TRIUMF

Machine Learning Techniques to Enhance Event Reconstruction in Water Cherenkov Detectors

Nick Prouse

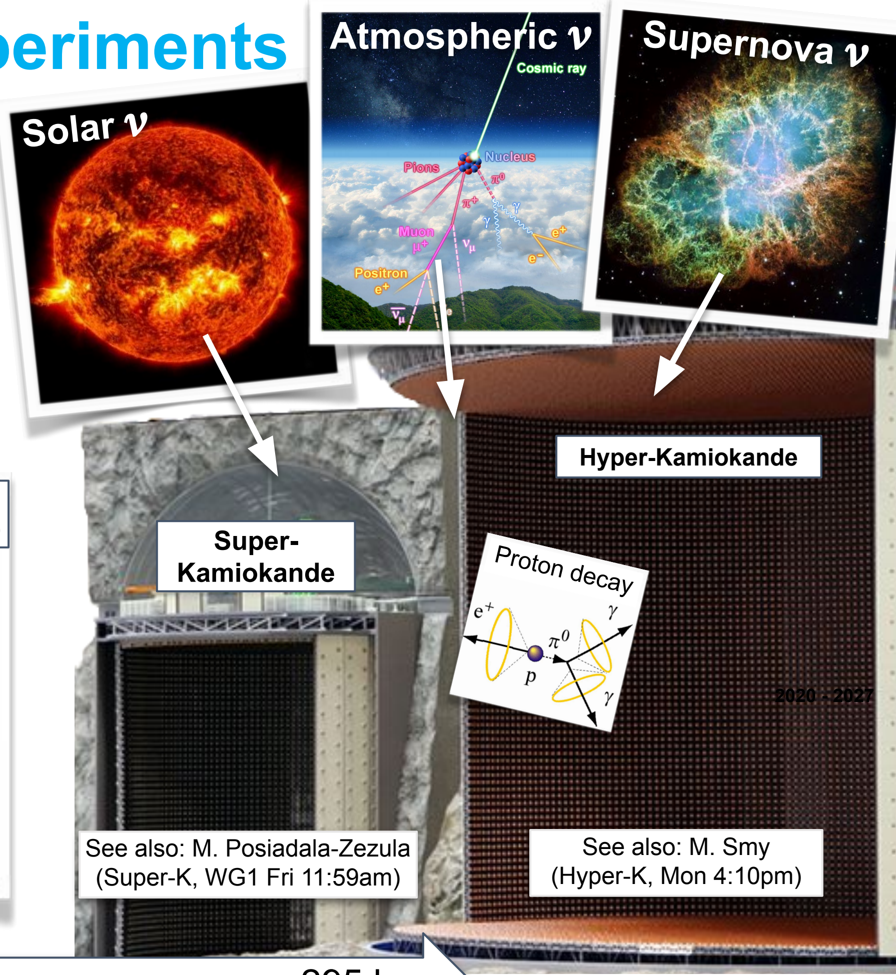
NuFact 2022, Salt Lake City, Utah, 4th August 2022

Water Cherenkov Neutrino Experiments

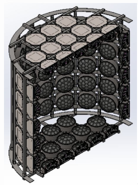
Current generation **Super-K** and **T2K** and next generation **Hyper-K** are world-leading neutrino experiments.

Broad & ambitious physics programmes covering many neutrino sources as well as proton decay measurements.

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

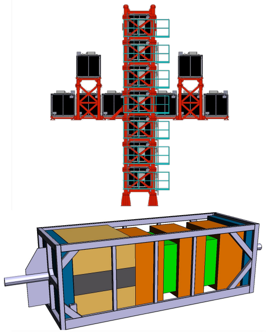


See also: L. Kormos (T2K, Mon 2:20pm)
M. Friend (J-PARC, Wed 8:30am)

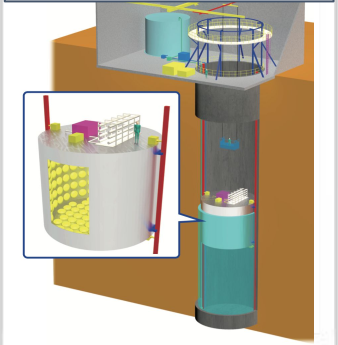


Water Cherenkov Test-beam Experiment (WCTE) at CERN

Near detectors



Intermediate Water Cherenkov Detector (IWCD)



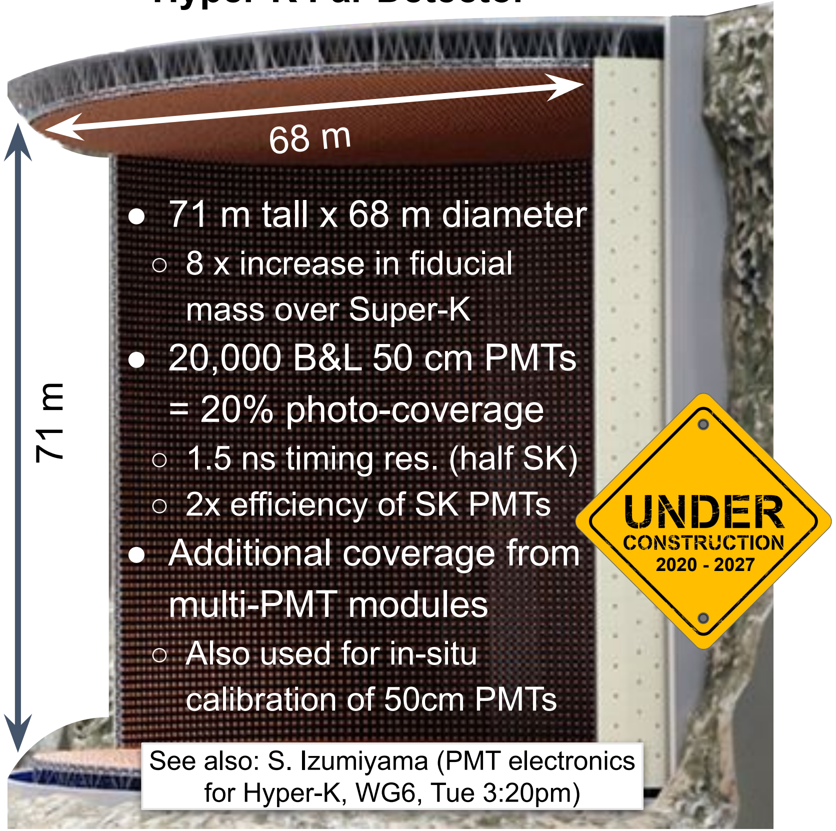
See also: M. Posiadala-Zezula (Super-K, WG1 Fri 11:59am)

See also: M. Smy (Hyper-K, Mon 4:10pm)

280 m → ~ 1 km → ~ 295 km

Hyper-K's WC Detectors

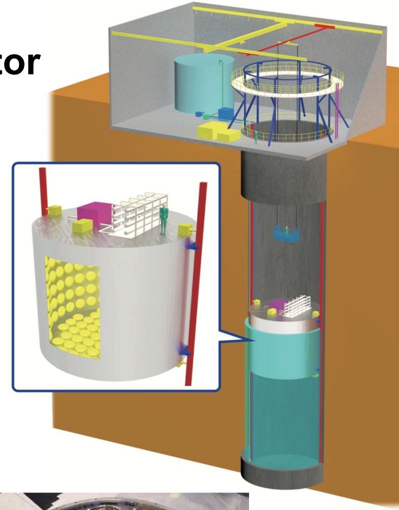
Hyper-K Far Detector



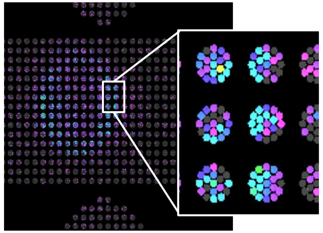
- 71 m tall x 68 m diameter
  - 8 x increase in fiducial mass over Super-K
- 20,000 B&L 50 cm PMTs = 20% photo-coverage
  - 1.5 ns timing res. (half SK)
  - 2x efficiency of SK PMTs
- Additional coverage from multi-PMT modules
  - Also used for in-situ calibration of 50cm PMTs

See also: S. Izumiyama (PMT electronics for Hyper-K, WG6, Tue 3:20pm)

Intermediate Water Cherenkov Detector



- Measures ν flux and cross-section of beam at ~1 km from source
- Moves vertically in ~50 m tall pit
  - spans off-axis angles of ν beam for different ν energy spectra
- 6 m tall x 8 m diameter tank 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

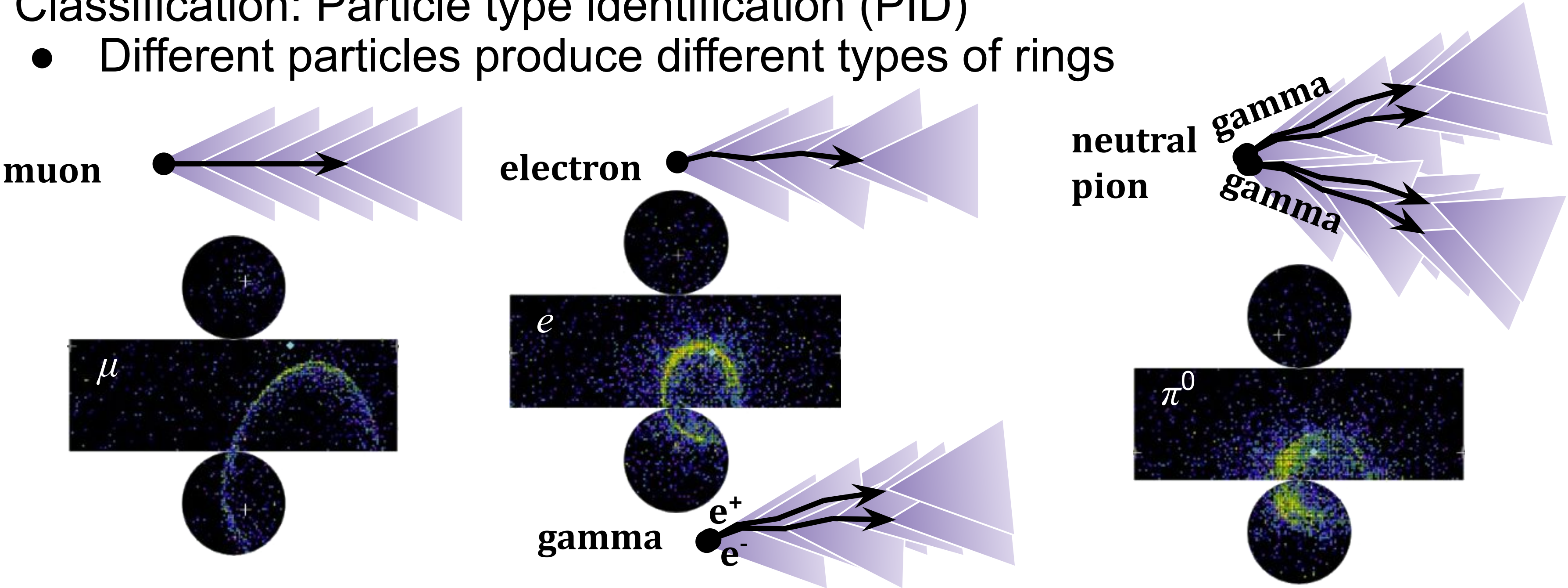


See also: R. Akutsu (mPMTs for IWCD & WCTE, WG6 Tue 3pm)

Reconstruction in WC detectors

Classification: Particle type identification (PID)

- Different particles produce different types of rings



Regression: reconstructing particle's properties:

- Location and time of PMT hits allows triangulating position and direction
- Amount of charge observed at PMTs gives estimate of energy

Machine learning reconstruction for WC

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

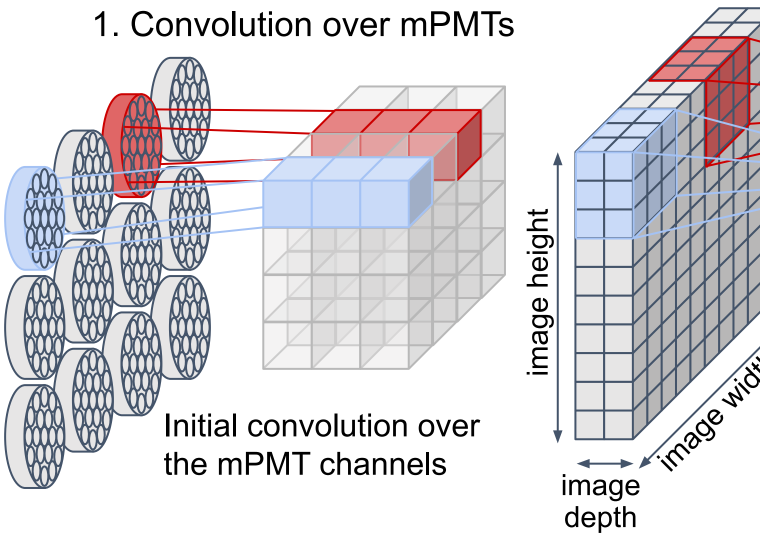
- Computation time is becoming a limiting factor
  - Larger far detector with more PMTs increases computation time
  - Smaller intermediate detector requires scaled down resolutions
  - Improving resolutions requires more complex algorithms with fewer approximations

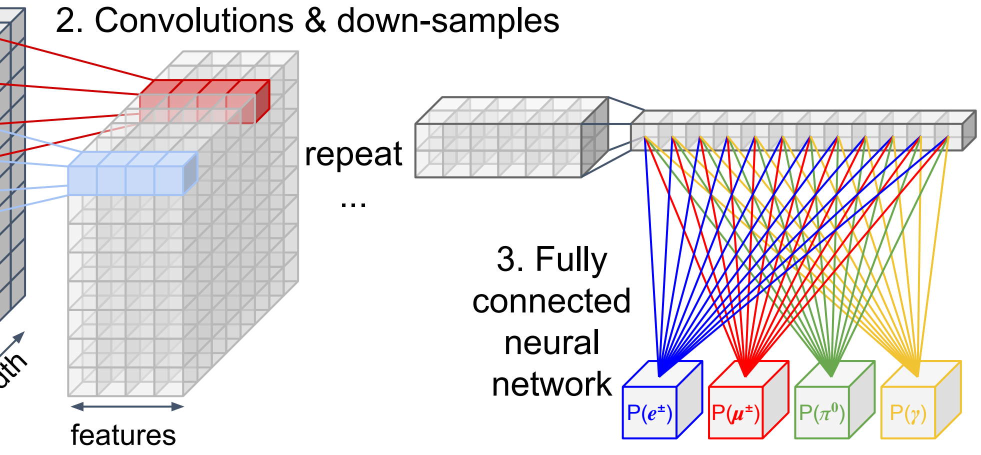
ML and deep neural networks have potential to push reconstruction further

- Very successful in areas of computer vision and image processing
- Potential to use all information without detector model approximations
- Very fast to run once neural networks have been trained
  - fiTQun on CPU: 1 event takes more than 1 minute
  - ML reconstruction on GPU: 100,000 events per minute
  - Opens opportunities for analyses with huge datasets not currently possible

See also: A. Yankelevich (ML for solar ν in SK, WG1+WG6 Thu 3:04pm)

Deep network architectures for IWCD



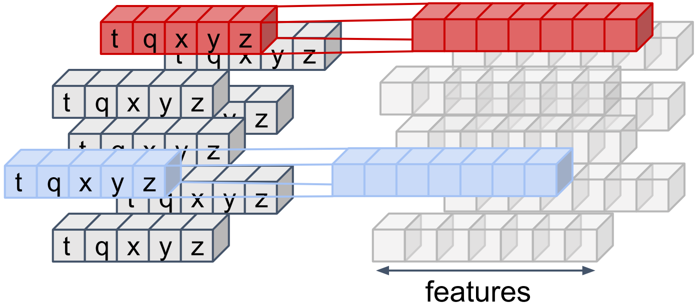


Convolutional Neural Network based on ResNet-18

- Full cylinder of mPMTs is unwrapped onto flat image
- One pixel per multi-PMT

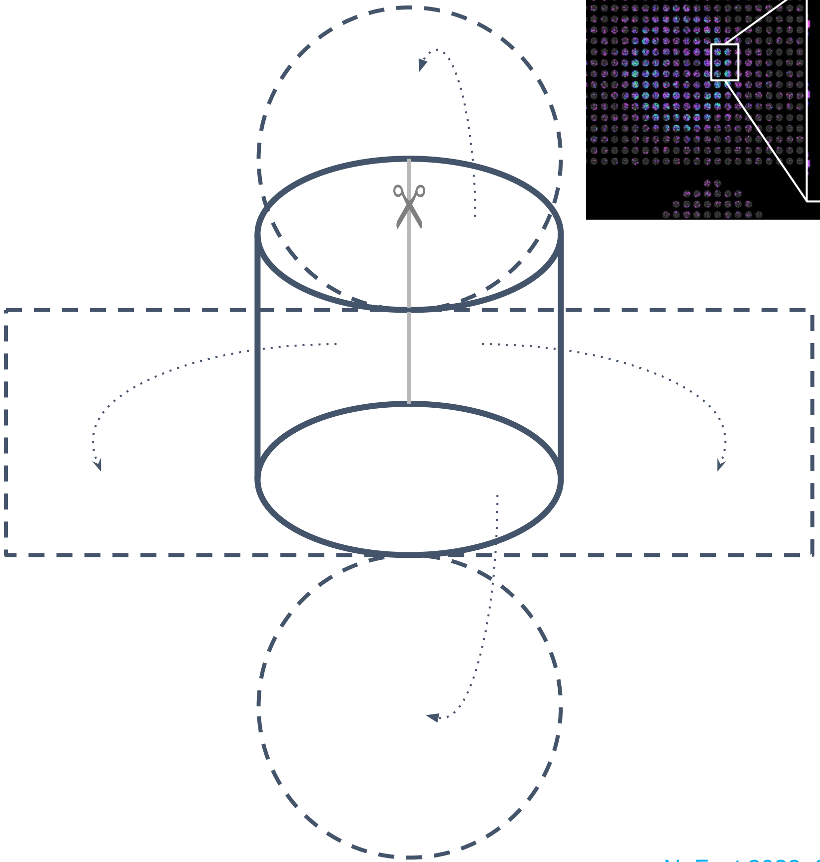
Point Cloud Neural Network based on PointNet

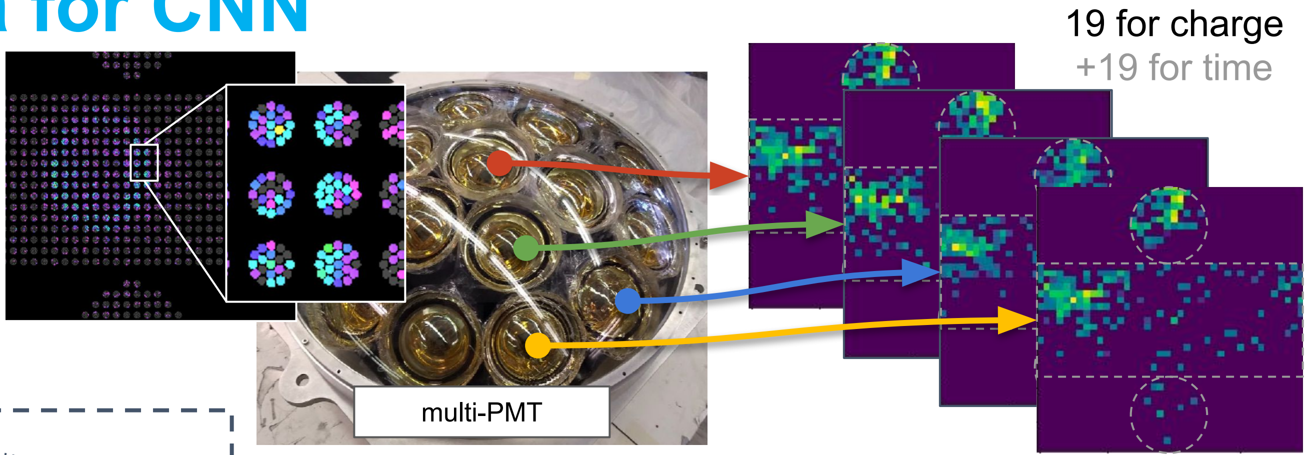
- Applies to point-cloud of PMT hits in 3D space
- Uses 1x1 convolutions and learns transformations applied to points



PointNet MLP (convolution over point cloud features)

Image-like data for CNN



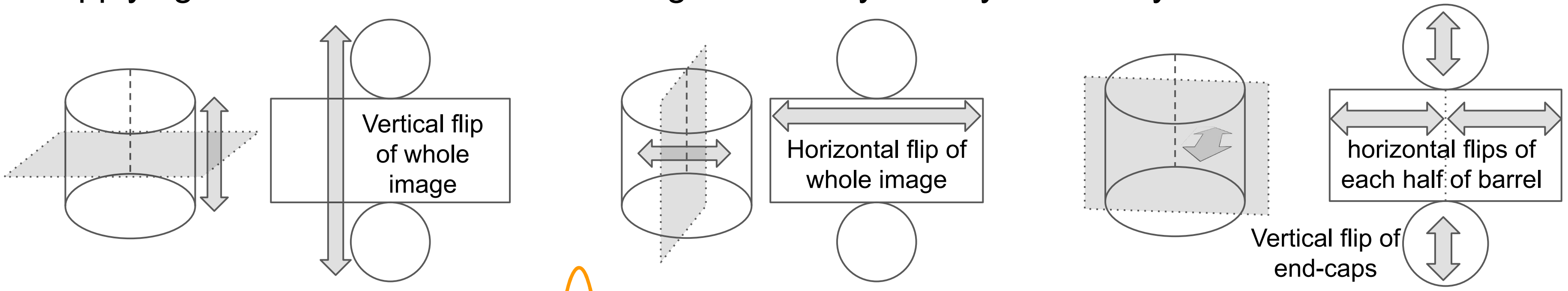


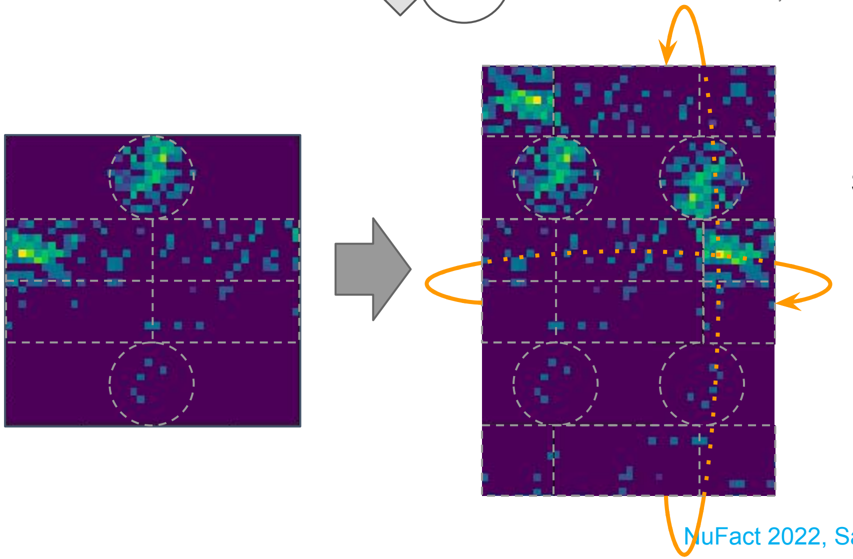
Full cylinder of mPMTs is unwrapped onto flat image

- 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

Data Transformations and Augmentation

Applying random transformations using detector symmetry effectively increases dataset

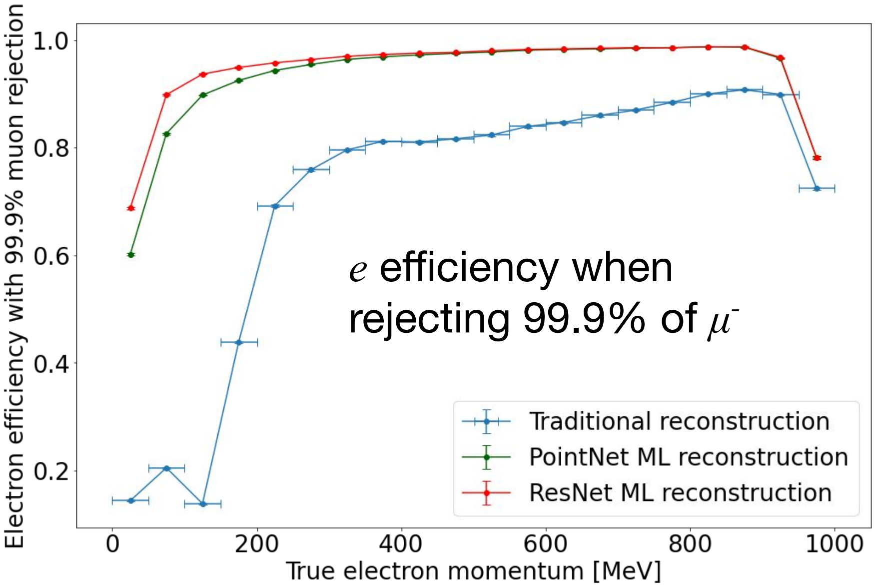


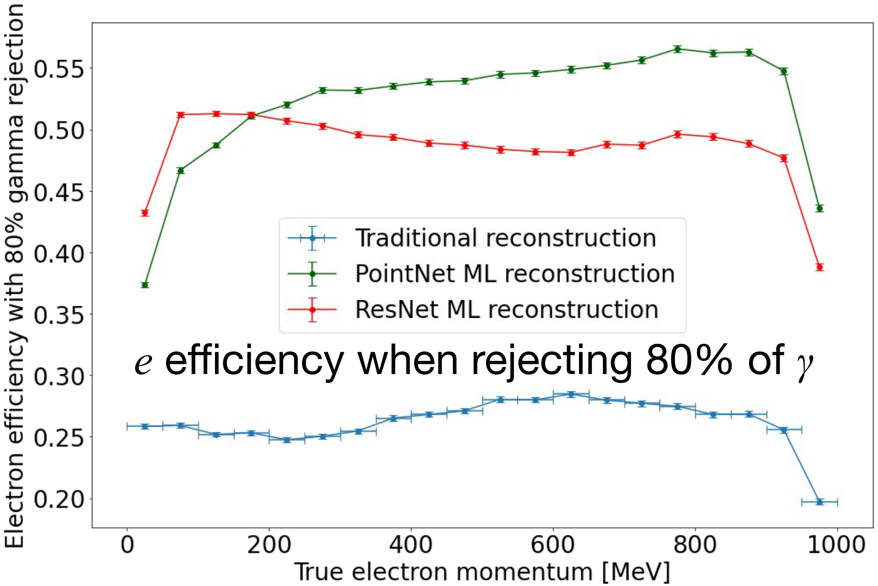


Rearranging and duplicating the geometrical surface has several advantages

- Less dependence on “choice” of slice along barrel to unwrap cylinder
- All segments appear exactly twice with minimal blank space
- Circular boundary conditions in both directions

Classification for PID in IWCD





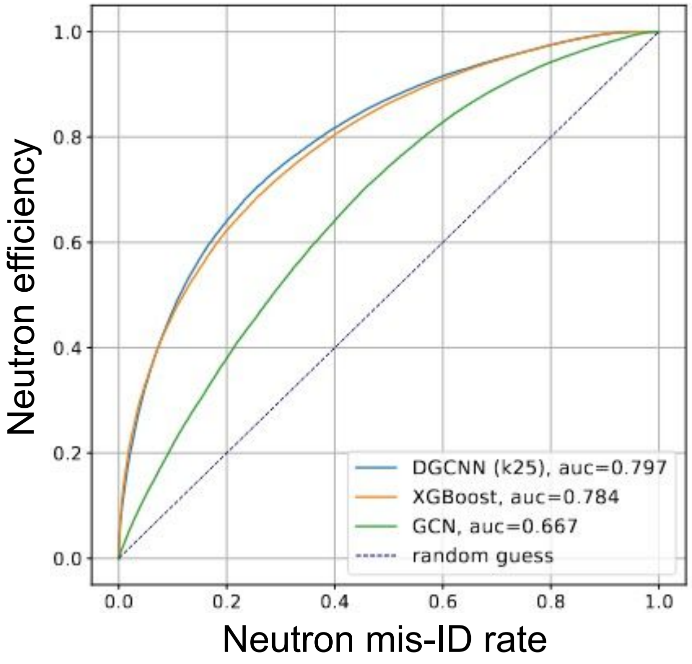
- ν_μ beam produces mostly μ, need rejection factor of 1000 for ν_e measurement
- Improved performance across energy range
- ResNet performs slightly better than PointNet for e vs μ classification

- γ and e almost indistinguishable in water Cherenkov detectors
- Discrimination has not been possible before
- PointNet performs better than ResNet for e vs γ classification

Classification for neutron captures

- At lower energies, images can be very sparse and CNNs tend to perform less well
- Alternative networks like graph networks may be more useful
  - Each PMT is a node on a graph
  - Time, charge, position are node features
  - Graph can be defined by nearest neighbors in Graph Convolution Network (GCN, arXiv:1609.02907)
  - Graph can be learned dynamically in Dynamic Graph Convolutional Neural Network (DGCNN, arXiv:1801.07829)
- Tested classifying neutron captures vs electron background
  - Signal: ~8 MeV gamma cascade
  - Background: beta decays of isotopes produced by cosmic muon spallation
  - Compared performance to baseline using BDT (XGBoost) with features including number of hits, hit isotropy, etc.

B. Jamieson, et al., arXiv:2206.12954

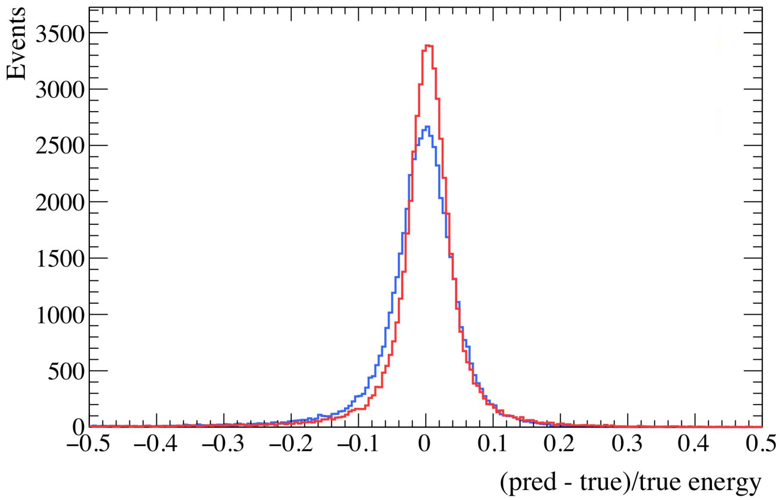


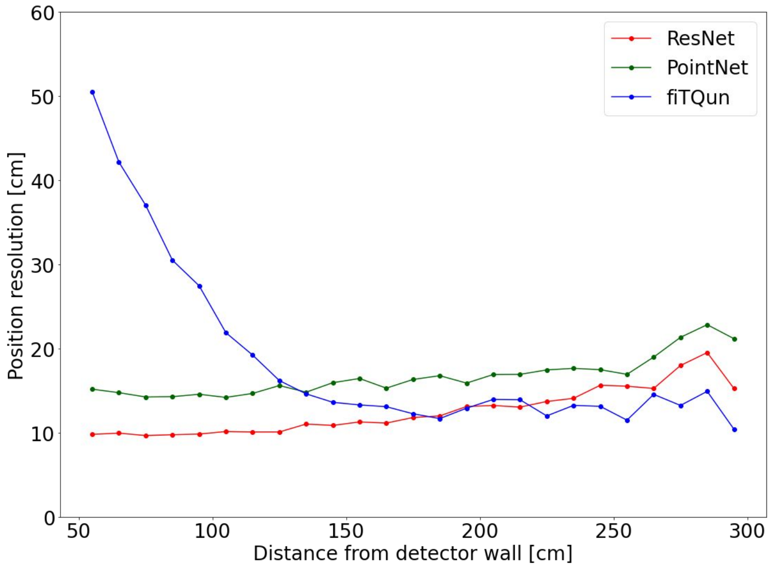
DGCNN outperforms BDT baseline, while GCN underperforms

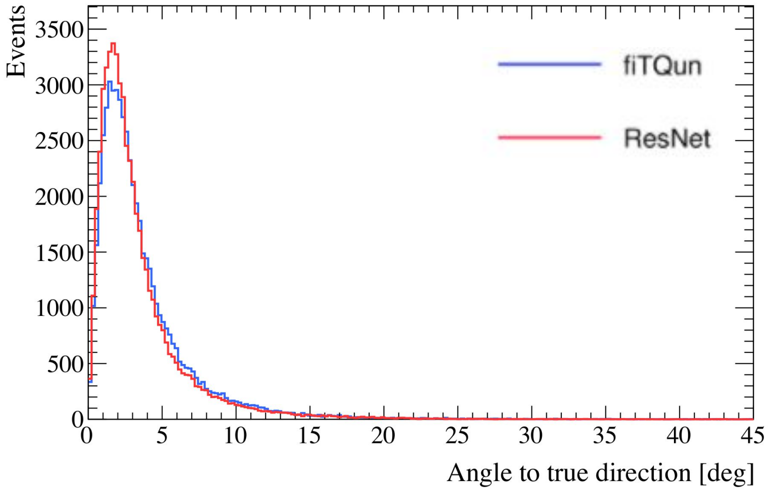
Position, direction, energy reconstruction

Using same IWCD data and networks as classification

- Output reconstructed quantities instead of PID variables
- Improved performance over traditional reconstruction (fiTQun), particularly for particles close to detector wall



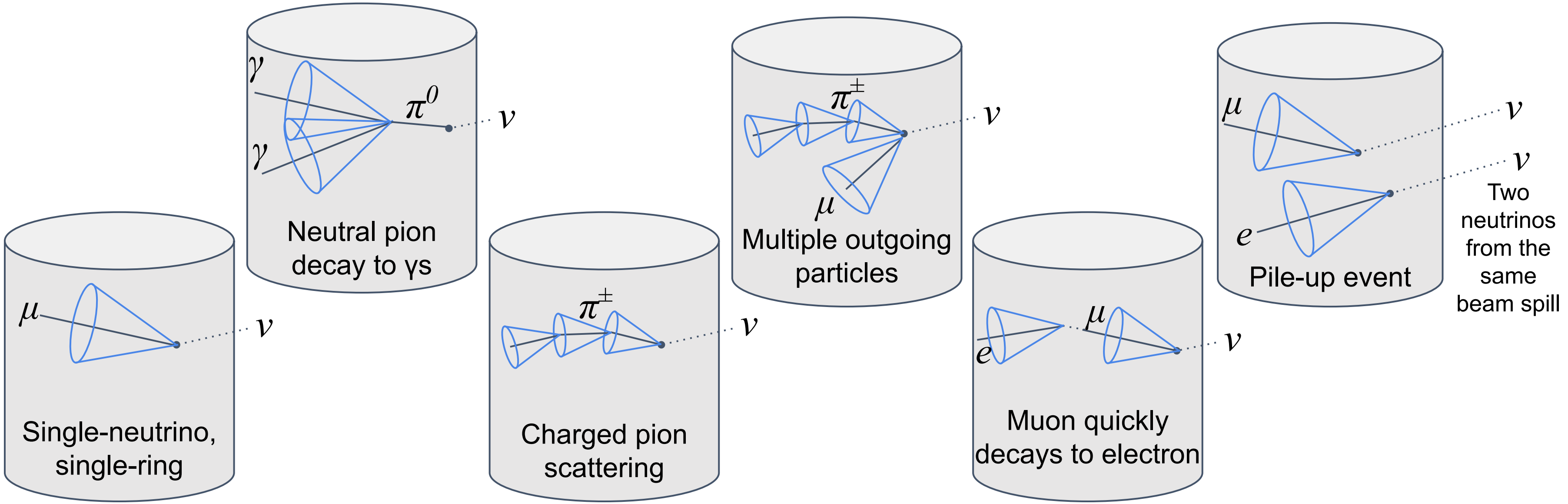




Multi-ring and multi-vertex events

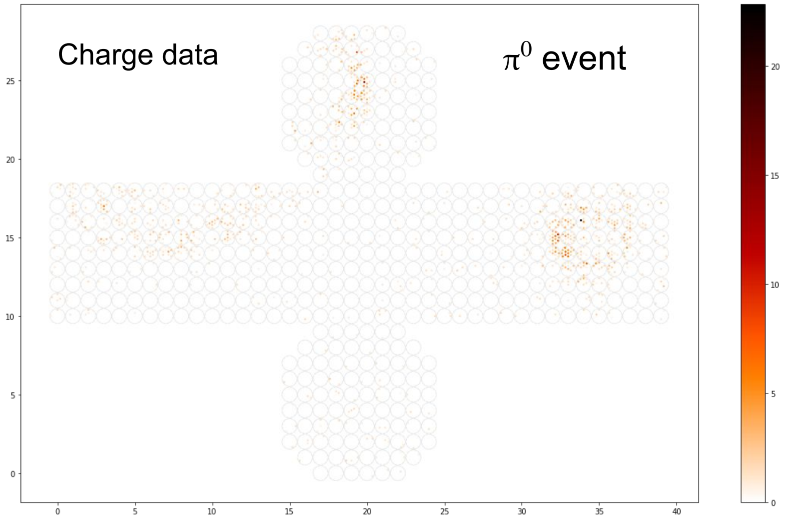
Need to develop ability to identify and reconstruct multi-ring and multi-vertex events

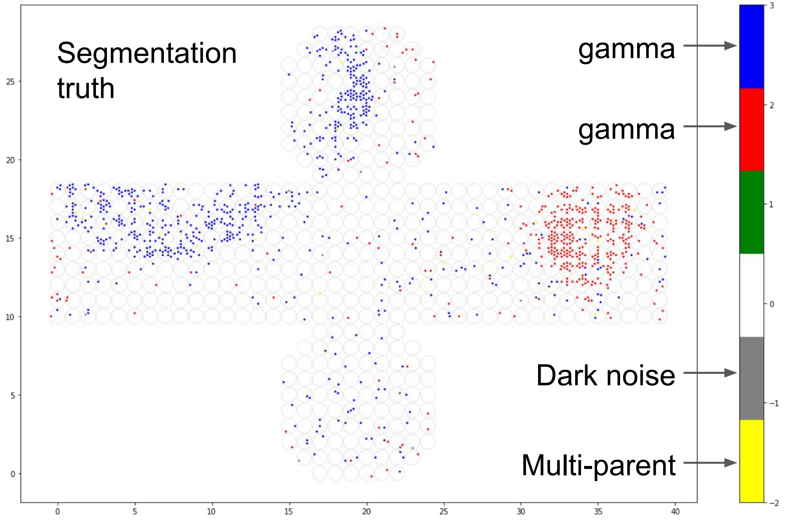
- Single-neutrino interactions can produce various multi-ring event topologies
- Pile-up of neutrino interactions is possible for IWCD due to proximity to beam source



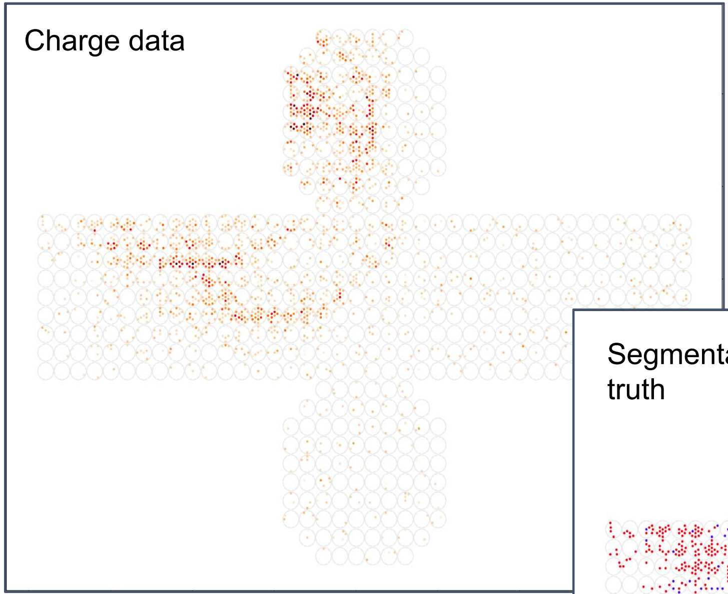
Segmentation networks

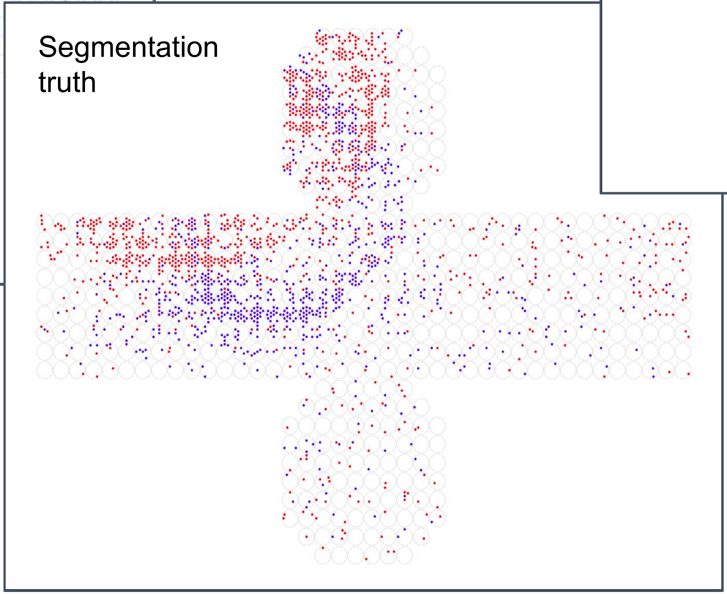
- Classification networks can be extended to perform segmentation
  - Deconvolutions and upsampling reverse convolutions and downsampling
  - Provides output value for each pixel
  - Currently using U-Net and FRRN
- Starting development with π^0 events
  - π^0 decay to produce two γ rings
  - Higher energy π^0 have overlapping rings

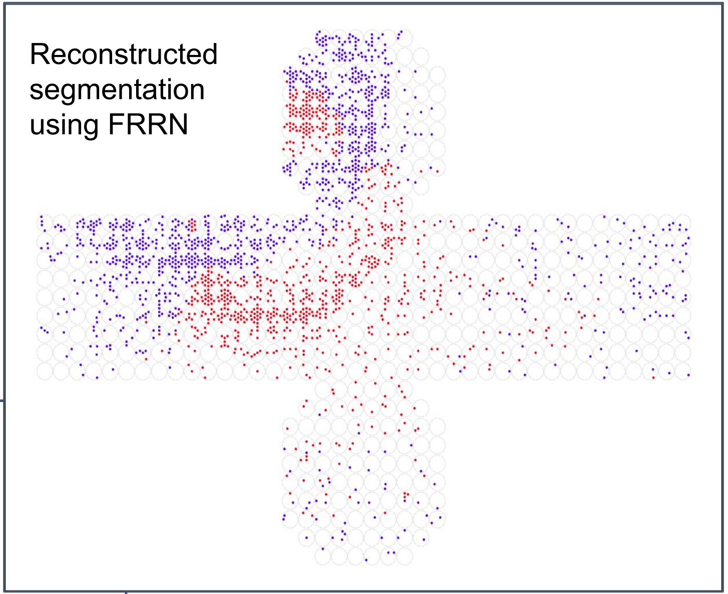




Segmentation results

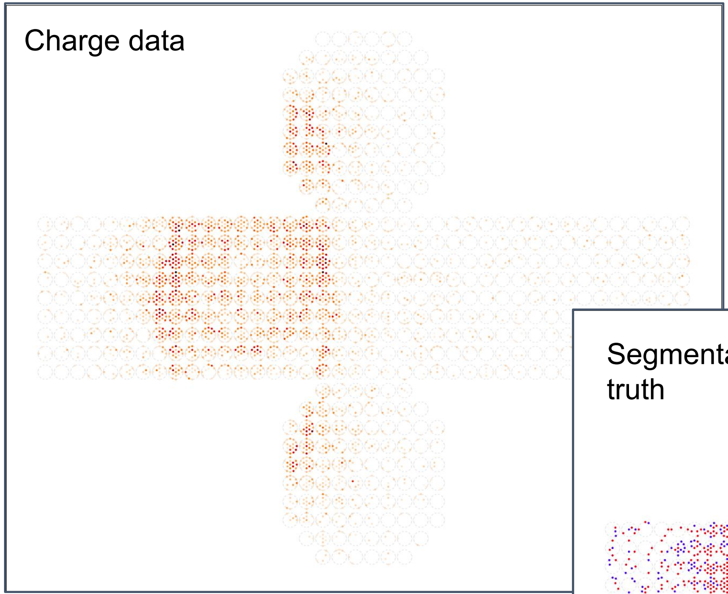


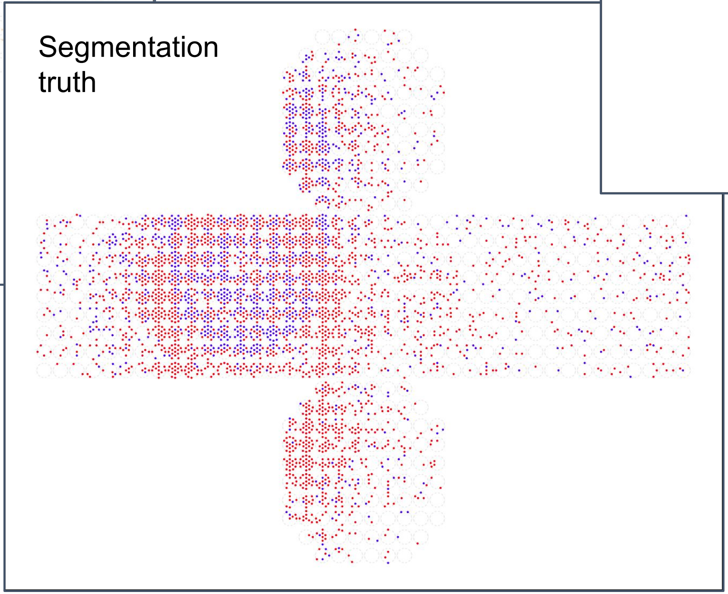


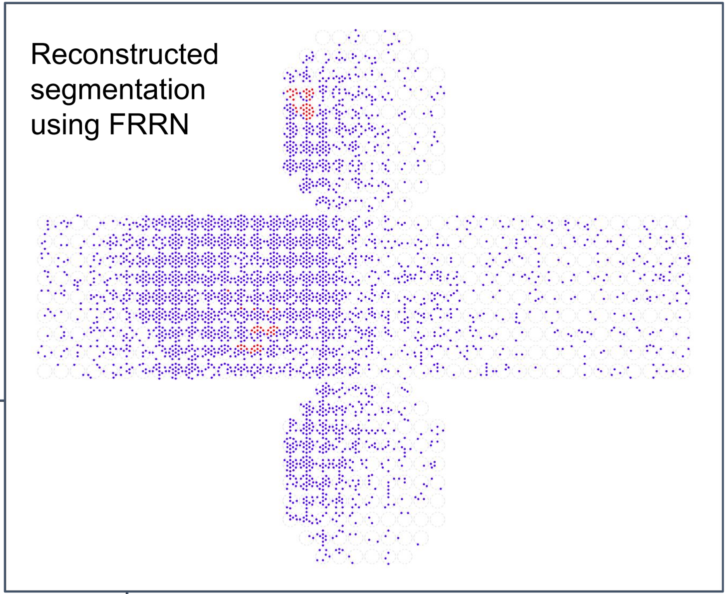


Works well with separated or partially overlapping rings

Segmentation results







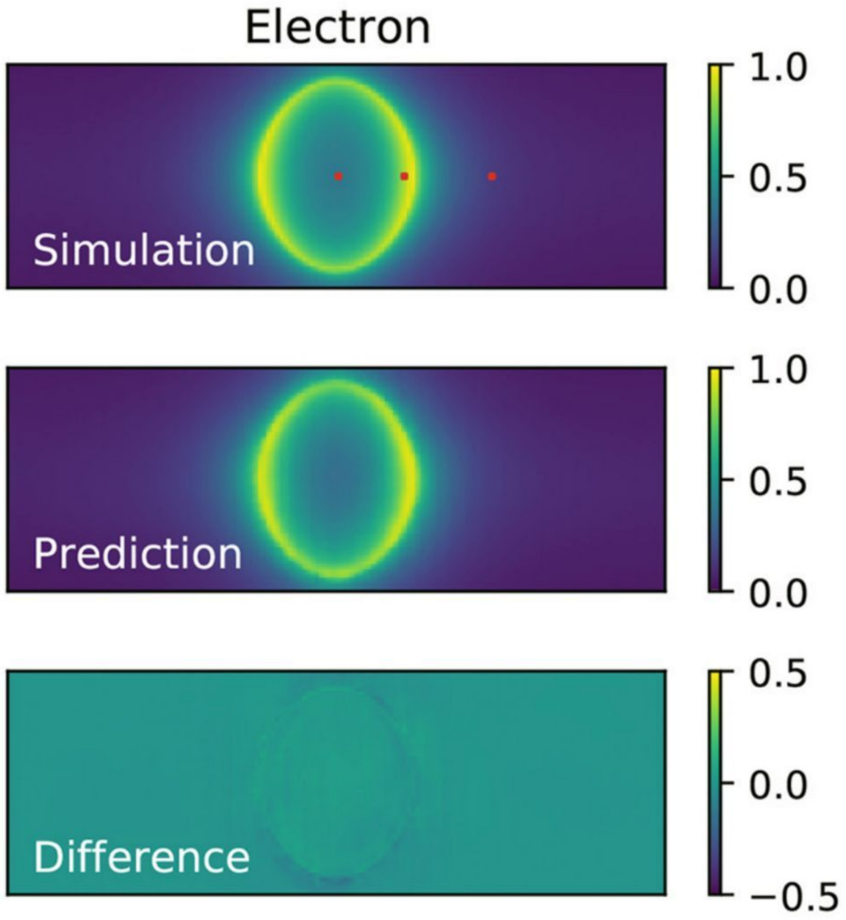
Poor reconstruction with some more overlapping rings

Generative networks

Generative networks create synthetic data, with several possible applications

- Often used for faster or more accurate simulations or modifying data for different purposes
- Novel use as part of hybrid approach with maximum likelihood event reconstruction
  - Limitations of traditional reconstruction arise from computational complexity of likelihood function
  - Generative network can quickly produce Cherenkov rings used in likelihood calculation without physics model approximations
  - Predict parameters of Gaussian mixture model for charge & time likelihood functions at each PMT

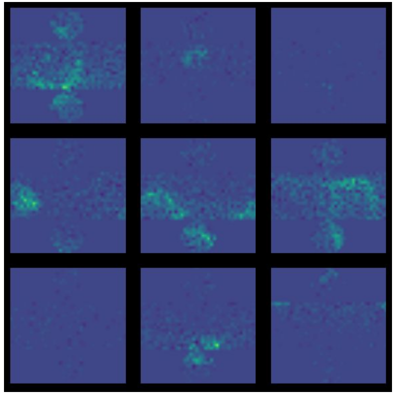
M. Jia, et al., arXiv:2202.01276



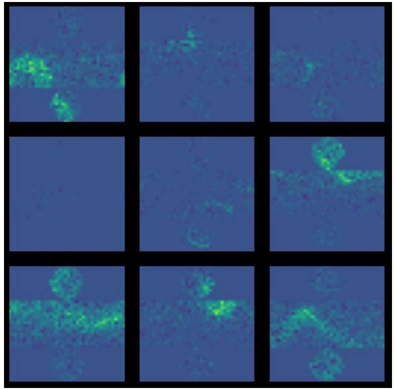
Generative networks

Generative networks create synthetic data, with several possible applications

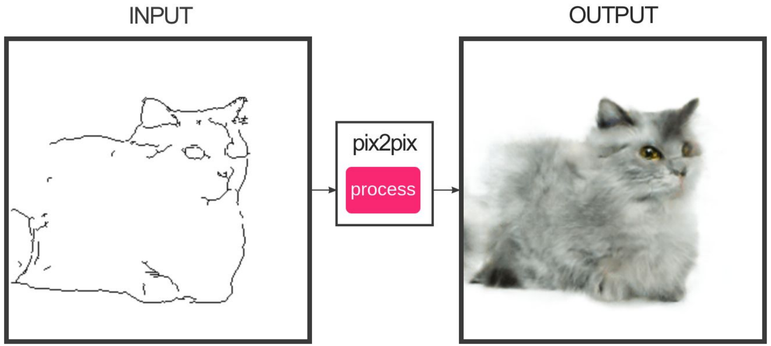
- Generative Adversarial Networks (GANs)
  - Train generative network and classifier together
  - Classifier is trained to distinguish generated data from training data
  - Generative network is trained to generate data the classifier cannot distinguish
- Can train using real data (e.g. calibration data, control samples, or general ‘unlabelled’ data)
  - Avoid biases / systematics from imperfect detector simulation models
- Potential use for noise reduction
  - Train generative network to produce de-noised events from noisy events
- Potential uses for detector calibration
  - Train network to modify simulated events to more closely match real data



GAN generated events



Geant4 simulated events



Inaccurate drawing of cat (simulated event)

Photo-realistic image of cat (looks like real data)

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 and multi-vertex events looks promising
- Generative networks allow hybrid ML/traditional reconstruction and novel approaches to handle detector calibration and modelling

Hyper-Kamiokande

WatChMaL.org

Appendix

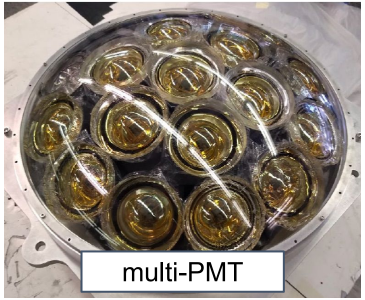
Hyper-K Detector

8 x increase in fiducial mass over Super-K

- 71 m tall x 68 m diameter = 258 kt total mass
  188 kt fiducial mass

New photo-detector technology for increased sensitivity

- 20,000 B&L 50 cm PMTs = 20% photo-coverage
  - 1.5 ns timing resolution (half that of SK PMTs)
  - Double quantum efficiency of SK PMTs
- Additional photo-coverage from multi-PMT modules
  - 8 cm PMTs grouped in modules of 19 PMTs
  - Improved position, timing, direction resolution
  - Also used for in-situ calibration of 50cm PMTs



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

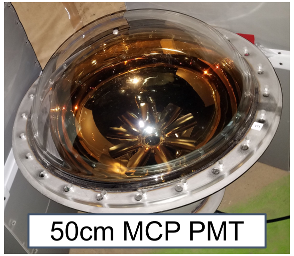
Hyper-K far detector

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

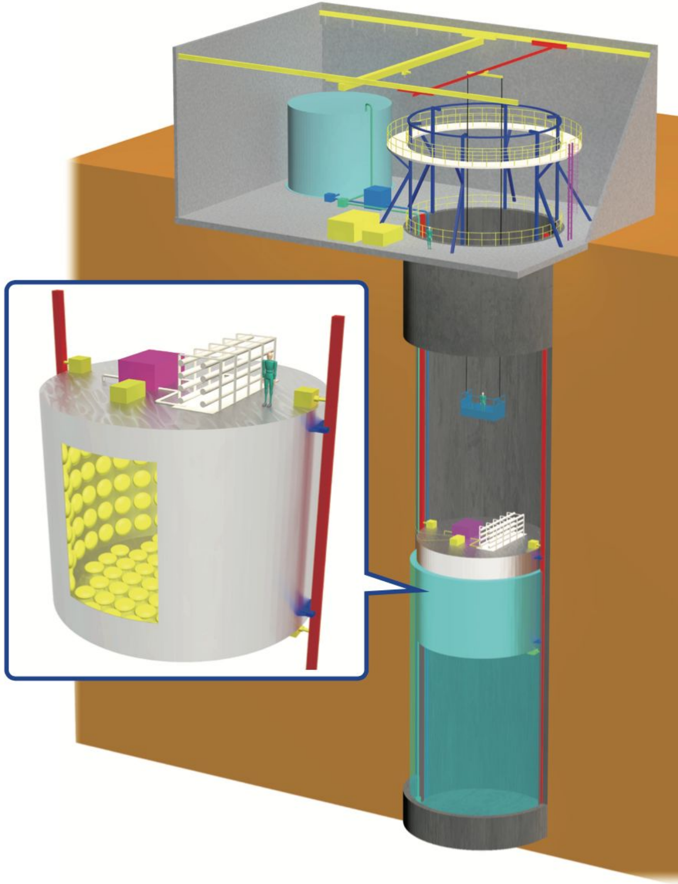
Intermediate detector (IWCD)

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

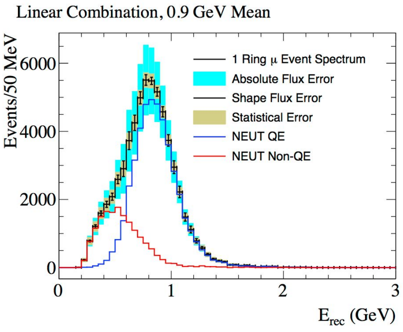
Off-axis spanning detector

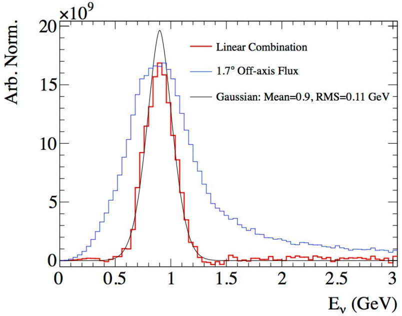
ν energy spectrum depends on angle off-axis to the neutrino beam

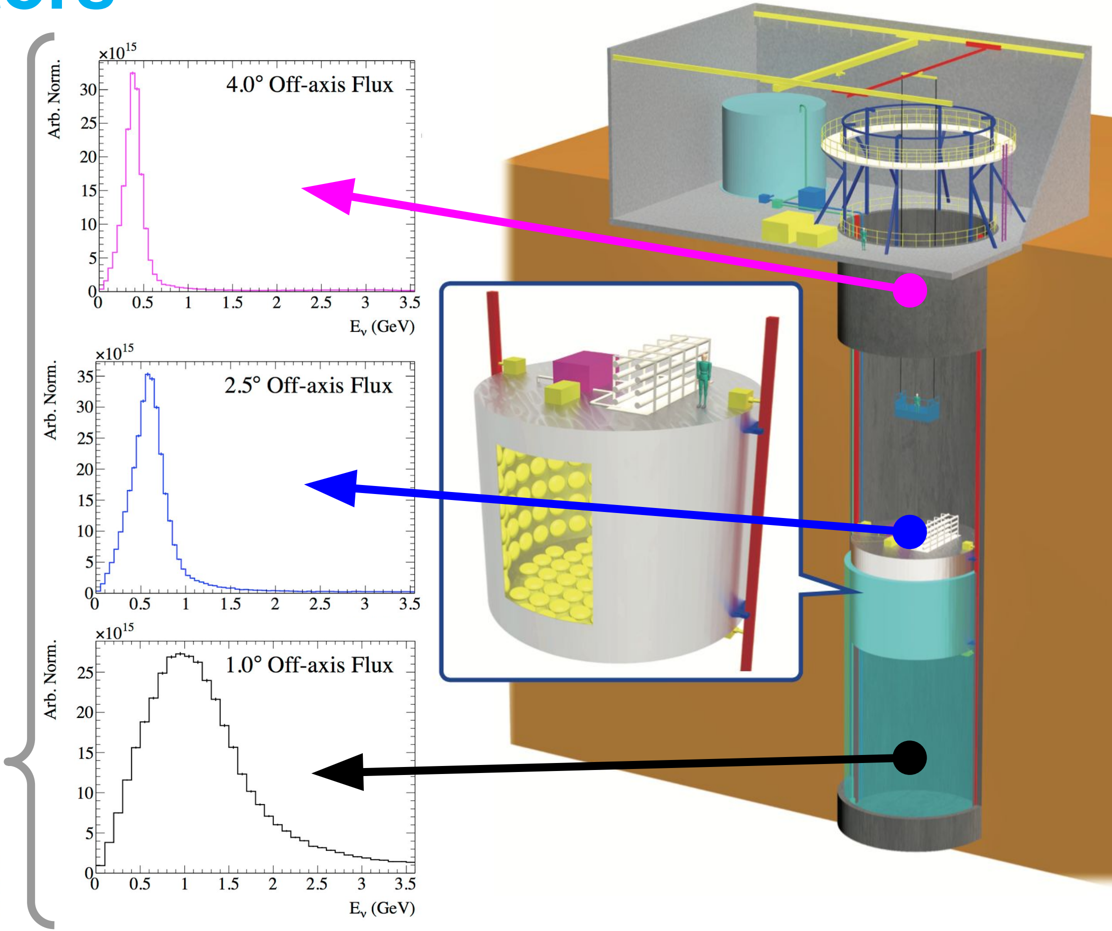
Far detector @ 2.5° for peak at ~600 MeV

Moving IWCD varies angle, allowing measurements at different energies

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







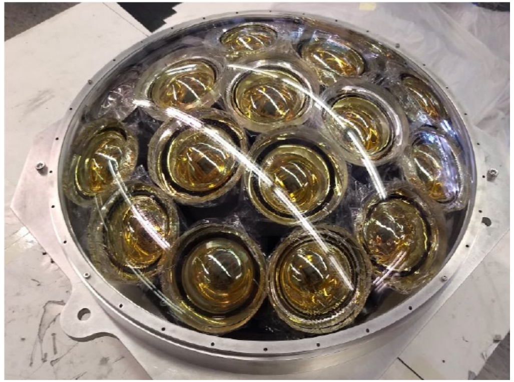
Hyper-K's WC detectors

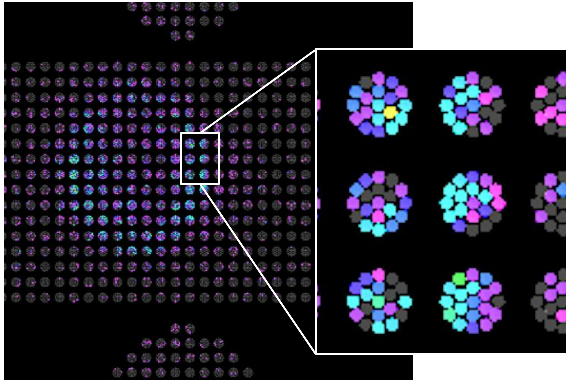
Multi-PMT modules

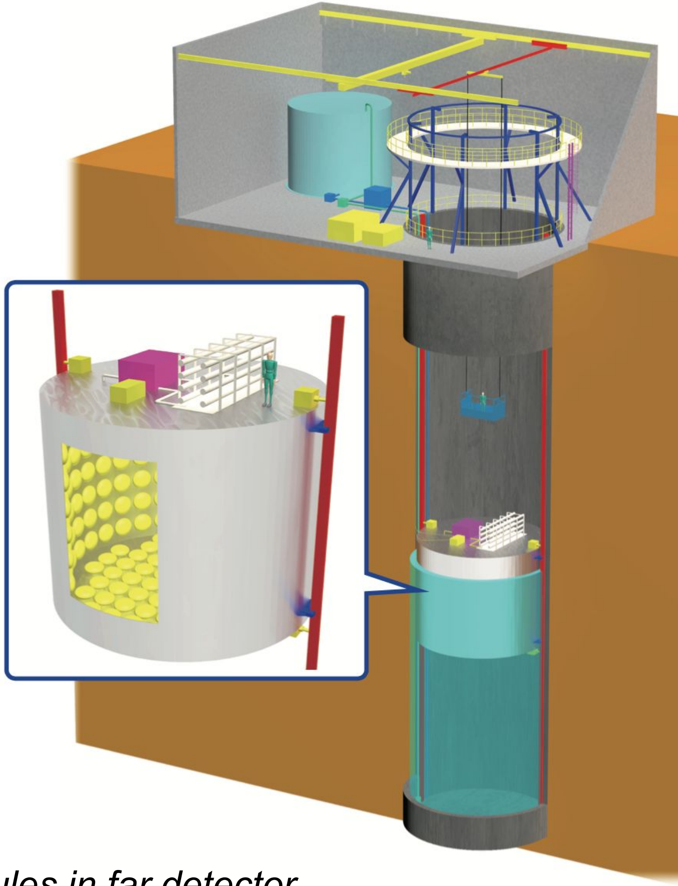
8 cm PMTs: Better position resolution
< 1 ns timing resolution
Additional directionality information

Need reconstruction to exploit additional information

Necessary for smaller detector size



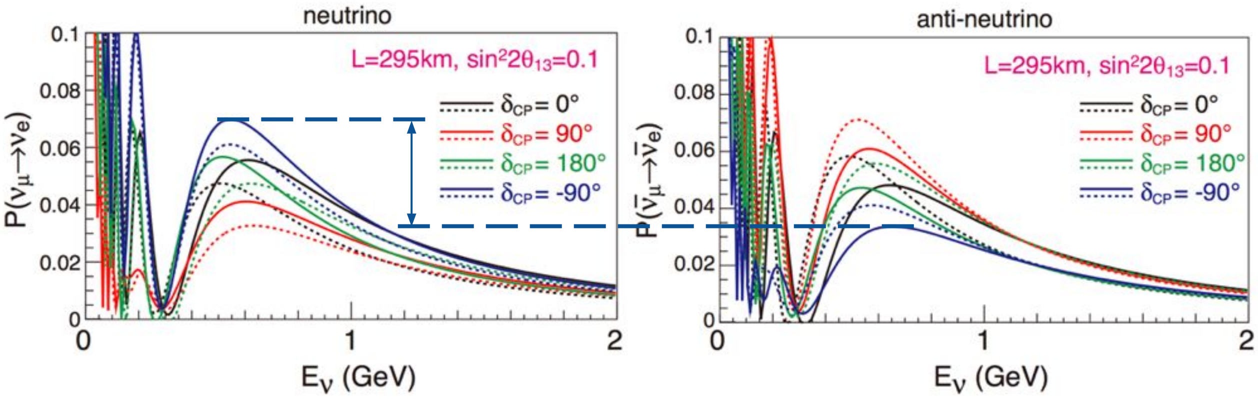




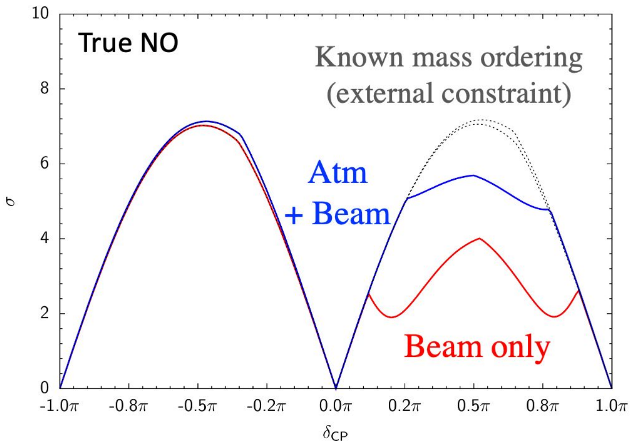
Also under investigation: Combining 50 cm PMTs + multi-PMT modules in far detector

Hyper-K's physics goals

Long-baseline neutrino oscillations: CP violation



10 years with 1.3MW, T2K 2018 systematic error

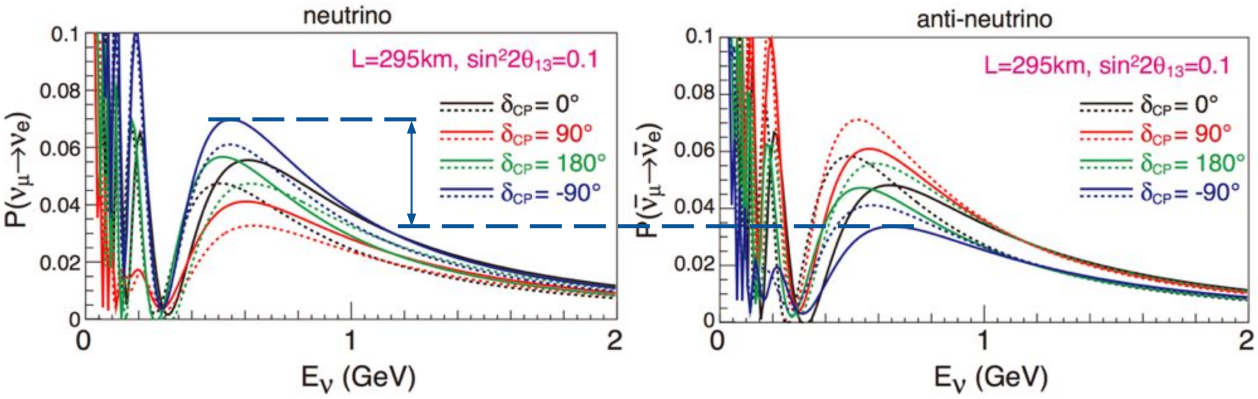


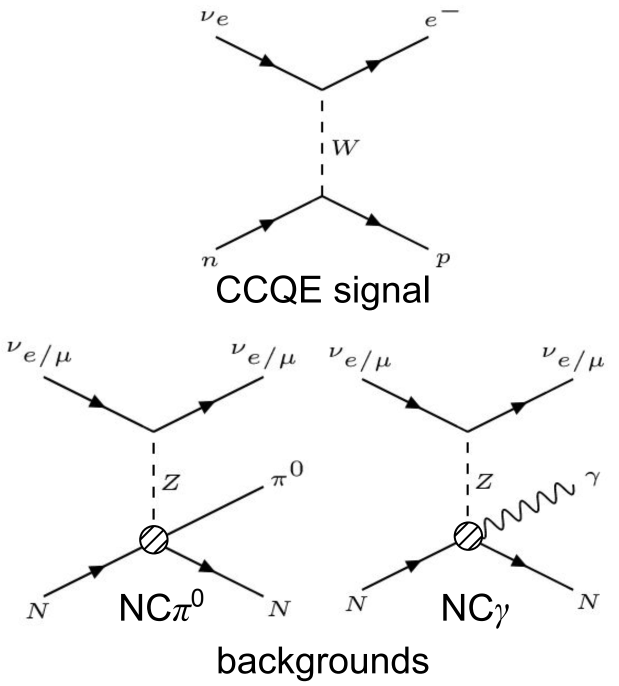
Combine beam and atmospheric neutrino observations for maximum sensitivity

- 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





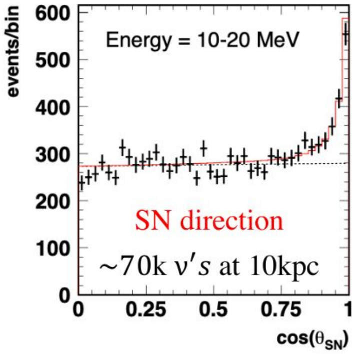
Oscillation maximum is at around 0.6 GeV

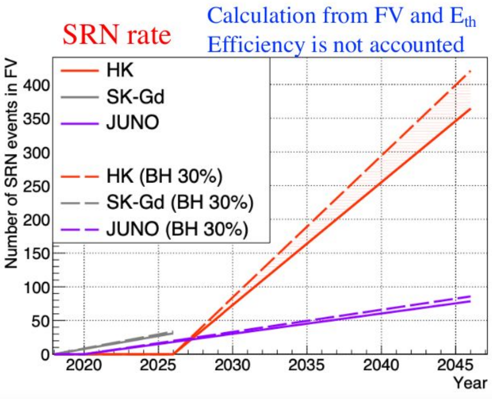
- Dominant signal v_e interaction is charged current quasielastic (CCQE)
- Potential background sources:
  - Neutral current interactions (v_e or v_μ) producing neutral pions or gammas
  - Muons from v_μ misidentified as electrons from v_e

Hyper-K's physics goals

Neutrino astrophysics

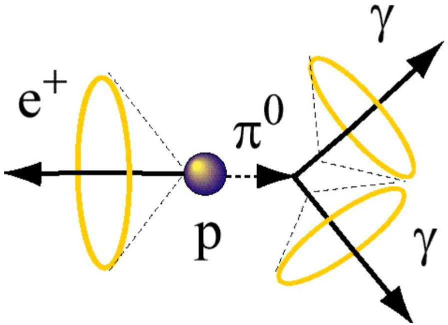
- Solar ν's: day/night asymmetry; hep ν's; ^{8}B ν spectrum upturn
- Supernova ν's: 1000's ν events for nearby supernova pointing, time & spectrum analysis; search for supernova relic ν's



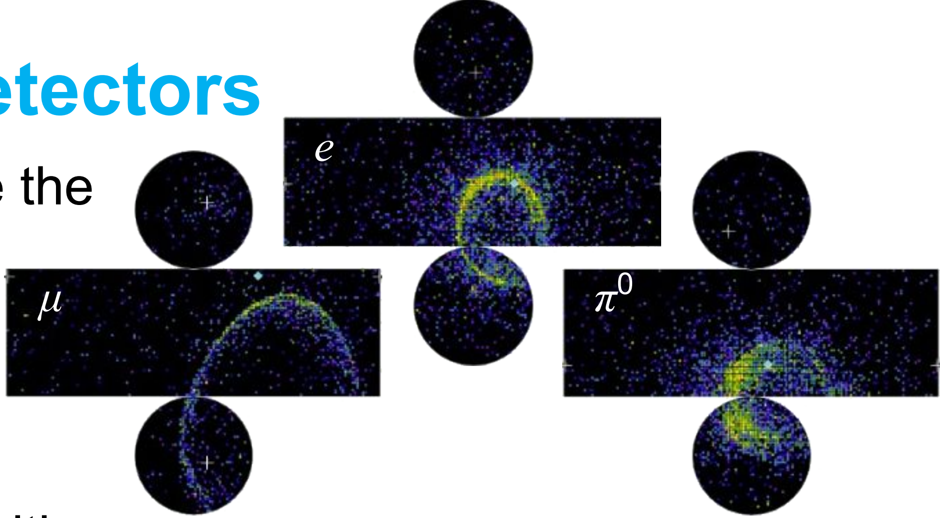


Proton decay

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

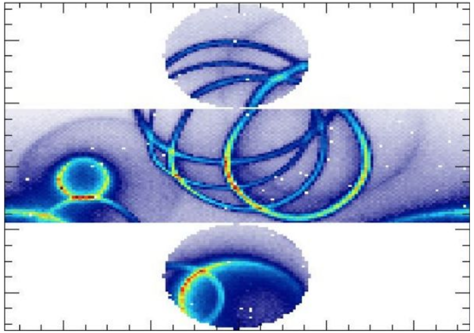


Reconstruction for WC detectors



Take raw detector data and determine the physics that occured

- Particle type identification
  - Separate signal events from background
- Particle momentum, direction, position
  - Kinematics essential to determine incoming neutrino energy for neutrino oscillation probability
- Separating & reconstructing multi-ring events
  - Events with multiple particles / rings contribute to both signal & background
  - Multiple neutrinos can interact around the same time in IWCD

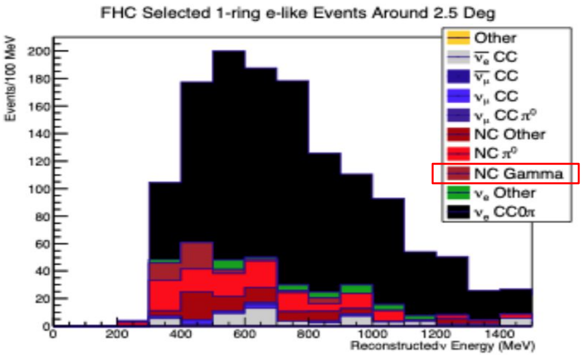


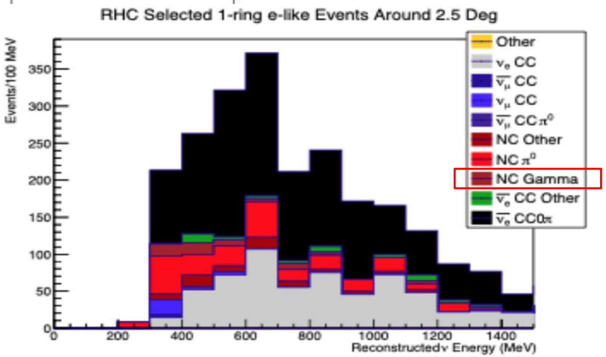
Physics Motivations

New opportunities beyond simple reconstruction improvement

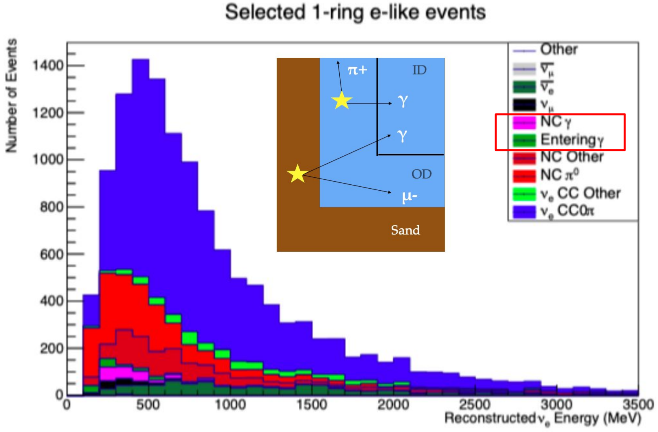
- NC γ discrimination and measurement

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

Machine learning reconstruction

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

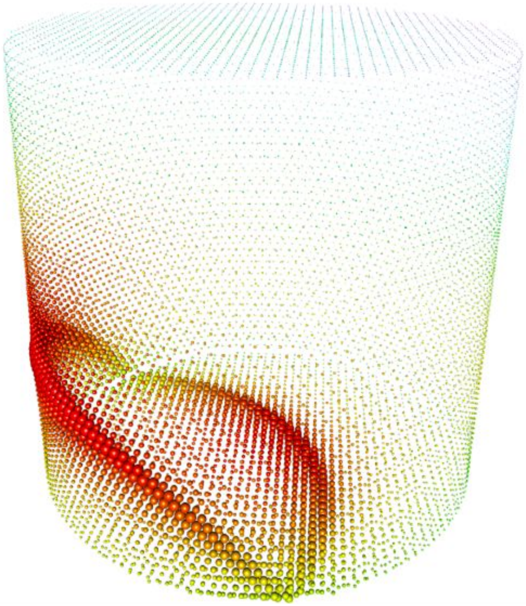
Common challenges for ML with WC detectors

- Cylindrical geometry
- High-resolution, sparse data

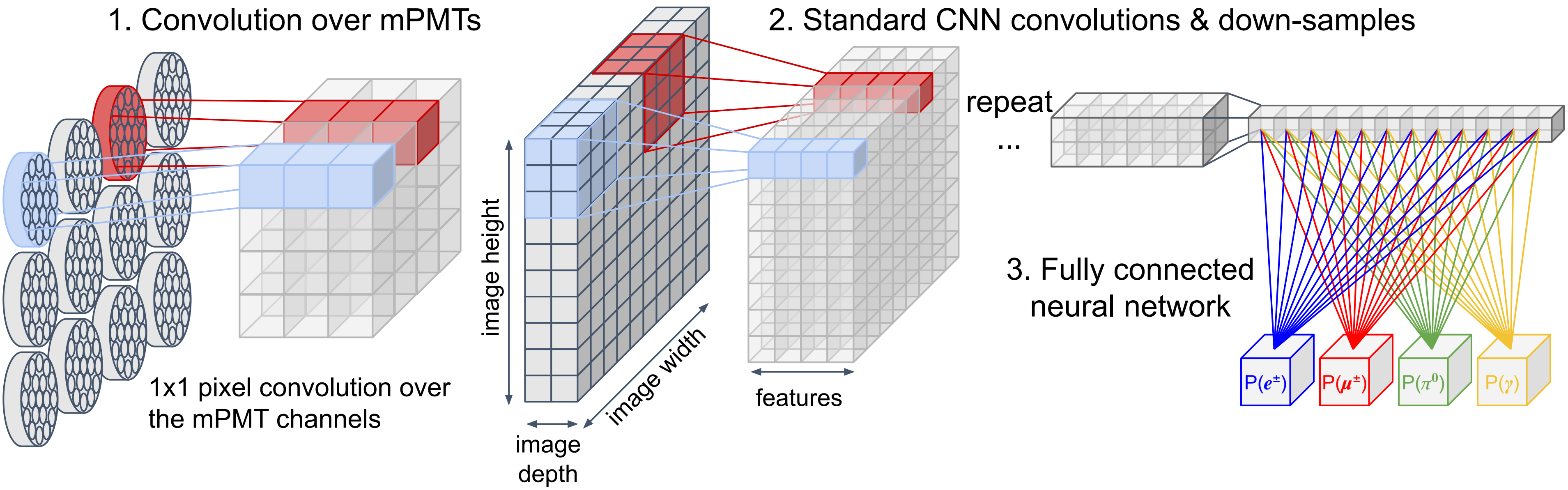
Many physics goals

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

WatChMaL.org



CNN Architecture

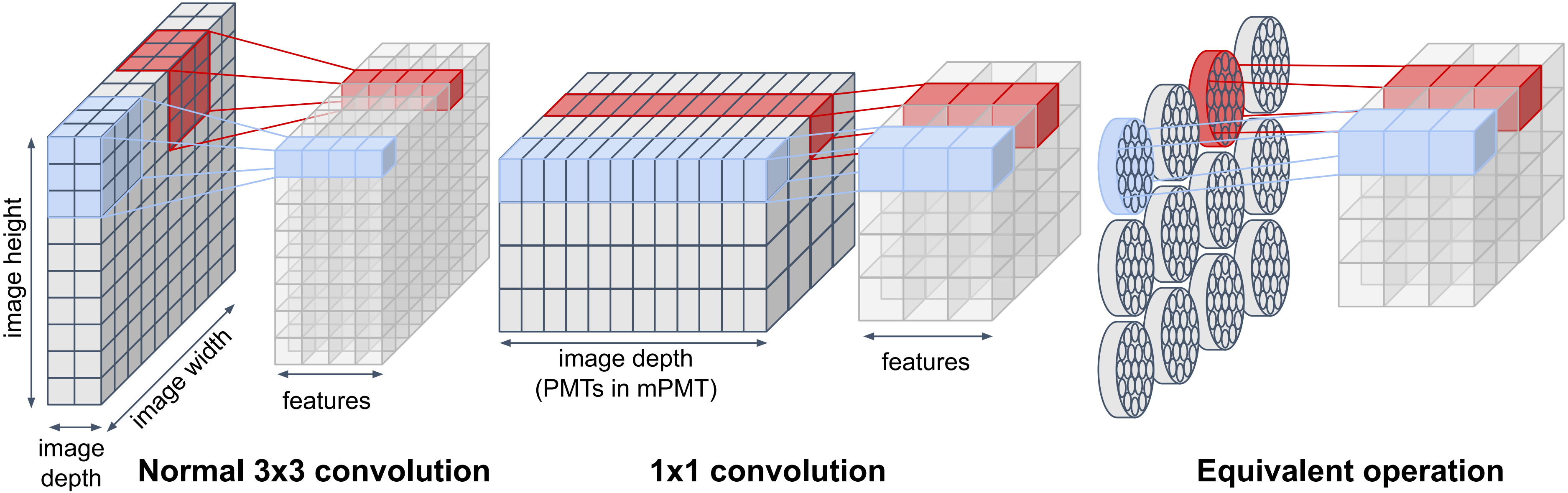


Network based on ResNet-18 CNN architecture [arXiv:1512.03385]

- Replaced initial 7x7 pixel convolution with 1x1 convolutions over all channels
  - Equivalent to convolution over the 19 PMTs within each mPMT

CNN architecture

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



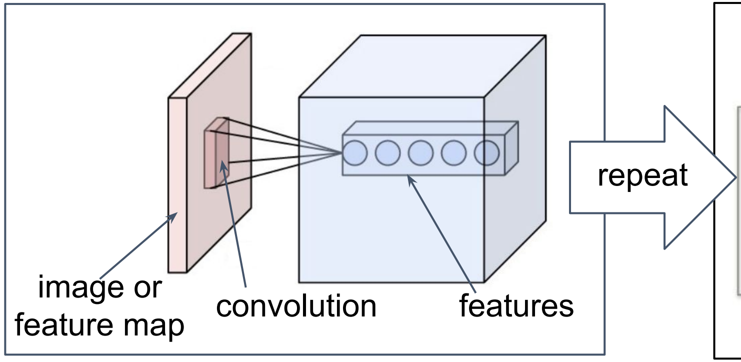
Normal 3x3 convolution

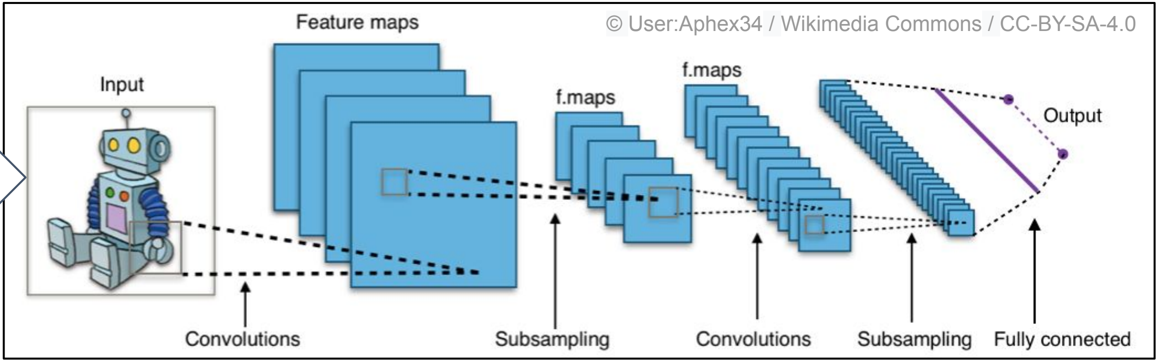
1x1 convolution

Equivalent operation

CNN architecture

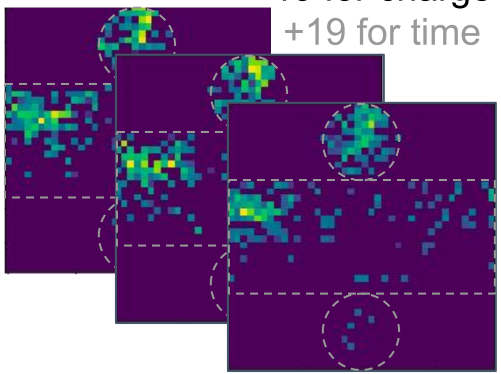
Convolutional neural networks hugely successful in image processing





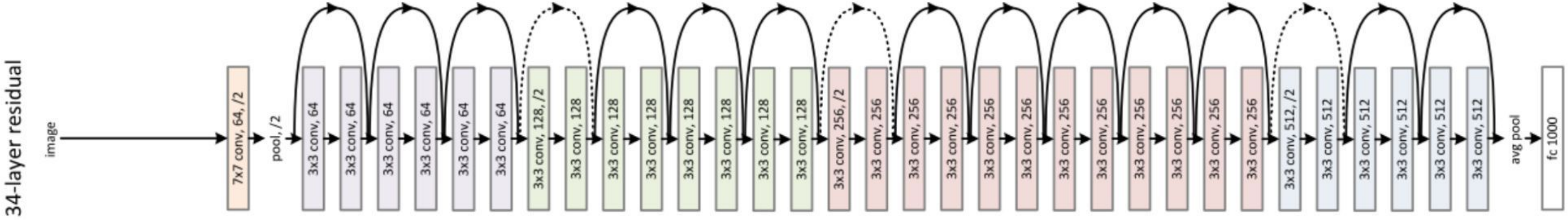
- Start with image with pixel values (‘features’): T and Q at each PMT
- Scan many small (e.g. 3x3) convolution kernels across image
  - Increases number of features
- Downsample image (e.g. 2x2 max-pooling)
  - Decreases number of pixels
- End with 1-D array of features, feed into traditional fully-connected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

CNN architecture



Full cylinder of mPMTs is unwrapped onto 40x40 image

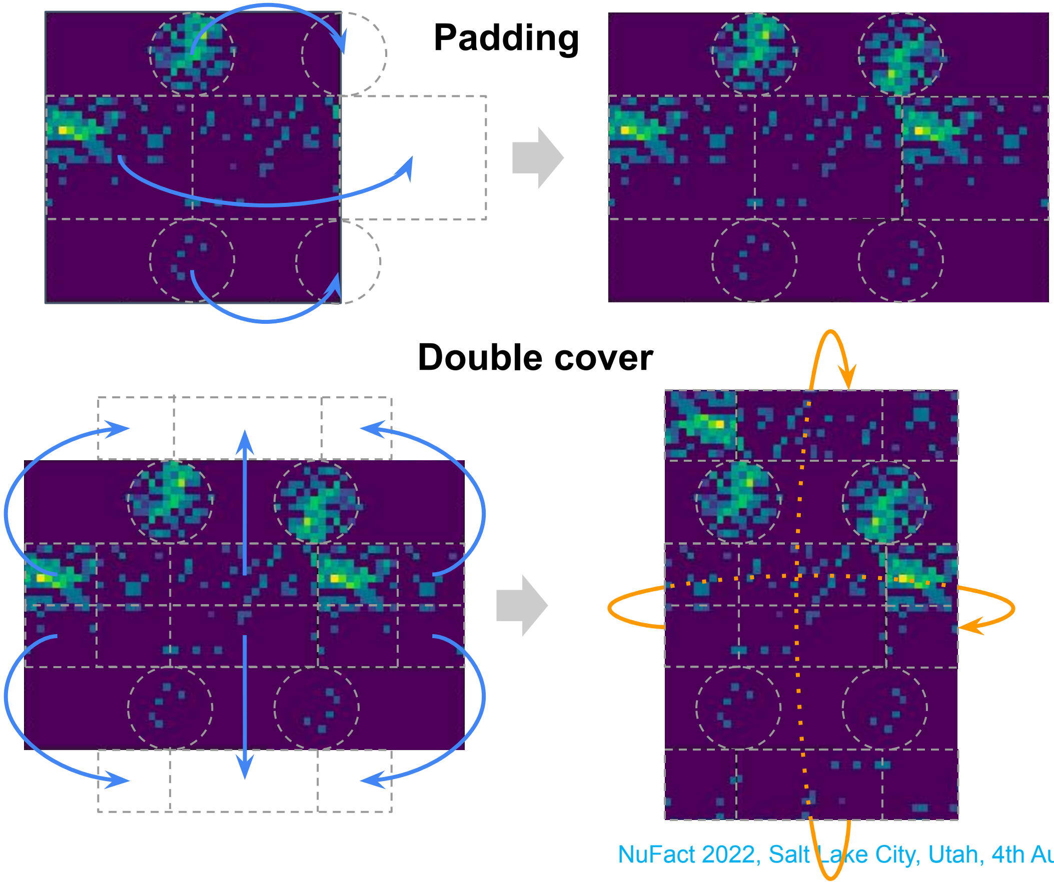
- 38 channels: charge & time of 19 PMTs per mPMT
- No special treatment for geometrical effects at boundary between barrel and end-caps
- Data augmented by reflecting / rotating around tank axis



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

‘Double cover’ images



‘Padding’ the image improves accuracy for some events

- Original image ‘slices’ along barrel at arbitrary position
- Some events have rings that span this slice
- Repeat part of the image after rotating tank to help CNN learn events where ring is sliced

Rearranging and duplicating in a more complex pattern has additional advantages

- All segments appear exactly twice
- Circular boundary conditions in both directions
- Minimal blank space

Topological map to square

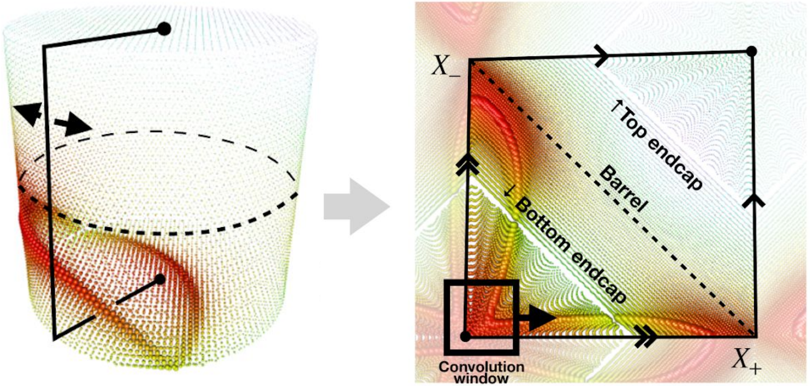
$$X_{\pm} = W(\rho, z)\frac{\pi \pm \phi}{2\pi} \qquad W(\rho, z) = \sqrt{\frac{\rho^2 + 2Rz + RH}{R^2 + RH}}$$

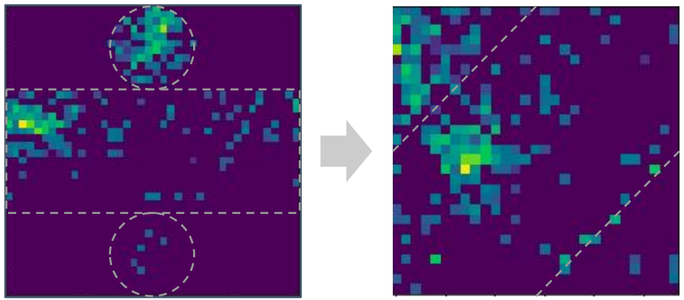
Solve differential eq. for constant Jacobian

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

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

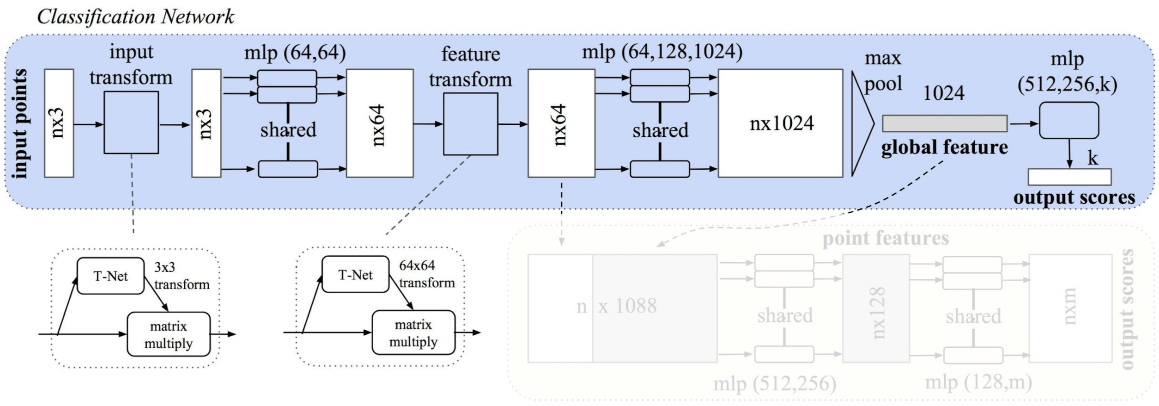


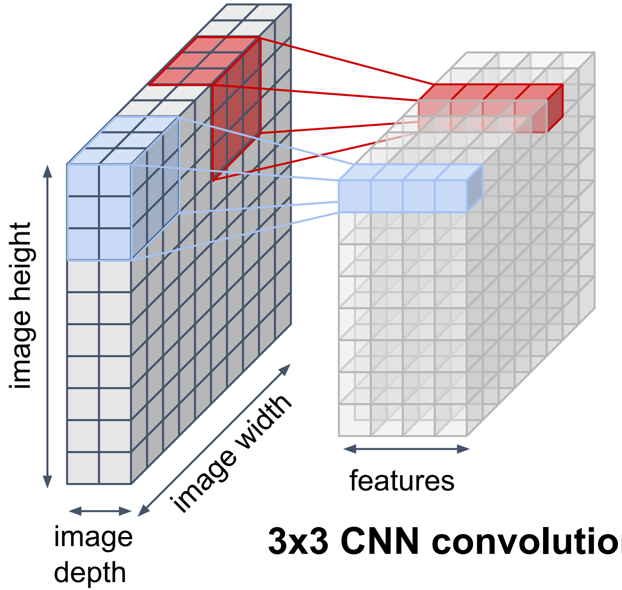


PointNet architecture

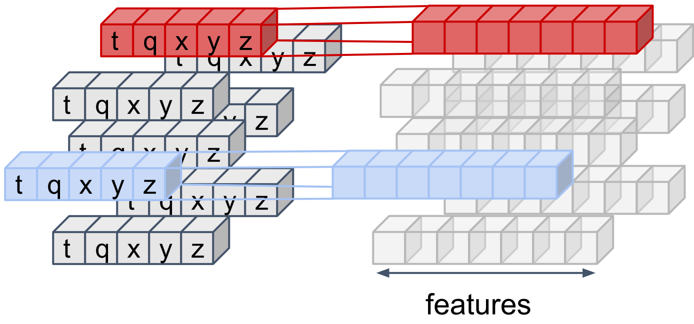
PointNet designed to work on 'point clouds' rather than images of pixels

- Each hit PMT is a 'point' with time, charge & position, not fixed to grid
- Convolution-like operations act on each point's charge, time and position
- Learn global transformations applied to all points
- Single pooling layer from all points to 1D array
- Can apply to any detector geometry





3x3 CNN convolution

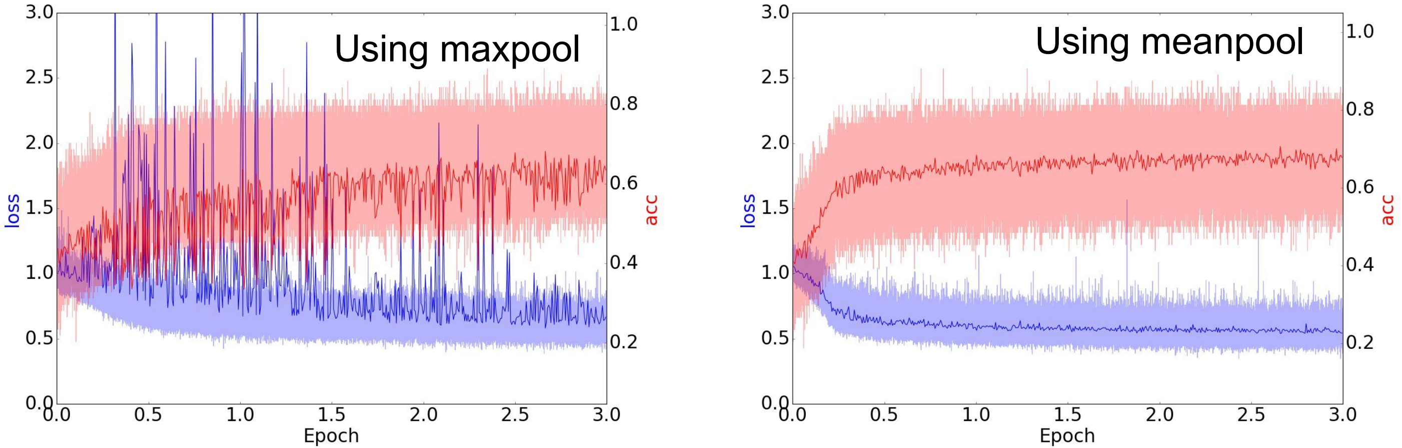


PointNet MLP (1x1 convolution on point cloud)

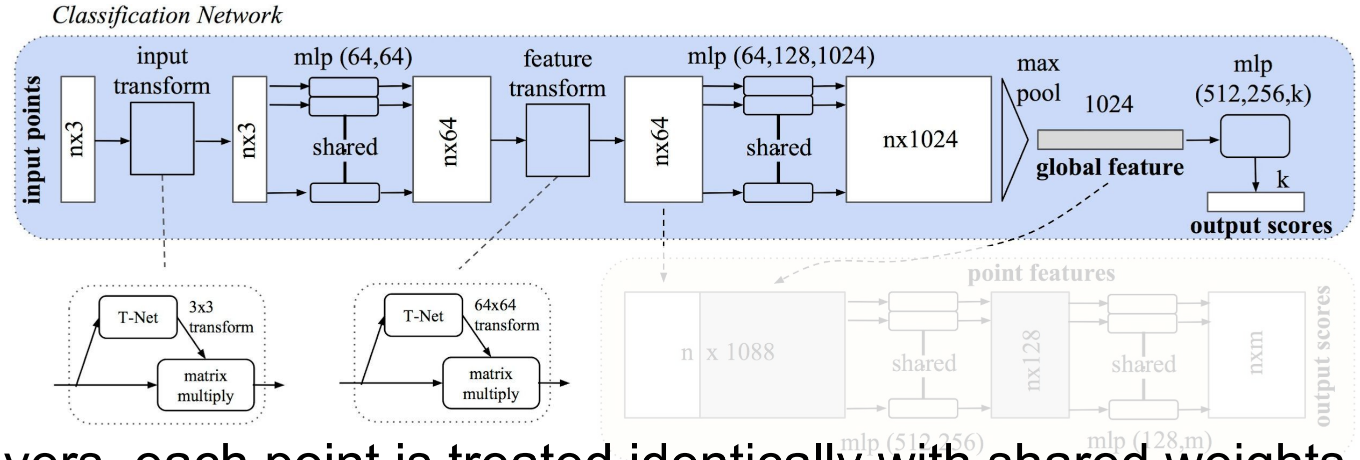
PointNet architecture

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

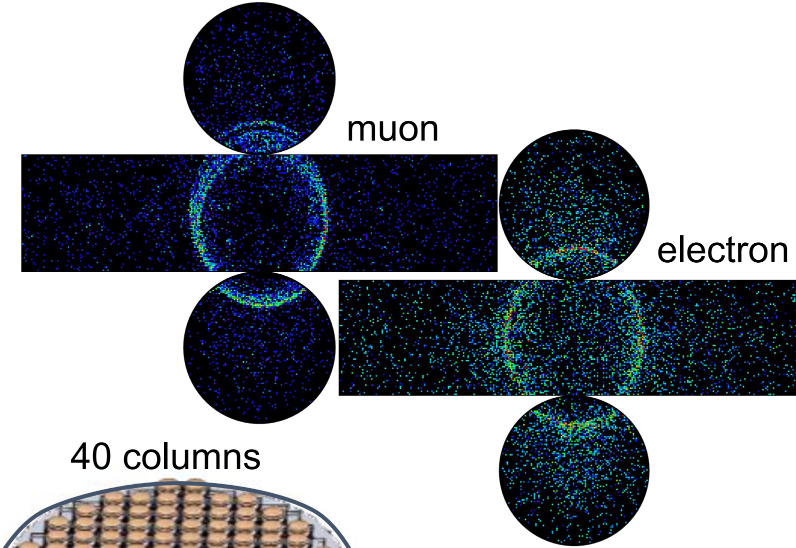
Particle type classification

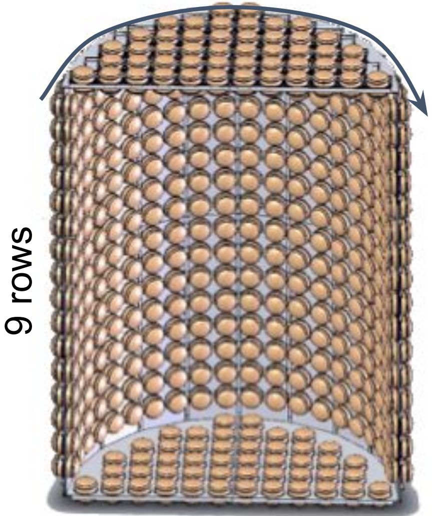
Initial studies to classify $\mu\,/\,\pi^0\,/\,e\,/\,\gamma$ particle types

- 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 type in IWCD detector

- 0 - 1 GeV energy above threshold
- Uniform positions, isotropic directions
- Split full dataset into 50% : 10% : 40% for training : validation : testing



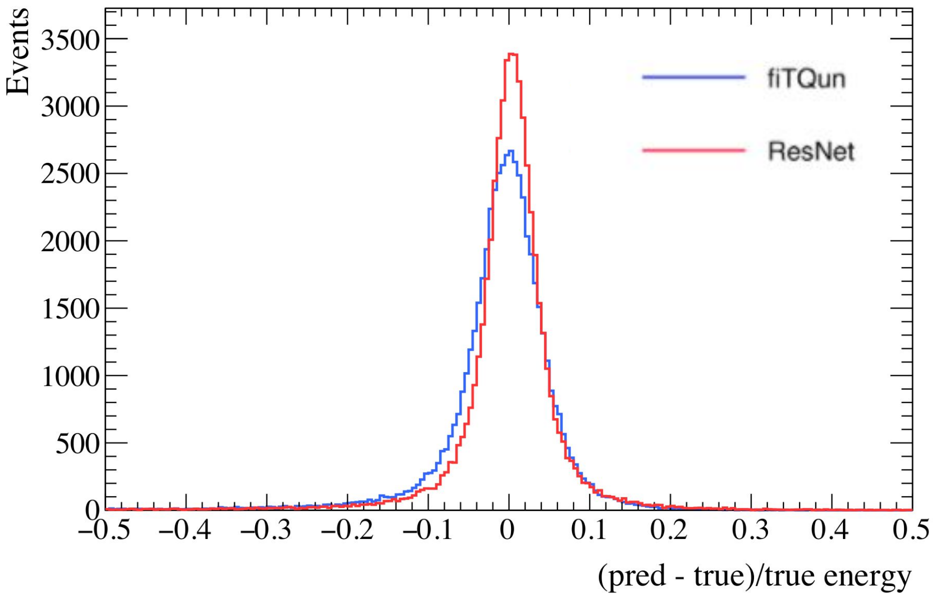


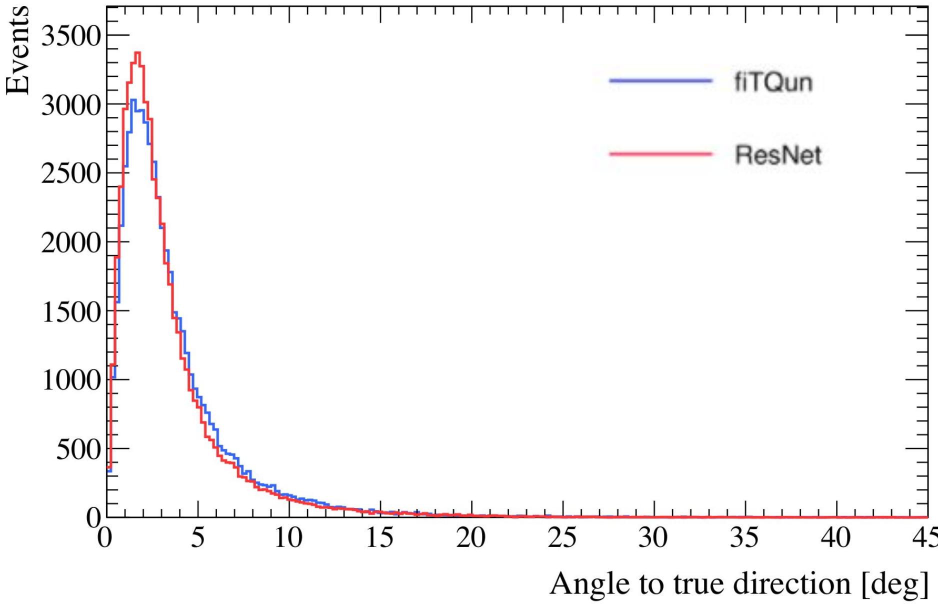
mPMT: 19 x 3” PMTs

Position, direction, energy reconstruction

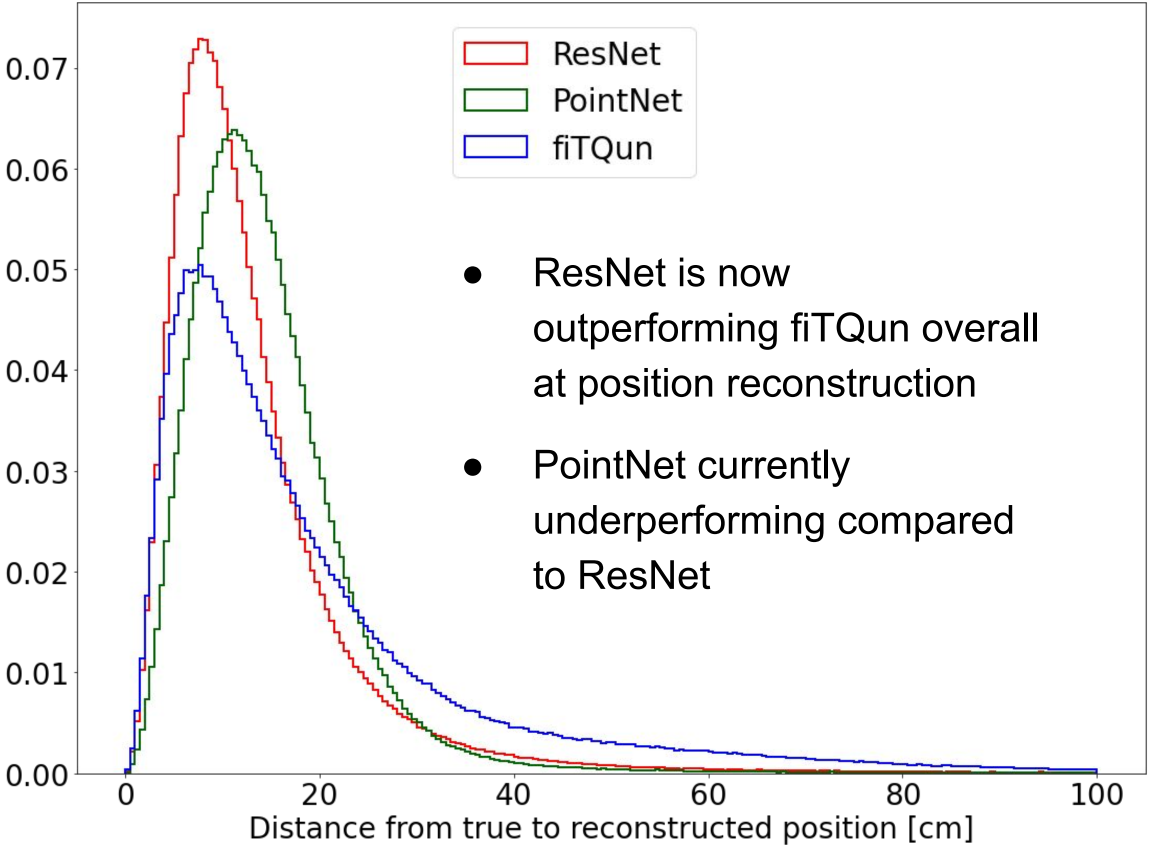
Similar ResNet and PointNet architectures as used for classification

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

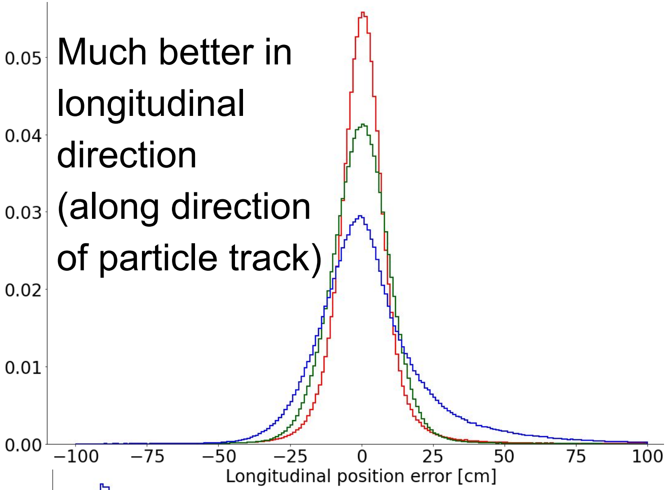


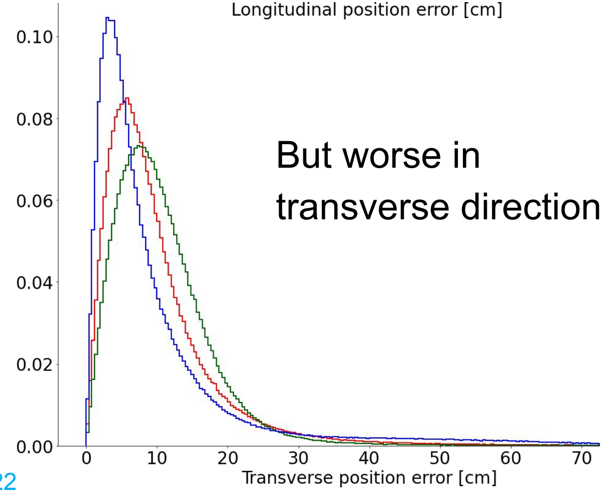


Position, direction, energy reconstruction

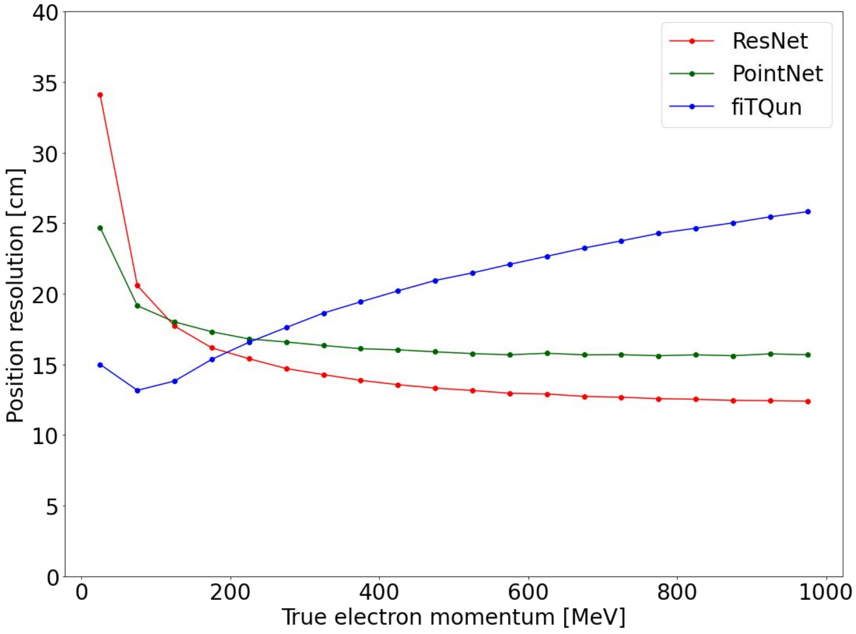


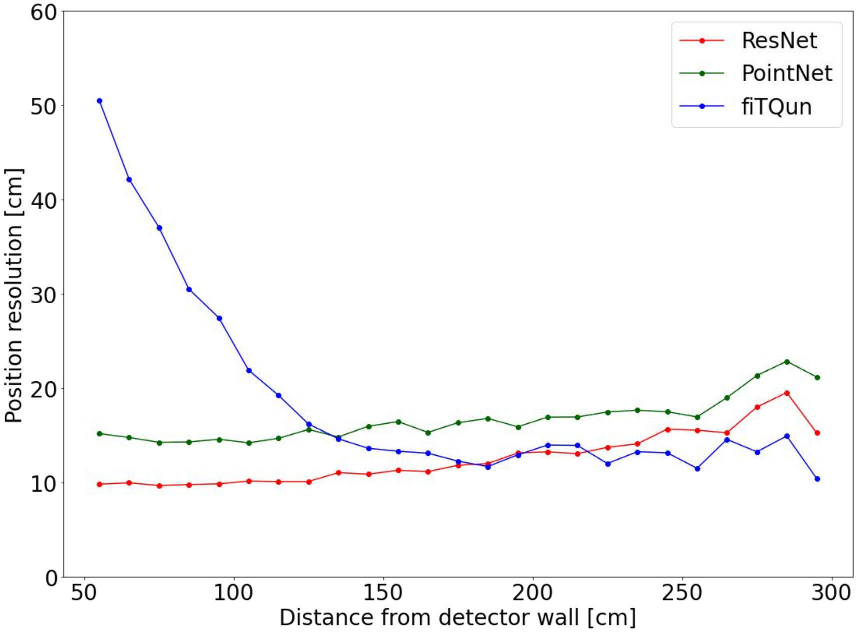
- ResNet is now outperforming fiTQun overall at position reconstruction
- PointNet currently underperforming compared to ResNet





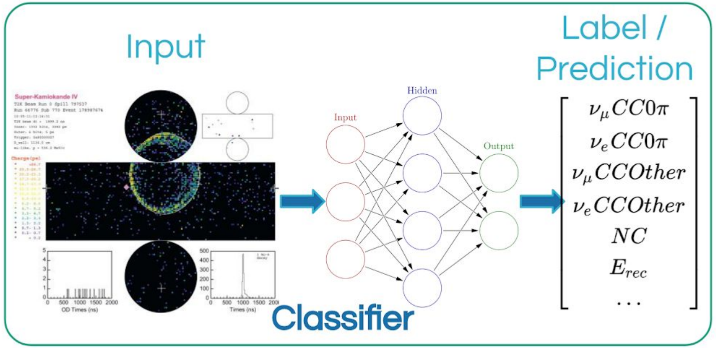
Position, direction, energy reconstruction

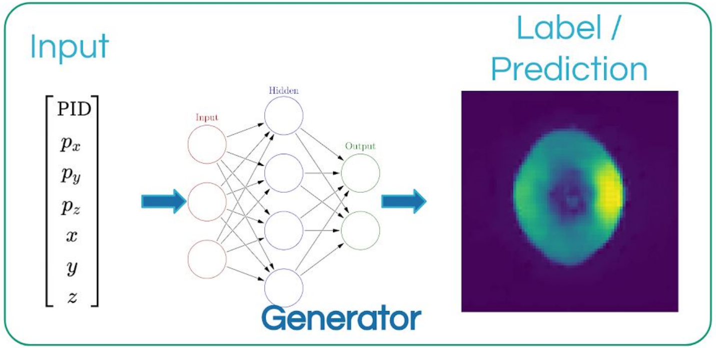




- Improvement in reconstruction with ML mainly in events close to detector wall
  - Approximations in likelihood calculation break down when close to PMTs
  - Could allow expansion of detector fiducial volume to allow increased statistics
- ML reconstruction could be improved at lower energy
  - potentially struggles to learn reconstruction of sparse events

Cherenkov ring generator

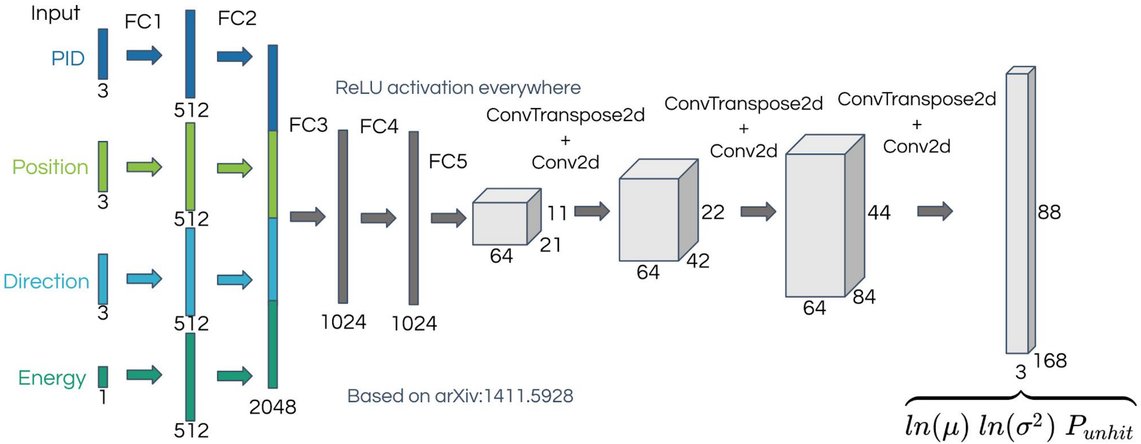




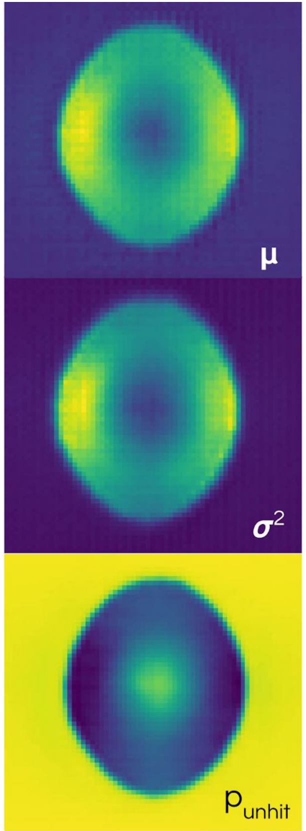
Investigating hybrid method using generative network

- 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



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



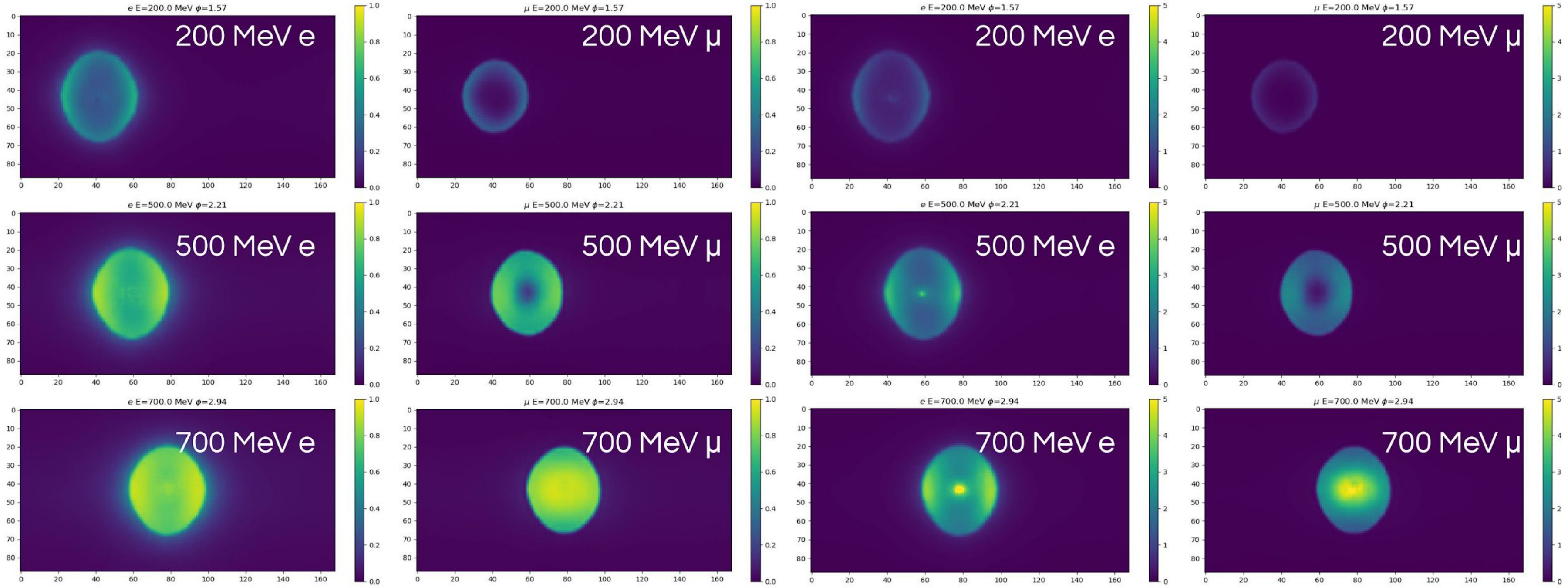
Network outputs likelihoods for hits observed at PMT

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

Cherenkov ring generator

Hit probability

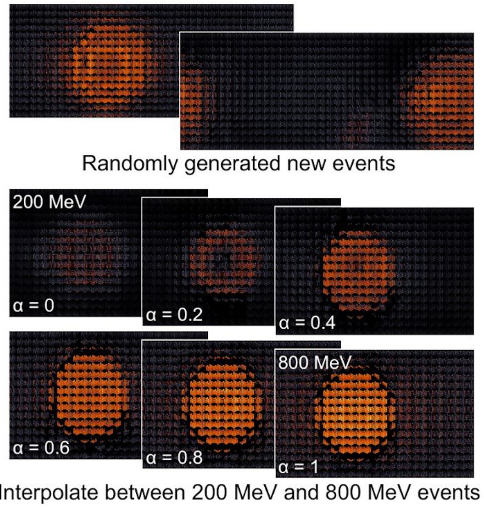
Hit probability
x
Mean charge



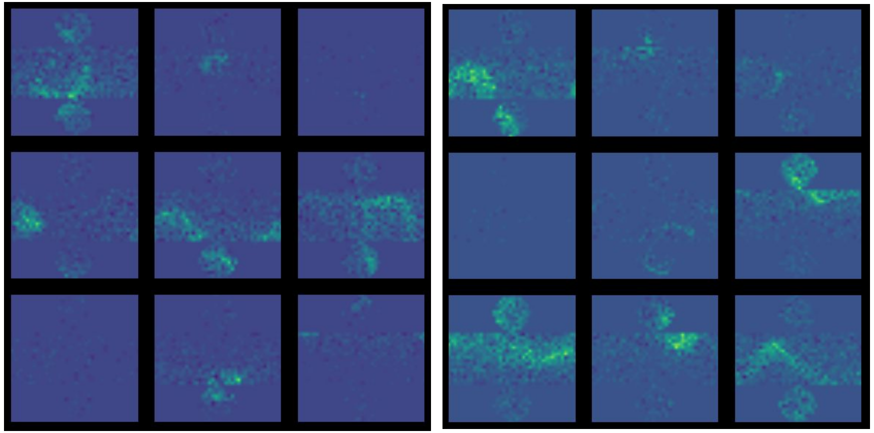
Generative networks

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



arXiv: 1911.02369



GAN generated events Geant4 simulated events