Search for Dark Neutrinos in SAND

Alessandro Ruggeri (INFN and University of Bologna)
on behalf of the RDS@DUNE group

DUNE-Italia Collaboration Meeting

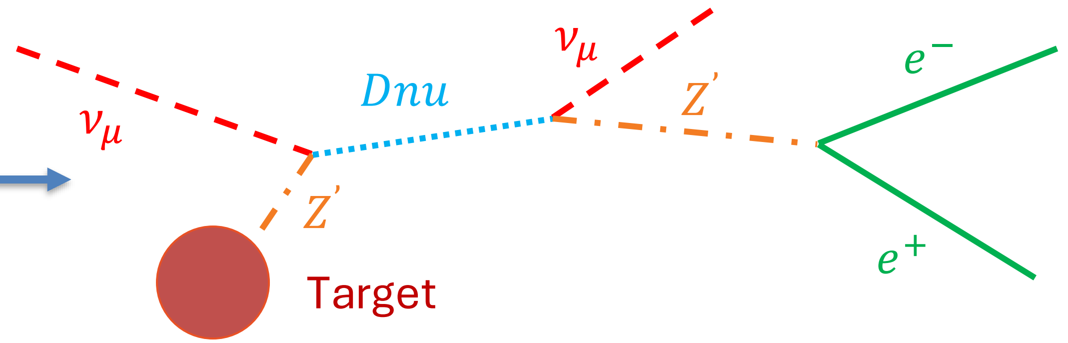
Frascati, 11/11/2025

Istituto Nazionale di Fisica Nucleare
Sezione di Bologna

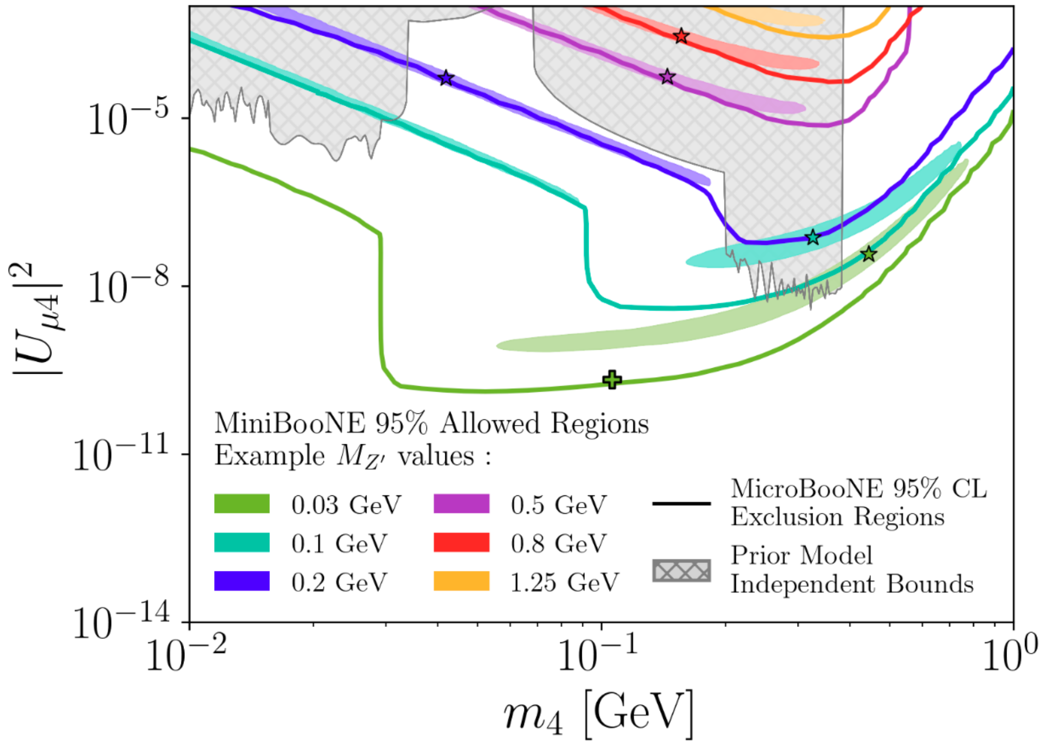
Dark Neutrino in a nutshell

Models with HNL states + Dark Photon Z' :

- **Dark neutrino** (DNu) production in ν scattering
- Coherent scattering: low-Q^2 nuclear recoil
- ***Dnu*** decays to neutrino and Z'
- Z' finally decays to di-lepton pair at a distance from the primary interaction (model dependent)



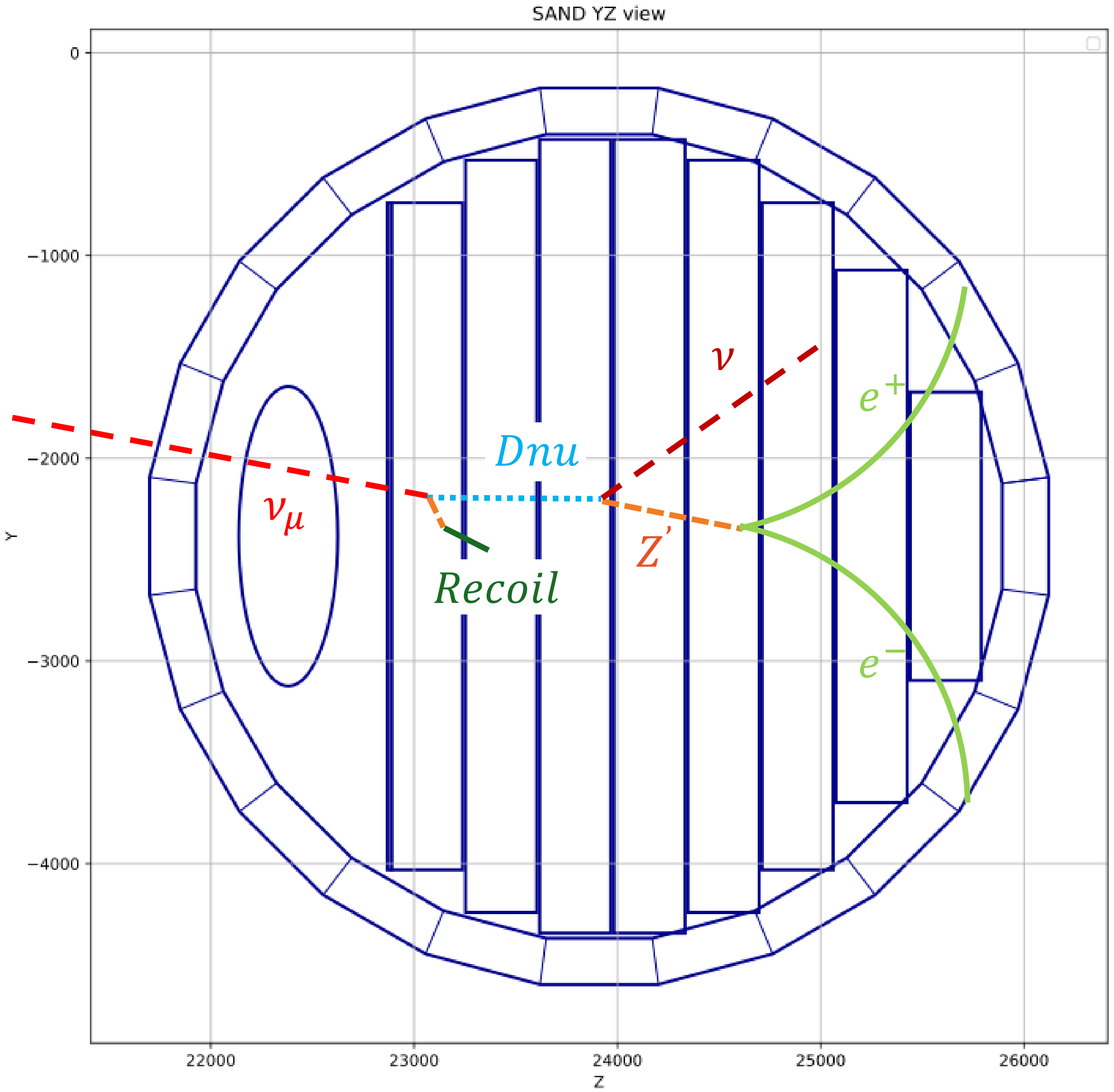
- Current limits are from MicroBooNE (arXiv:2502.10900v1)
- Z' decays to e+e- have been explored so far
- The higher flux and low-density tracker at SAND may allow an improvement



Dark neutrinos in SAND

What can be simulated in SAND with the available generators:

- Coherent scattering: no vertex trajectories
- **Dnu** decays to on-shell Z':
  - prompt decays, $\ell \leq 1\ mm$
  - di-lepton pairs with a fixed invariant mass $M_{Z'}$
- Z' decays to e+e- pairs



Signature in SAND

- e+/e- pair production in the tracker volume, in time with the beam and with the Z' invariant mass (model dependent)

Backgrounds

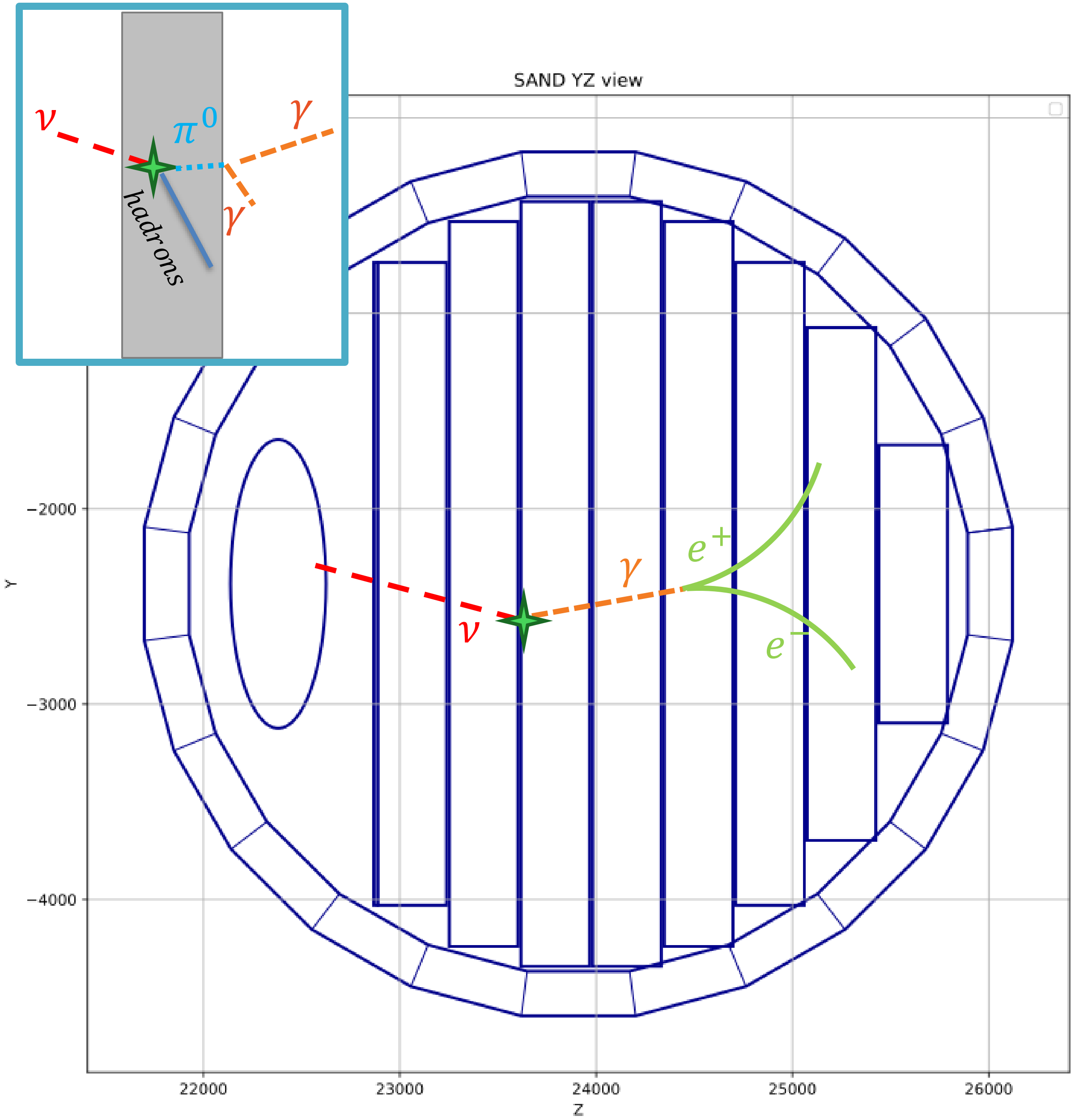
e^+e^- pairs produced in the tracker without an associated vertex activity.

Potential sources:

- primary vertices outside the tracker volume (matching is not performed at the moment)
- Short or absent vertex tracks (e.g. NC interactions with $\pi^0 \rightarrow \gamma\gamma$)

Discrimination possible via pair kinematics:

- Pair invariant mass
- Pair energy and transverse momentum
- Pair-beam angle and aperture angle

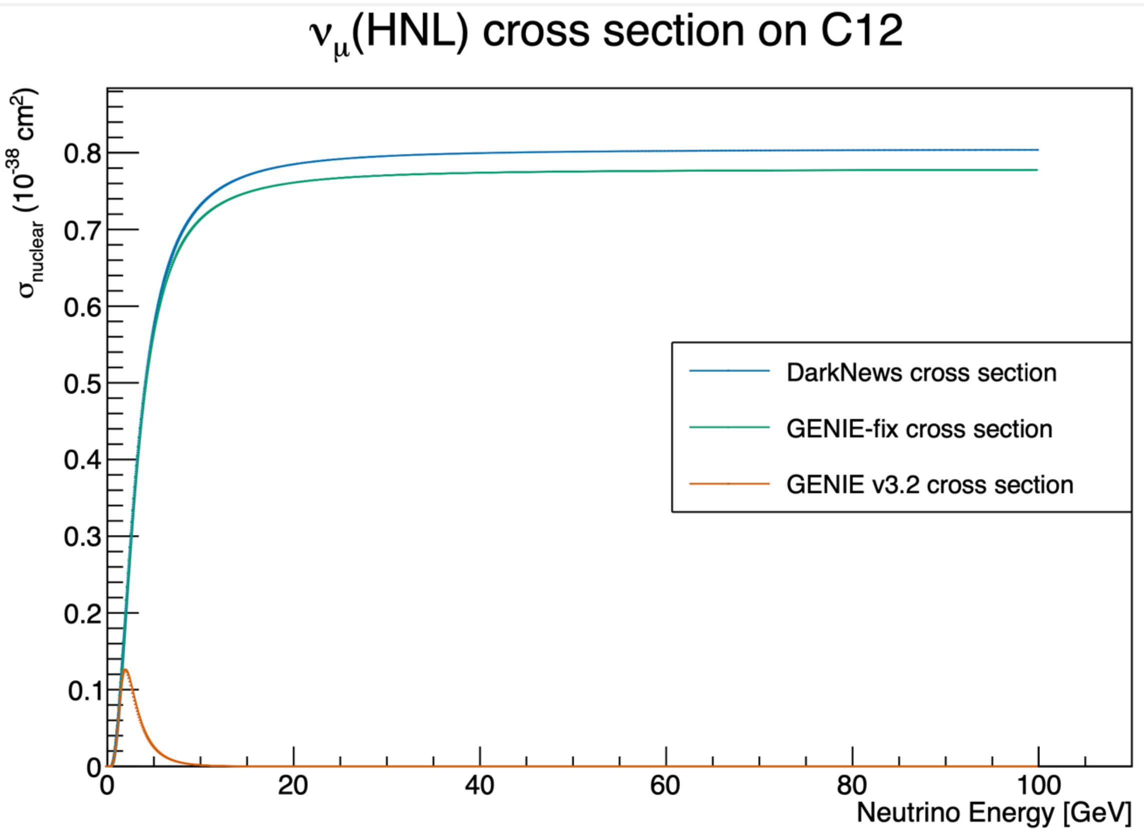


Event simulation framework

SAND reco chain adapted to enable **Dnu** simulation.

A) **Event generation with GENIE** with **Dnu** production and decay
   - Fixed a bug in the cross-section evaluation (available in GENIE 3.6.2)

B) **Particle propagation with edep-sim:** fixed an issue with displaced vertex



Limitations in GENIE (needs long-term development):

- scattering on free nucleons (H) not implemented
- only real (on shell) Z' are generated
- decay channels other e^+e^- are missing ($Z' \to \mu^+\mu^-$)

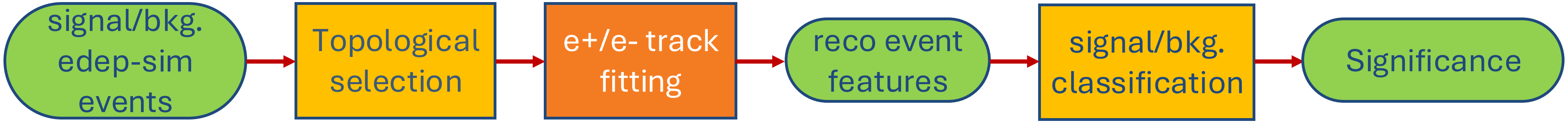
Data Production

Simulated processes:

- **Signal**: Dnu models spanning **Dnu** masses M_N between 35 MeV and 1 GeV.
  Fixed $M_{Z'} = 30\ MeV$, $|U_{\mu 4}|^2 = 2 \times 10^{-10}$
- **Background**: simulated with the standard tune (AR23_20i_00_000)

- Productions made with the Drift-chamber geometry (**SAND_opt3_DRIFT1**)
- ~500k events in GRAIN+TRACKER for each sample
- Beam: ν_μ, $\bar{\nu}_\mu$ and ν_e, $\bar{\nu}_e$ in FHC mode (**gsimple** format)
- Fixed versions of genie and edep-sim discussed before
  (signal H1 interactions are neglected)

Analysis pipeline



1. Topological cuts based on edep-sim info:

- Selecting candidate e^+e^- pairs produced in the tracker volume
- For bkg. events: no visible vertex activity (**Dnu** scattering is coherent, 50 keV recoil)

2. Fitting of the candidate pair trajectories:

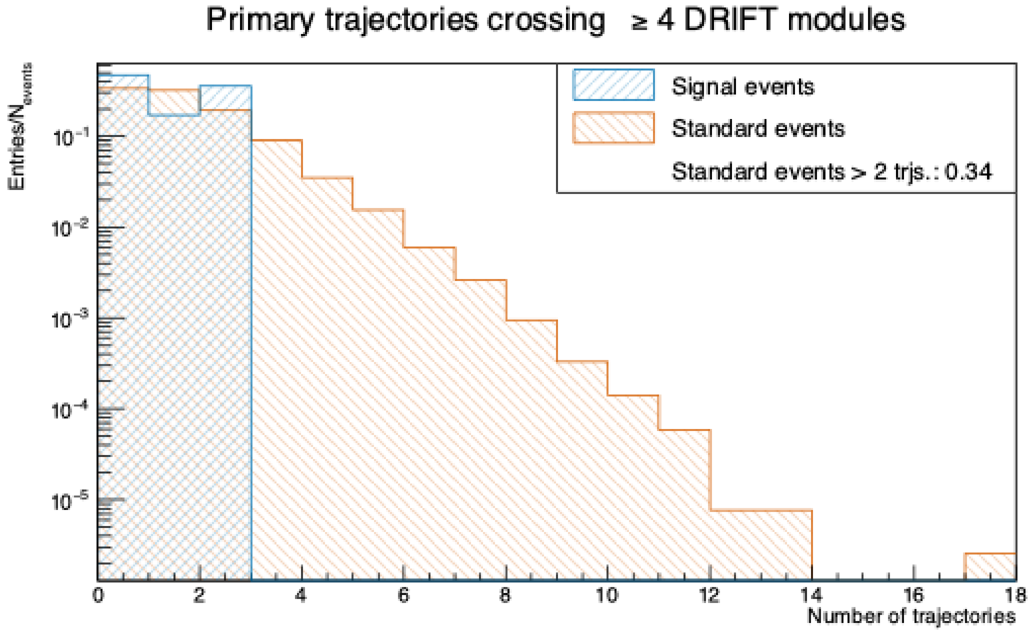
- Fitting of the e^+e^- trajectory hits to reconstruct momenta and extract candidate event features
- Selection on fit-quality for the pair trajectories

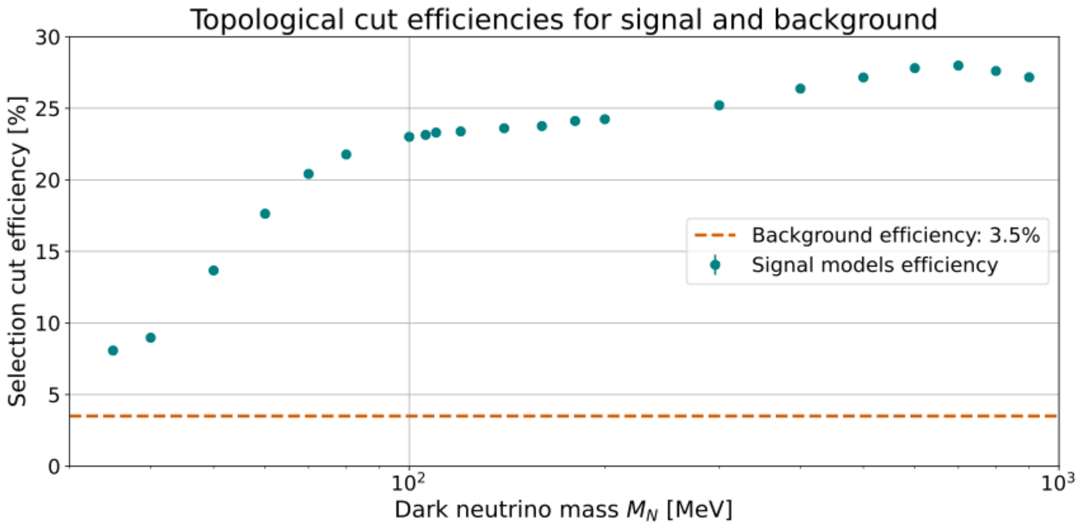
3. Multivariate classification:

- Signal/background separation based on the the reconstructed features
- Compute the significance Z from the resulting number of signal and bkg. events (S and B)

1. Topological cut

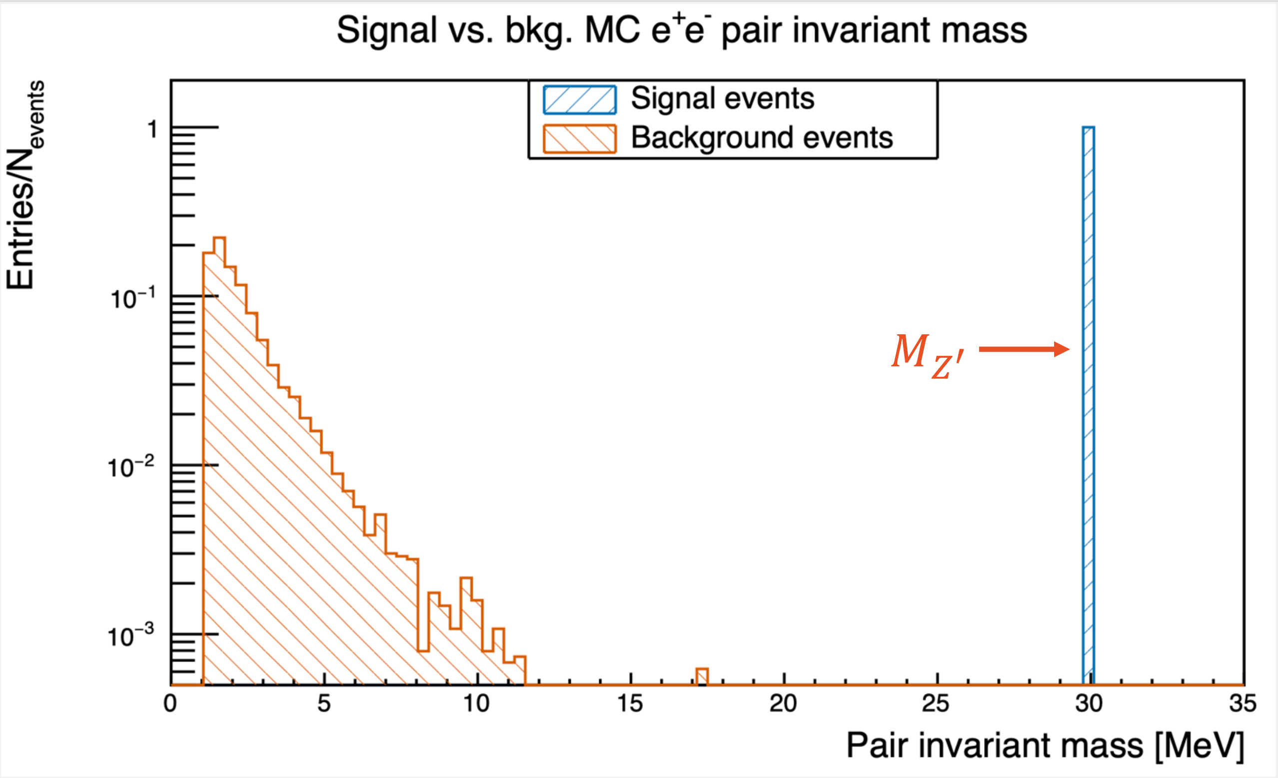
- Selection based on a standard quality cut for the tracker: ≥ 4 crossed tracker modules
- Candidate events have ≥ 1 pairs produced in the tracker and passing the cut.
- Cut on vertex activity based on primary trajectories above the threshold
- Two bkg. topologies:
  - No vertex trjs. above threshold, but ≥ 1 candidate pairs → selecting the most upstream
  - Single primary pair above threshold
- Signal events: requiring signal pair to pass the candidate cut





MC pair kinematics features

- Characterised the candidate pairs by a set of kinematic features:
  - Invariant mass
  - Pair energy and transverse momentum
  - Pair opening angle and pair-beam angle
- Clear separation in the invariant mass (same $M_{Z'}$ in all signal models)
- The kinematic distributions vary across the model range.



- Implemented a track fitting algorithm to provide reconstructed features.

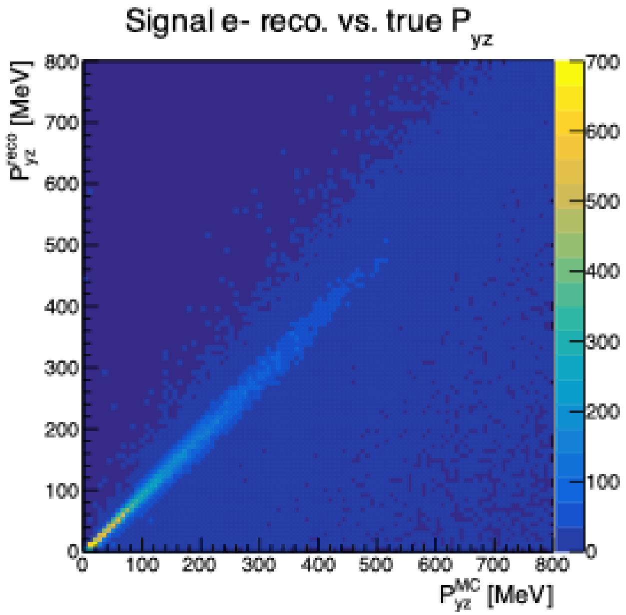
2. e+/e- track reconstruction

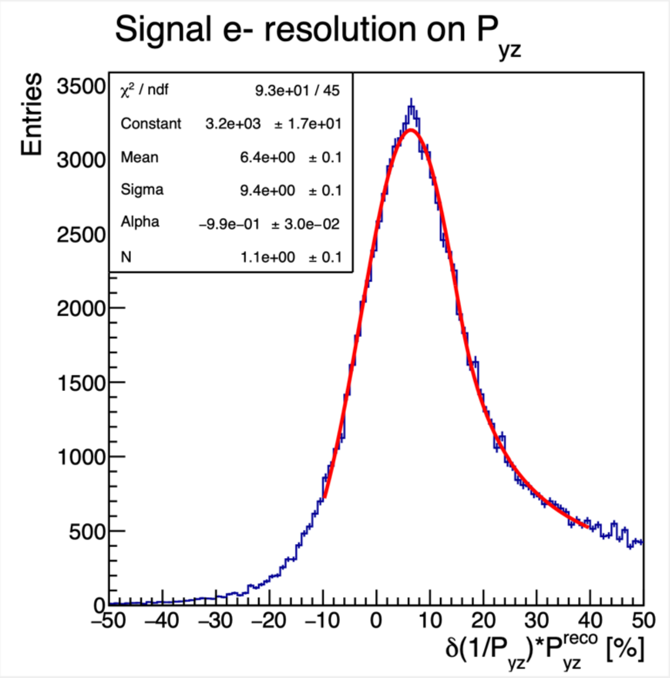
Preliminary track reconstruction using a helix-fit:

- Fitting the edep-sim hits of the e^+e^- trajectories
  → 0.2 mm smearing is applied
- Radiative losses in e+/e- tracks:
  - Fit limited to the first 25% of hits
  - Path-dependent weighting of χ^2 in the bending plane to account for reduced R
- Reconstructed momentum from fit parameters

- The helicoidal fit underestimates e+/e- momenta:

~9.5% bending plane resolution, 6% bias



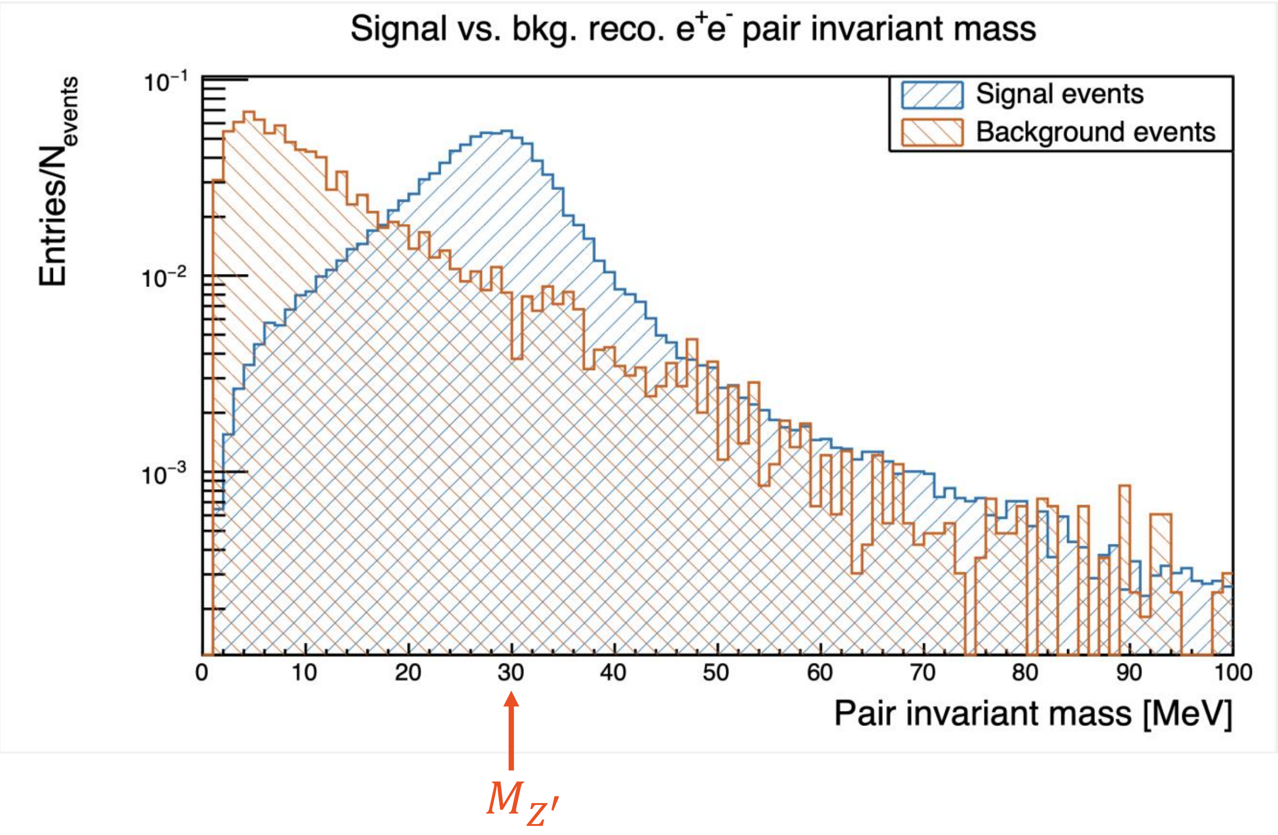


Reconstructed pair features

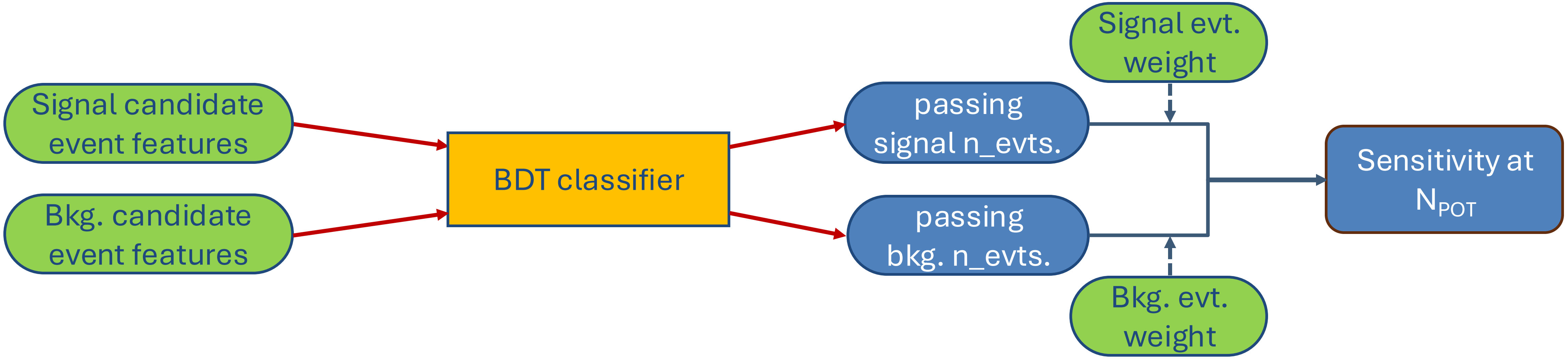
- Pair kinematics features extracted from the reconstructed tracks.

- Reconstructed invariant masses are not as well separated.
- Best resolutions are on the pair energy and pair-beam angle (7.9% and 8.2%)

- Decided for a multivariate classifier trained on the kinematics to perform the final selection step.



3. Signal/bkg. discrimination with BDTs



- Boosted Decision Tree classification on reco-level e+/e- pair features (energy, aperture angle, pair-beam angle, invariant mass, P-transverse, etc…)
- Set up a data preparation/training/testing pipeline in TMVA (native TMVA models or XGBoost)
- Counting experiment approach to extract the sensitivity: obtain Z from the passing signal and background events from the classifier (S and B) + weights

Discovery potential computation

1. Single BDT is trained on all signal model samples against the background events
2. Working point score *t** found by optimising the significance on the BDT test set
3. Applied the full chain on samples of ~500k event for all signal models + bkg and computed the significance at the t^* score cut:

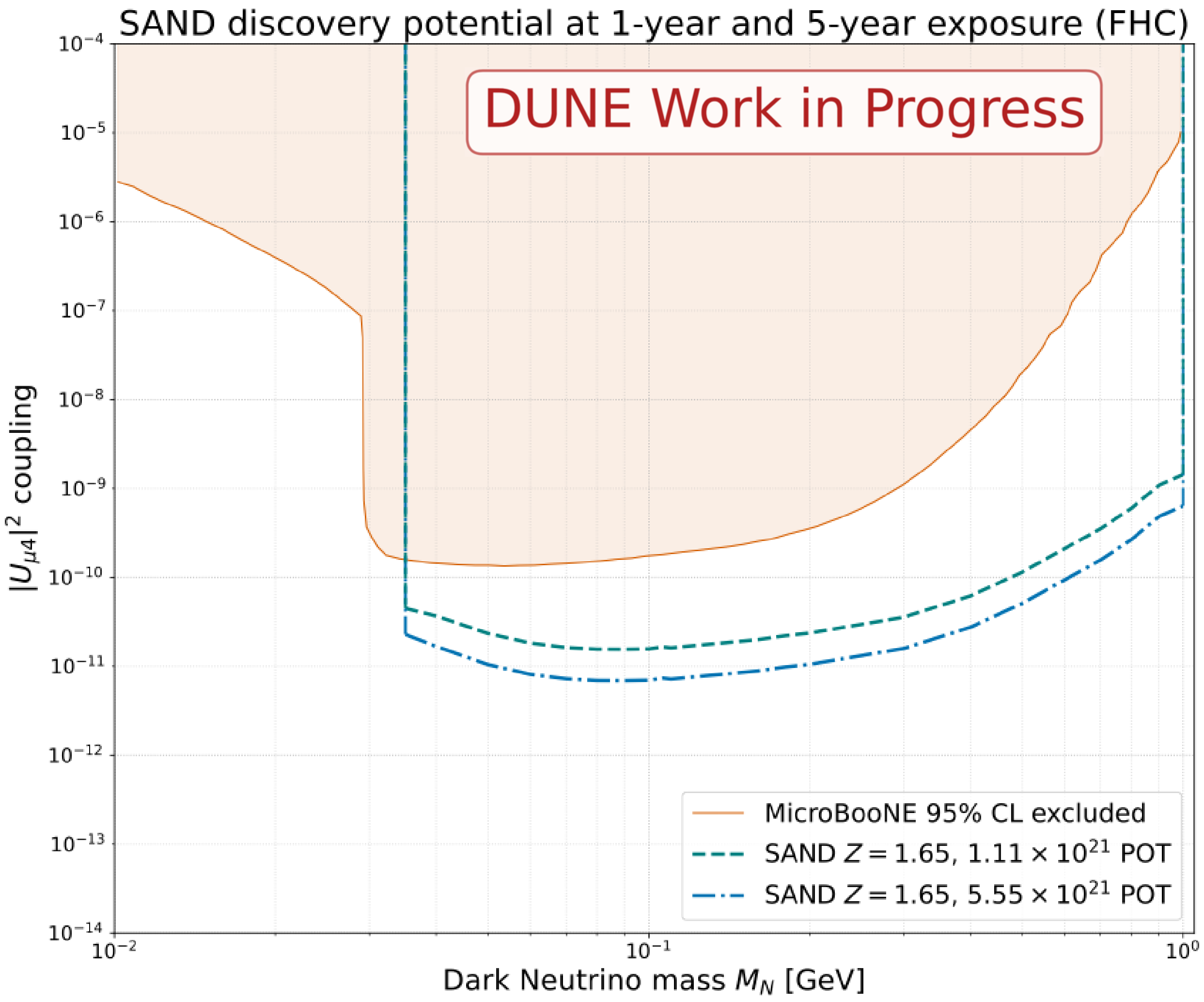
$$Z_{tgt}(t^*) \equiv \sqrt{2\left[\left(S_{tgt}(t^*) + B_{tgt}(t^*)\right)\ln\left(1 + \frac{S_{tgt}(t^*)}{B_{tgt}(t^*)}\right) - S_{tgt}(t^*)\right]},$$

- General Z formula to account for potentially small B.
- Scaling S and B to the exposure and significance chosen for the curve (+ additional corrections for acceptance, cross-section, etc…)

Discovery potential curve

- Reference exposures at 1e21 and 5e21 POT
- Discovery potential at Z=1.65
  (95% C.L. one-sided)
- Reference MicroBooNE excluded region for the same $M_{Z'}$ (~6.8e20 POTs)

- Off-shell decays ($M_{DNu} < M_{Z'}$) are not included in GENIE but may be particularly interesting in SAND:
  - signal stat. depends on external mass
  - best case for low density detectors



Next steps

First performance assessment used a preliminary analysis pipeline.

Further developments are needed:

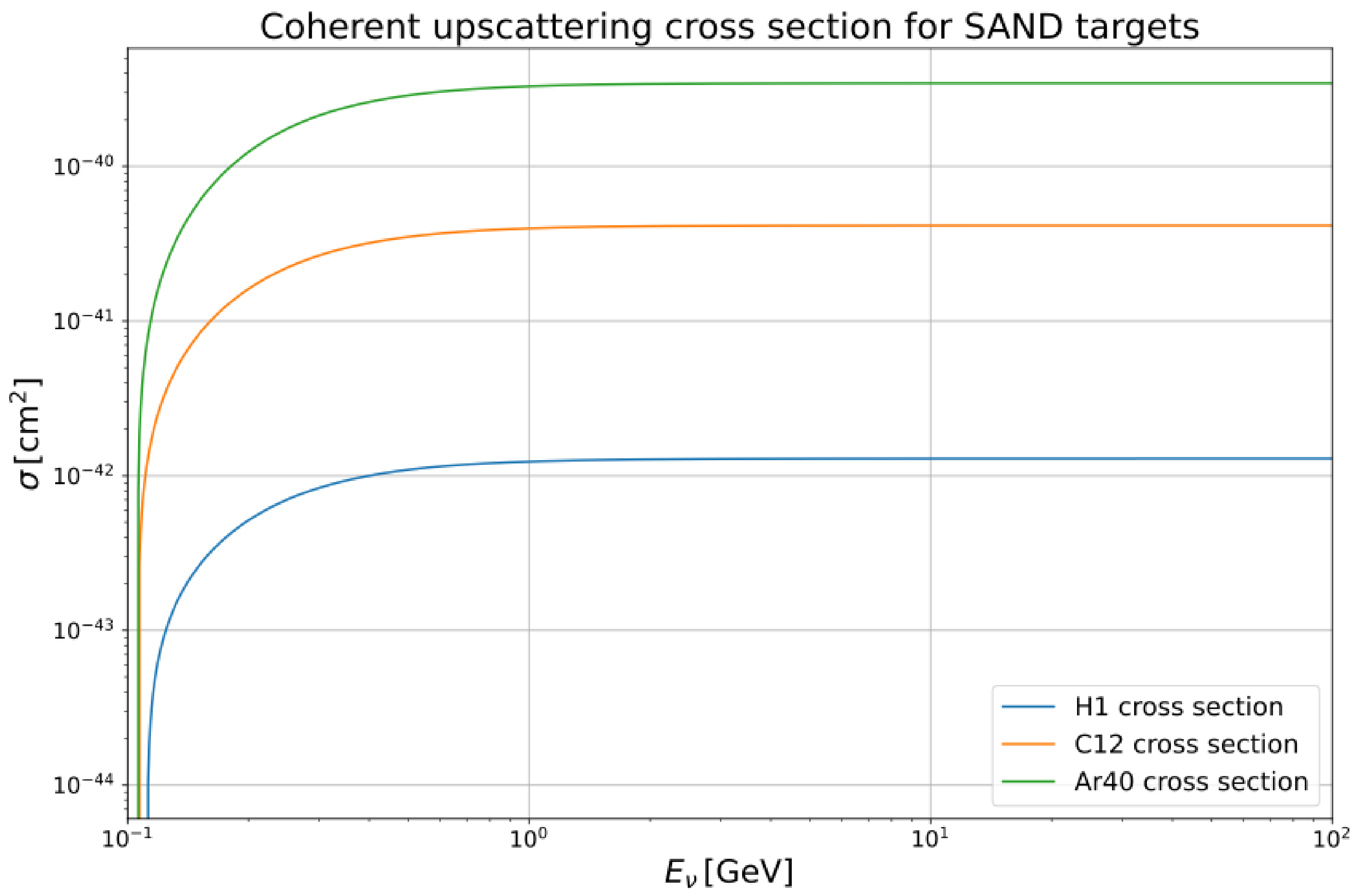
- On the SAND software side:
  - Integrate the sandreco track digitisation and reconstruction (Kalman filter or new algorithm)
  - Perform the topological selection through a vertexing algorithm
  - Investigate the use of sandreco ECAL reconstruction in the analysis (selection on cluster multiplicity, energy reconstruction, e+e- tagging via E/p)

- On the GENIE side as well:
  - Introduce a cross section for H
  - Implement off-shell Z' decays
  - Implement the decay channel to μ+μ-

Thank you for the attention!

Backup and additional items

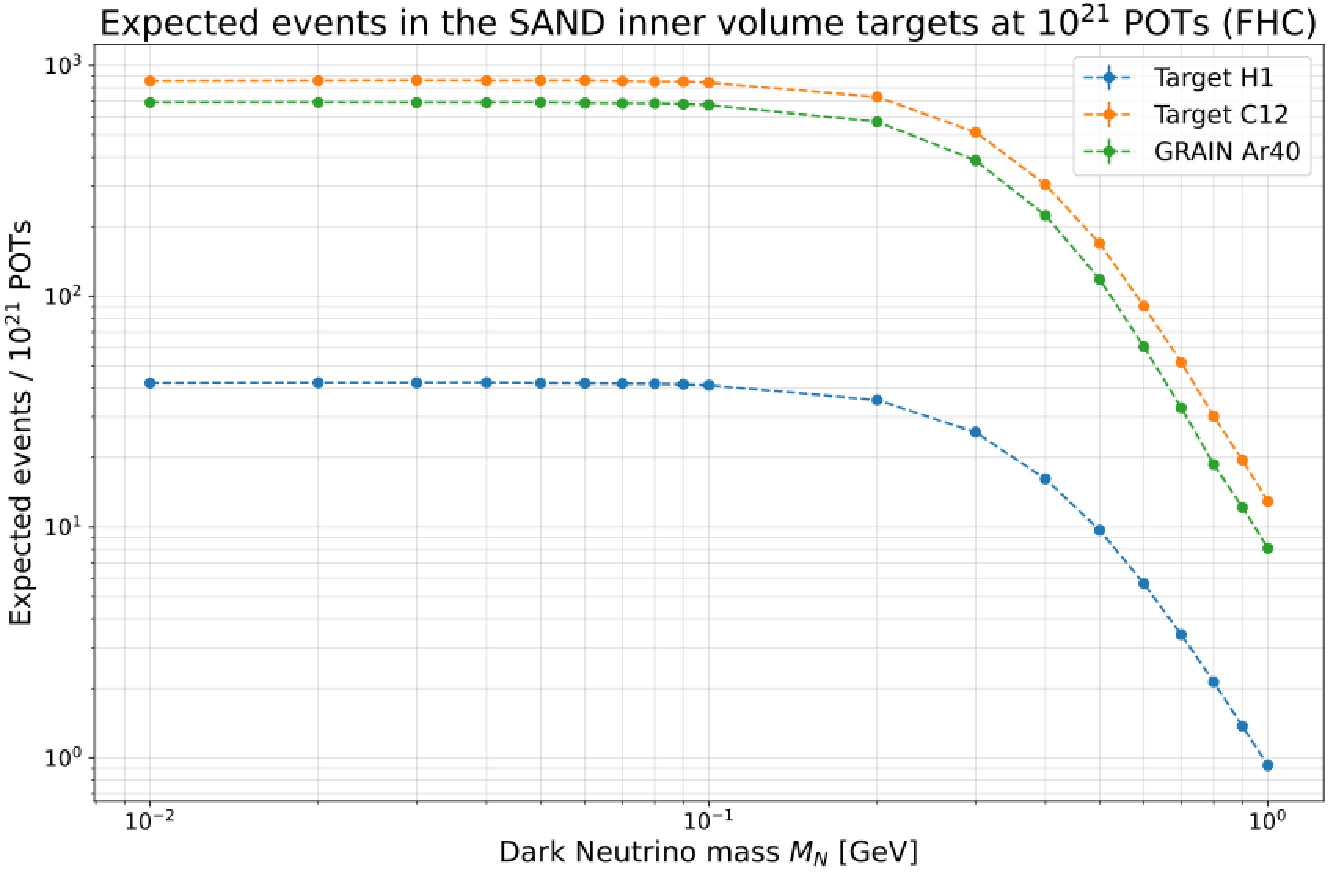
Dark neutrino cross section on H1



$$M_{DNu} = 106\ MeV,\ M_{Z'} = 30\ MeV,\ |U_{\mu 4}|^2 = 2 \times 10^{-12}$$

Missing statistics from H1

- Used DarkNews to estimate the event rate on several SAND targets:
  - Tracker target H1 (~488 kg)
  - Tracker target C12 (3.7 tons)
  - GRAIN Ar40 (1.2 tons)
- 5k weighted event generations at different M_{Dnu} for 10^{21} POT (FHC)



- Event rate lower by a factor of ~10
- Neglected for the moment

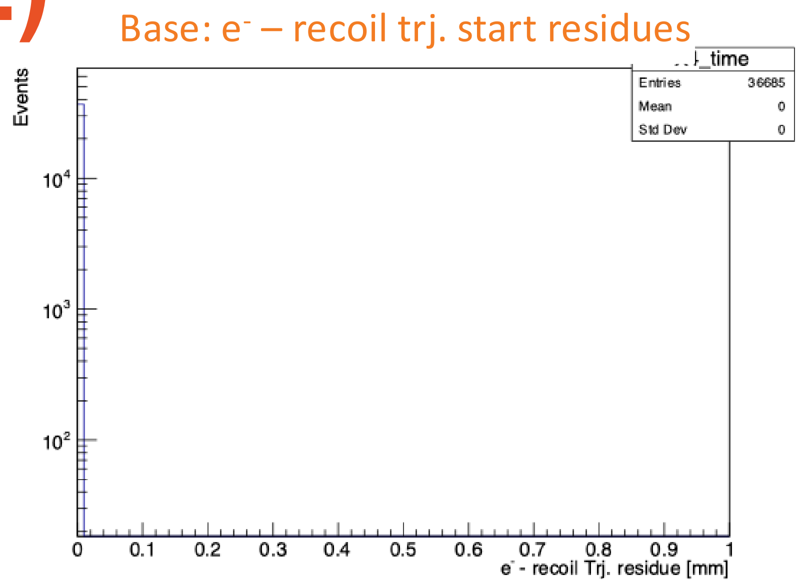
Event Simulation Framework (2)

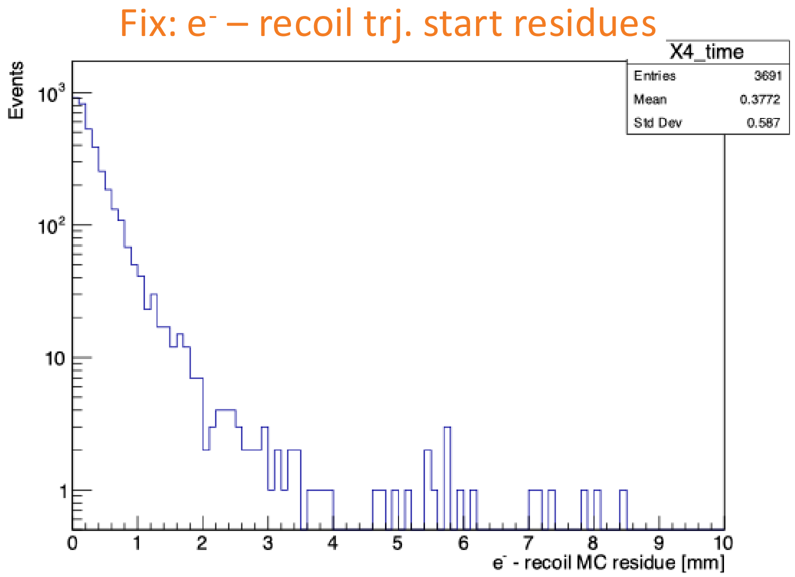
B) **Particle Propagation:**

- GENIE output files contain all particles: both from primary interaction and Z' decays, originating in different positions → edepsim primaries
- All primary tracks in edepsim are propagated them from the same point (the primary vertex)

- Displaced decays are a distinctive feature of Dark Neutrino models (in principle)

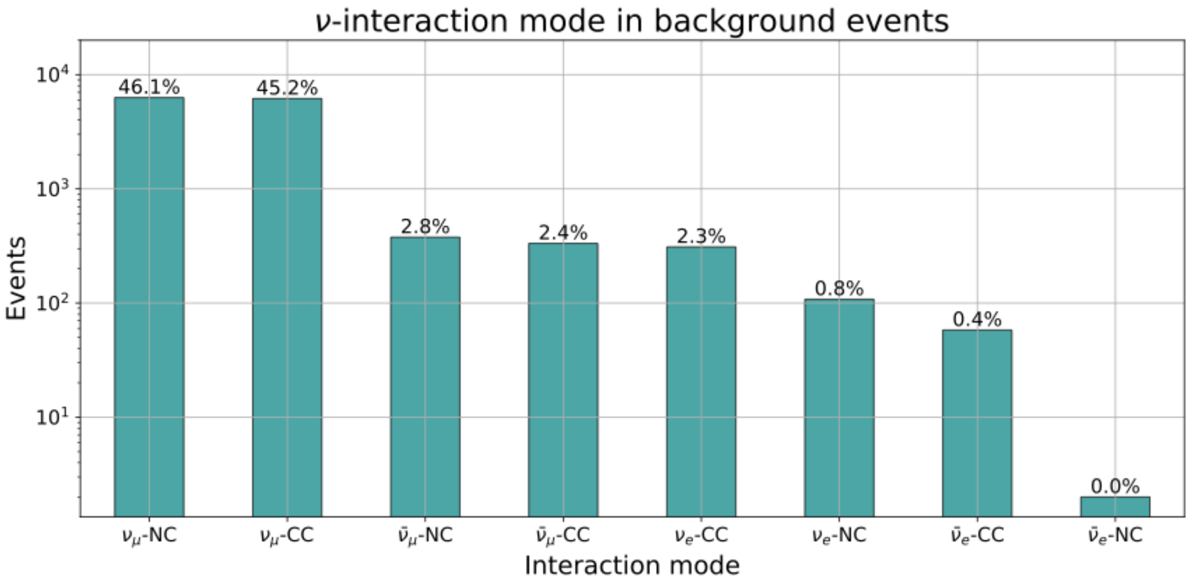
→ We fixed edep-sim to group input particles by position and time and fill a vector of TG4PrimaryVertex with different positions
(available at github.com/AlessandroRuggeri/edep-sim)



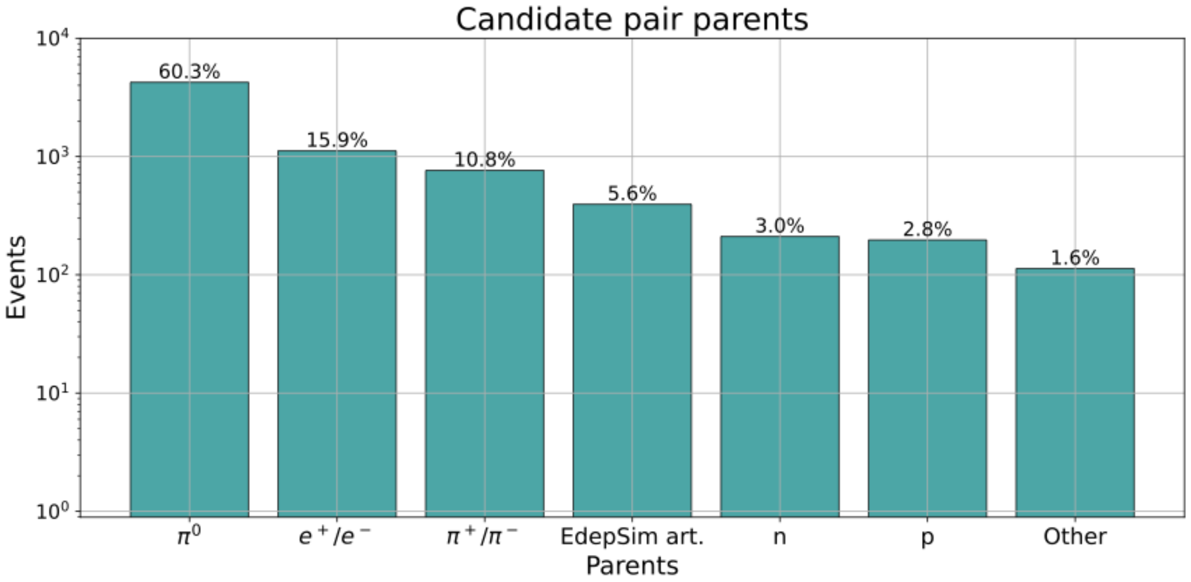


Background sources

Largest fraction of bkg. interactions are NC



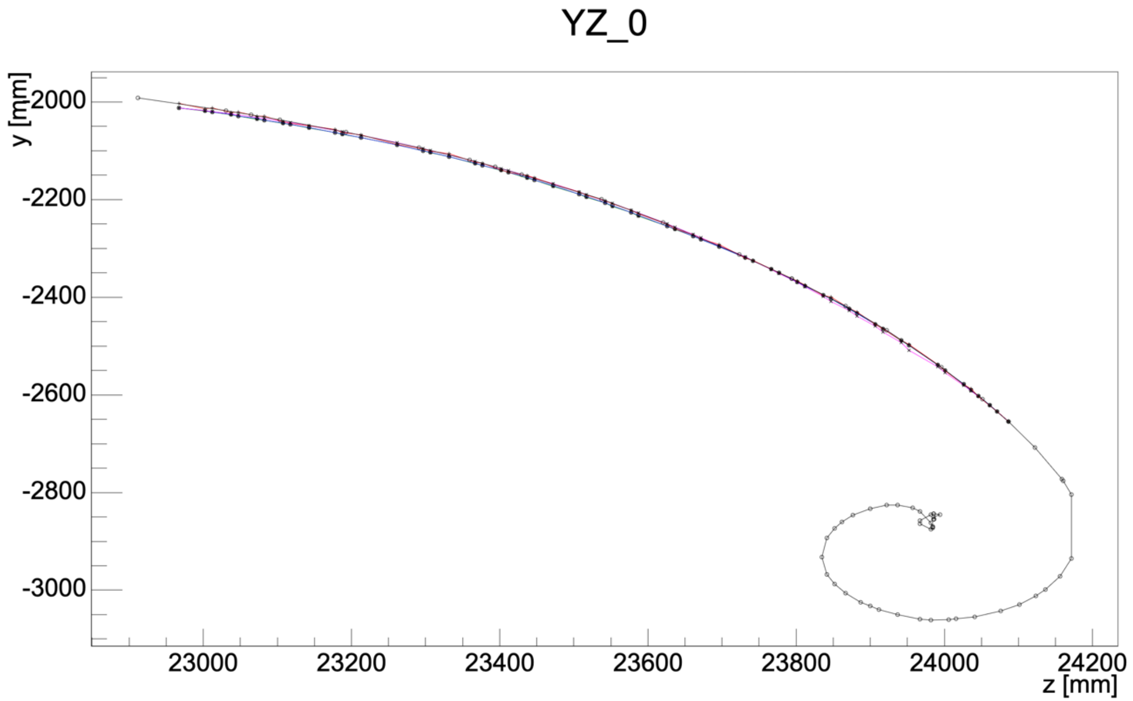
Pi0 are the most common bkg. pair parents (via $\gamma\gamma$)

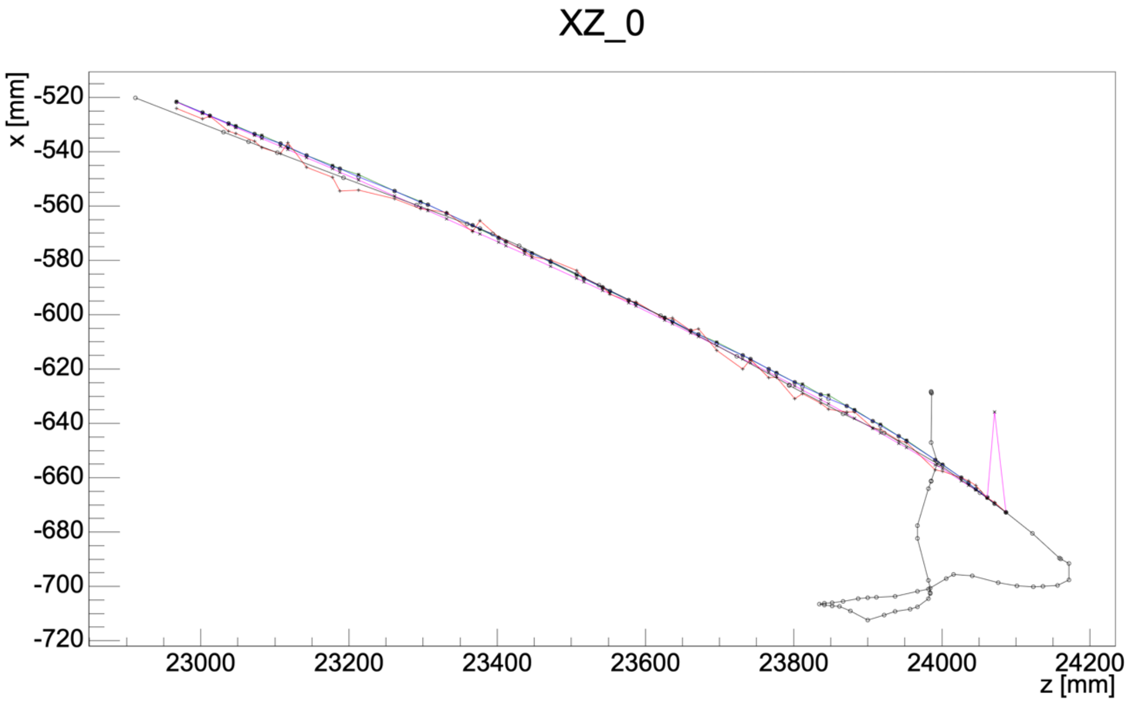


Pair parents taken from edep-sim: some parent γ trajectories are split

Signal e- trajectory examples

- Signal pair e- trajectories from a test with the Kalman filter ($M_N = 106$ MeV benchmark)
- The trajectories loop back within the tracker volume → underestimation of R



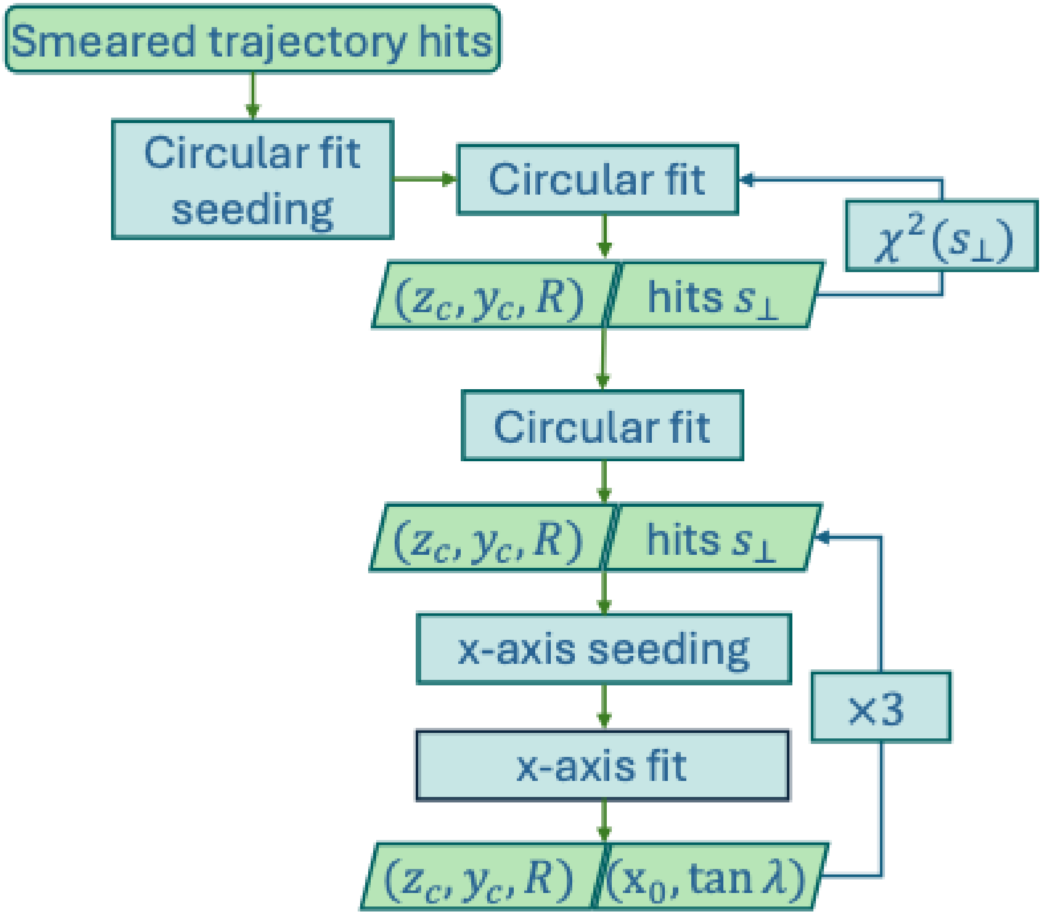


Track fitting algorithm

- Hit coordinates smearing (~0.2 mm)
- Circular + straight line fit:

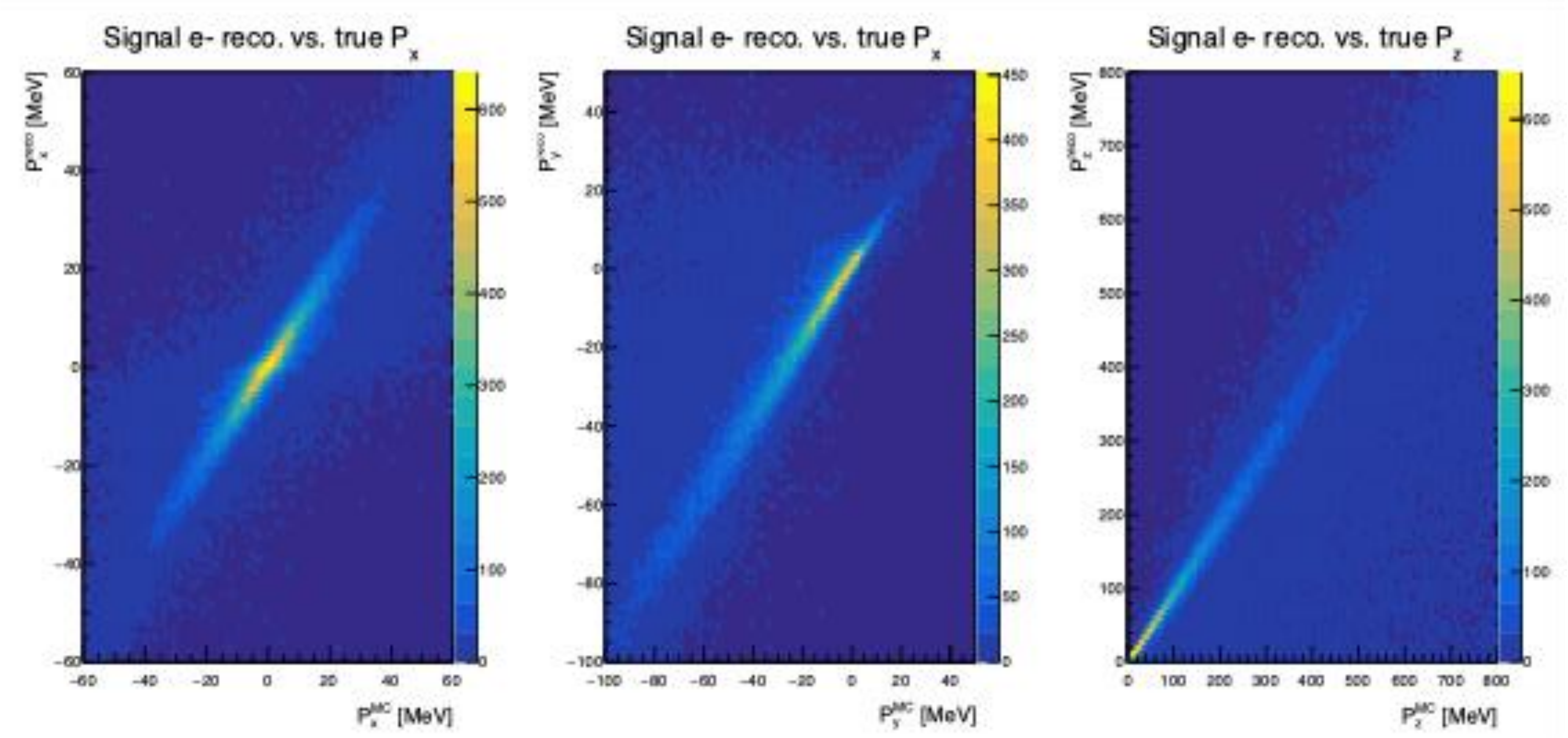
$$\begin{cases}(z-z_c)^2+(y-y_c)^2=R^2\\ x=x_0+s_t\cdot\tan\lambda\end{cases}$$

- Accounting for reduction in R in the circular fit by:
  1. Limited fit to initial 1/4 of the track hits
  2. s_t-dependent χ^2 : $\sigma=\sigma_{res}+\alpha s_t+\beta s_t^2$
  3. χ^2 weights depending on R-residue:
     $$w=c/\Delta R$$



Reco. vs. true momenta

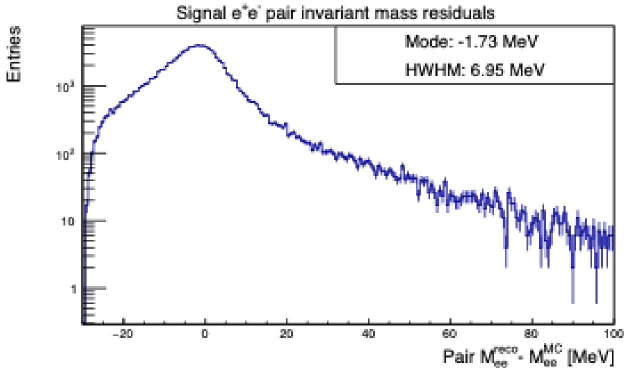
- Momentum extrapolated from fit parameters at the MC start of the trajectory
- In signal events: $p_z \gg p_y > p_x$



Momenta tend to be underestimated → energy loss in e+/e-

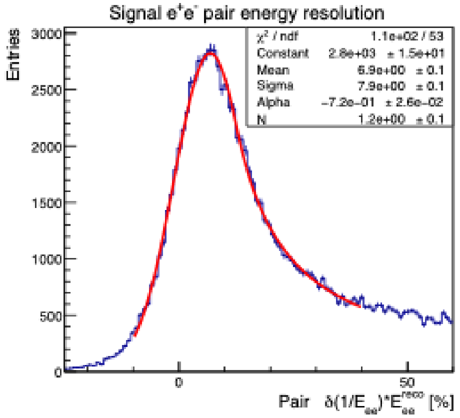
Reco. features resolutions

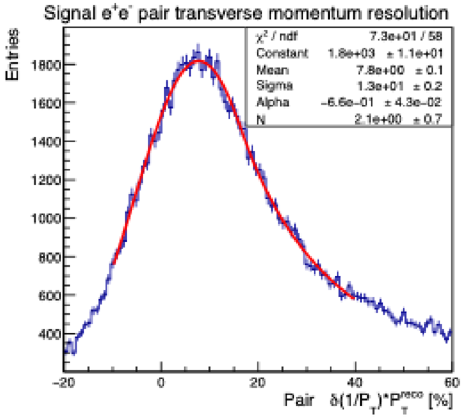
- Resolutions on the reconstructed pair kinematic variables (M_N=106 MeV benchmark)



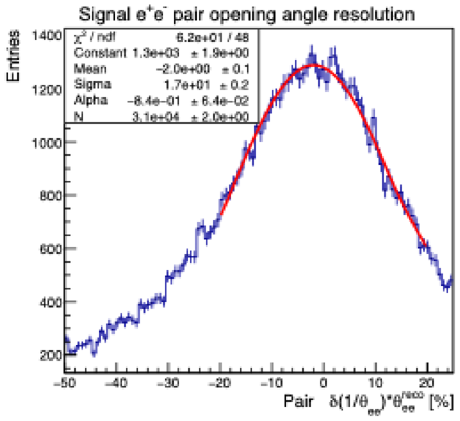
- Reconstructed invariant masses are not as well separated.

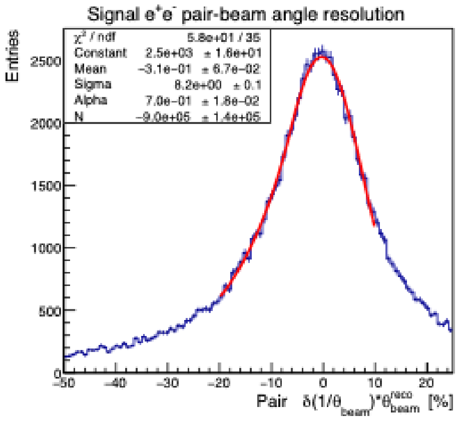
- Best resolutions are on the pair energy and pair-beam angle (7.9% and 8.2%)





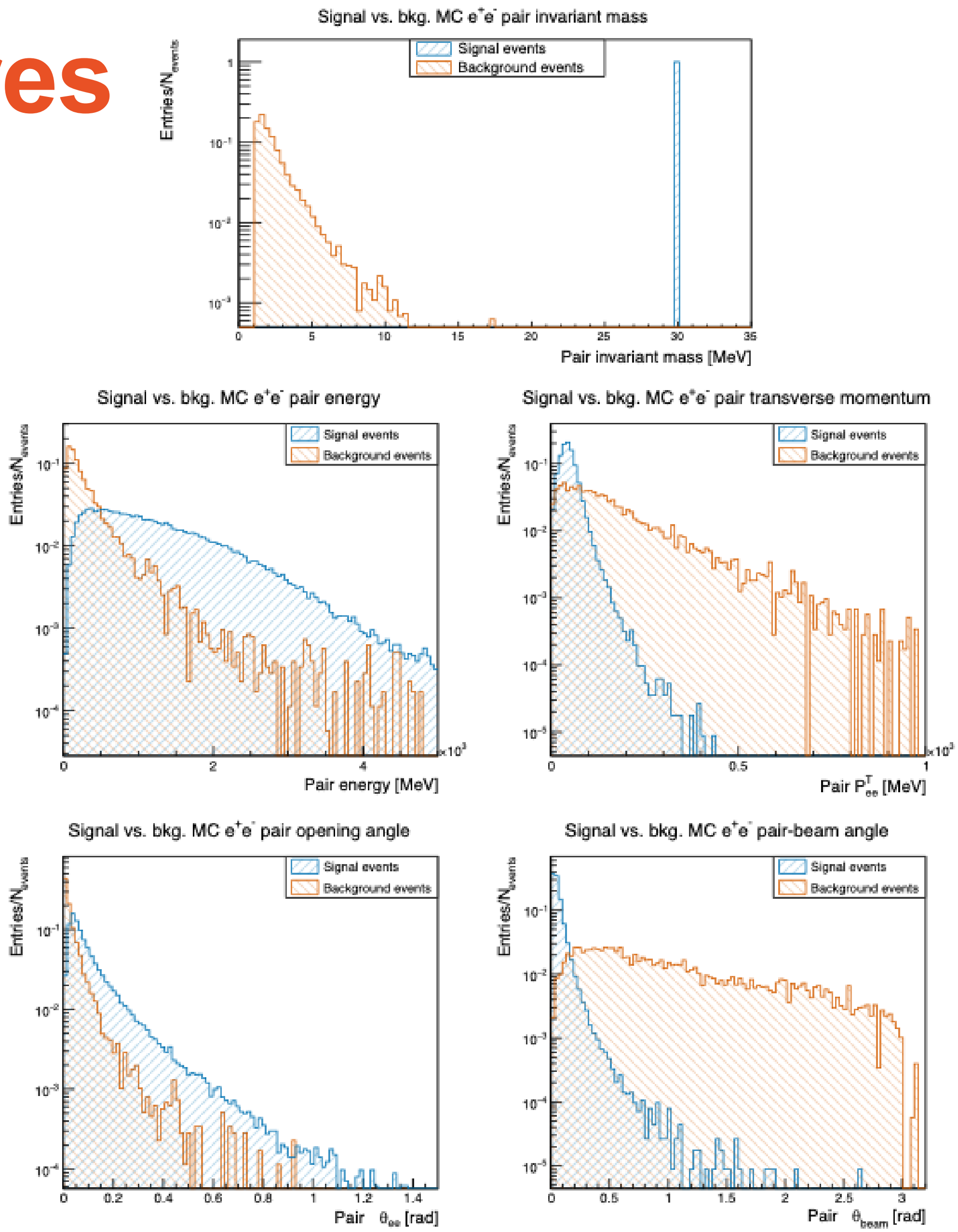
- Decided to go with a multivariate classifier for the final selection step





MC pair kinematics features

- Characterised the candidate pairs by a set of kinematic features:
  - Invariant mass
  - Pair energy and transverse momentum
  - Pair opening angle and pair-beam angle
- Clear separation in the invariant mass (same $M_{Z'}$ in all signal models)
- The kinematic distributions vary across the model range.
- Implemented a track fitting algorithm to provide reconstructed features.



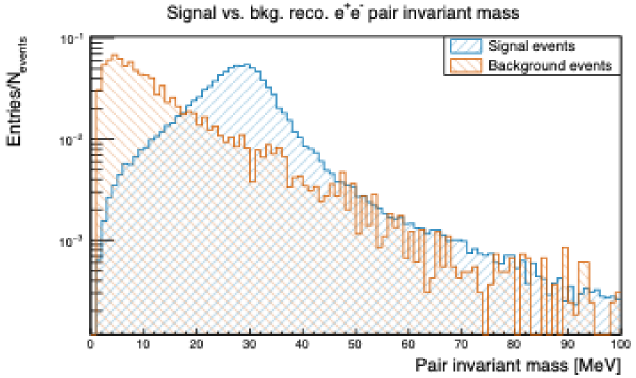
Reconstructed pair features

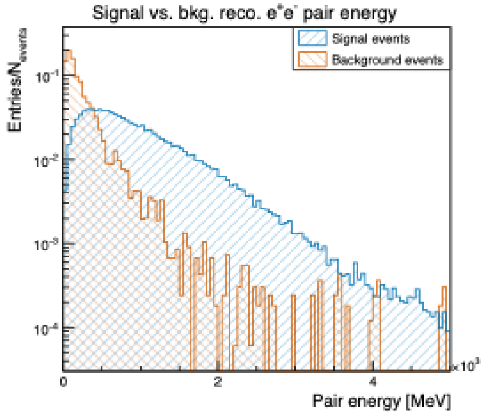
- Pair kinematics features extracted from the reconstructed tracks.

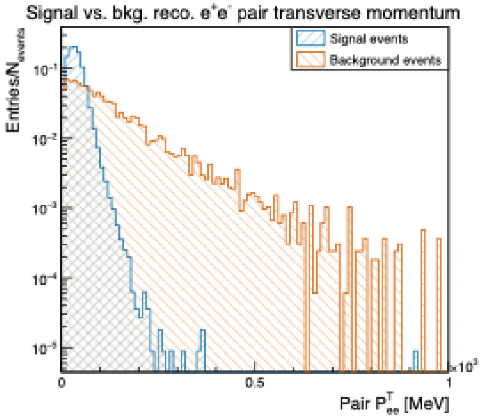
- Reconstructed invariant masses are not as well separated.

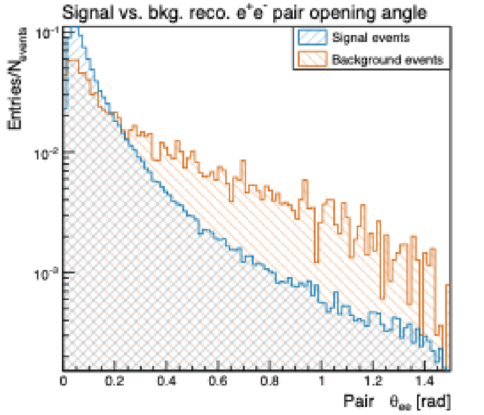
- Best resolutions are on the pair energy and pair-beam angle (7.9% and 8.2%)

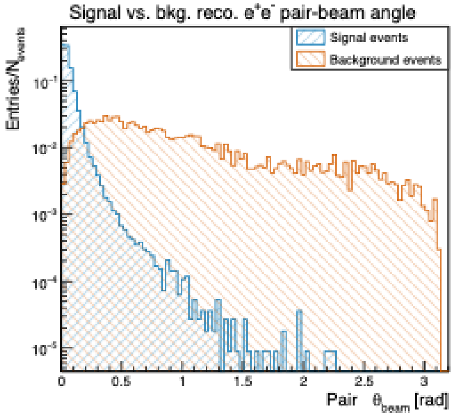
- Decided for a multivariate classifier trained on the kinematics to perform the final selection step.











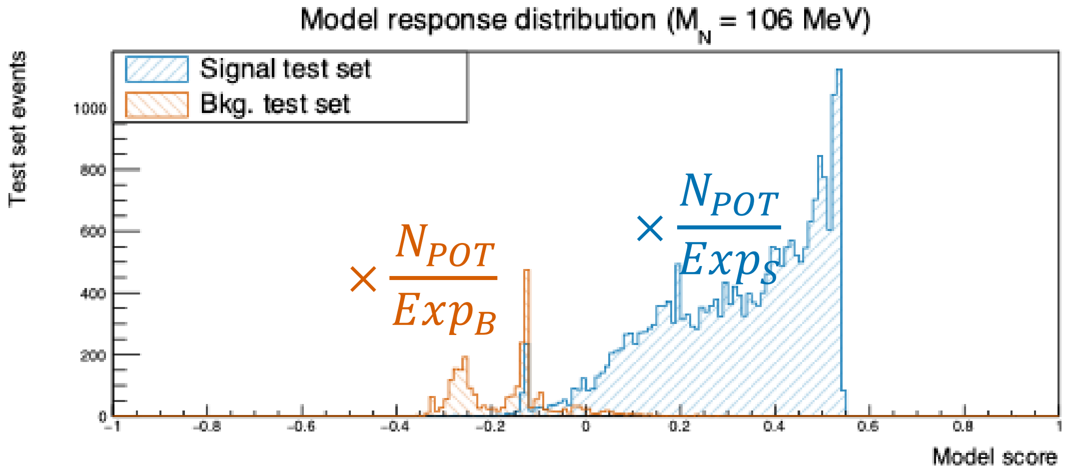
BDT and cut selection

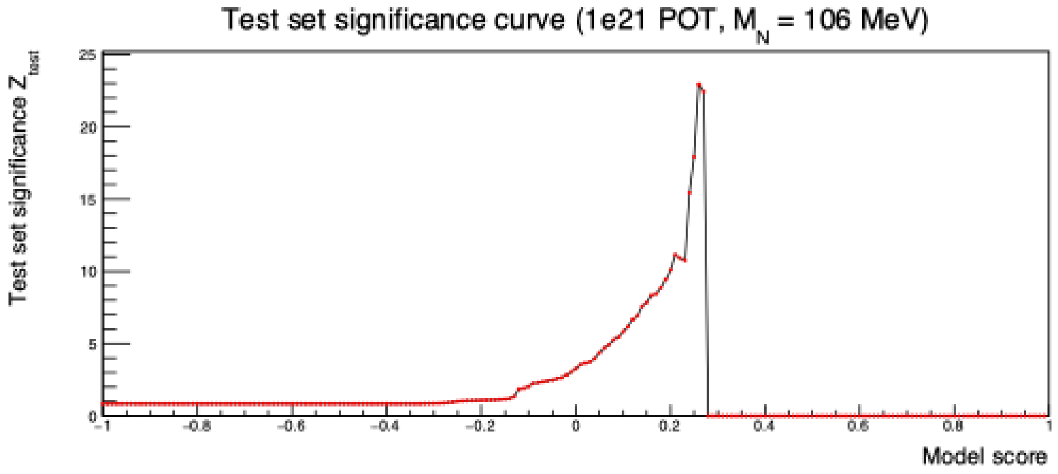
- Single BDT is trained on all signal model samples against the background events
- For each signal model: computed the test-set significance as a function of the cut *t*:

$$Z_{test}(t) \equiv \sqrt{2\left[\left(S_{test}(t)+B_{test}(t)\right)\ln\left(1+\frac{S_{test}(t)}{B_{test}(t)}\right)-S_{test}(t)\right]},$$

- S_{test} and B_{test} scaled from the production exposures
- Global BDT score t^* found as the one maximising the worst test-set significance out of all signal models

- Tested several BDTs: chosen the one that maximises the significance on a benchmark model



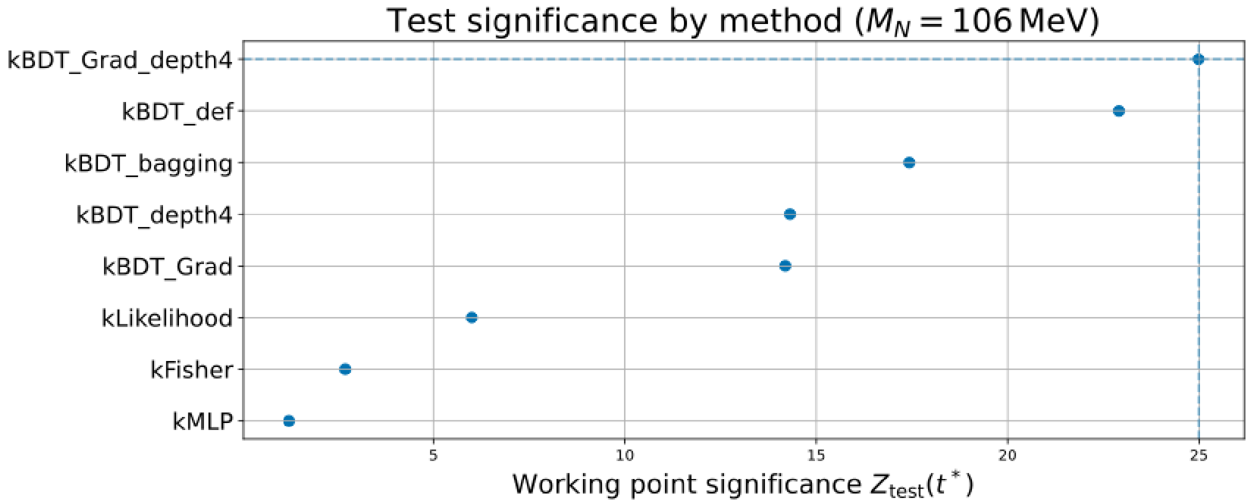


Choice of classifier

- Tested several classifier configurations, focusing on BDTs
- Reference classifiers of different types

Tested BDT models	MaxDepth	nCuts	BaggedBoost	BoostType
kBDT_def	3	20	no	AdaBoost
kBDT_Grad	3	50	no	GradBoost
kBDT_depth4	4	50	no	AdaBoost
kBDT_Grad_depth4	4	50	no	GradBoost
kBDT_bagging	3	50	yes	AdaBoost
Reference classifiers				
kLikelihood	Projective likelihood classifier			
kFisher	Fisher discriminant classifier			
kMLP	Multi-Layer Perceptron model			

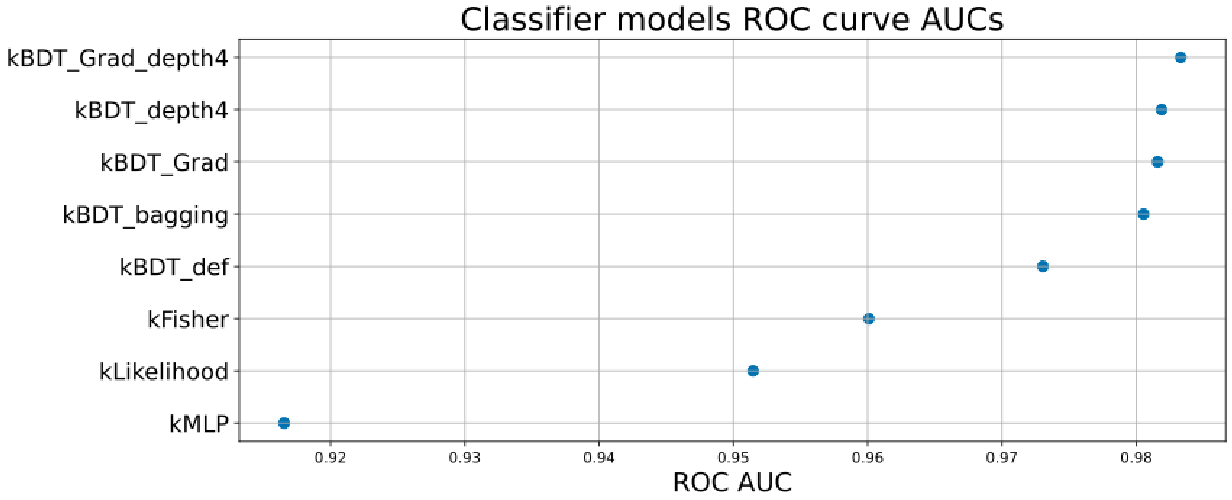
- Selected the classifier with the best test-set significance $Z_{test}(t^*)$



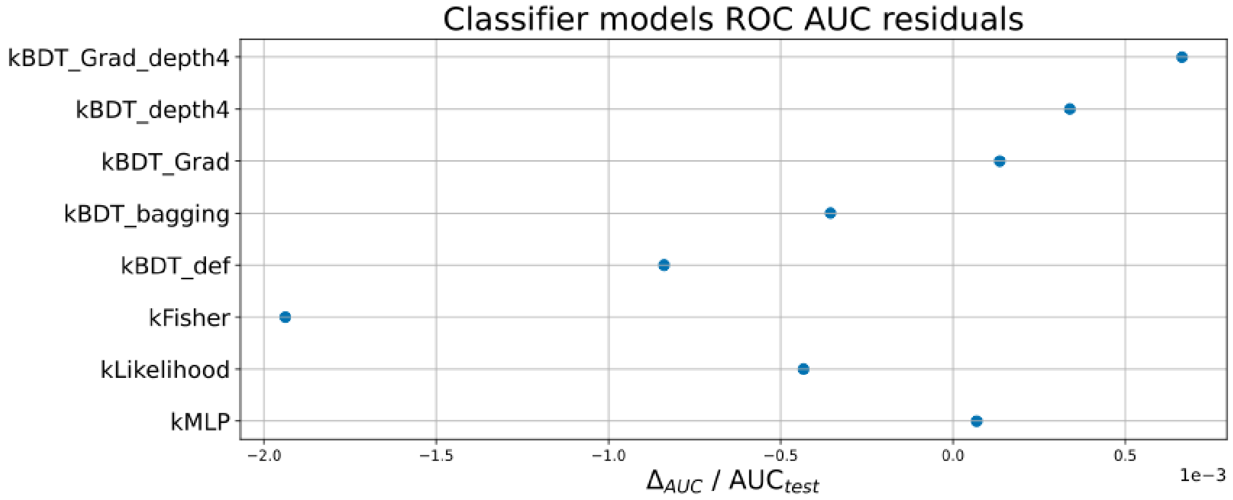
Best working-point significance found for kBDT_Grad_depth4

Initial classifier assessment

- Basic performance assessment considering the ROC curve



Both BDT and reference classifiers are correctly trained



Relative train-test AUC residual: 0.05% residuals indicate negligible overtraining

Decay length correction

- It's not just the x-sec that scales with U^2, the Dnu decay rate goes as:

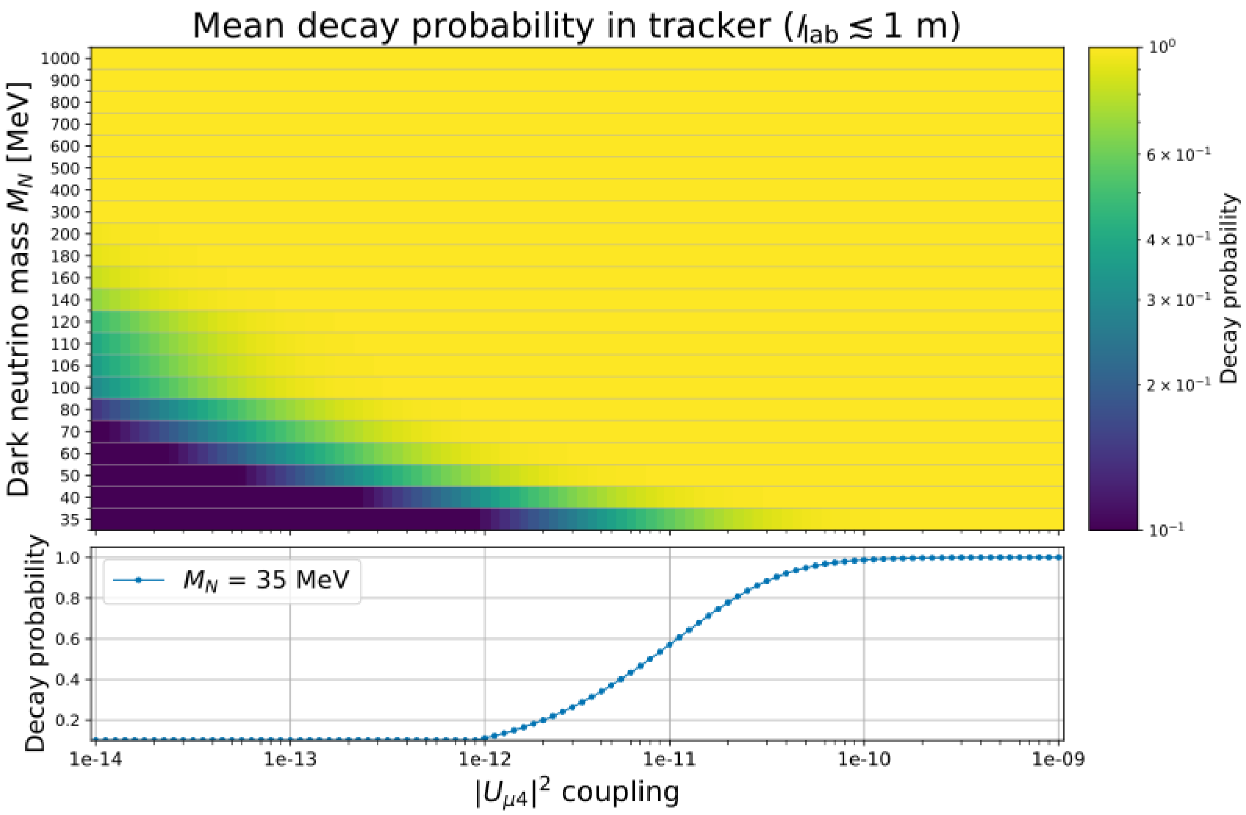
$$\Gamma_{N\to Z'+\nu} \propto U^2\,(1-U^2)$$

- At smaller couplings, the Dnu may decay outside the detector.

- Estimated the mean probability to decay within 1 m of the production point, relative to the production coupling:

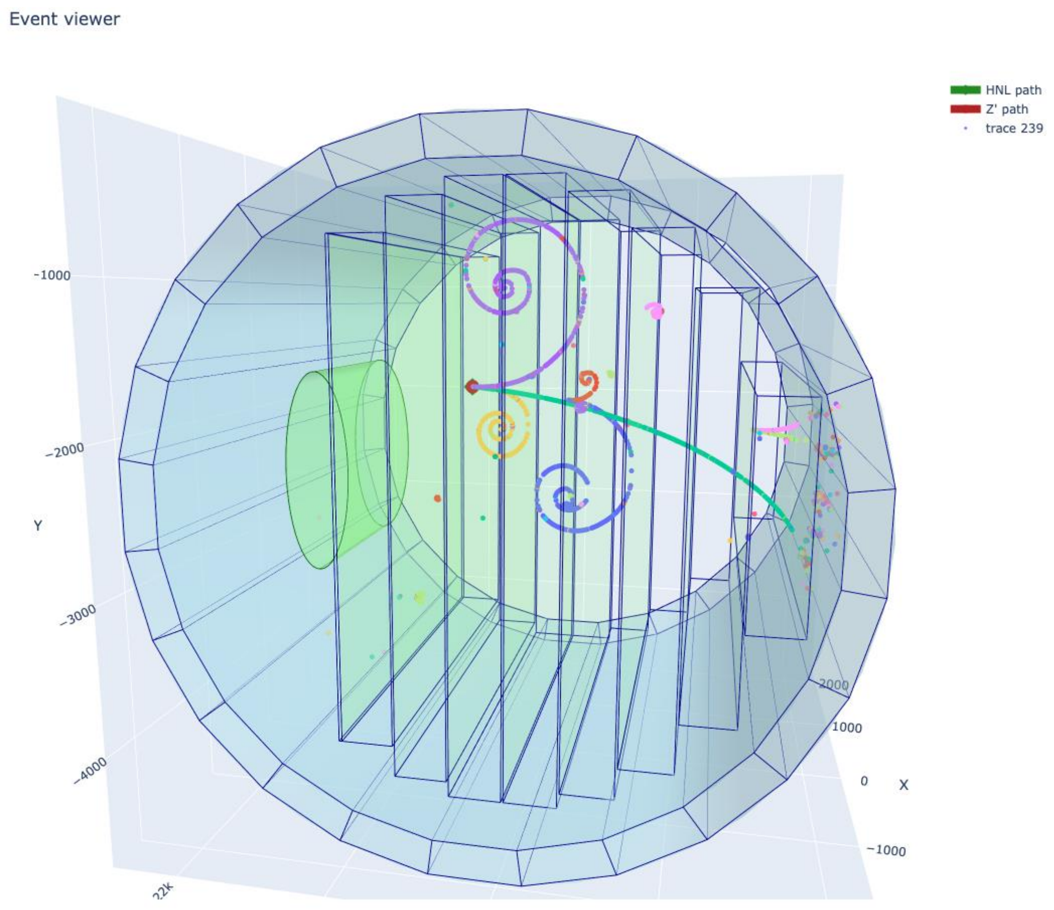
$$A_\lambda(M,U^2) \equiv \frac{P_M(l_{det},U^2)}{P_M(l_{det},U^2_{prod})}$$

- The resulting matrix is applied when scaling U^2 (~quite small effect)



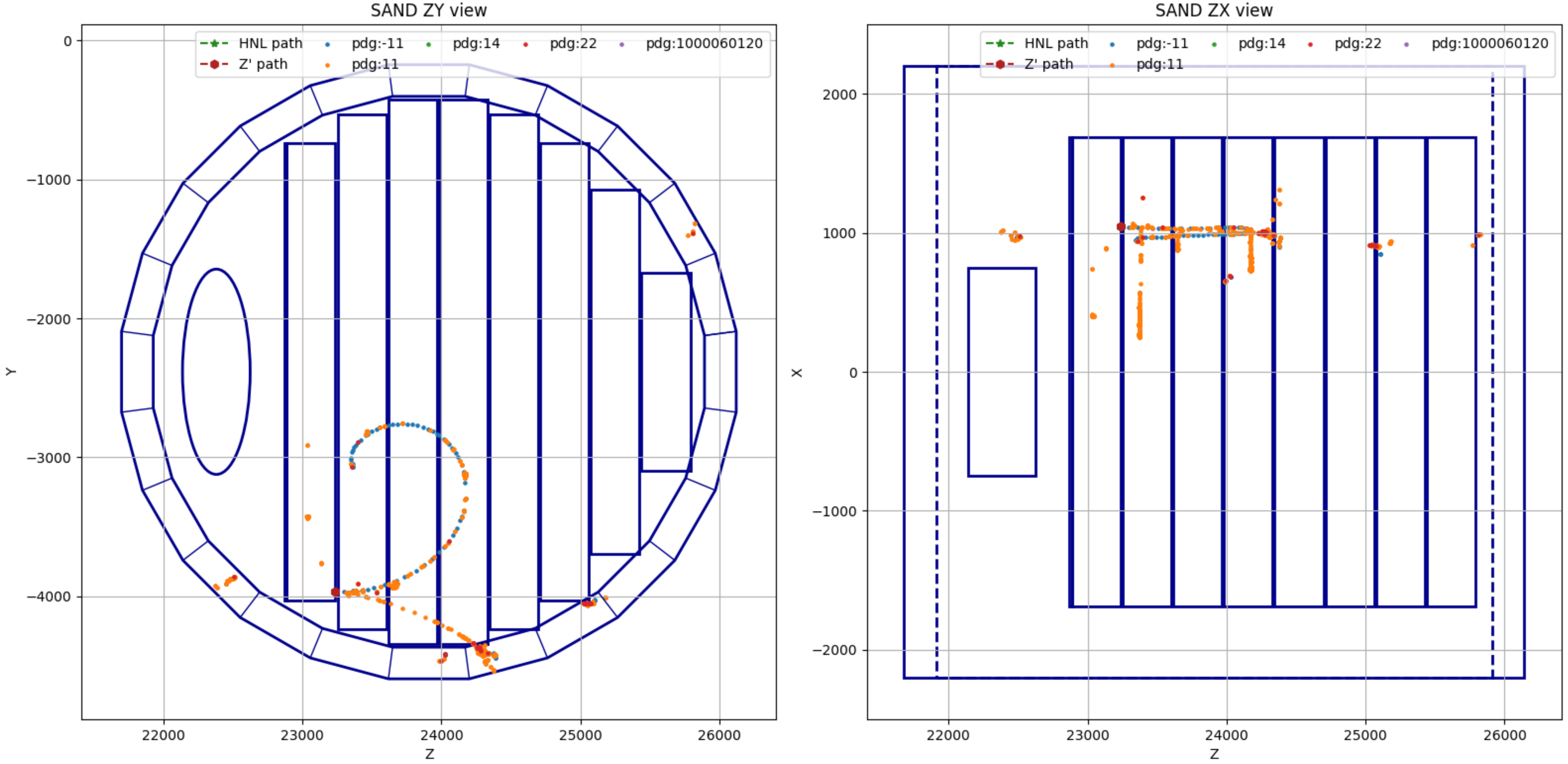
Event display

- Using the plotly Python library
- Generate outlines for:
  - GRAIN active volume
  - Drift supermodules
  - ECAL (hardcoded)
- 2D version in Matplotlib
- Data series added in custom functions
- Organised functions into a Python module (depends on code from analysis library)



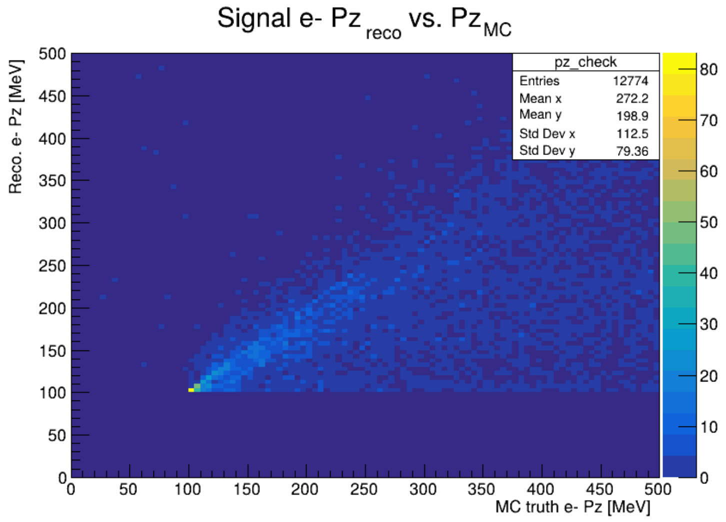
TrajectoryPoints by Trajectory ID

2D Display


SAND ZY view
HNL path
Z' path
pdg:-11
pdg:11
pdg:14
pdg:22
pdg:1000060120
0
−1000
−2000
−3000
−4000
Y
22000
23000
24000
25000
26000
Z
SAND ZX view
HNL path
Z' path
pdg:-11
pdg:11
pdg:14
pdg:22
pdg:1000060120
2000
1000
0
−1000
−2000
X
22000
23000
24000
25000
26000
Z

Preliminary tests of KF reconstruction

- KF does not yet include a Bremsstrahlung term
- Potentially keeping badly reconstructed tracks among the statistics
- Results do not seem significantly better than helicoidal fit
- Tried increasing the initial Covariance matrix terms (initial factor of 3x)




```
TMatrixD initial_cov_matrix(5, 5);
initial_cov_matrix[0][0] = 15*pow(200E-6, 2);
initial_cov_matrix[1][1] = 15*pow(200E-6, 2);
initial_cov_matrix[2][2] = 15*pow(0.1, 2);
initial_cov_matrix[3][3] = 15*pow(0.1, 2);
initial_cov_matrix[4][4] = 15*pow(0.1, 2);
```

Increased matrix terms by a factor of x3