Particle Identification with a Machine Learning Technique in the Hyper-Kamiokande Detector

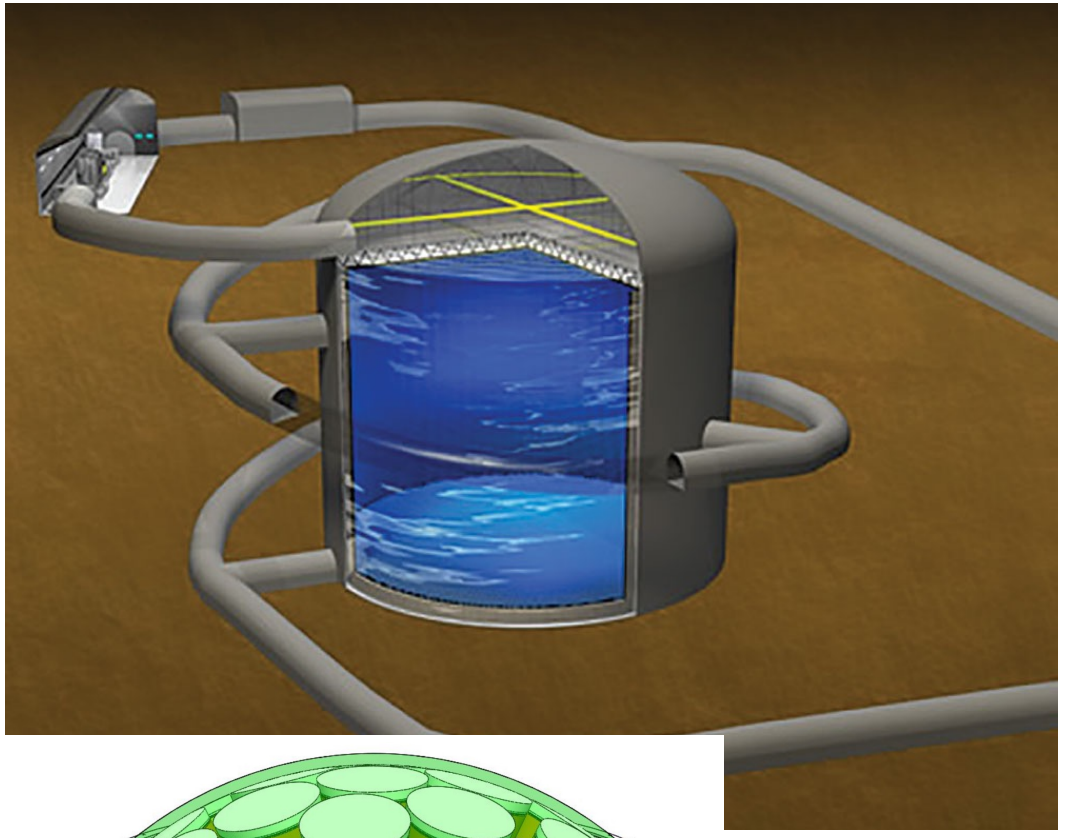
22 November 2022

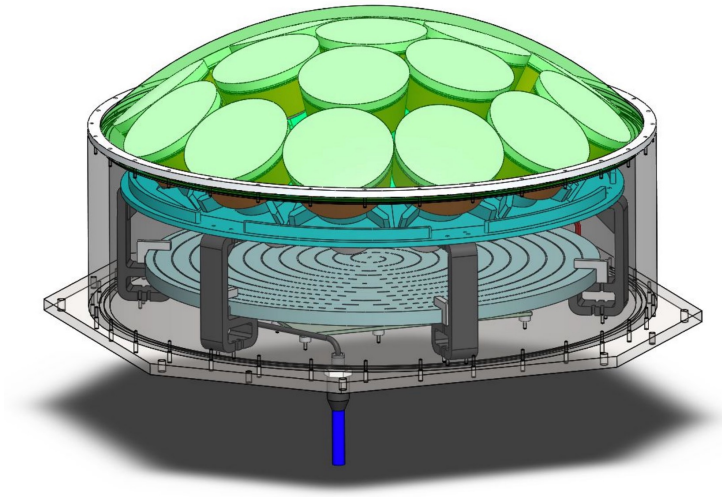
Joanna Gao

KING'S College LONDON

The Hyper-Kamiokande

- A water Cherenkov detector, the successor of Super-Kamiokande (Super-K)
- A cylinder with ~70 meter in height and diameter (roughly 260 ktons), 8 times the fiducial volume of Super-K
- Housing around 20k 20 inch PMTs and additional multi-PMTs for more signal granularity
- The signals are mostly in the form of Cherenkov rings

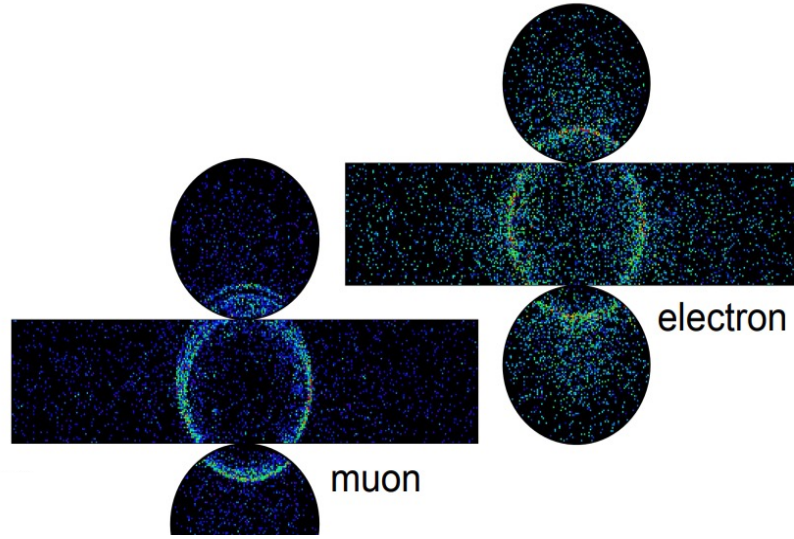


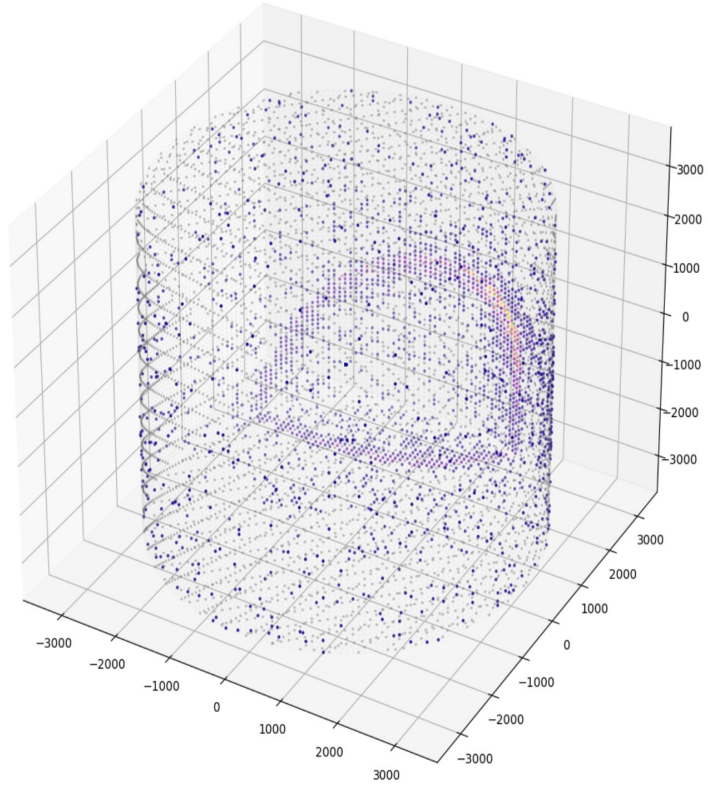


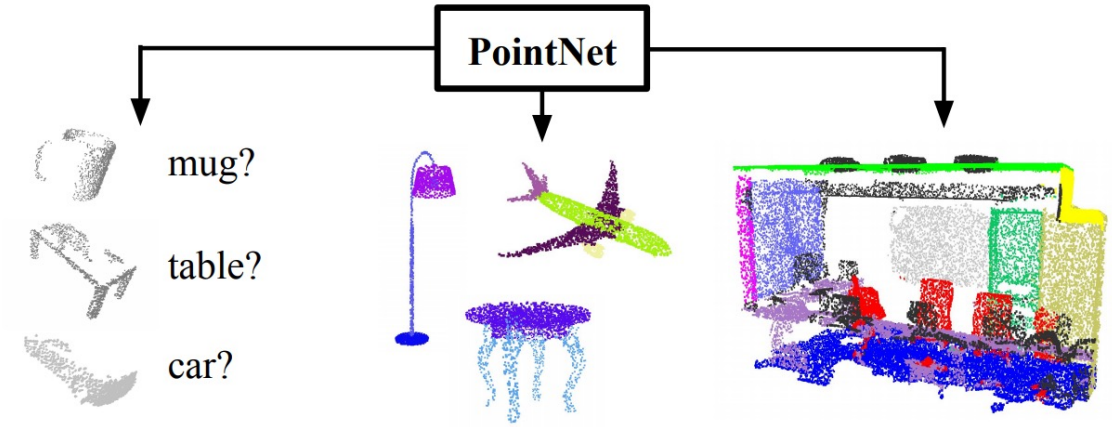
KING'S College LONDON

PointNet

- Unlike convolutional neural network (CNN), which unwraps a 3D model into a 2D image, PointNet is a 3D classification and segmentation tool
- Advantage:
  - retaining location, timing and charge relation between hit PMTs;
  - can apply to any detector size and geometry
  - faster than statistical reconstruction tool



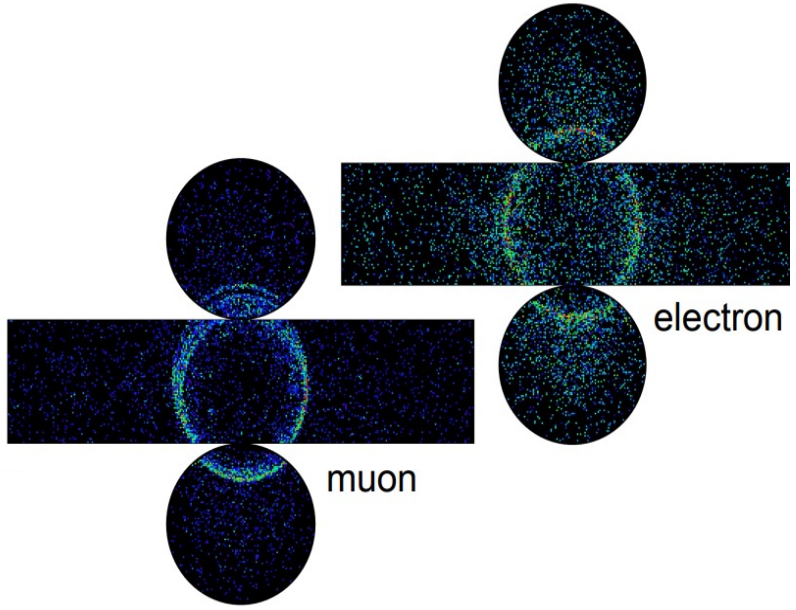




[1]

FiTQun – the Try and Tested Reconstruction Tool

- Negative log-likelihood based method
- Widely used in Super-K for O(100) MeV to O(10) GeV signal reconstruction
- Been migrated to the Hyper-K software platform
- Reliable but very time consuming (~2 minute per event) comparing to machine learning model evaluation (0.1 second per event)



$$L(\mathbf{x}) = \prod_{j}^{\text{unhit}} P_j(\text{unhit}|\mathbf{x}) \prod_{i}^{\text{hit}} \{1 - P_i(\text{unhit}|\mathbf{x})\} f_q(q_i|\mathbf{x}) f_t(t_i|\mathbf{x})$$

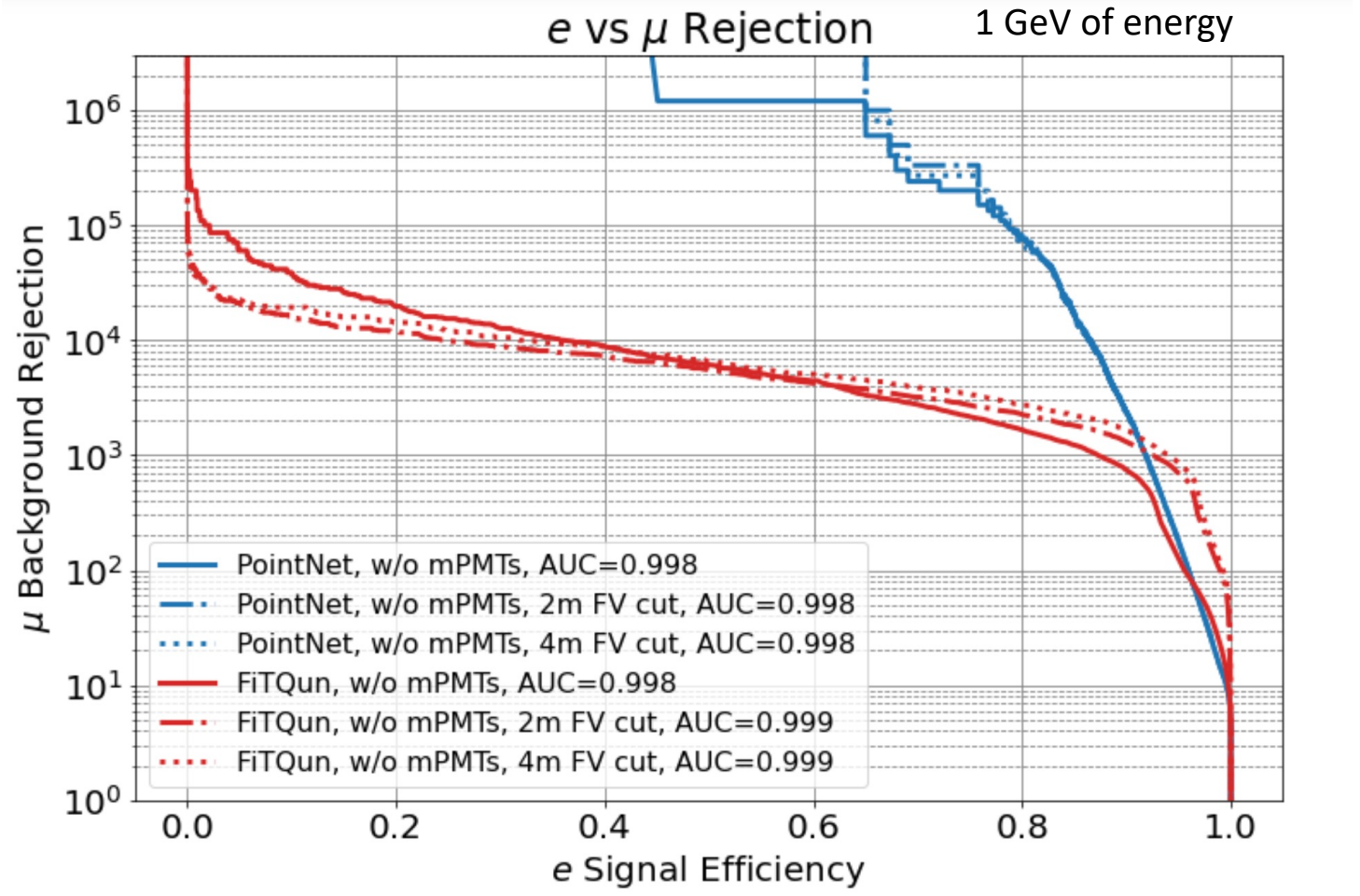
i-th PMT's charge, time

Fit parameters — Unhit probability — Hit probability — Charge likelihood — Time likelihood

Current Results – e/mu Separation

- Figure of merit: area under the ROC curve (AUC) $\in (0,1]$
- There's issue with mPMT geometry which confuses PointNet which leads to worse results -> ignoring mPMTs in this plot
- FiTQun responds better to different fiducial volume (FV) cuts than PointNet, but understandable due to the nature of the fiTQun model

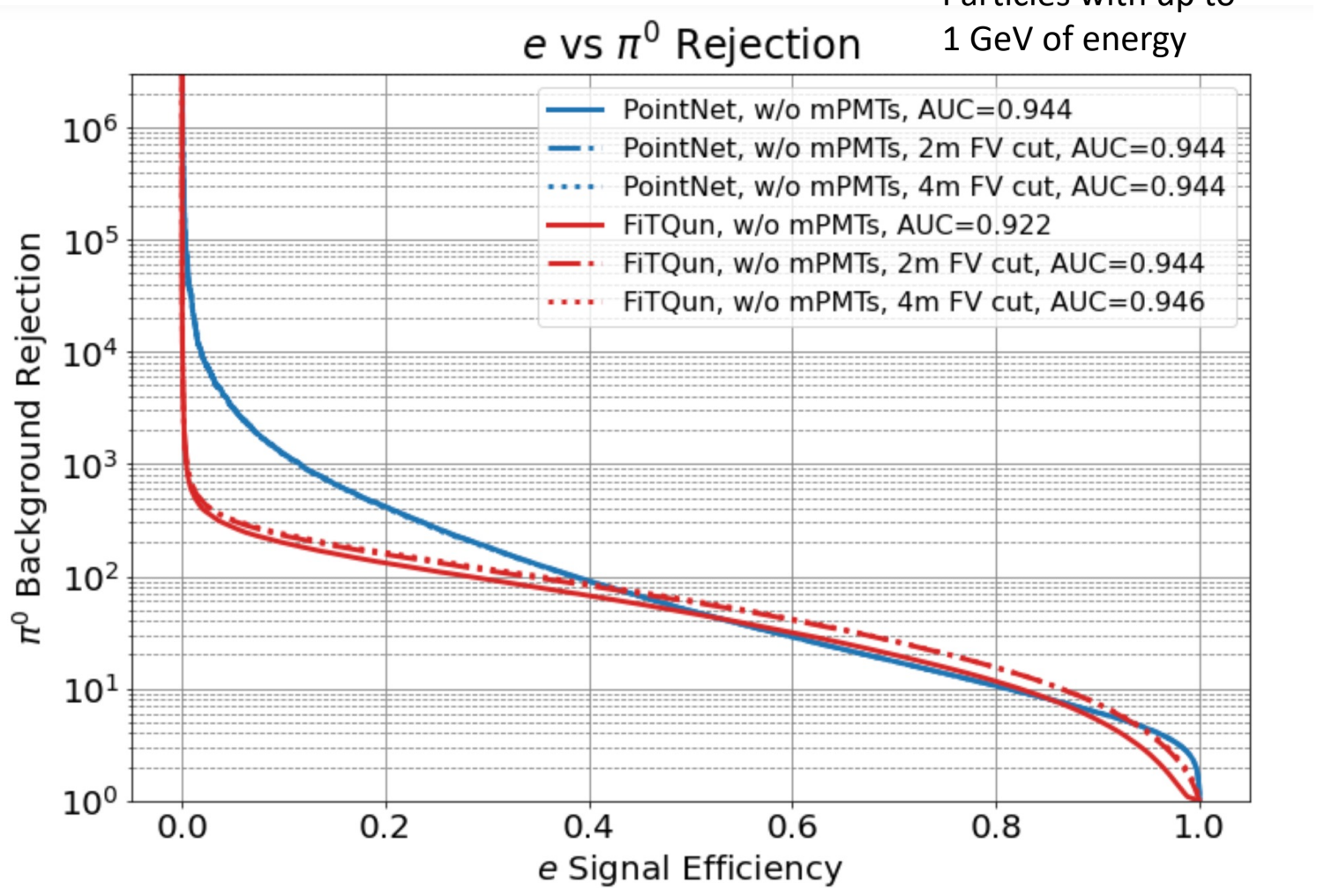
Particles with up to 1 GeV of energy



Current Results – e/pi0 Separation

- There are 2 fiTQun variables that help to classify pi0, pi0 NLL and reconstructed pi0 mass
- Only using pi0 NLLs from fiTQun in this plot, working on having both variables

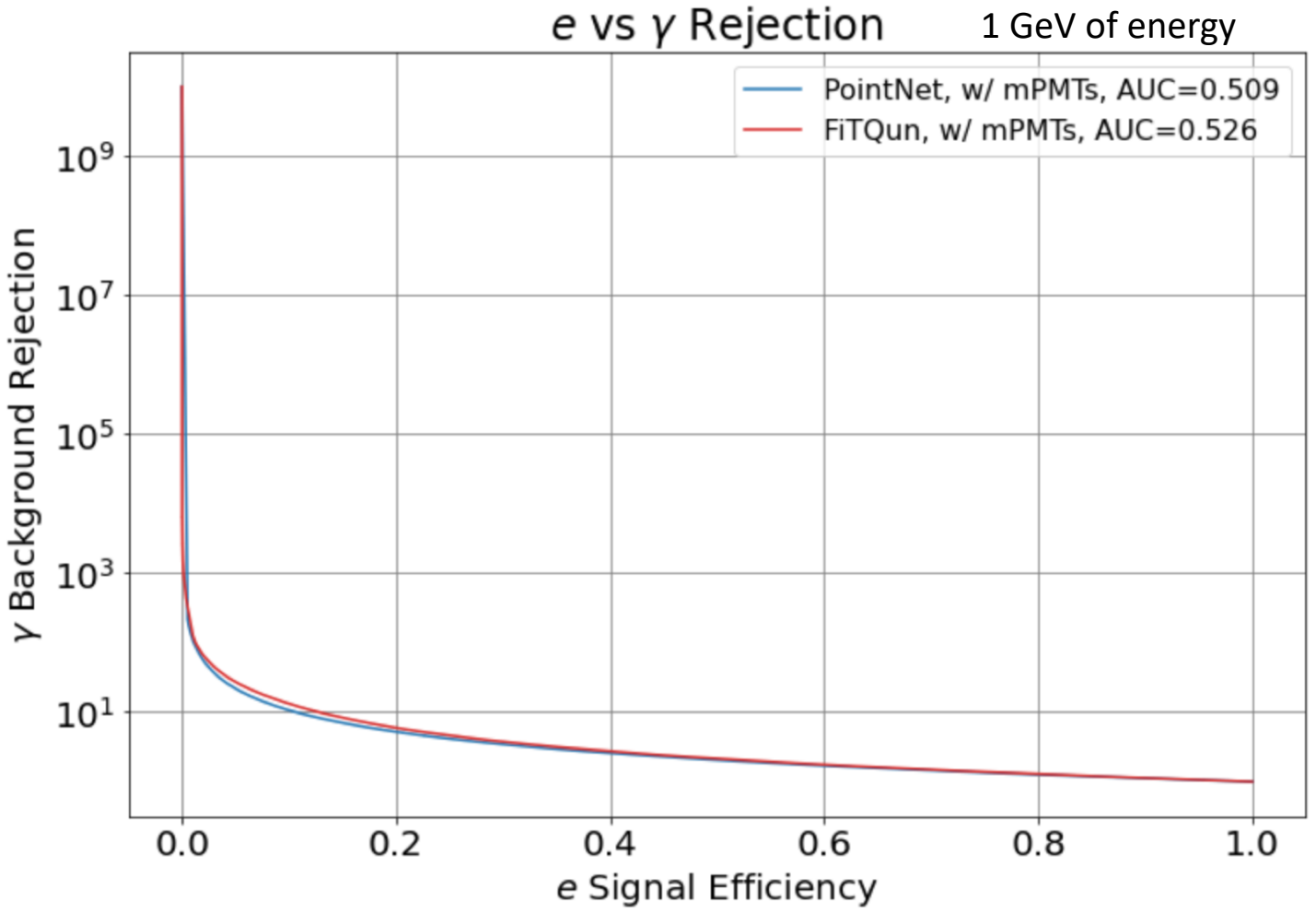
Particles with up to 1 GeV of energy



Current Results – e/gamma Separation

- Both PointNet and fiTQun has performed extremely poorly (AUC = 0.5 basically means it's a coin flipping)
- Currently investigating the reason

Particles with up to 1 GeV of energy



Conclusion

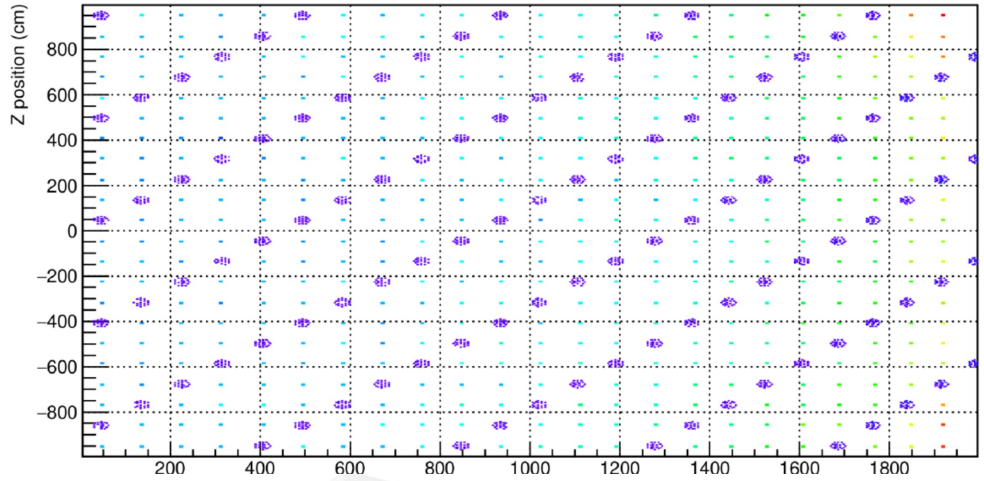
- Using a novel CNN, PointNet, for particle identification and comparing the results to fiTQun
- The current version of PointNet is performing on par with fiTQun for e/mu and e/pi0 classification without mPMT signals
- Currently working on incorporating mPMT signals into the training data; apply PointNet on SK simulation and check its efficacy with real data; introducing kinematics reconstruction using PointNet into HK hybrid geometry

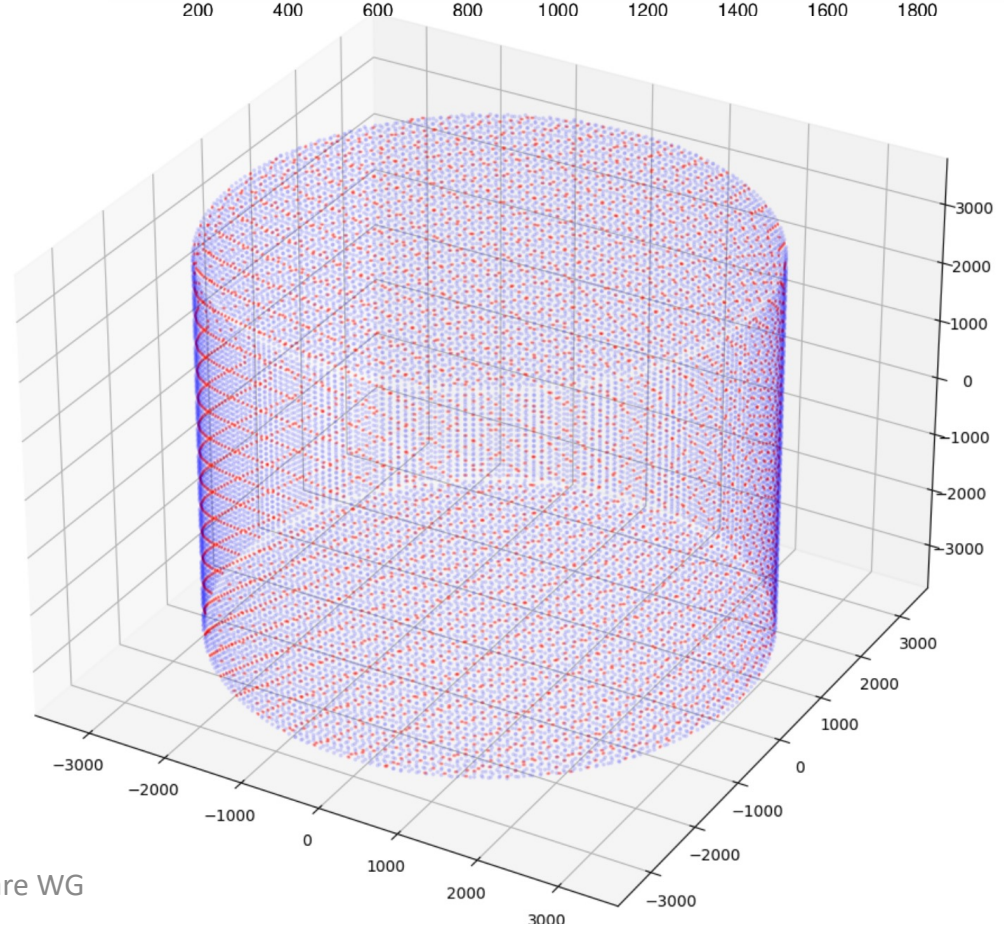
KING'S College LONDON

Backup Slides

Hyper-K Data Simulation Setup

- Example geometry on the right
- Way too many mPMTs in the simulation – masking them off
- Simulated 4 types of particles – e/mu/π^0/γ
- 3 million events each, approximately 0-1 GeV of energy, uniformly distributed and going in isotropic directions





FitQun Timing for HK Hybrid

UNIT: seconds		Mean	RMS
E	Total Time	139.97	54.82
	1R Fit	81.85	30.98
	Pi0 Fit	54.33	30.04
Mu	Total Time	142.72	67.77
	1R Fit	80.89	40.66
	Pi0 Fit	59.18	38.49
Pi0	Total Time	163.38	64.11
	1R Fit	83.80	31.17
	Pi0 Fit	75.09	40.93
Gamma	Total Time	183.81	70.19
	1R Fit	108.42	38.25
	Pi0 Fit	71.53	38.71

KING'S College LONDON