% of the hits it contains have already been picked by full traces, it is discarded.
    2. Otherwise, a prediction is made of the position of the first hit of the seed in the next station by projecting a line.
    3. Based on the value, an intersection is performed in the tree using certain ranks: $(PredX - 100,\ PredX + 100,\ mean - 150,\ mean + 150,\ 1,\ 1)$ with format $(X_{min}, X_{max}, Y_{min}, Y_{max}\ Z_{min}, Z_{max})$
        - $mean = \frac{(\mathrm{Y}_{X_0 - U} + \mathrm{Y}_{X_1 - V})}{2}$

SEED UNIFICATIONS

4. Of all the *seeds* obtained by the previous intersection, the one with the best probability of being valid is selected using the ANN model called *two_seeds_validation*.
5. If this has more than 25% of *hits* already selected by another combination, it is discarded and returns to step 2.
6. Based on the previous result, an intersection is again performed at the last station by predicting the position using the projection of a line.
7. From all the *seeds* obtained by the previous procedure, the one with the best probability of being valid is selected together with the result of section 4 and the same as section 5 is checked.
8. Finally, if the combination of the three seeds passes all the checks, it is saved as a valid track and all the hits used are marked.

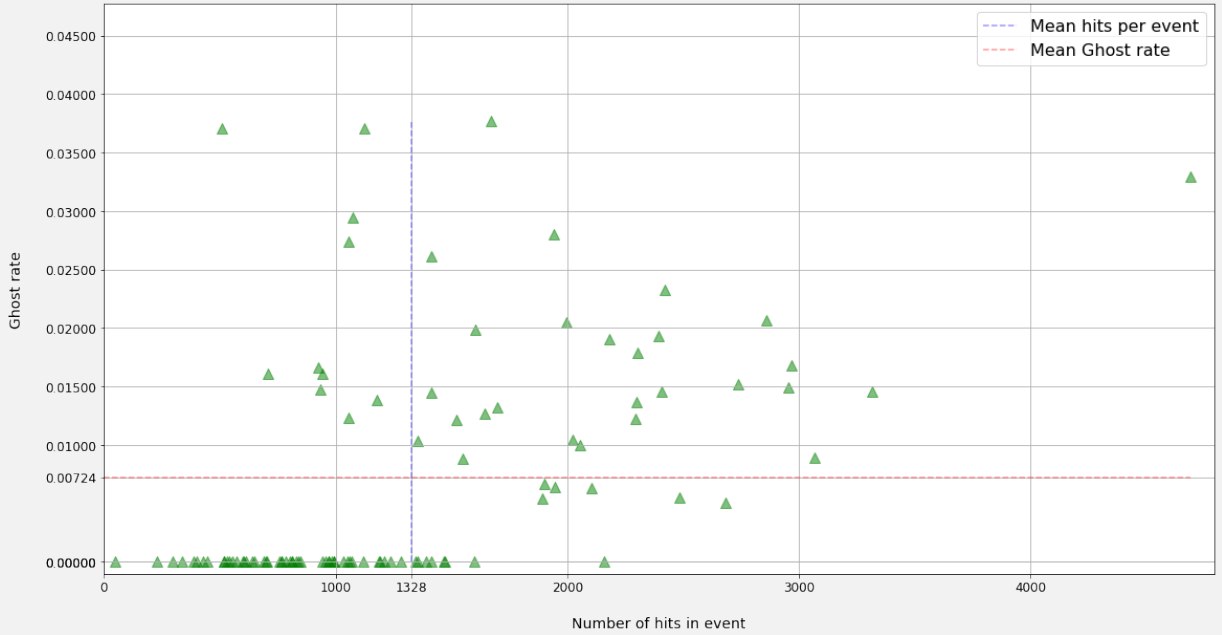
PERFORMANCE

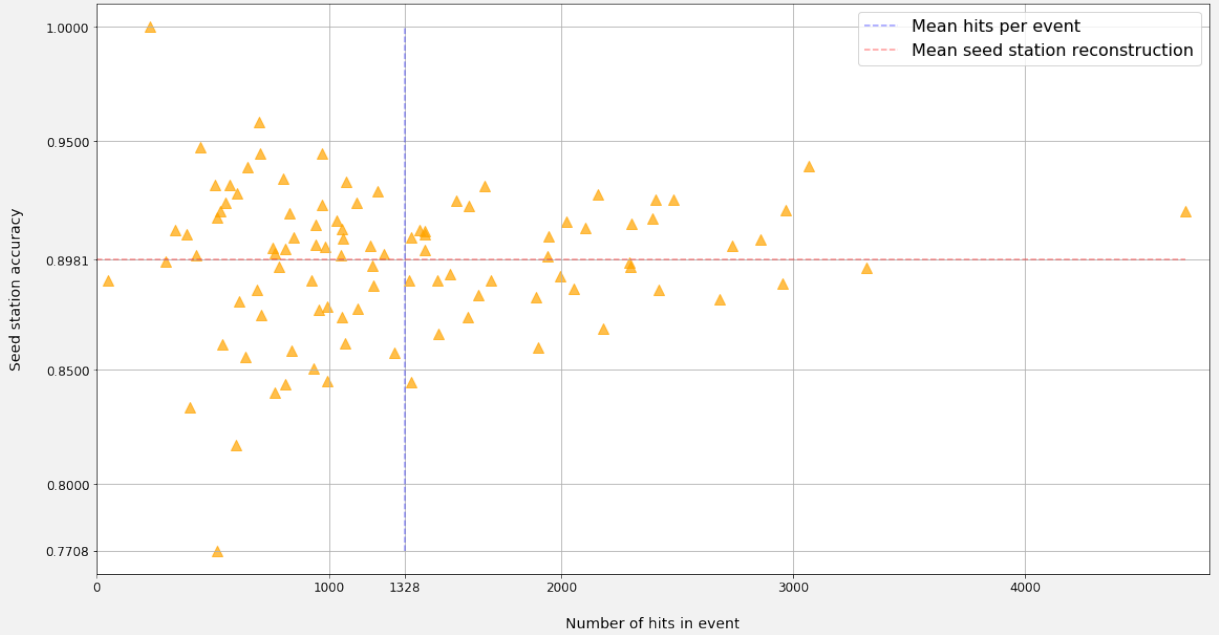
INDICATORS AND CONSIDERATIONS

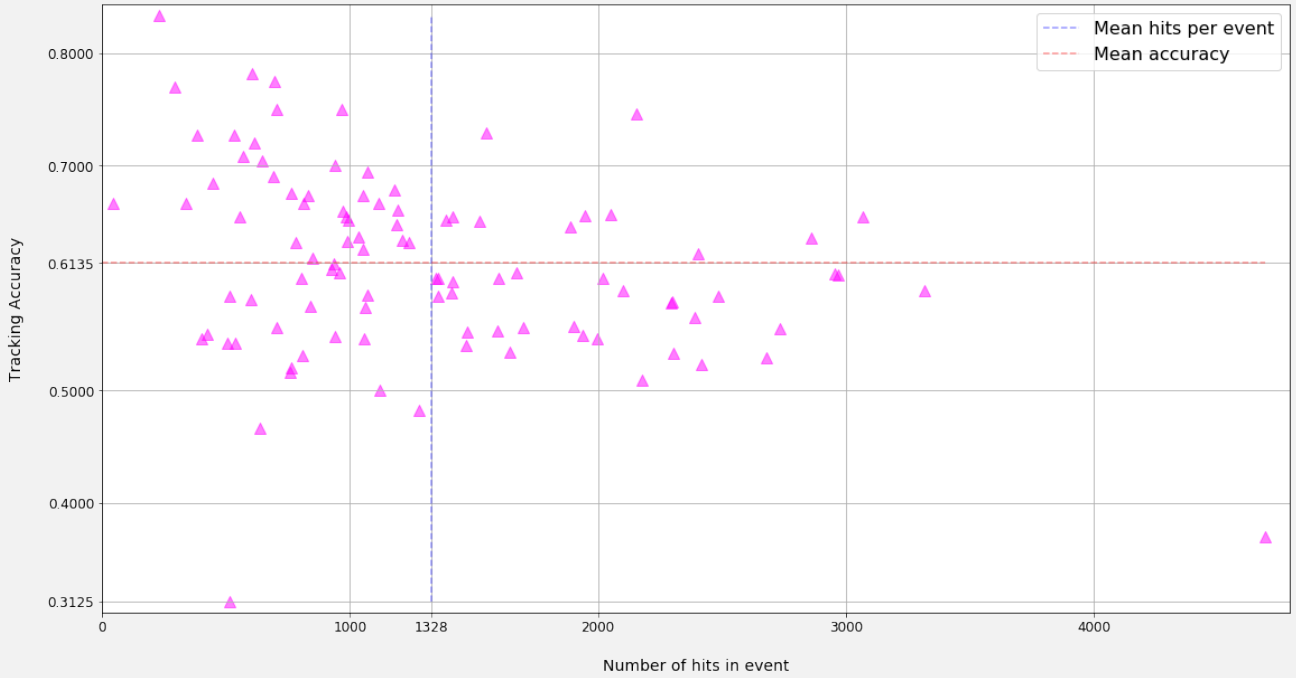
- Reconstructed traces: total number of traces that the algorithm is able to reconstruct.
- *Reconstructable traces*: Total number of traces that can be reconstructed.
- Tracking efficiency = $\frac{valid\ reconstructed}{reconstructables}$
- *Ghost rate* = $\frac{invalid\ reconstructed}{reconstructables}$
- For a reconstructed trace to be valid, it must match at least 70% of the *hits* with a reconstructable one.
- A particle is reconstructable if it has hits in at least one of the two layers X and U or V.

RESULTS

	Efficiency	Ghost rate	Seed reconstruction
Current	(67.2 ± 0.1) %	(7.9 ± 0.1) %	-
New	(47.3 ± 2.1) %	(1.1 ± 0.2) %	(86.5 ± 0.1) %
New – Traces without missing *hits*	(74.8 ± 2.8) %	(0.7 ± 0.2) %	(96.6 ± 0.1) %
New – Traces with max one *hit less*	(53.2 ± 2.6) %	(1.2 ± 0.2) %	(90.4 ± 0.1) %







laSalle
Ramon Llull University

CONCLUSION

End-to-end efficiency around ~20% below existing approach.

Reduction of *ghost rate* by 86% compared to existing approach.

The lower seed reconstruction efficiency happens in the unification part and not in the reconstruction of seeds.

Higher number of hits per event reduces efficiency, mainly caused by using fixed ranges when searching trees.

FUTURE

1

Implementation of an improved system for the unification of seeds based on the use of certain heuristics.

2

Create a methodology of dynamic ranges (for larger outdoor and smaller indoor areas).

3

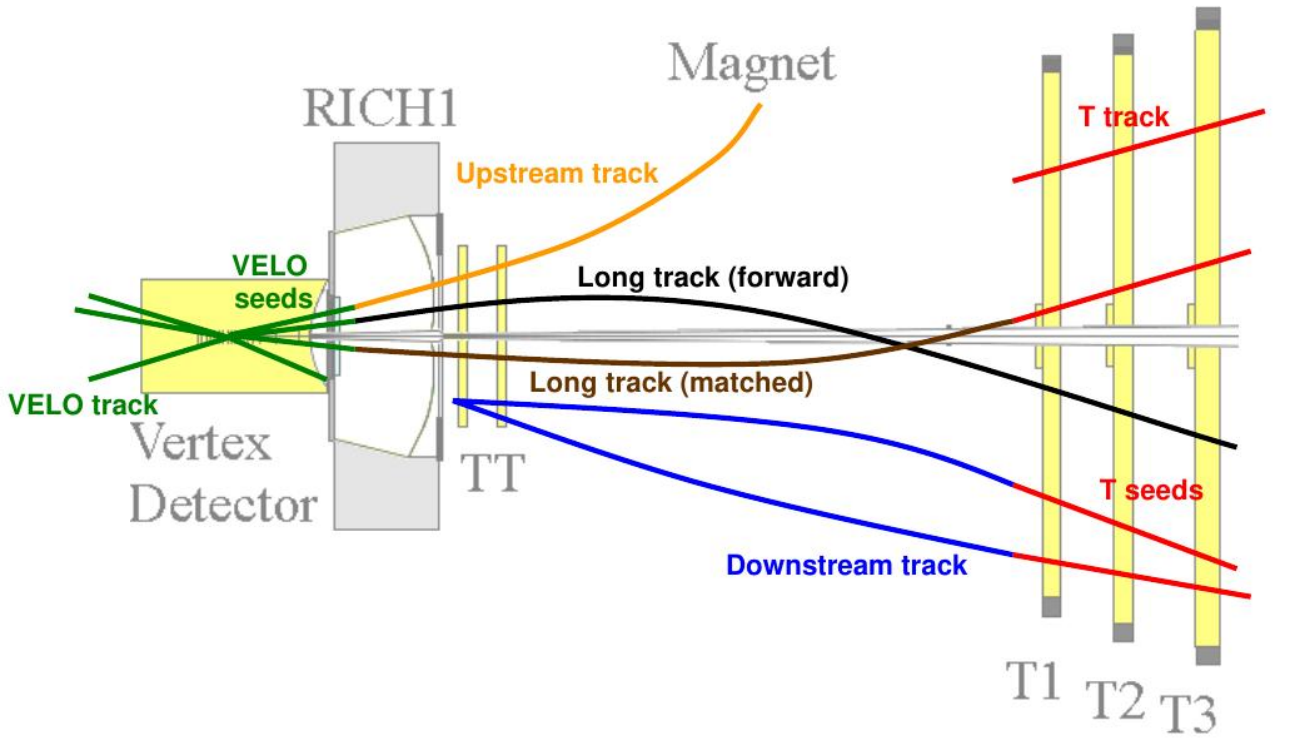
Carry out extensive research on ANN models, given their impact on the effectiveness of the algorithm.

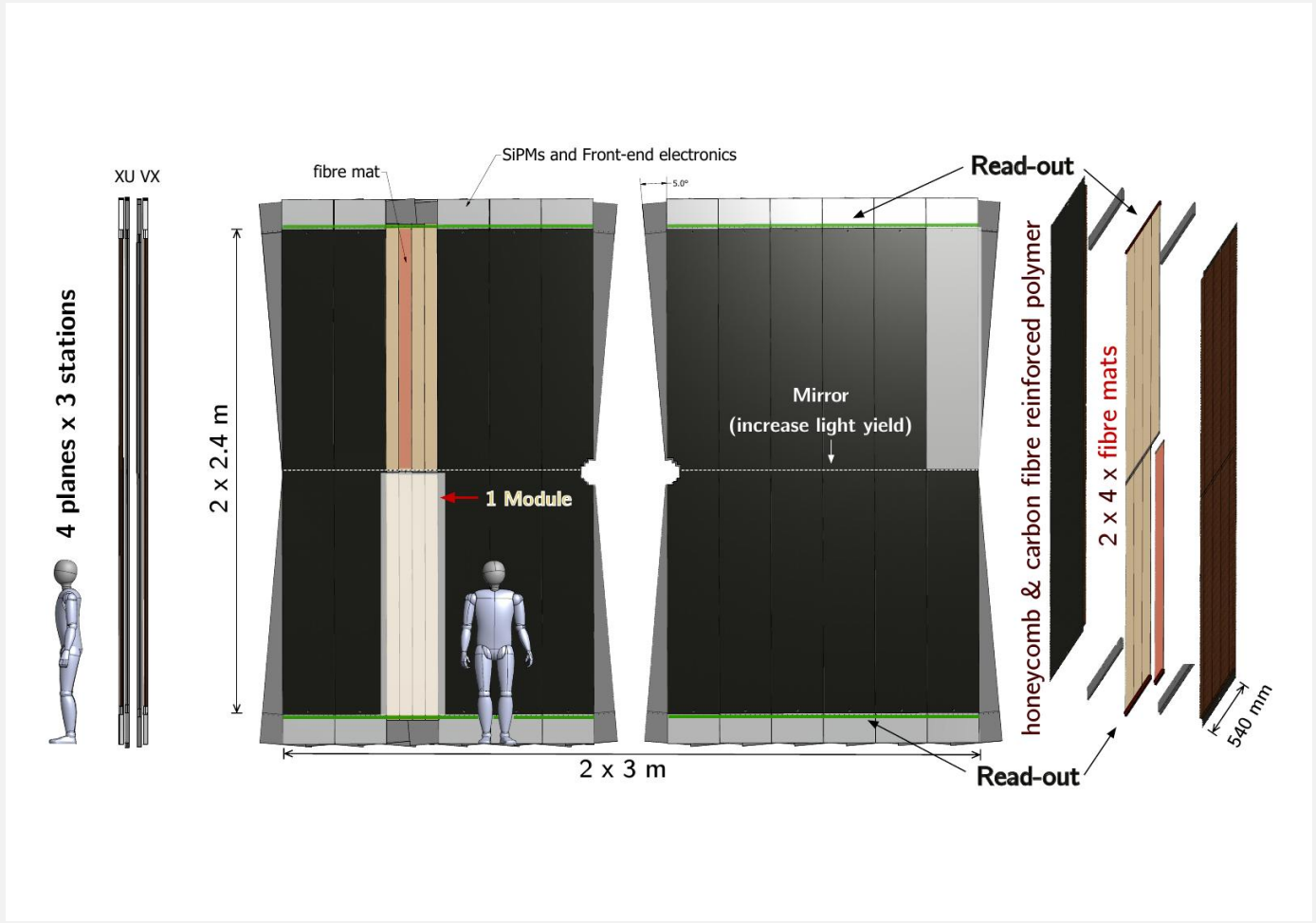
THANK YOU

Questions?

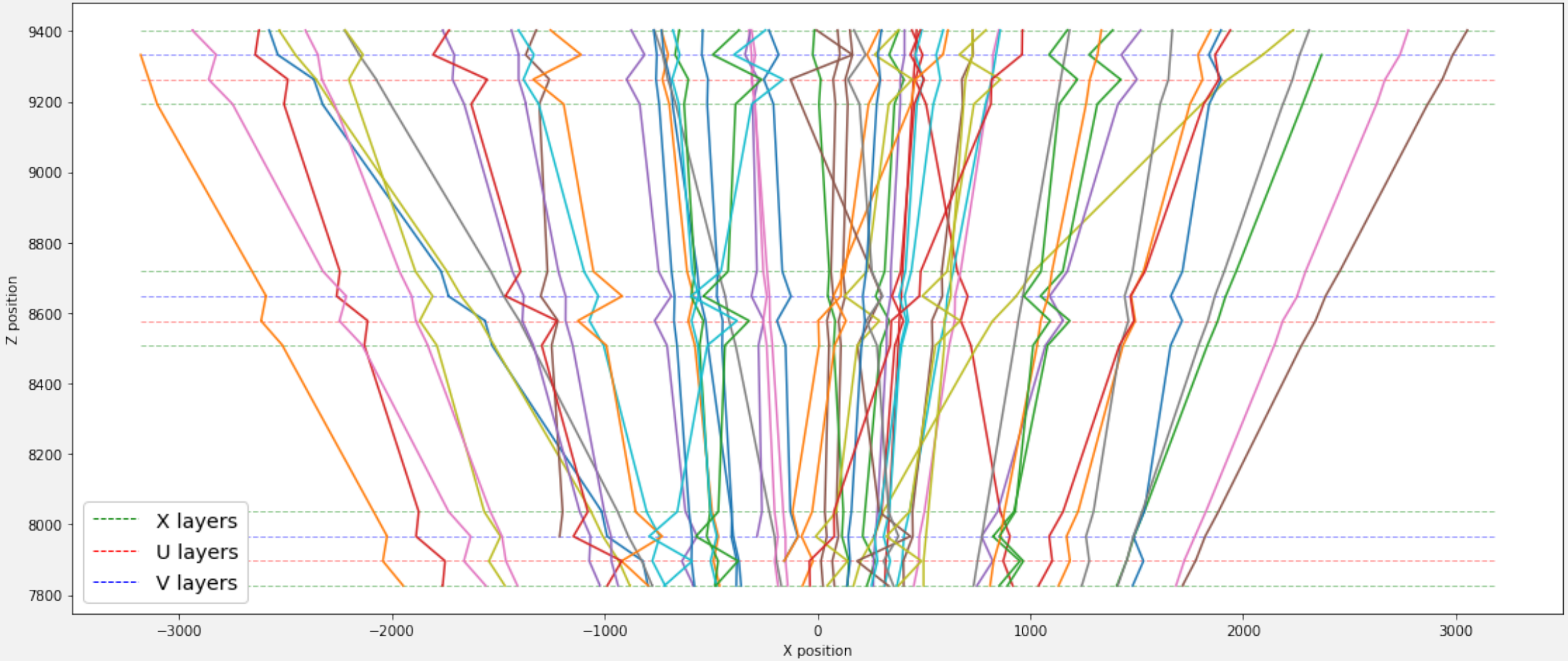
BACKUP

TRACKS


Magnet
RICH1
T track
Upstream track
VELO
seeds
Long track (forward)
Long track (matched)
VELO track
Vertex
Detector
TT
T seeds
Downstream track
T1
T2
T3

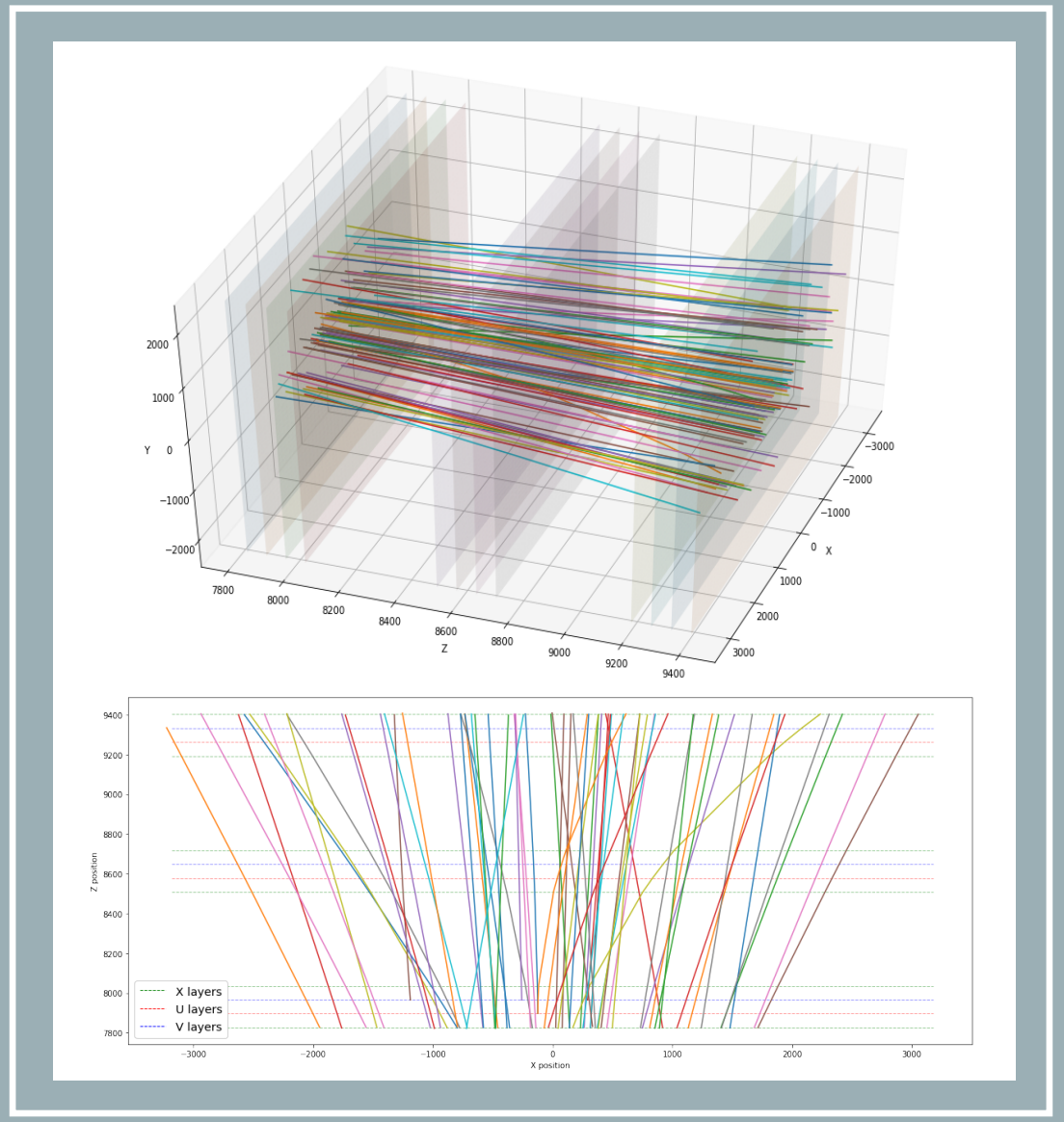


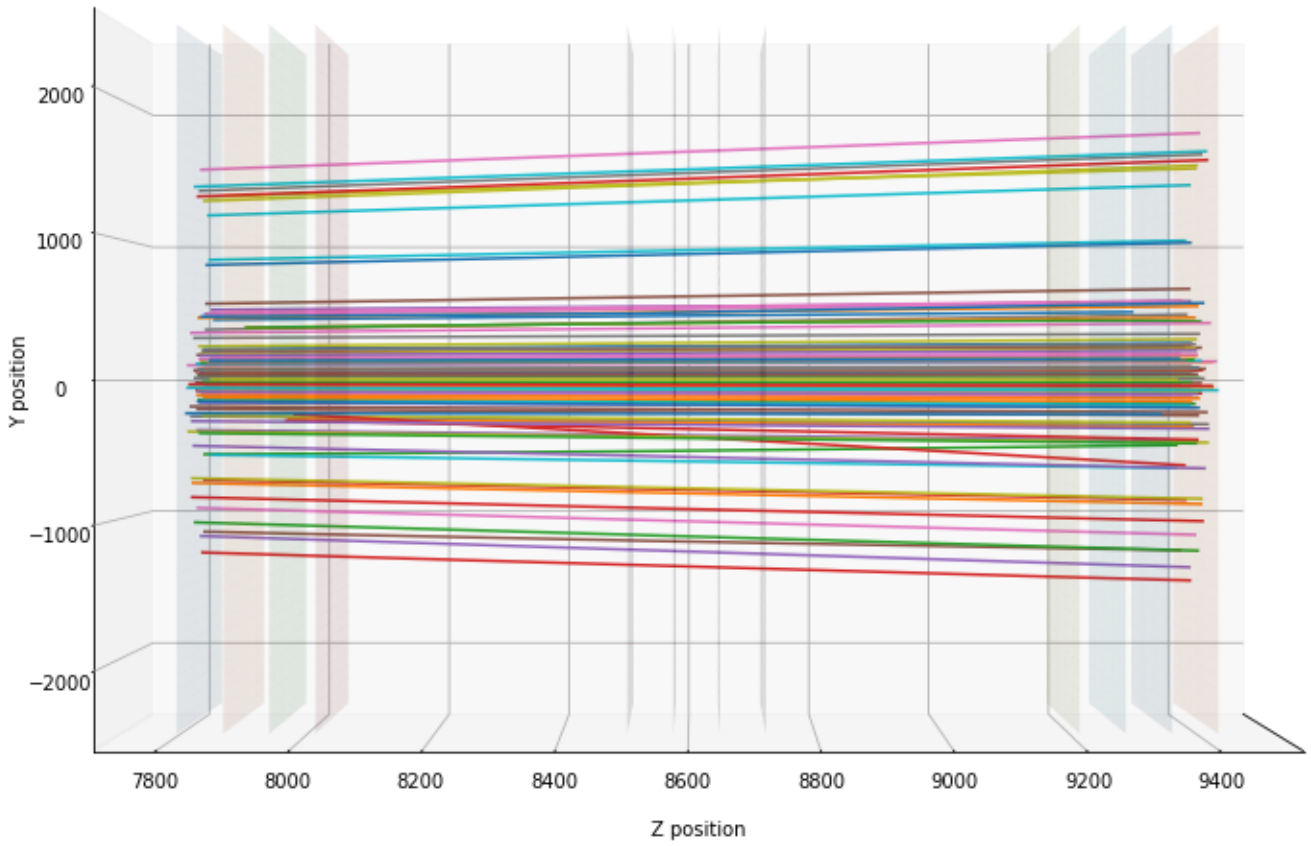
- Each station is made up of 4 different layers (X-U-V-X).
- X layers are oriented vertically (x position) and U-V forms about +/- 5 degrees (y position).
- Each layer - 12 modules, each of which contains 8 fibrillated matrices with 6 stacked layers of fibers 2.5 m long.
- Silicon PhotoMultipliers (SiPM) are located at the edges of these modules in order to collect the light produced and transported by the fibers.

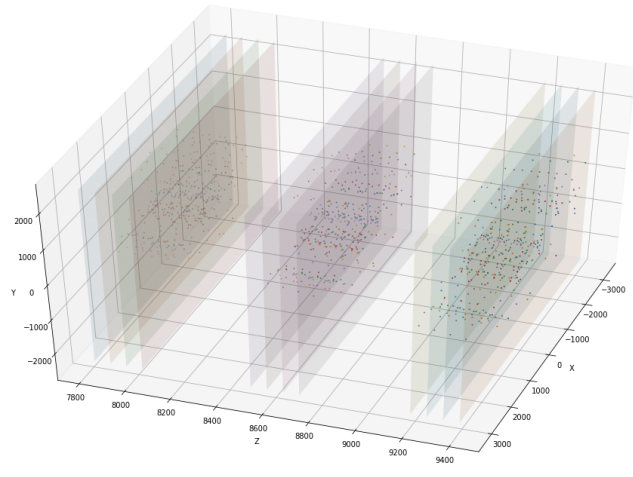

9400
9200
9000
8800
8600
8400
8200
8000
7800
Z position
−3000
−2000
−1000
0
1000
2000
3000
X position
X layers
U layers
V layers

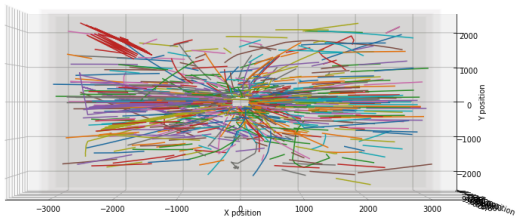
MCHITS

- *Hits* are three-dimensional points that represent the energy deposits of the particles in the sub-detector.
- Three-dimensional points that represent particle hits in the different sub-detectors.
- They are used to validate the performance of pattern recognition algorithms.




2000
1000
0
−1000
−2000
Y position
7800
8000
8200
8400
8600
8800
9000
9200
9400
Z position


2000
1000
Y
0
−1000
−2000
7800
8000
8200
8400
8600
Z
8800
9000
9200
9400
−3000
−2000
−1000
0
X
1000
2000
3000

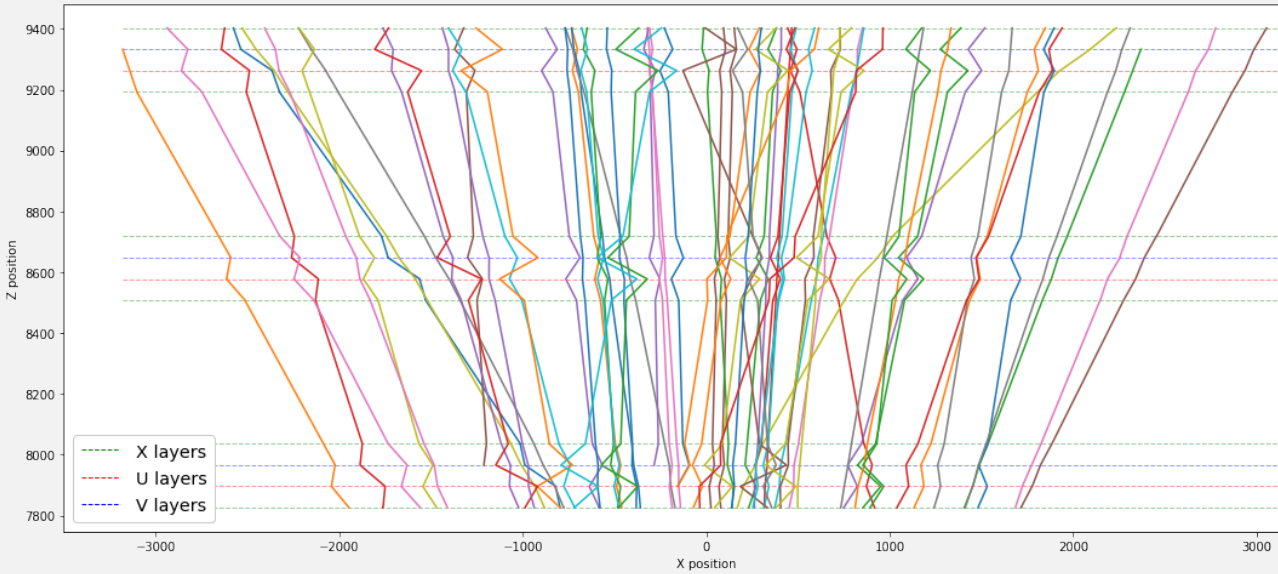

2000
1000
0
−1000
−2000
Y position
−3000
−2000
−1000
X position
1000
2000
3000

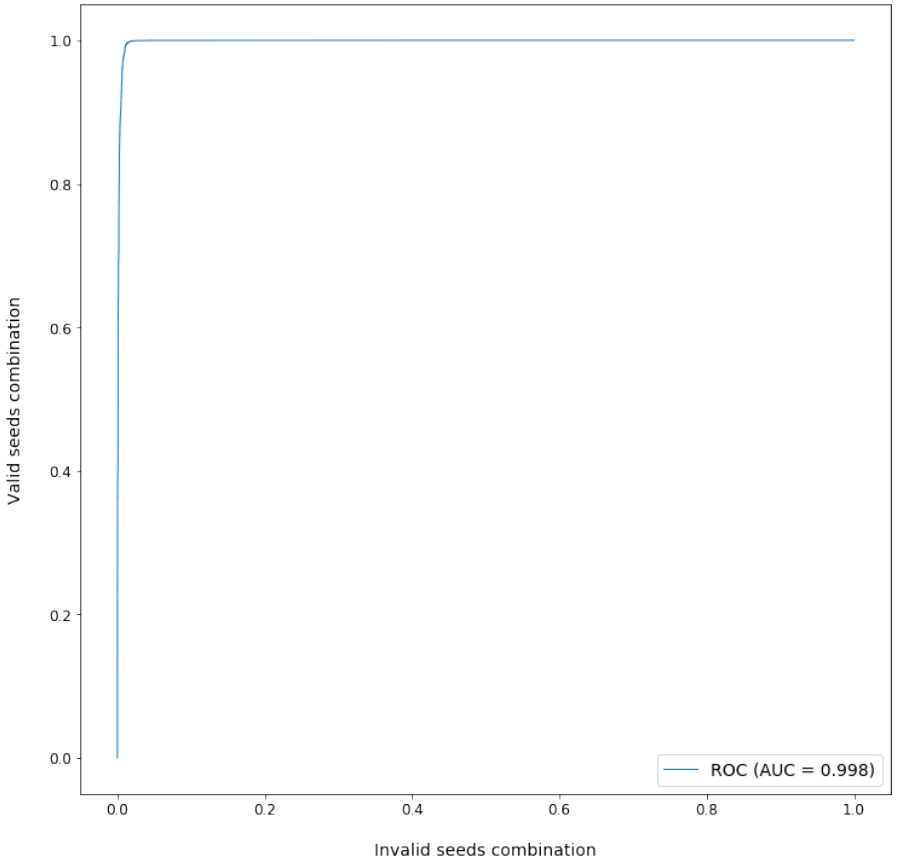
PRHITS

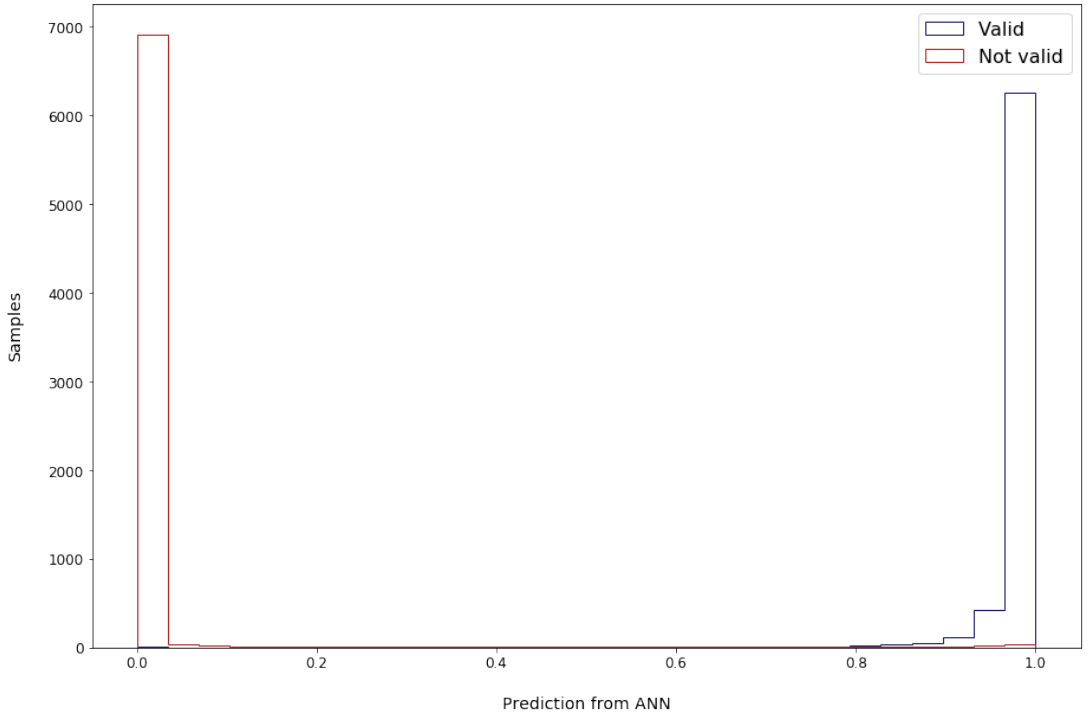
They represent a real approximation of SciFi output at a real event.

They are the processed positions obtained from the reading electronics.

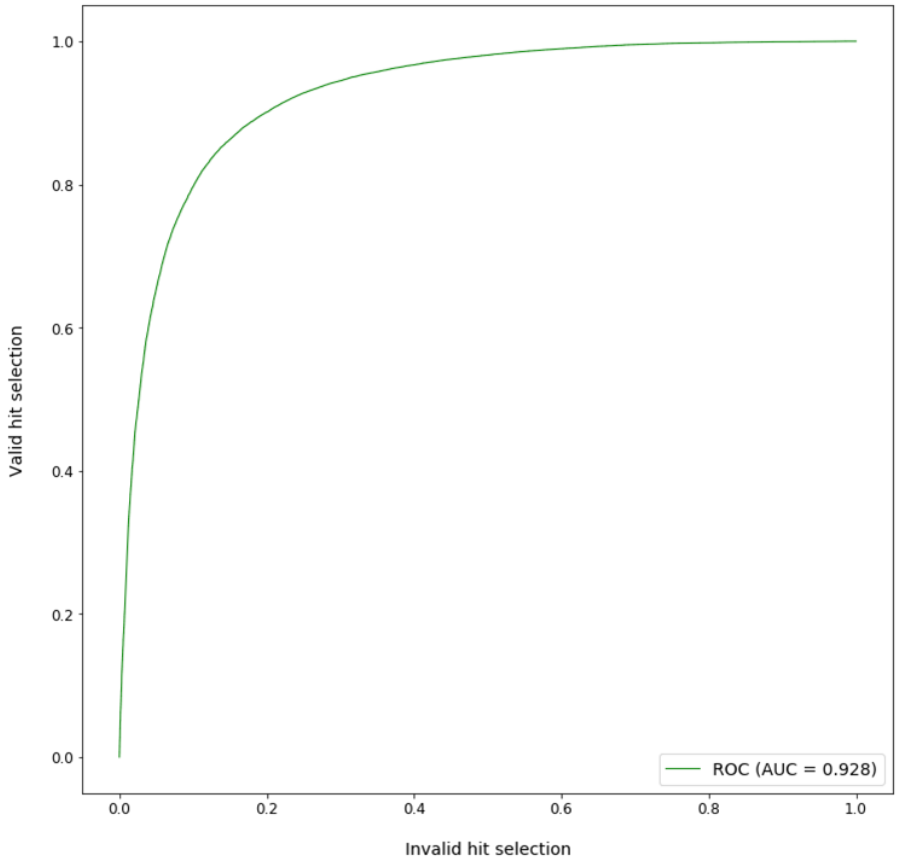
Only a 2D view of each hit can be obtained, as we cannot determine the y-position directly from the raw data.

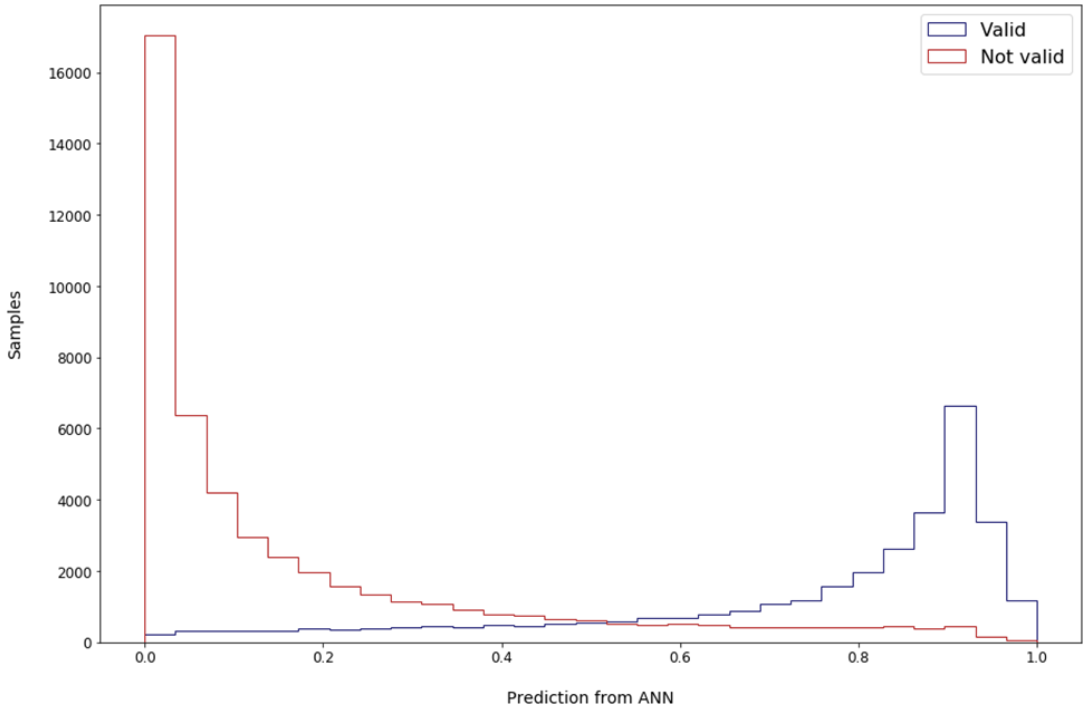




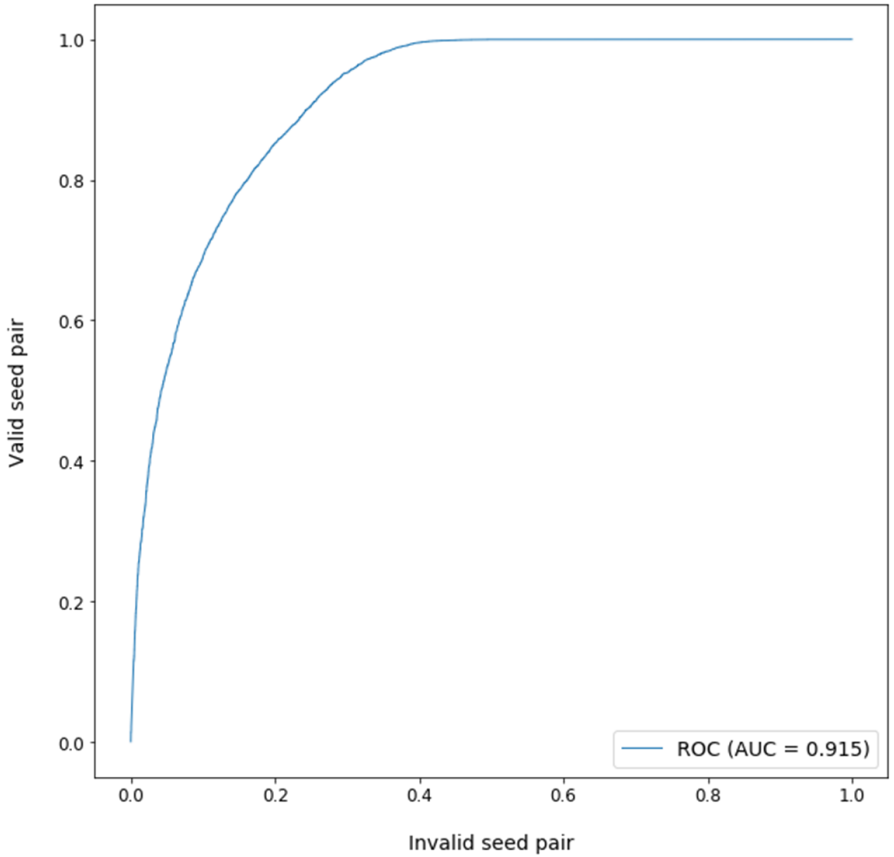


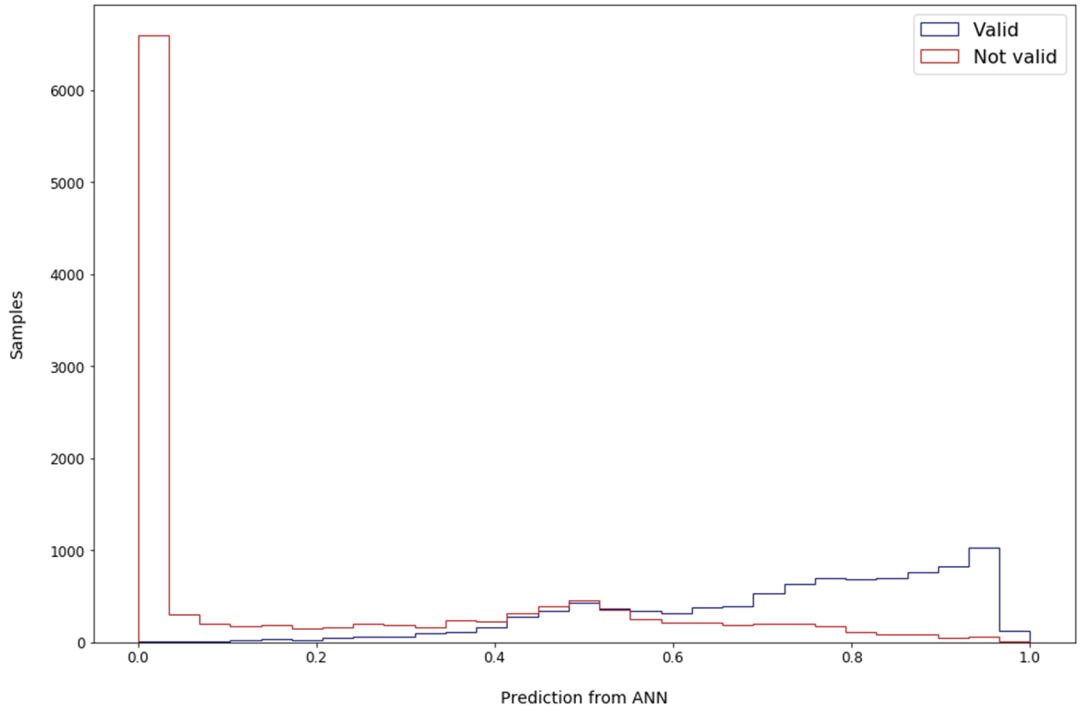
MODEL ALL SEED VALIDATION





MODEL STATION SEED VALIDATION





MODEL TWO SEEDS VALIDATION

CURRENT EFFICIENCY

Tipus de traça	Actual
hasT	(67.2 ± 0.1) %
long	(90.6 ± 0.1) %
long P >5 GeV/c	(94.8 ± 0.1) %
long from B	(93.4 ± 0.1) %
long from B P >5 GeV/c	(95.4 ± 0.1) %
UT + SciFi strange	(89.7 ± 0.1) %
UT + SciFi strange P >5 GeV/c	(95.2 ± 0.1) %
noVELO + UT + SciFi strange	(89.4 ± 0.1) %
noVELO + UT + SciFi strange P >5 GeV/c	(95.0 ± 0.1)%
ghost rate	(7.9 ± 0.1) %