Demonstration of background rejection using deep neural networks in the NEXT experiment

C. Adams, M. Del Tutto, J. Renner, K. Woodruff,
M. Kekic

Neutrino Physics and Machine Learning (NPML), July 21, 2020

Neutrino **E**xperiment with a **X**enon **T**PC

DIPC • U. de Girona • IFIC • U. Autónoma de Madrid • U. de Santiago de Compostela • U. Politécnica de Valencia • U. de Zaragoza

ANL • U. Texas at Arlington • FNAL • Harvard U. • Iowa State U. • LBNL • Texas A&M U.

U. de Aveiro • U. de Coimbra

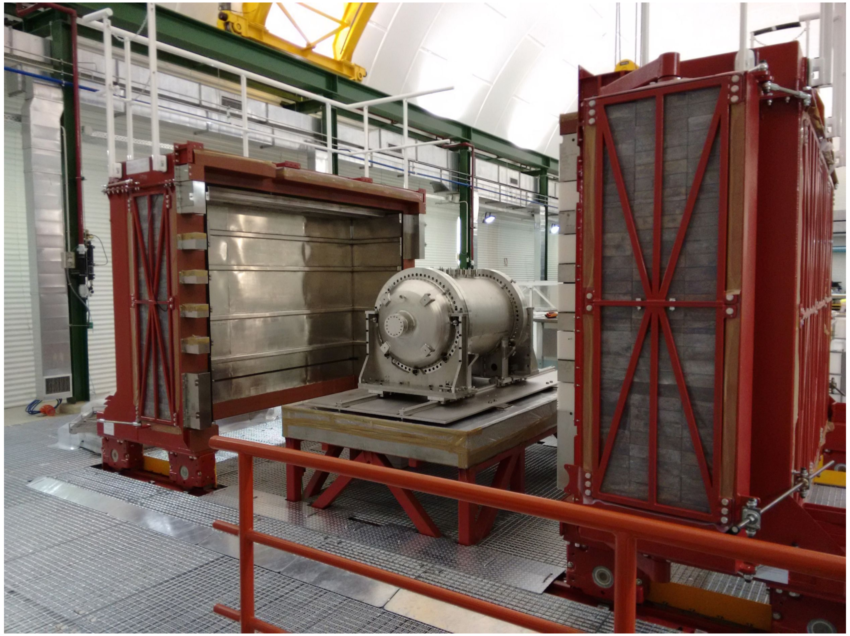
Ben-Gurion U. of the Negev

U. Antonio Nariño

Co-Spokespeople: Prof. J.J. Gómez Cadenas (DIPC)
Prof. D. R. Nygren (UTA)

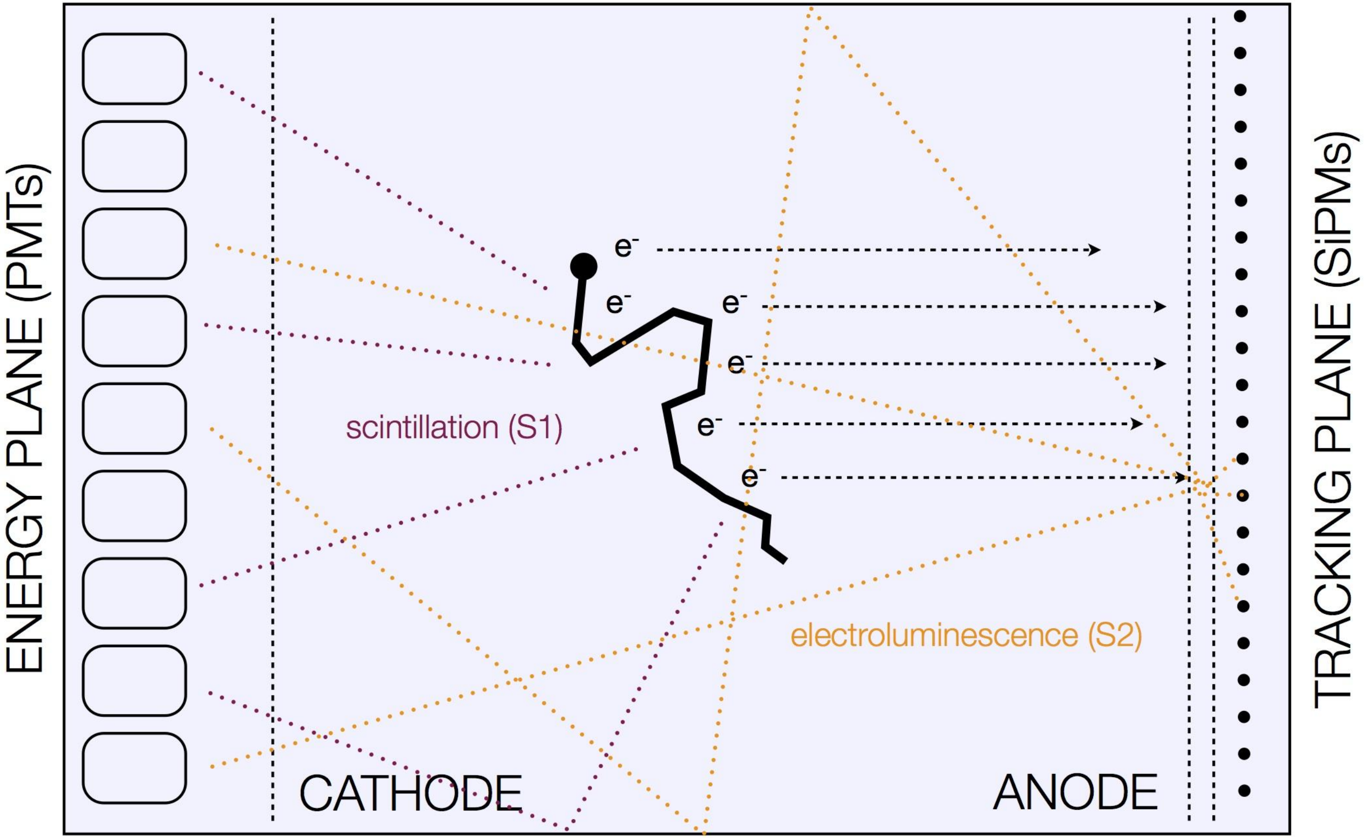
Funded by: erc • MARIE CURIE ACTIONS • GOBIERNO DE ESPAÑA MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD • GENERALITAT VALENCIANA • FCT Fundação para a Ciência e a Tecnologia • U.S. DEPARTMENT OF ENERGY • UNIVERSITY OF TEXAS ARLINGTON

Neutrino **E**xperiment with a **X**enon **T**PC



- NEXT-White (NEW) operating a 5 kg-scale demonstrator at the Canfranc Underground Laboratory (LSC)
- NEXT-100 to be commissioned in 2021: 100 kg Xe, enriched to ^{136}Xe(90%)

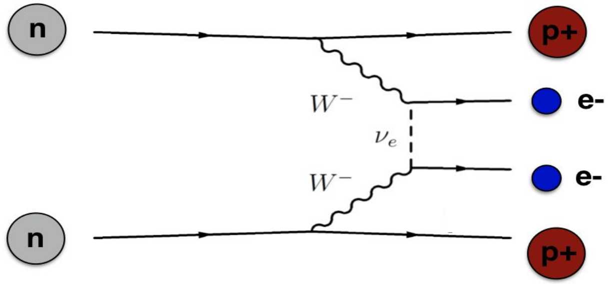
NEW : simulation and reconstruction

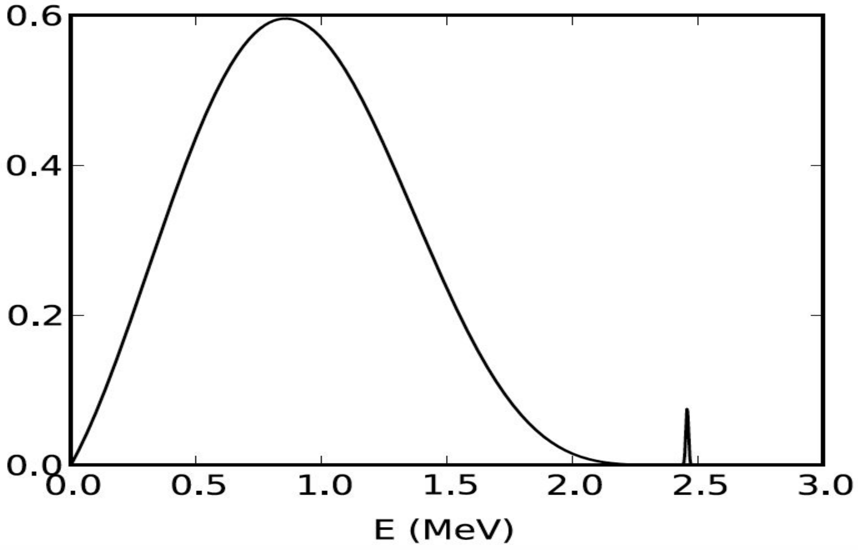


Geant4 based simulation:
- simulate energy deposits ('hit') of charged particles in the Xe gas
- simulate PMTs/SiPMs responses

Reconstruction (same for data and MC):
- find XYZ position of hits based on SiPM signal; assign energy measured at PMTs plane
- correct and calibrate hit energy
- voxelize event such that voxel energy is the sum of hits energy inside the voxel

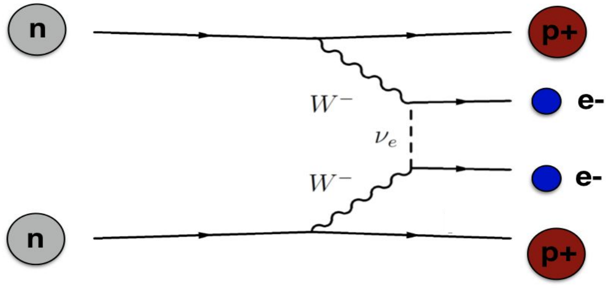
Neutrinoless double beta decay

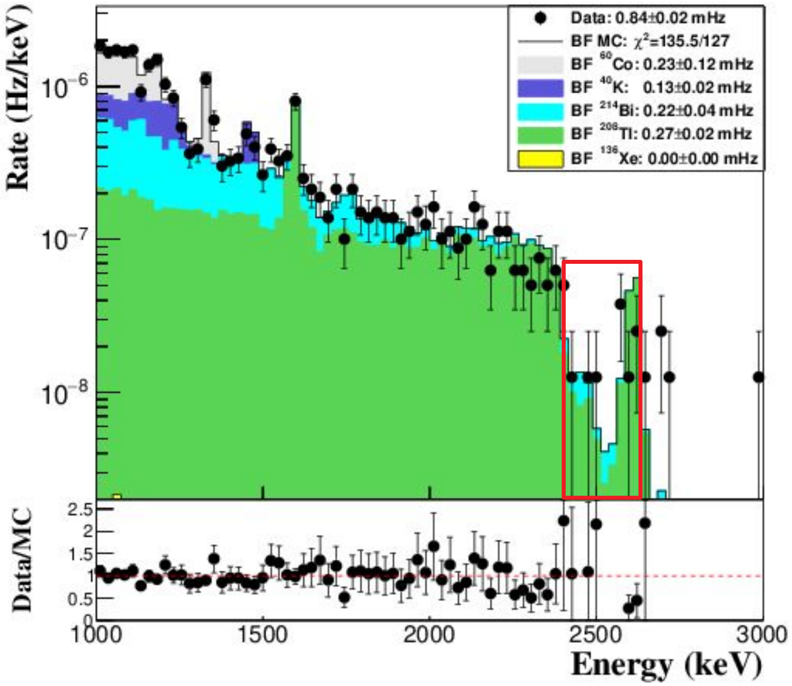




Essential:
1. Good energy resolution

Neutrinoless double beta decay



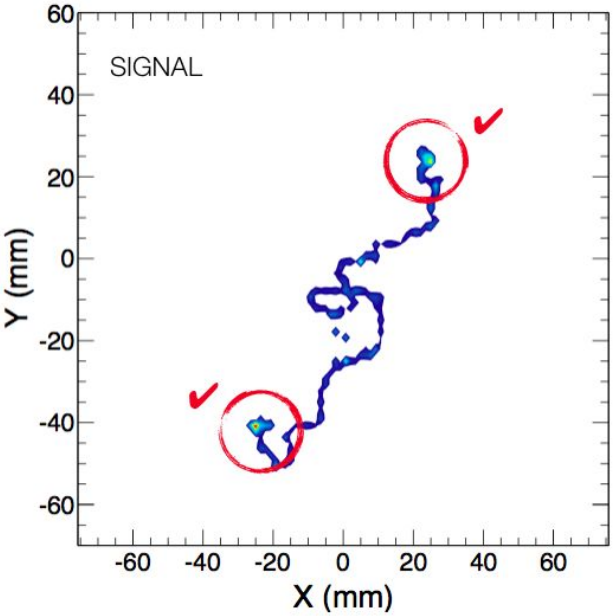


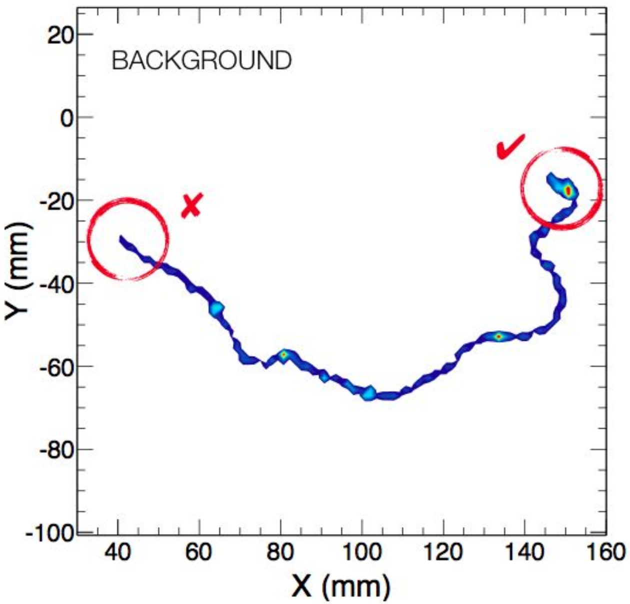
Essential:

1. Good energy resolution
2. Good background identification

arXiv:1905.13625
JHEP 10 (2019) 051

Background identification





At the end of the track the energy deposited per unit length increases- Bragg peak (blob):

- Signal : 2 blobs
- Background : 1 blob

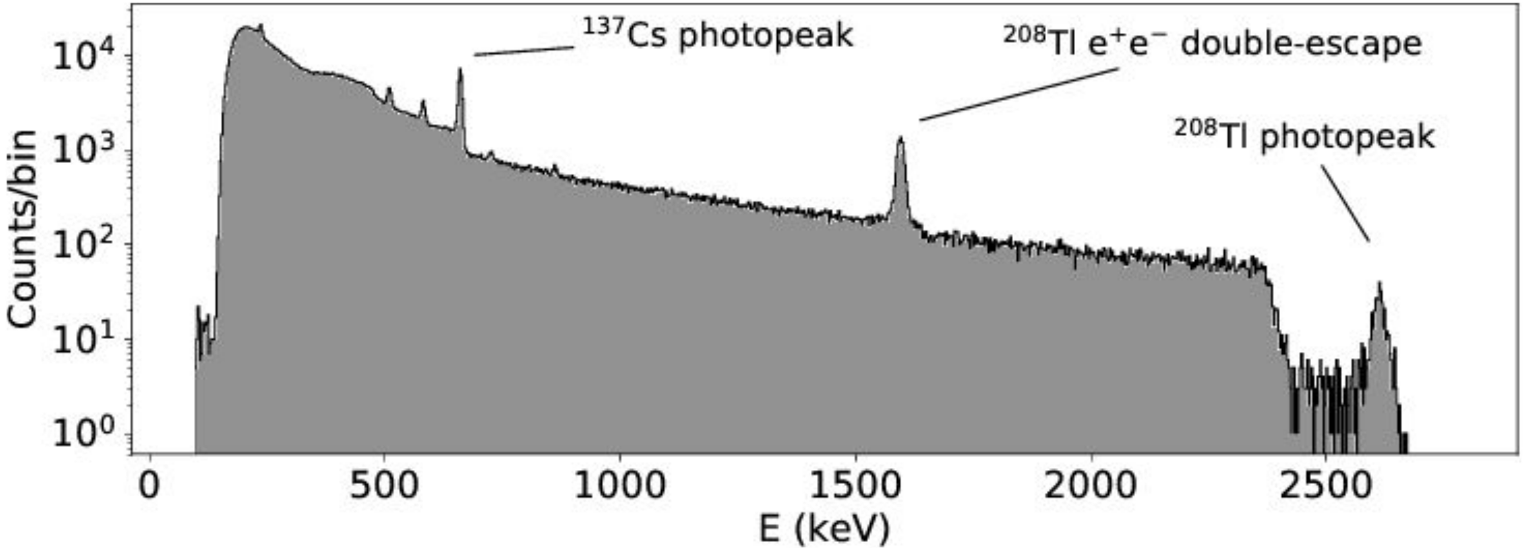
Previous study

- **Initial Monte Carlo-based study:** JINST 12 (2017) 01, T01004 (arXiv:1609.06202)

Cut	Signal Events		BG Events (^{208}Tl)		BG Events (^{214}Bi)	
	$2\times2\times2$	$10\times10\times5$	$2\times2\times2$	$10\times10\times5$	$2\times2\times2$	$10\times10\times5$
(Initial events)	1.0	1.0	1.0	1.0	1.0	1.0
Energy	7.59×10^{-1}	7.59×10^{-1}	2.27×10^{-3}	2.27×10^{-3}	1.42×10^{-4}	1.42×10^{-4}
Fiducial	6.71×10^{-1}	6.68×10^{-1}	1.19×10^{-3}	1.17×10^{-3}	8.62×10^{-5}	8.54×10^{-5}
Single-Track	3.75×10^{-1}	4.79×10^{-1}	7.90×10^{-6}	1.81×10^{-5}	3.84×10^{-6}	8.75×10^{-6}
Classification*	3.23×10^{-1}	3.67×10^{-1}	7.70×10^{-7}	2.41×10^{-6}	2.90×10^{-7}	9.59×10^{-7}
Classification (DNN)	**3.23×10^{-1}**	**3.67×10^{-1}**			**1.80×10^{-7}**	**8.22×10^{-7}**

Now we can test it on data

Proof-of-concept in NEXT-white with e^+e^- track



arXiv:1905.13110 (2019)
JHEP 10 (2019) 230

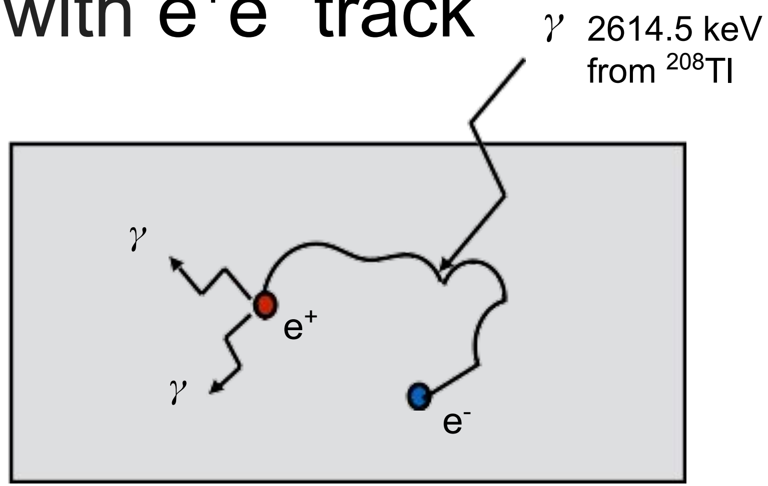
Calibration with ^{137}Cs and ^{228}Th sources
Achieved <1% Resolution at FWHM near $Q_{\beta\beta}$

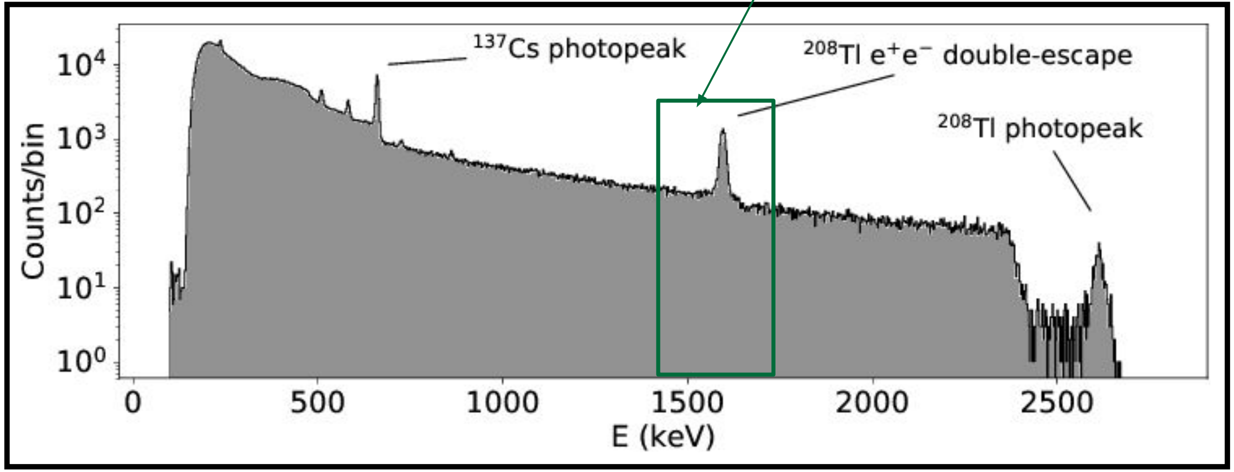
Proof-of-concept in NEXT-white with e^+e^- track

We can test our model on DE (double escape) peak:

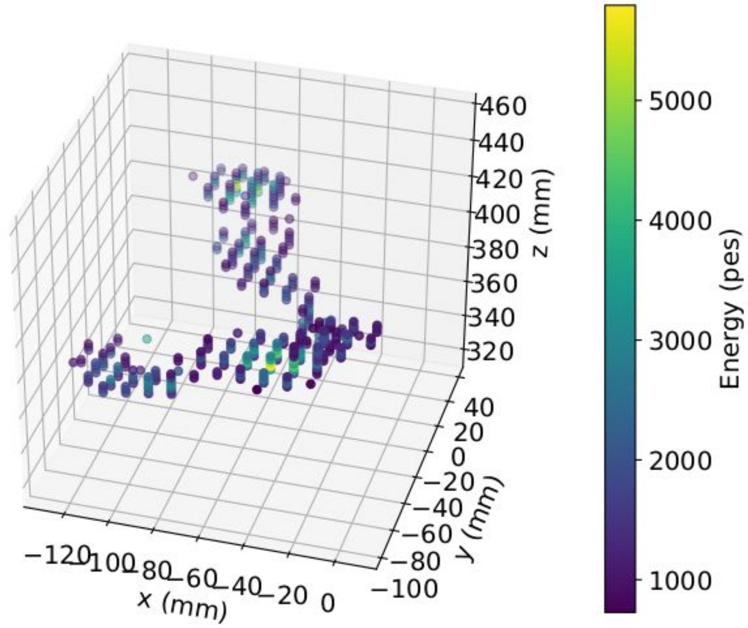
- Signal : e^+e^- track
- Background : single e^- track

Can calculate ratio of the signal (gaussian) and the background (exponential)

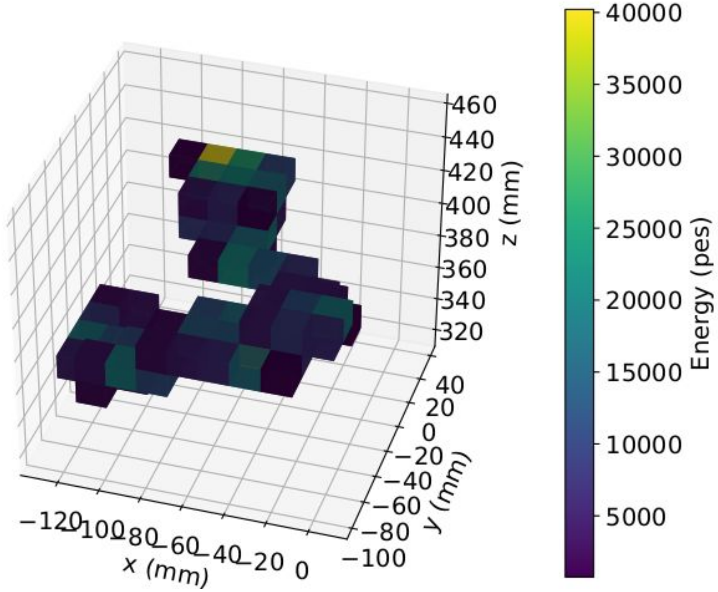




Event example



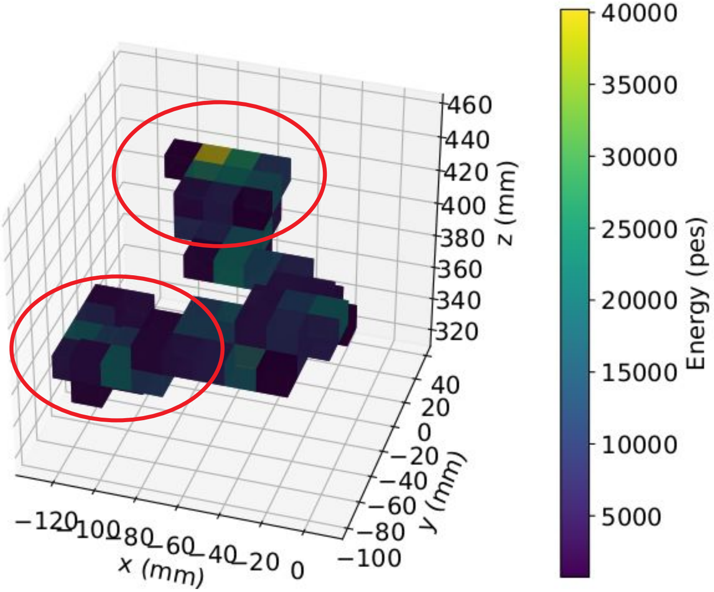
Reconstructed hits



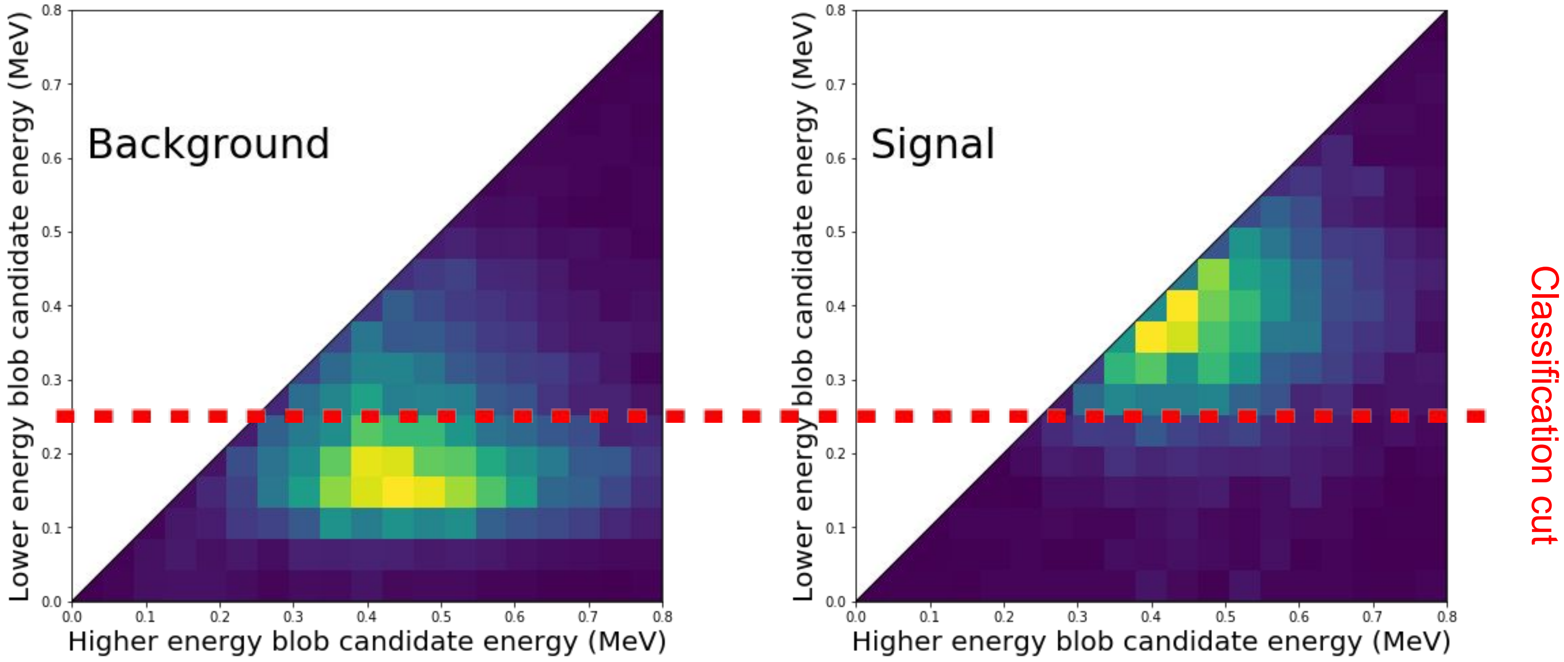
Voxelized hits

Classical approach

1. Find the track based on graph theory
2. Identify track extremes
3. Calculate energy inside the blobs



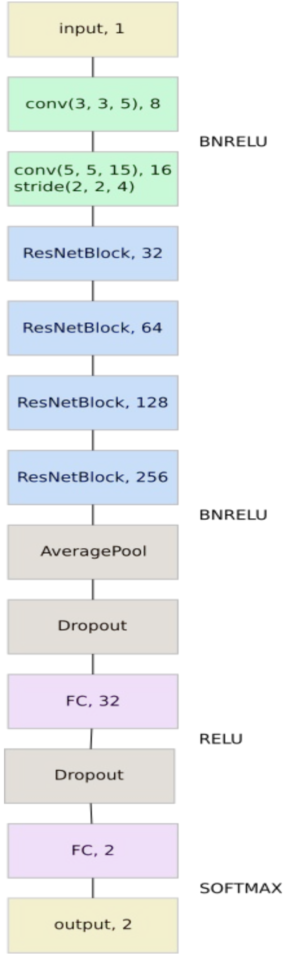
Classical approach

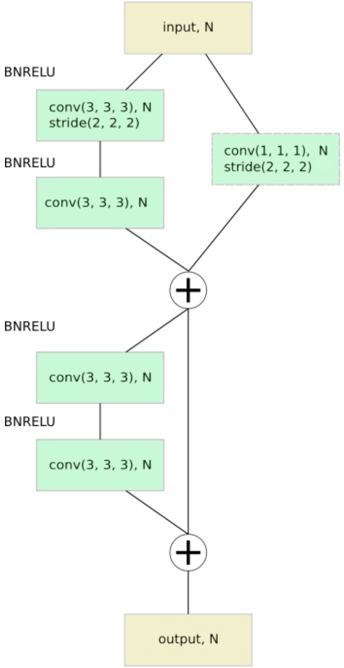


Geant4 Monte Carlo:

background (single electrons) and signal (electron-positron pair) blob energies distributions

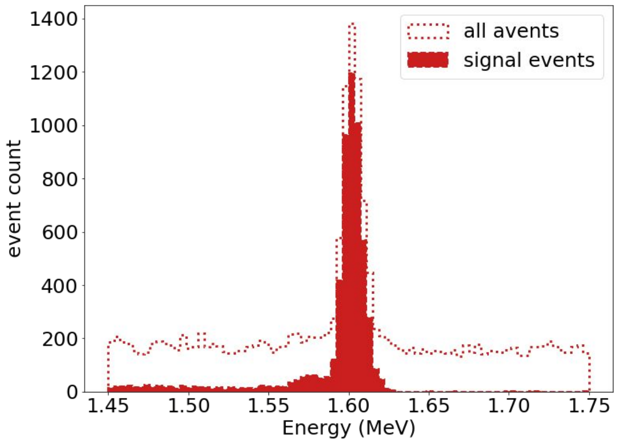
DNN based approach





- Input size 40X40x110 (voxel size 10x10x5 mm^3)
- Energy of every event normalized to 1 (so the network does not have information about total event energy)
- ~500000 fiducial events, 35% signal

validation MC energy distribution



We experimented with both standard dense networks (keras/tf implementation) and sparse networks (Submanifold Sparse Convolutional Networks in pytorch) achieving similar results

Plots in this talk are for sparsenet

Evaluation metrics

1. **AUC-ROC*** : Degree of distinguishability between classes – higher is better

True Negative rate $(\frac{\text{rejected \ background}}{\text{total \ background}})$

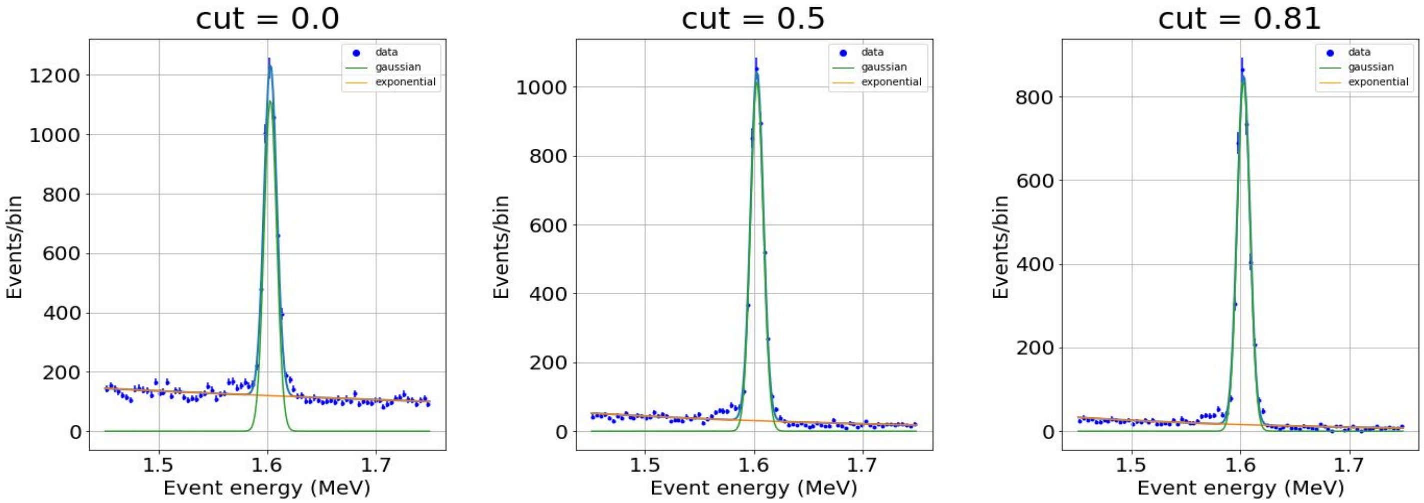
vs

True Positive rate $(\frac{\text{accepted \ signal}}{\text{total \ signal}})$

2. **Figure of merit** : $\frac{\epsilon_{sig}}{\sqrt{\epsilon_{bck}}}$ – higher is better

The sensitivity to the half-life of the ββ0v decay is proportional to f.o.m
in background-limited experiments (arXiv:1010.5112v4)

Evaluation based on energy spectrum



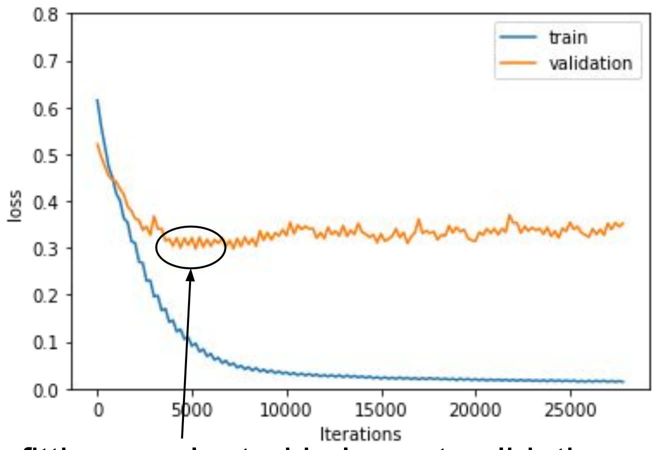
1. Fit the histogram to gaussian (signal) and exponential (background) $N^0_{bck},\ N^0_{sig}$

2. Integrate to calculate total number of signal and background $N^i_{bck},\ N^i_{sig}$

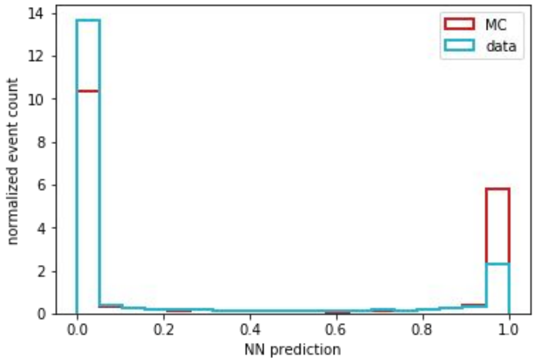
3. Apply i^{th} cut on DNN prediction and calculate

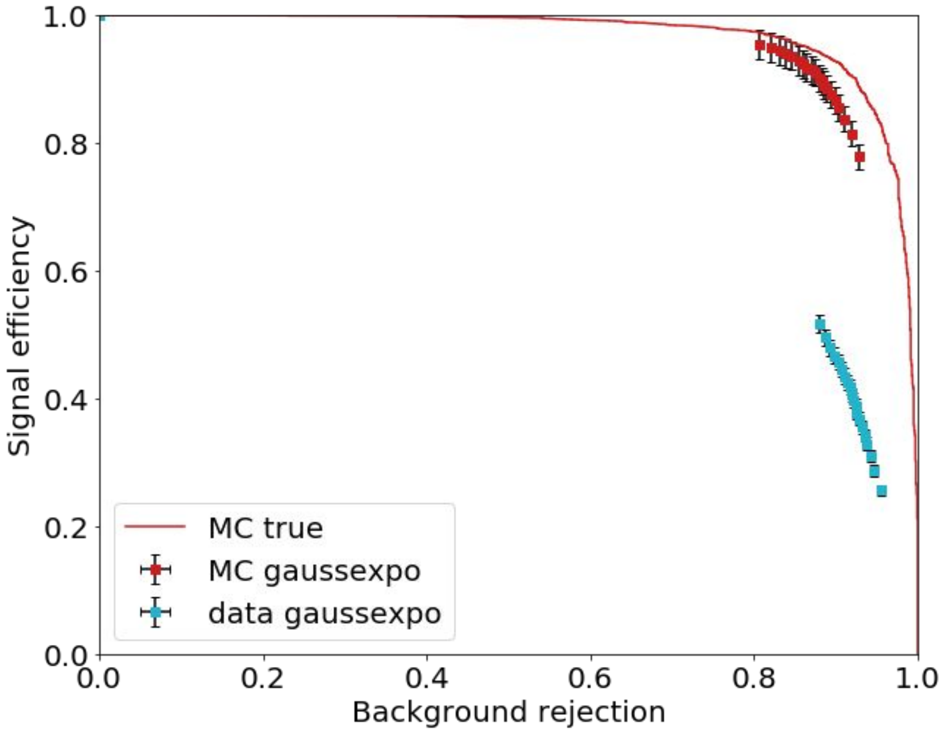
$$\epsilon^i_{sig} = \frac{N^i_{sig}}{N^0_{sig}} \qquad \epsilon^i_{bck} = \frac{N^i_{bck}}{N^0_{bck}} \qquad \mathrm{f.o.m} = \frac{\epsilon^i_{sig}}{\sqrt{\epsilon^i_{bck}}}$$

Training on MC, evaluating on data



overfitting, evaluated in lowest validation point

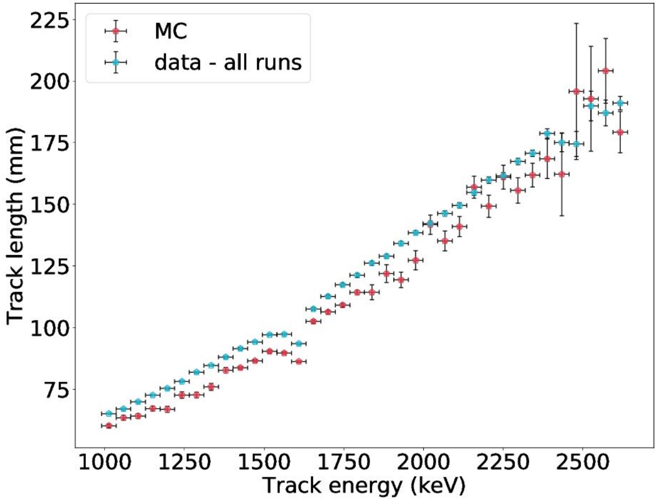




Prediction on data is biased towards lower values

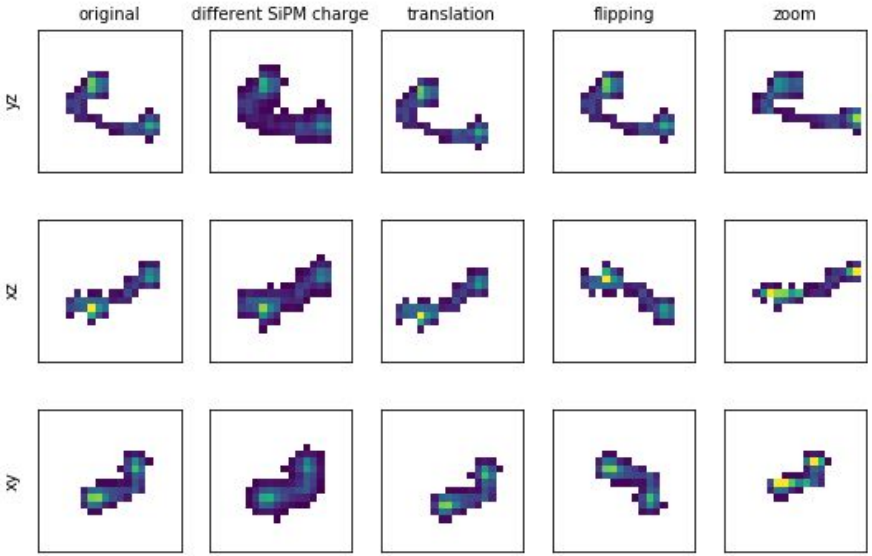
Known MC/data differences

From classical analysis we know there are some MC/data differences, e.g. track length, bob energy...



arXiv:1905.13141
***JHEP* 10 (2019) 052**

To make network robust to those differences we apply on-fly augmentation



Can we predict performance on data without using the data?

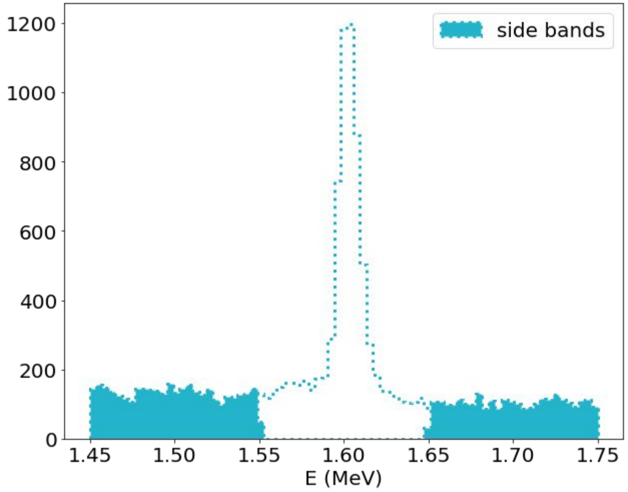
Try to predict data/MC disagreement looking at disagreement of features distributions on sidebands.

Assumption: domain shift does not depend on type of events — if network is robust to data/MC differences on background events it will be on signal events as well.

To quantify the distances of distributions in Ndim space we used Energy test statistics:

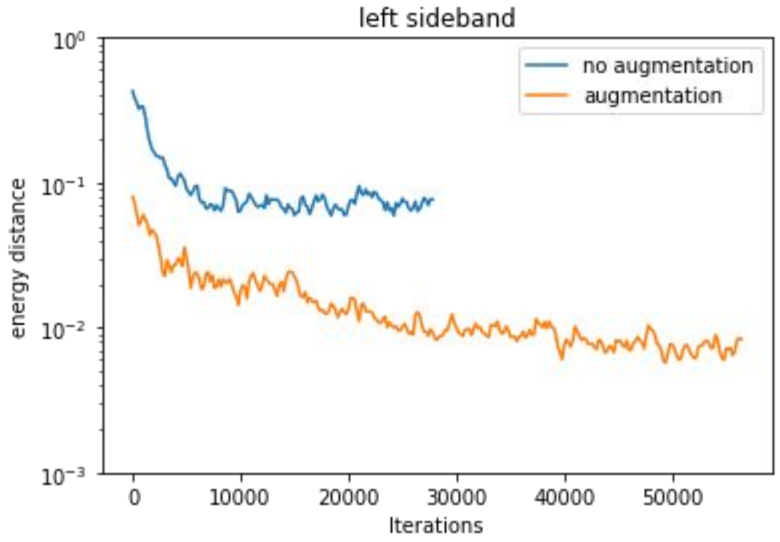
$$A := \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}\|x_i - y_j\|, B := \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|x_i - x_j\|, C := \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m}\|y_i - y_j\|$$

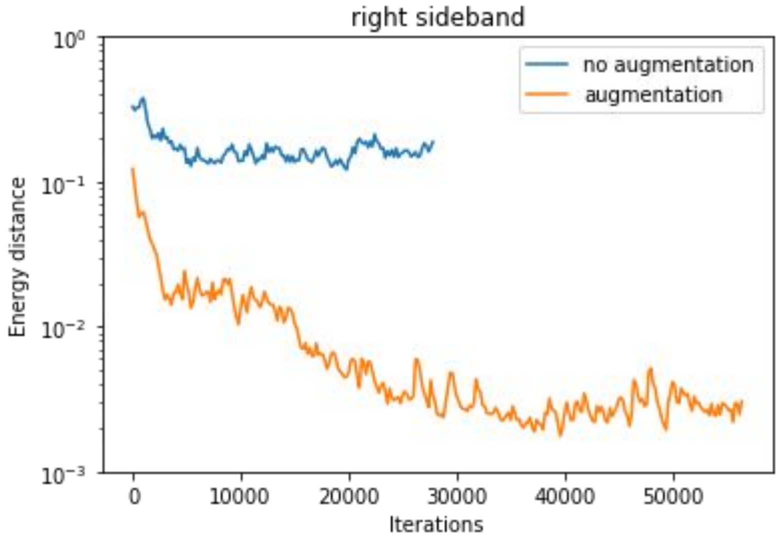
$$E_{n,m}(X,Y) := 2A - B - C$$



Sideband discrepancy tracking

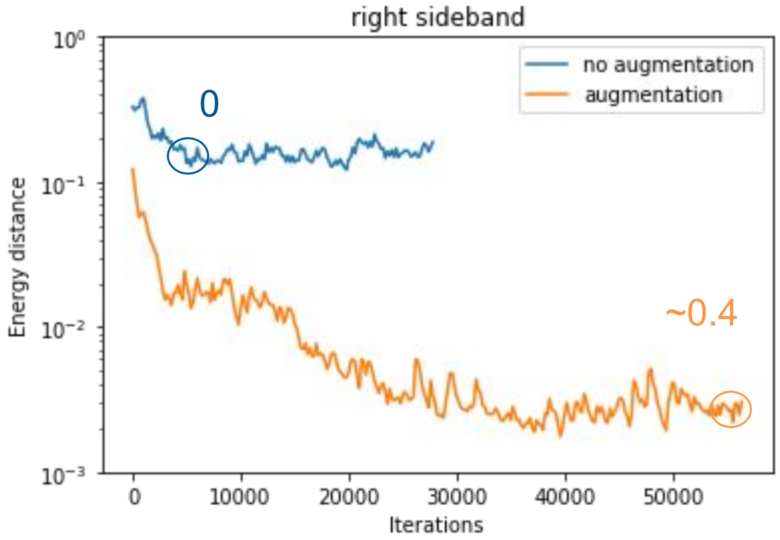
Compare distributions of extracted features - first flatten layer





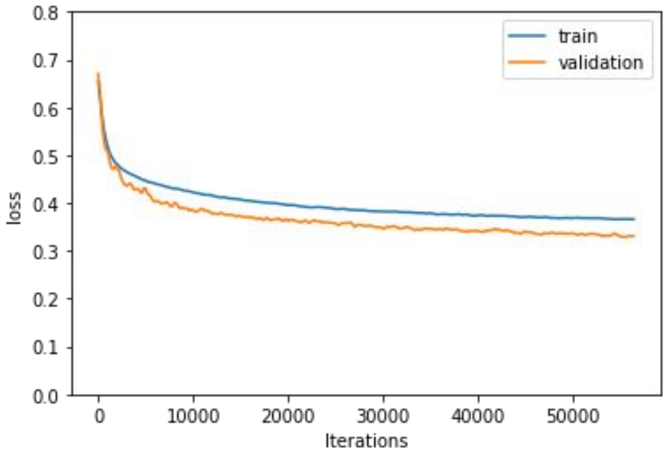
Sideband discrepancy tracking

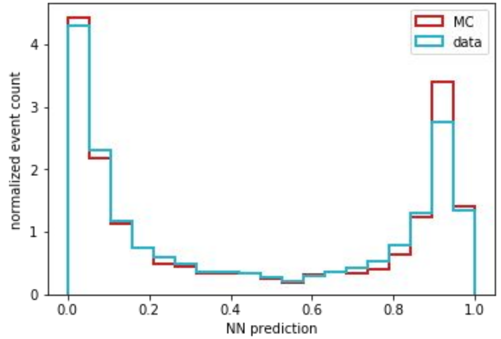
Compare distributions of extracted features - first flatten layer

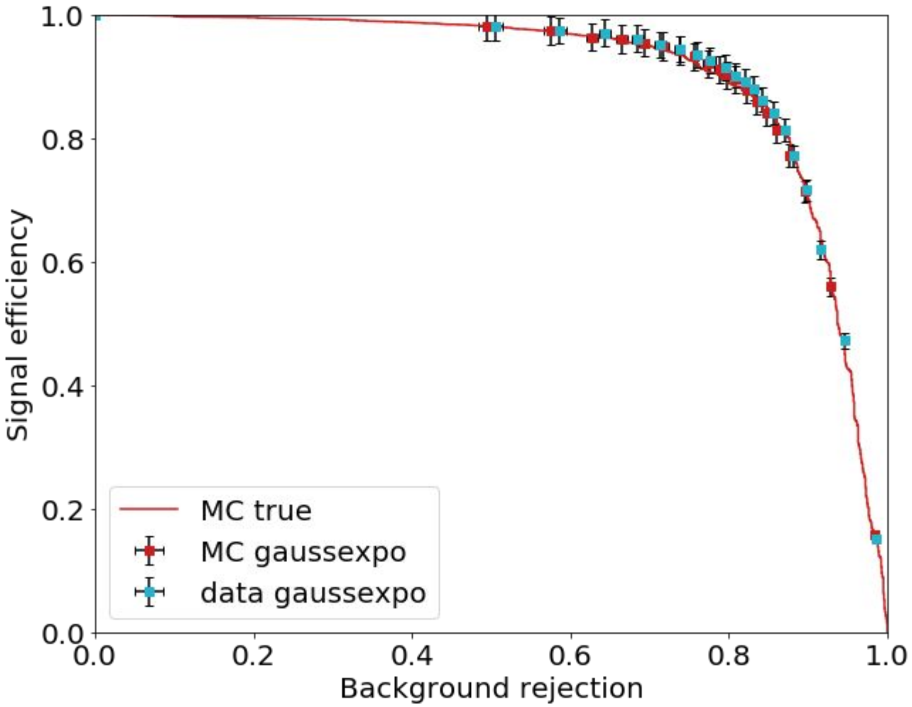


Permutation test to extract p-value, mix MC/data samples and calculate the fraction of trials (1000 trials) in which the distance of original MC/data is smaller than obtained one (small p-value allows us to reject the null hypothesis - MC and data come from the same distribution)

Training with augmentation

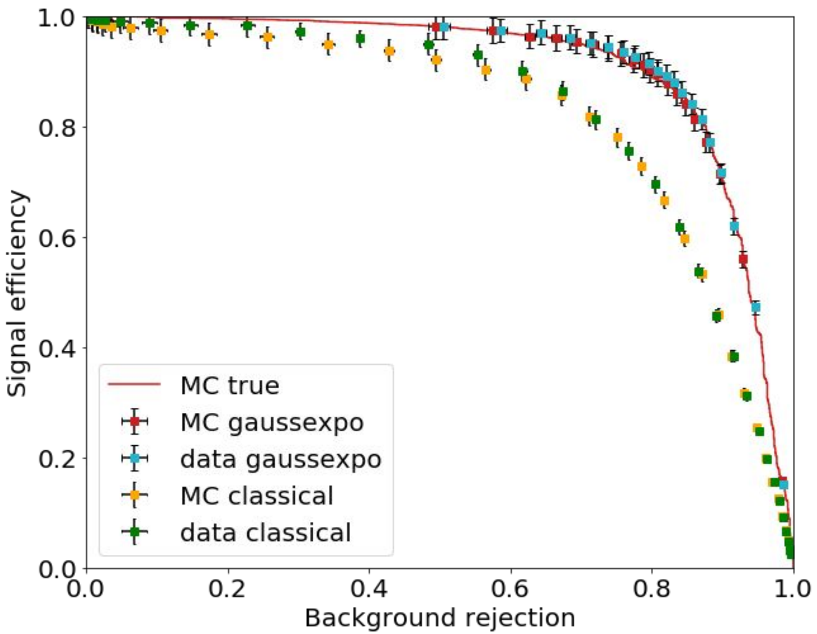


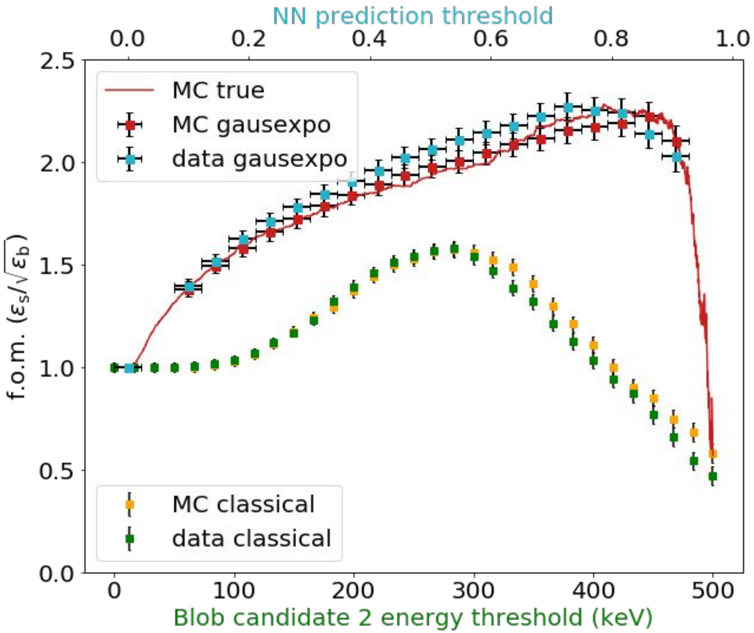




Much nicer result!

Comparison with classical analysis



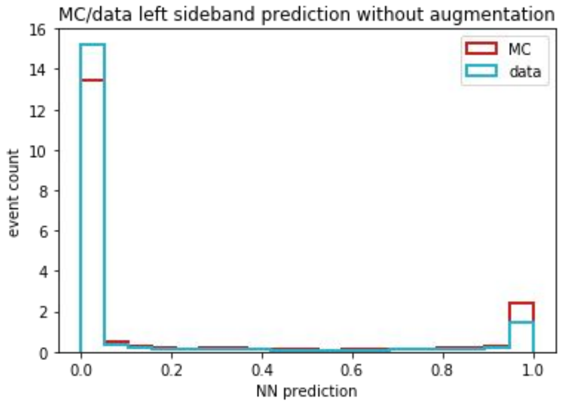


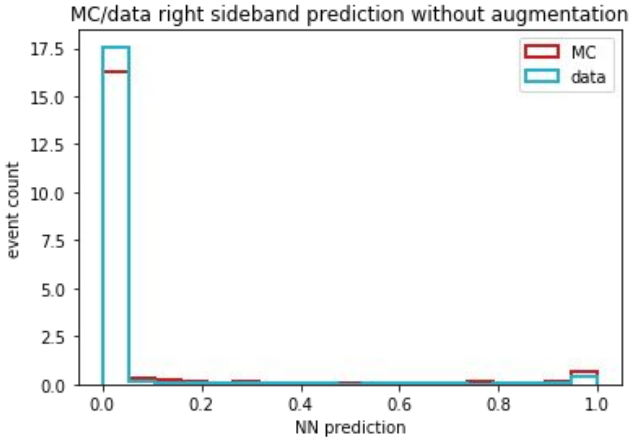
Summary

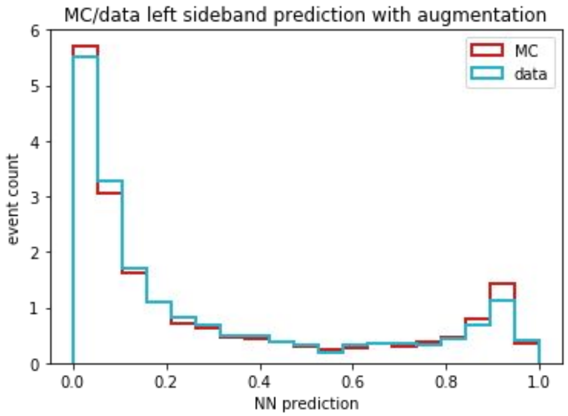
- We have used NEXT-White double escape peak calibration data to demonstrate topological discrimination of signal vs background
- There are MC/data discrepancies that affect the performance on data
- The network can be made more robust with the help of data augmentation
- Analysis of background sidebands helps determining optimal point for evaluation on data
- Future plans : next talk!

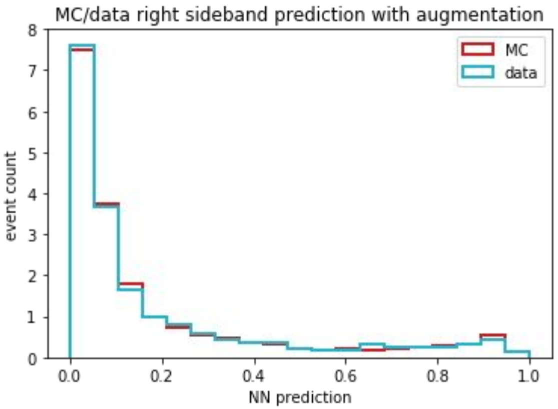
BACKUP SLIDES

NN prediction distributions

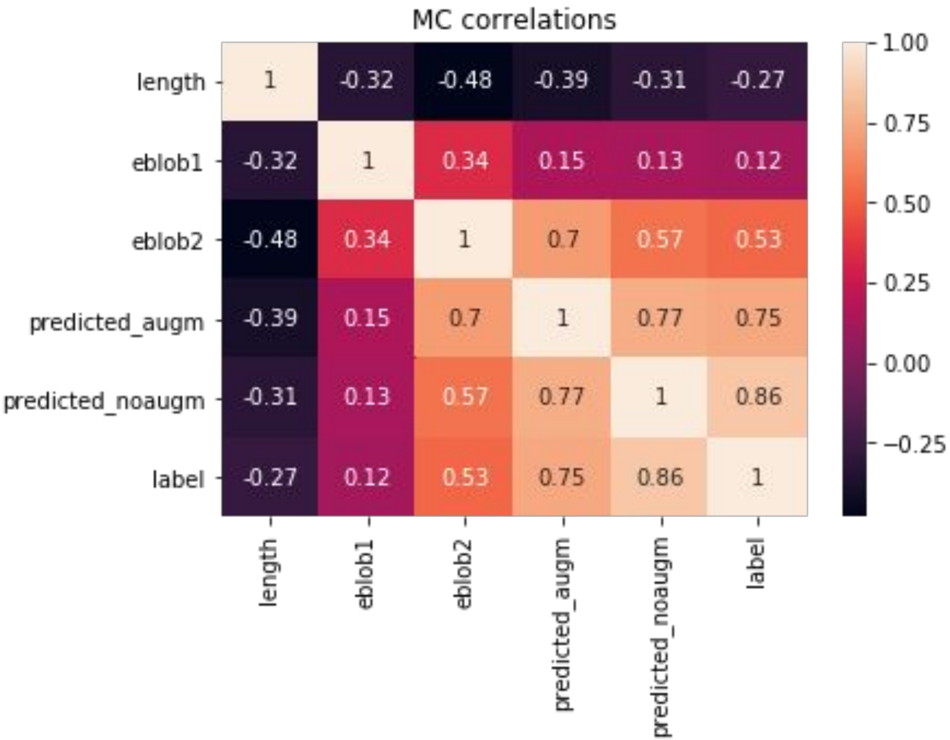


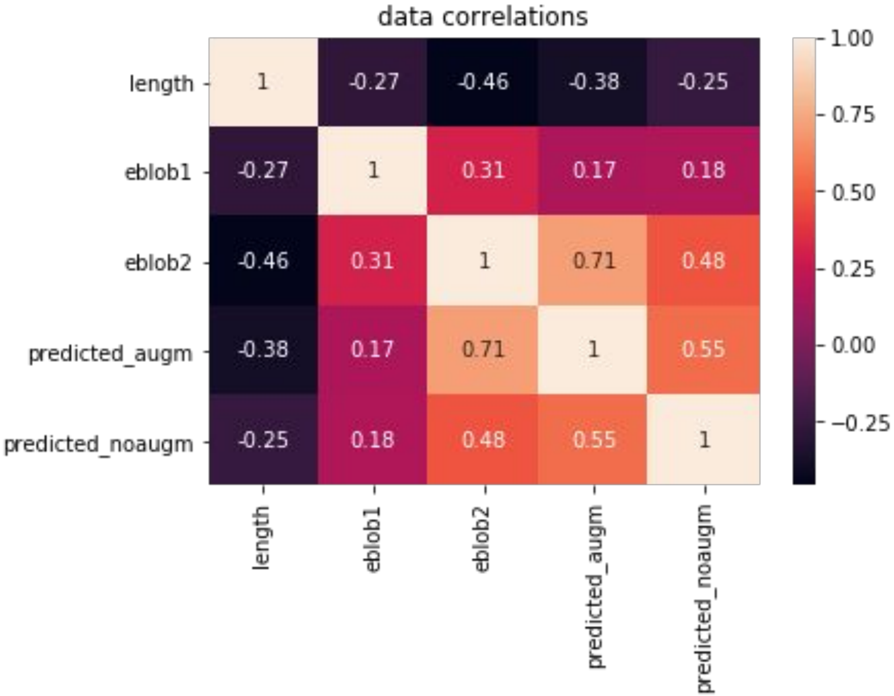




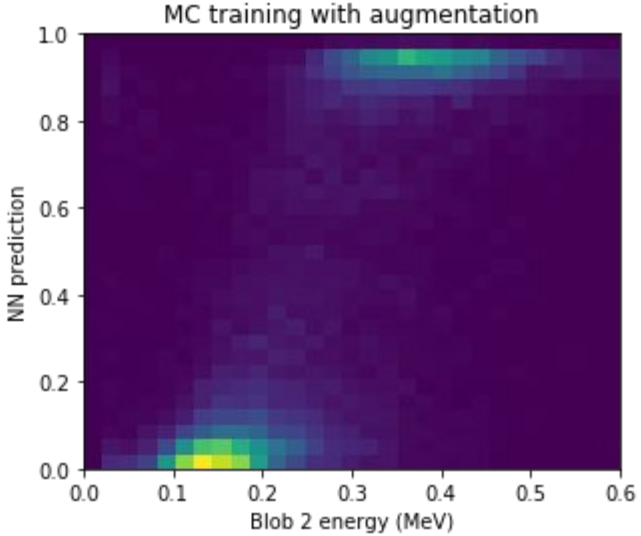


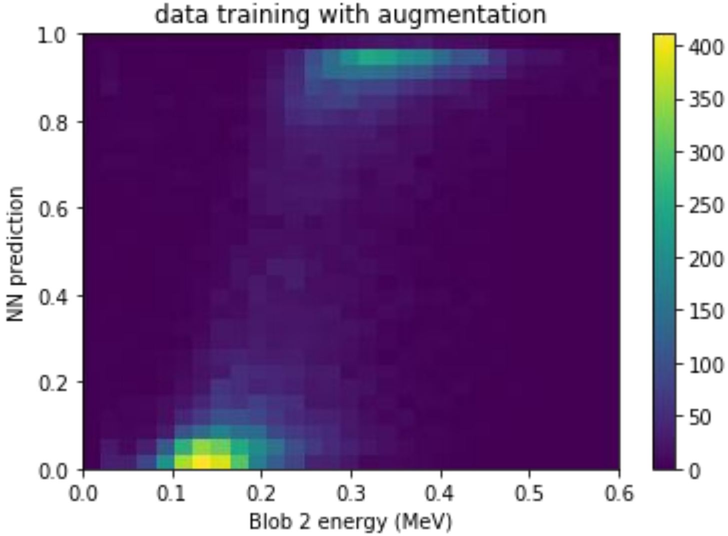
Correlations





Correlation with eblob2





Some events predicted signal with augmented training and background without augmentation