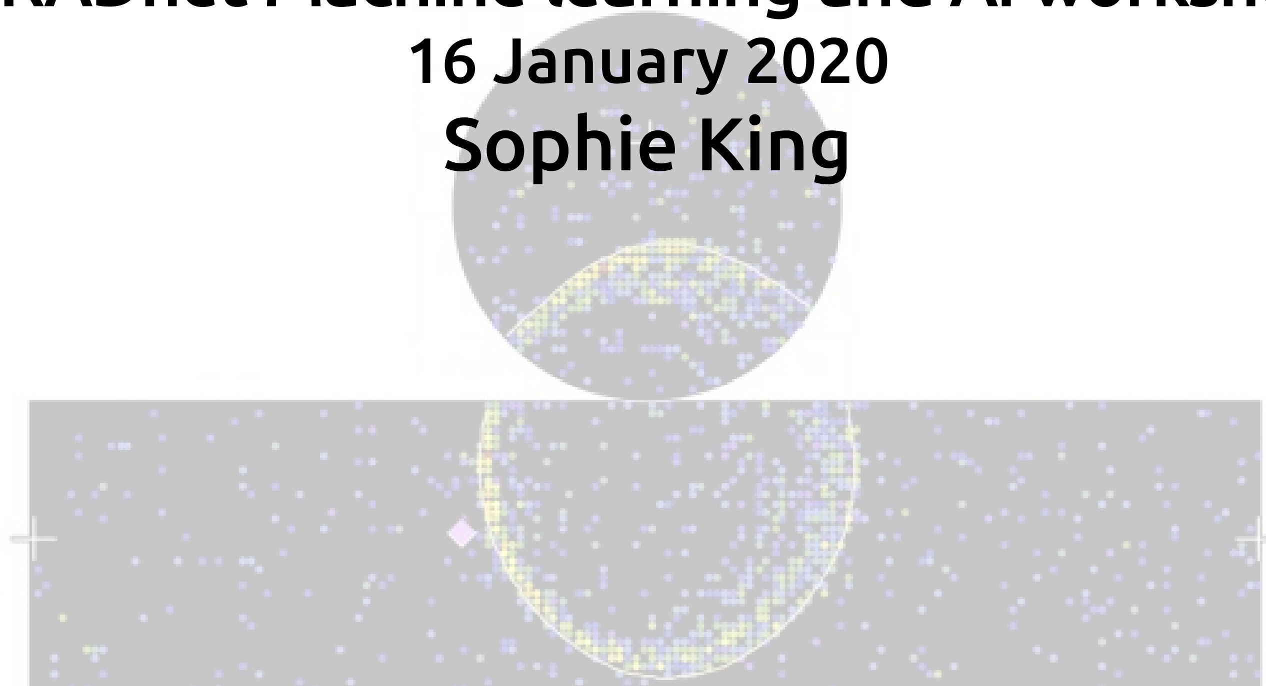
Machine learning applied to neutrino physics

GRADnet Machine learning and AI workshop

16 January 2020

Sophie King



Hello

* Joint PhD with Queen Mary and University of Southampton
  → Leptogenesis
  → T2K, Neutrino cross sections

* Post doc at Queen Mary and King's College
  → T2K and Hyper-K experiments
    → cross sections
    → selection and detector systematic development
    → computing / GRID

My interest in machine learning (ML)

* The detector I work on (T2K near detector) is being upgraded next year

* Hyper-K will introduce two new detectors

  → Lots of development for upcoming detectors
    → good time to start looking into the potential role ML can take in the new detectors

Contents

Nova: Particle identification with 2D Convolutional Neural networks
→ final product, used in their main analysis

superFGD: Voxel classification in the T2K near detector

WatCHMaL: Particle identification for the Hyper-K Intermediate Water Cherenkov Detector

→ both examples are very much in the development stage

Other examples/potentials of ML in neutrino physics

Summary

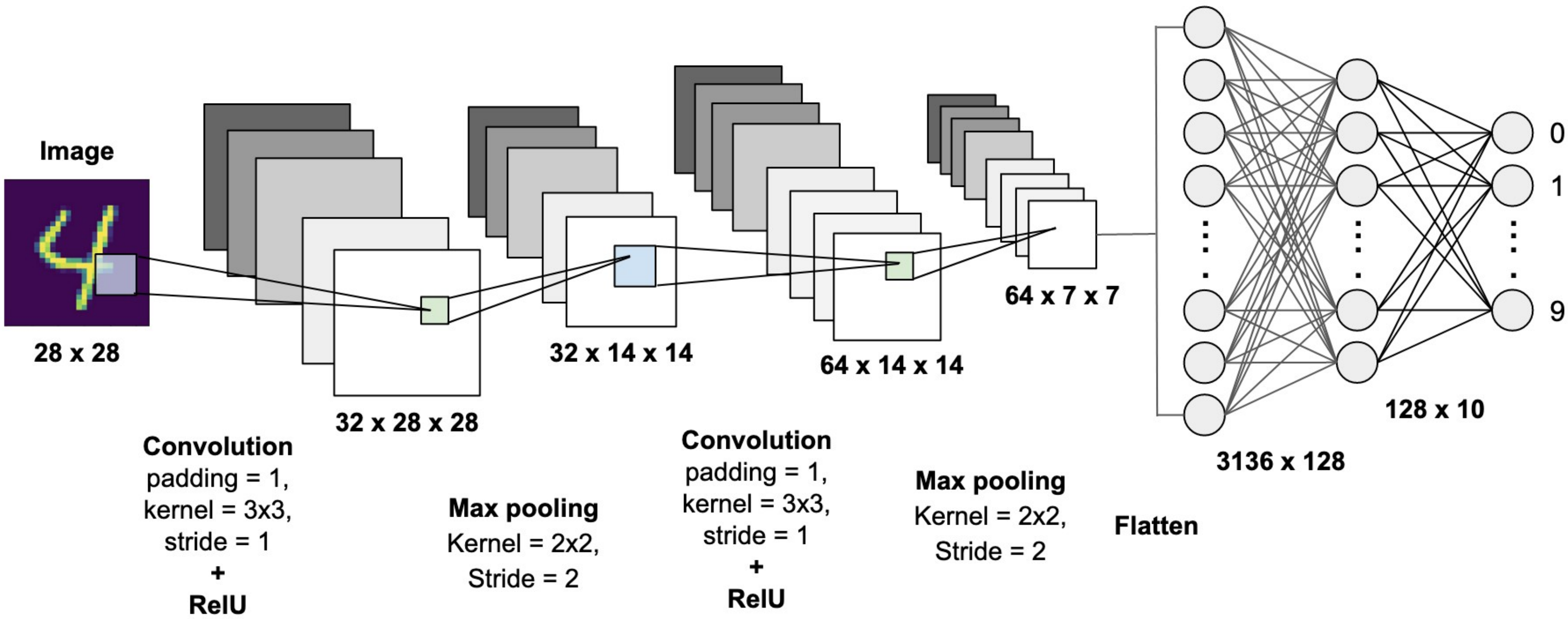
Nova:

Particle Identification with

2D Convolutional Neural networks

A. Aurisanoet al., “A Convolutional Neural Network Neutrino Event Classifier”, arXiv:1604.01444v3

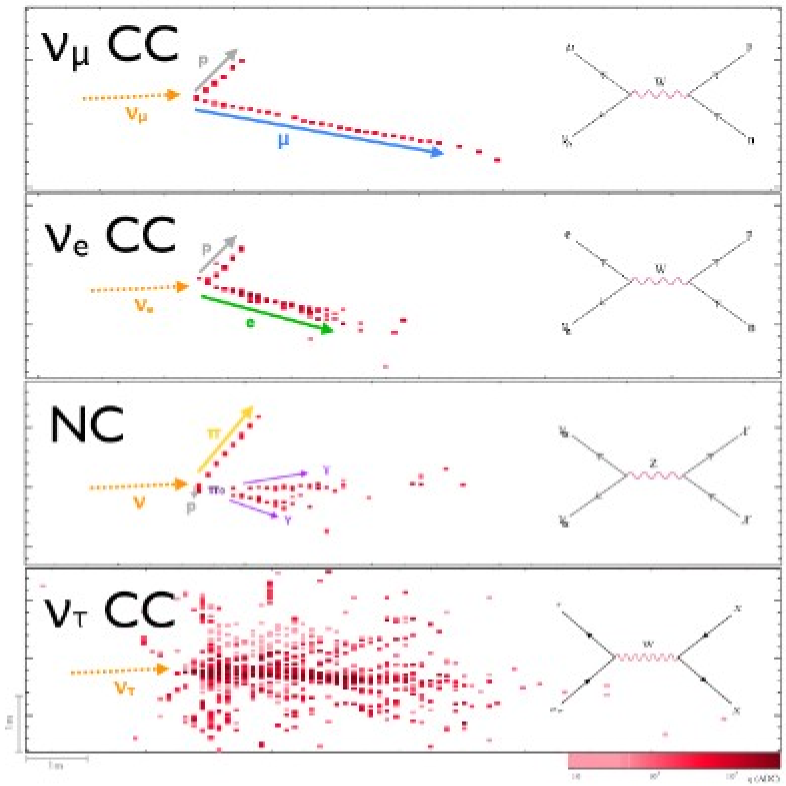
Convolutional neural networks

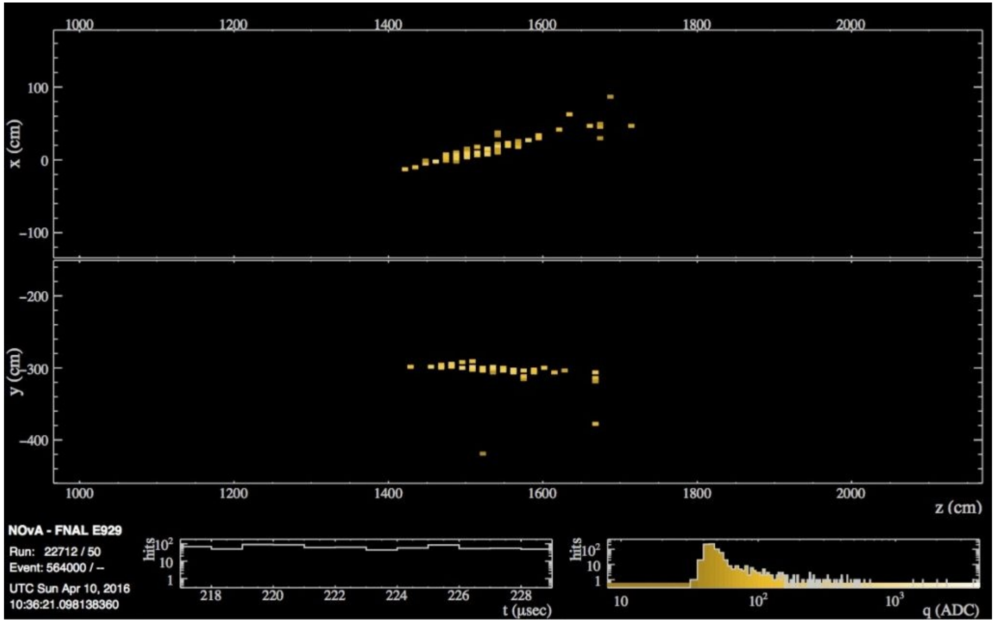


Graphic: https://towardsdatascience.com

Nova

Fine grained liquid scintillator
neutrino detector – 2D planes



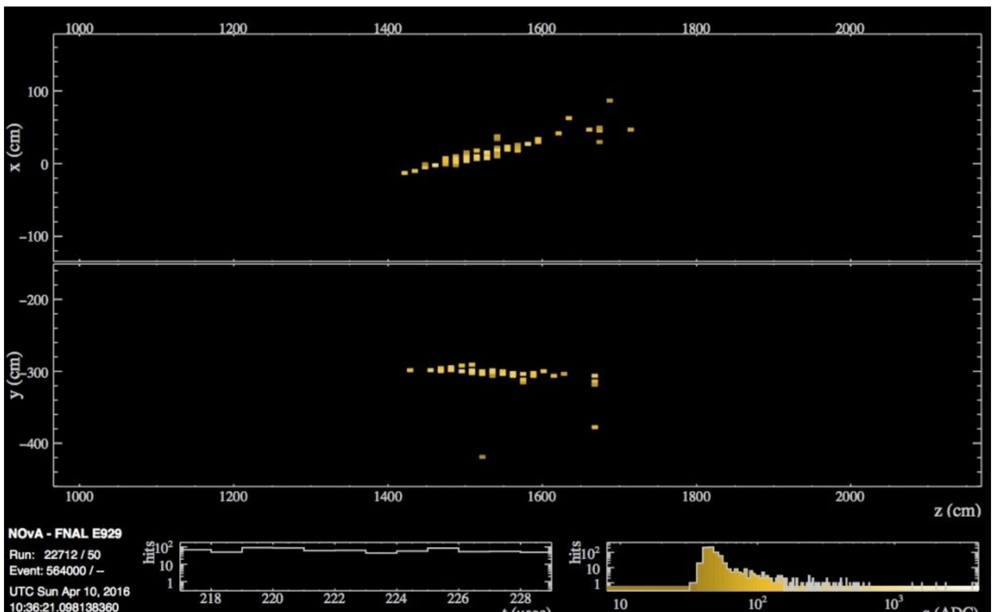


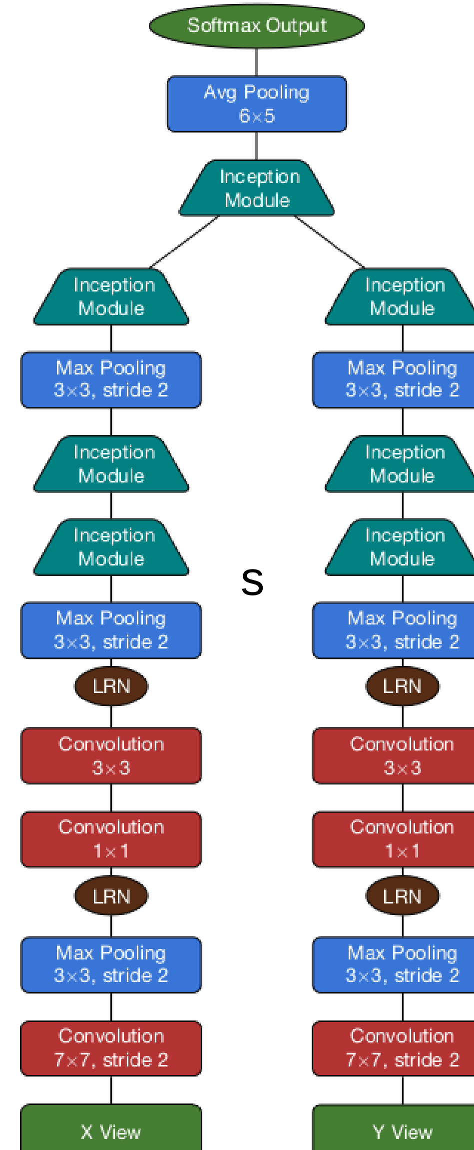
A. Aurisanoet al., “A Convolutional Neural Network Neutrino Event Classifier”, arXiv:1604.01444v3

Nova

First neutrino oscillation experiment to fully embrace machine learning techniques!

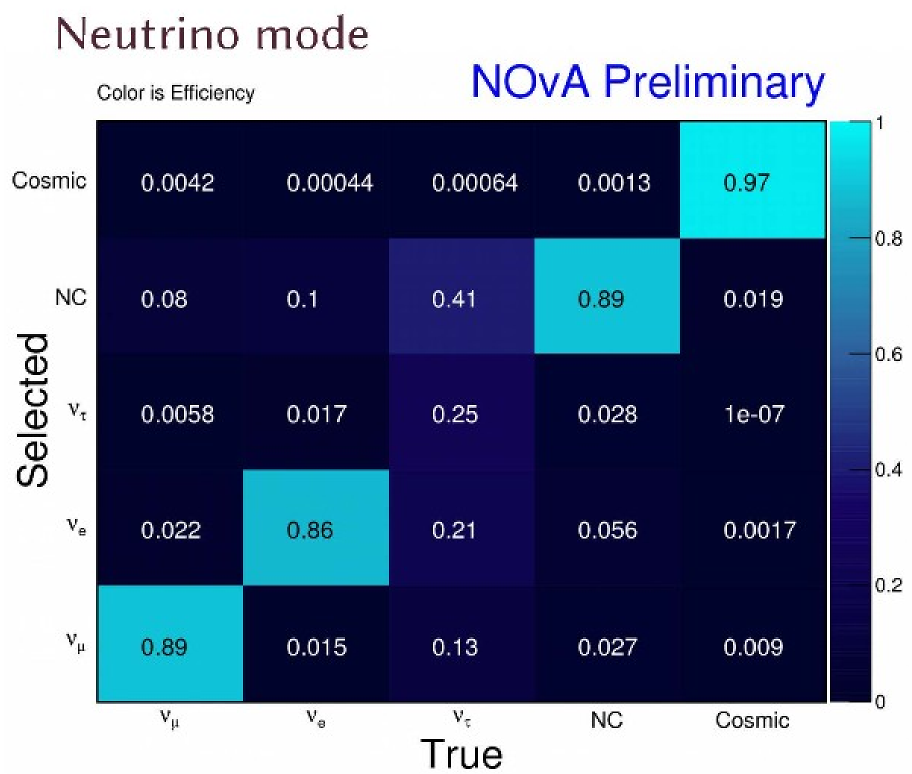
2D grid readout is perfect for CNN

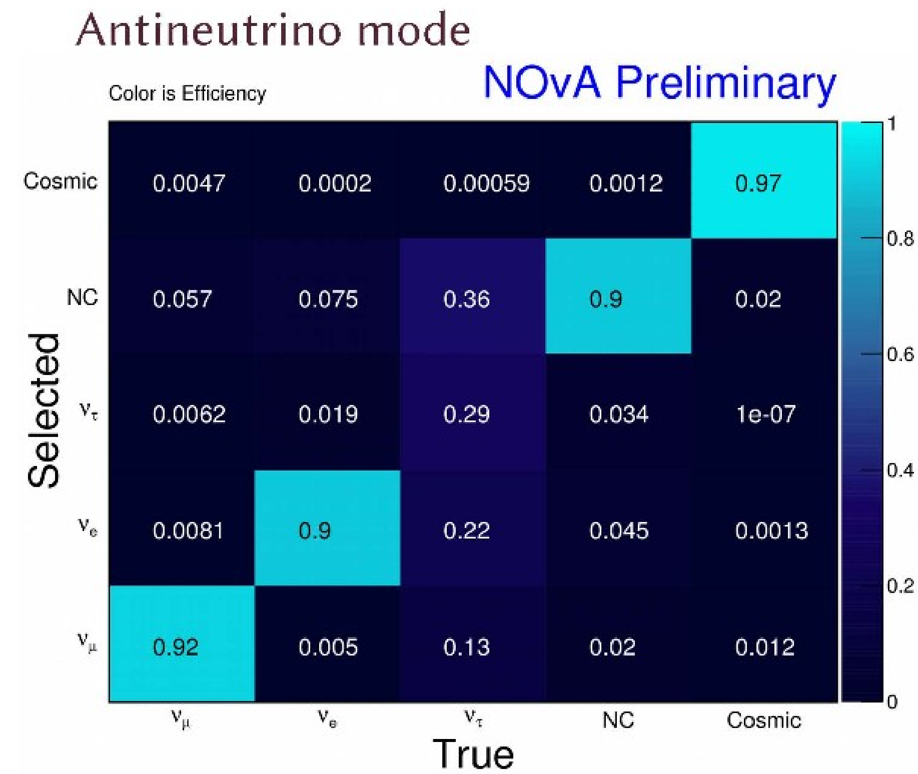




A. Aurisanoet al., "A Convolutional Neural Network Neutrino Event Classifier", arXiv:1604.01444v3

Nova

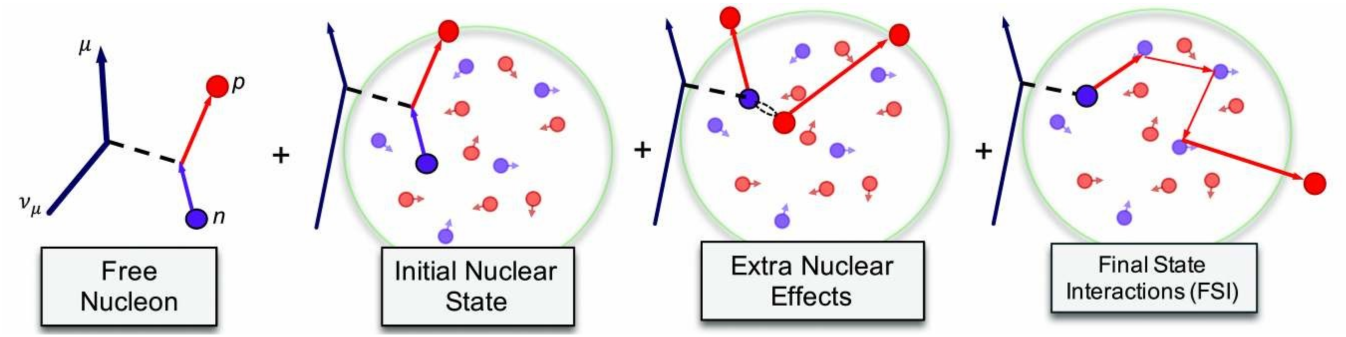




Out performs the ‘standard’ methods previous used
→ now fully adopted into oscillation analysis!

A. Aurisanoet al., “A Convolutional Neural Network Neutrino Event Classifier”, arXiv:1604.01444v3

Neutrino interactions and generators
→ A warning!



Interactions occur with nucleons bound inside a nucleus
→ **Nuclear effects!!**

We know our neutrino interaction generators are 'dodgy' at best
→ **be very careful using them for training !!**

It is not expected that they model energy deposit around the vertex/interaction point well

Is much safe to e.g. do particle ID on an electron (rather than a nue interaction)

SuperFGD:

Voxel classification

in the T2K near detector

The T2K experiment

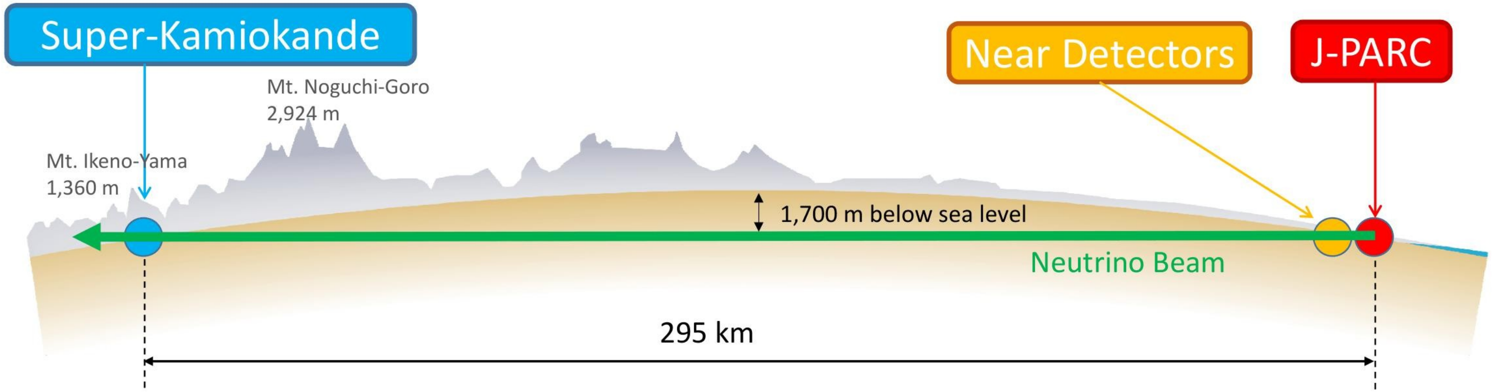
* Long-baseline neutrino oscillation experiment in Japan

* High intensity neutrino beam, predominantly ν_μ ($\bar{\nu}_\mu$)

* Primary goal is to measure neutrino oscillation properties

→ ν_e ($\bar{\nu}_e$) appearance and ν_μ ($\bar{\nu}_\mu$) disappearance

Main goal: CP violation in the lepton sector!



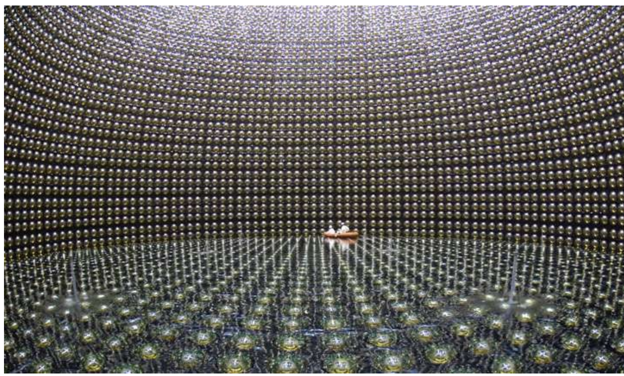
The T2K experiment

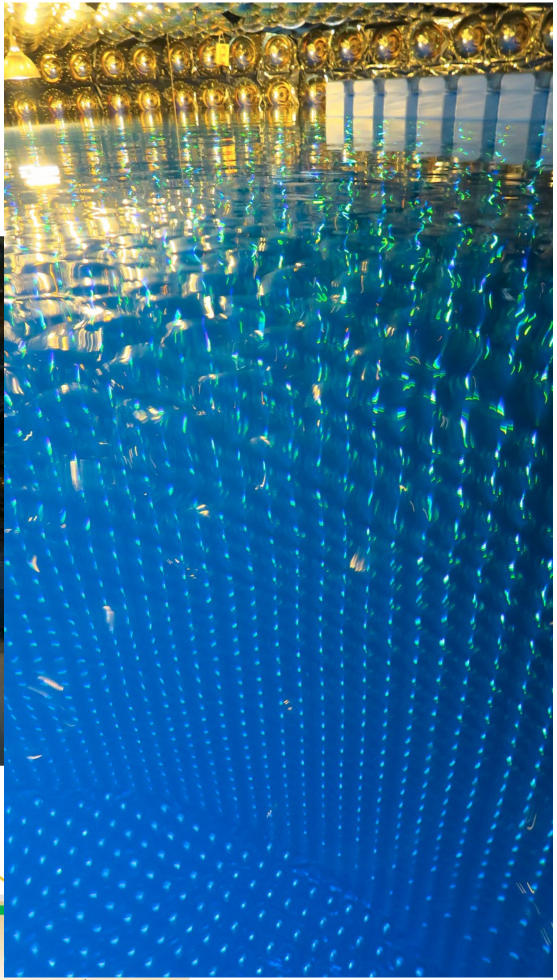
Far detector: **Super-Kamiokande (SK)**

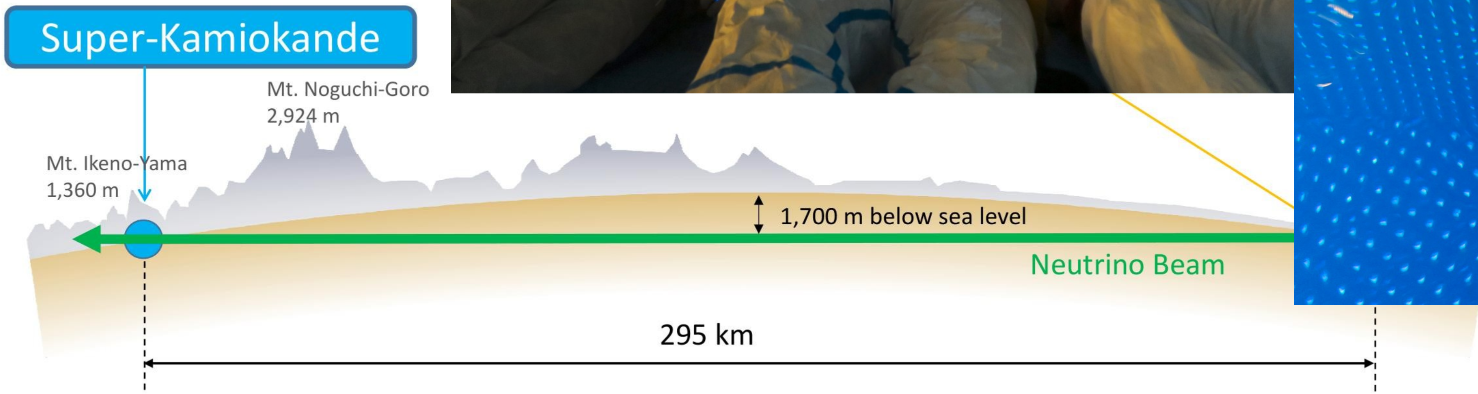
→ measures oscillated neutrino spectrum

→ Cylindrical Water Cherenkov detector

→ 40m tall, 40m radius

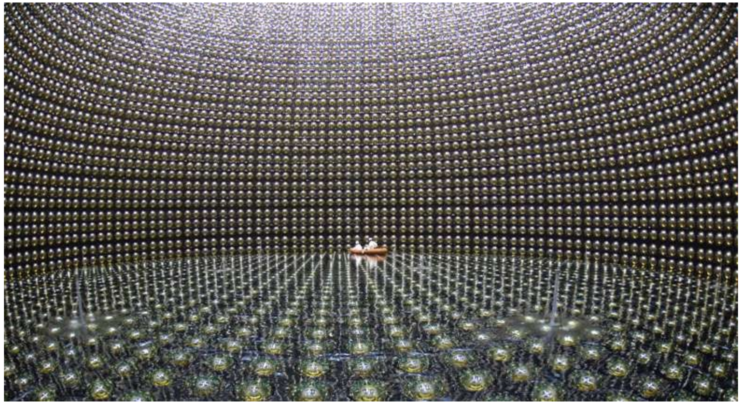






The T2K experiment

Far detector: **Super-Kamiokande (SK)**
→ measures oscillated neutrino spectrum
→ Cylindrical Water Cherenkov detector
→ 40m tall, 40m radius

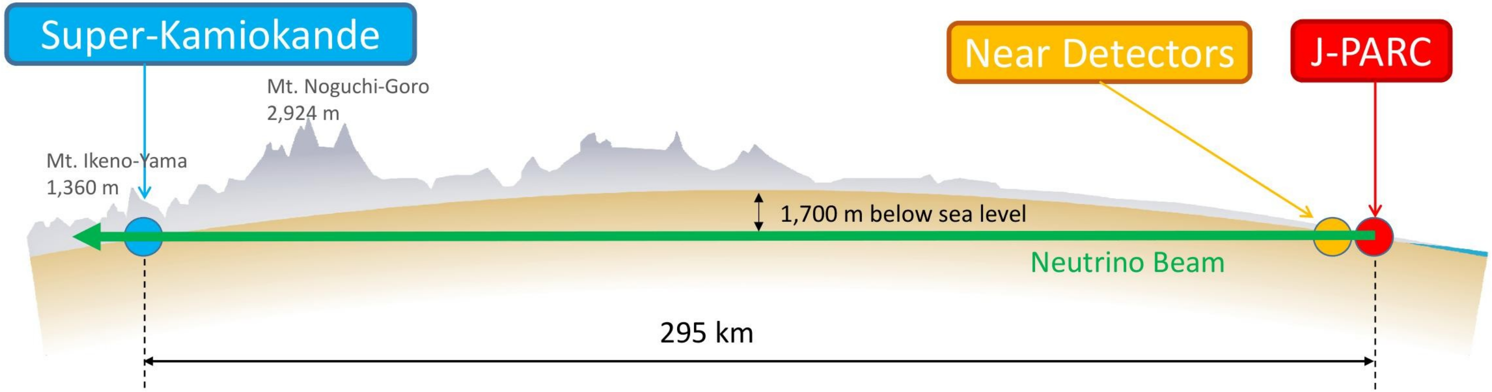


Near detector: ND280
→ composite detector
→ scintillator, EM calorimeters
time projection chambers

Constrains flux and neutrino interaction models

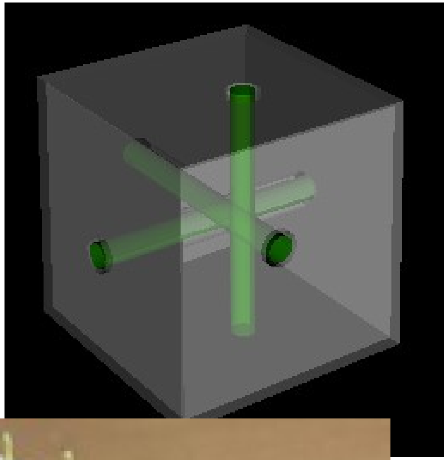
→ Undergoing an upgrade in 2021

→ **New 'SuperFGD'**



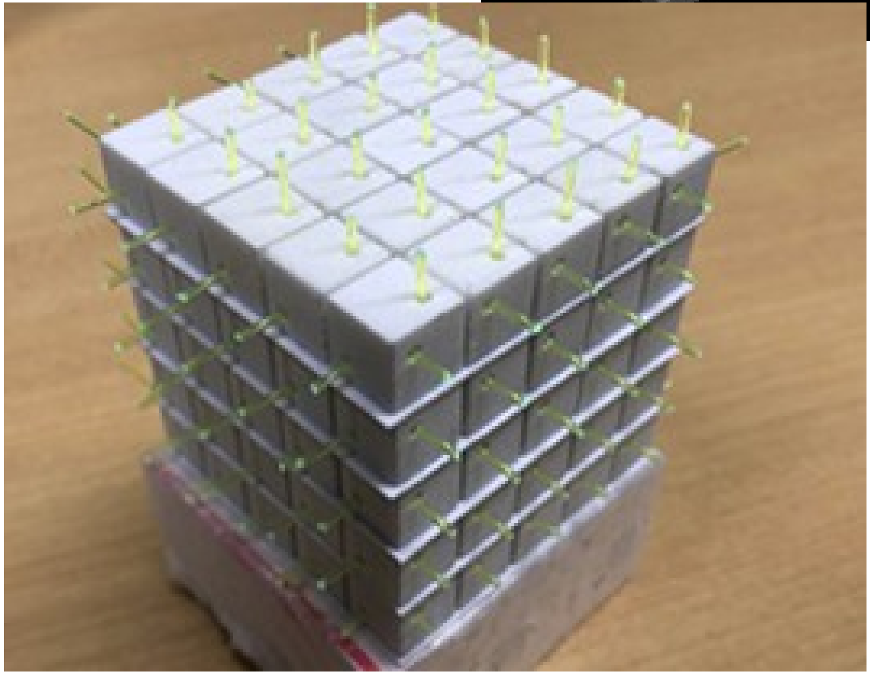
SuperFGD (Super Fine Grained Detector)

New sub-detector for the T2K composite near detector
To be installed in 2021 - currently being tested in a neutron beam



Made of scintillating cubes in a 3D grid
Optical fibres pass through cubes in all 3 planes

Motivation:
→ increases active target **mass**
→ improved **angular acceptance**
→ reconstruct **low energy short tracks**
  → improved hadronic information
  → better $\gamma \rightarrow e^{+} e^{-}$ identification



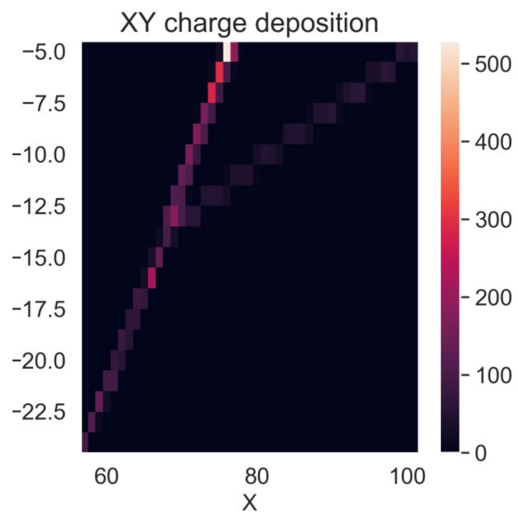
SuperFGD size: 192×192×56 cubes

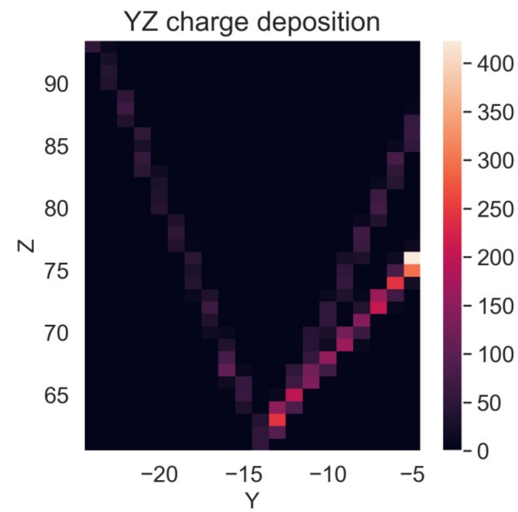
Technical Design Report for nd280 upgrade:
arXiv:1901.03750

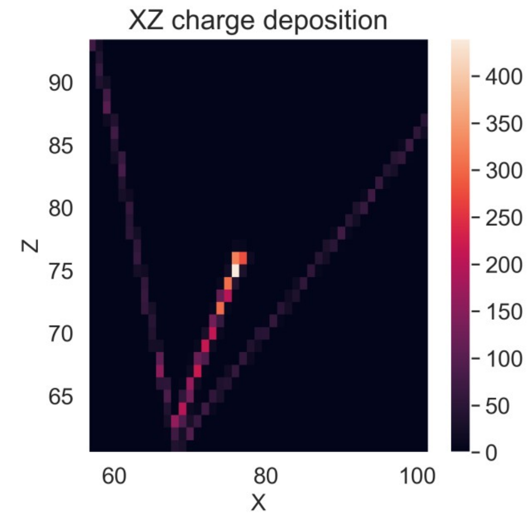
The superFGD

2D readout from each plane

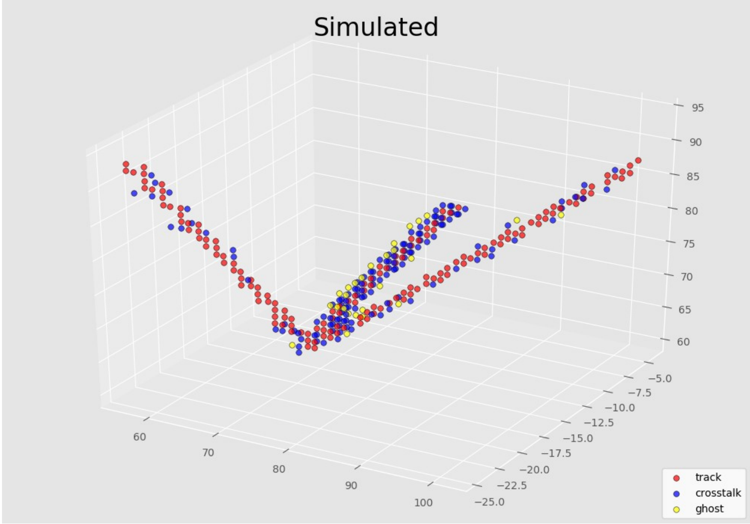
* charge
* time



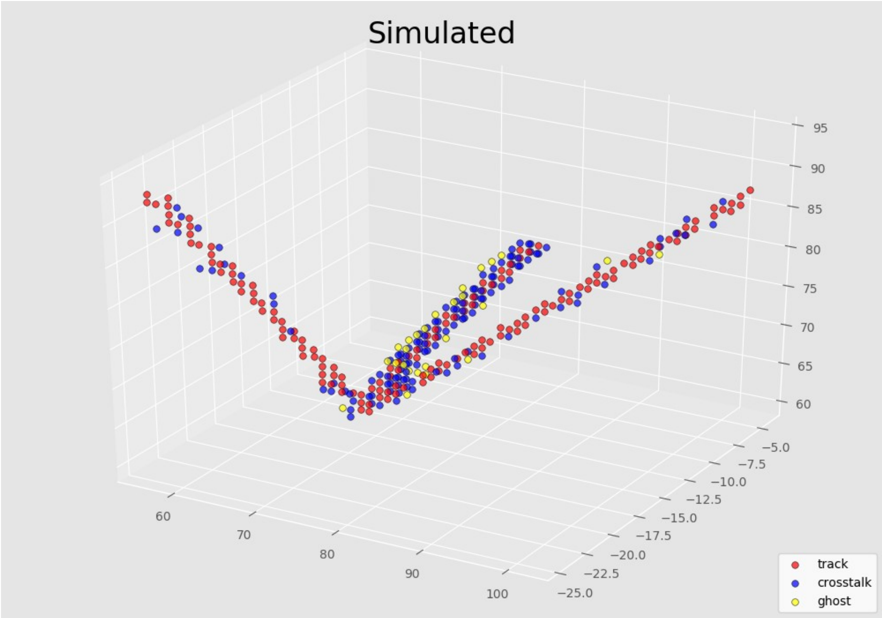




Construct 3D hit info from the 2D planes



SuperFGD: Voxel classification



Construct 3D hits (voxels') from the 2D planes
(some basic recon involved)

- **Track voxel**:
  - a cube with a real deposition
  - a particle has passed through
- **Crosstalk voxel**:
  - a cube with a real deposition
  - but no particle has passed through it
  - physical effect → cube-to-cube optical cross-talk
- Ghost voxel:
  - a cube that does NOT have any real deposition
  - no particle has passed through
  - reconstruction ambiguity when going from 2D to 3D

→ **Plan: Use machine learning technique to classify the voxels**

SuperFGD: Voxel classification

Desires

* Classification of individual nodes/voxels
  (rather than e.g. image recognition, segmentation)

* Works well on unseen data (different numbers of nodes, different config)

Graph Neural Networks (GNNs)

* Suited to individual node classification
* Uses neighbourhood/adjacency of node (suited for ghost, cross talk classification)
* graph representation lightweight (e.g. compared to full 3D grid of the detector)

GraphSAGE (type of GCN)

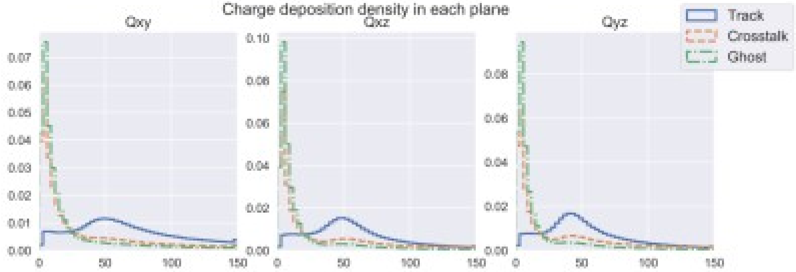
* samples nodes neighbourhood, trains on formations
* sampling and aggregating technique
  → less computationally intense
  → generalises to unseen data, graphs of varying sizes etc.

SuperFGD: Voxel classification

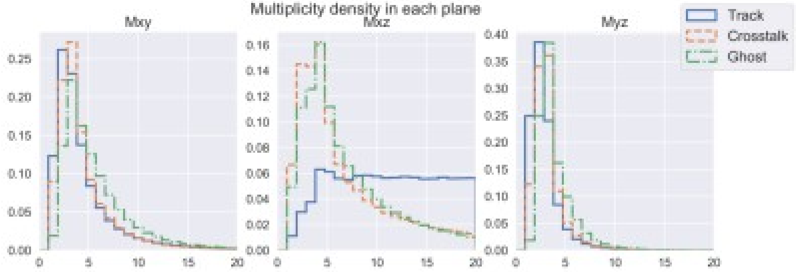
We attach variables to each node

Fundamental / Low level variables

- Voxel position (X, Y, Z).
- # of photoelectrons in each plane.



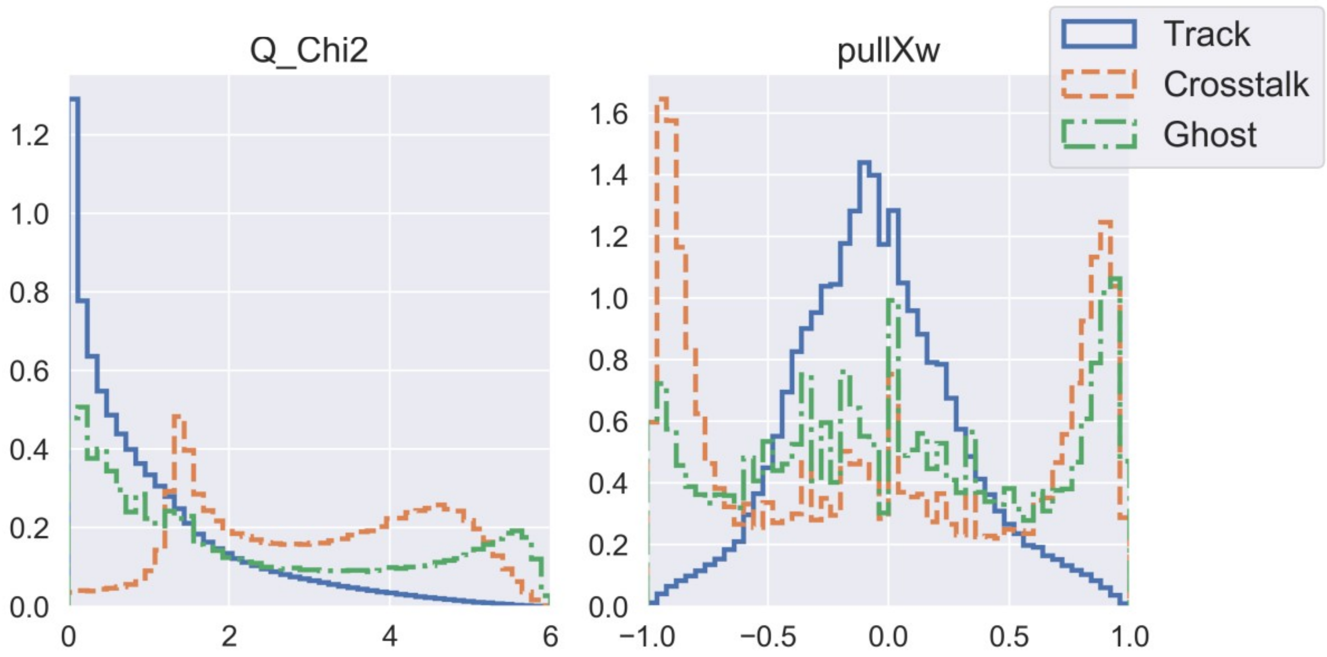
- Multiplicity in each plane.



SuperFGD: Voxel classification

We attach variables to each node

Constructed variables



Can play around adding/removing different constructed variables to see which help the most

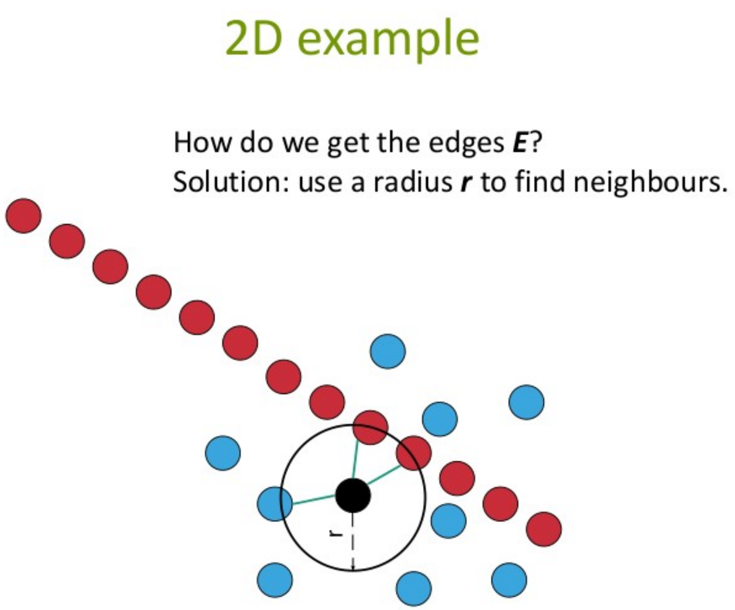
SuperFGD: Voxel classification

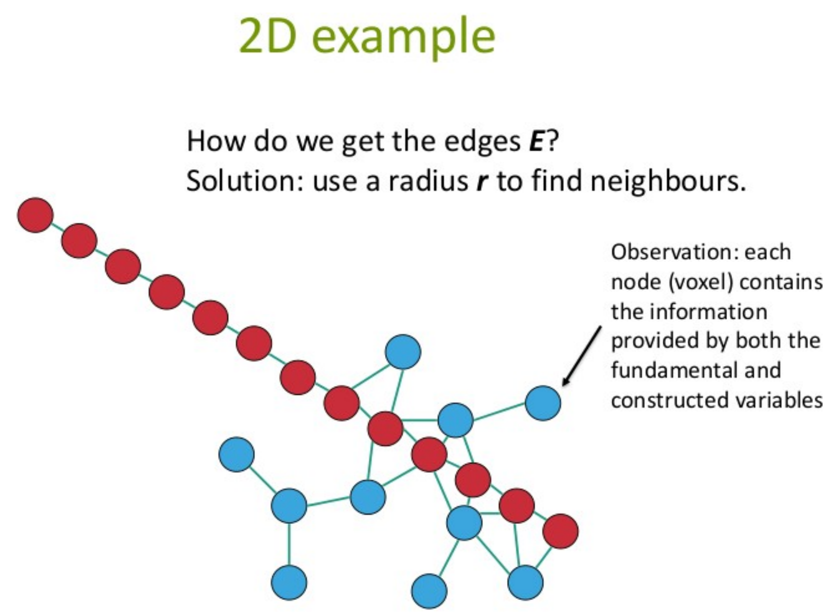
Graphs are a **set of nodes and edges/connections**

For some graphs, you will naturally have the connections/edges
e.g. citation links, chemical bonds

In this case we need to **define connections/edges**

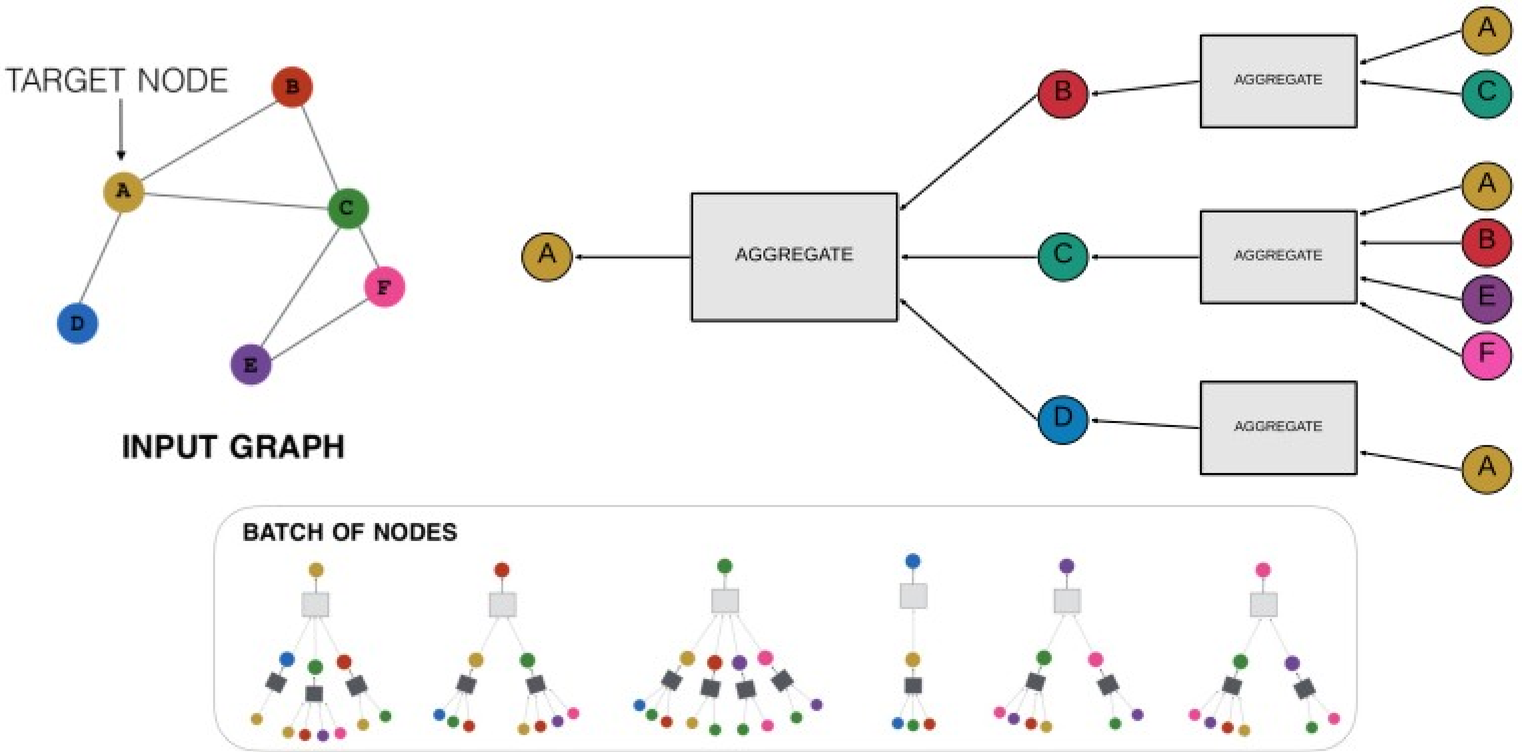
→ you can play around with different ways depending on the problem you are trying to solve





SuperFGD: Voxel classification

Each node is defined by sampling and aggregating its neighbourhood
- can play around with your method of sampling and method of aggregating to suit your problem

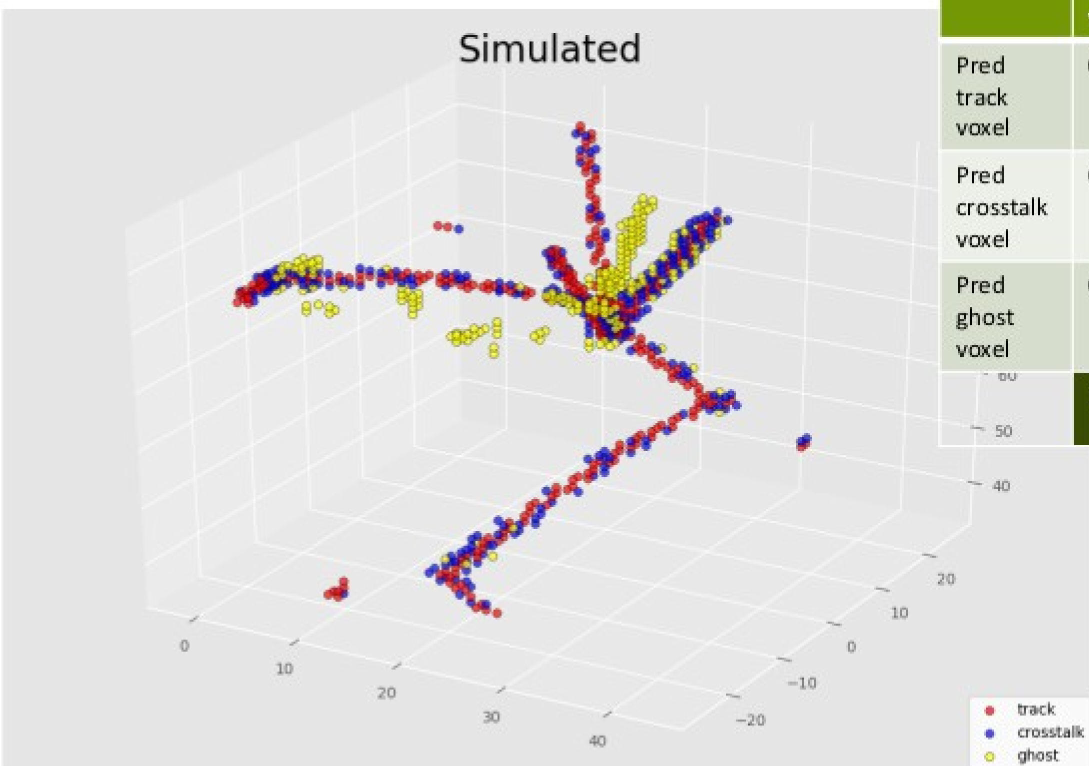


Can think of sampling and aggregating a bit like in CNN when you take a patch of the whole image, and transform/ conv and aggregate/max -pool it

SuperFGD: Voxel classification

Results (GraphSAGE)

Event 1: simulated vs pred. (GIF image*):



	True track voxel	True crosstalk voxel	True ghost voxel
Pred track voxel	**0.898**	0.070	0.029
Pred crosstalk voxel	0.099	**0.890**	0.041
Pred ghost voxel	0.003	0.040	**0.930**
	1.000	1.000	1.000

SuperFGD: Voxel classification

GraphSAGE performing well at classifying voxels

Still in development stage

- defining connections, sampling
    - some hits (0.04%) have no nearest neighbour under current system
- adding/removing constructed variables

Future

- add timing information
- considering systematic uncertainties
    - cross talk model
    - interaction generators (hopefully just a sanity check)
- extend to tasks such as vertex reconstructions

WatCHMaL:

Particle identification

for the Hyper-K

Intermediate Water Cherenkov Detector

arXiv:1911.02369
Variational Autoencoders for Generative Modelling of Water Cherenkov Detectors

https://indico.cern.ch/event/835190/contributions/3613920/attachments/1941211/3218735/WatChMaL_NNN19.pdf

Abhishek Abhishek, Wojciech Fedorko, Patrick de Perio, Nicholas Prouse, Julian Z. Ding

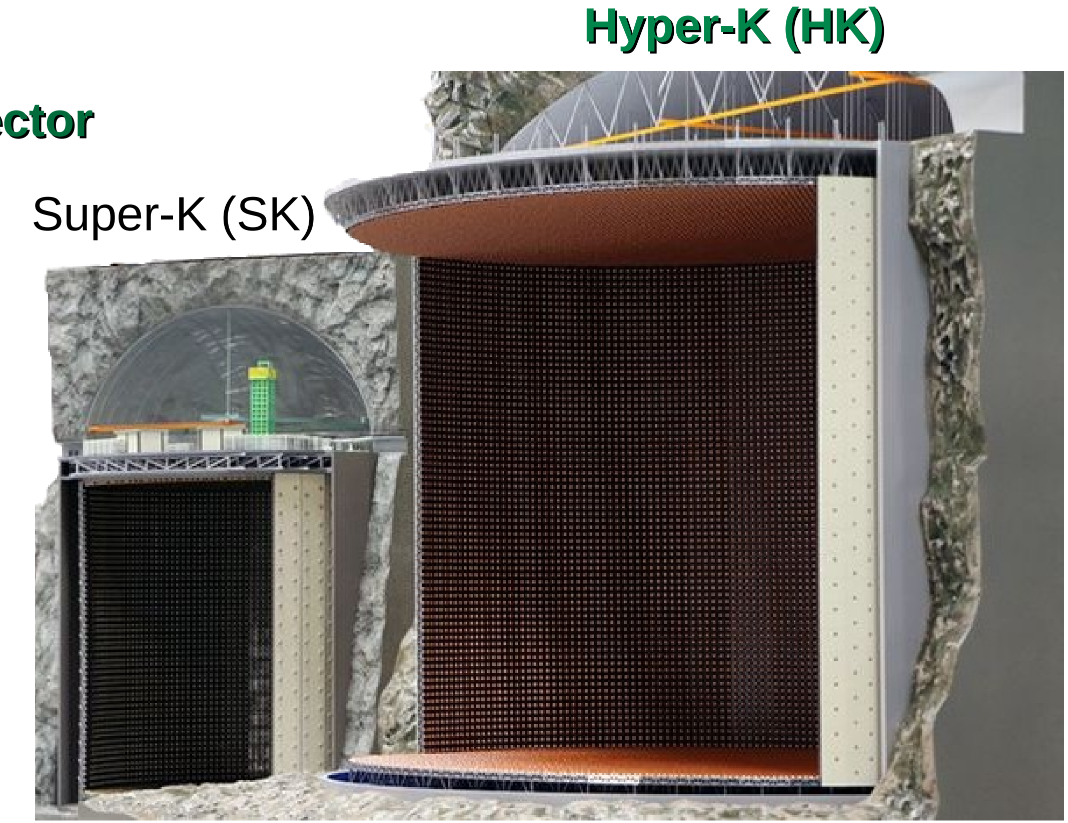
The Hyper-K experiment

Bigger and better version of T2K

* T2K beam

* T2K detectors

*** New Water Cherenkov (WC) far detector**



Size comparison

SK: height 40m, diameter 40m
→ 50 kton

HK: height 72m, dimeter 68m
→ 258 kton

The Hyper-K experiment

Bigger and better version of T2K

* T2K beam
* T2K detectors
* New Water Cherenkov (WC) far detector
* **New intermediate WC detector (IWCD)**

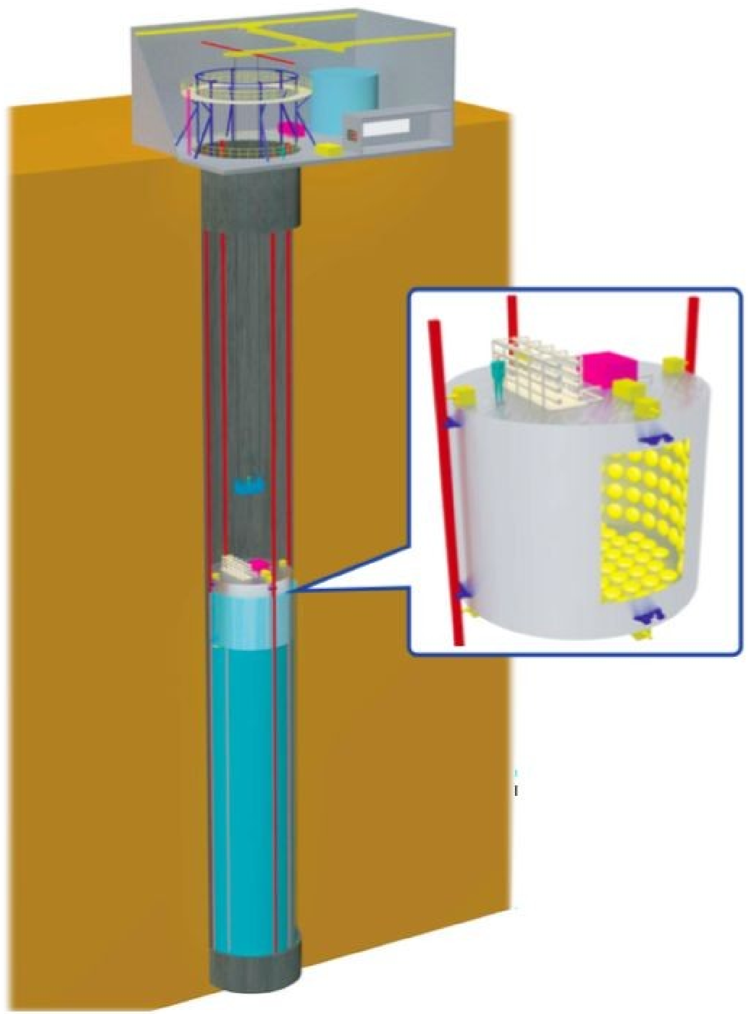
Additional near/intermediate detector (0.75km)

Also designed to constrain flux and neutrino interactions

Same target (water) as far detector

Ability to move up and down
→ samples flux at different angles
→ sample flux with different energy peaks/profiles
→ measure interactions across range of energies

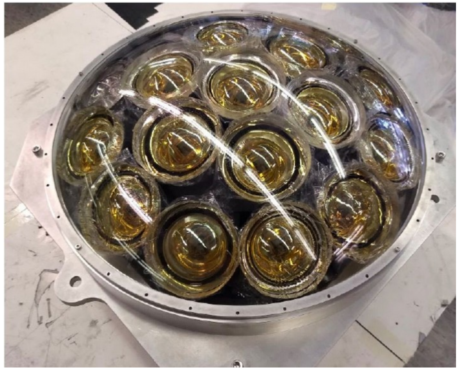
IWCD



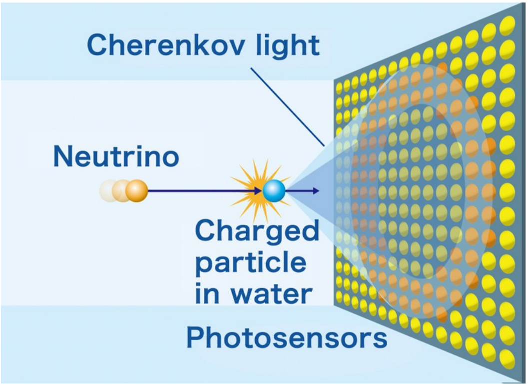
Hyper-K: Intermediate Water Cherenkov Detector

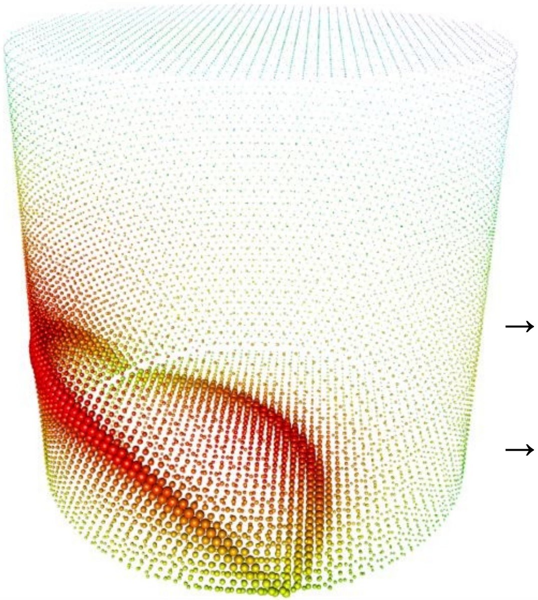
Detector walls lined with photomultiplier tubes (PMTs)
→ multi-PMT module contains 19 PMTs (3 inch)

→ detected light creates 2D image on the tank walls
→ **Can we use ML for particle identification**



multi-PMT





→ each multi-PMT can act as a pixel/node for neural network methods

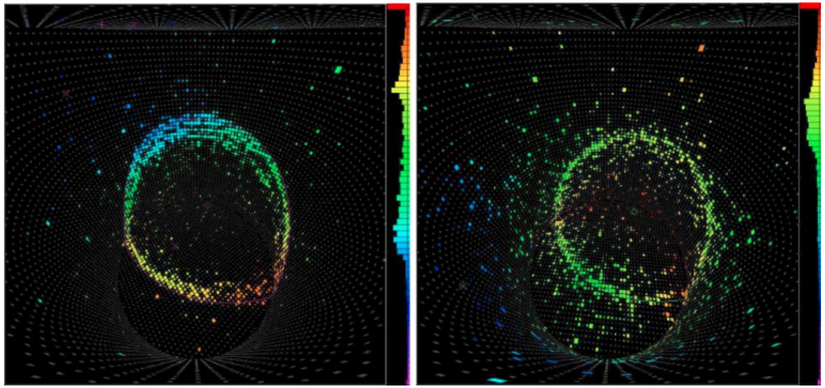
→ each pixel has 19 channels (charge of each PMT)
(can extend to 38 channels if you include time)

Hyper-K: Intermediate Water Cherenkov Detector

Can ML help with gamma(s) Vs electron ?
Events with pions?

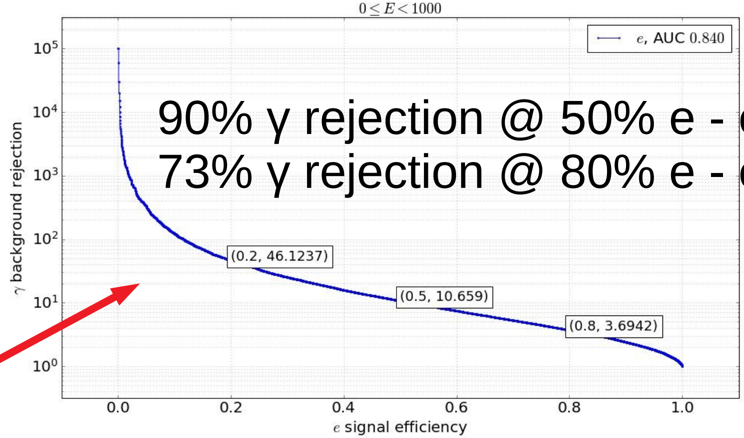
2D patterns in grid characterize particle type
→ 2D CNN obvious place to start

Issue: How to deal with the cylindrical shape?
- → For now, **ignore top/bottom** of detector
- → Simulate **particles from centre** of tank, perpendicular **towards the walls**
  - → particle gun: e, mu, gamma



mu rings – clear   e rings - 'fuzzy'

e Vs gamma



90% γ rejection @ 50% e - efficiency
73% γ rejection @ 80% e - efficiency

2D CNN results:

(excludes top/bottom of tank)

e Vs mu

>99% mu/e discrimination

This is very promising!

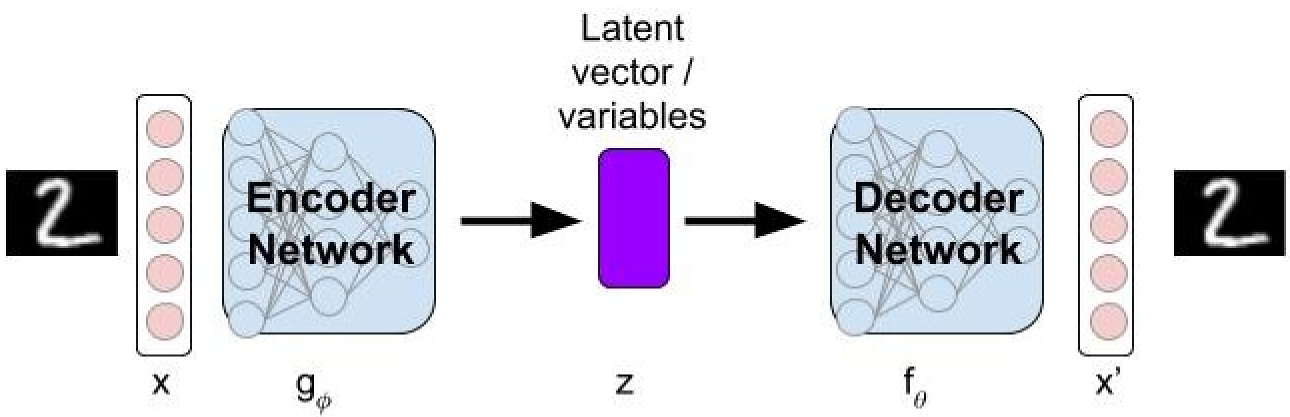
Hyper-K: Intermediate Water Cherenkov Detector

Fully supervised CNN showed very promising potential

→ encouraged further exploration of ML techniques

→ Variational Autoencoders (VAE)

Autoencoder



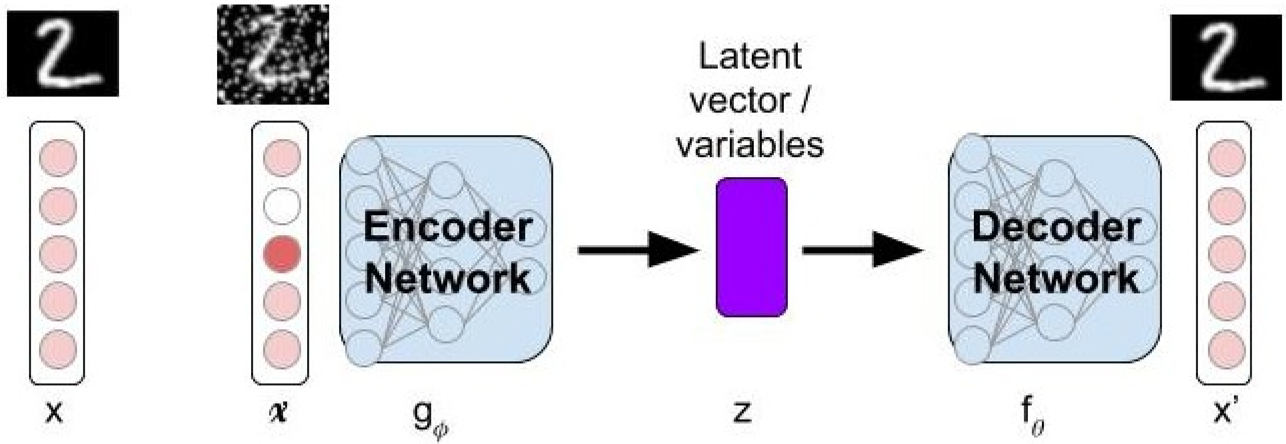
$$L(\theta,\phi) = 1/n \; \Sigma_i \, [\; x^i - f_\theta(\, g_\phi(x^i)\,)\,]^2$$

* Training your network to find the reduced set of latent variables
* Train such that the decoder network can reproduce the original image

→ No 'truth' needed, i.e. can you unlabelled data / unsupervised

→ Uses: compression / decompression of data!

(denoising) Autoencoder



$$L(\theta,\phi) = 1/n \;\; \Sigma_i \, [\; x^i - f_\theta(\, g_\phi(\boldsymbol{x}^i)\,)\,]^2$$

* Modify input, but train on the original

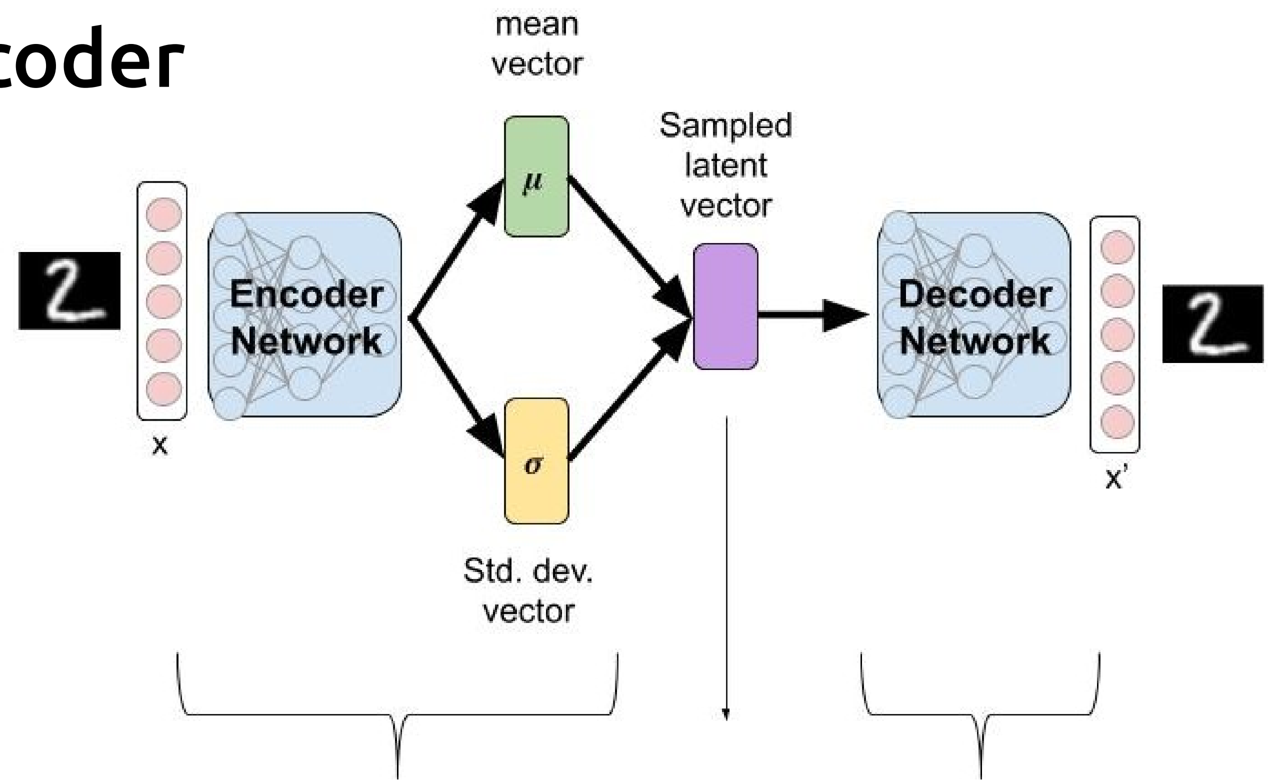
→ Denoising/cleaning images

→ Object removal in images

Variational Autoencoder

Replace bottleneck latent variable with probabilistic distributions

Sample from this to form latent vector to feed to the decoder



$q_\phi(z|x)$ $\quad$ $z = \mu \,.\, \varepsilon$ $\quad$ $p_\theta(x|z)$

Generative model

→ Can use unlabelled data to learn the latent variables

→ sample from the latent space to generate new images

$$L = -E_{q\phi(z|x)}[\log\{ P_\theta(x|z) \}] + D_{KL}(q_\phi(z|x) \mid P_\theta(z))$$

Reconstruction term $\quad$ Divergence / regularisation

Note: Can add a parameter to control regularization strength

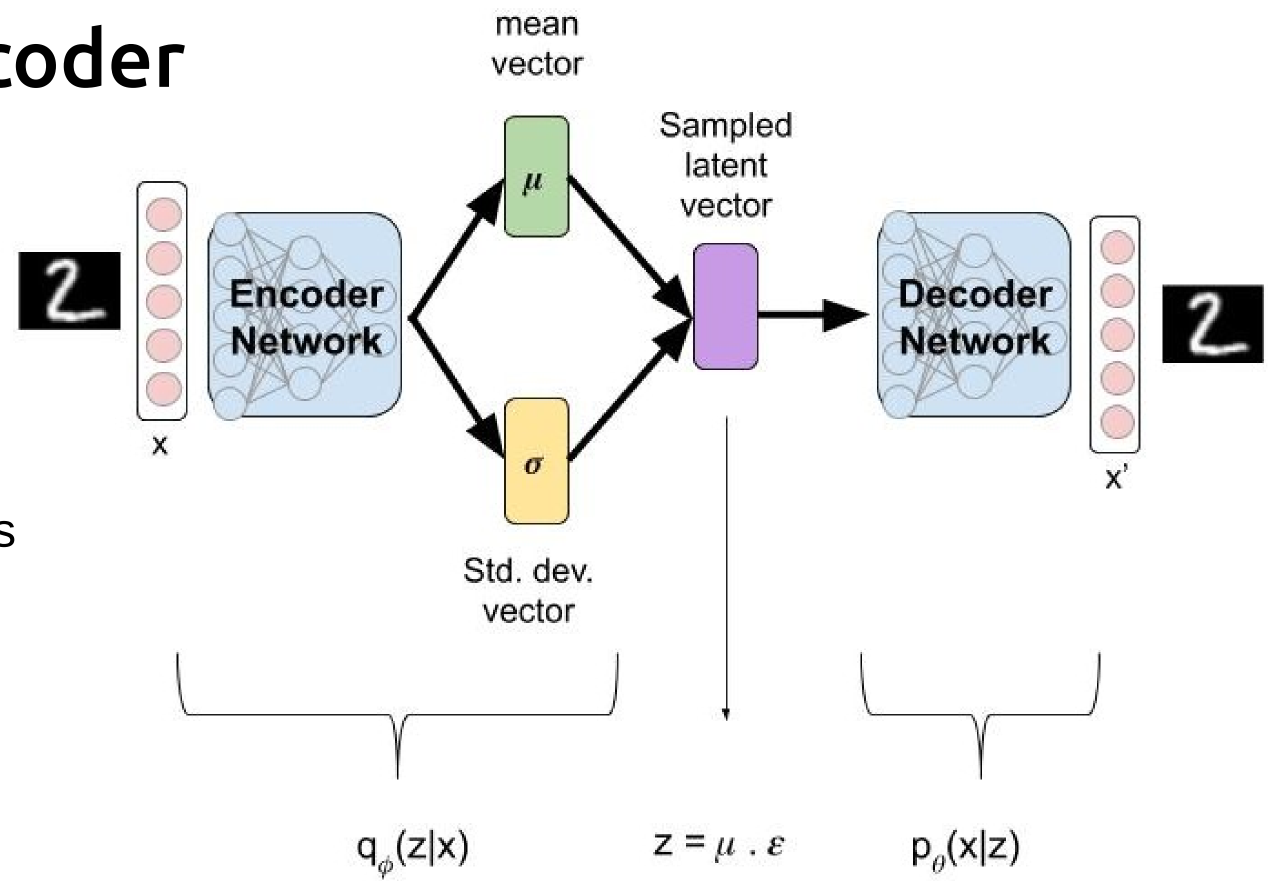
Variational Autoencoder

Generative model

Hope to train your latent variables to have physical meaning

e.g. moving in one direction in your latent space can equate to your image rotating, or getting fatter/thinner etc.

→ sample from the latent space to generate new images

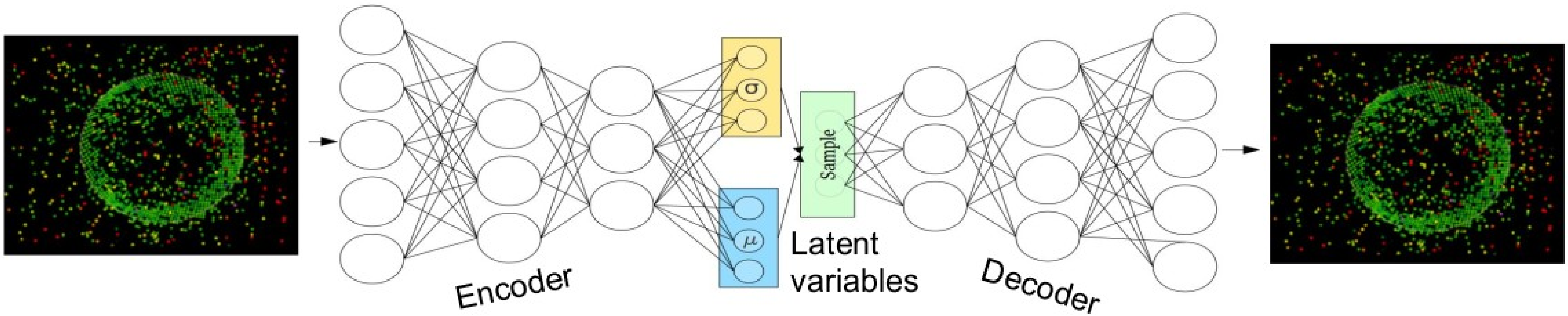


$$L = -E_{q\phi(z|x)}[\log\{ P_\theta(x|z) \}] + D_{KL}(q_\phi(z|x) \mid P_\theta(z))$$

Reconstruction term

Divergence / regularisation

Hyper-K: Intermediate Water Cherenkov Detector



Initial test:

- Unsupervised / unlabelled training to learn latent space

- Generate new event images

- Test if directions in latent space correspond to physical interpretations

Hyper-K: Intermediate Water Cherenkov Detector

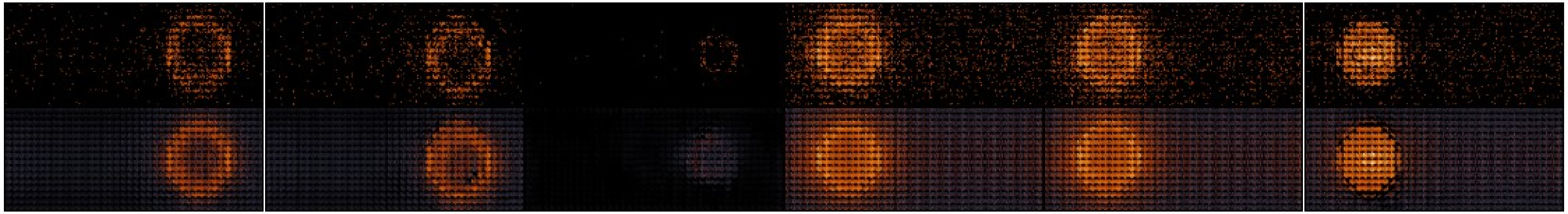


Figure 2: Cherenkov ring images comparing actual simulated events (top) with their corresponding VAE reconstructed events (bottom).

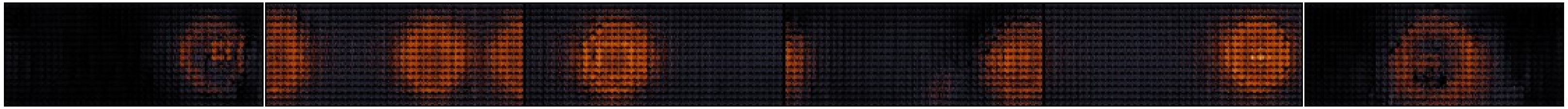


Figure 3: Cherenkov ring images for events randomly sampled from the latent prior p(z) = $\mathcal{N}(0, I)$.

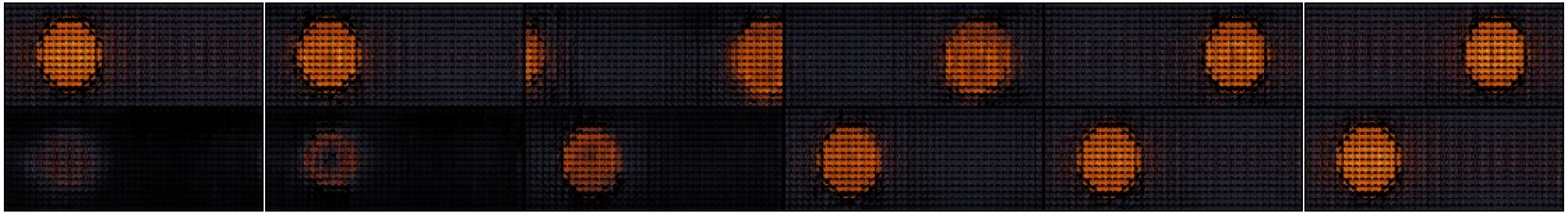


Figure 4: Linear interpolation in the latent space for μ^- events along the angle axis from $\phi = 0$ to $-\pi$ (top) and energy axis from 200 MeV to 800 MeV (bottom).

Hyper-K: Intermediate Water Cherenkov Detector

Next test: Introduce small sample of labelled data

Test semi-supervised and supervised learning for particle identificaiton (PID)

Number of training examples	**γ background rejection (%) at 50% e^- signal efficiency**		**γ background rejection (%) at 80% e^- signal efficiency**	
	SS-CNN	**CNN**	**SS-CNN**	**CNN**
$11,250$	**77.6**	76.4	**50.7**	46.3
$22,500$	**80.4**	78.1	**54.3**	48.5
$45,000$	**80.7**	79.4	**55.9**	49.9

→ Semi-supervised learning outperforms fully supervised method

Hyper-K: Intermediate Water Cherenkov Detector

Potential

* Train on unlabelled data, calibration data, control samples

* Using direction in latent space to extrapolate to phase space with limited data

* Training on real calibration data
  → use VAE for part of MC generation
  → possible to circumvent detector model/syst for certain aspects

Possible difficulties

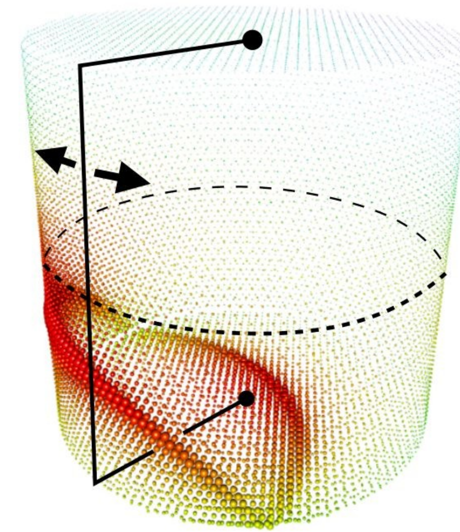
* Image (re)construction still needs work
  - ring sharpness, replicating dark noise and scattered/reflected light

* Low energy events / neutron capture expected to be difficult
  - sparse PMT hits

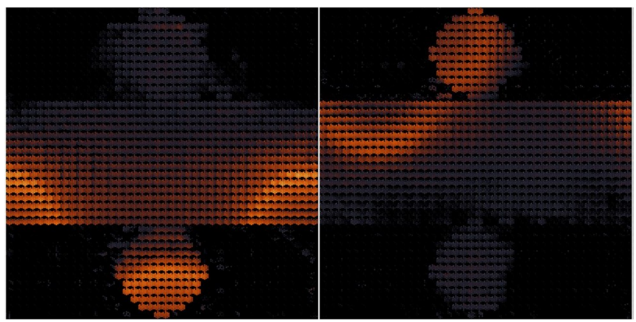
Hyper-K: Intermediate Water Cherenkov Detector

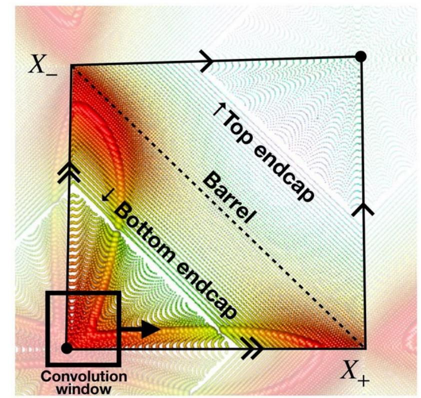
Biggest issue so far: Detector geometry

- test so far have been based on tank wall only, not the top/bottom



Ways to ‘flatten’ geom and construct CNN





Methods that focus just on the pixels

Point cloud: PointNET

Each pixel is a member of a list pointNET uses symmetric functions to avoid problem of ordering in the list

GraphSAGE

Test options for defining edges and sampling method

Summary

Other uses of ML in nu physics

Current

T2K near detector we use BDTs for pi0 tagging
→ low level variables from multiple detectors
→ effective at improving efficiency

Liquid argon TPC images are perfect for CNN
- MicroBooNE
- Dune

Potential

HK
→ possibilities to use ML for basic recon info for DAQ
→ Current reconstruction is incredibly slow… can we use ML to speed things up?

Automating shifts (to some extent) – identifying problems/solutions

Summary

Lots of different uses for ML in neutrino physics

NoVa led the way with 2D CNN work for particle ID

Other techniques being explored such as voxel classification: GNN

Potential for reconstruction technqiues
- microboone leading the way
- nd280 upgrade and HK heading towards that diraction

New/Current generation of Liquid argon detectors well suited to CNN

Warning: Be careful with trusting hadronic/vertex information from neutrino generators!

References

ArXiv: 1901.03750 T2K ND280 upgrade technical design report

MicroBooNE: Image based reconstruction
https://indico.desy.de/indico/event/21853/session/2/contribution/46/material/slides/0.pdf

ArXiv: 1406.5298   Semi-supervised Learning withDeep Generative Models

Backup Slides

SK Systematic errors

Error (%)

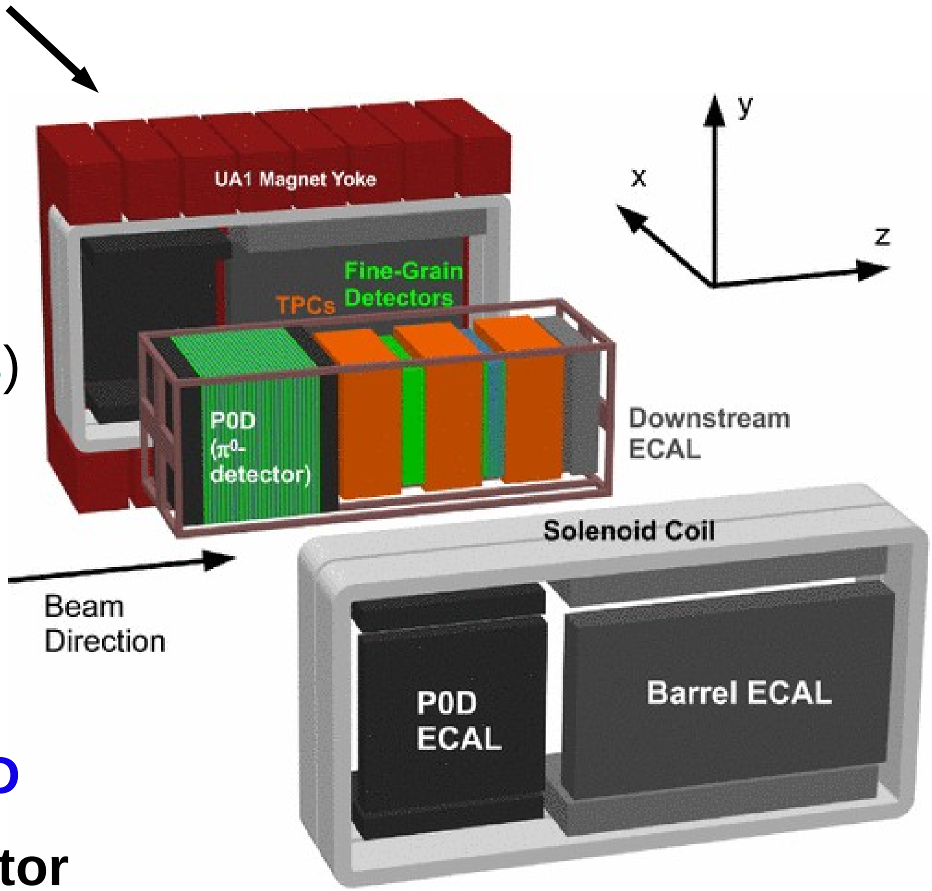
	1-Ring μ		1-Ring e			
Error source	FHC	RHC	FHC	RHC	FHC 1 d.e.	FHC/RHC
SK Detector	2.40	2.01	2.83	3.80	13.15	1.47
SK FSI+SI+PN	2.21	1.98	3.00	2.31	11.43	1.57
Flux + Xsec constrained	3.27	2.94	3.24	3.10	4.09	2.67
E_b	2.38	1.72	7.13	3.66	2.95	3.62
$\sigma(\nu_e)/\sigma(\bar{\nu}_e)$	0.00	0.00	2.63	1.46	2.61	3.03
NC1γ	0.00	0.00	1.09	2.60	0.33	1.50
NC Other	0.25	0.25	0.15	0.33	0.99	0.18
Osc	0.03	0.03	2.69	2.49	2.63	0.77
All Systematics	5.12	4.45	8.81	7.13	18.38	5.96
All with osc	5.12	4.45	9.19	7.57	18.51	6.03

Table 5: Percentage error on event rate by error source and sample. Final column is the percentage error on the r
of FHC/RHC events in the one-ring e sample.

Near Detectors

280m from the ν ($\bar{\nu}$) source

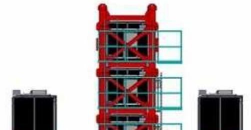
Magnetised



ND280

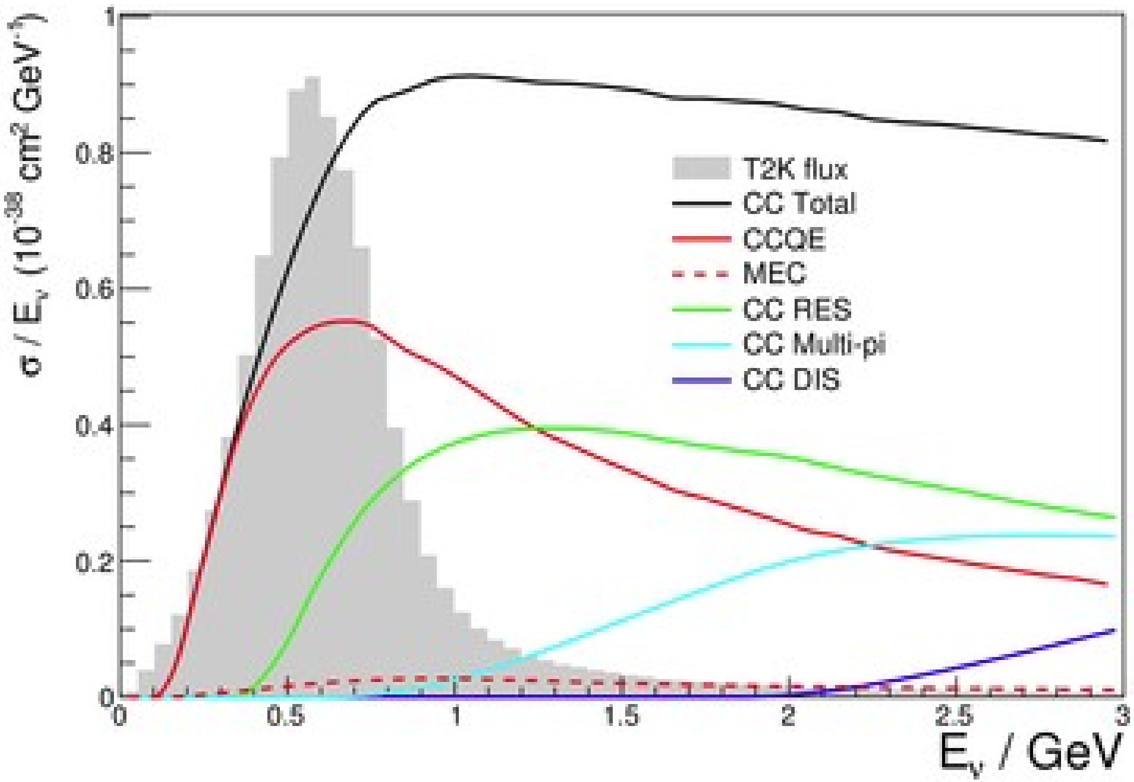
Same off-axis angle as SK

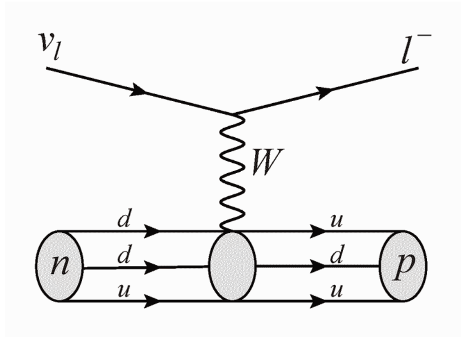
- Active target mass → 2 x scintilltors (**FGDs**)
  → vertex reconstruction
- 3 Time projection chambers (**TPC**)
  → **momentum** reconstruction
  → **charge** identification
  → Particle identification (**PID**)
- Electromagnetic calorimeters (**Ecal**) → **PID**
- **π^0 detector** and **side muon range detector**



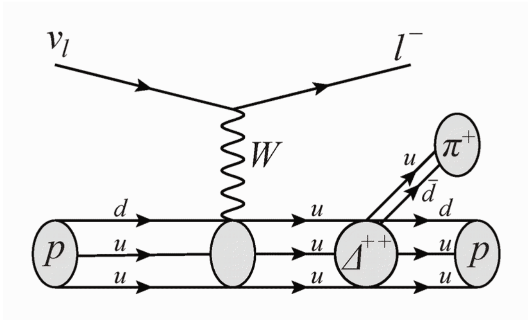
INGRID

T2K flux and neutrino cross sections

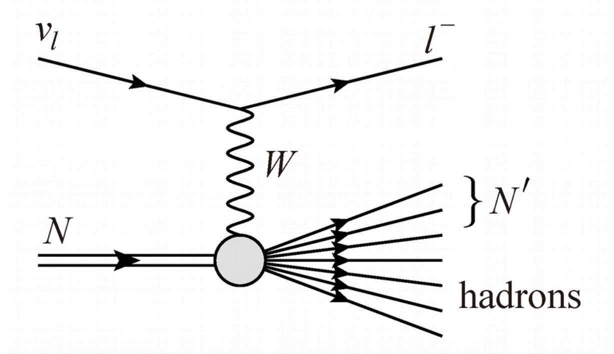




CCQE

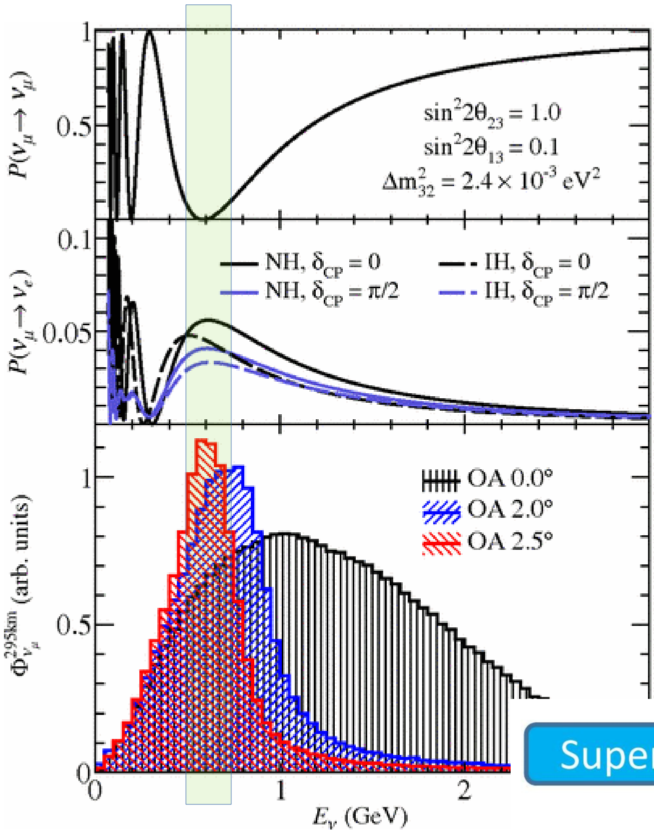


CCRES



CCDIS

The T2K experiment



* Long-baseline neutrino oscillation experiment

* High intensity neutrino beam, predominantly ν_μ ($\bar{\nu}_\mu$)

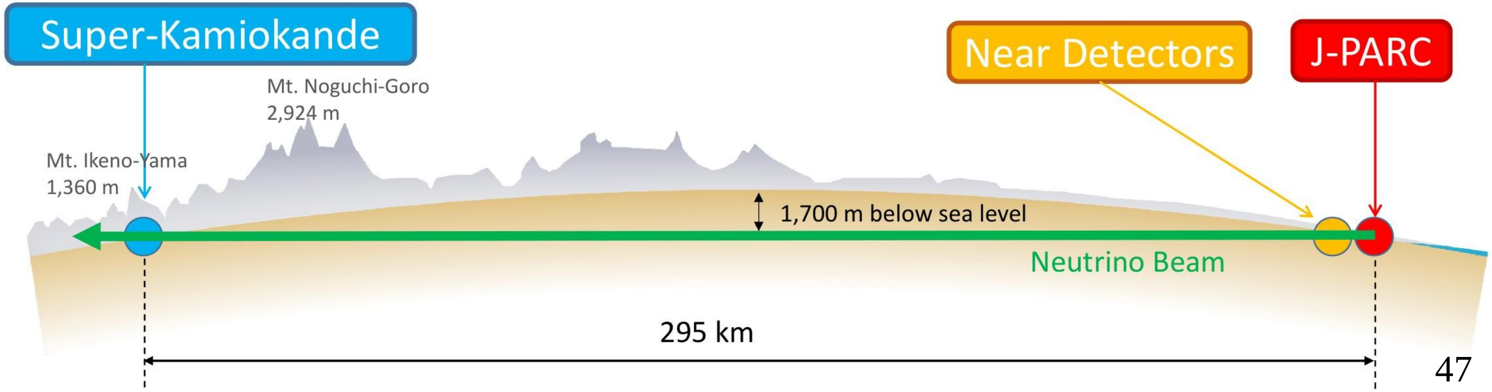
* Primary goal is to measure neutrino oscillation properties

  → ν_e ($\bar{\nu}_e$) appearance

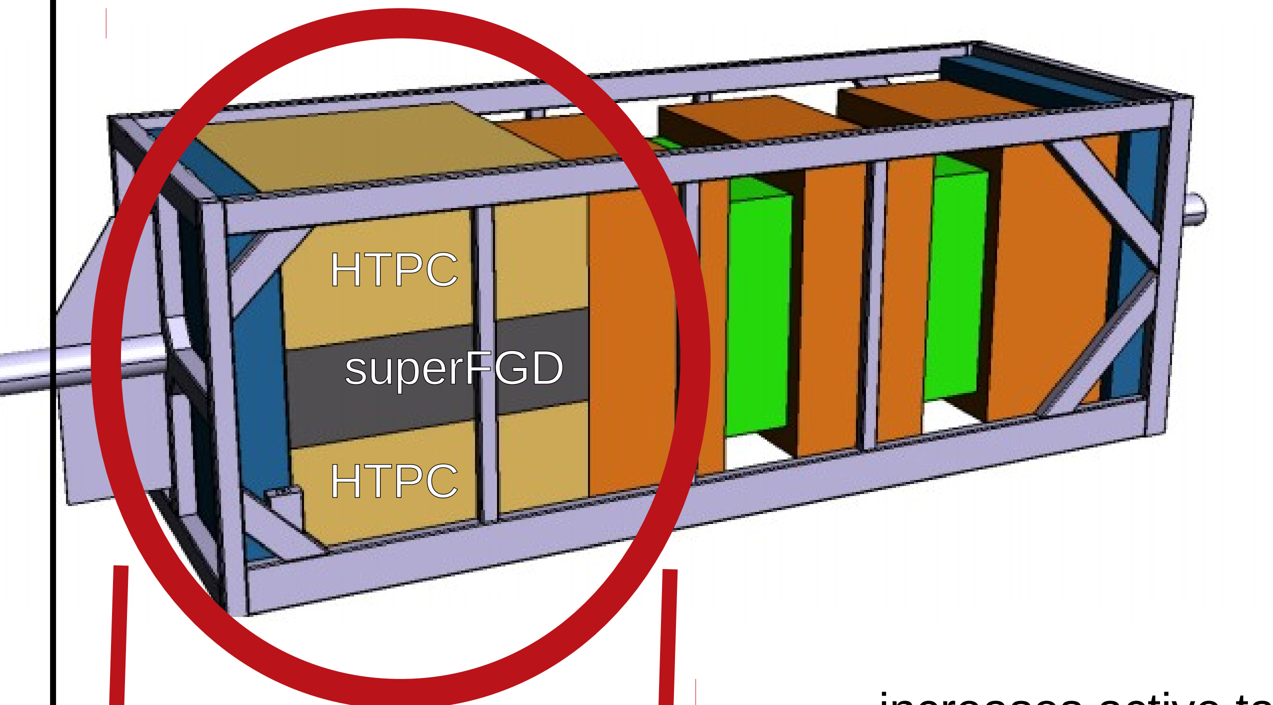
  → ν_μ ($\bar{\nu}_\mu$) disappearance

* Off-axis far detector → oscillated neutrinos (295km)

* On/Off-axis near detectors → unoscillated beam (280m)



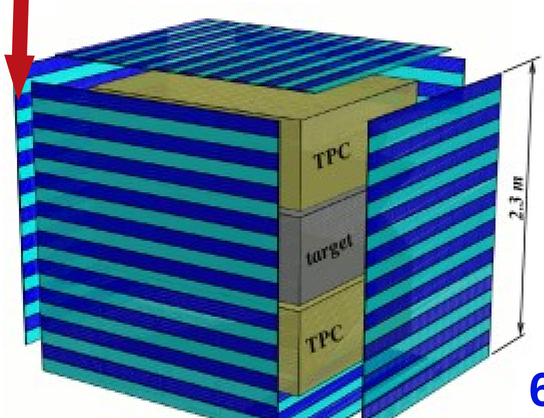
ND280 upgrade



ND280 will be upgraded in 2021 during the beam upgrade

Pi0 detector is being replaced by
- **SuperFGD**
  - higher granularity
  - 3D readout
- **Horizontal TPCs (HTPCs)**
- **Time of Flight (ToF) planes**

→ increases active target **mass** for oscillation analysis
→ improved **angular acceptance**
→ able to reconstruct **low energy short tracks**
  → improved hadronic information
  → better $\gamma \rightarrow e^+ e^-$ identification



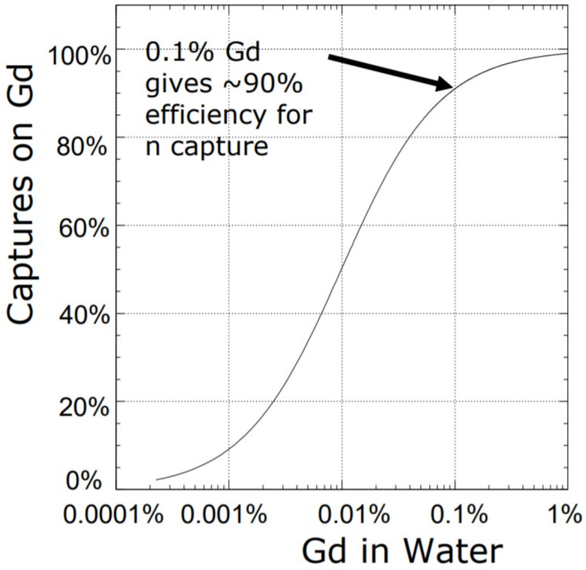
6 ToF planes

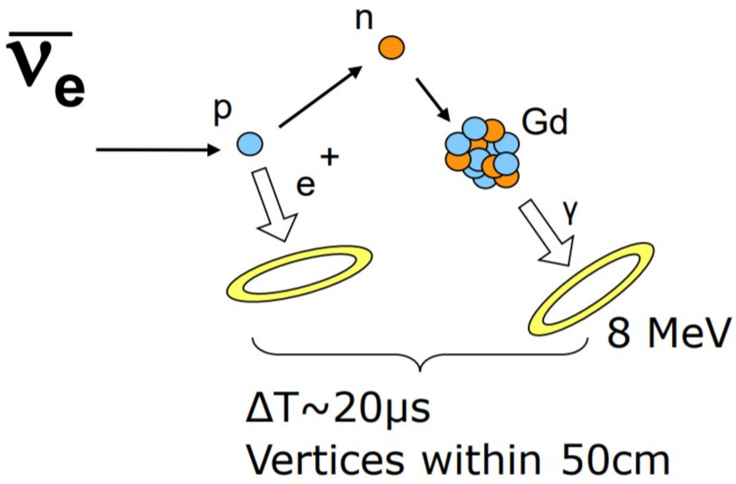
T2K II: SK upgrade

* SK repairs performed in 2018
- detector drained and cleaned
- reinforcement of water sealing
- improved tank piping
- PMTs replaced

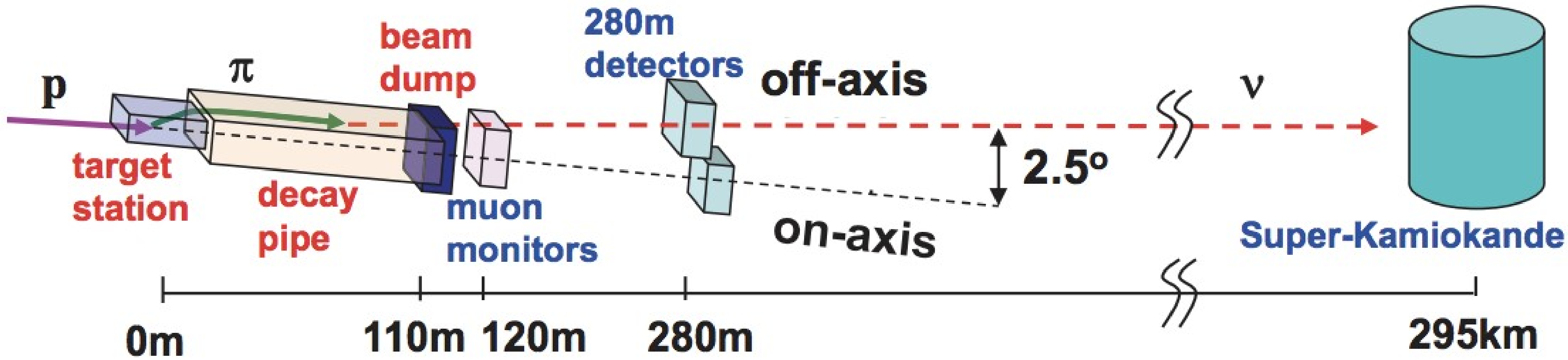
* Plan to add **Gadolinium** to the water
- 0.01% next year
- increase to 0.1% eventually

→ **Better ν / $\bar{\nu}$ separation**





T2K



p
π
beam
dump
280m
detectors
off-axis
ν
target
station
decay
pipe
muon
monitors
2.5°
on-axis
Super-Kamiokande
0m
110m
120m
280m
295km

PointNet – Point cloud neural network

arXiv:1612.00593

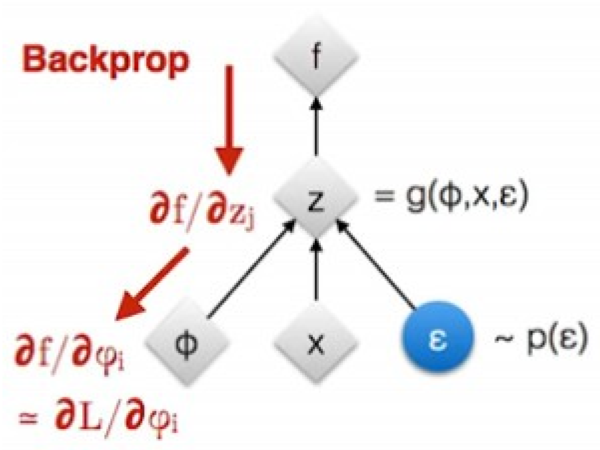
Data stored as unordered set of points
→ less cumbersome than creating 3D grid of voxels
→ no combinatorial issues

Requires points to be invariant to permutations
→ Done by using a symmetric function to aggregate the info from each point/neuron

Invarient under geomtric transformations
- point cloud roatation should not alter result

Construct a family of symmetric functions by neural networks

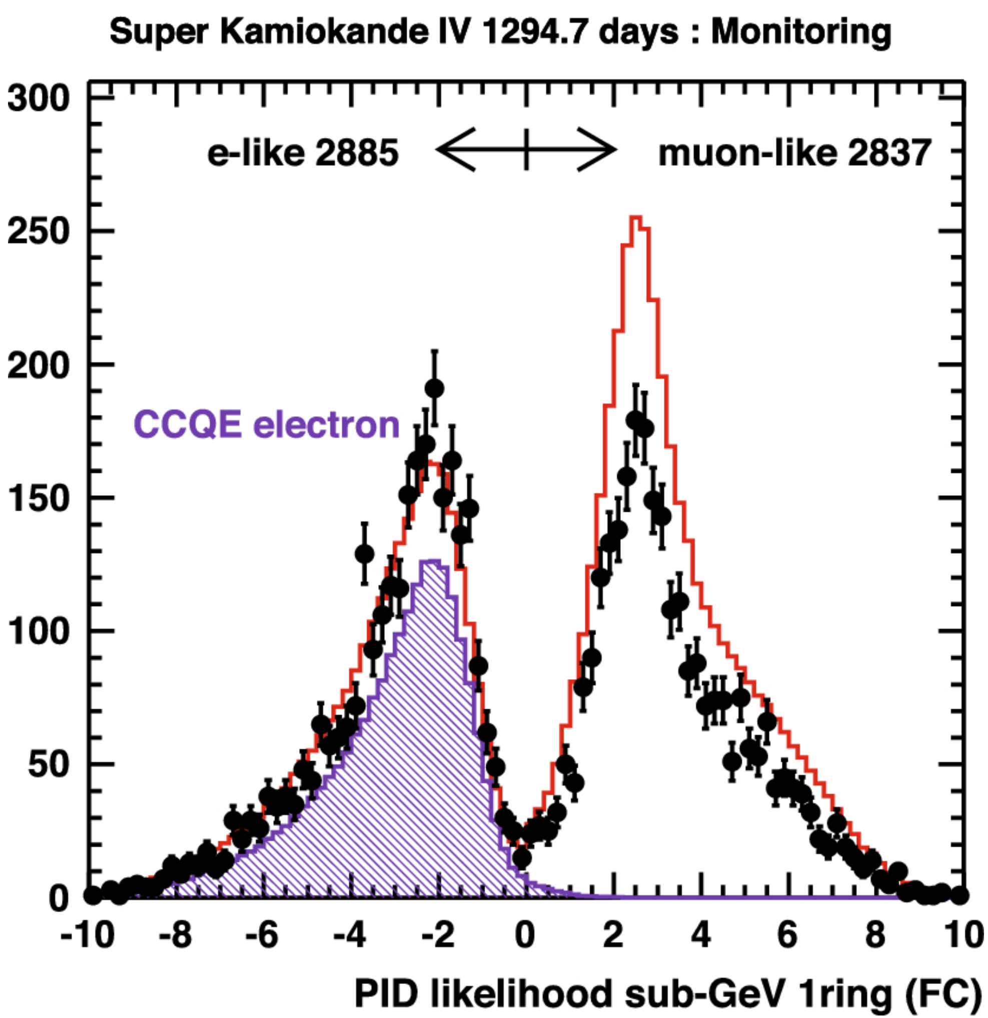
Variational Autoencoder



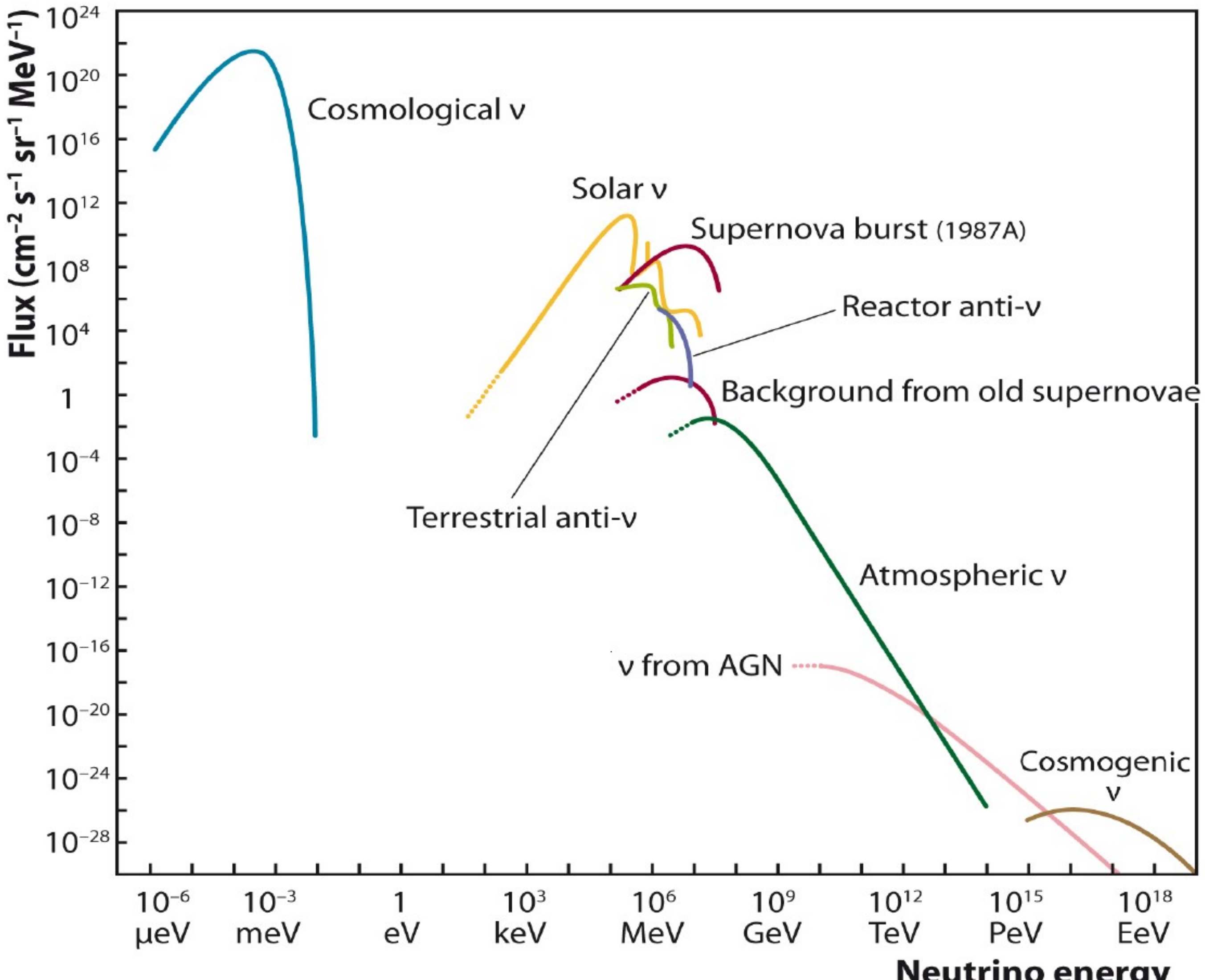
Don’t need to back propagate through the stochastic node

→ we are not trying to modify epsilon

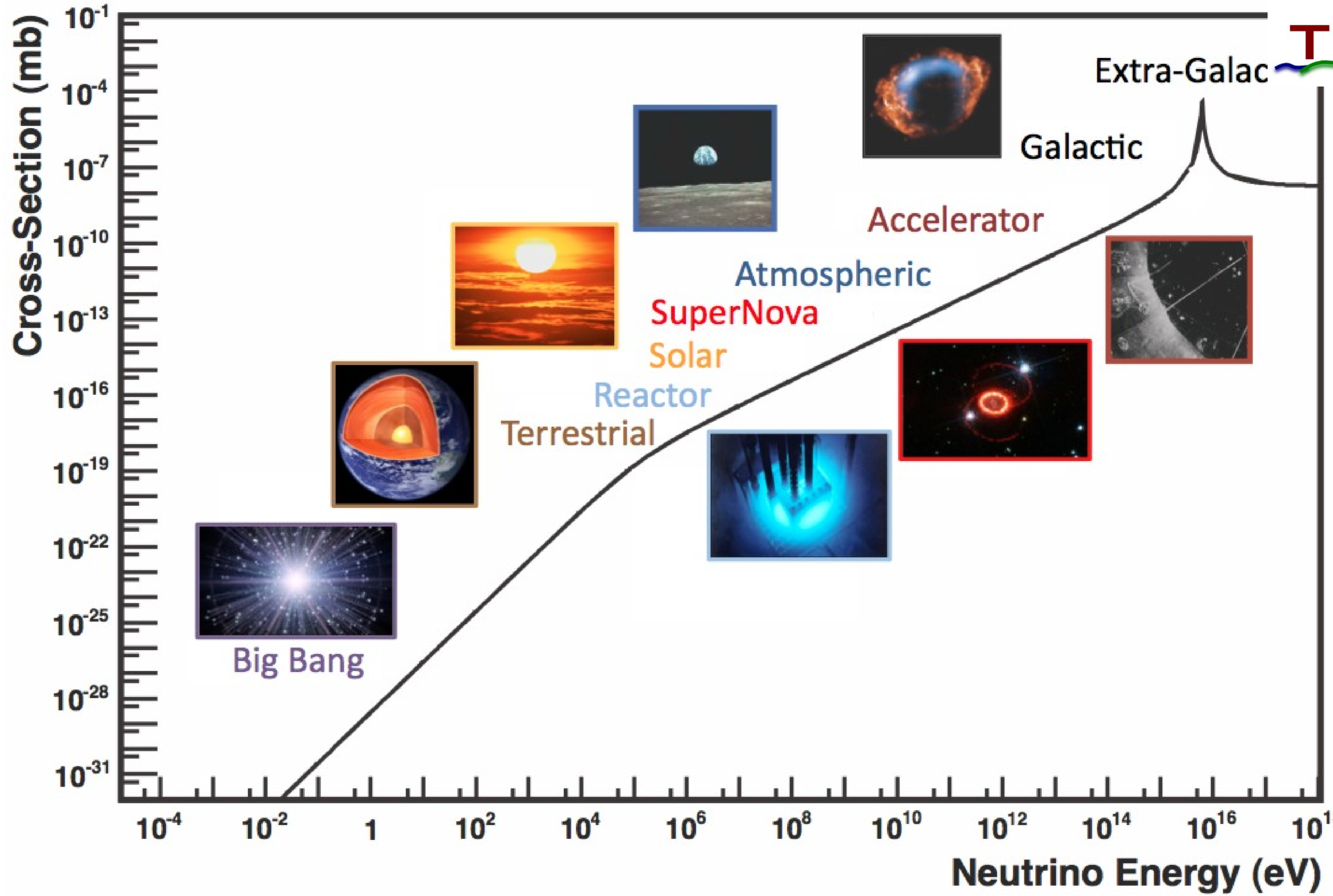
SK e Vs mu PID



Super Kamiokande IV 1294.7 days : Monitoring
e-like 2885
muon-like 2837
CCQE electron
300
250
200
150
100
50
0
-10
-8
-6
-4
-2
0
2
4
6
8
10
PID likelihood sub-GeV 1ring (FC)



Flux (cm⁻² s⁻¹ sr⁻¹ MeV⁻¹)
10^24
10^20
10^16
10^12
10^8
10^4
1
10^-4
10^-8
10^-12
10^-16
10^-20
10^-24
10^-28
Cosmological ν
Solar ν
Supernova burst (1987A)
Reactor anti-ν
Background from old supernovae
Terrestrial anti-ν
Atmospheric ν
ν from AGN
Cosmogenic ν
10^-6 μeV
10^-3 meV
1 eV
10^3 keV
10^6 MeV
10^9 GeV
10^12 TeV
10^15 PeV
10^18 EeV
Neutrino energy



Cross-Section (mb)
10^{-1}
10^{-4}
10^{-7}
10^{-10}
10^{-13}
10^{-16}
10^{-19}
10^{-22}
10^{-25}
10^{-28}
10^{-31}
10^{-4}
10^{-2}
1
10^{2}
10^{4}
10^{6}
10^{8}
10^{10}
10^{12}
10^{14}
10^{16}
10^{18}
Neutrino Energy (eV)
Extra-Galac
Galactic
Accelerator
Atmospheric
SuperNova
Solar
Reactor
Terrestrial
Big Bang