Using Machine Learning to Identify Neutron Captures in Gd Loaded Water Cherenkov Detectors

Matthew Stubbs
University of Winnipeg
Feb9, 2021

Outline

1) Introduction
- Water Cherenkov Detectors
- Neutron Tagging
- Why ML?

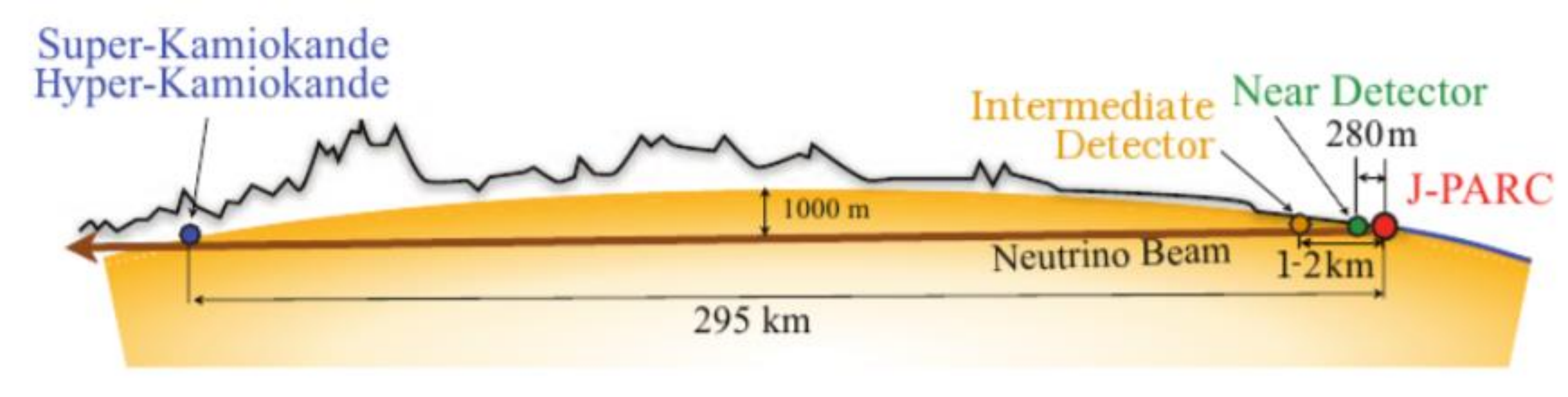
2) Likelihood Analysis
- Statistical Baseline

3) Machine Learning Methods
- MLP
- XGBoost
- Feature Engineering

4) Deep Learning: Graph Neural Networks
- GCN, AGNN, SGConv

5) Conclusions / Future Work

Overview: IWCD

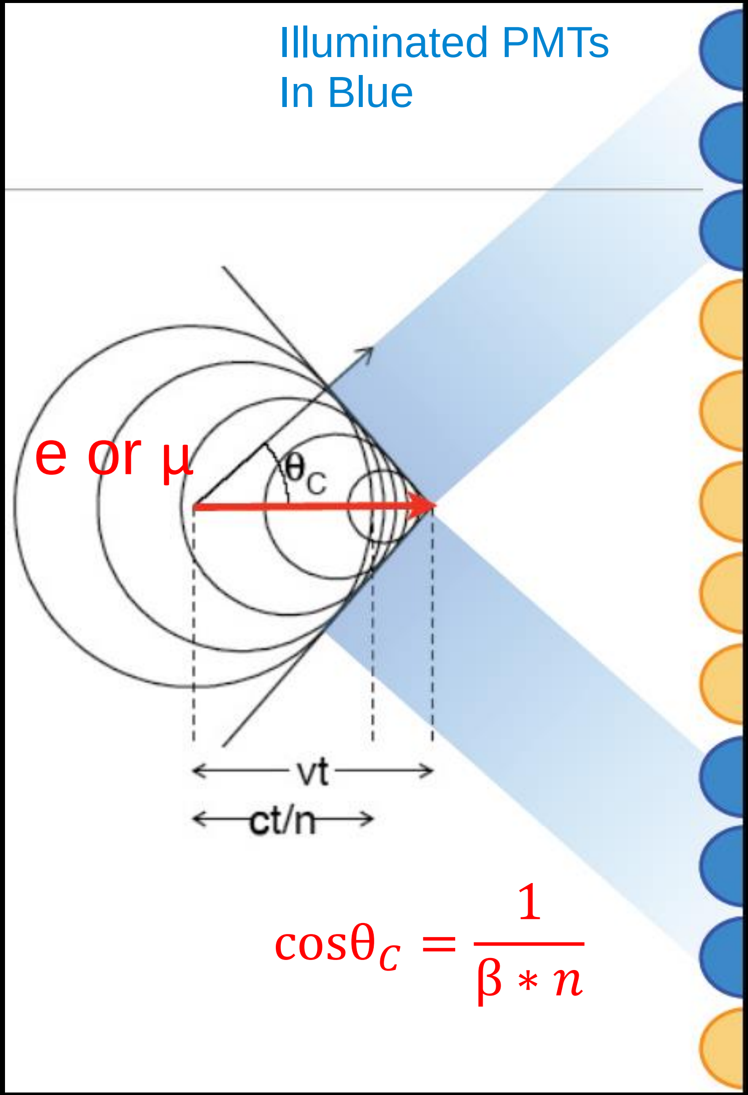


- Currently: T2K neutrino beam produced at J-PARC in Tokai, Japan and measured by Super-Kamiokande in Kamioka.

- Hyper-Kamiokande: next generation WC detector, to be far detector for upgraded J-PARC beam. Order of magnitude larger fiducial mass, order of magnitude more events

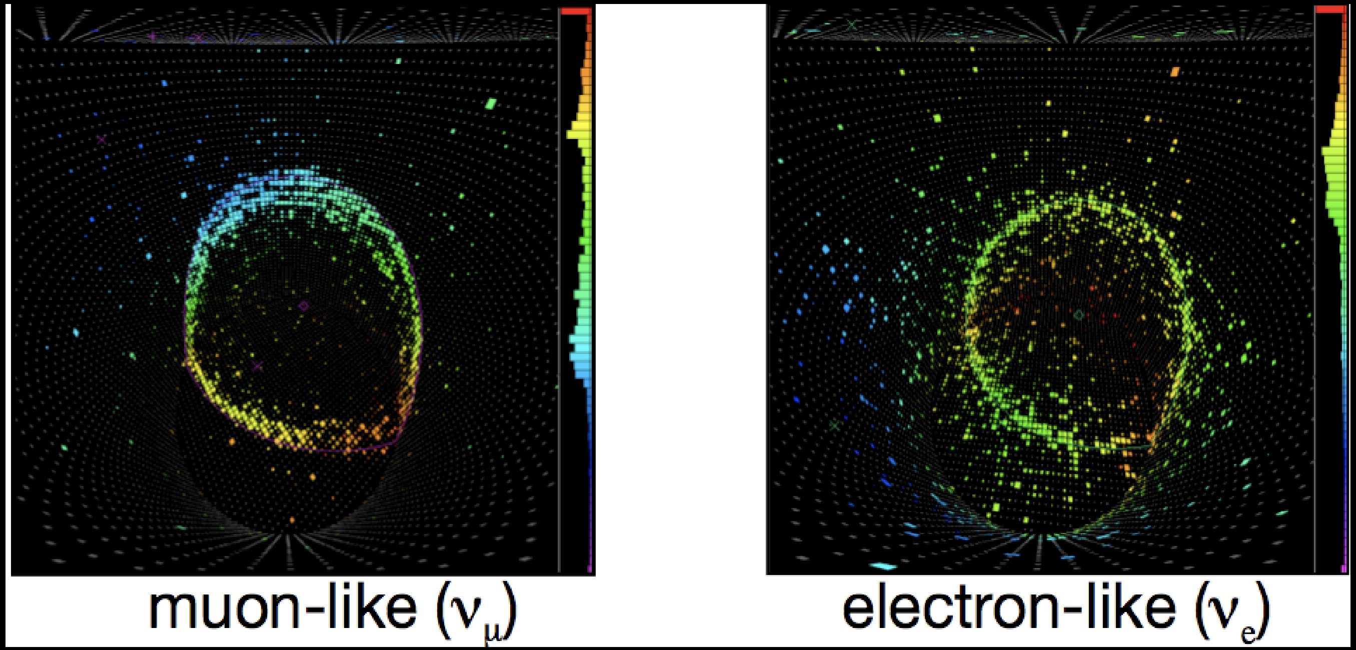
- Intermediate Water Cherenkov Detector: **proposed** near detector at Tokai (1-2km from J-PARC) for Hyper-Kamiokande. Goal, reduce systematic uncertainties:
  - Reduce flux uncertainty
  - By spanning off-axis angles 1 to 4 degrees, constrains relationship between observed lepton kinematics and incident neutrino energy.

Neutrino detection in Water Cerenkov Detectors



$$\cos\theta_C = \frac{1}{\beta * n}$$

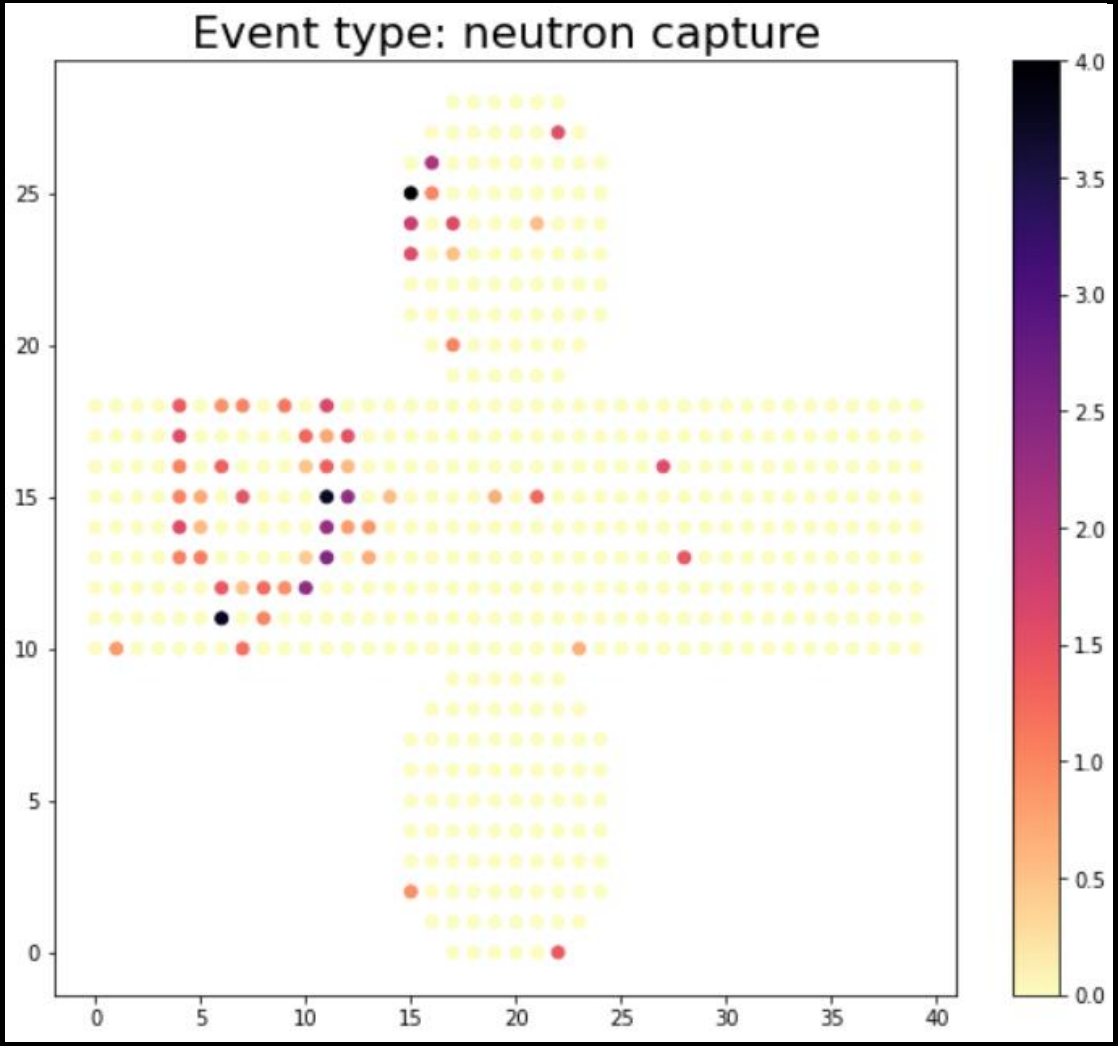
- Neutrino interaction produces a charged particle
- Cerenkov light (EM version of sonic boom) when charged particle moves > speed of light in water
- Cone of blue light at angle related to v/c and material index of refraction

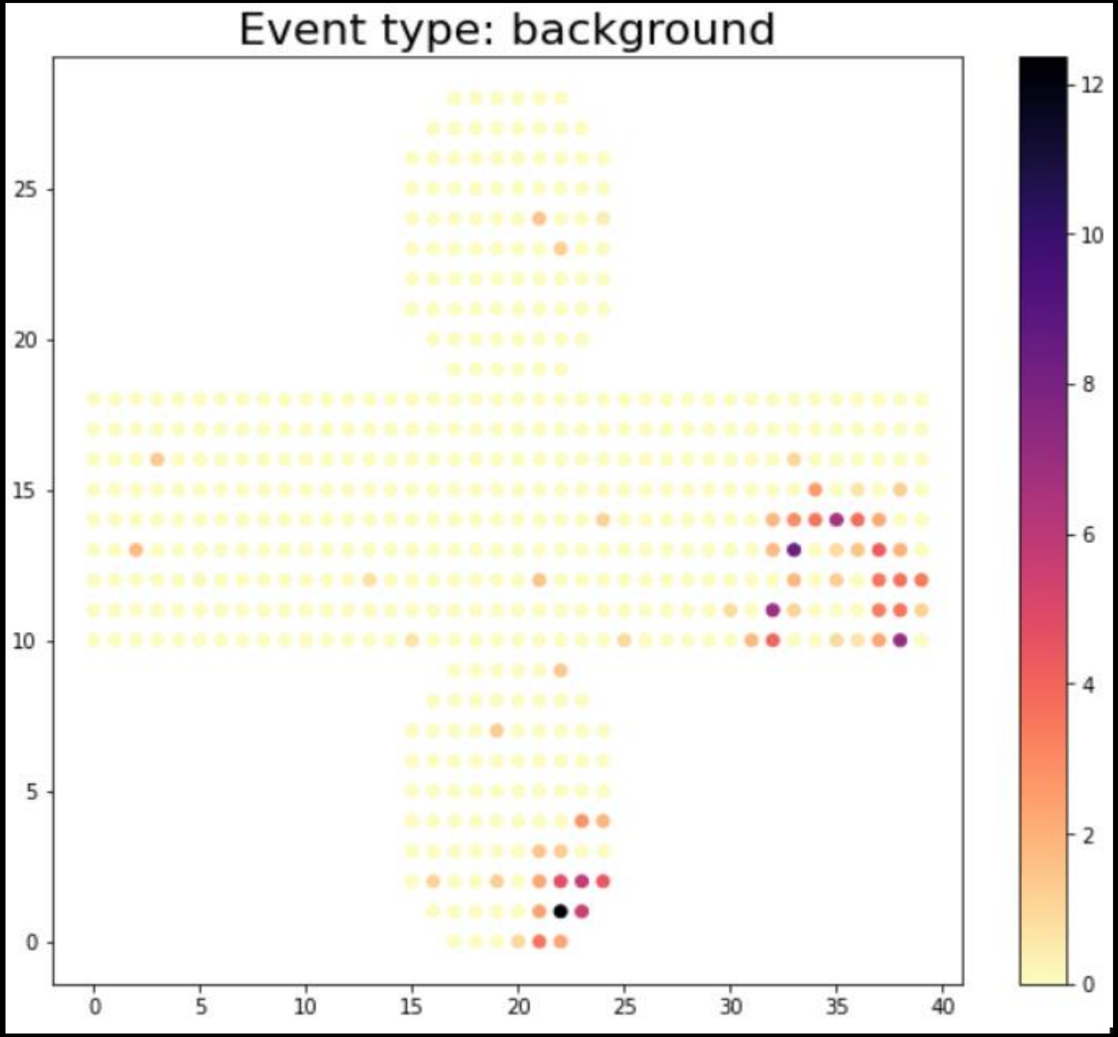


Dataset

- Data generated using WCSim software for IWCD geometry
- ~2 million events (half neutron capture, half background). For each event, up to 250 hit PMTs. For each hit PMT, 8 feature values: charge, time, 3D PMT position (x, y, z) and 3D PMT orientation (x, y, z).
- Simulations include 0.1% Gadolinium doping in simulated water (eq. to 0.2% Gd sulphate)
  - Gadolinium doping: 8MeV gammas, higher neutron capture cross section than hydrogen nuclei (2.2MeV gammas, smaller ncapture cross section)
  - about 88% of neutrons will capture on Gadolinium, remainder capture on hydrogen
- background source: **electron energy distribution**
  - uniform energy distribution from 0 to 20 MeV, events distributed uniformly in tank
  - Background electron rings look similar to neutron capture rings. Let's see that…

IWCD geometry mapped to 2d event display. Shows charge over PMT modules

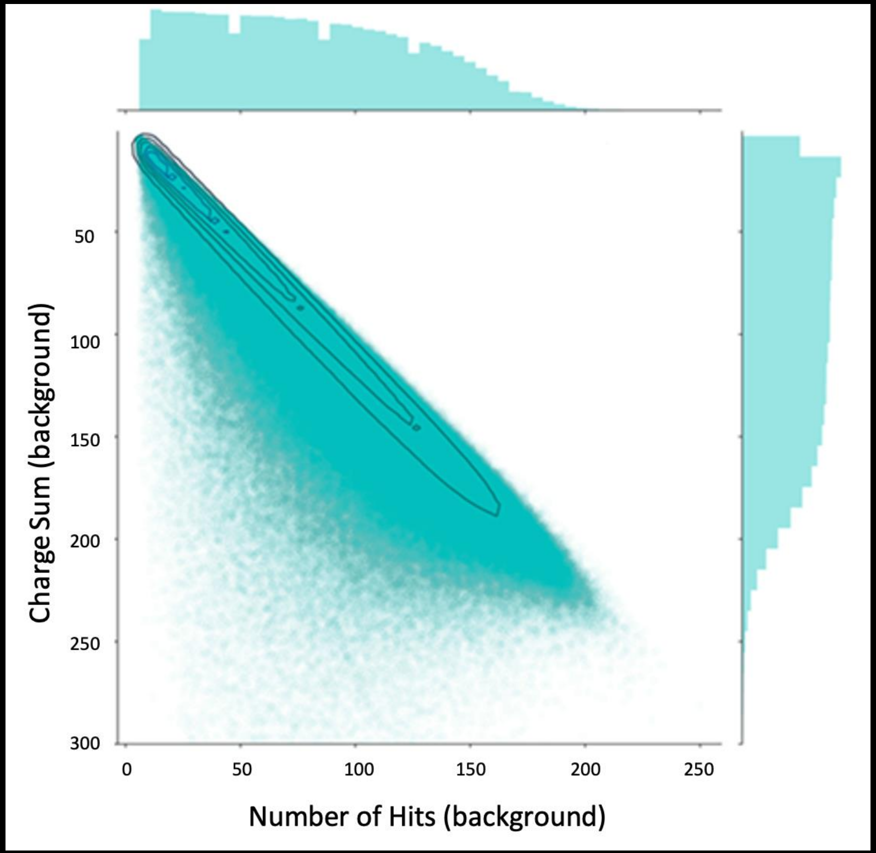


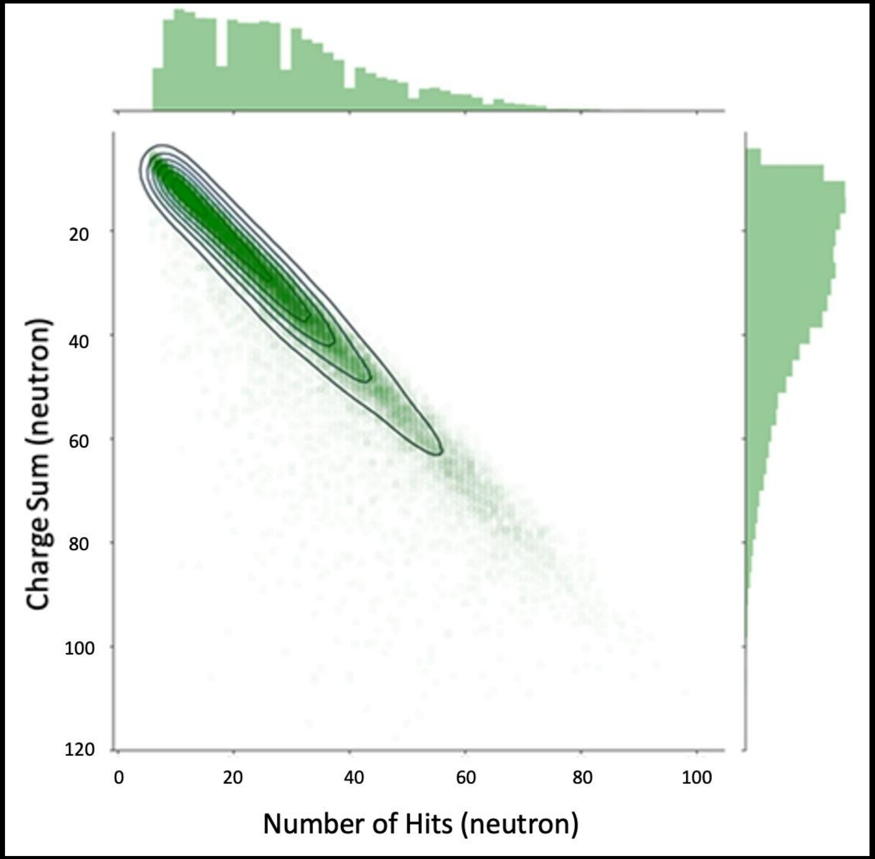


Task: (using ML), can we improve the differentiation of Neutron captures from radioactive background electrons?

Likelihood Analysis

Purpose: non-ML baseline. Good separation from number of hits, charge sums alone
Likelihood test: fit nhits, charges distributions with smoothed KDEs, classify event based on highest probability





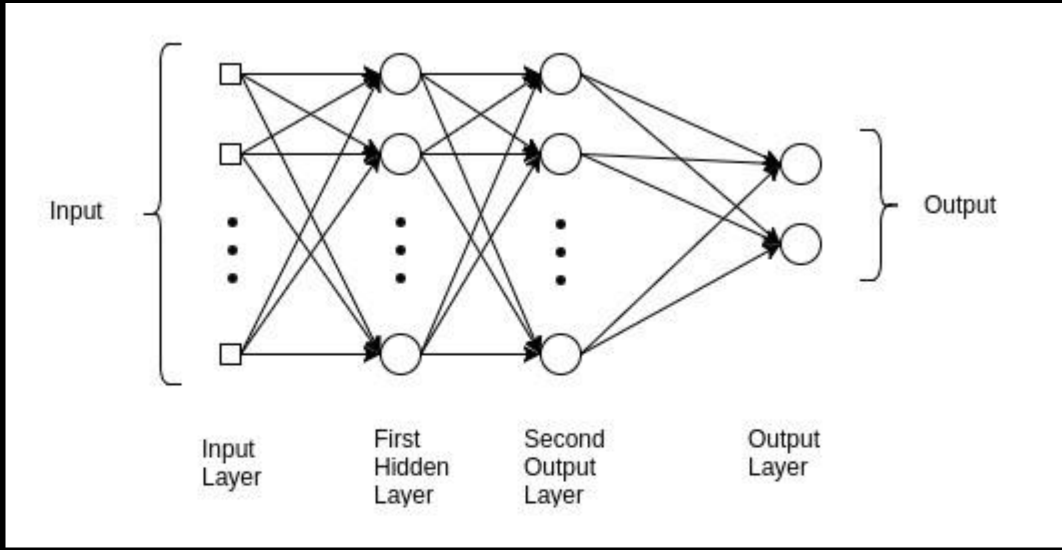
Statistical Baseline:

~79%

MLP/GBM

Multi-Layer Perceptron (MLP)

- "Standard Neural Network"
- Series of layers with various numbers of compute units ("neurons"), each receiving input from previous layer, computing output
- Loss computed at output layer, backpropagated back through network, weights adjusted using optimization algorithm



Gradient Boosting Machines (GBM)

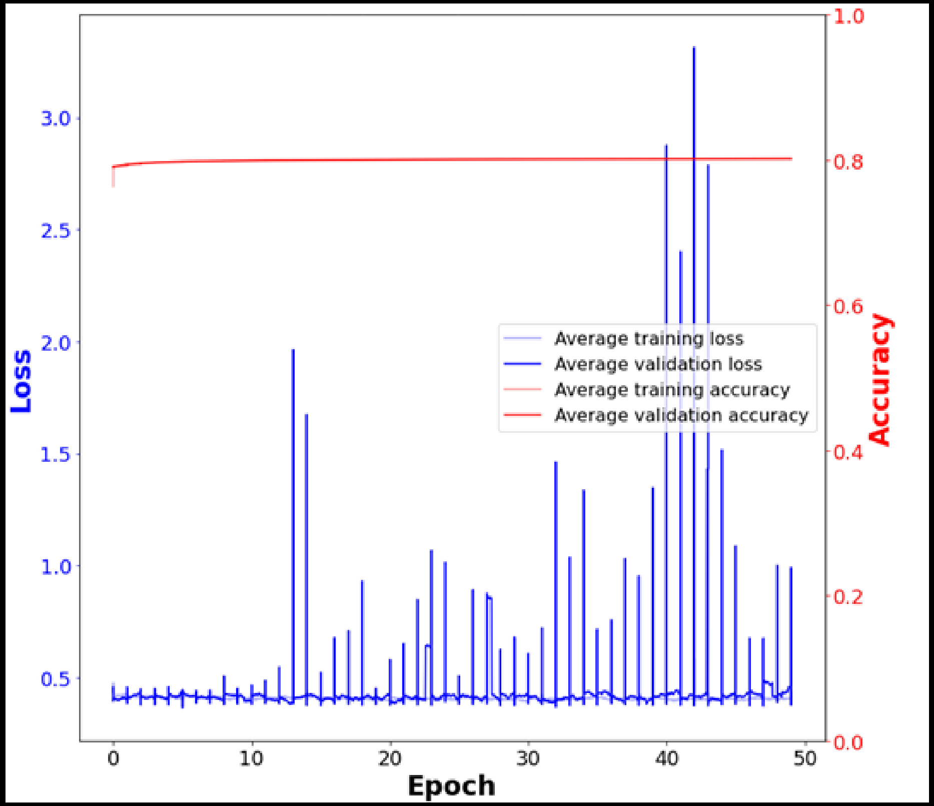
- **Ensemble method**: final model constructed iteratively based on many individual models (weak learners). Idea: combination of weak models in an ensemble leads to improved result
- Decision trees are most common type of weak model in gradient boosting machines
- Gradient used to minimize loss function
- XGBoost & LightGBM are two popular GBM methods
  - XGBoost looks at individual features and makes branching decisions based on what yields the highest information gain

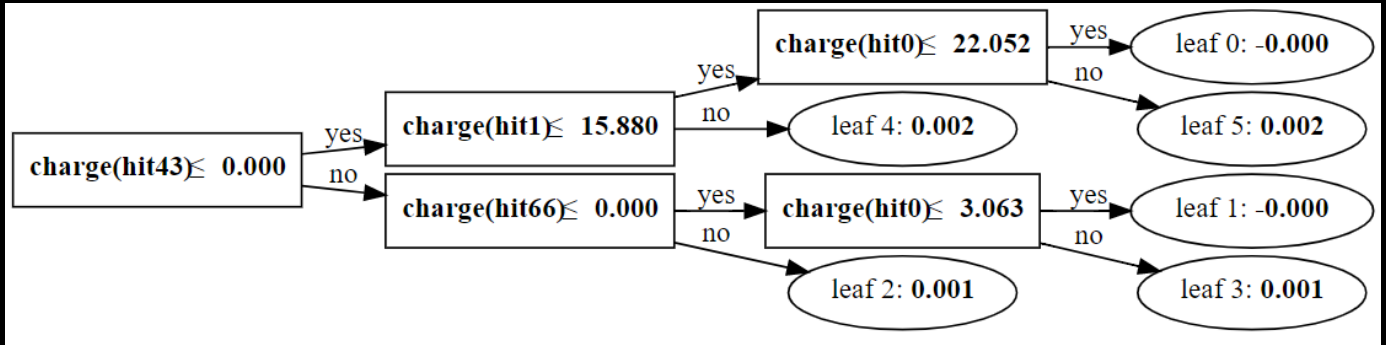
MLP/GBM

Original features (charge, time, PMT position, PMT orientation)

- Performance best with all features
- Both methods: close to 1% accuracy improvement over likelihood baseline
- Difficult to improve from baseline performance!
- Both models find it more difficult to detect background

MLP training progression





XGBoost sample tree

- Adam optimizer
- 50 epochs
- Learning rate = 0.01
- 5 layers (8 -> 64 -> 128 -> 64 ->12 -> 2)

Feature Engineering

Try engineering features more informative for a network to learn from

Largely inspired by:

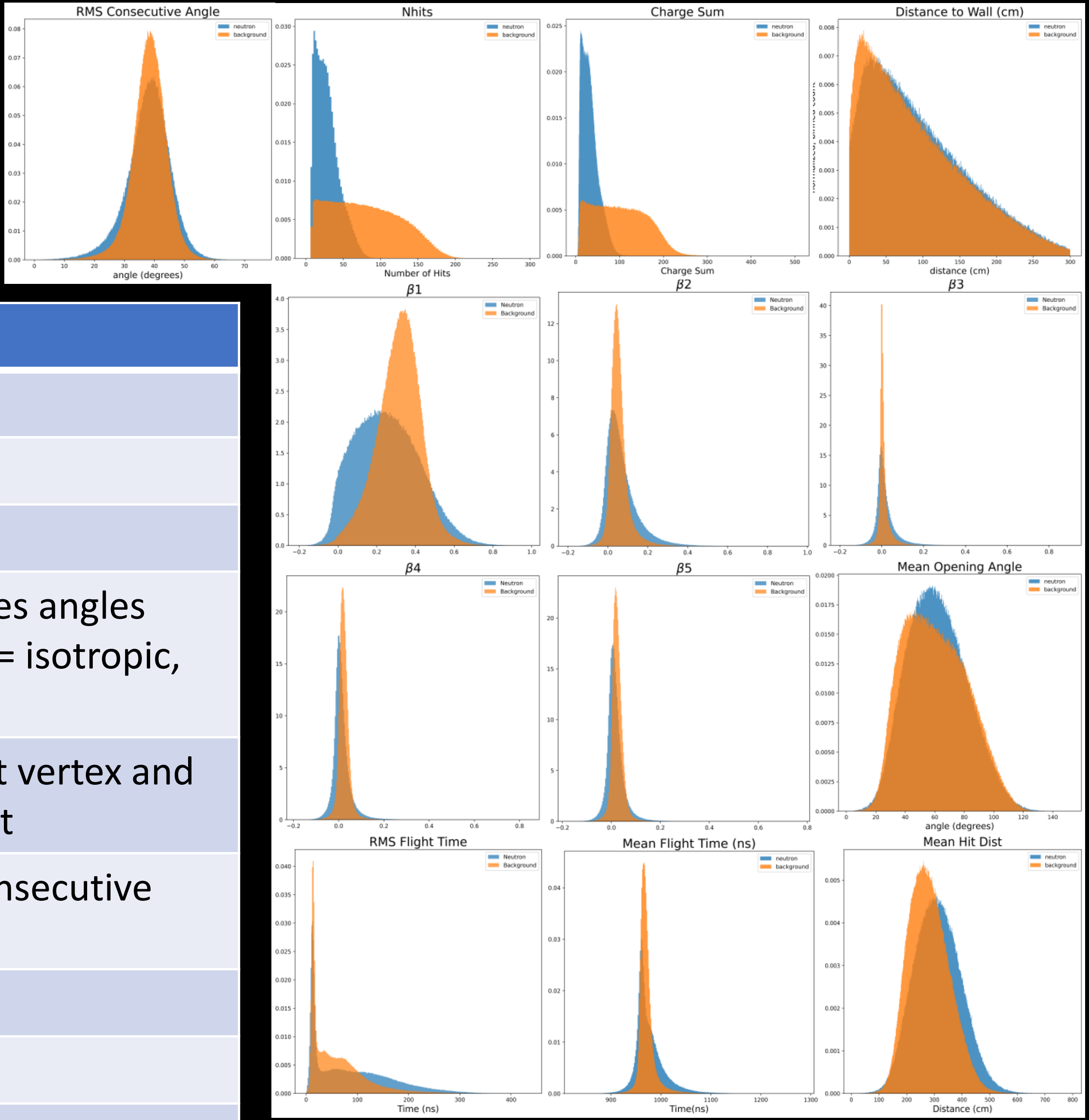
- Abe, K., Haga, Y., Hayato, Y., Ikeda, et al. (2013). Neutron Tagging following Atmospheric Neutrino Events in a Water Cherenkov Detector. Prog. Theor. Exp. Phys. PTEP (and similar studies)

What separability is there in the dataset?

- Different number of hits and charge
- Isotropy: Different amounts of scattering and reflections
- Time of Flight: longer time due to neutron gamma pair production at 180°
    - Different average distance between hits in an event?
    - Average angle between event vertex and sum of hit PMT positions?
    - Background clustering? Distance from detector wall?

Engineered Features

Feature	Description
Nhits	Number of hits per event
Charge Sum	Sum of charges in an event
DWall	Distance from event vertex to wall
β1 – β5 $\beta_l = \langle P_l(\cos\theta_{ik})\rangle_{i\neq k}$	Measure of event isotropy. Computes angles between all pairs of hits in event. 0 = isotropic, higher numbers -> anisotropic
Opening Angle (µ)	vector sum of angles between event vertex and hit position for every hit in the event
Hit Dist (µ)*	Mean average distance between consecutive hits in an event
Flight Time (RMS)	RMS of time of flight per event
Flight Time (µ)	Mean time of flight per event
RMS Consecutive Angle	RMS of angles between consecutive hits in event



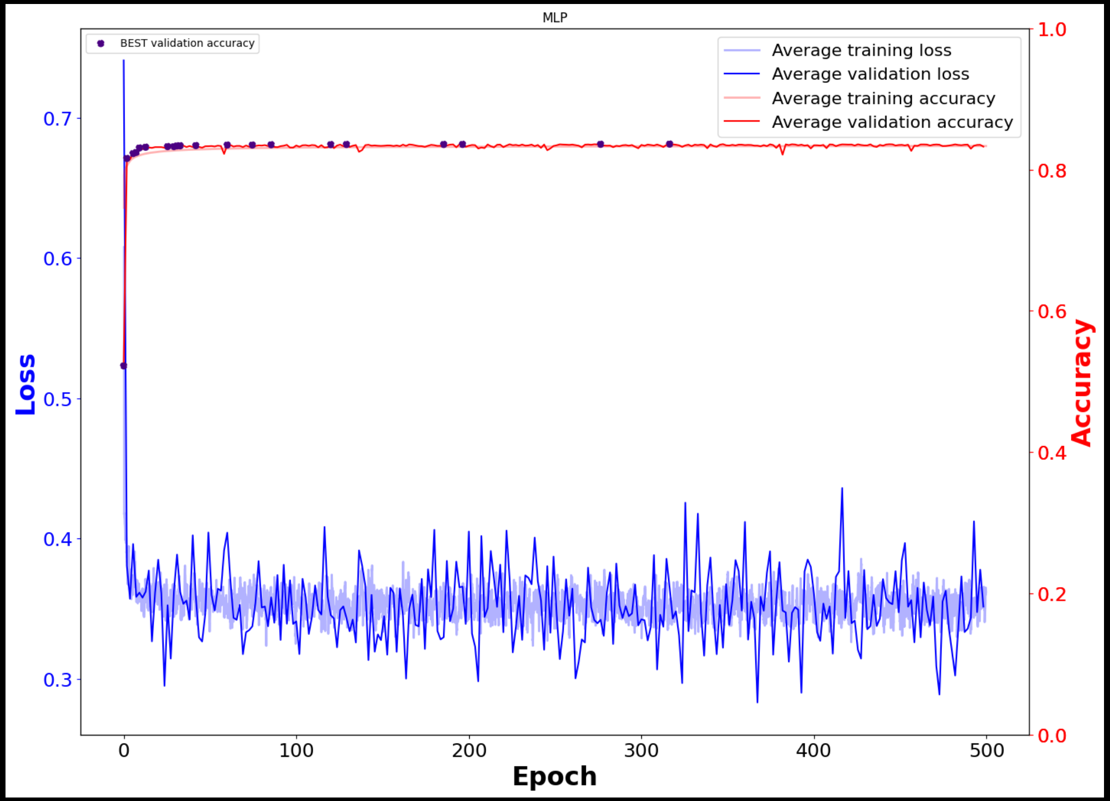
* new

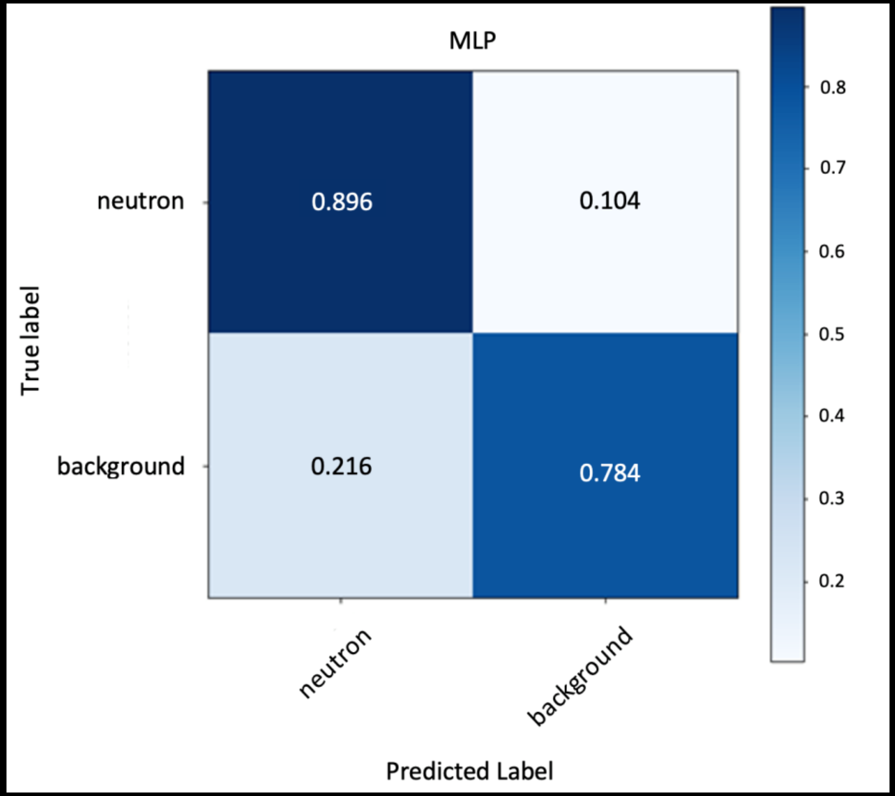
ML Methods

Feature Engineering: Aggregate Features

- Activation function: ReLU
- Classifier: SoftMax
- Layers/Units: 12 -> 36 -> 64 -> 64 -> 24 -> 10 -> 2
- epochs 500
- lr 0.003 (no lr decay)
- Batch size: 512

	MLP (all features)
Test accuracy	**83.7**
Best Dev Accuracy	**83.8**
ROC AUC (test dataset)	**0.915**





ML Methods

Graph Networks

Can model particle event as a graph:

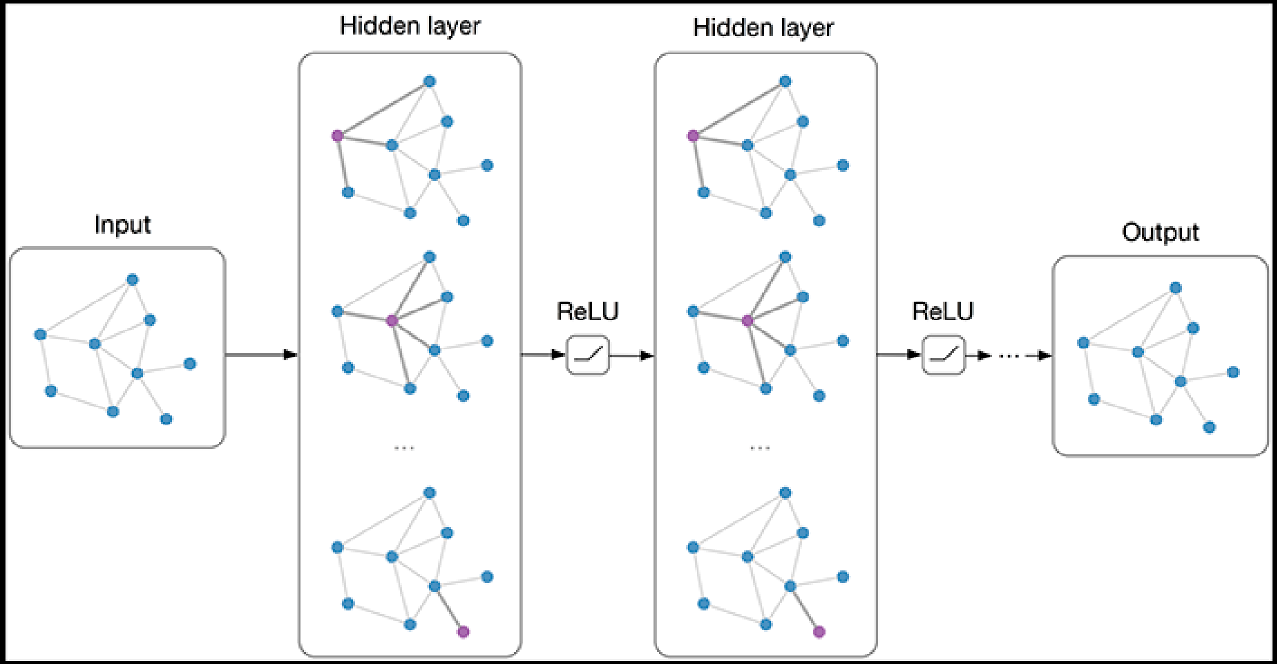
- Nodes -> hit PMTs
- Edges -> connections between nodes

Can be used for:

- **node classification**
- **Graph classification** (apply pooling layer at output) – this is our case

GCN (Graph Convolution Network) example:

- Step 1: Each node aggregates neighbourhood feature representations ('smoothing step')
- Step 2: Each node updates activations, passing through MLP network layer
  - Rinse and repeat



Graph Networks: Models (fully connected)

	AGNN	GCN	SG
Test accuracy	79.7	79.5	79.5
Best Validation Accuracy	79.9	79.7	79.8
ROC AUC (test dataset)	0.855	0.851	0.853

Graph models (Dataset 1):

AGNNConv: Attention-based Graph Neural Network for Semi-supervised Learning
https://arxiv.org/abs/1803.03735

GCNConv: SEMI-SUPERVISED CLASSIFICATION WITH GRAPH CONVOLUTIONAL NETWORKS
https://arxiv.org/pdf/1609.02907.pdf

SGConv: Simplifying Graph Convolutional Networks
https://arxiv.org/pdf/1902.07153.pdf

Graph Networks: Graph Construction Methods

GCN (fully connected, distance weighted)

Test accuracy	79.7
Best Validation Accuracy	79.9
ROC AUC (test dataset)	0.863

GCN (k nearest neighbours graph)

	K = 5	K = 15	K = 19	K = 23	K = 100
Test accuracy	78.2	77.0	78.4	73.4	77.3
ROC AUC	0.842	0.833	0.846	0.784	0.836

Synopsis

Overall	**Next Steps**
• Model Saturation *Comparison to human-level performance* • Feature Engineering • Graph Networks • "Dark Noise" application	• Super Kamiokande data • Comparison to other architectures • Other Physics

Thank you

Any questions?

Bibliography

- Abe, K., Haga, Y., Hayato, Y., Ikeda, et al. (2013). Neutron Tagging following Atmospheric Neutrino Events in a Water Cherenkov Detector. Prog. Theor. Exp. Phys. PTEP
- Dunmore, Jessica. “The Separation of CC and NC Events in the Sudbury Neutrino Observatory.” *University of Oxford Thesis*, 2004
- Sekiya, Hiroyuki. "The Super Kamionade Gadolinium Project." Journal of Physics: Conference Series. Vol. 1342. No. 1. IOP Publishing, 2020.
- Prouse, Nick. “Neutron Tagging in an Intermediate Water Cherenkov Detector for the J-PARC Neutrino Beam.”, 2017, idm2016.shef.ac.uk/event/1/contributions/278/attachments/237/242/iop2017.pdf
- Thekumparampil, K. K., Wang, C., Oh, S., & Li, L. J. (2018). Attention-based graph neural network for semi-supervised learning. arXiv preprint arXiv:1803.03735.
- Kipf, T. N., & Welling, M. (2016). Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907.
- Wu, F., Zhang, T., Souza Jr, A. H. D., Fifty, C., Yu, T., & Weinberger, K. Q. (2019). Simplifying graph convolutional networks. arXiv preprint arXiv:1902.07153.

Extra Slides

Gadolinium Loading

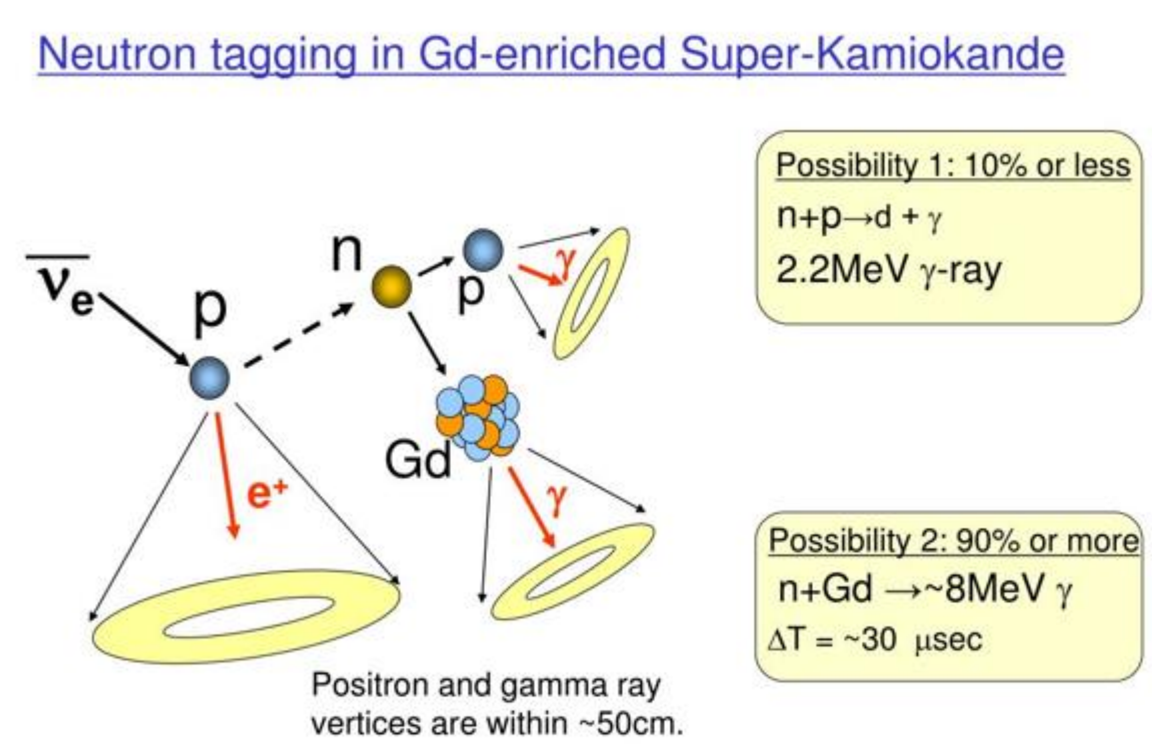
Neutrons produced through inverse beta decay

$$\bar{\nu}_e + p \rightarrow e^+ + n$$

- Neutrons scatter in water until they reach low enough ("thermal") energy. Then captured by nucleus, which becomes excited. Nucleus de-excites by producing gamma rays which produce electrons.

Advantages of Gadolinium:

- Much higher neutron capture cross-section
- Higher energy gamma rays (8MeV > 2.2MeV)
- More neutrons detectable



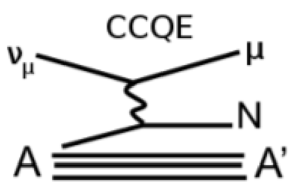
Vagins, M., Ishino, H., & Collaboration, S. K. (2012). GADZOOKS!. Phys. Rev. Lett, 108, 052505.

Neutron Tagging

1) Distinguishing neutrino/antineutrino events

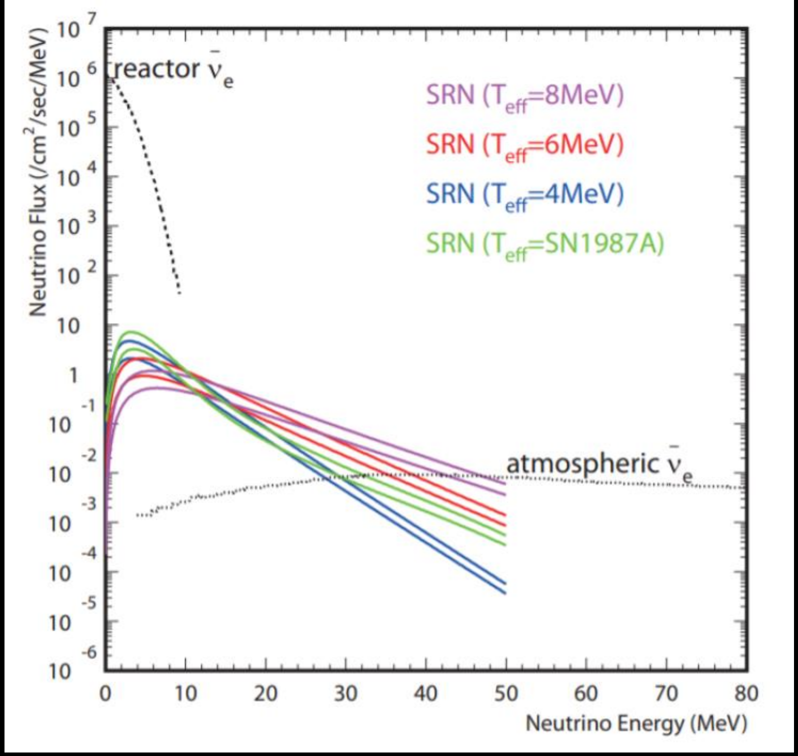
Charged Current Quasielastic (CCQE):

ν_μ :	$\nu_\mu + n \rightarrow \mu^- + p$	0 neutrons	
$\bar{\nu}_\mu$:	$\bar{\nu}_\mu + p \rightarrow \mu^+ + n$	1 neutron	



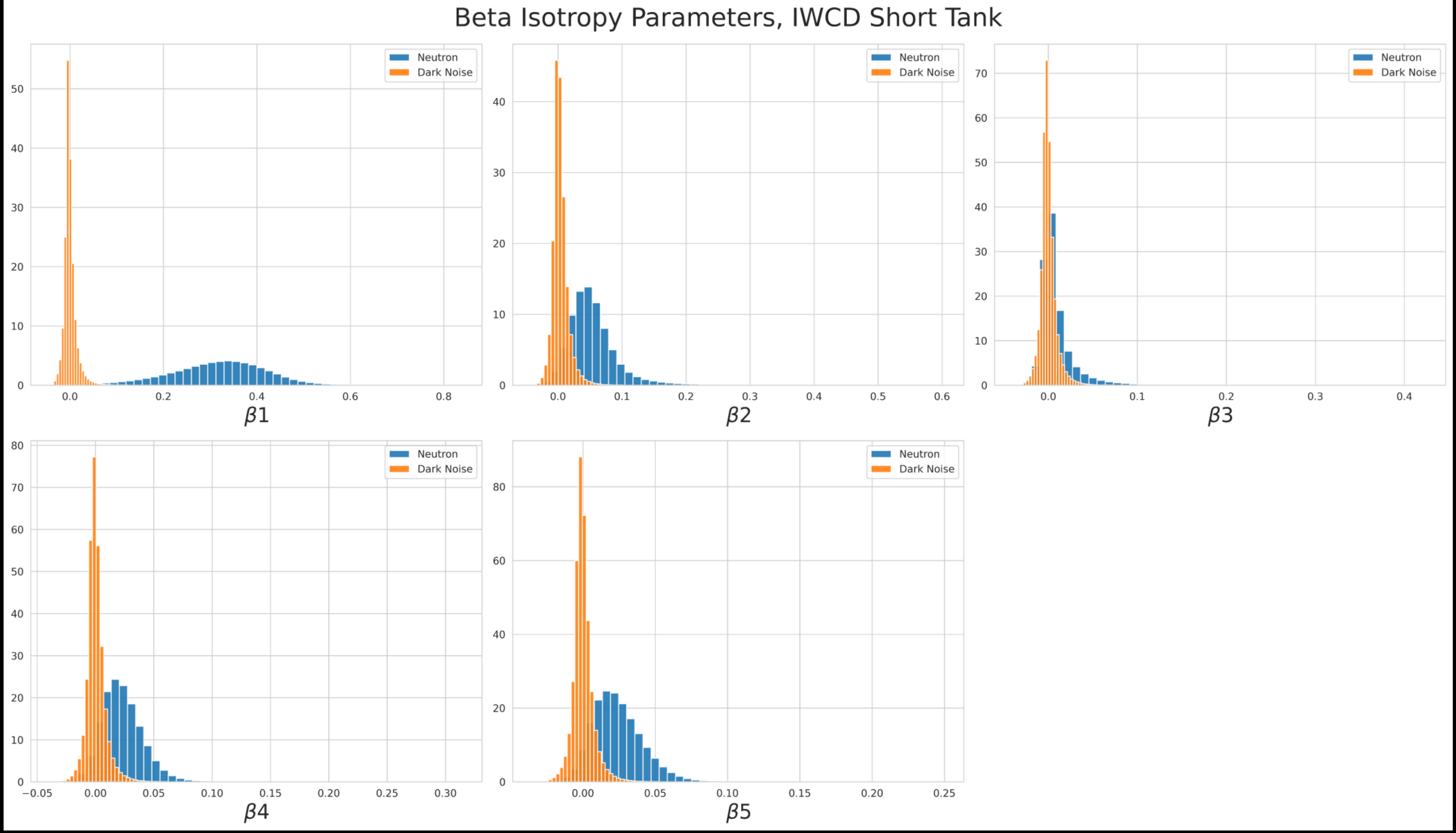
2) Diffuse supernova neutrino background (DSNB)
   - open window at low energy (0-30MeV)

   *Less applicable for IWCD, more a far detector task



Sekiya, H. (2017, April). The Super-Kamiokande Gadolinium Project. In 38th International Conference on High Energy Physics (Vol. 282, p. 982). SISSA Medialab.

Dark Noise Dataset (IWCD)



Dark Noise Dataset (IWCD)