TRIUMF

Machine Learning Techniques for Water Cherenkov Event Reconstruction

N. Prouse, TRIUMF

CAP Congress, 7th June, 2021

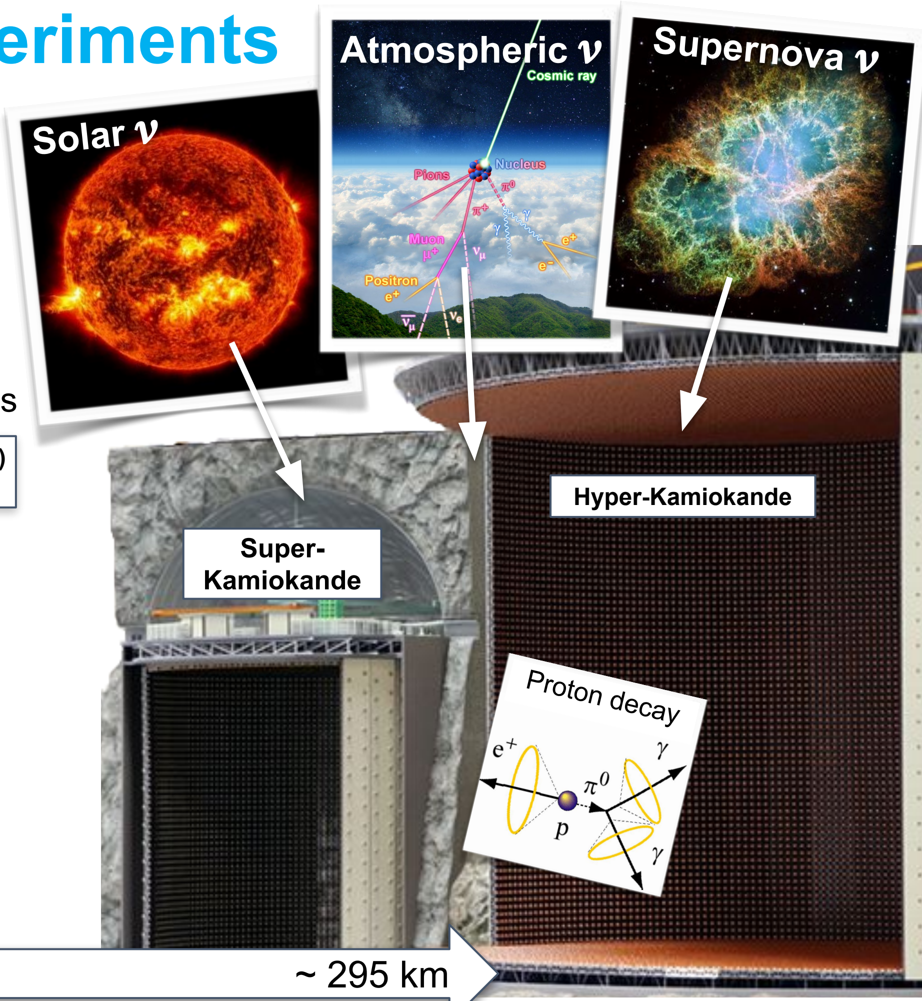
The Super-K & Hyper-K Experiments

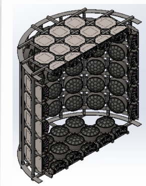
Current generation **Super-K** and next generation **Hyper-K** are world-leading neutrino experiments

Broad & ambitious physics programmes covering many neutrino sources and proton decay measurements

Water Cherenkov detector technology provides huge target mass with excellent particle ID and reconstruction capabilities

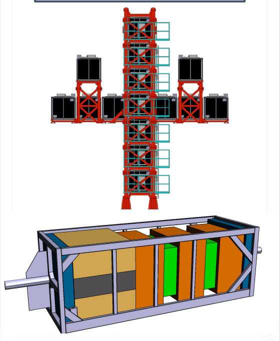
See also talks by: P. de Perio (T2K, SK & HK, Tue), M. Yu (T2K CCπ^0, Thu) B. Jamieson (Photogrammetry, Thu), V. Gousy-Leblanc (PTF, Thu)



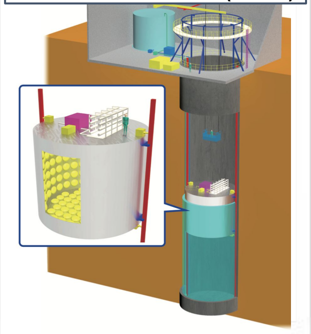


Water Cherenkov Test Experiment (WCTE) at CERN

Near detectors



Intermediate Water Cherenkov Detector (IWCD)

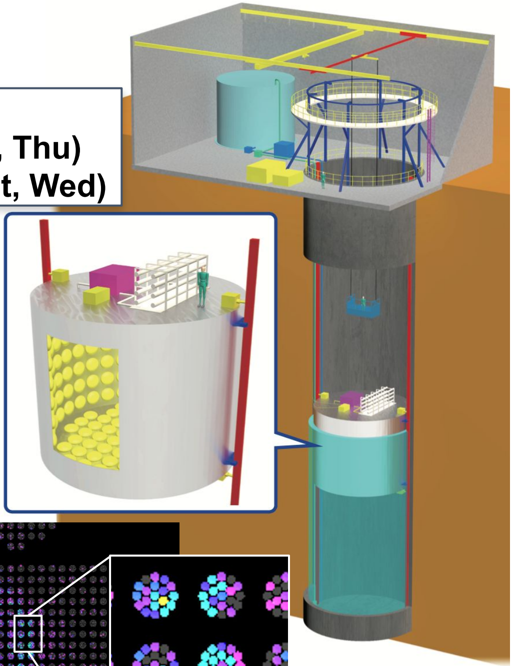


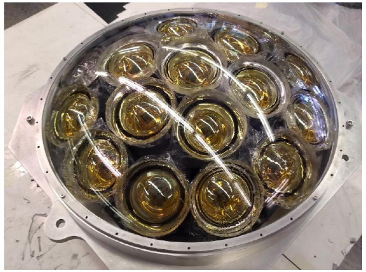
280 m → ~ 1 km → ~ 295 km

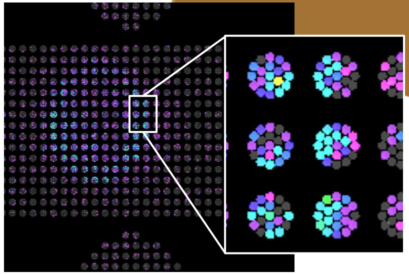
The Intermediate Water Cherenkov Detector

- Measures of flux and cross-section of mostly un-oscillated beam to reduce systematics at far detector
- Located ~ 1 km from ν beam source
- Moves vertically in ~50 m tall pit
  - spans range of angles off axis from ν beam for different ν energy spectra
- 6 m tall x 8 m diameter surrounded with ~ 500 multi-PMT modules (mPMTs)
  - 8 cm PMTs: Better position resolution
    < 1 ns timing resolution
  - Additional directionality information
  - mPMTs will also be used for WCTE
  - Also in consideration for portion of far detector photo-coverage

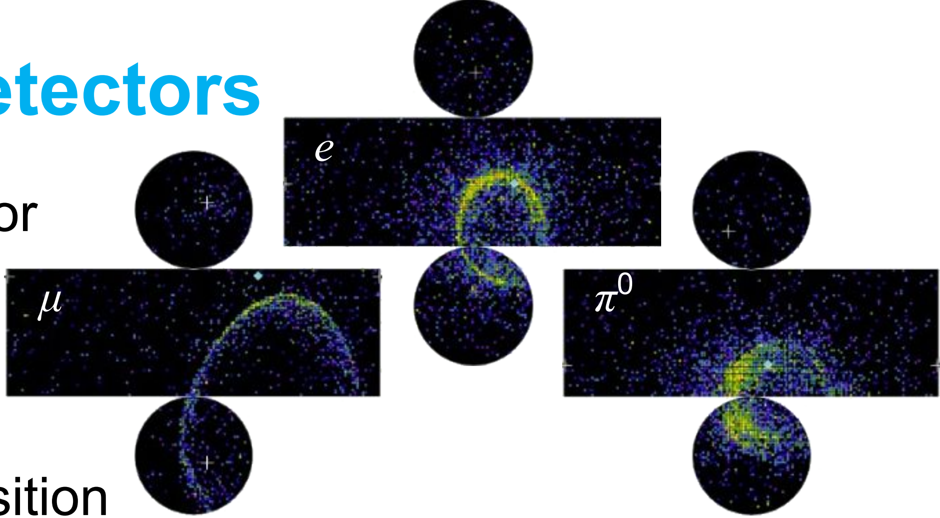
See also talks by
R. Akutsu (neutrons in IWCD, Thu)
L. Koerich (mPMT development, Wed)





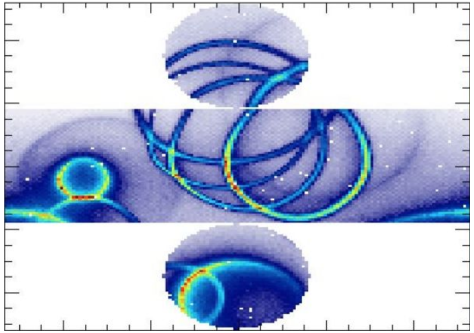


Reconstruction for WC detectors



Reconstruction software is essential for

- Particle type identification
  - Separate signal events from background
- Particle momentum, direction, position
  - Kinematics essential to determine incoming neutrino energy
  - Neutrino energy affects oscillation probability, for oscillation parameter measurements
  - Kinematics also useful for signal / background classification
- Separating & reconstructing multi-ring events
  - Events with multiple particles / rings contribute to both signal & background
  - Pile-up of events will be significant in IWCD detector



Machine learning reconstruction for WC

Limit of traditional maximum-likelihood reconstruction methods (fiTQun) is being reached

- Computation time is becoming a limiting factor
  - Larger far detector with more PMTs
  - Smaller intermediate detector requires scaled down resolutions
  - Improving resolutions requires more complex algorithms with fewer approximations

ML and deep neural networks have potential to push reconstruction further

- Very successful in areas of computer vision and image processing
- Becoming common in HEP applications beyond just e.g. event selections
- Potential to use all information without detector model approximations
- Very fast to run once neural networks have been trained
  - fiTQun on CPU: < 1 event per minute
  - ML reconstruction on GPU: 100,000 events per minute

See also talk by W. Fedorko (ML in particle physics, Thu)

Particle type classification

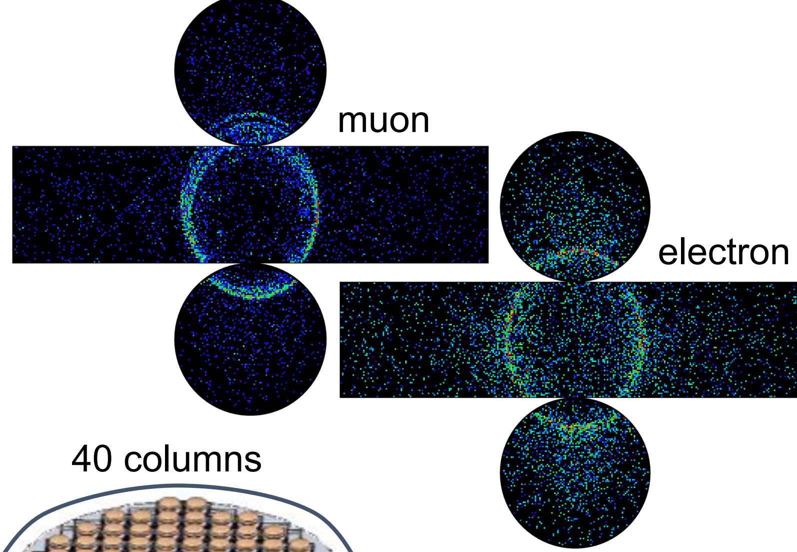
Initial studies to classify $\mu/\pi^0/e/\gamma$ particle types

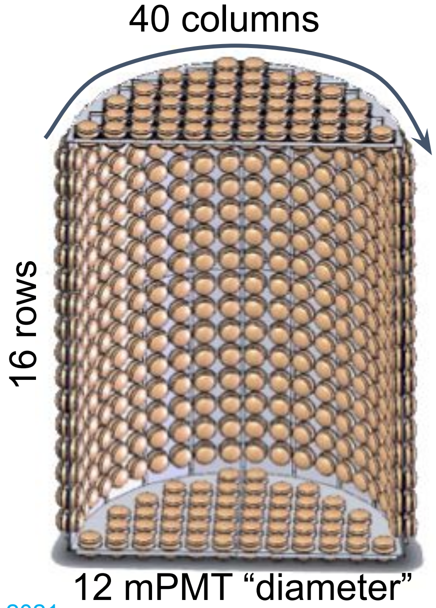
- μ vs e is classified extremely well by traditional methods (>99% accuracy)
- e vs π^0 works reasonably well, but could be improved
- e vs γ has not been used successfully with traditional methods

Simulated 3M of each in IWCD detector

- 0 - 1 GeV energy above threshold
- Uniform positions, isotropic directions
- Vertical and horizontal reflections for data augmentation

Exploring various network architectures



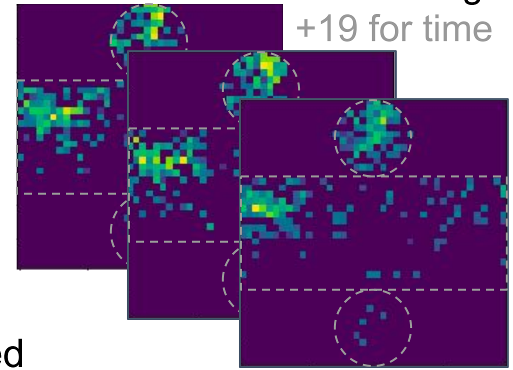


mPMT: 19 x 3” PMTs

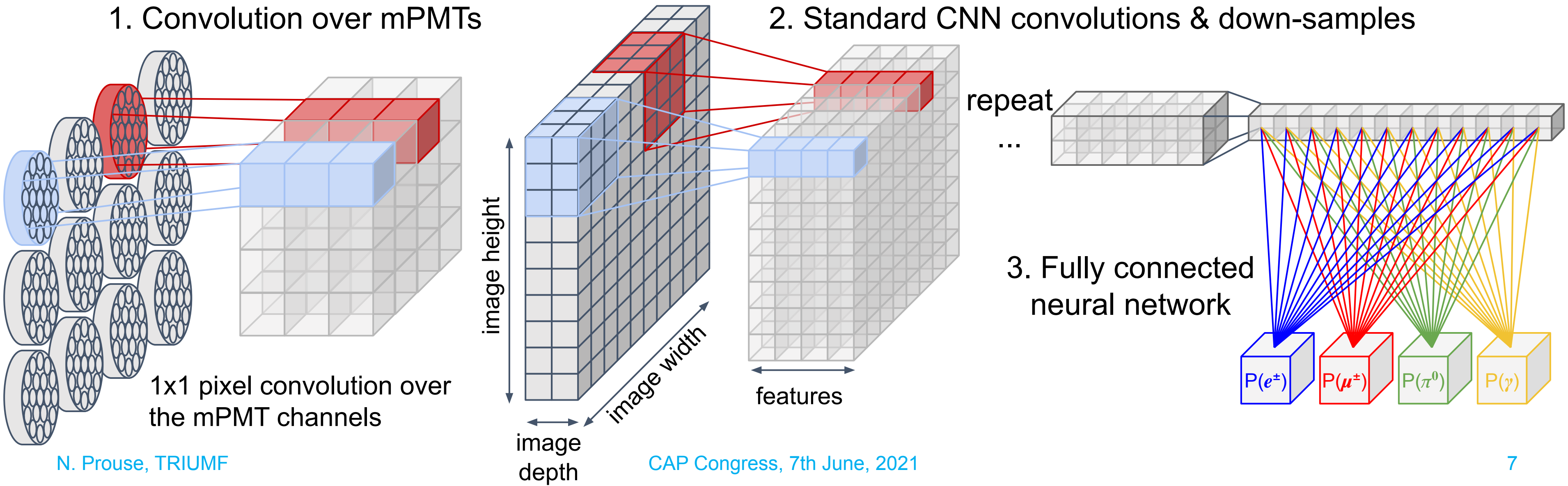
CNN architecture

Full cylinder of mPMTs is unwrapped onto flat image

- One pixel per multi-PMT
- Charge (& time) of 19 PMTs per mPMT
- No special treatment at barrel / end-cap boundary
  - Alternative projections from cylinder to grid have also been explored



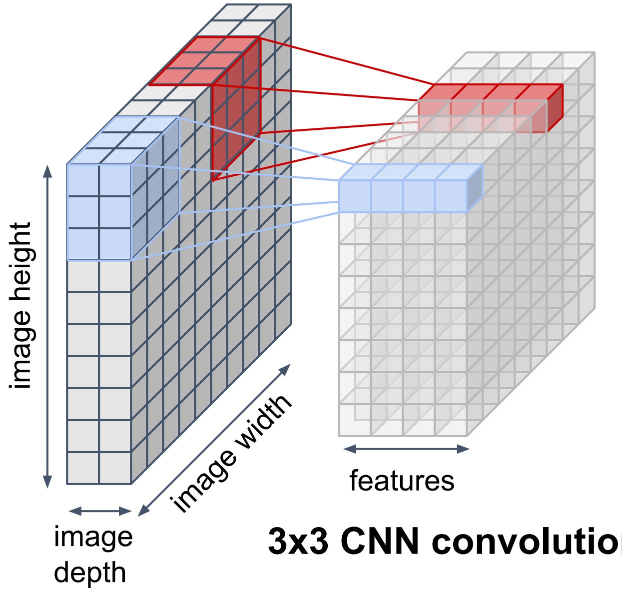
Network based on ResNet-18 CNN architecture[arXiv:1512.03385]



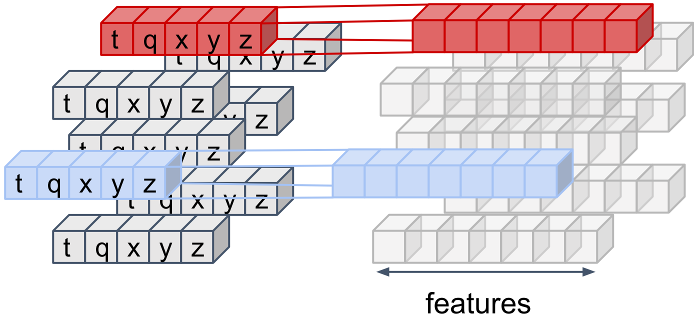
PointNet architecture

PointNet is designed to work on ‘point clouds’ rather than images

- Each hit PMT is a ‘point’ with time, charge & position, not fixed to grid
    - CNN learns translation-invariant functions on image
    - PointNet learns symmetric functions on point clouds
        - Symmetric: ordering of points cannot affect outcome
- Convolution-like operations act on each point’s charge, time and position
- Information flows between points by learning global transformations applied to all points
- Can apply to any detector geometry



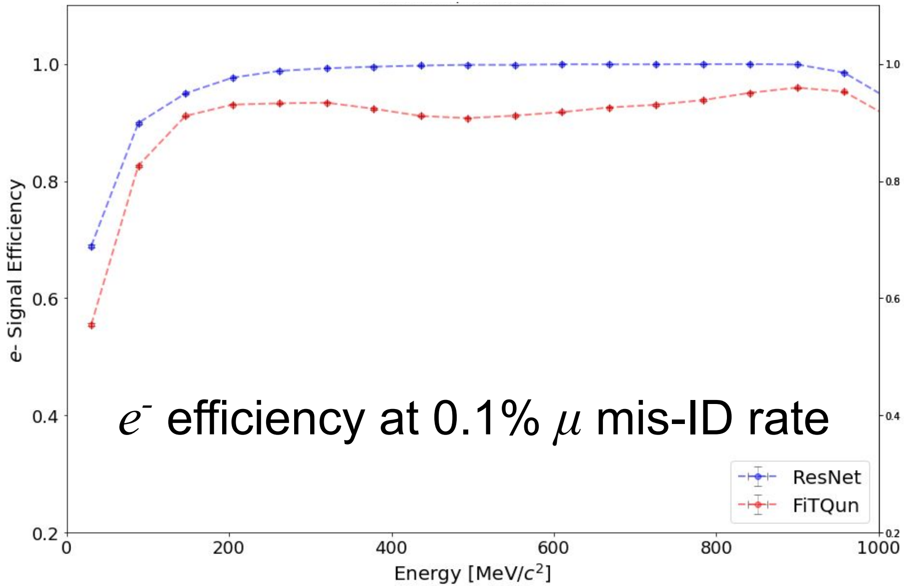
3x3 CNN convolution

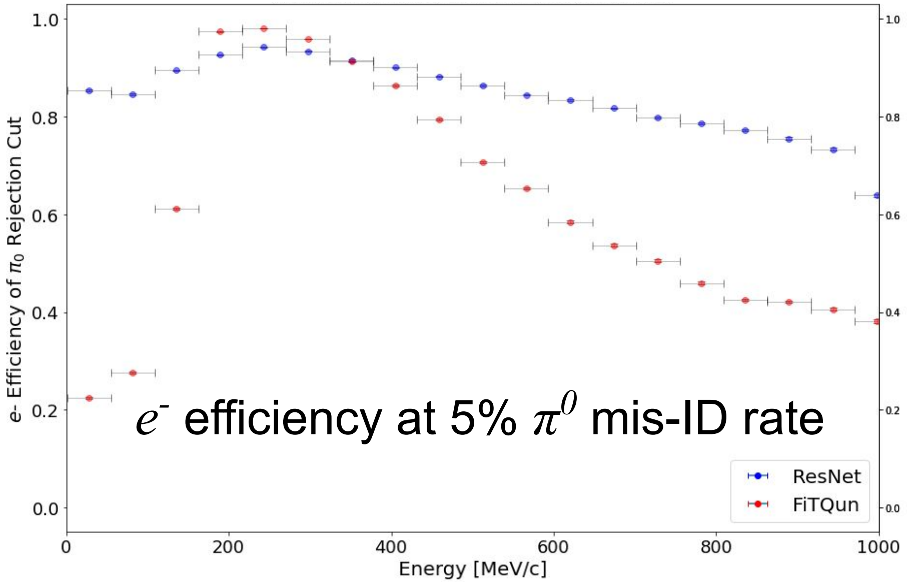


PointNet MLP (1x1 convolution on point cloud)

Classification results

Comparison of ResNet to traditional maximum-likelihood method (fiTQun)



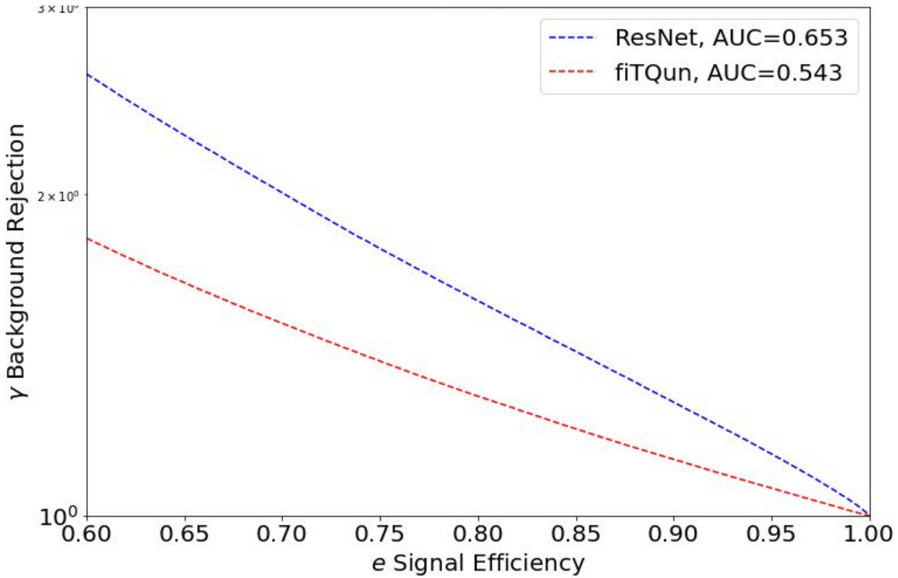


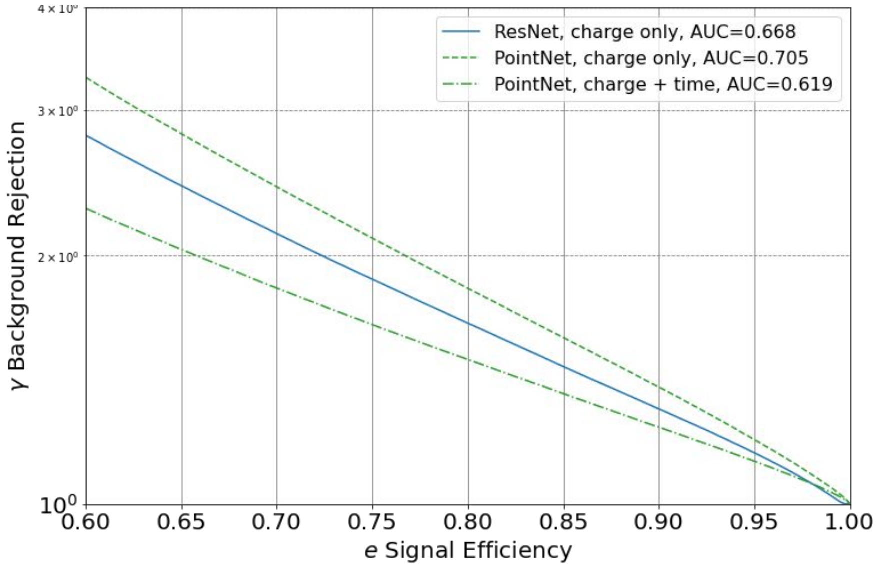
- ν_μ beam produces mostly μ, need rejection factor of 1000 for ν_e measurement
- Increased e^-/μ discrimination across energies

- π^0 is significant background to e^- signal
- Increased e^-/π^0 discrimination, particularly at challenging energies

Classification results

γ and e^- almost indistinguishable in water Cherenkov detectors



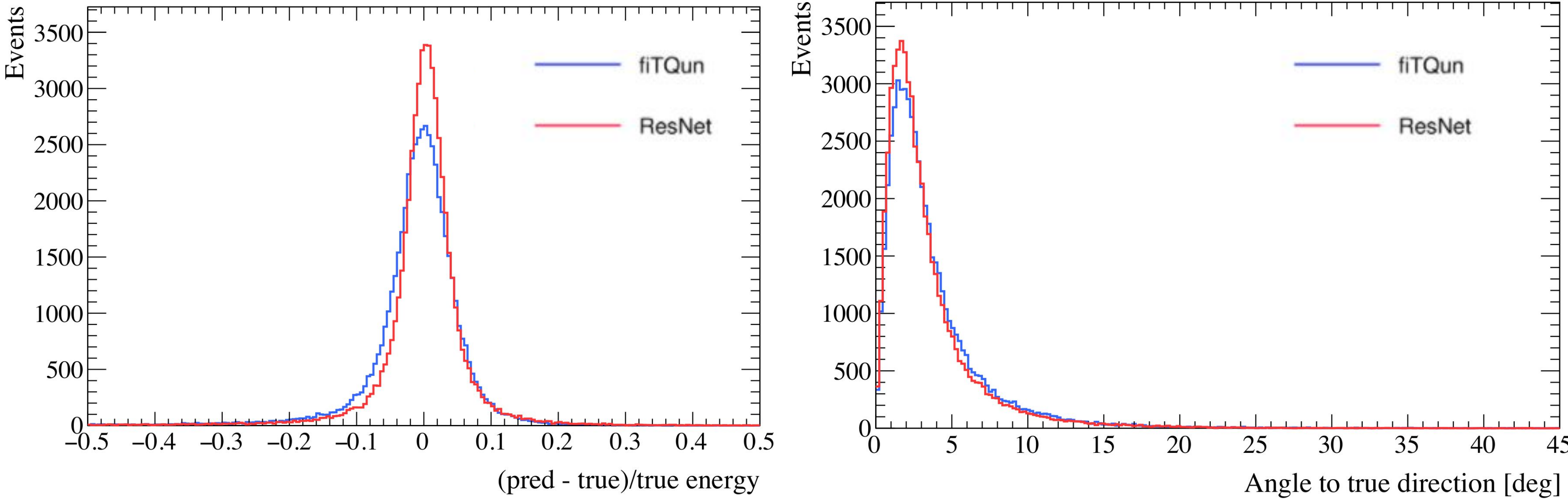


- e^- / γ discrimination with fiTQun not been successfully used
- Statistical separation significantly improved with ResNet
- ResNet has only been used with charge channels so far
- PointNet with charge+time gives significant advantage

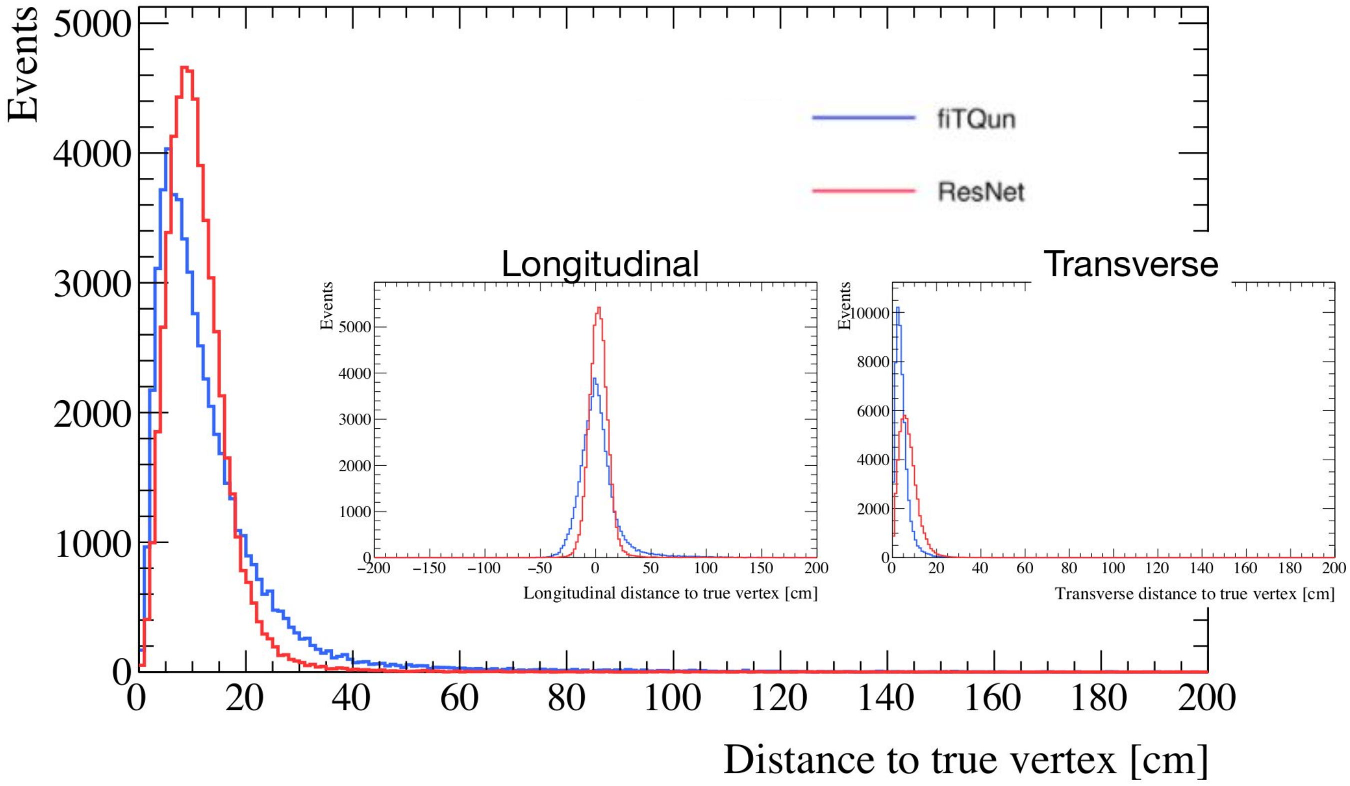
Position, direction, energy reconstruction

Similar ResNet architecture as used for classification

- Output reconstructed quantities instead of classification variables
- Use Huber loss to minimise true-reconstructed residuals
- ResNet is outperforming fiTQun at energy and direction reconstruction



Position, direction, energy reconstruction

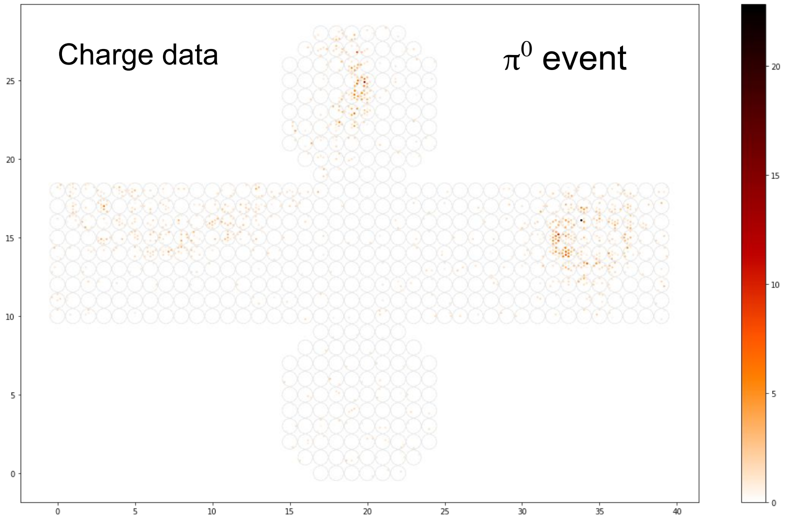


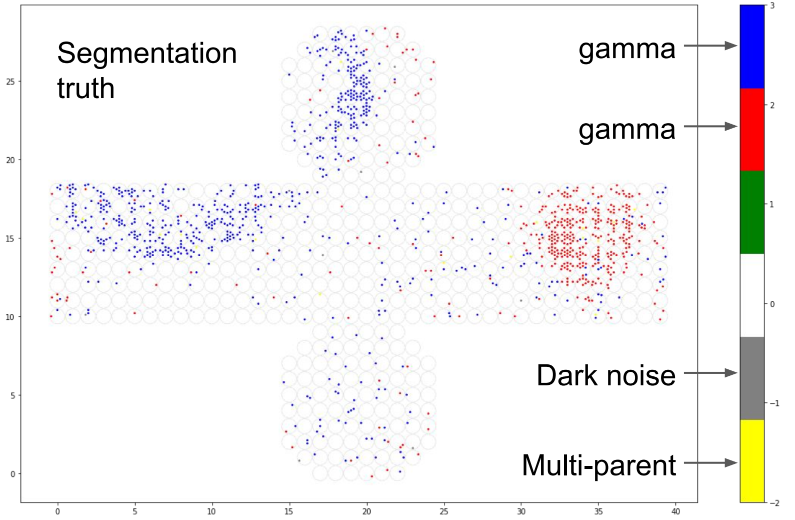
- ResNet is outperforming fiTQun overall at position reconstruction
  - Better in longitudinal direction (along direction of particle track)
  - But worse in transverse direction

Segmentation networks

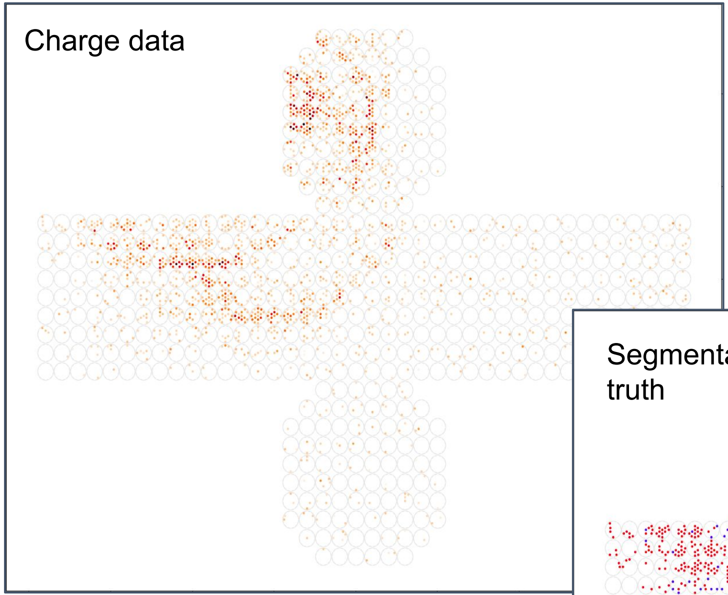
Need to separate out complex multi-ring events or multi-vertex pile-up events

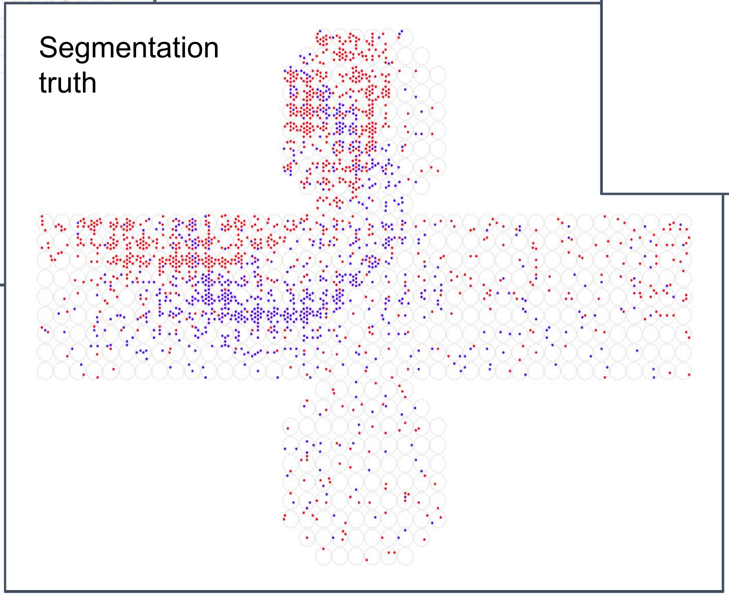
- Classification networks can be extended to perform segmentation
  - Encoding part of the network is the similar to classification
  - Segmentation part of the network provides output for each pixel
    - Deconvolutions and upsampling reverse convolutions and downsampling
  - Currently using U-Net and FRRN
- Starting development with π^0 events
  - π^0 decay to produce two γ rings
  - Higher energy π^0 have overlapping rings

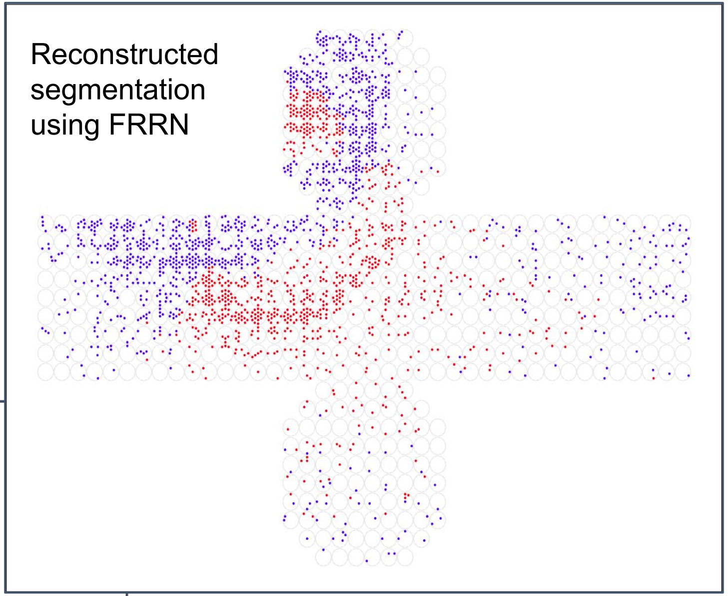




Segmentation results

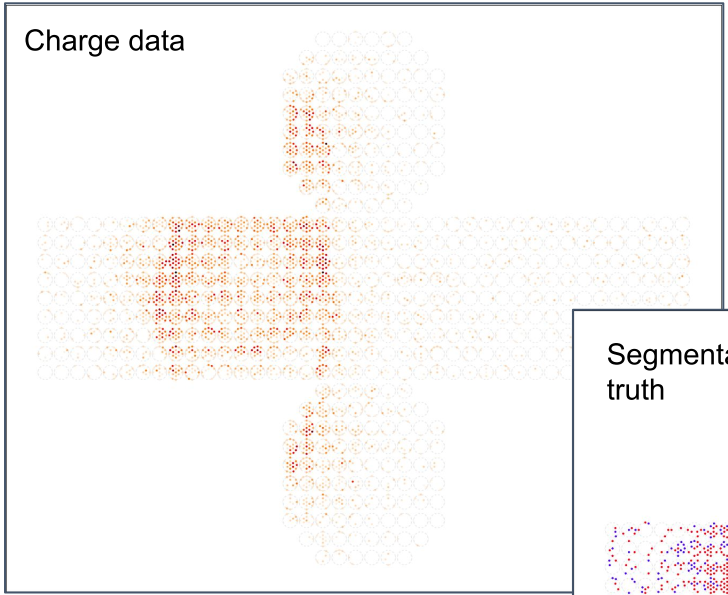


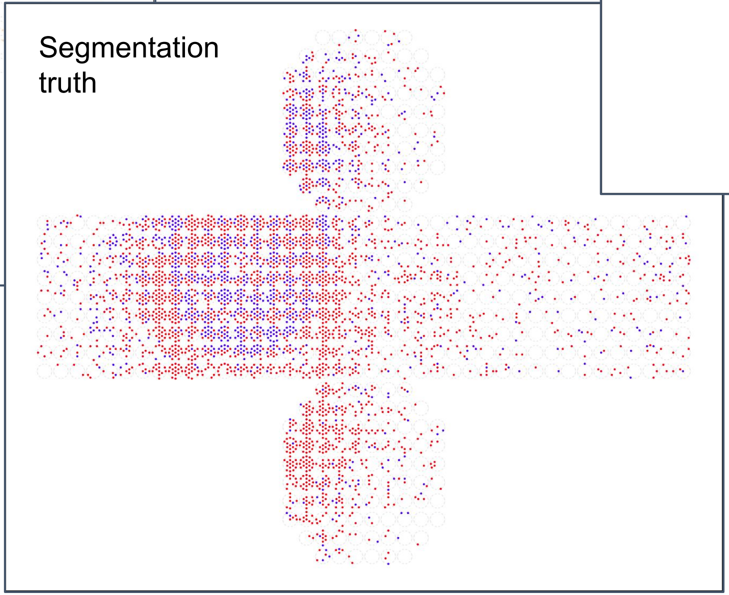


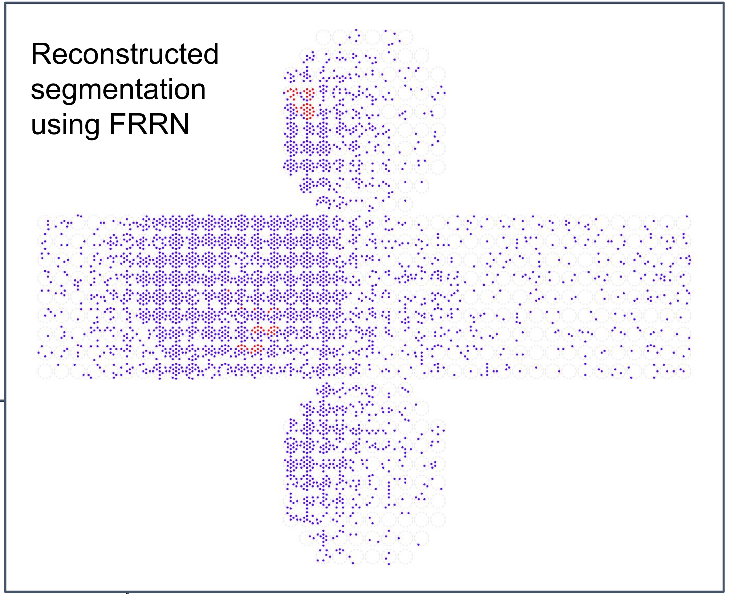


Works well with separated or partially overlapping rings

Segmentation results







Poor reconstruction with some more overlapping rings

Summary

Hyper-kamiokande, the next-generation water Cherenkov neutrino detector has begun construction to start operation in 2027

- Both the far detector and IWCD will require new techniques to improve reconstruction, suppress backgrounds and reduce systematics

Machine learning can bypass the model approximations of old methods

- ResNet CNN and PointNet architectures already outperforming traditional methods
  - Improved reconstruction of particle position, direction and energy
  - Classification of particle types improves on existing selections and enables new analyses
- Additional benefit of huge increase in speed of reconstruction

Exploring other areas where machine learning can provide benefits

- Segmentation of multi-ring events looks promising
- Extending IWCD studies to Super-K and Hyper-K far detectors
- More ideas and studies in the pipeline

Hyper-Kamiokande

WatChMaL.org

Appendix

Machine learning reconstruction

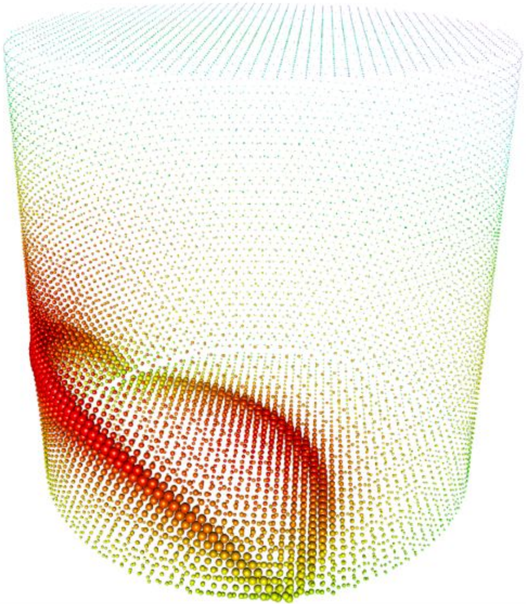
WatChMaL: cross-collaboration group formed to explore ML for WC

Common challenges for ML with WC detectors

- Cylindrical geometry
- High-resolution, sparse data

Many physics goals

- Maximise precision of new detectors
- Reconstruct complex event topologies
- Discriminate electron and gamma rings
- Improving detector calibration & systematics



WatChMaL.org

The Hyper-K Experiment

February 2020: Budget approved by Japanese government

May 2020: Univ. of Tokyo President and KEK Director General signed MOU:

Univ. of Tokyo to construct & operate Hyper-K detector

KEK to upgrade & operate J-PARC neutrino beam

Hyper-K's WC detectors

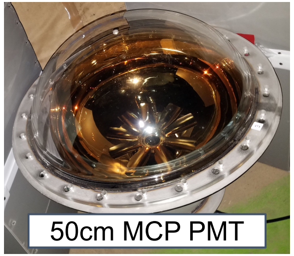
Hyper-K far detector

3rd generation of WC detectors at Kamioka

8 x increase in fiducial mass over Super-K
72 m tall x 68 m diameter = 258 kt total mass
188 kt fiducial mass

Baseline design: 40,000 B&L 50 cm PMTs
= 40% photo-coverage

New photo-detector technology to provide increased sensitivity



Hyper-K's WC detectors

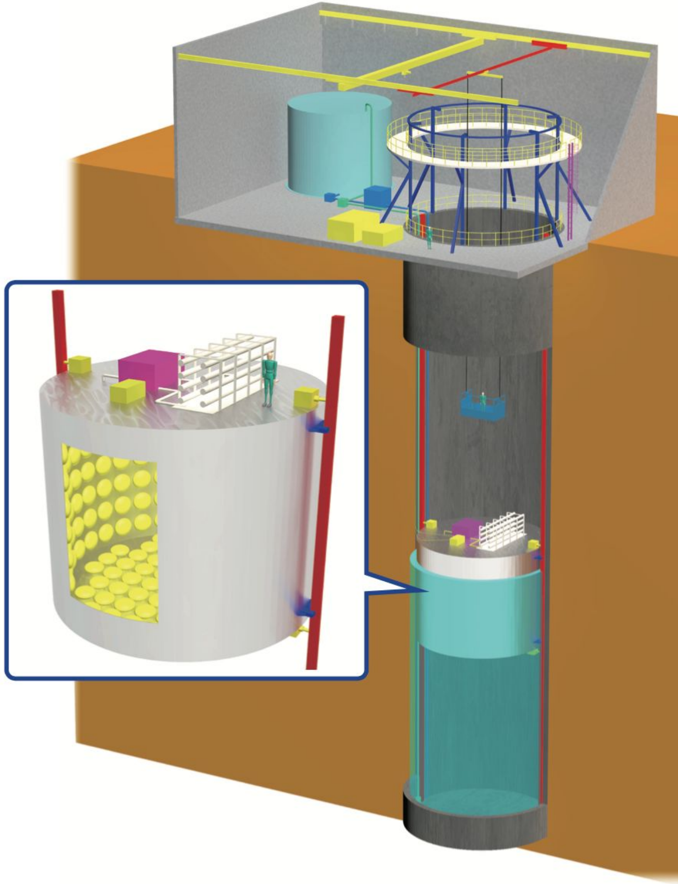
Intermediate detector (IWCD)

Located ~ 1 km from beam source
6 m tall x 8 m diameter inner detector
~ 500 multi-PMT modules

Measure combination of flux and cross-section to reduce systematics at far detector

High event rate, same detector technology and target nuclei as far detector

Moves vertically in ~50 m tall pit
measuring different off-axis angles gives different ν energy spectra



Hyper-K's WC detectors

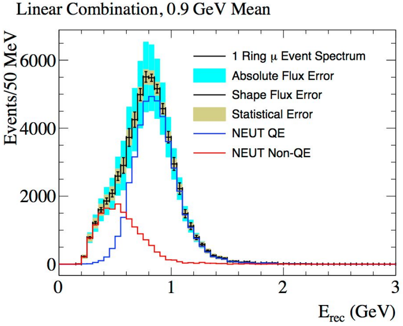
Off-axis spanning detector

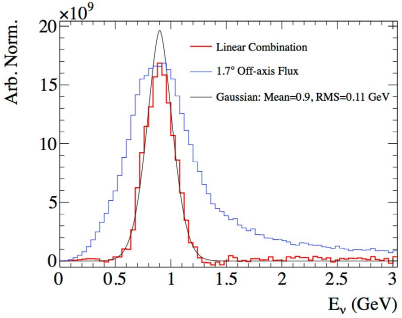
ν energy spectrum depends on angle off-axis to the neutrino beam

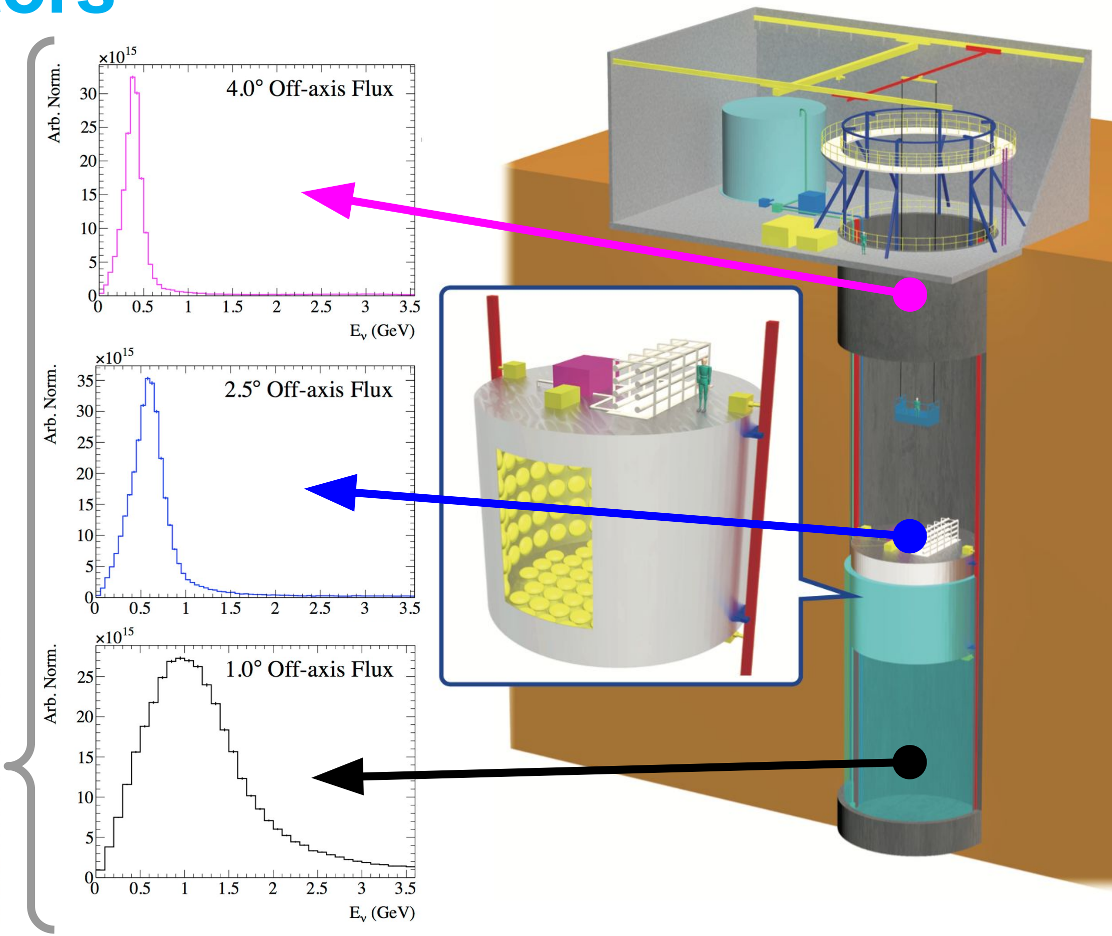
Far detector @ 2.5° for peak at ~600 MeV

Moving IWCD varies angle, allowing measurements at different energies

Linear combinations allows mimicking mono-chromatic beam or far-detector spectrum







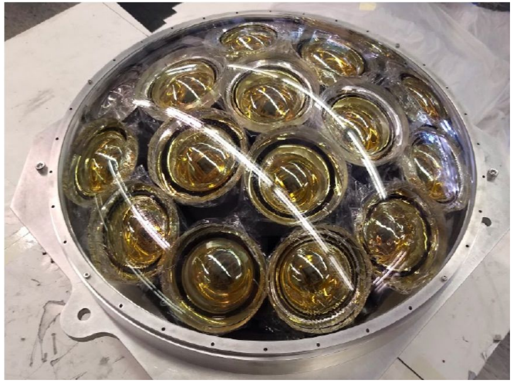
Hyper-K's WC detectors

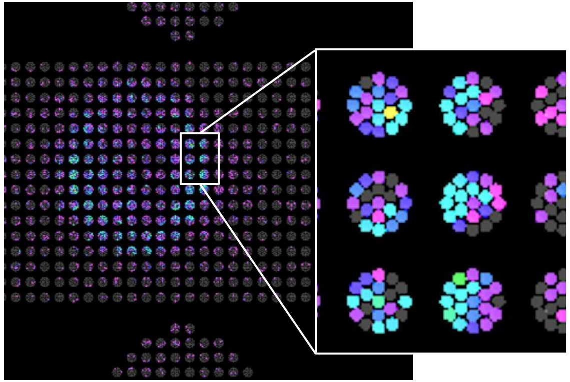
Multi-PMT modules

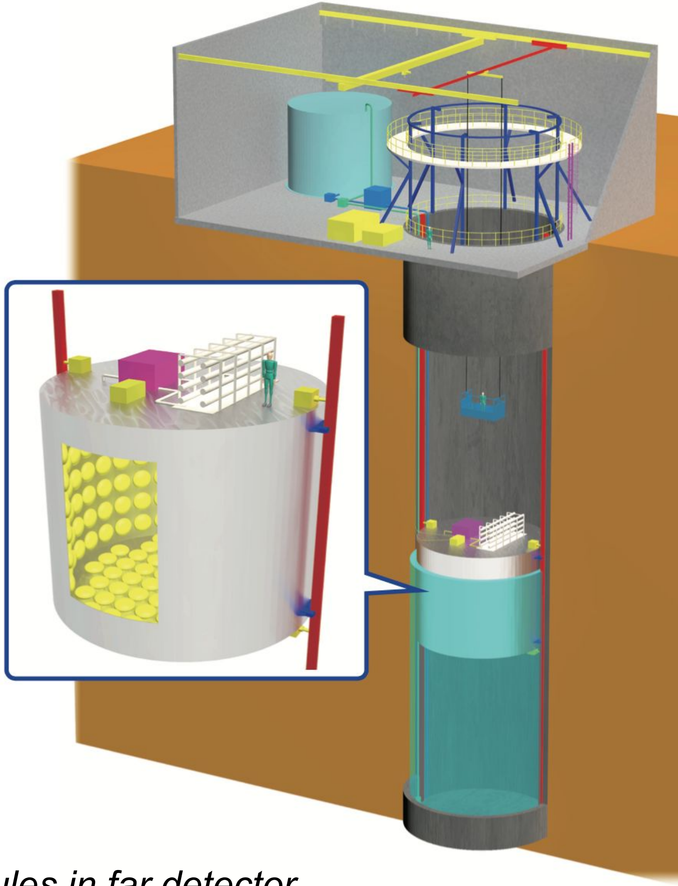
8 cm PMTs: Better position resolution
< 1 ns timing resolution
Additional directionality information

Need reconstruction to exploit additional information

Necessary for smaller detector size



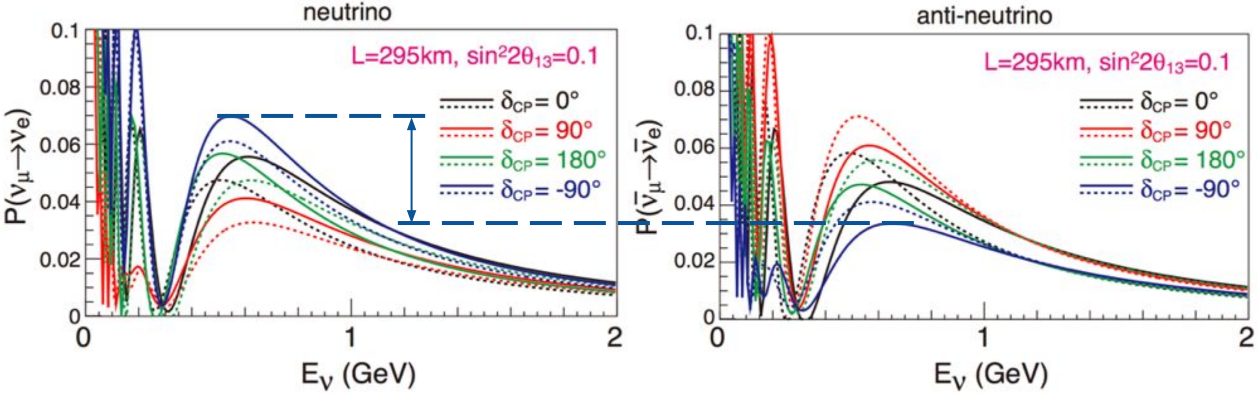




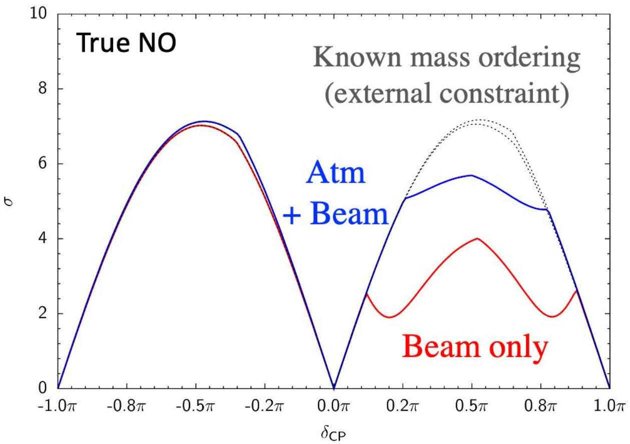
Also under investigation: Combining 50 cm PMTs + multi-PMT modules in far detector

Hyper-K's physics goals

Long-baseline neutrino oscillations: CP violation



10 years with 1.3MW, T2K 2018 systematic error

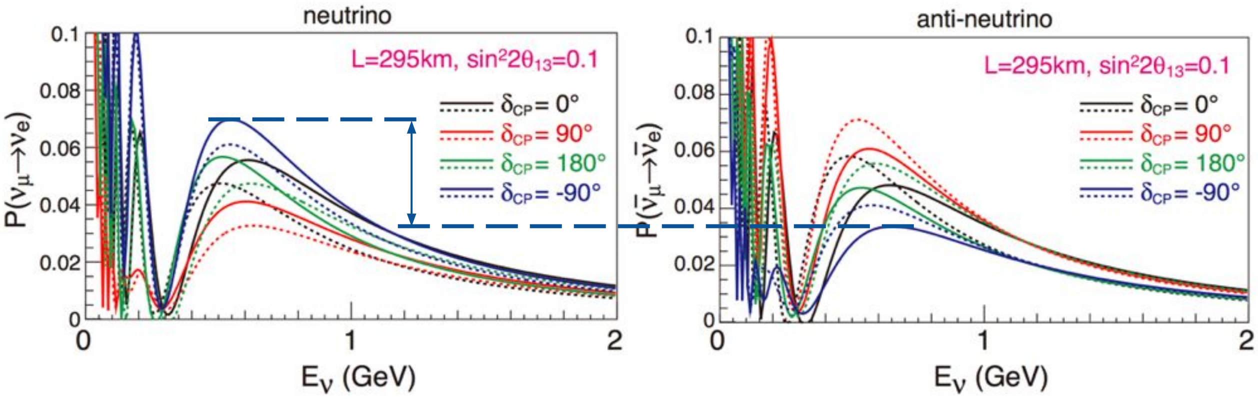


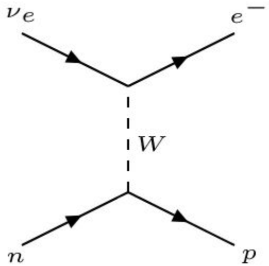
Combine beam and atmospheric neutrino observations for maximum sensitivity

- δ_{CP} precision comes mostly through difference in $P(\nu_\mu \rightarrow \nu_e)$ vs $P(\bar{\nu}_\mu \rightarrow \bar{\nu}_e)$
- Effect of δ_{CP} can be degenerate with normal vs inverted mass ordering
- Atmospheric ν's gain sensitivity to mass ordering by exploiting matter effect of Earth on oscillations

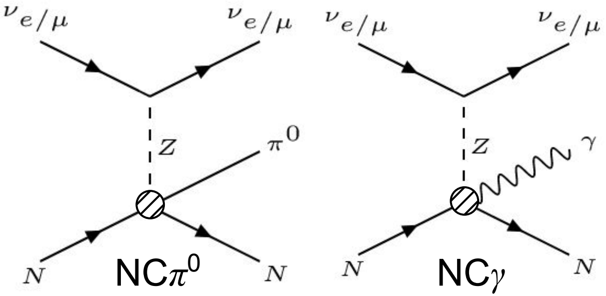
Hyper-K's physics goals

Long-baseline neutrino oscillations: CP violation





CCQE signal



backgrounds

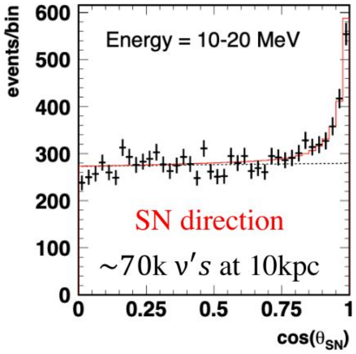
Oscillation maximum is at around 0.6 GeV

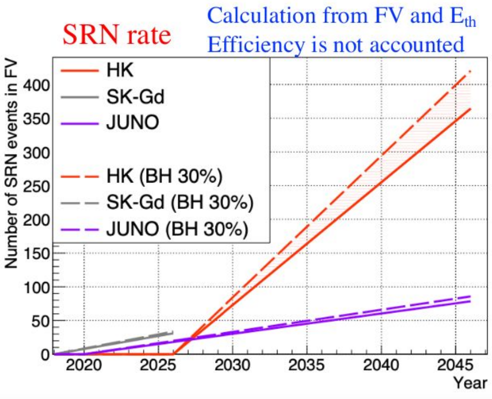
- Dominant signal ν_e interaction is charged current quasielastic (CCQE)
- Potential background sources:
  - Neutral current interactions (ν_e or ν_μ) producing neutral pions or gammas
  - Muons from ν_μ misidentified as electrons from ν_e

Hyper-K's physics goals

Neutrino astrophysics

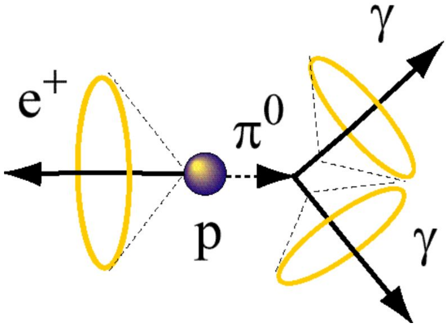
- Solar *ν*'s: day/night asymmetry; hep *ν*'s; ^{8}B *ν* spectrum upturn
- Supernova *ν*'s: 1000's *ν* events for nearby supernova pointing, time & spectrum analysis; search for supernova relic *ν*'s





Proton decay

- Search to order of magnitude greater lifetime than current limit
- 10^{35} years for $p \longrightarrow e^{+} + \pi^{0}$
- $3{\times}10^{34}$ years for $p \longrightarrow \bar{\nu} + K^{+}$

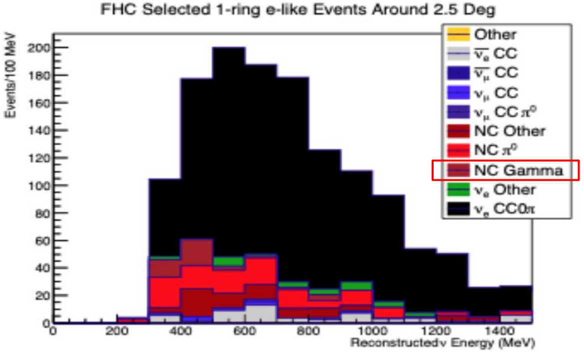


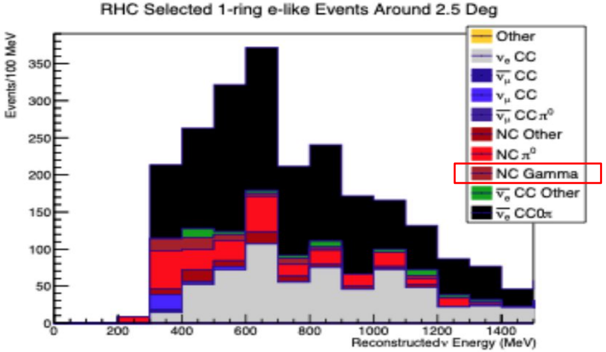
Physics Motivations

New opportunities beyond simple reconstruction improvement

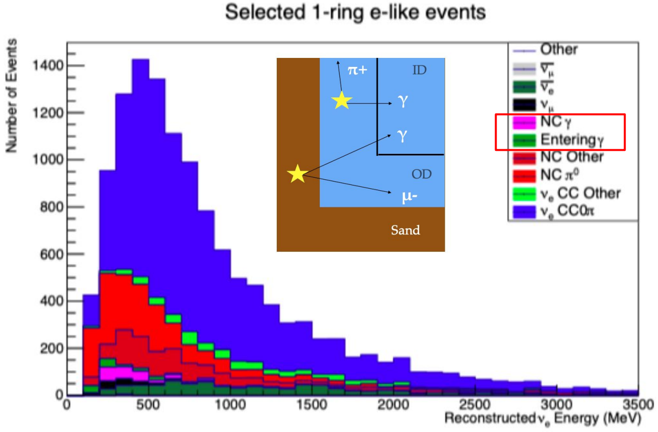
- NC γ discrimination and measurement

Systematic Type	Requirement	Motivation
$\boldsymbol{\sigma(\nu_e)/\sigma(\nu_\mu)}$ to $\boldsymbol{\sigma(\bar{\nu}_e)/\sigma(\bar{\nu}_\mu)}$ ratio	2.9-3.7%	CP violation search and precision δ_{CP} measurement.
$\boldsymbol{\sigma(\nu_e)/\sigma(\nu_\mu)}$, $\boldsymbol{\sigma(\bar{\nu}_e)/\sigma(\bar{\nu}_\mu)}$	3-5%	θ_{23} octant and precision θ_{23} measurement.
Intrinsic $\nu_e, \bar{\nu}_e$ and NC background normalizations	3-5%	CP violation search and precision δ_{CP} measurement.





2.7-4.0 degree off-axis range



- Potential neutron tagging application

- Bottom-up calibration: Enable multitude of detector parameter variations

Traditional reconstruction method

fiTQun: Likelihood-based reconstruction for higher energies

- Originally developed for Super-K detector
  - Based on algorithm of MiniBooNE: https://arxiv.org/abs/0902.2222
- Uses full information of unhit PMTs + time & charge of hit PMTs:

$$\boxed{L}(\boxed{\mathbf{x}}) = \prod_{j}^{unhit} \boxed{P_j\left(unhit|\mathbf{x}\right)} \prod_{i}^{hit} \boxed{P_i\left(hit|\mathbf{x}\right)} \boxed{f_q\left(q_i|\mathbf{x}\right)} \boxed{f_t\left(t_i|\mathbf{x}\right)}$$

Likelihood to maximise

Candidate event hypothesis

Probability of no hit at PMT

Probability of hit at PMT

Hit charge probability density

Hit time probability density

- Probabilities calculated based on direct + scattered + reflected light
- Likelihood ratios used to distinguish particle types and single-ring / multi-ring event topology hypotheses

Machine learning reconstruction

WatChMaL: cross-collaboration group formed to explore ML for WC

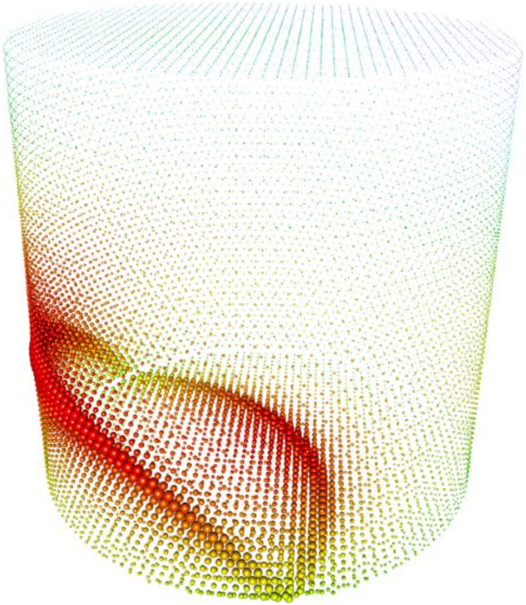
Common challenges for ML with WC detectors

- Cylindrical geometry
- High-resolution, sparse data

Many physics goals

- Maximise precision of new detectors
- Reconstruct complex event topologies
- Discriminate electron and gamma rings
- Improving detector calibration & systematics

WatChMaL.org



The IWCD detector

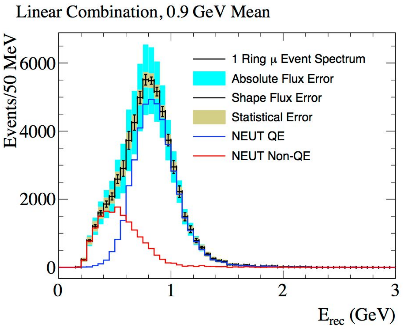
Off-axis spanning detector

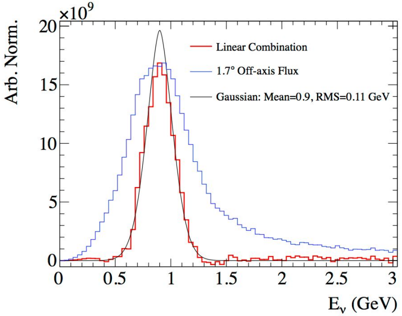
ν energy spectrum depends on angle off-axis to the neutrino beam

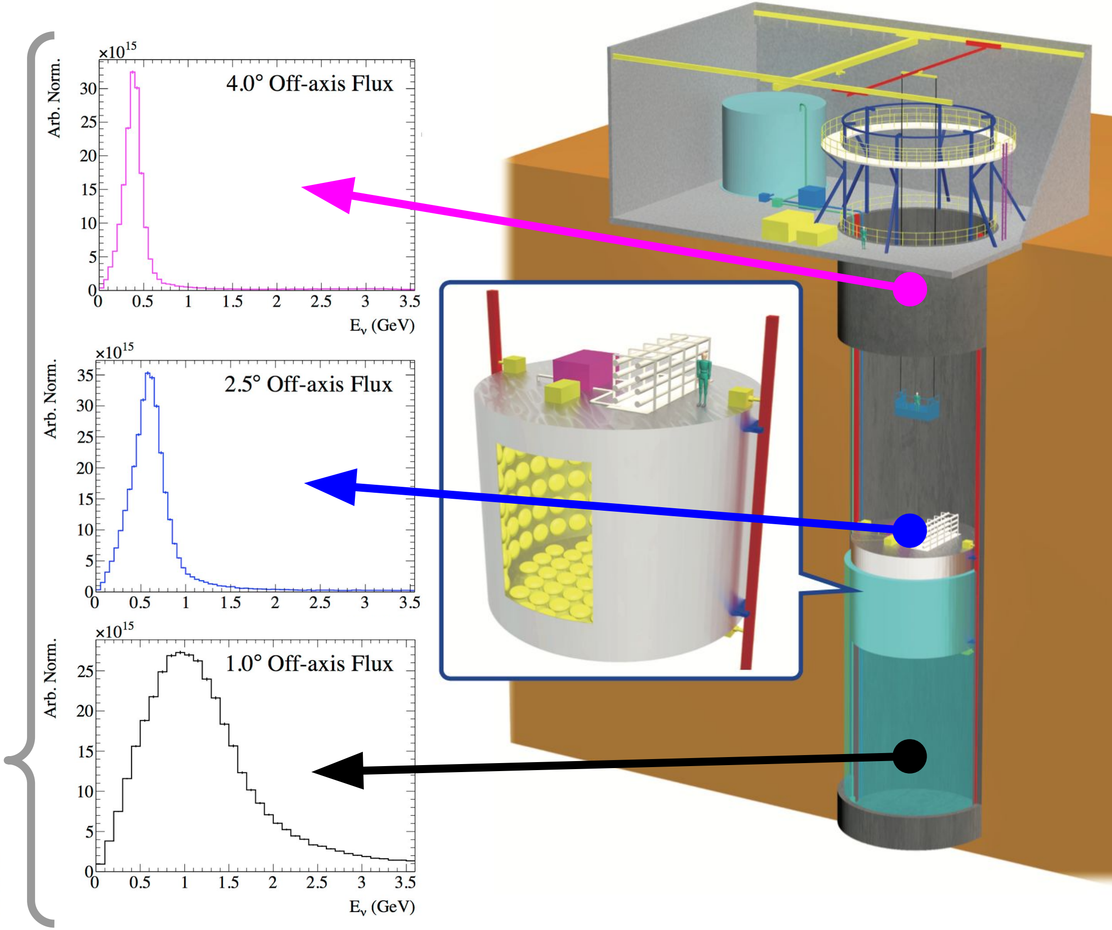
Far detector @ 2.5° for peak at ~600 MeV

Moving IWCD varies angle, allowing measurements at different energies

Linear combinations allows mimicking mono-chromatic beam or far-detector spectrum

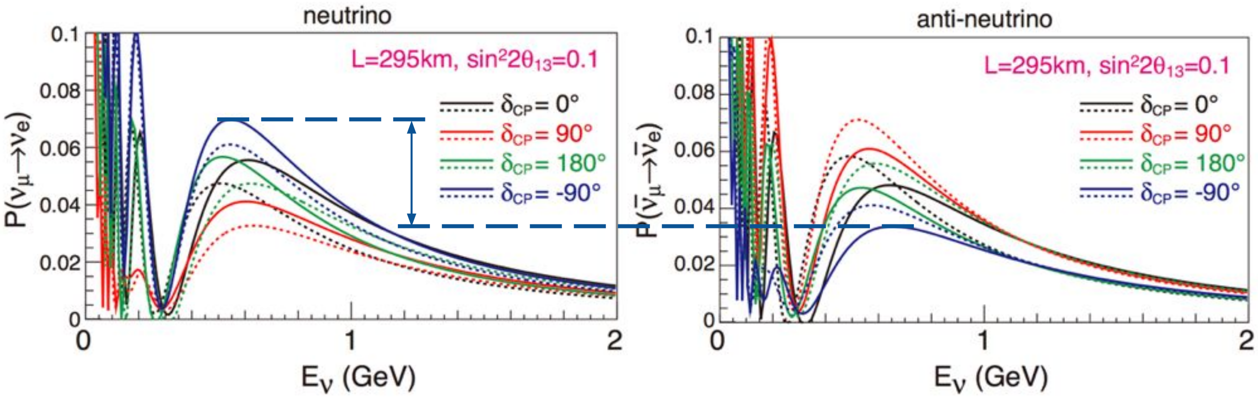




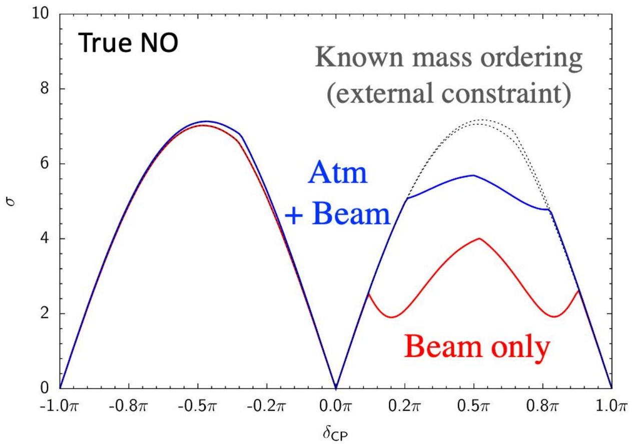


Hyper-K's physics goals

Long-baseline neutrino oscillations: CP violation



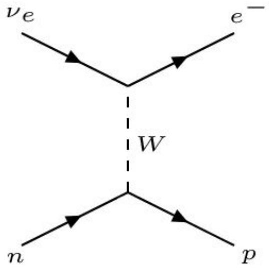
10 years with 1.3MW, T2K 2018 systematic error



Combine beam and atmospheric neutrino observations for maximum sensitivity

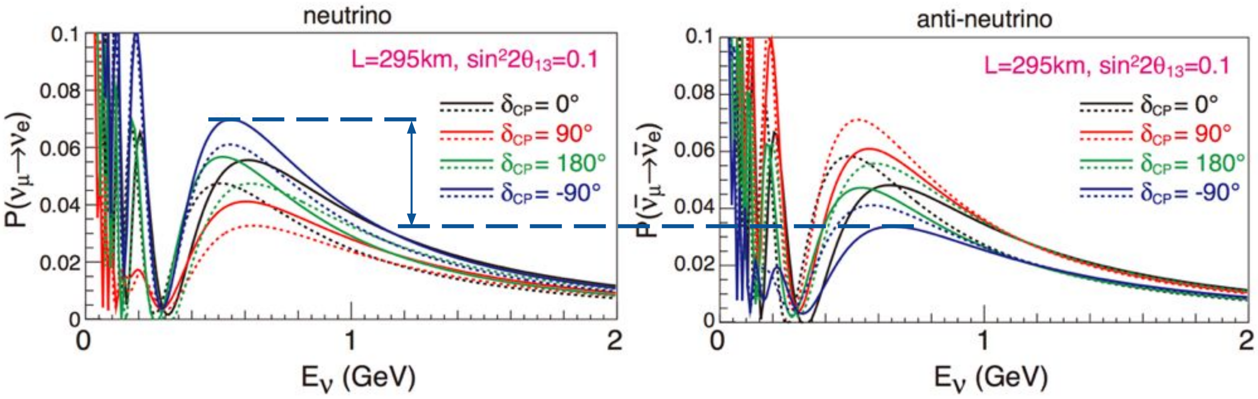
- δ_{CP} precision comes mostly through difference in $P(\nu_\mu \to \nu_e)$ vs $P(\bar{\nu}_\mu \to \bar{\nu}_e)$
- Effect of δ_{CP} can be degenerate with normal vs inverted mass ordering
- Atmospheric ν's gain sensitivity to mass ordering by exploiting matter effect of Earth on oscillations

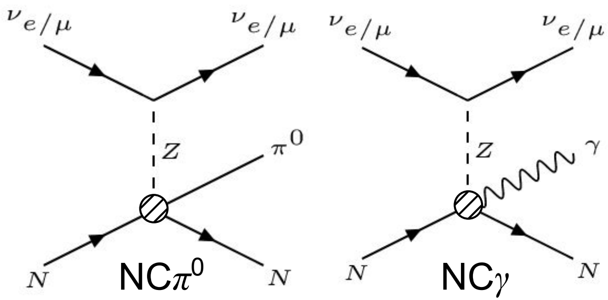
Hyper-K's physics goals



CCQE signal

Long-baseline neutrino oscillations: CP violation





backgrounds

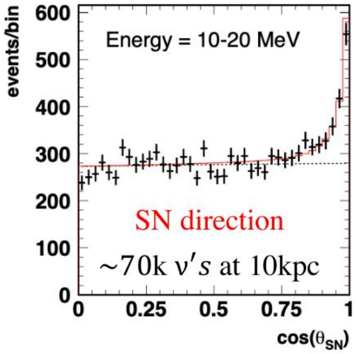
Oscillation maximum is at around 0.6 GeV

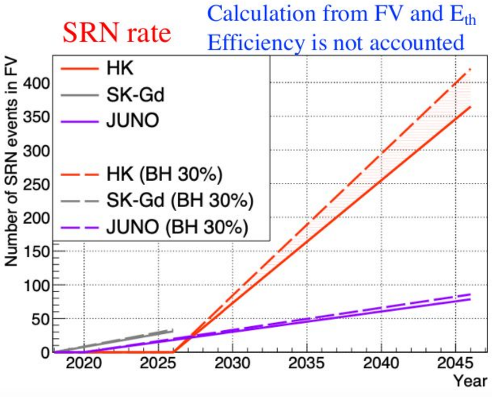
- Dominant signal ν_e interaction is charged current quasielastic (CCQE)
- Potential background sources:
  - Neutral current interactions (ν_e or ν_μ) producing neutral pions or gammas
  - Muons from ν_μ misidentified as electrons from ν_e

Hyper-K's physics goals

Neutrino astrophysics

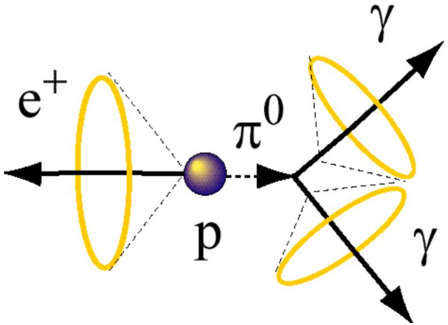
- Solar ν's: day/night asymmetry; hep ν's; ^{8}B ν spectrum upturn
- Supernova ν's: 1000's ν events for nearby supernova pointing, time & spectrum analysis; search for supernova relic ν's





Proton decay

- Search to order of magnitude greater lifetime than current limit
- 10^{35} years for $p \longrightarrow e^+ + \pi^0$
- 3×10^{34} years for $p \longrightarrow \bar{\nu} + K^+$

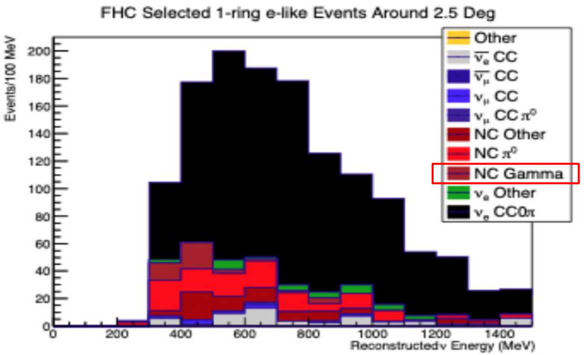


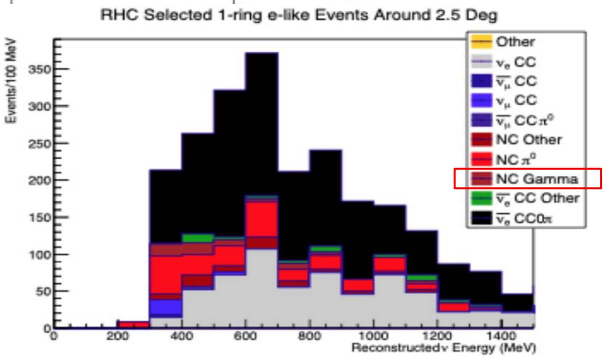
Particle type classification

New opportunities beyond simple reconstruction improvement

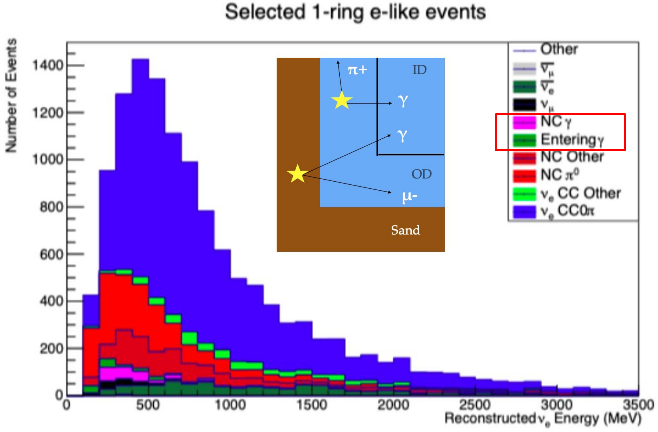
- NC γ discrimination and measurement

Systematic Type	Requirement	Motivation
$\sigma(\nu_e)/\sigma(\nu_\mu)$ to $\sigma(\bar{\nu}_e)/\sigma(\bar{\nu}_\mu)$ ratio	2.9-3.7%	CP violation search and precision δ_{CP} measurement.
$\sigma(\nu_e)/\sigma(\nu_\mu)$, $\sigma(\bar{\nu}_e)/\sigma(\bar{\nu}_\mu)$	3-5%	θ_{23} octant and precision θ_{23} measurement.
Intrinsic $\nu_e, \bar{\nu}_e$ and NC background normalizations	3-5%	CP violation search and precision δ_{CP} measurement.





2.7-4.0 degree off-axis range



- Potential neutron tagging application

- Bottom-up calibration: Enable multitude of detector parameter variations

Traditional reconstruction method

fiTQun: Likelihood-based reconstruction for higher energies

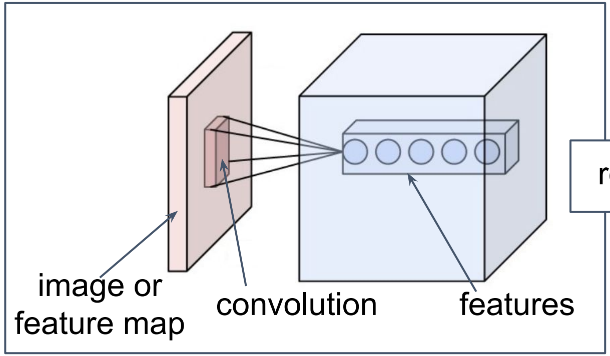
- Originally developed for Super-K detector
  - Based on algorithm of MiniBooNE: https://arxiv.org/abs/0902.2222
- Uses full information of unhit PMTs + time & charge of hit PMTs:

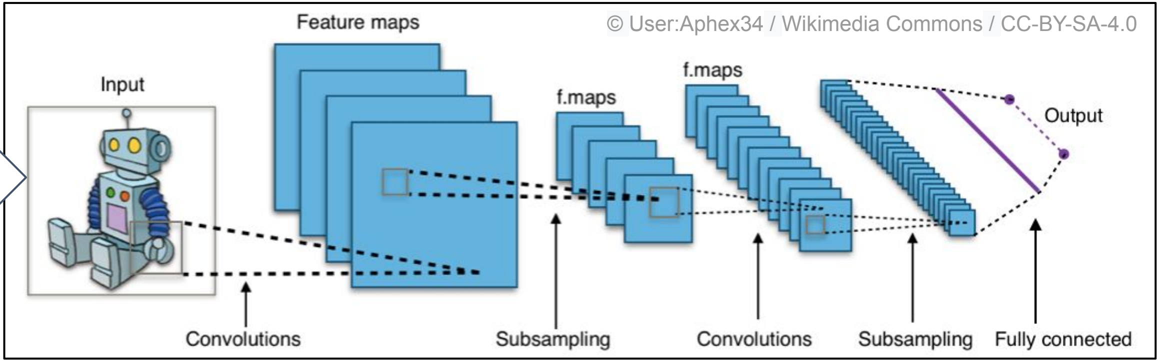
$$\underbrace{\boxed{L}}_{\text{Likelihood to maximise}}(\underbrace{\boxed{\mathbf{x}}}_{\text{Candidate event hypothesis}}) = \prod_{j}^{unhit} \underbrace{\boxed{P_j\left(unhit|\mathbf{x}\right)}}_{\text{Probability of no hit at PMT}} \prod_{i}^{hit} \underbrace{\boxed{P_i\left(hit|\mathbf{x}\right)}}_{\text{Probability of hit at PMT}} \underbrace{\boxed{f_q\left(q_i|\mathbf{x}\right)}}_{\text{Hit charge probability density}} \underbrace{\boxed{f_t\left(t_i|\mathbf{x}\right)}}_{\text{Hit time probability density}}$$

- Probabilities calculated based on direct + scattered + reflected light
- Likelihood ratios used to distinguish particle types and single-ring / multi-ring event topology hypotheses

CNN architecture

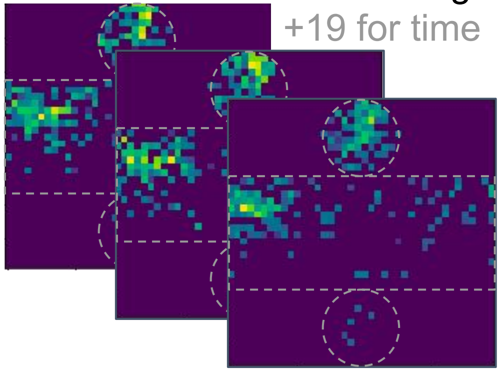
Convolutional neural networks hugely successful in image processing





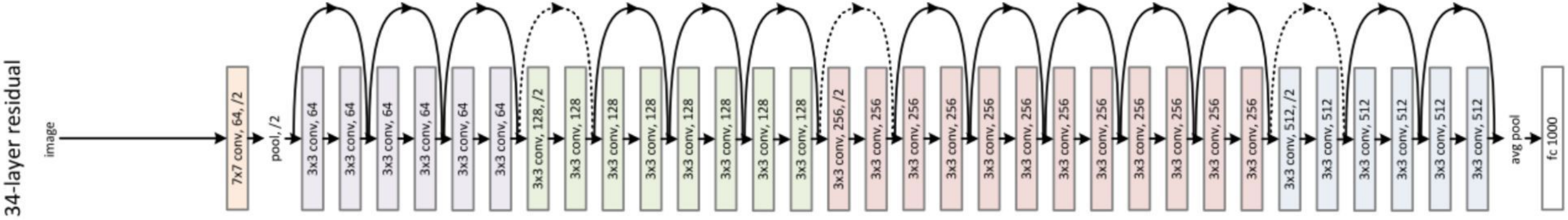
- Start with image with pixel values (‘features’): T and Q at each PMT
- Scan many small (e.g. 3x3) convolution kernels across image
  - Increases number of features
- Downsample image (e.g. 2x2 max-pooling)
  - Decreases number of pixels
- End with 1-D array of features, feed into traditional fully-connected neural network
- Learn convolution and final network weights through ‘back-propagation’ of loss

CNN architecture



Full cylinder of mPMTs is unwrapped onto 40x40 image

- 38 channels: charge & time of 19 PMTs per mPMT
- No special treatment for geometrical effects at boundary between barrel and end-caps
- Data augmented by reflecting / rotating around tank axis

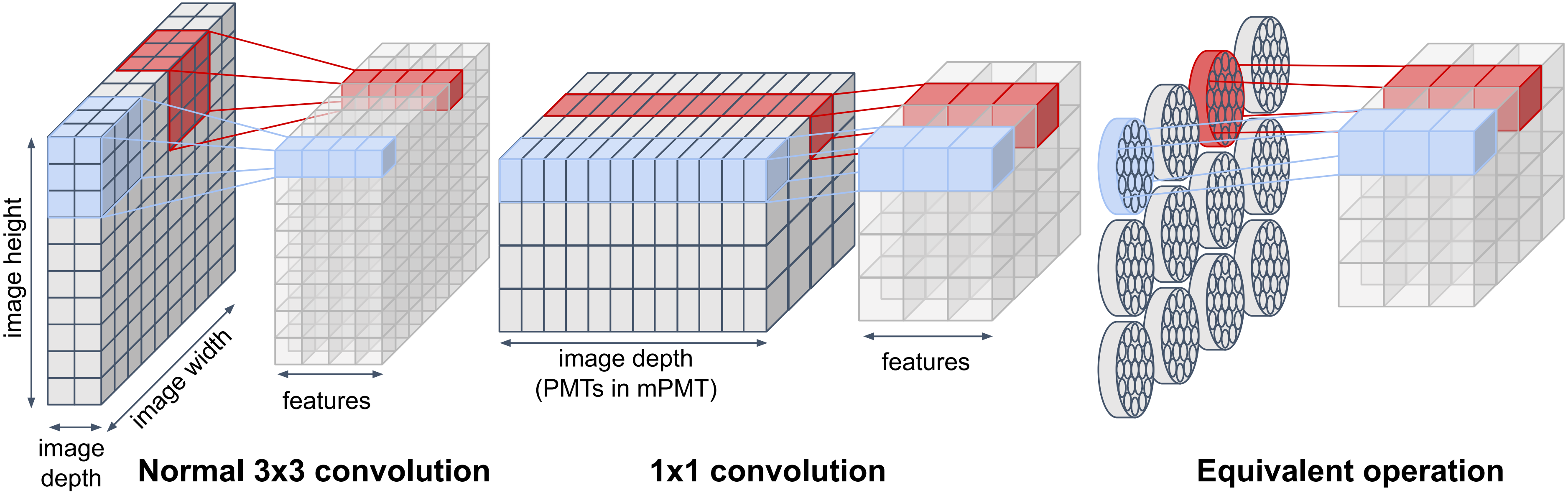


Mostly using ResNet-18 architecture [arXiv:1512.03385]

- Initial 1x1 convolution added to act on the 19 PMTs of each mPMT
- Also explored deeper networks with small improvement

CNN architecture

Treating each PMT inside mPMT as a channel, starting with 1x1 convolution
➔ equivalent to doing a ‘convolution’ over each mPMT

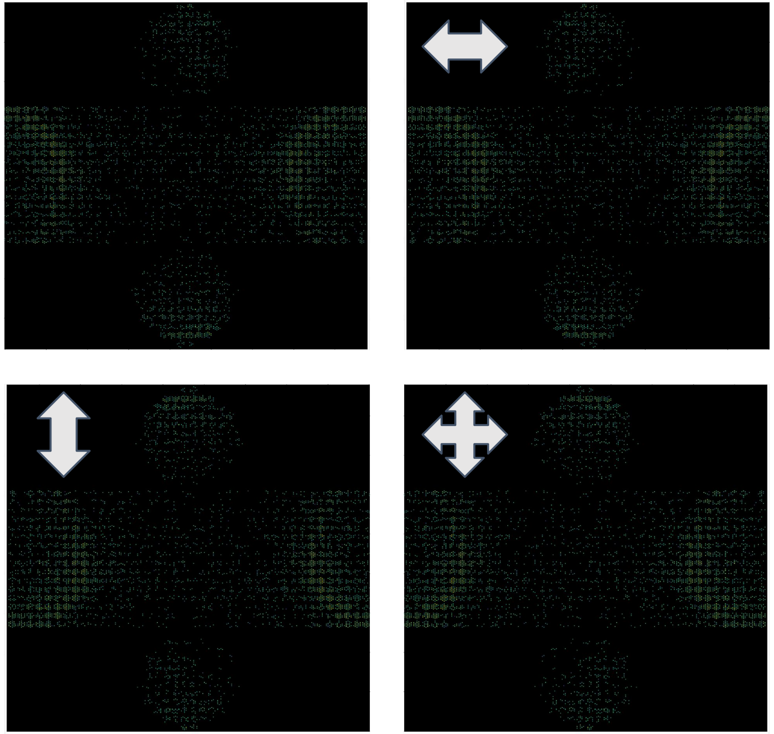


Normal 3x3 convolution **1x1 convolution** **Equivalent operation**

Data augmentation

Effectively increase dataset size by implementing transformations of existing events (data augmentation)

- Started by simplest transformations
  - Horizontal flip of image
  - Vertical flip of image
  - Both horizontal and vertical flip
- In future could also do rotation of 180° around tank axis
  - Can only do 180° since end-cap mPMT positioning is symmetric by 90° rotations but PMTs within mPMT symmetric by 60° rotations



4x increase in dataset size allows network to learn more real features without overfitting

Topological map to square

$$X_{\pm} = W(\rho, z)\frac{\pi \pm \phi}{2\pi}$$

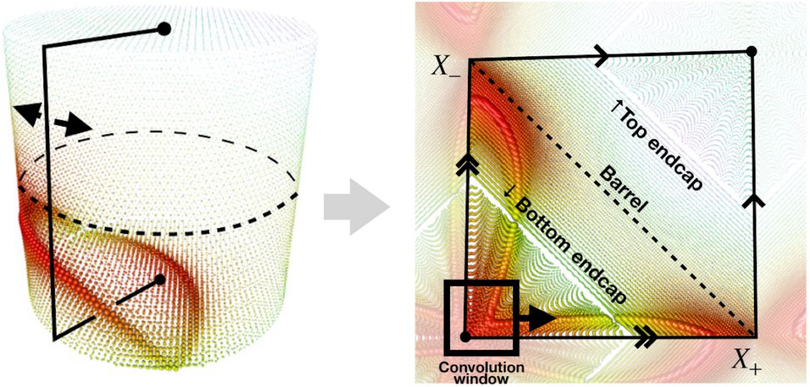
$$W(\rho, z) = \sqrt{\frac{\rho^2 + 2Rz + RH}{R^2 + RH}}$$

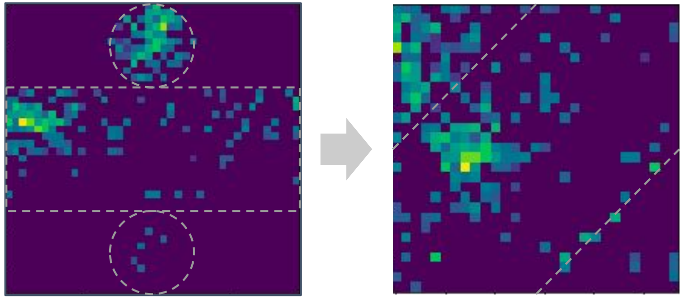
Solve differential eq. for constant Jacobian

$$\mathrm{d}X_+ \mathrm{d}X_- = \left|\frac{\partial(X_+, X_-)}{\partial(\rho, \phi)}\right| \mathrm{d}\rho\, \mathrm{d}\phi$$

Alternative map onto square with boundary conditions preserving topology of cylinder

- Cut open along barrel to centre of end caps (solid line)
- Deform onto square, keeping density of PMTs constant
- Place mPMTs onto nearest pixel
- Use boundary conditions identifying edges of square (indicated by arrows)
    - Pad image with copy of pixels from the corresponding edge

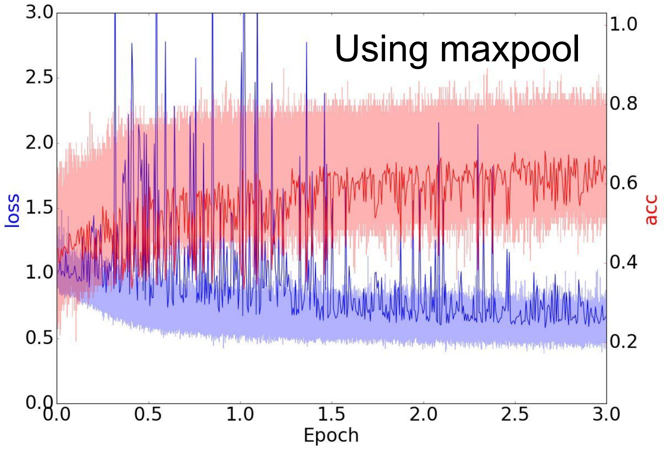


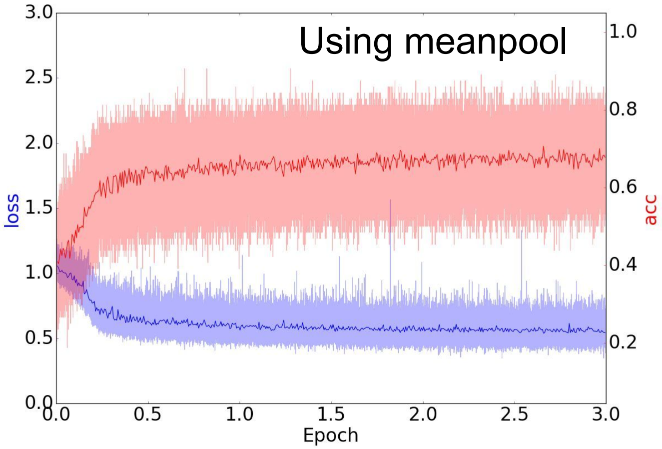


PointNet architecture

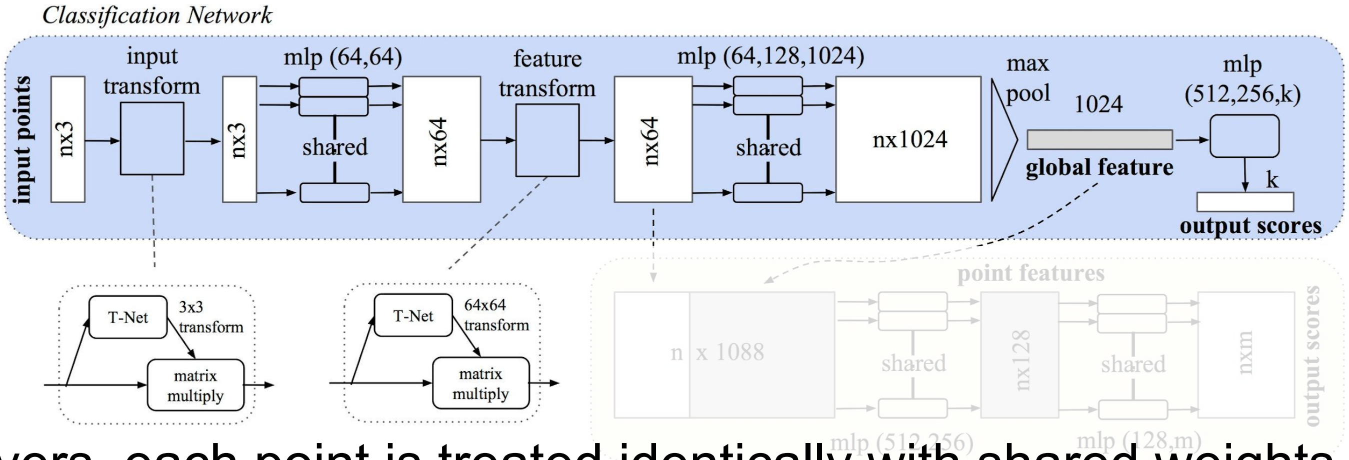
Some changes to standard PointNet give improvements

- Severe overfitting until max. features reduced from 1024 to 256
  - Possibly due to limited batch size with larger network
  - Data augmentation could also help
- We find that mean pool works better than standard max pool here
  - PointNet usually picks key points to learn features, but aggregating information from all points seems better for our tasks





PointNet architecture



In MLP layers, each point is treated identically with shared weights

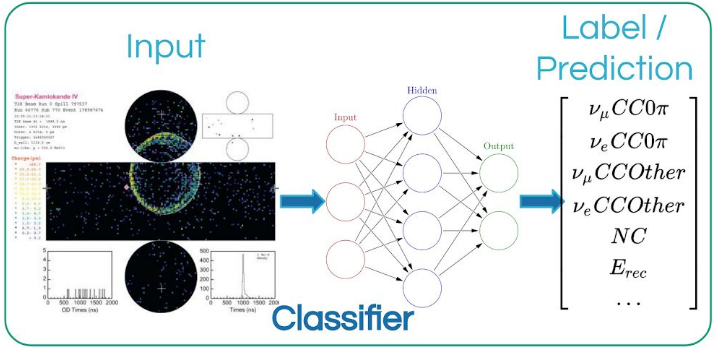
- Similar to each pixel treated the identically in a CNN
- But without downsampling, information does not transfer between points

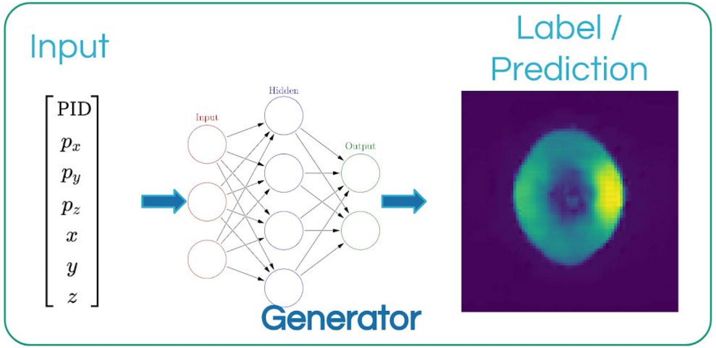
Instead ‘T-Nets’, resembling PointNet, learn transformations of the points

- Linear transformation is learnt to e.g. rotate all input vectors
- Feature transform allows global information to affect individual points

Single downsampling layer at the end of the network collapses all points

Cherenkov ring generator

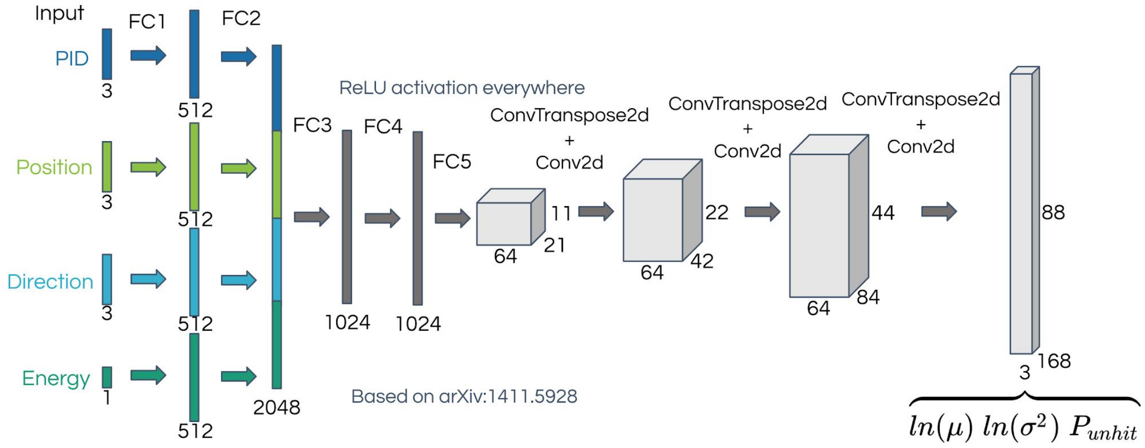




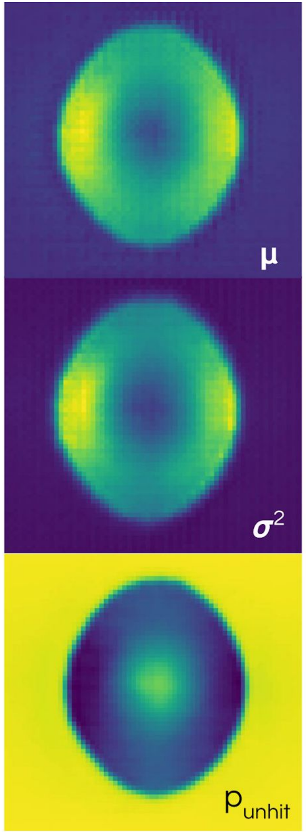
Investigating hybrid method using generative network

- Generative network can predict PMT hit charge and time
- Use to replace likelihoods in traditional reconstruction
- Combine learning ability of CNN with physics domain knowledge of traditional reconstruction
- Simple replacement for existing reconstruction in full analysis chain

Cherenkov ring generator



$$\text{Loss} = -ln(\mathcal{L}) = -\sum_{unhit} ln(P_{unhit}) - \sum_{hit} ln(1-P_{unhit}) - \sum_{hit} \frac{1}{2}\left[ln(2\pi\sigma^2) + \frac{(q_{obs}-\mu)^2}{\sigma^2}\right]$$



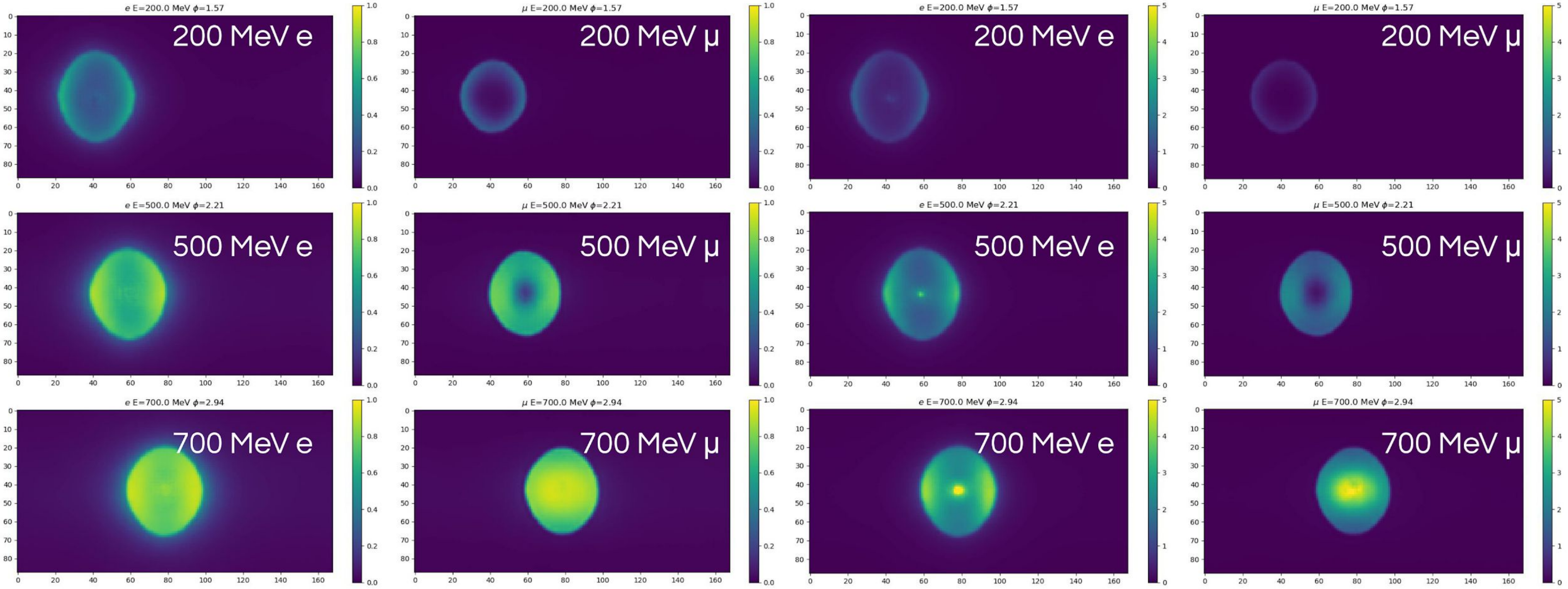
Network outputs likelihoods for hits observed at PMT

- Probability of PMT being hit
- Gaussian pdf (μ, σ) for charge

Cherenkov ring generator

Hit probability

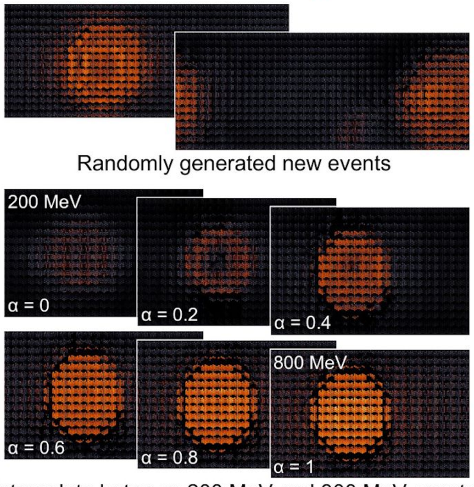
Hit probability
x
Mean charge



Generative networks

Also considering using generative networks for improved detector systematics

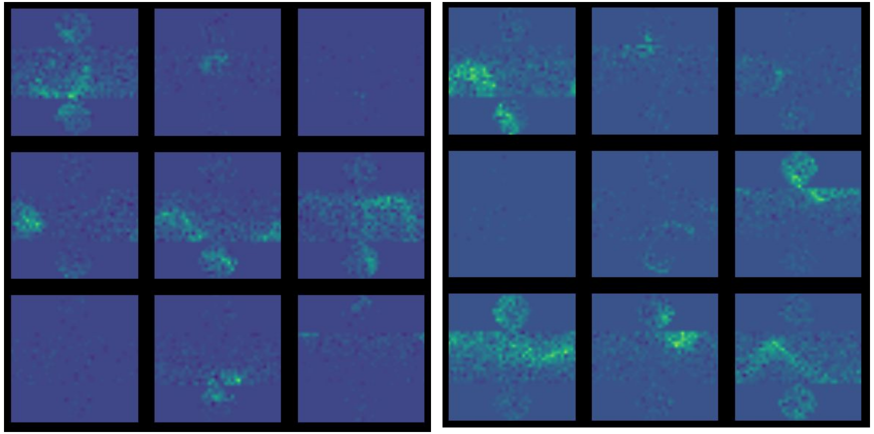
- Train generative network to reproduce real data: removed dependence on MC
- Train GAN to take simulated event and make it look like real data
    - Reduce detector systematics by ‘fixing’ mismodelled detector simulation
- Initial work on VAE showed some promise, but struggled with noise and sharp details
- Now we are investigating GANs



Randomly generated new events

Interpolate between 200 MeV and 800 MeV events

arXiv: 1911.02369



GAN generated events   Geant4 simulated events