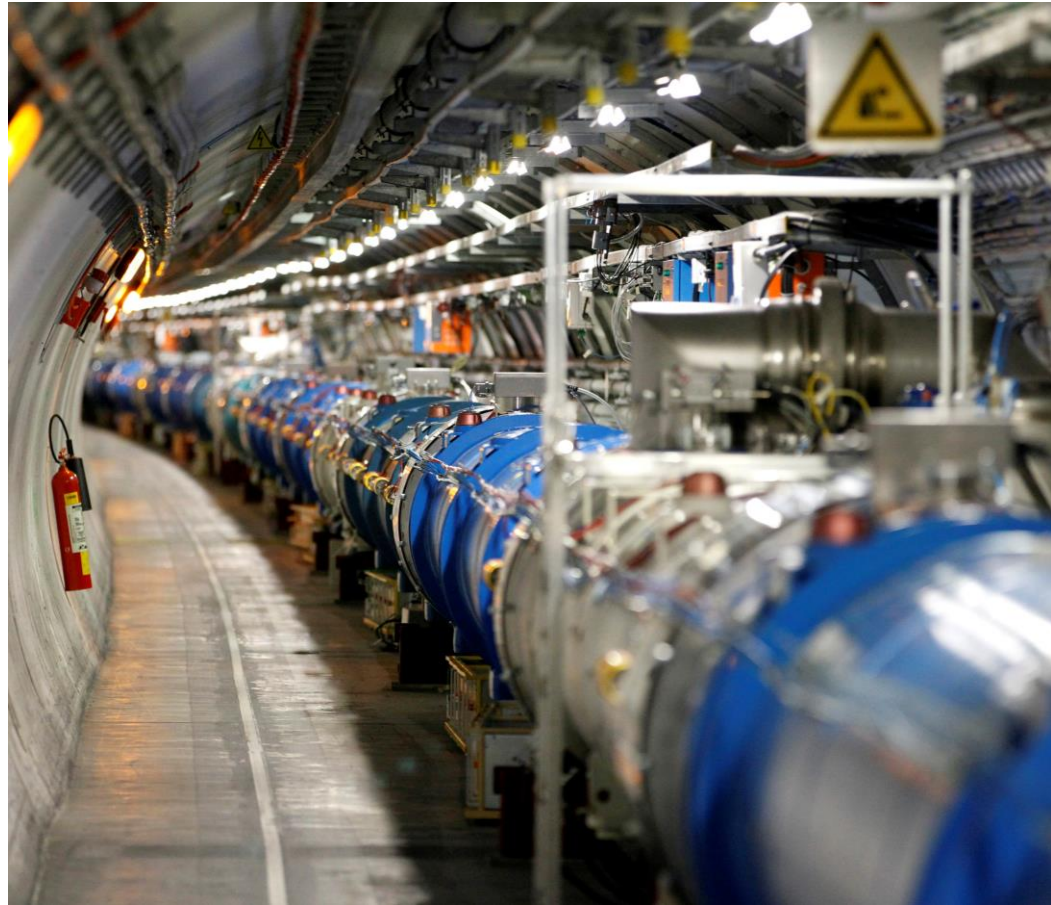




AN APPROACH TO TRACK RECONSTRUCTION FOR THE SCIFI TRACKER AT LHCb USING ANNs AND SPATIAL INDEXING

Gabriel Cammany Ruiz

CERN | LHC





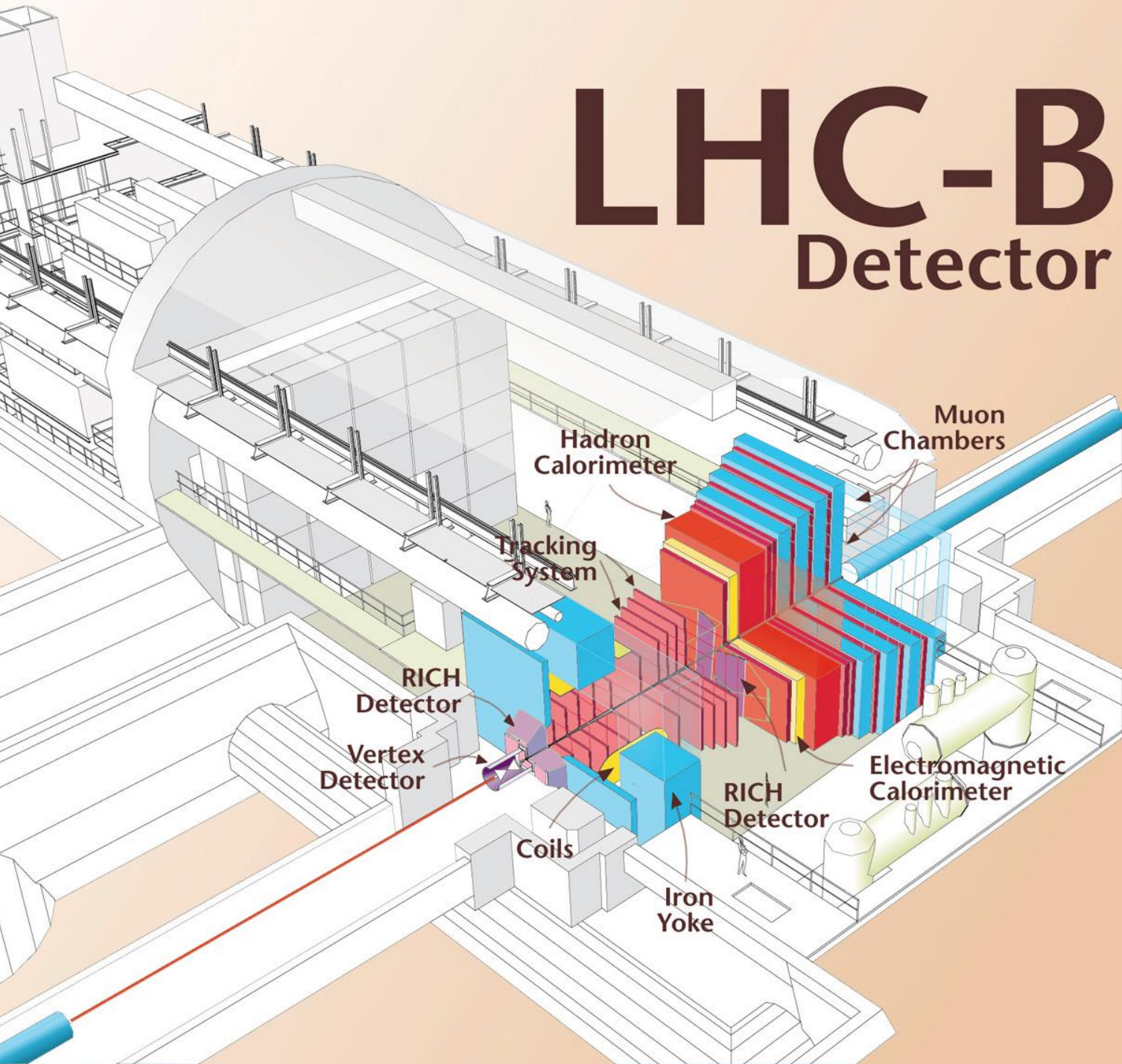
- CERN – *Conseil Européen pour la Recherche Nucléaire*.
- El LHC (en anglès Large Hadron Collider) és el col·lisionador de partícules més gran i poderós del món.
- Túnel de 27 km de circumferència situat entre les fronteres de França i Suïssa
- Detectors principals són ATLAS, CMS, ALICE i LHCb.



DETECTOR LHCB



LHC-B Detector



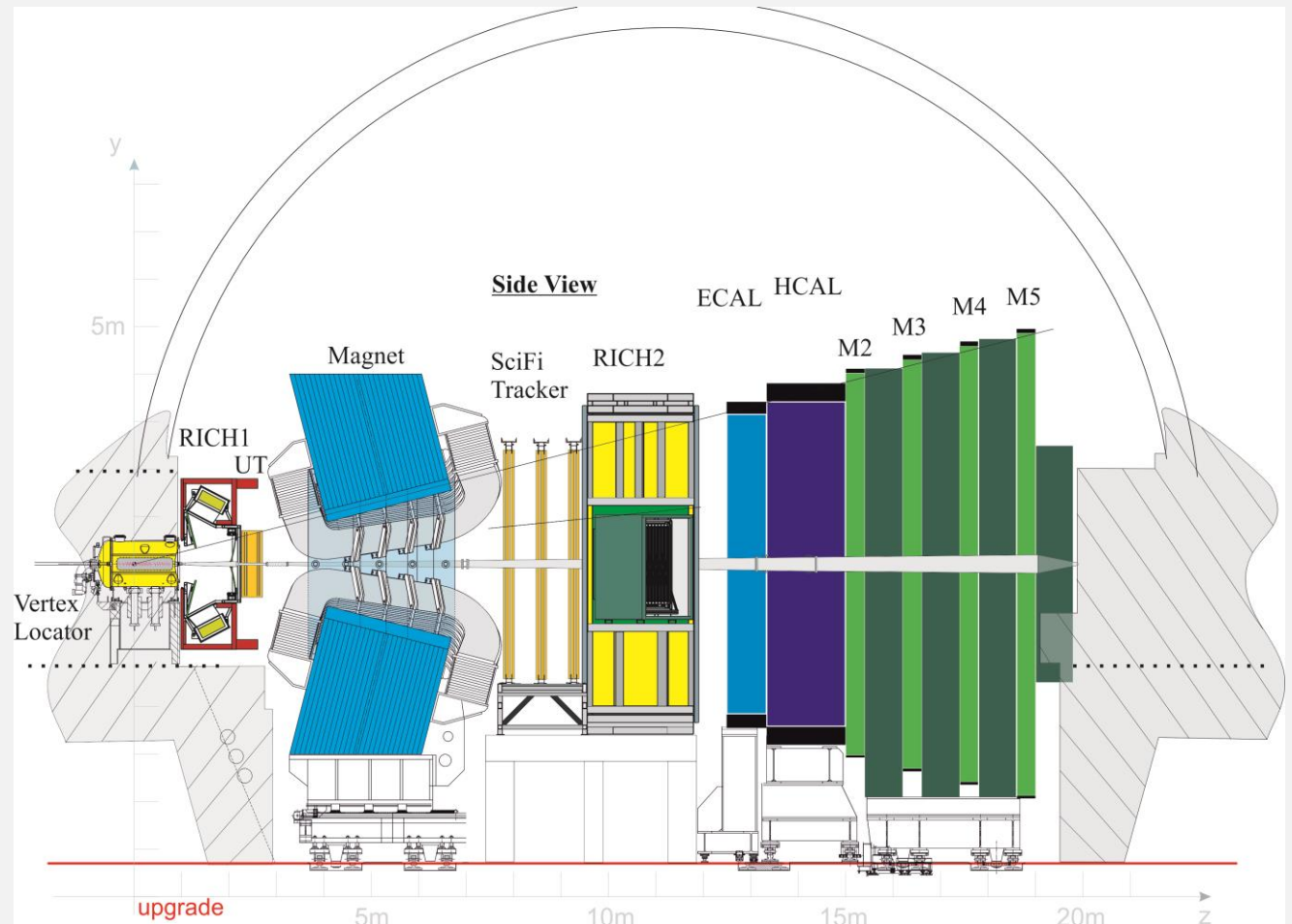
Explicar l'asimetria entre matèria i antimatèria observada a l'Univers.

Utilitza una sèrie de subdetectors per detectar les partícules llançades cap endavant per la col·lisió en una direcció de 18 metres de longitud.

Es realitzarà una actualització que afectarà els subdetectors i al sistema de trigger.

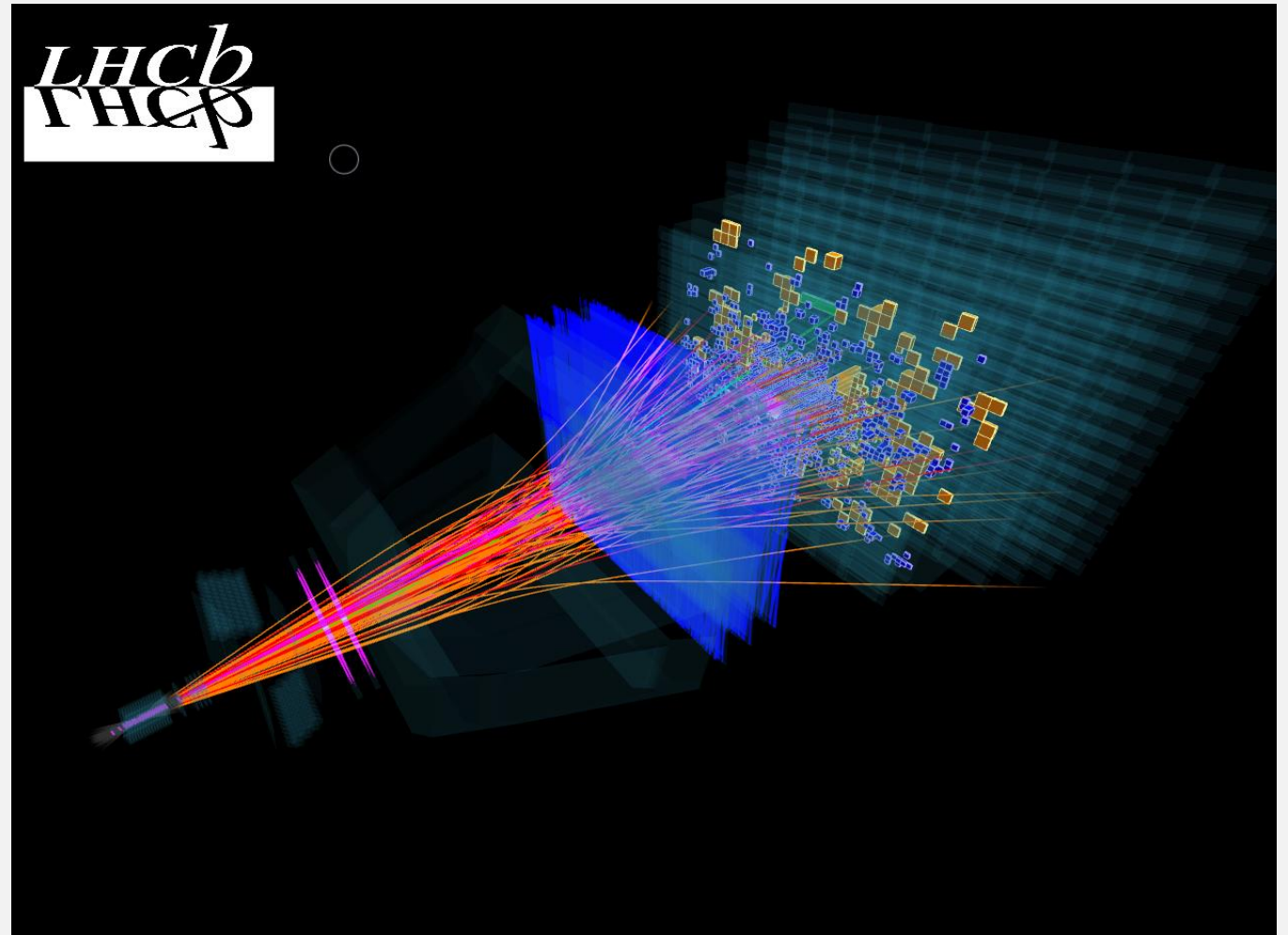
SISTEMES DEL LHCb

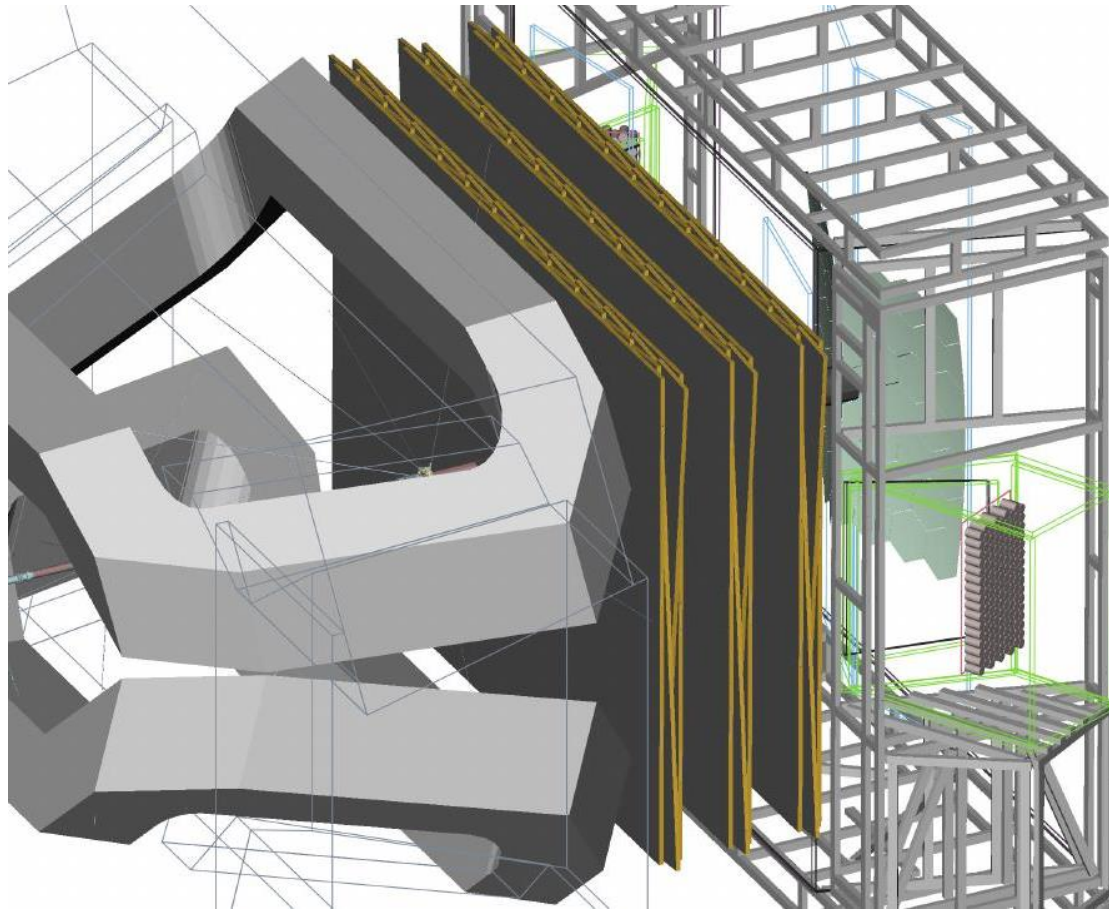
- **Seguiment – Tracking**
 - VELO + UT + Iman + SciFi
- **Identificació**
 - RICH1, 2
 - ECAL i HCAL
 - M1,2,3,4,5
- **Trigger**
 - HLT (High level trigger)



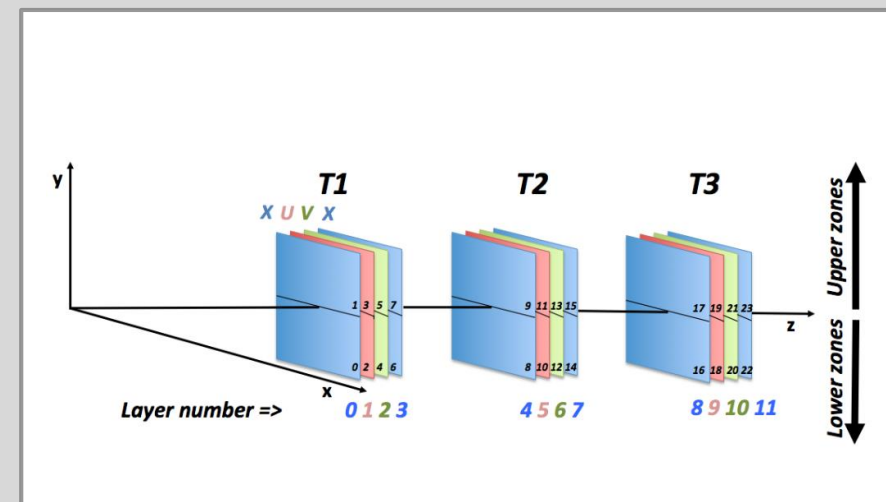
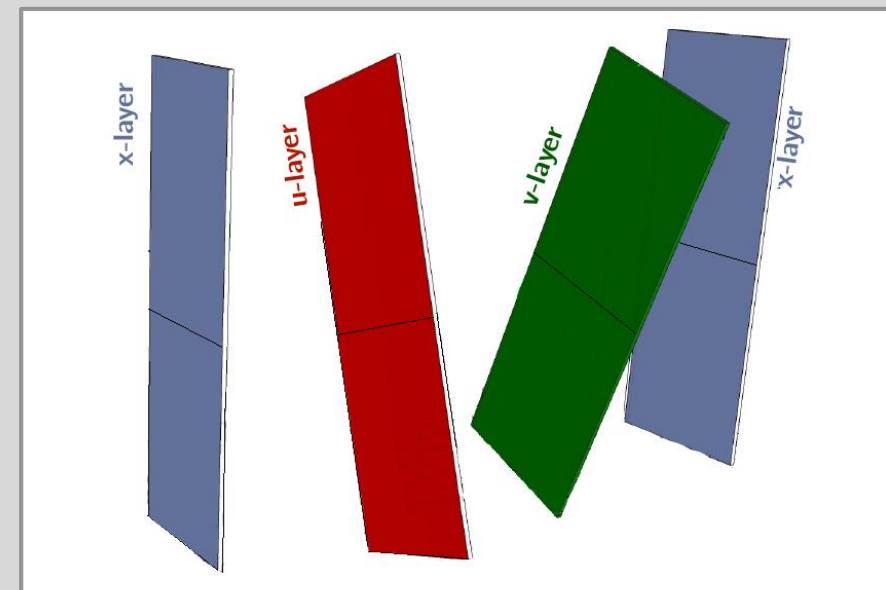
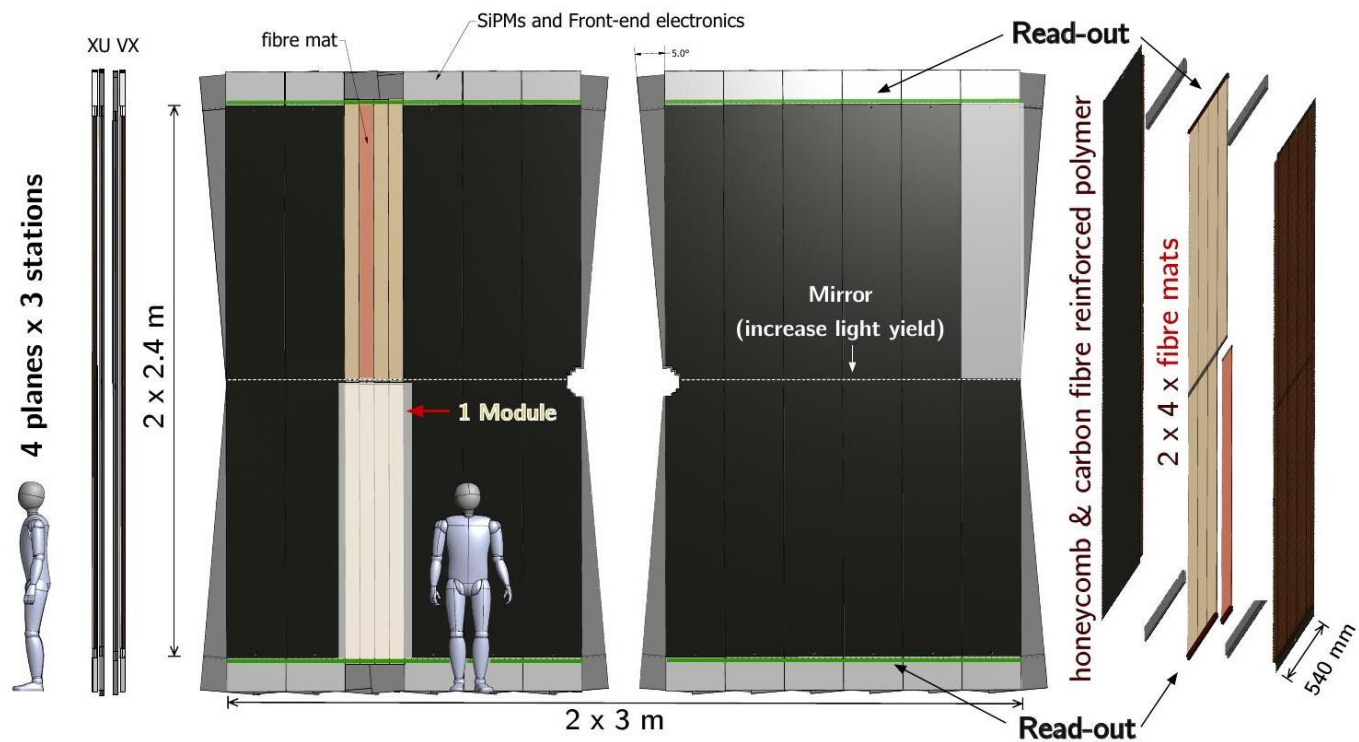
SISTEMA DE SEGUIMENT DE PARTICULES

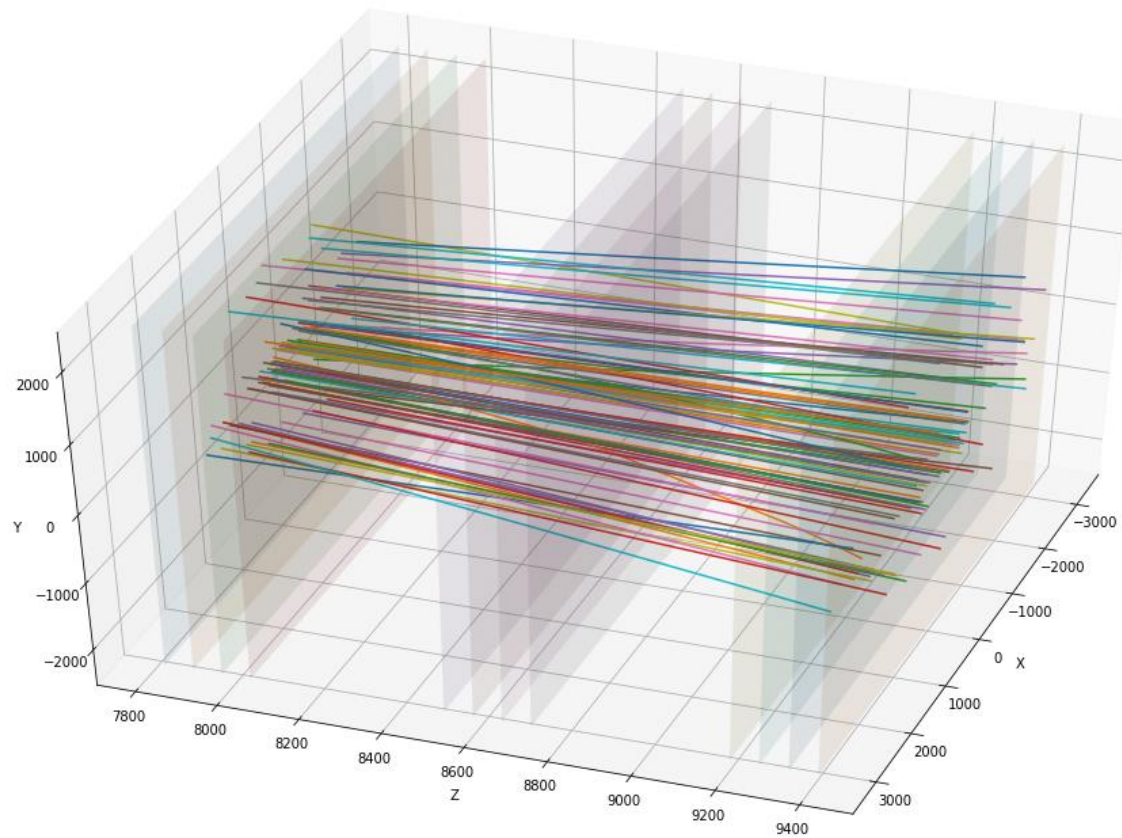
- Proporcionar una reconstrucció eficient de les trajectòries de partícules carregades.
- La seqüència de reconstrucció consta de dues parts.
 - Pattern recognition.
 - Validació dels resultats.





SCINTILLATING FIBRE TRACKER



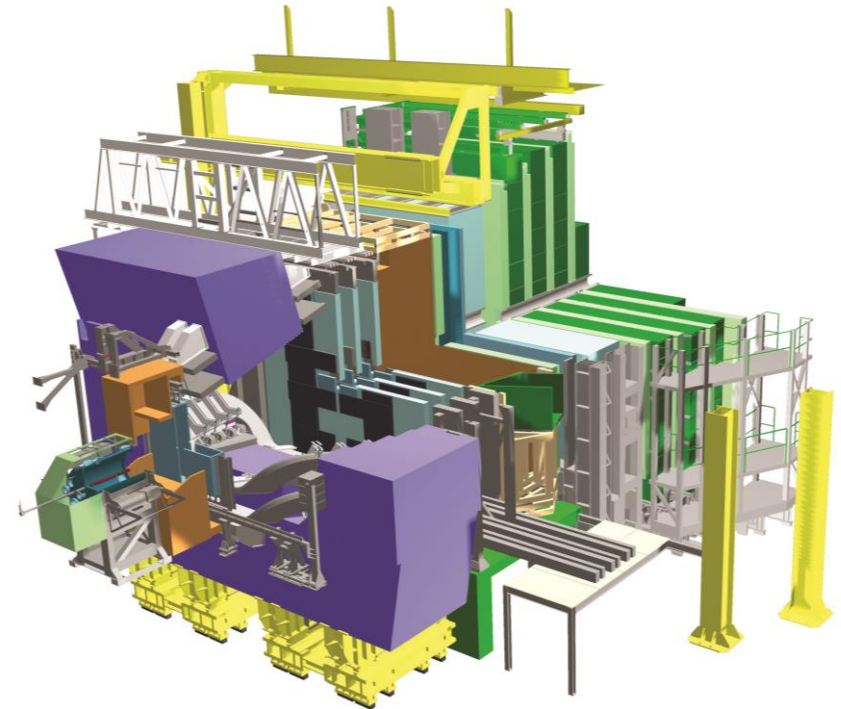


DADES DEL SCIFI

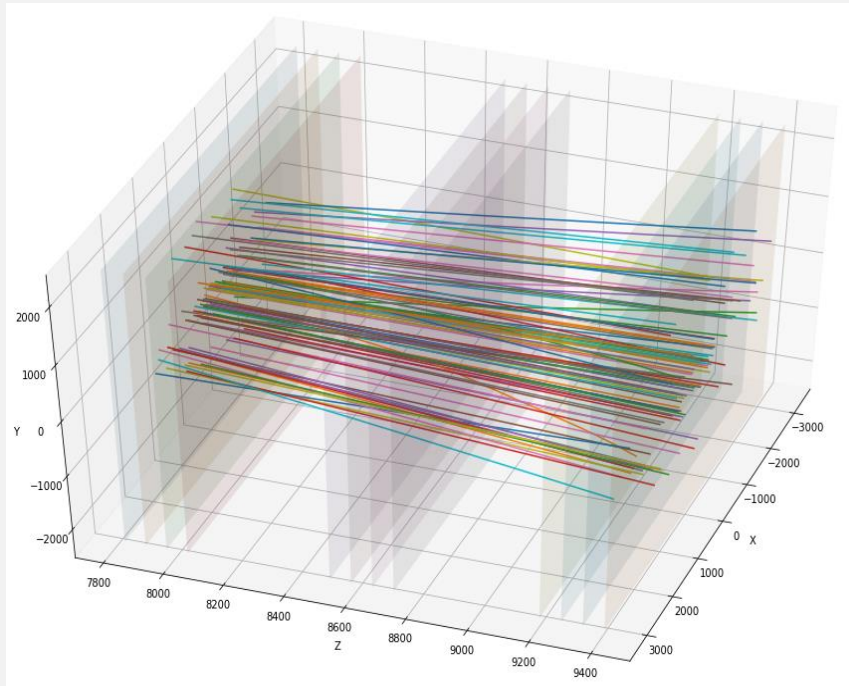
MONTE CARLO

El mètode Monte Carlo permet simular el detector del món real i la seva resposta a una col·lisió de determinades partícules.

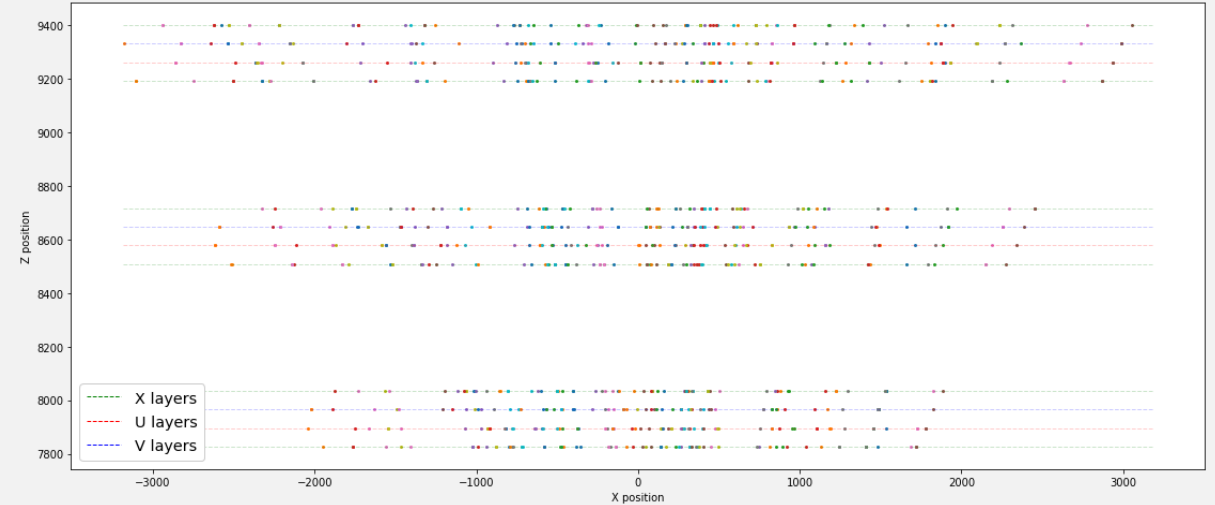
Descripció precisa, des del seu material sensible fins a la cola que enganxa cada mòdul per a una estació de T determinada.



MCHITS



PRHITS

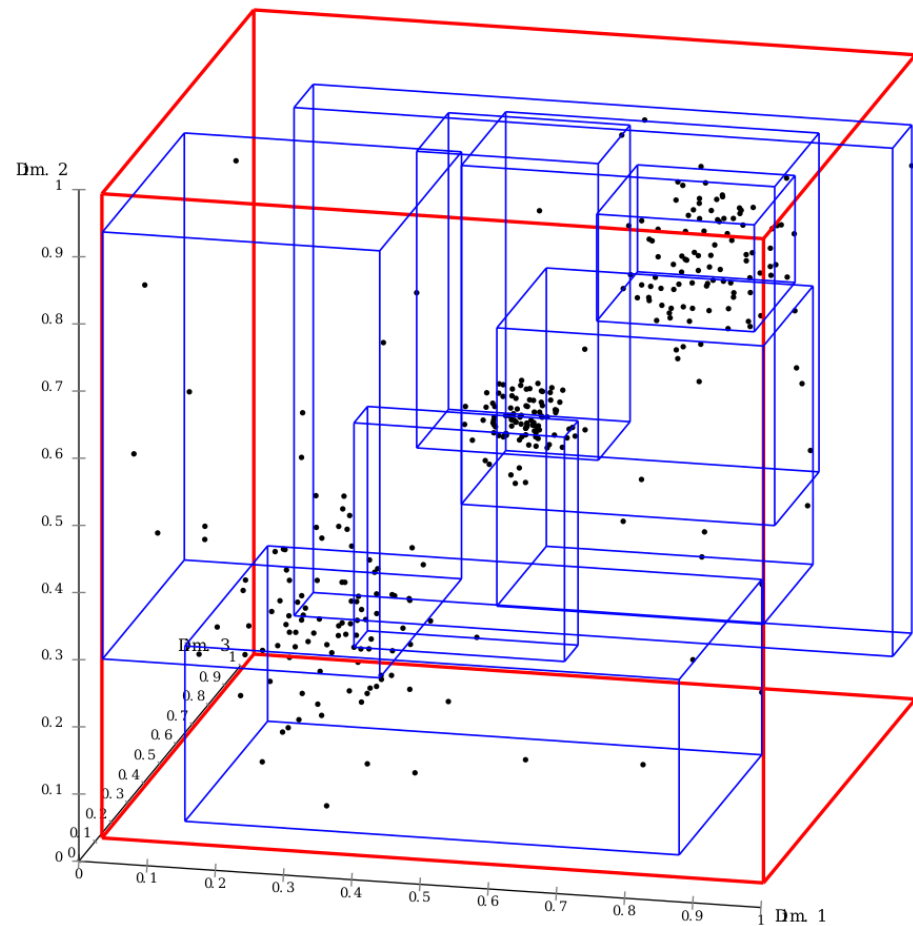


* *Hits* són punts tridimensionals que representen els dipòsits d'energia de les partícules en el subdetector.

ALGORISME

CONCEPTES

- *Hit*: Interacció d'una partícula amb una part sensible d'un detector i processat per tal de ser usat pels algorismes de *pattern recognition*.
- *Seed*: Unió de 2 a 4 hits d'una estació on cada un pertany a una capa determinada.
- *Track*: Conjunt de *seeds* que representa la trajectòria d'una partícula en el detector.
- *Index*: instància d'un r-tree.
- *Ghost rate*: Ràtio de *seeds* incorrectament categoritzades com vàlides però inexistents en la realitat.
- *Clone rate*: Ràtio de *seeds* duplicades del total reconstruït.
- *Esdeveniment*: Col·lisió protó-protó.



R-TREE

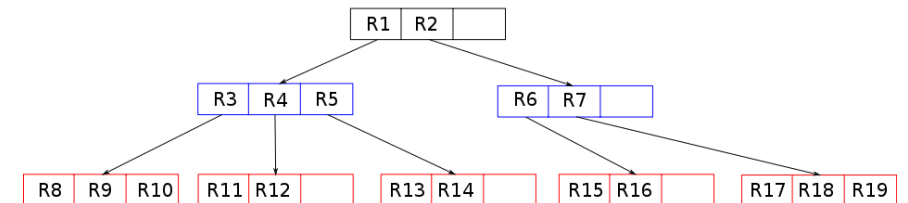
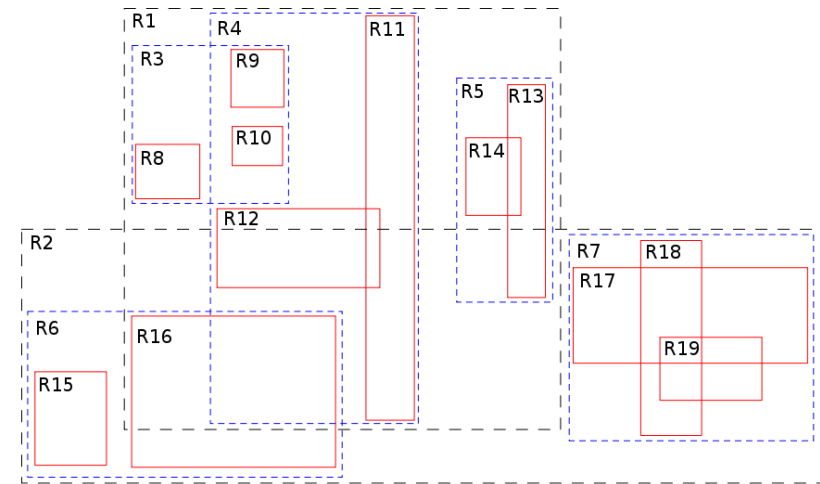
QUE SÓN?

Estructures de dades
forma d'arbre.

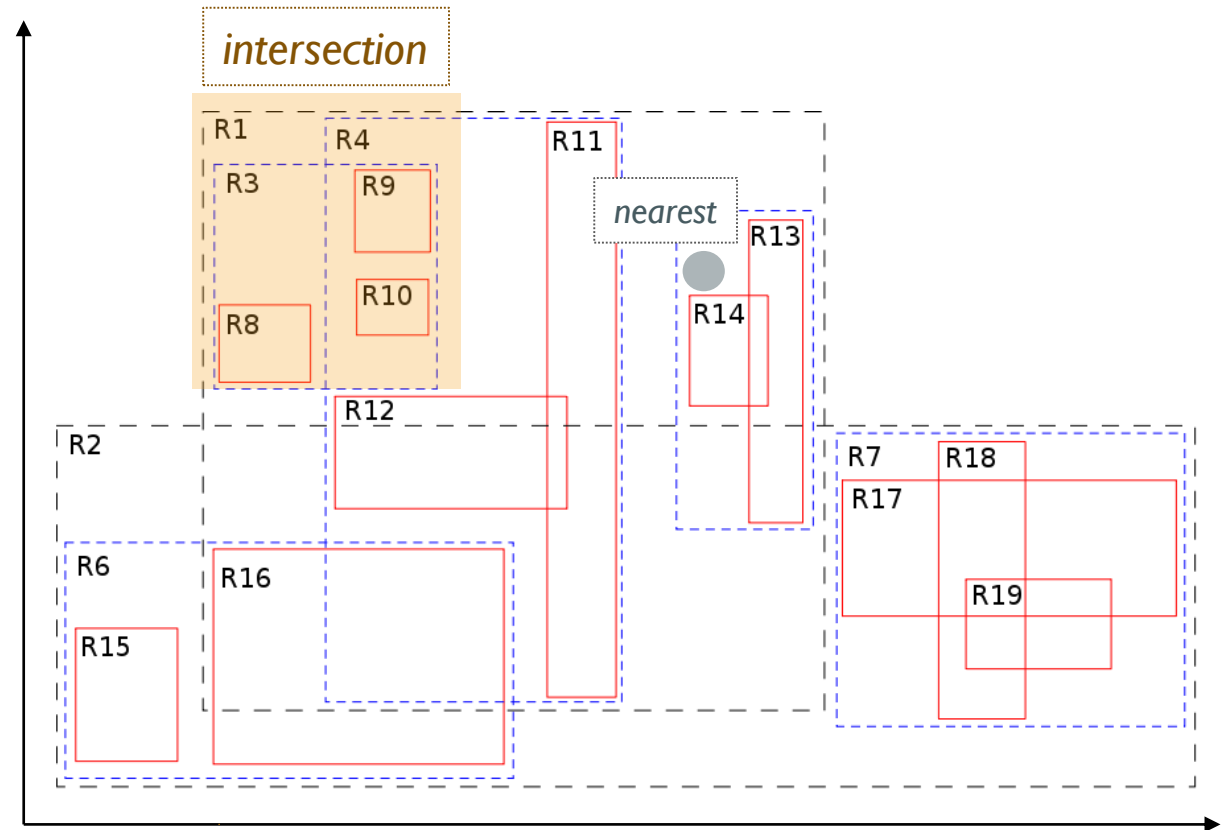
Indexar informació
multidimensional, com
ara coordenades
geogràfiques.

Agrupar els objectes
propers i representar-
los amb el seu
rectangle delimitant
mínim.

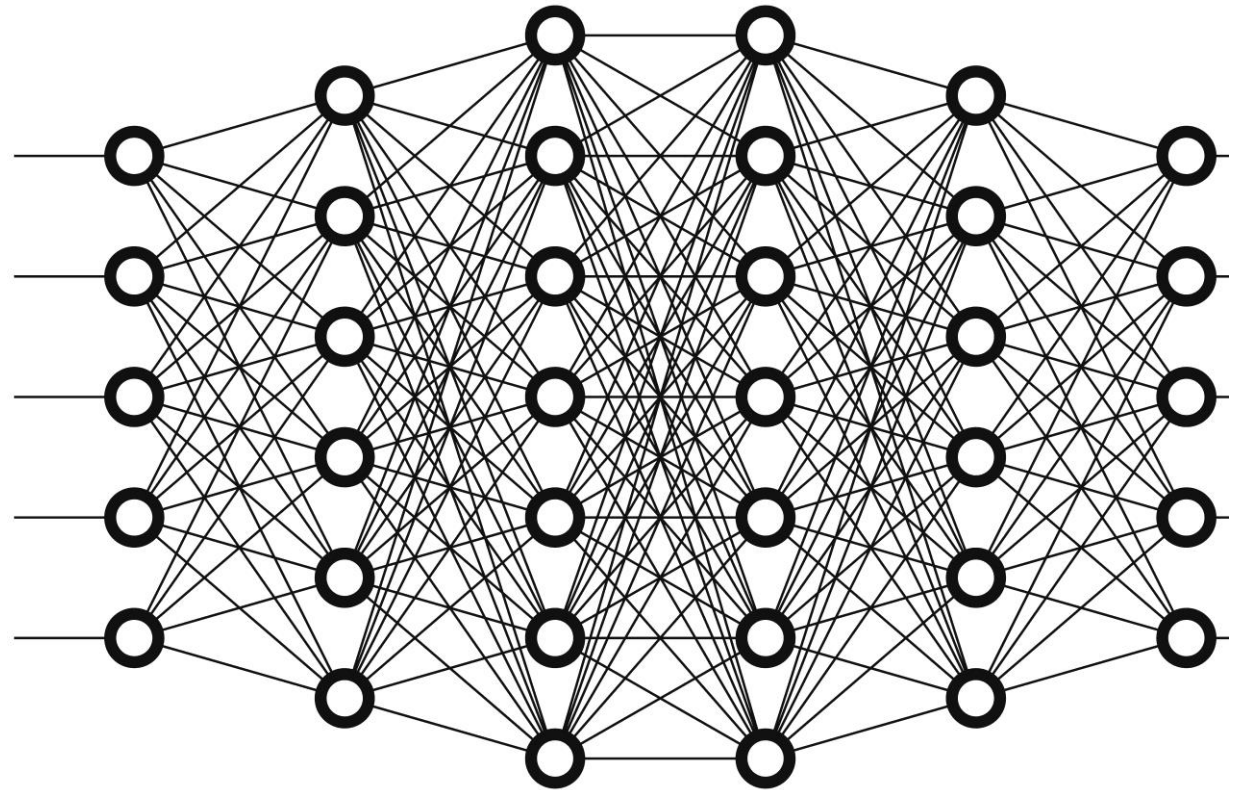
Cost de cerca de
 $O(\log_M(N))$ sent M
entrades per cada
node i N el numero de
nodes.



OPERACIONS



XARXES NEURONALS ARTIFICIALS



QUE SÓN?

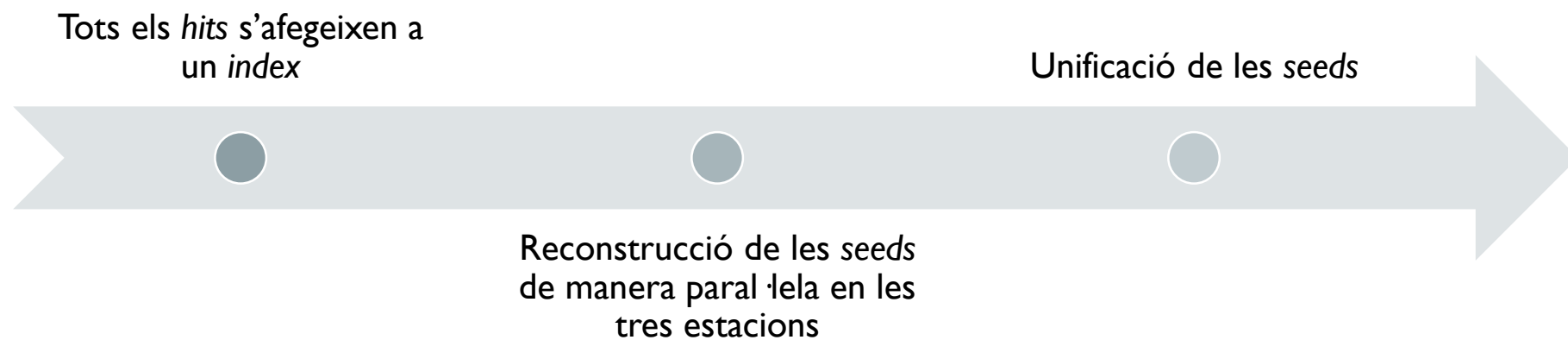
Sistemes computacionals que s'inspiren en les xarxes neuronals biològiques.

“Aprenen” a realitzar tasques considerant exemples, generalment sense ser programats amb cap regla específica de tasques.

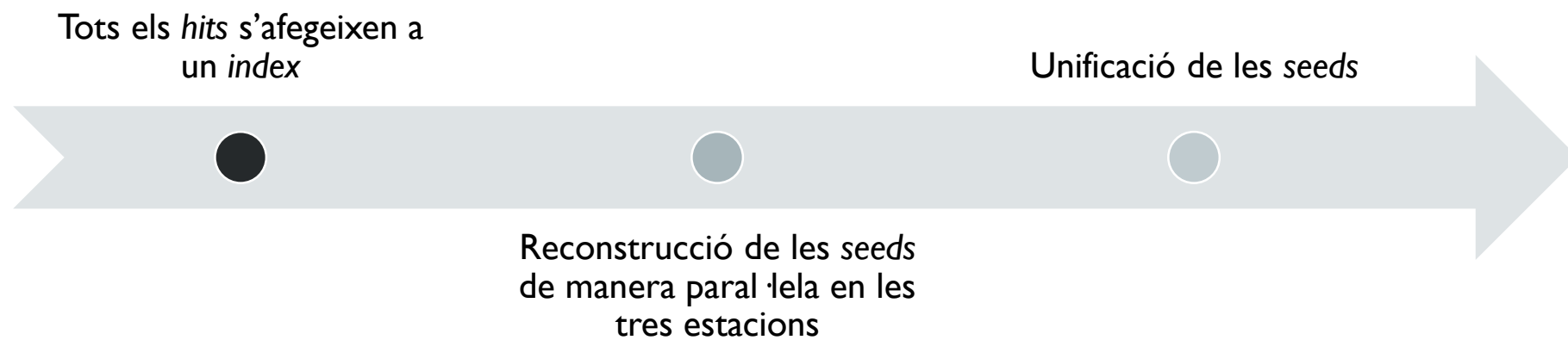
Generen automàticament característiques identificatives del material d'aprenentatge que processen.

Dues de les principals aplicacions de les xarxes neuronals artificials són predir o classificar les dades.

VISIÓ GENERAL



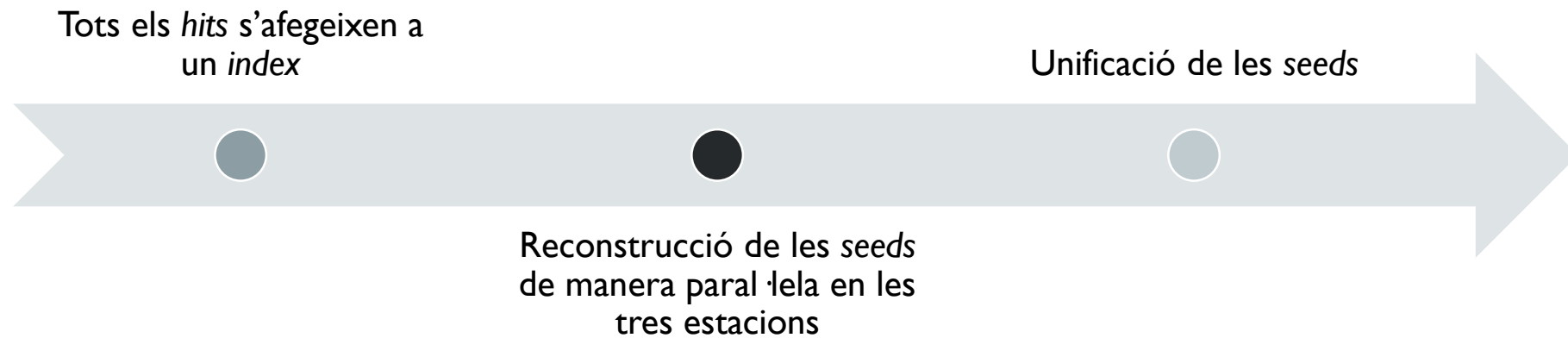
VISIÓ GENERAL



OMPLIR *HITS_INDEX*

- Afegir tots els *hits* a *hits_index* amb les següents característiques:
 - Sistema de coordenades amb 3 dimensions.
 1. Posició X en mil·límetres del hit.
 2. Codi numèric del pla.
 3. Zona superior o inferior en el pla.
- Descartar punts que tinguin una diferència igual o menor a $\pm 1\text{mm}$, per tal de disminuir el *clone rate*.

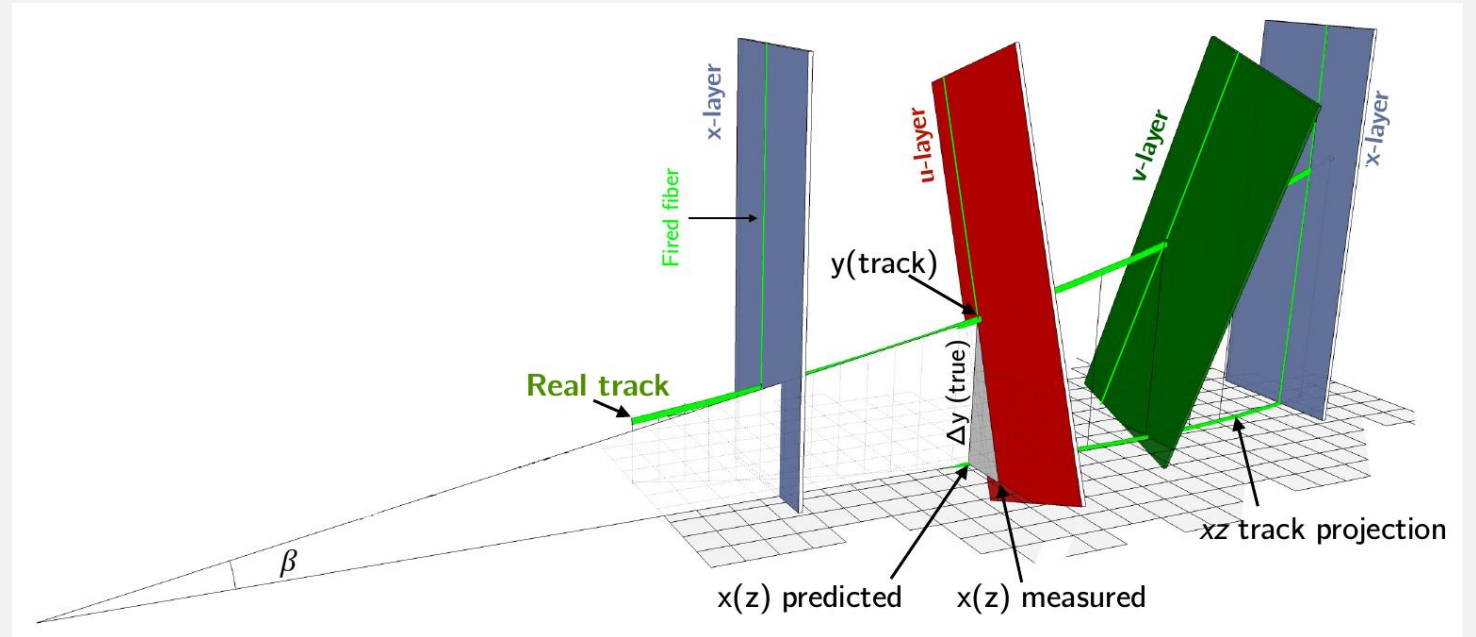
VISIO GENERAL



RECONSTRUCCIÓ DE SEEDS PER ESTACIÓ

- Unificar *hits* per formar possibles *seeds* en cada estació.
- Tres fils d'execució, un per cada estació.
- Les *seeds* reconstruïdes per cada un s'afegeixen a un *index* nou amb les següents característiques:
 - Sistema de coordenades amb 3 dimensions:
 1. Posició X en mil·límetres del primer *hit*.
 2. Predicció de l'altura de la *seed*.
 3. Numero d'estació (0 a 2).

PREDICCIÓ DE L'ALTURA



Projecció d'una línia a partir de la posició (0,0)

Obtenir l'angle beta:

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

Calcular la possible posició X de la partícula en una capa determinada.

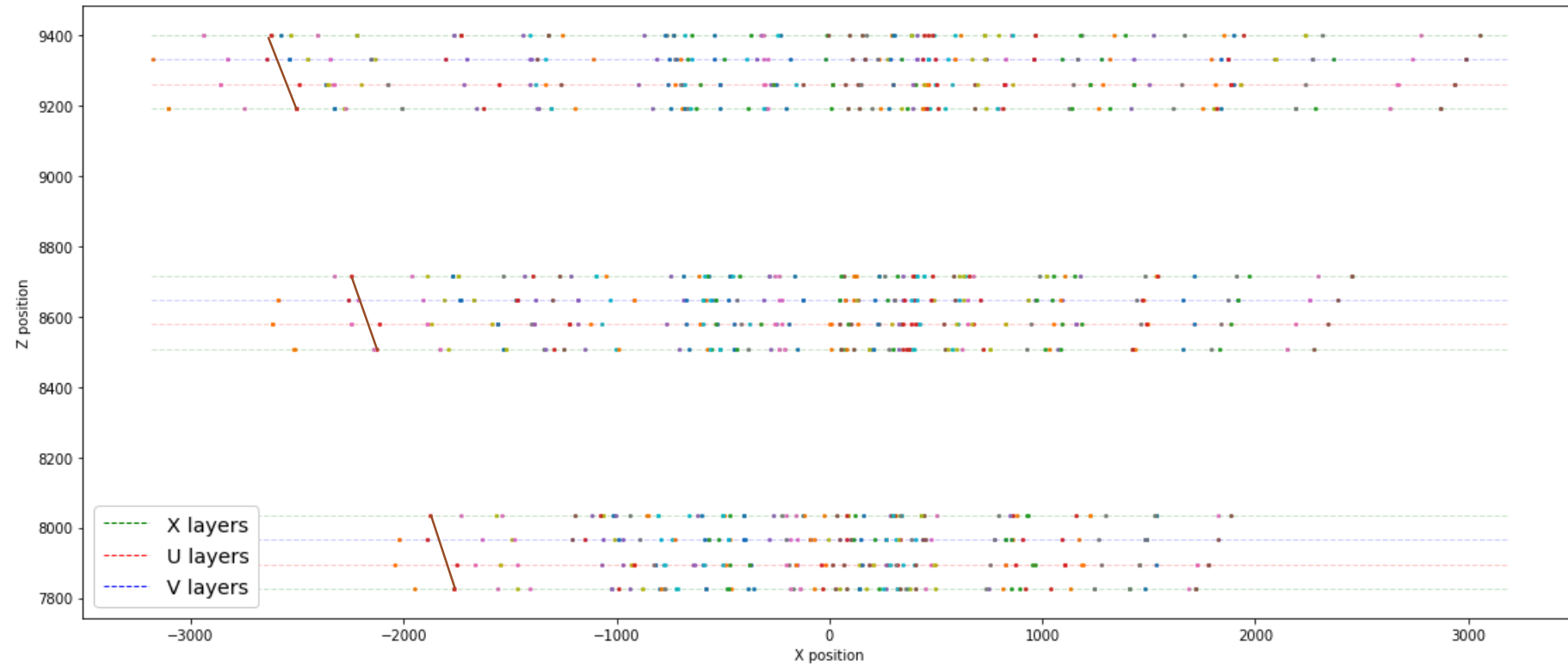
L'altura finalment s'obté mitjançant:

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

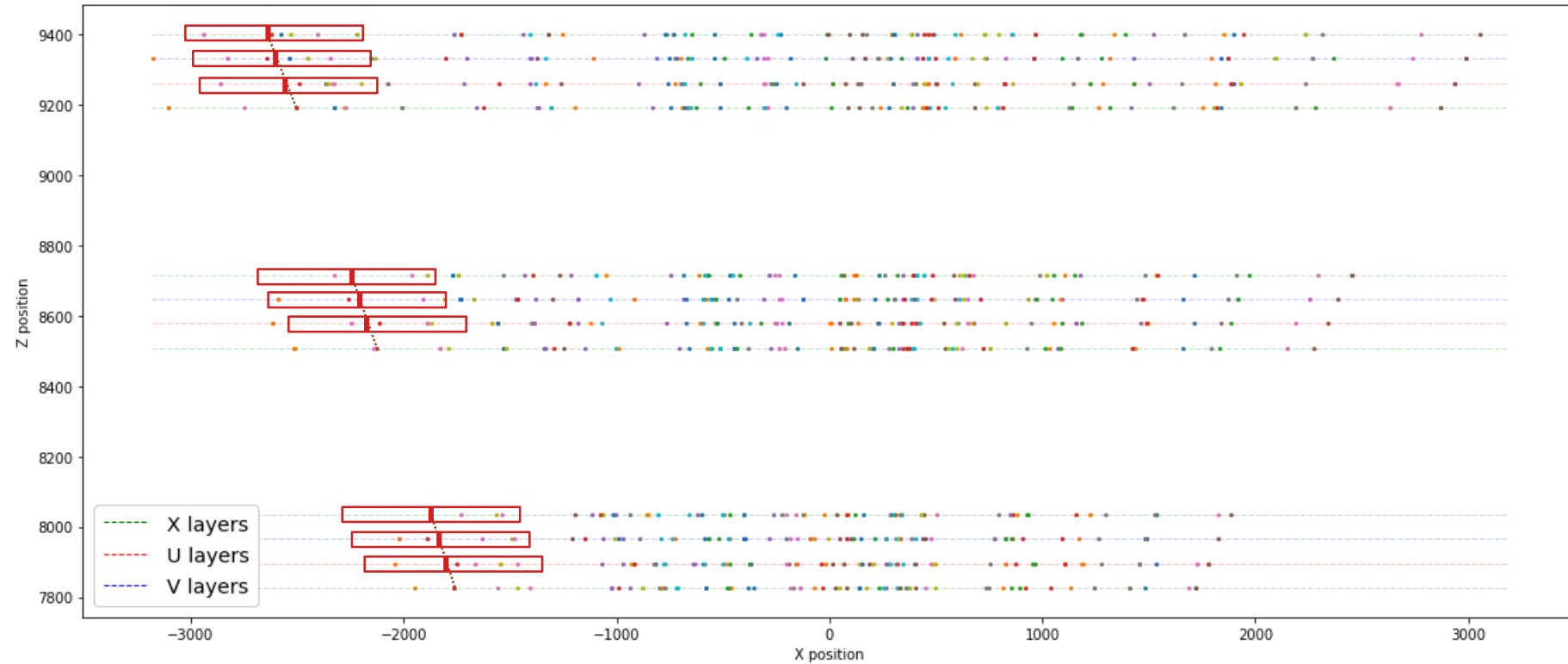
RECONSTRUCCIÓ DE SEEDS PER ESTACIÓ

1. Seleccionar tots els *hits* de *hits_index* mitjançant les següents coordenades (X_{min} , X_{max} , X_{0_layer} , X_{0_layer} , 0, 1).
2. Per cada un d'aquest:
 1. Predir mitjançant la projecció bàsica d'una línia la posició de la partícula en les següents capes U-V-X.
 2. Aquests valors $PredX_U$, $PredX_V$ i $PredX_{X1}$ s'utilitzen per accedir al *hits_index* per obtenir els possibles *hits* de la partícula en les capes U-V- X_0 :
 - Intersecció a ($PredX_U - limit$, $PredX_U + limit$, $U_{planecode}$, $U_{planecode}$, Hit_{zone} , Hit_{zone}) amb format (X_{min} , X_{max} , Y_{min} , Y_{max} , Z_{min} , Z_{max})

RECONSTRUCCIÓ DE SEEDS PER ESTACIÓ



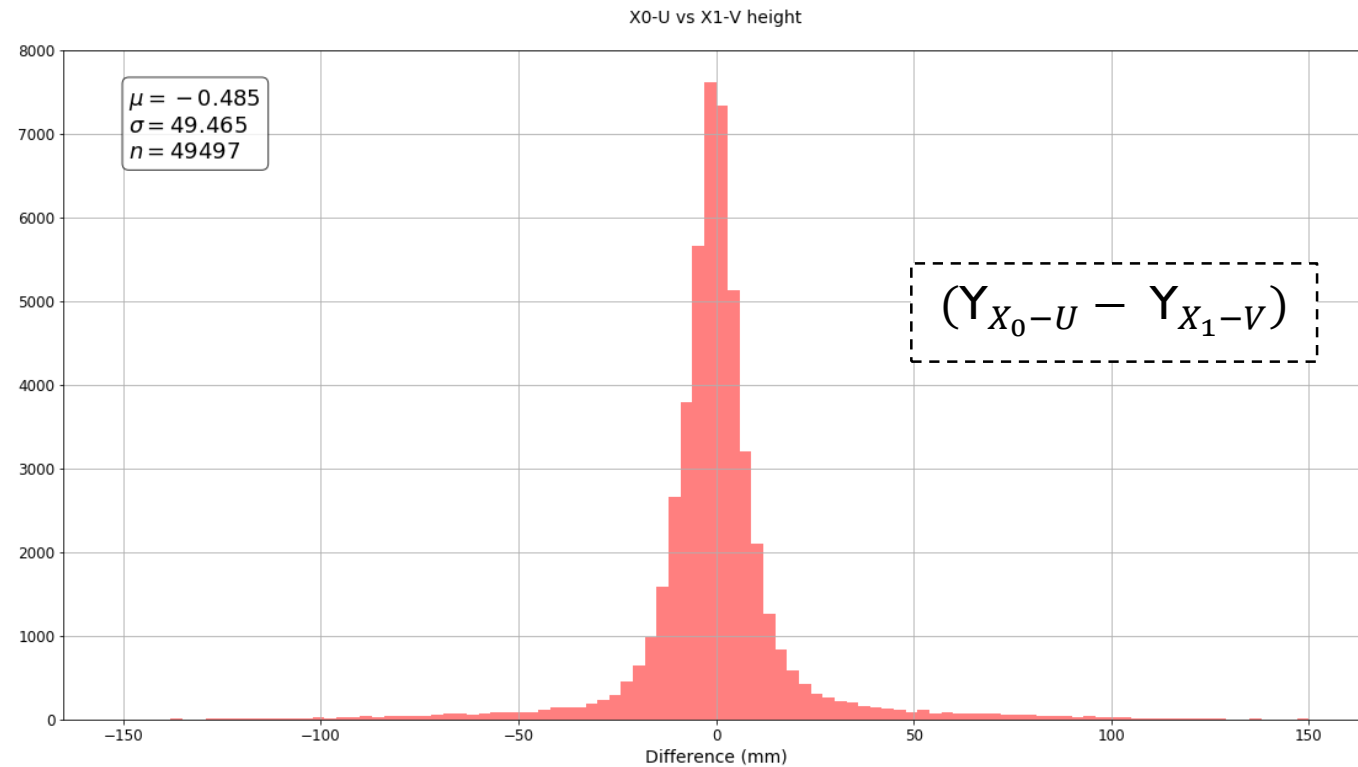
RECONSTRUCCIÓ DE SEEDS PER ESTACIÓ



RECONSTRUCCIÓ DE SEEDS PER ESTACIÓ

3. Un cop obtingut els *hits* per cada capa:
 1. Si en alguna no n'ha retornat cap, es creen *seeds* incompletes.
 2. En cas de tenir un nombre major a 0 de *hits* per cada capa, es calcula l'altura de cada possible combinació i es formen possibles *seeds*.
 3. Totes aquestes que hi hagi una diferència entre Y_{X_0-U} i Y_{X_1-V} major a 50 mm son descartades.
4. Se selecciona la *seed* que millor probabilitat de combinació vàlida té mitjançant el model d'ANN anomenat *seed_station_validation*.
5. En cas d'obtenir una probabilitat major al 20%, s'introdueix en l'index *seeds_index*.

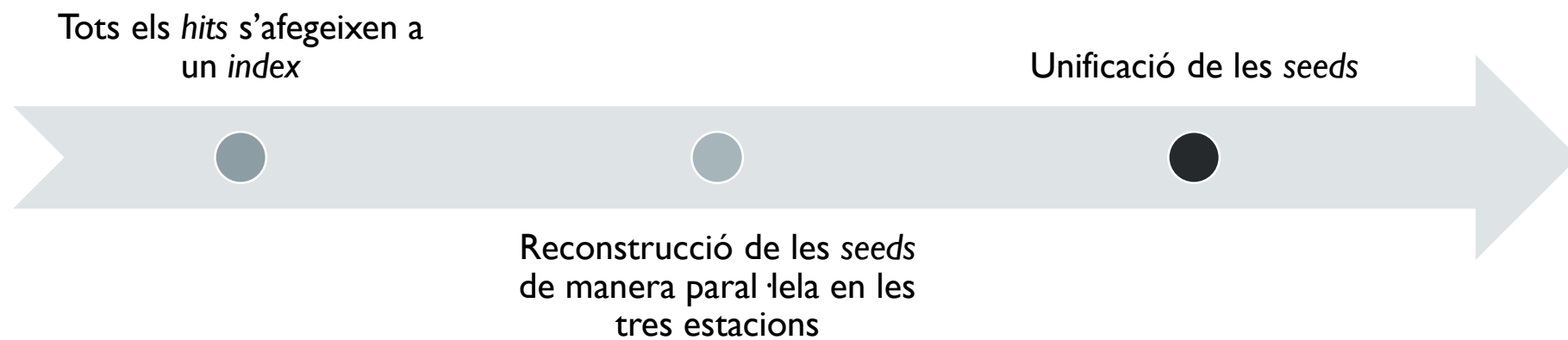
RECONSTRUCCIÓ DE SEEDS PER ESTACIÓ



RECONSTRUCCIÓ DE SEEDS PER ESTACIÓ

3. Un cop obtingut els *hits* per cada capa:
 1. Si en alguna no n'ha retornat cap, es creen *seeds* incompletes.
 2. En cas de tenir un nombre major a 0 de *hits* per cada capa, es calcula l'altura de cada possible combinació i es formen possibles *seeds*.
 3. Totes aquestes que hi hagi una diferència entre Y_{X_0-U} i Y_{X_1-V} major a 50 mm son descartades.
4. Se selecciona la *seed* que millor probabilitat de combinació vàlida té mitjançant el model d'ANN anomenat *seed_station_validation*.
5. En cas d'obtenir una probabilitat major al 20%, s'introdueix en l'index *seeds_index*.

VISIÓ GENERAL



UNIFICACIÓ DE SEEDS

1. Obtenir totes les seeds de la primera estació del *rtree* anomenat *index_seeds*.
2. Per cada una d'aquestes:
 1. Si el 25% dels hits que conté ja han estat escollits per traveses completes, es descarta.
 2. En cas contrari, es fa una predicció de la posició del primer *hit* de la *seed* en la següent estació mitjançant la projecció d'una línia.
3. A partir del valor, es realitza una intersecció en l'arbre utilitzant rangs determinats:
 - $(PredX - 100, PredX + 100, mean - 150, mean + 150, I, I)$ amb format $(X_{min}, X_{max}, Y_{min}, Y_{max}, Z_{min}, Z_{max})$
 - $mean = \frac{(Y_{X_0-U} + Y_{X_1-V})}{2}$

UNIFICACIÓ DE SEEDS

4. De tots els seeds obtinguts per la *intersecció* anterior, se selecciona el que millor probabilitat té de ser vàlid mitjançant el model d'ANN anomenat *two_seeds_validation*.
5. En cas que aquesta tingui més del 25% de *hits* ja seleccionats per una altra combinació, es descarta i torna al pas 2.
6. A partir del resultat anterior, es realitza de nou una intersecció a l'última estació mitjançant la predicció de la posició utilitzant la projecció d'una línia.
7. Se selecciona de totes les seeds obtingudes per l'anterior procediment, la que millor probabilitat té de ser vàlida juntament amb el resultat de l'apartat 4 i es comprova el mateix que l'apartat 5.
8. Finalment, si la combinació de les tres seeds passa tots els controls, es desa com una *track* vàlida i tots els *hits* usats es marquen.

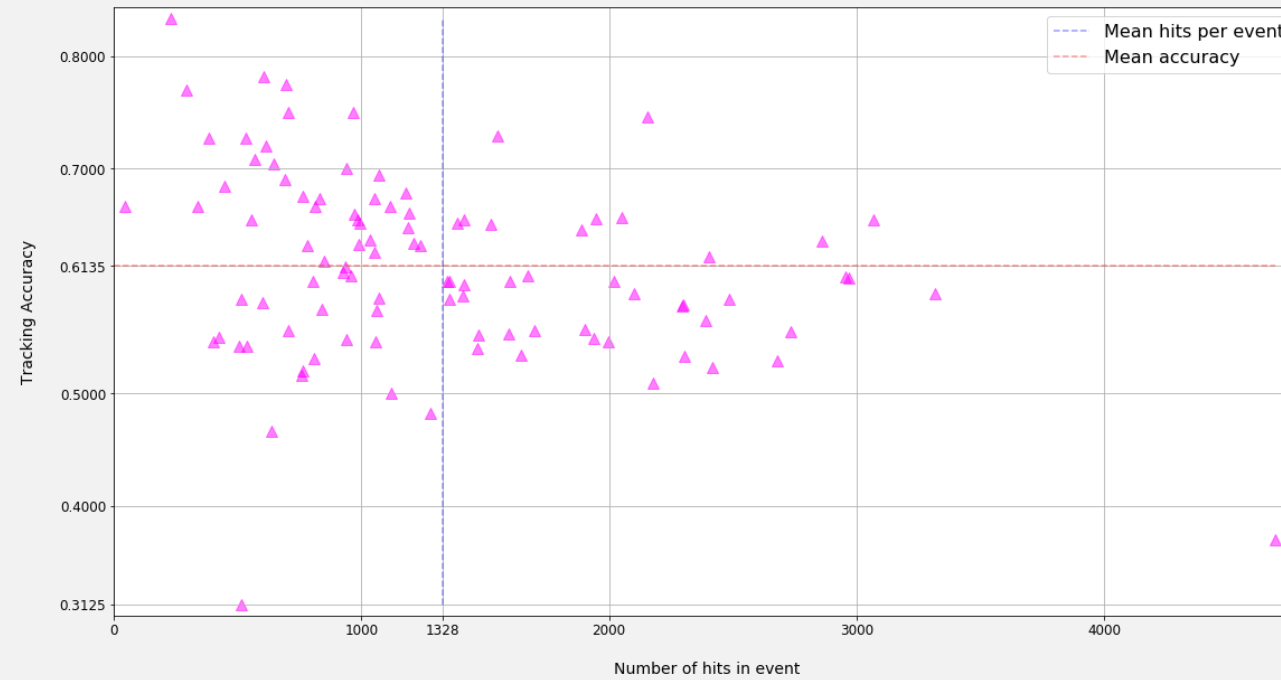
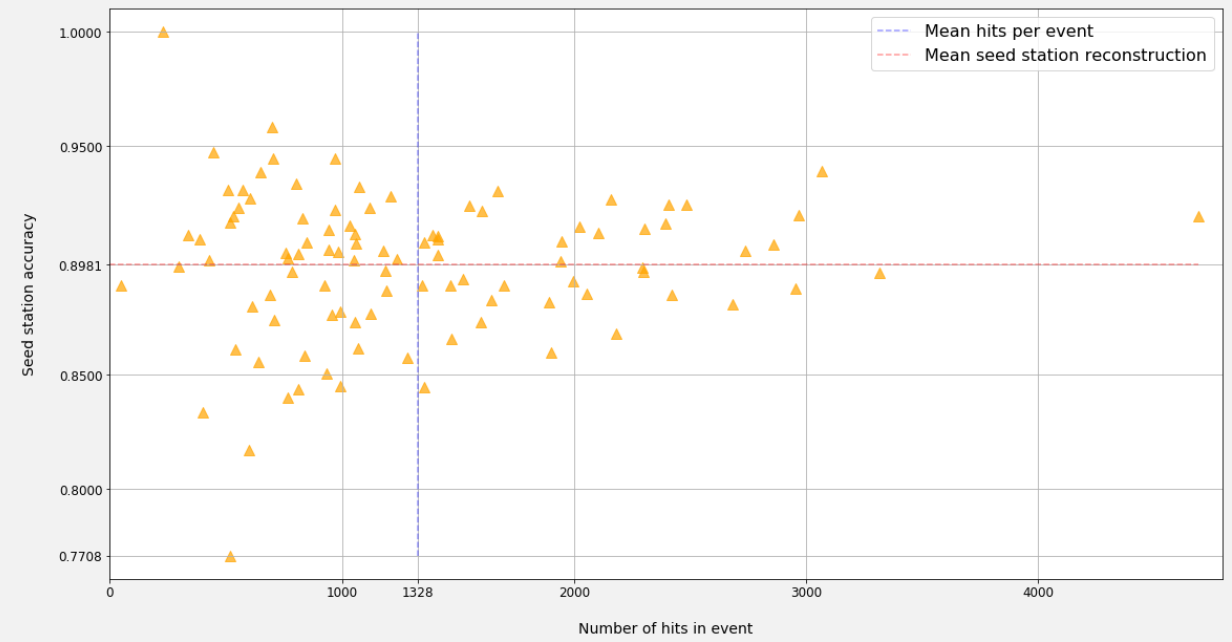
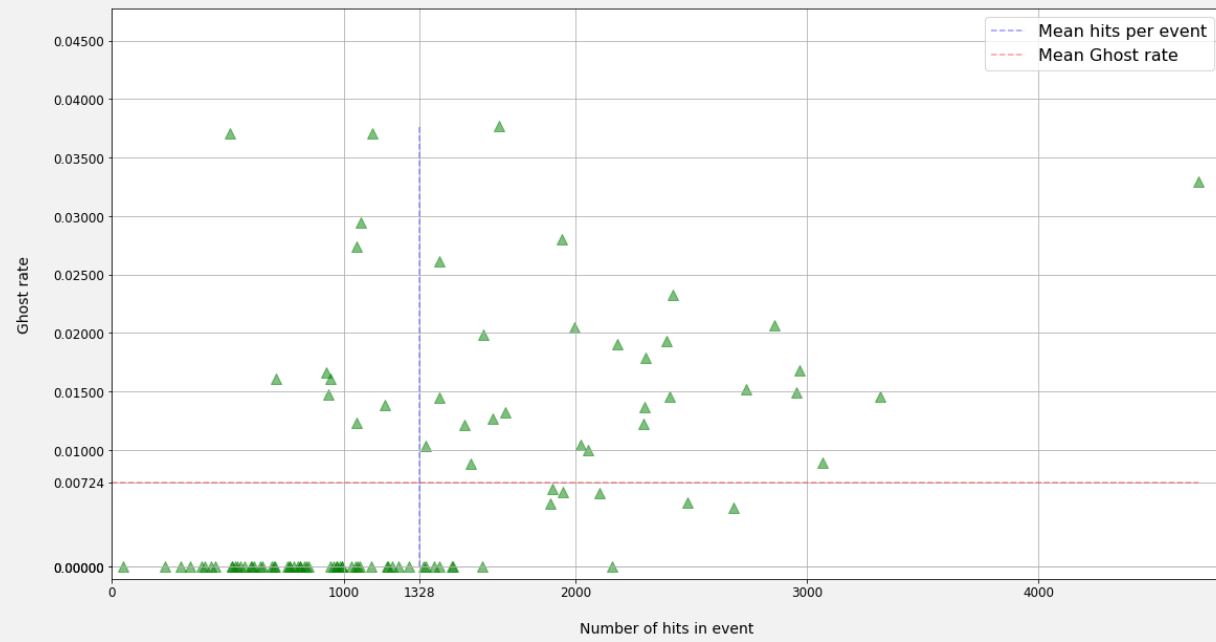
RENDIMENT

INDICADORS I CONSIDERACIONS

- Traces *reconstruïdes*: nombre total de traces que l'algoritme és capaç de reconstruir.
- Traces *reconstruïbles*: nombre total de traces que és possible reconstruir.
- Eficiència de *Tracking* = $\frac{\text{reconstruïdes \& valides}}{\text{reconstruïbles}}$
- *Ghost rate* = $\frac{\text{reconstruïdes not valides}}{\text{reconstruïdes}}$
- Perquè una traça reconstruïda sigui vàlida, ha de coincidir amb un mínim de 70% dels *hits* amb una *reconstruïble*.
- Una partícula és *reconstruïble* si té *hits* per lo menys en una de les dues capes X i U o V.

RESULTATS

	Eficiencia	Ghost rate	Reconstrucció de seeds
Actual	$(67.2 \pm 0.1) \%$	$(7.9 \pm 0.1) \%$	-
Algorisme	$(47.3 \pm 2.1) \%$	$(1.1 \pm 0.2) \%$	$(86.5 \pm 0.1) \%$
Algorisme – Traces sense falta de <i>hits</i>	$(74.8 \pm 2.8) \%$	$(0.7 \pm 0.2) \%$	$(96.6 \pm 0.1) \%$
Algorisme – Traces amb màxim un <i>hit</i> menys	$(53.2 \pm 2.6) \%$	$(1.2 \pm 0.2) \%$	$(90.4 \pm 0.1) \%$



CONCLUSIONS

El rendiment està 20% per sota de l'actual.

Decrement del *ghost rate* d'un 86% respecte a l'actual implementació.

El baix rendiment ve donat en la part d'unificació i no en la reconstrucció de *seeds*.

Major nombre de *hits* per esdeveniment redueixen l'eficiència, principalment causat per l'ús de rangs fixos quan es cerca en els arbres.

LÍNIES DE FUTUR

1

Implementació d'un sistema millorat per la unificació dels seeds a partir de l'ús d'heurístiques determinades.

2

Crear una metodologia de rangs dinàmics (per les zones exteriors més grans i interiors reduïts).

3

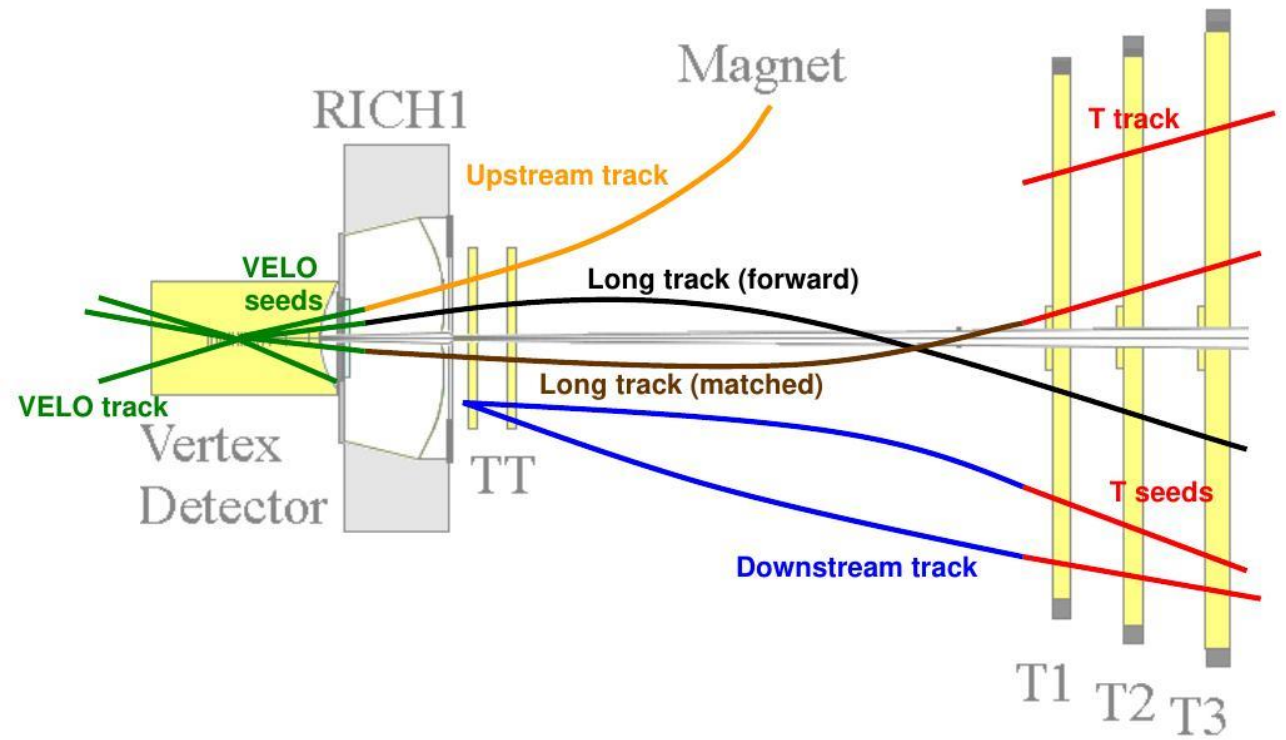
Realitzar una investigació estesa en els models d'ANN, ja que tenen un gran impacte en l'eficàcia de l'algorisme.

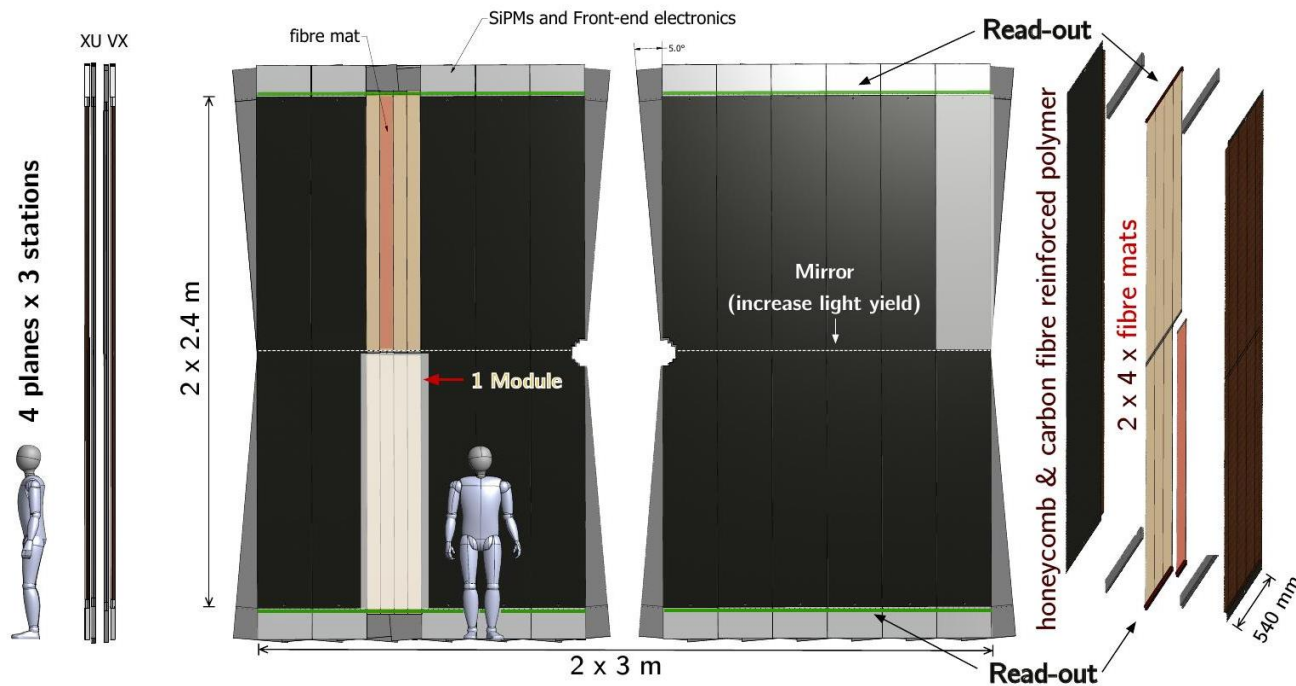
GRÀCIES PER LA VOSTRE ATENCIÓ

Alguna pregunta?

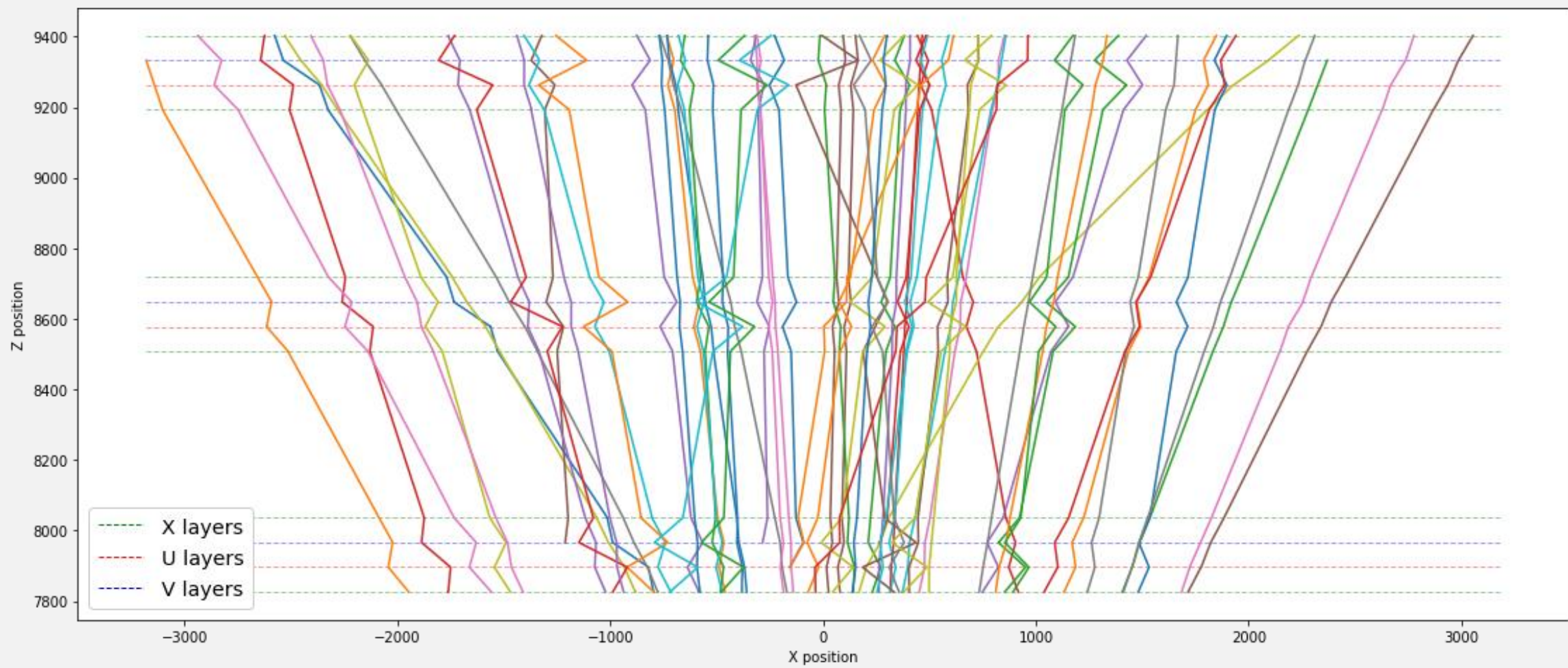
BACKUP

PISTES



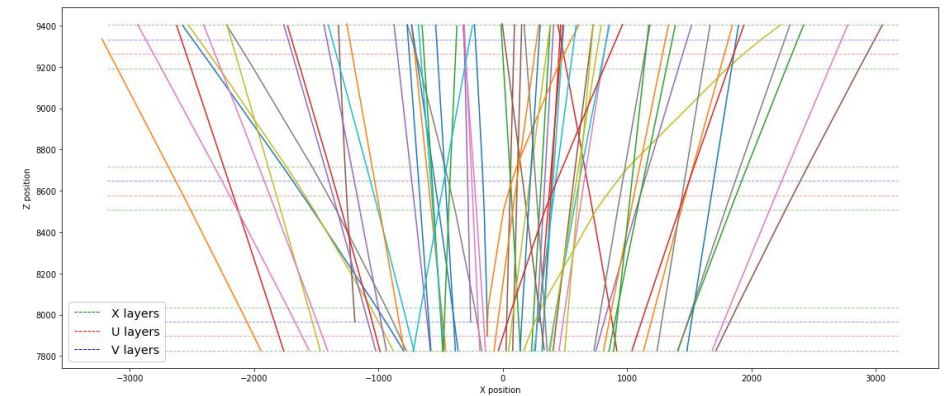
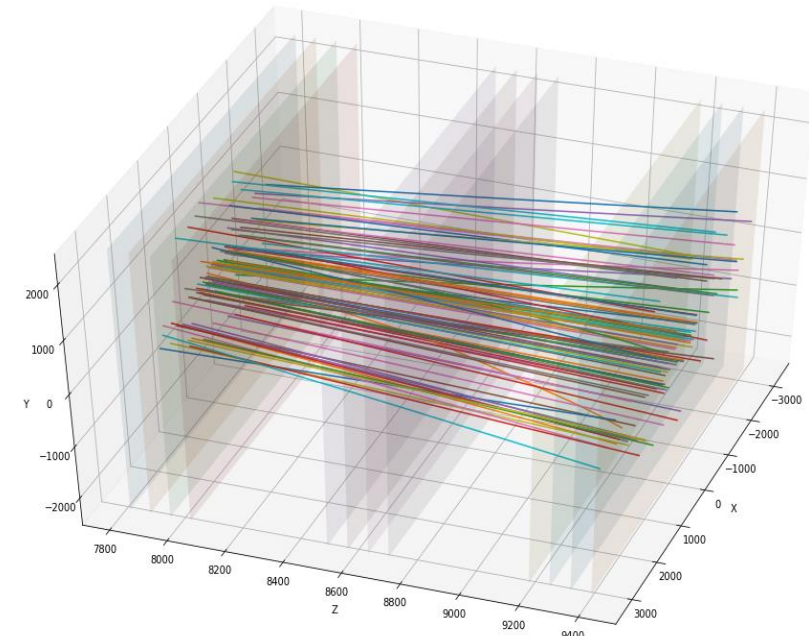


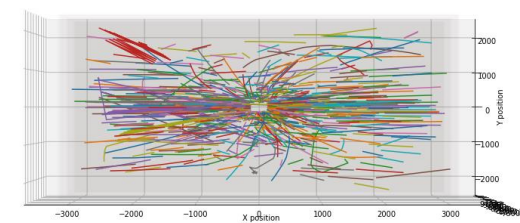
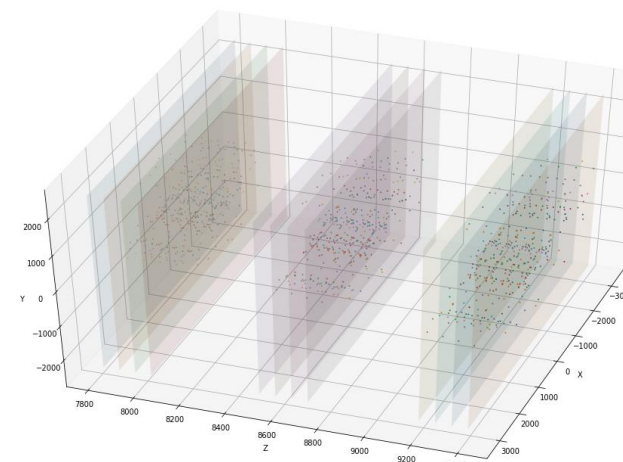
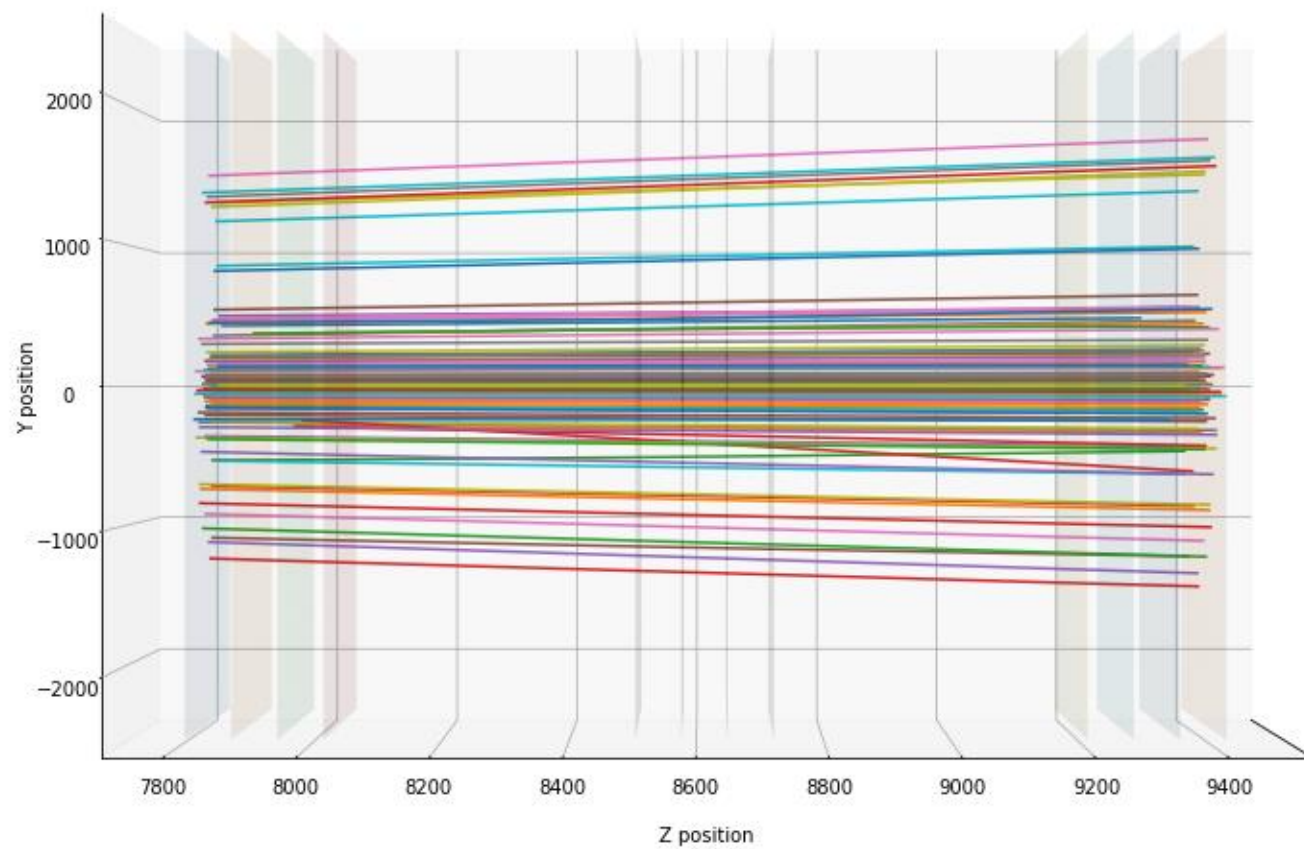
- Cada estació està formada per 4 capes diferents (X-U-V-X).
- Capes X estan orientades verticalment (posició x) i les U-V formen uns +/- 5 graus (posició y).
- Cada capa - 12 mòduls, cadascun d'aquests conté 8 matrius fibril·lats amb 6 capes apilades de fibres de 2,5 m de llarg.
- Els Silicon PhotoMultipliers (SiPM) es troben a les vores d'aquests mòduls per tal de recollir la llum produïda i transportada per les fibres.



MCHITS

- *Hits* són punts tridimensionals que representen els dipòsits de energia de les partícules en el sub-detector.
- Punts tri-dimensionals que representen *hits* de les partícules en els diferents sub-detectors.
- S'utilitzen per validar el rendiment dels algoritmes de *pattern recognition*.



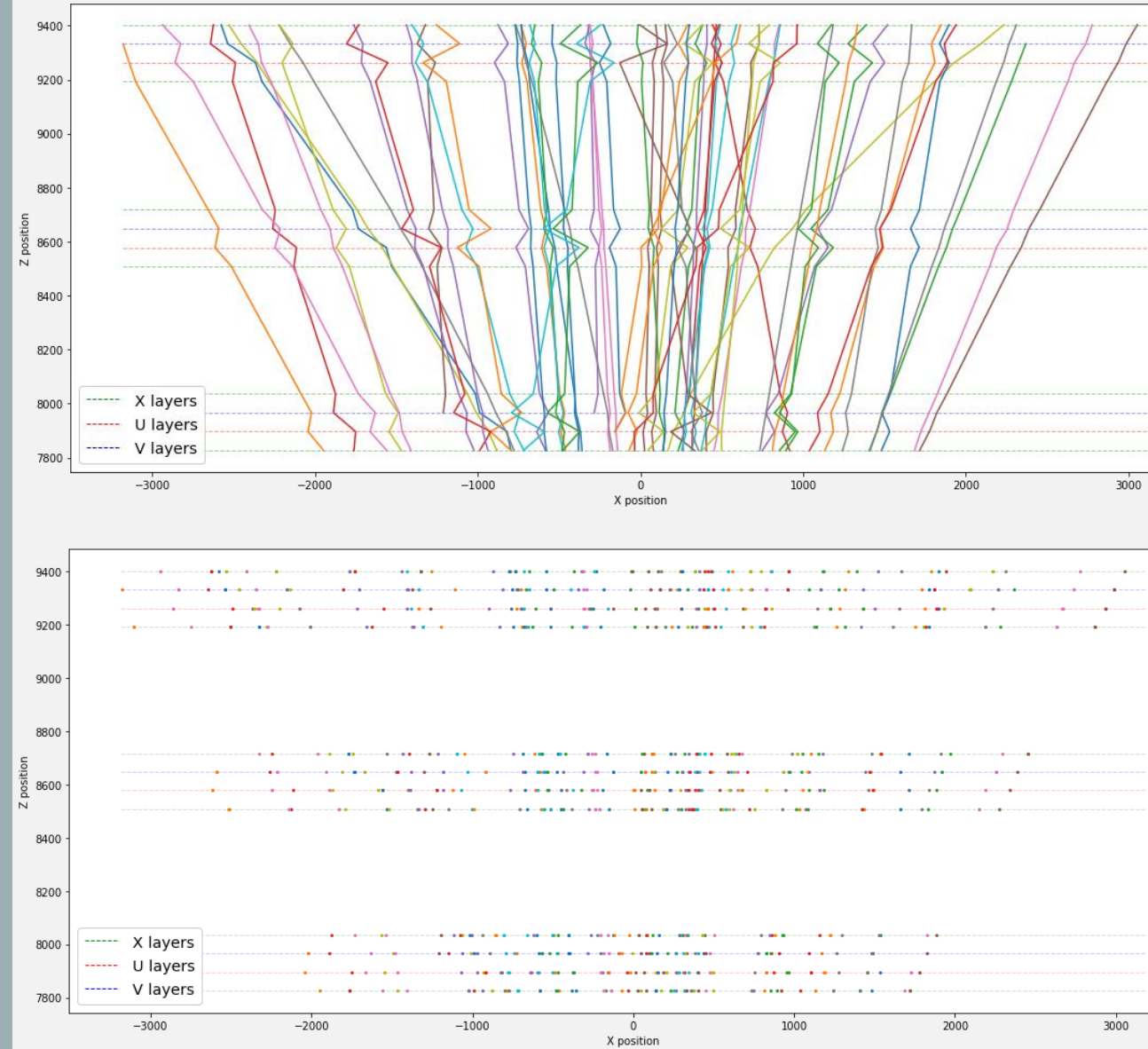


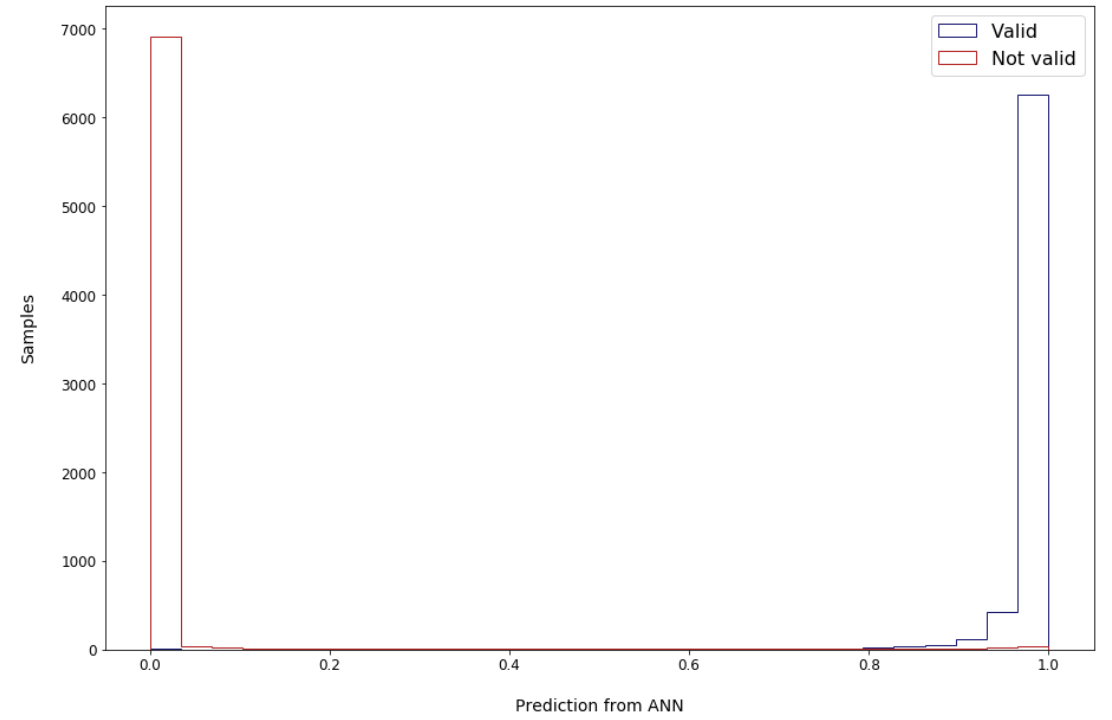
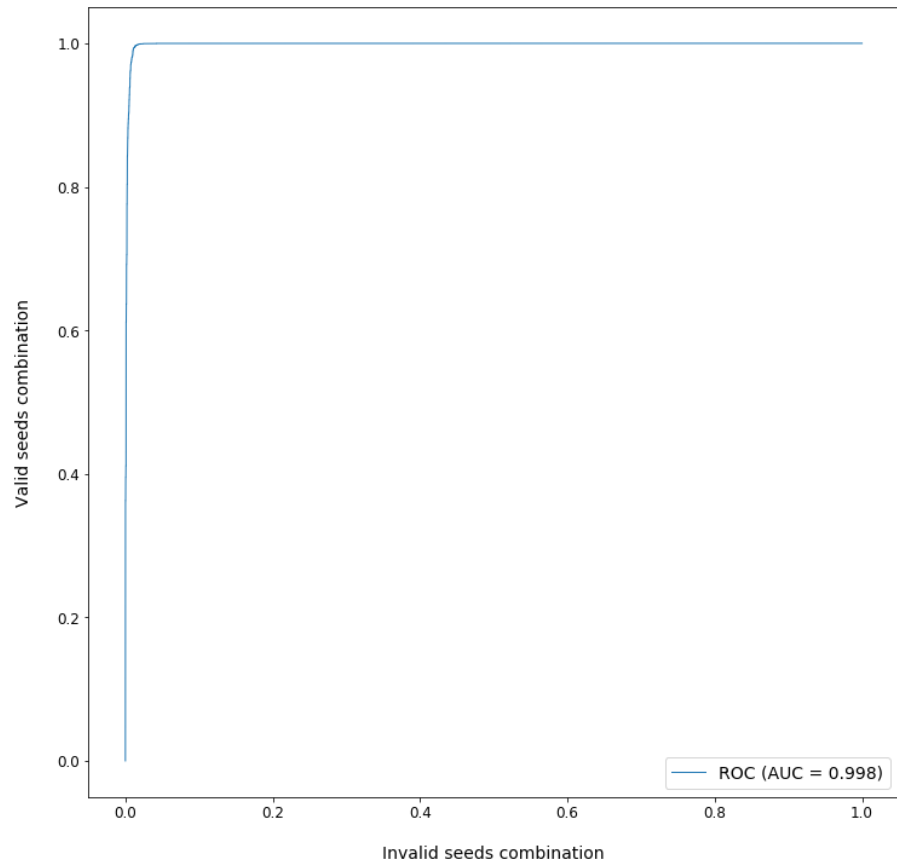
PRHITS

Representen una aproximació real de la sortida SciFi en un esdeveniment real.

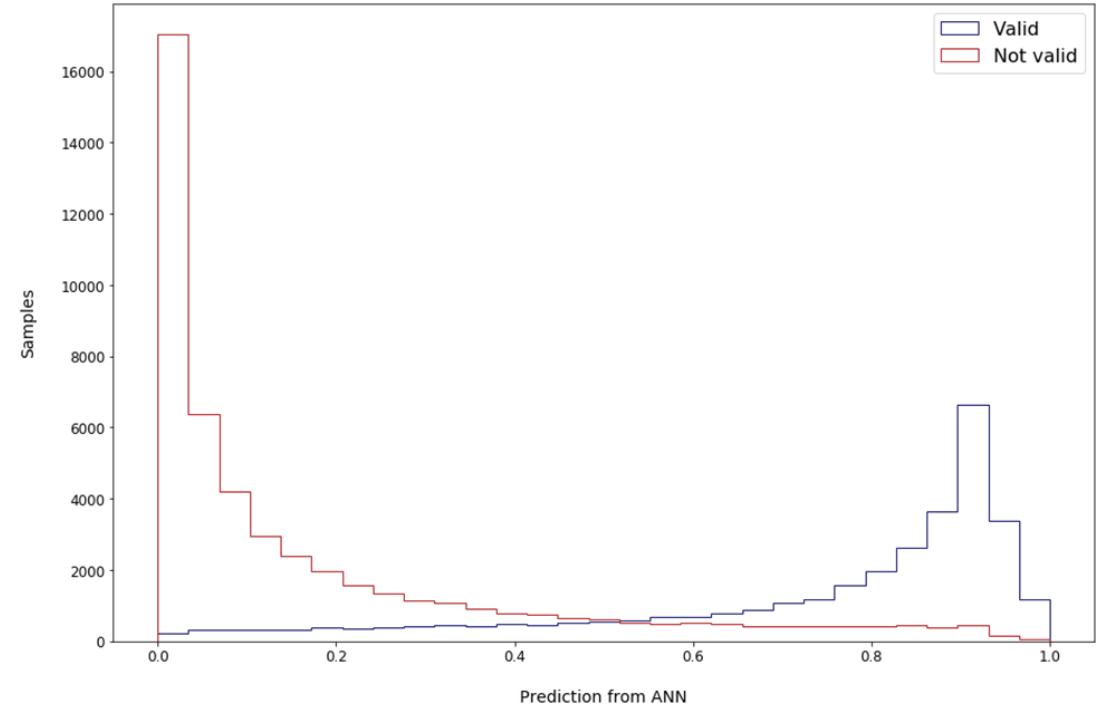
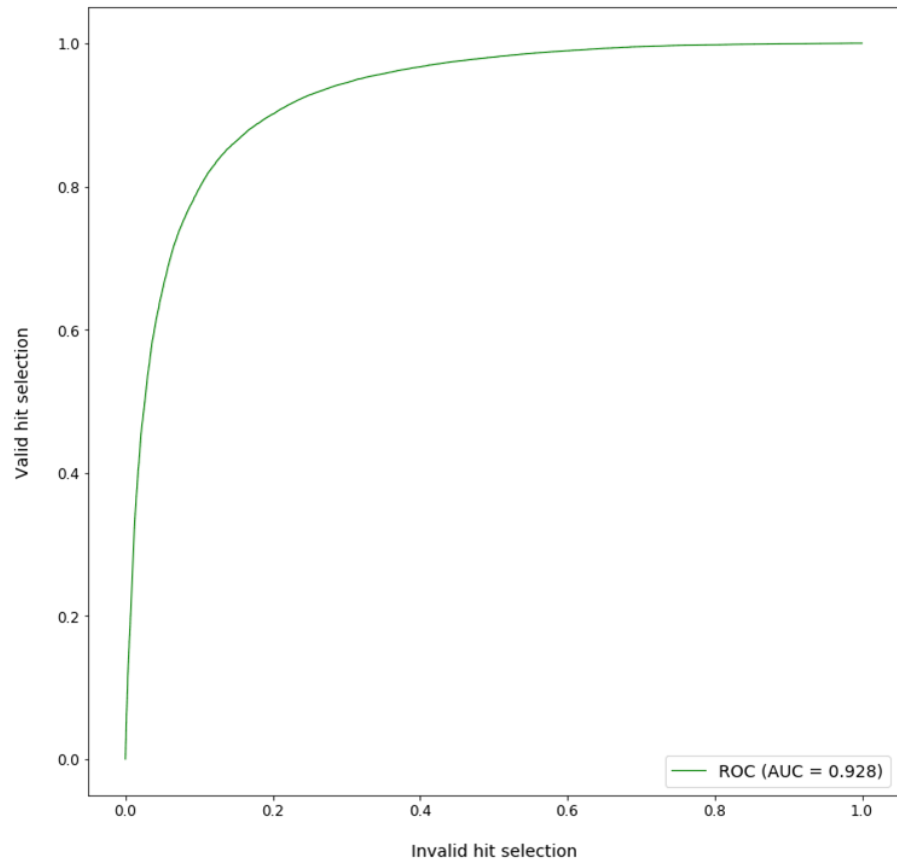
Són les posicions processades obtingudes de l'electrònica de lectura.

Només es pot obtenir una vista 2D de cada *hit*, ja que no podem determinar la posició y directament de les dades en brut.

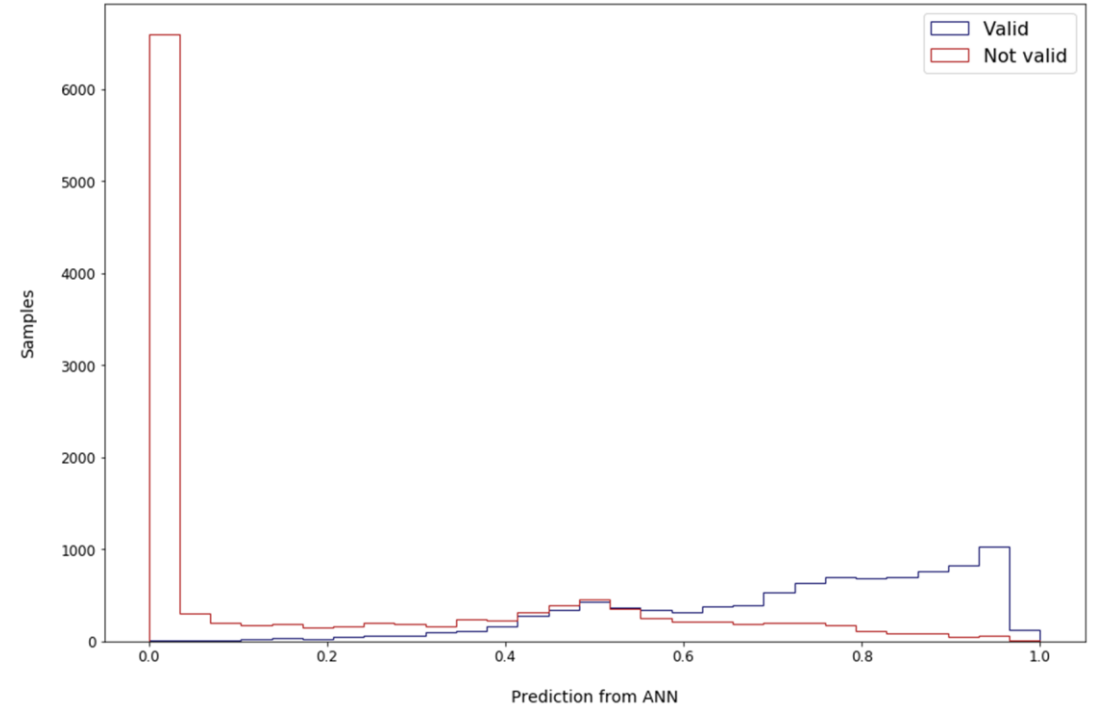
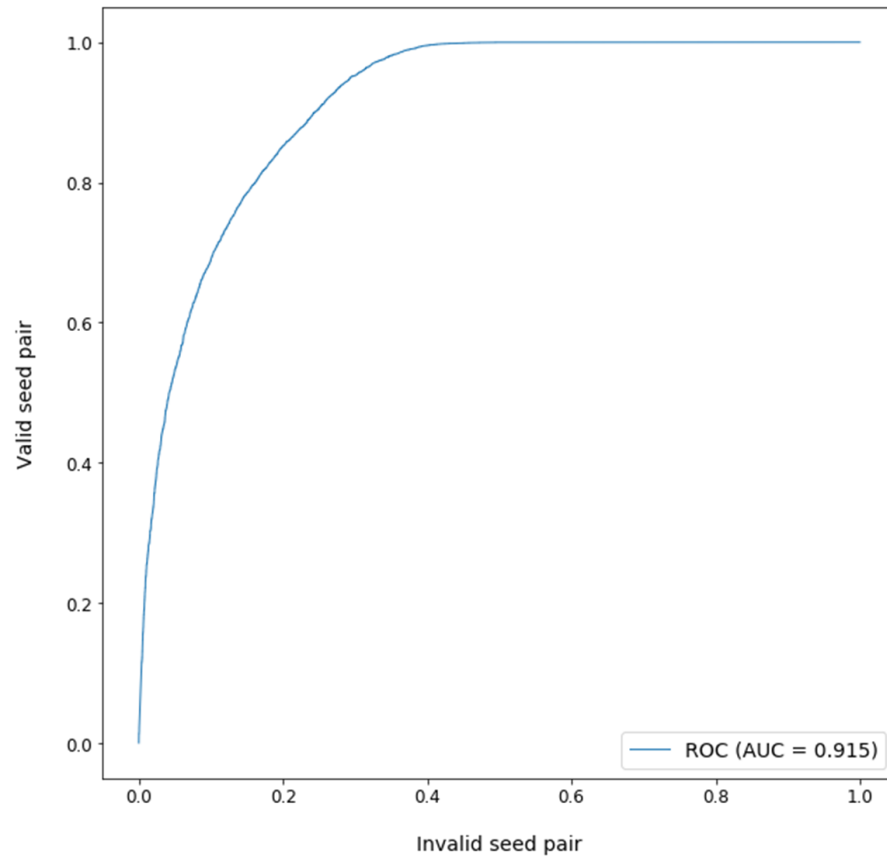




MODEL ALL SEED VALIDATION



MODEL STATION SEED VALIDATION



MODEL TWO SEEDS VALIDATION

RENDIMENT ACTUAL

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