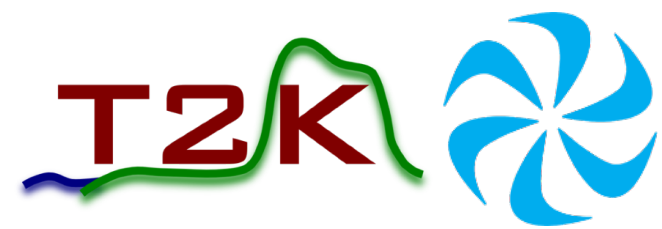


Successes and pitfalls: applying ML to T2K & SuperK

Félix Cormier
15/06/2026
NPML Workshop 2026



Introduction



- Tokai to Kamioka (T2K) is a long-baseline neutrino experiment in Japan
 - Muon neutrinos travel 295km to far detector Super-Kamiokande (SK)
- Measuring rates of neutrino appearance and disappearance at the far detector gives us access to physical quantities
 - **Dip** in energy spectrum gives us strength of oscillation parameters θ
 - Exact energy at **dip** gives us mass difference between neutrino flavours Δm
 - Differences in oscillation rates between neutrino and anti-neutrino gives us degree of CP violating phase δ_{CP}
- In this project, focus ML techniques for **muon neutrino disappearance at far detector, SK**
 - Need solid **classification performance between muons and charged pions**
 - Good **energy resolution**
 - **Extend fiducial area** for reconstruction by improving position/direction reconstruction

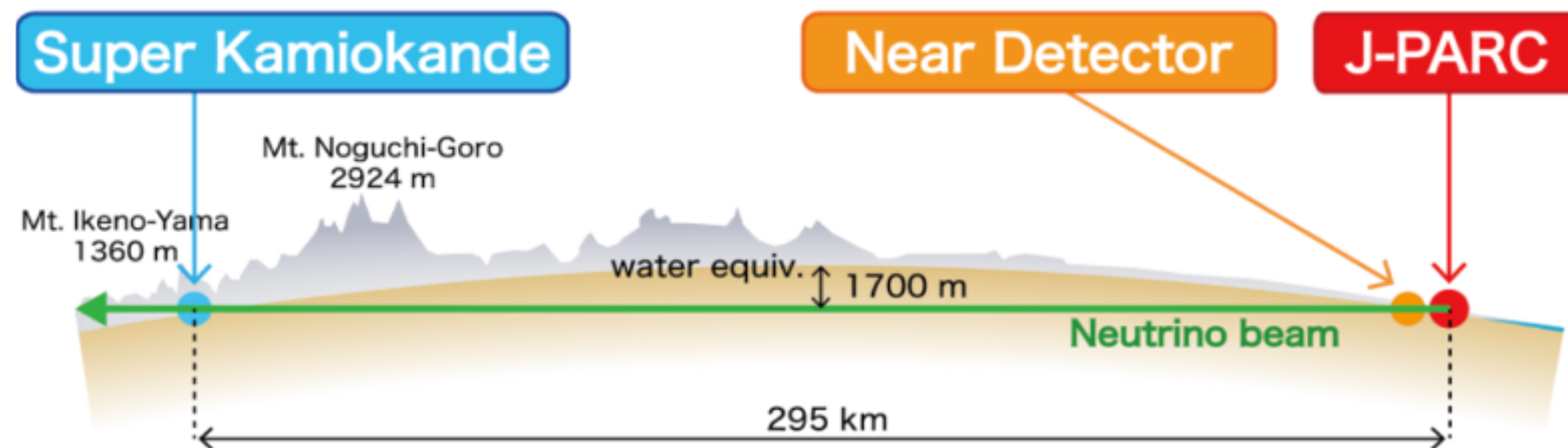
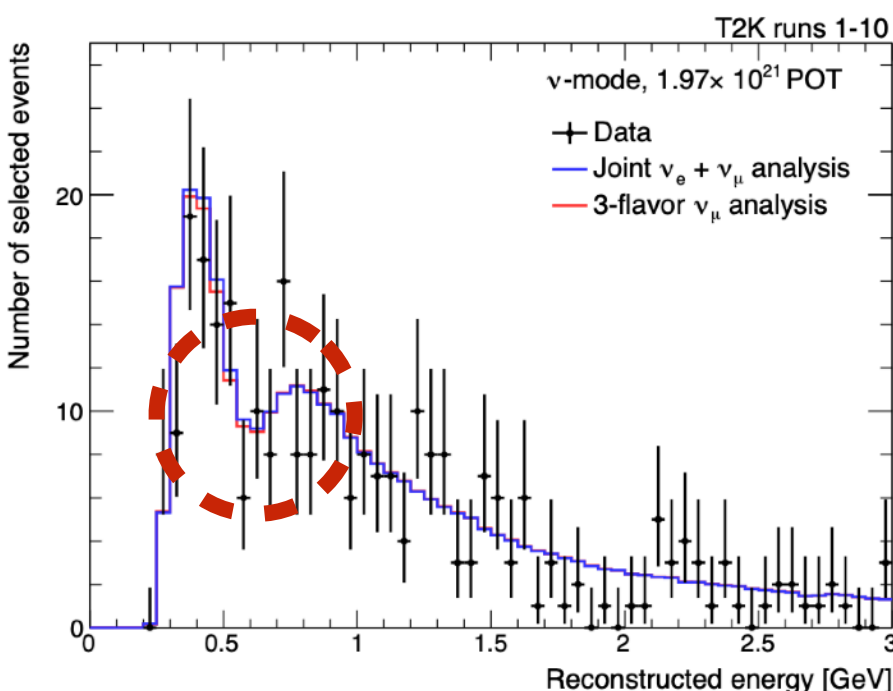
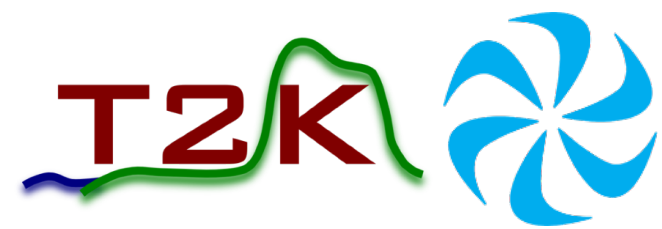
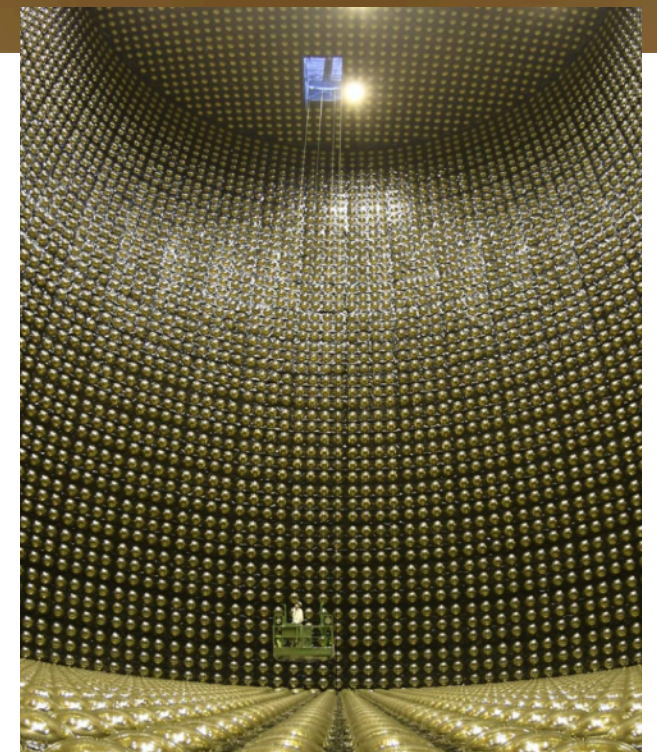
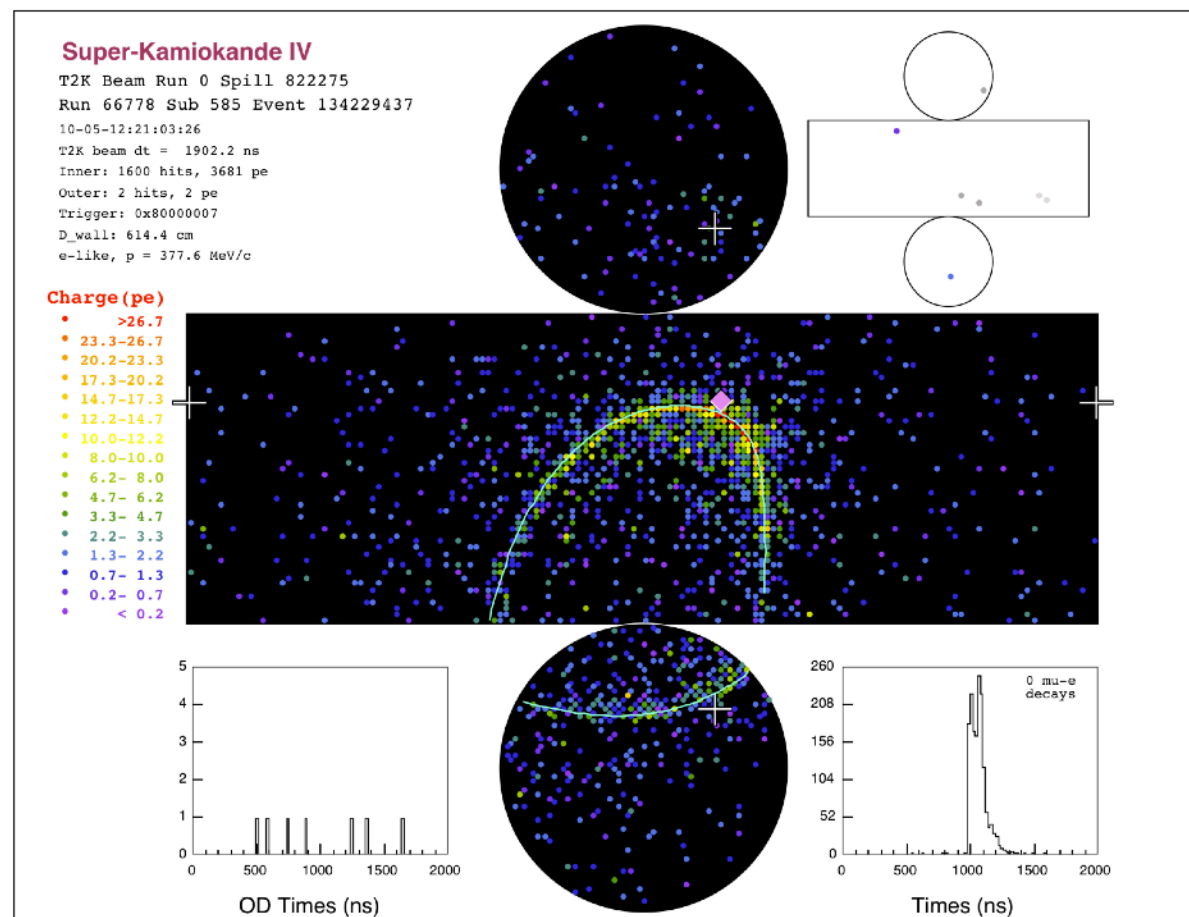
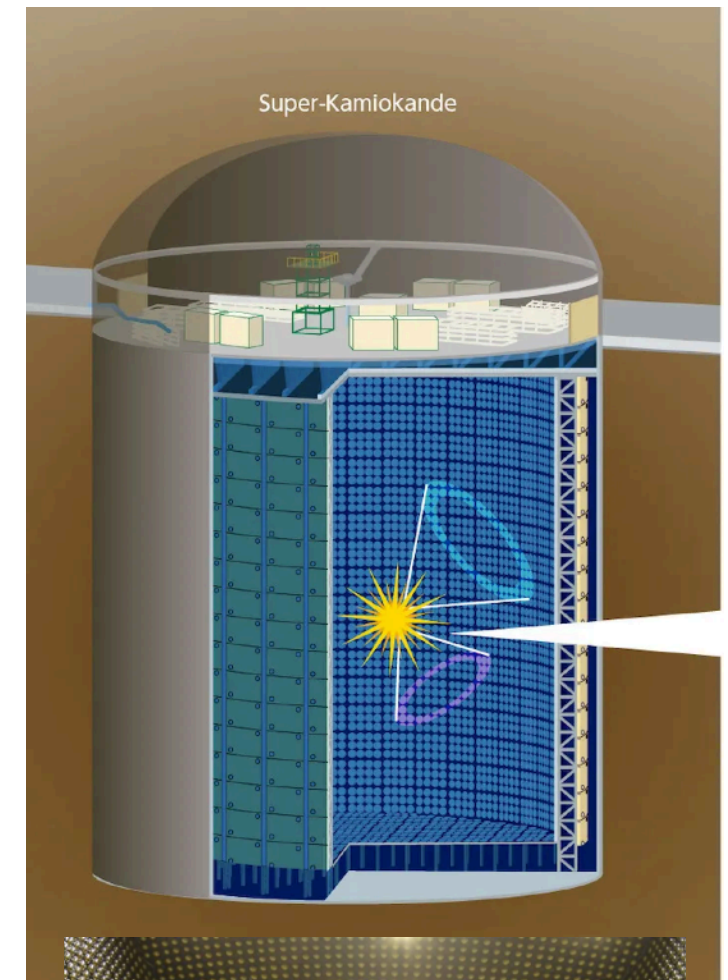


Diagram showing the experimental set-up of T2K

Water Cherenkov Detectors - SuperK

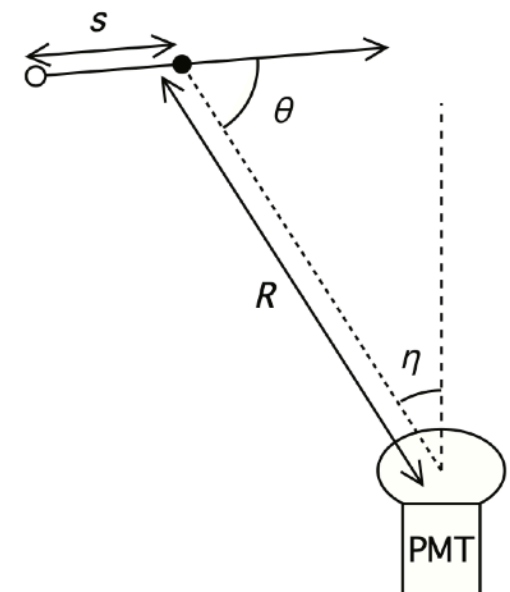


- Water Cherenkov detectors typically consist of an Inner Detector (ID) and Outer Detector (OD)
 - ID is ringed with inside-looking PMT, about 11500 50cm PMTs in SK
 - OD has outside-looking PMTs for background removal, about 1500 PMTs in SK
- Ultrapure water, with recent addition of Gadolinium for neutron ID
- Focus on accelerator neutrino program with T2K in this project

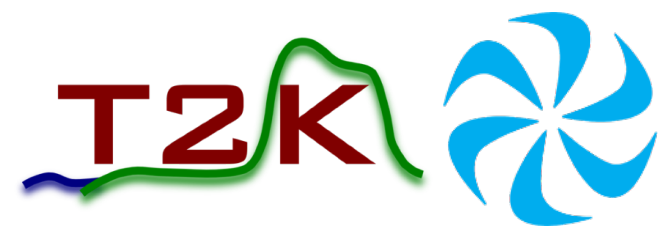


- Current reconstruction in high-energy T2K/SK is built in a maximum likelihood fit
- Calculates likelihood of charge and time on every PMT based on hypothesis of particles per ring
 - Hypothesis multiplicity is very high
 - Slow to compute
- Works very well for differentiating electrons from muons & kinematic reconstruction

$$L(\mathbf{x}) = \prod_j^{\text{unhit}} P_j(\text{unhit}|\mu_j) \prod_i^{\text{hit}} \{1 - P_i(\text{unhit}|\mu_i)\} f_q(q_i|\mu_i) f_t(t_i|\mathbf{x})$$



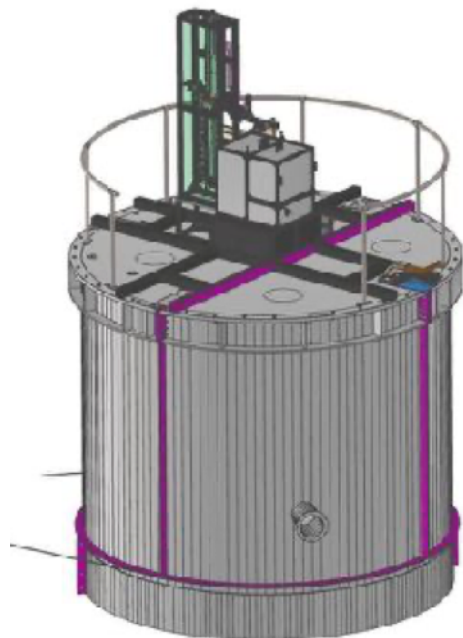
The WatChMaL Collaboration



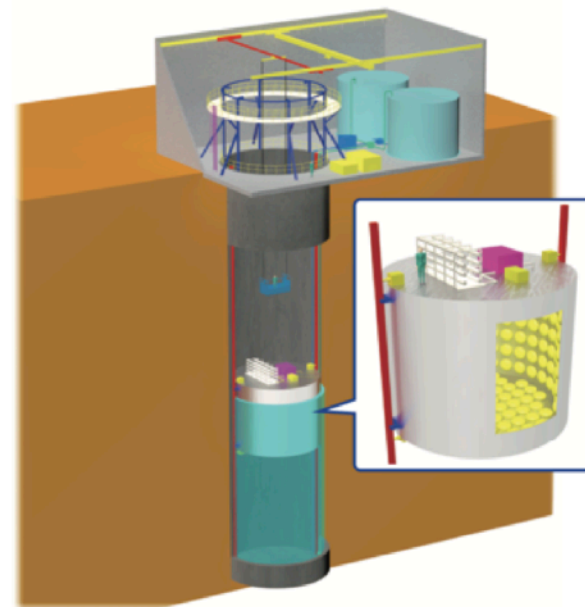
- This project is part of the **Water Cherenkov Machine Learning Collaboration**
- Common data generation, pre-processing and training frameworks are shared by members
- Currently have members in experiments
 - **T2K/SuperK**
 - Water Cherenkov Test Experiment (WCTE)
 - Intermediate Water Cherenkov Detector (IWCD)
 - Hyper-Kamiokande Far Detector



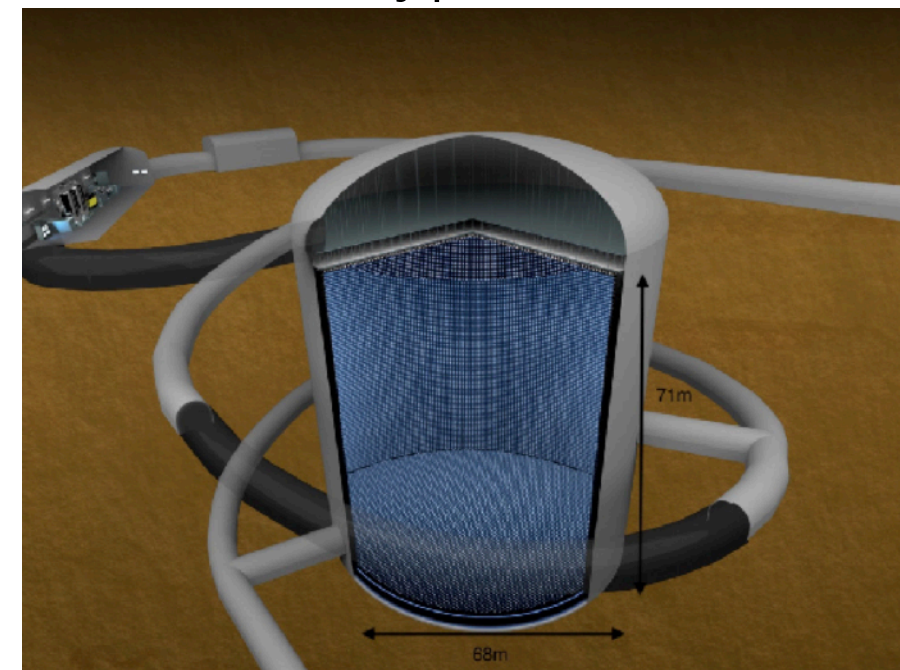
WCTE



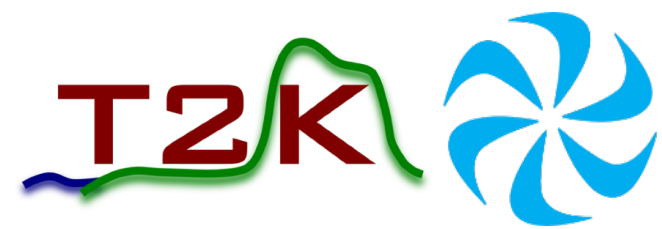
IWCD



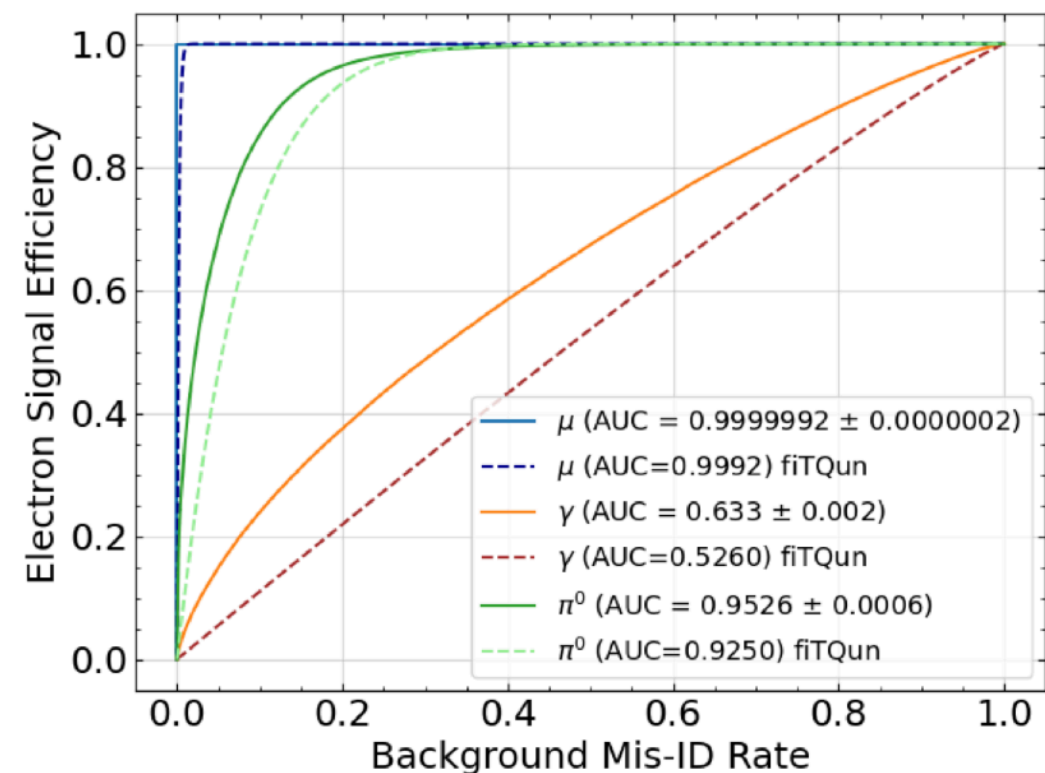
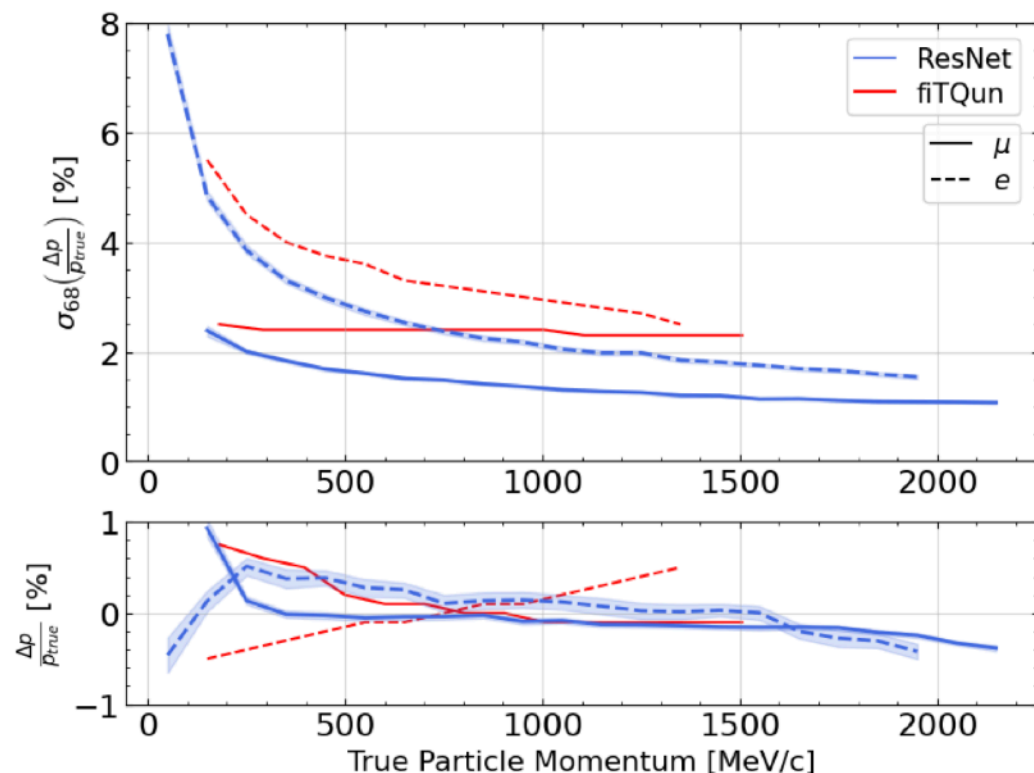
HyperK



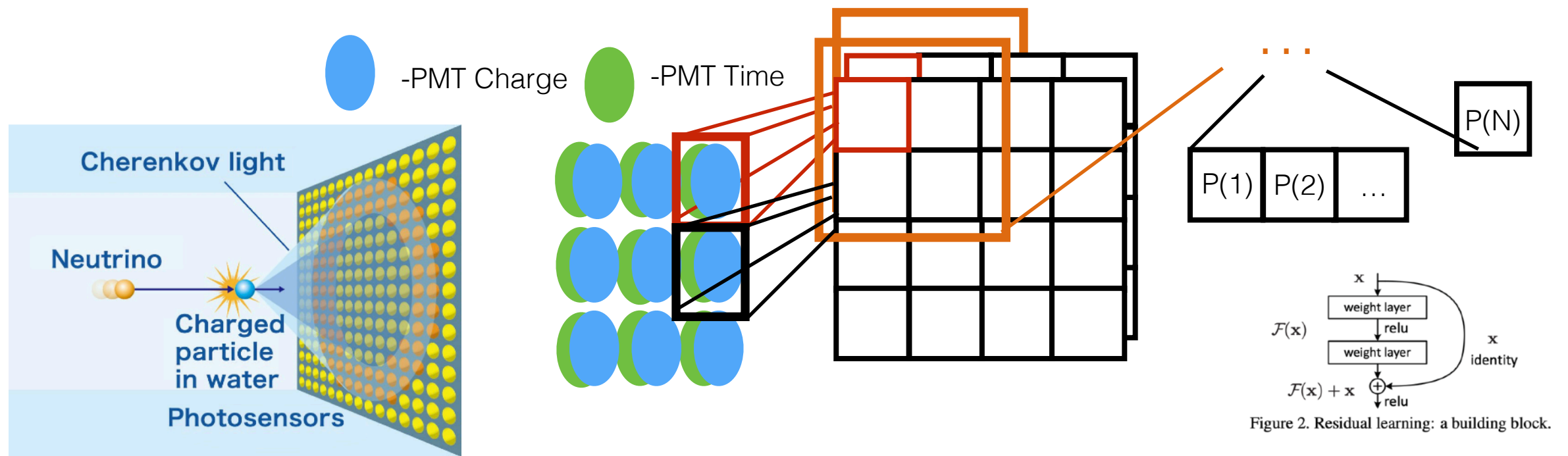
Ongoing WatChMaL Projects



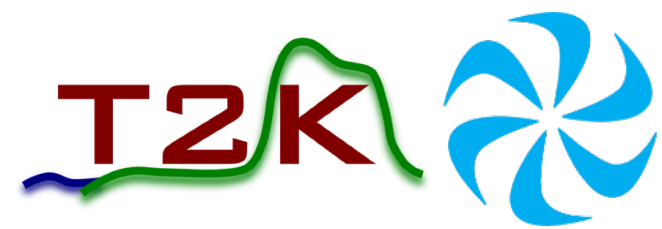
- Water Cherenkov Test Experiment
 - Robust suite of classification ($e/\mu/\pi^+/\pi^0/\gamma$) and regression networks trained ([link](#))
 - Current effort is on enhancing digitization in simulation to reduce data/MC differences
- Intermediate Water Cherenkov Detector
 - Optimization of networks for classification (focus on γ) ongoing
 - Testing ongoing on simulation of neutrino beam
- Hyper Kamiokande Far Detector
 - Results using ResNet for reconstruction recently put on [arxiv](#)
 - Since these results, enhancements on general performance & momentum bias by converting to prediction of 3-momentum
 - Transformer & Foundation model studies by Shuoyu. - see poster!



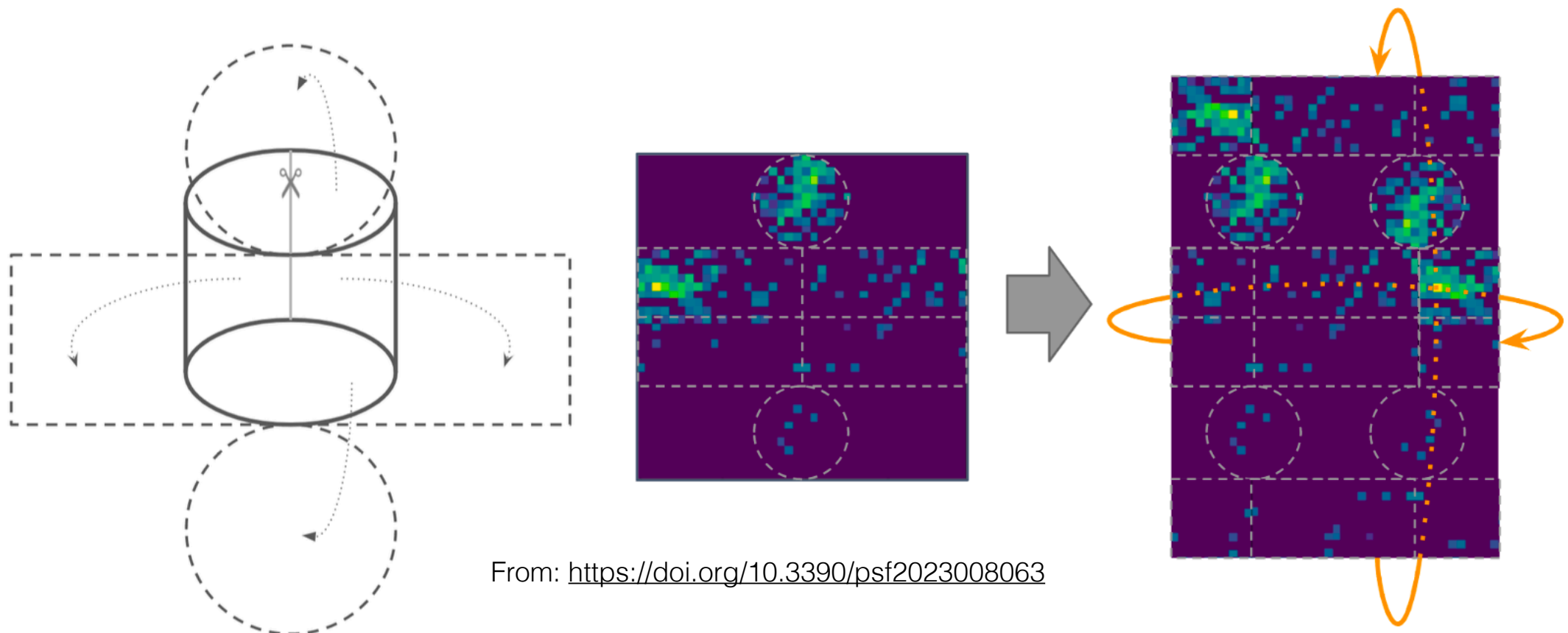
- Across WatChMaL collaboration, we tested many different types of networks
 - Residual Networks (ResNet - CNN)
 - Point cloud networks (PointNet)
 - Graph Neural Networks
 - Transformer-based Networks
- For Super-Kamiokande reconstruction, found that **ResNet** gave best results



Double Cover & Training dataset

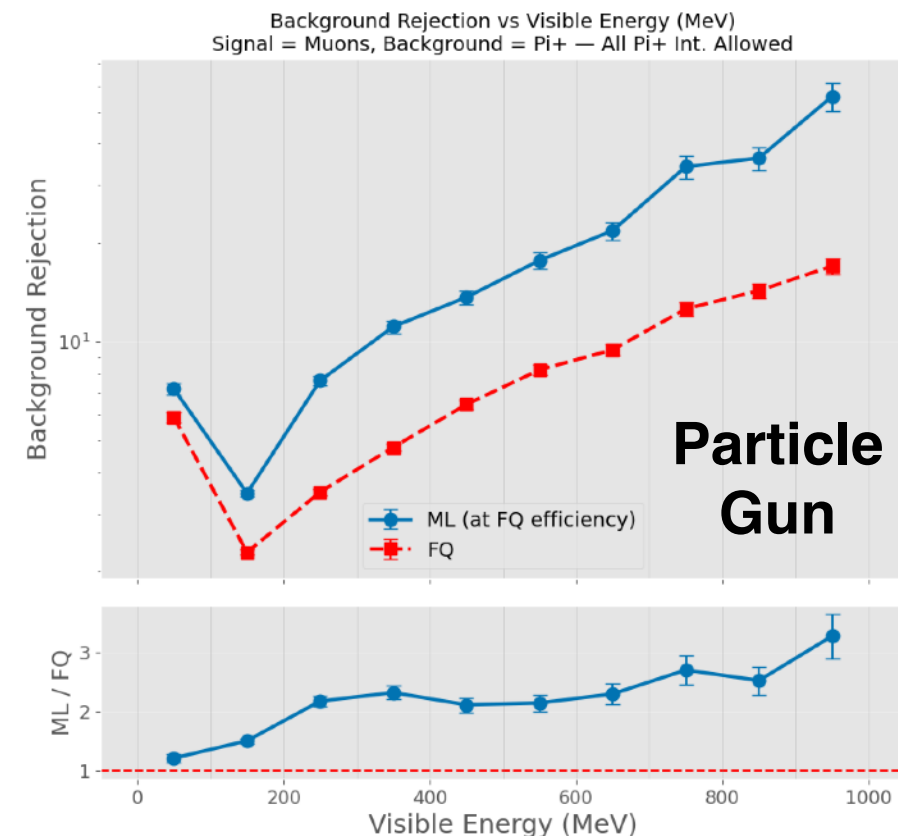
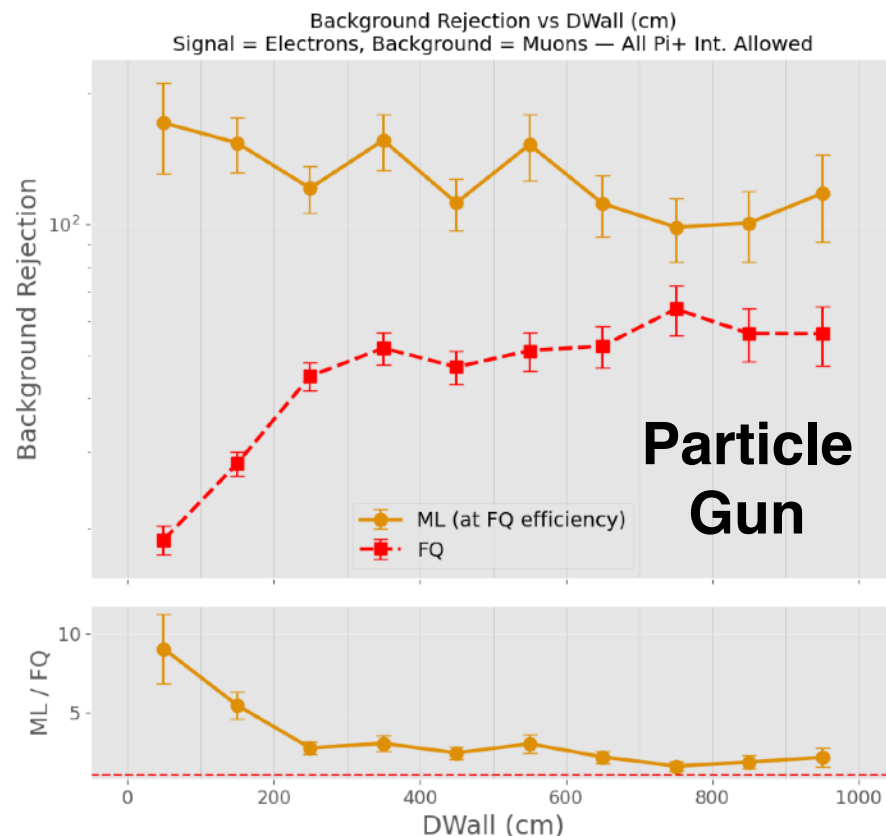
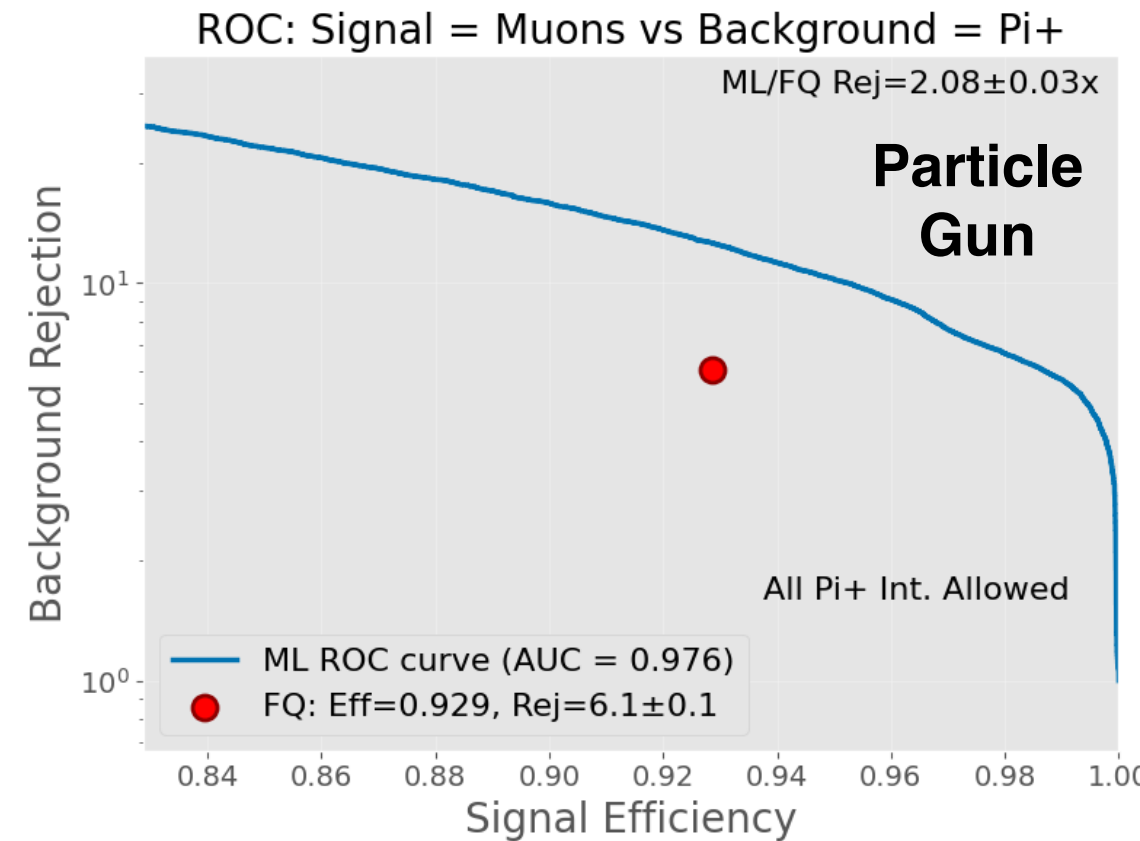


- One of main challenges of using CNN is ‘unrolling’ our cylindrical detector into 2D image
 - Double cover & circular padding - which repeats the detector, removed the arbitrary choice of unrolling location & reduced azimuthal dependency
- Training data simulation
 - Samples of 5M events each of μ , e , π^+
 - Kept general, uniform energy distribution above Cherenkov threshold to 2GeV, uniform position & isotropic direction simulation
 - Did a general hyper-parameter scan for both ResNet hyper parameters as well as pre-processing of input values



Successes: Classification

- 1 classification ResNet trained to differentiate $e/\mu/\pi^+$
- Focus on μ/π^+ as we want to target muon neutrino disappearance
- Can get over **2x background rejection** with same signal efficiency as fiTQun
- Can look at per-variable performance
 - Dwall: Much improved performance close to the ID wall - could loosen some fiducial cuts to increase signal
 - Visible energy: harder to improve at low visible energy

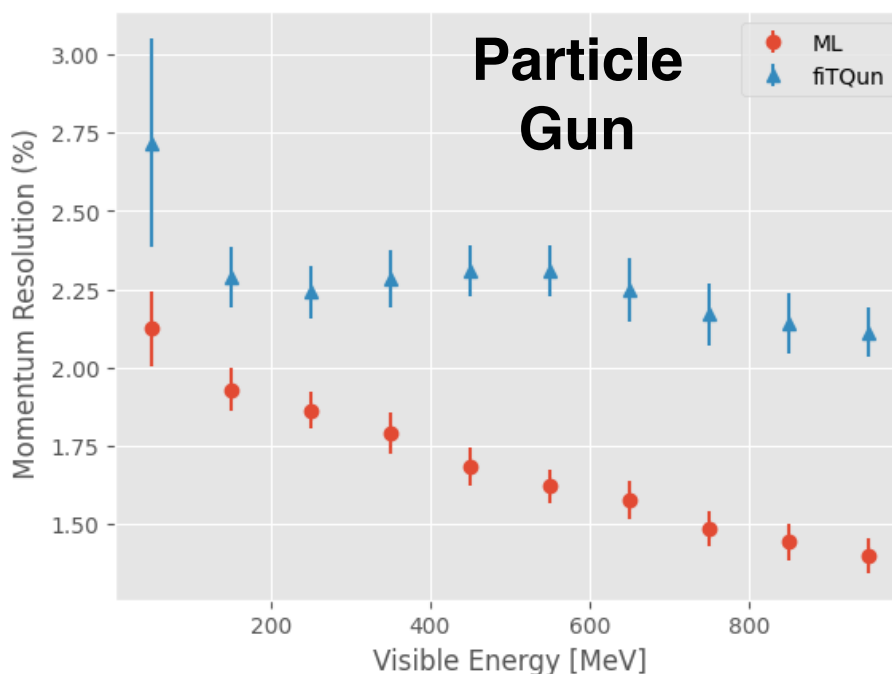
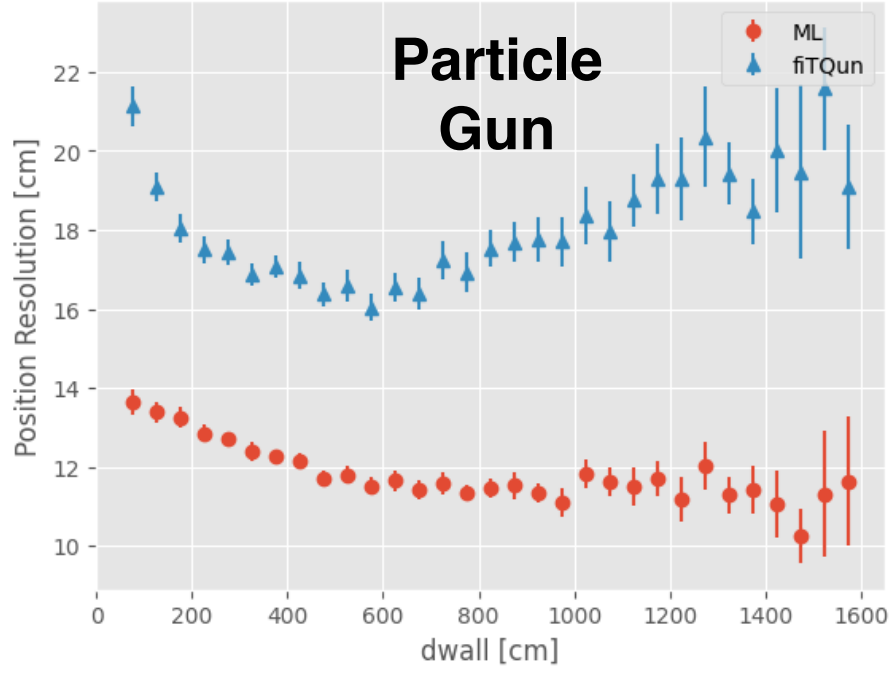
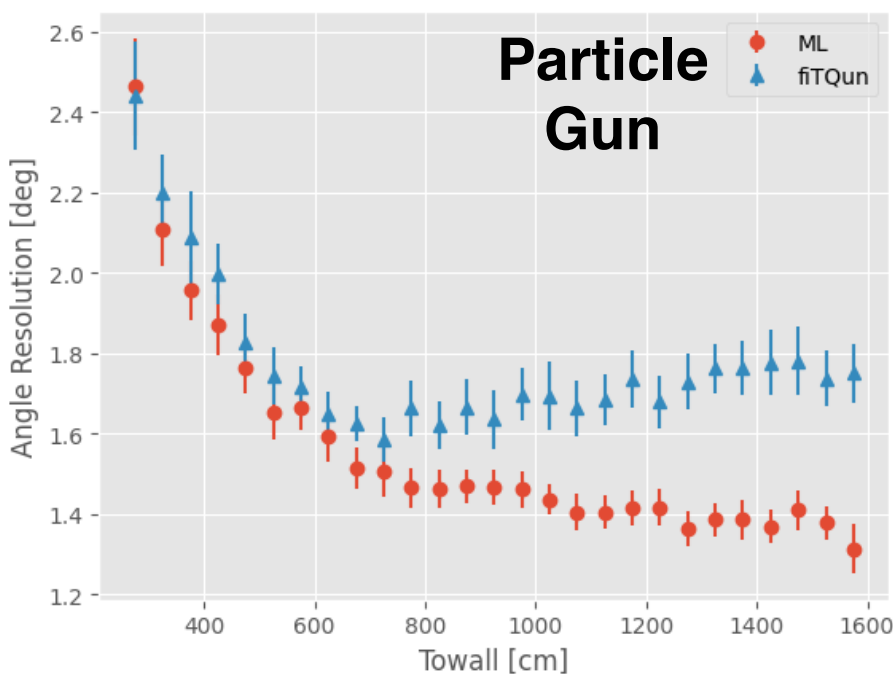


Successes: Regression

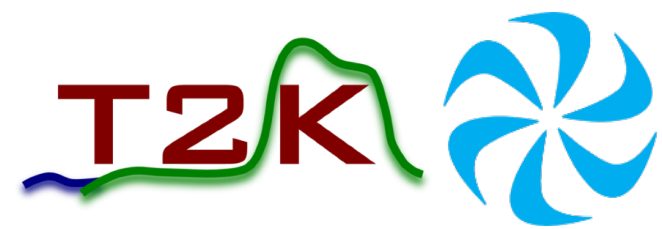


- Ultimately we want to reconstruct muons and electrons
- Train separate ResNet for e/μ for
 - Momentum/Position/Direction
 - 6 regression networks in all
- We can compare performance in particle gun sample in testing set
- Also investigate the performance as function of true particle properties

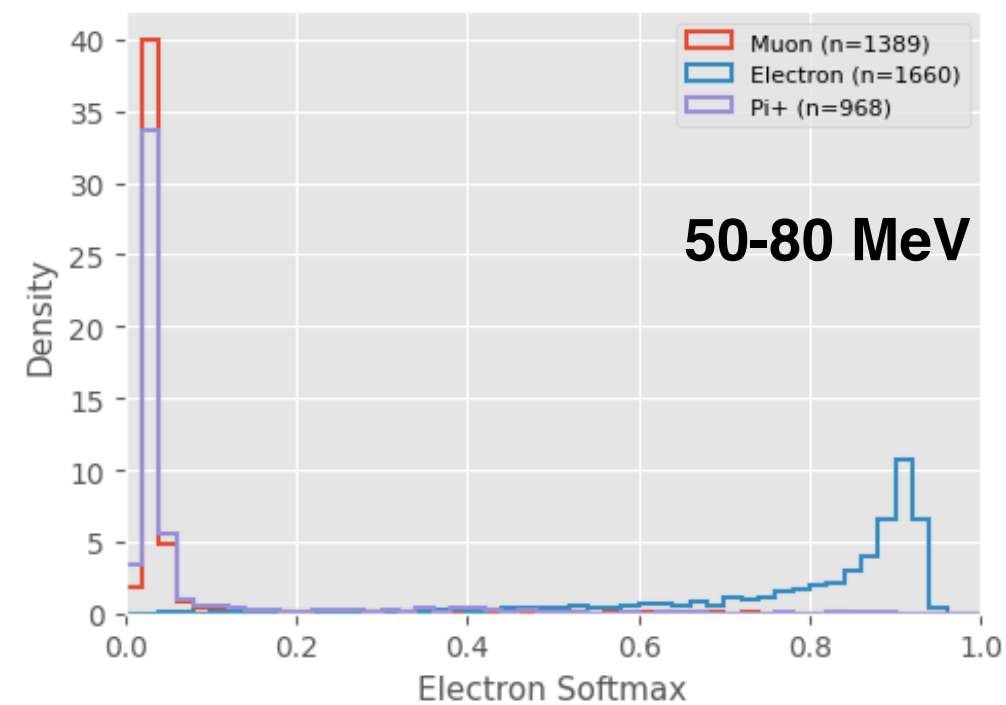
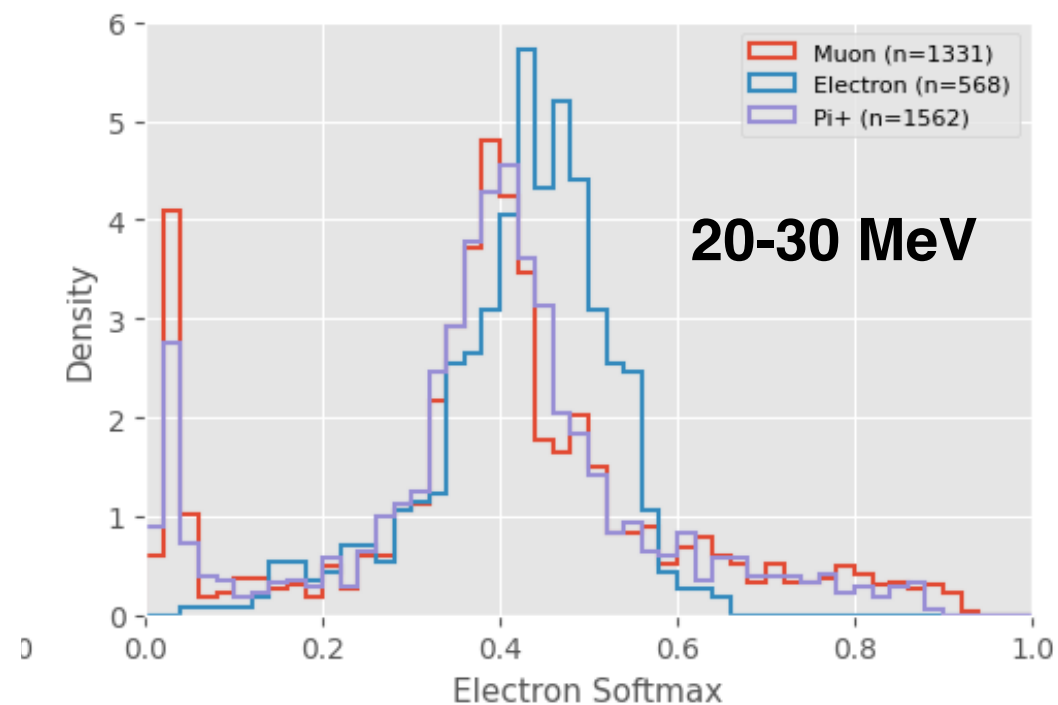
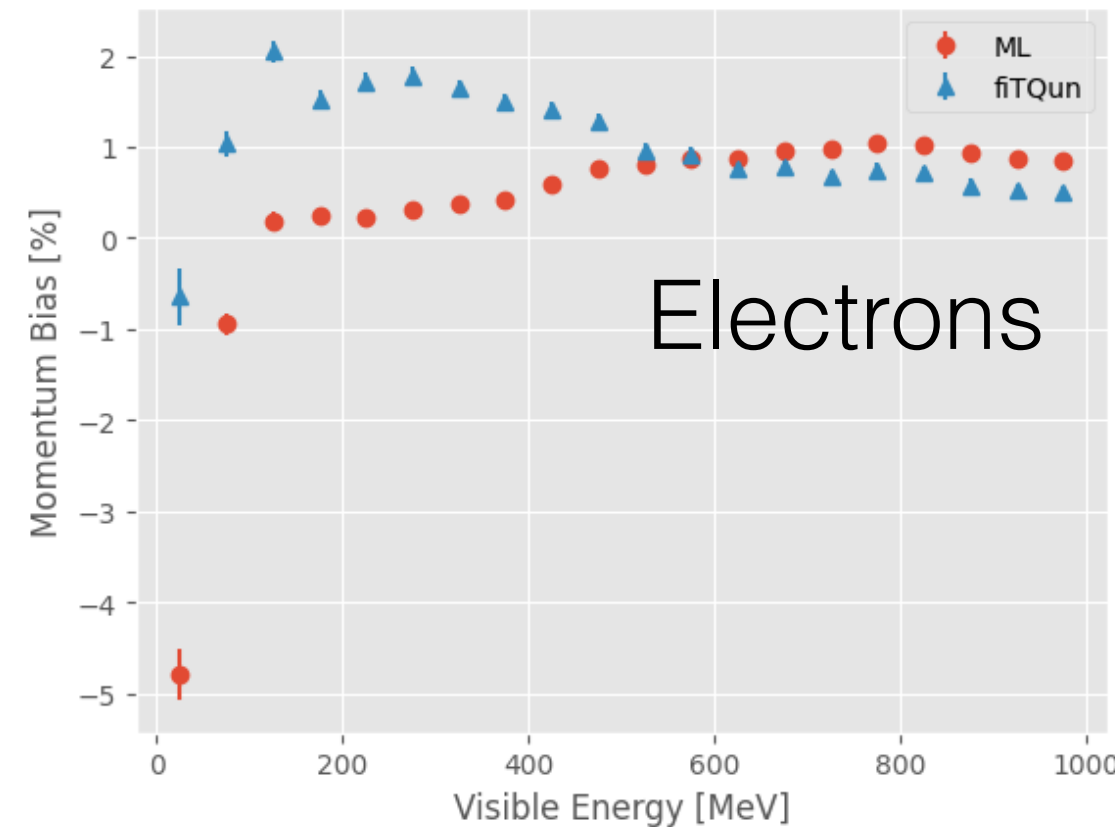
Particle Gun	ResNet	fiTQun
Momentum Resolution	1.80(2)%	2.27(3)%
Momentum Bias	0.02(1)%	0.30(1)%
3D Position Resolution	12.5(6) cm	18.4(8) cm
Direction Resolution	1.33(1)°	1.80(1)°



Challenges in Training



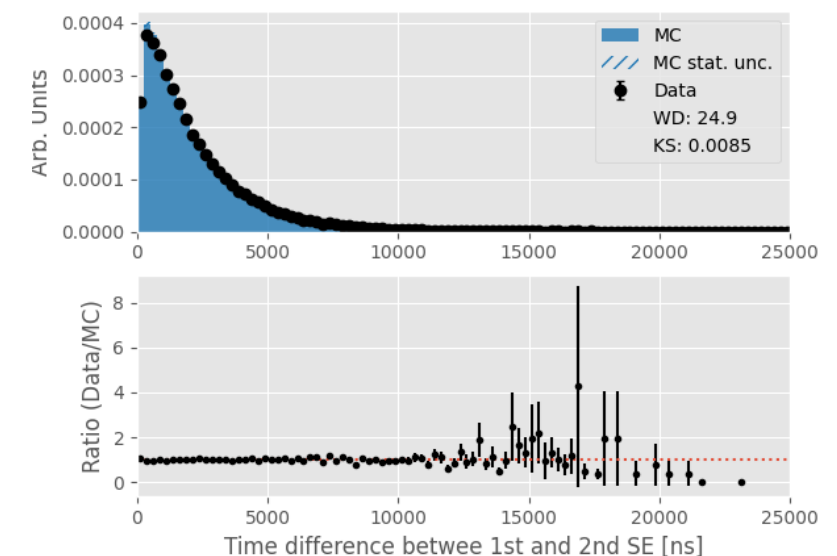
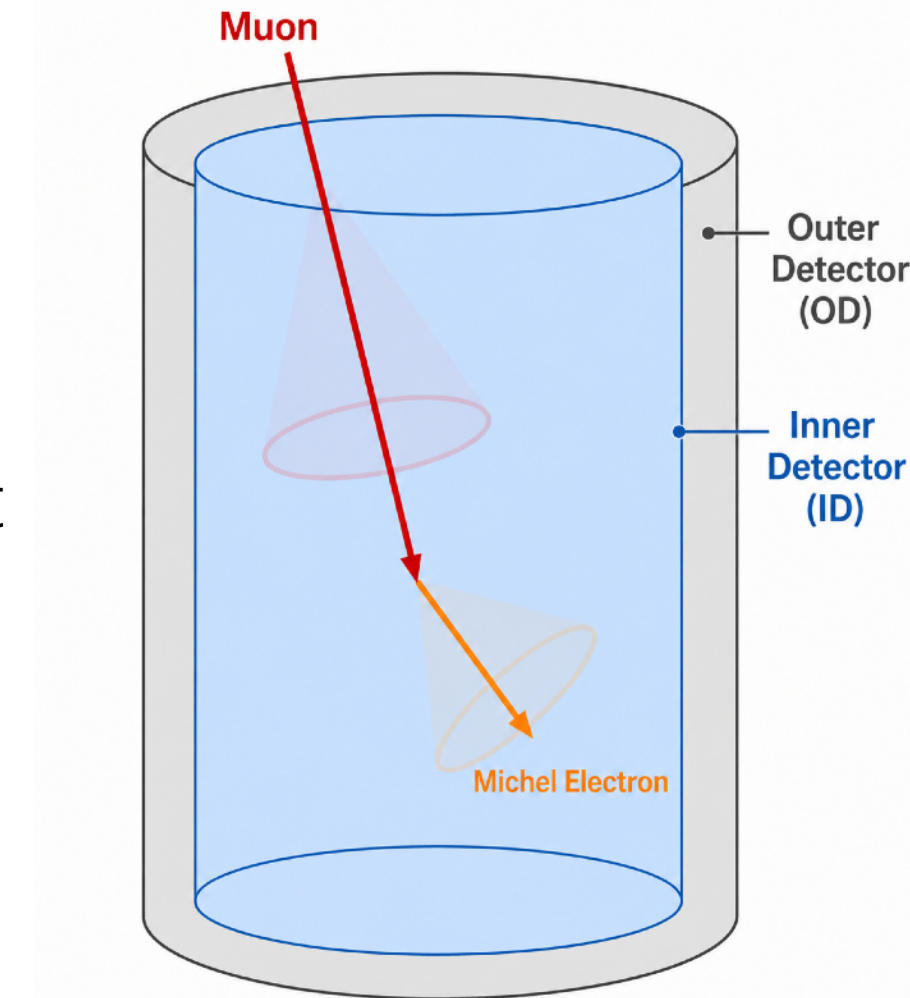
- Momentum bias in reconstruction:
 - Saw large swings in bias depending on loss function used
 - Used relative Huber loss as momentum spans many orders of magnitude
 - Led to asymmetric loss around target
- Difficulties in classification at low energy
 - Below 50 MeV, network finds it difficult to differentiate between classes
 - Motivation to develop low-E ML algorithms



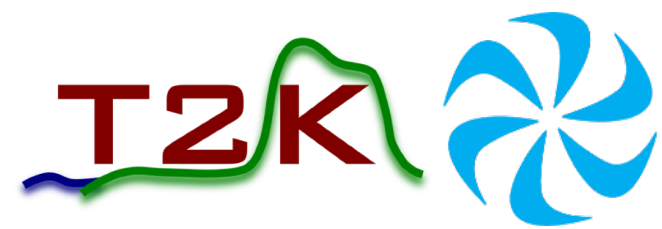
Control Sample - Stopping Muons



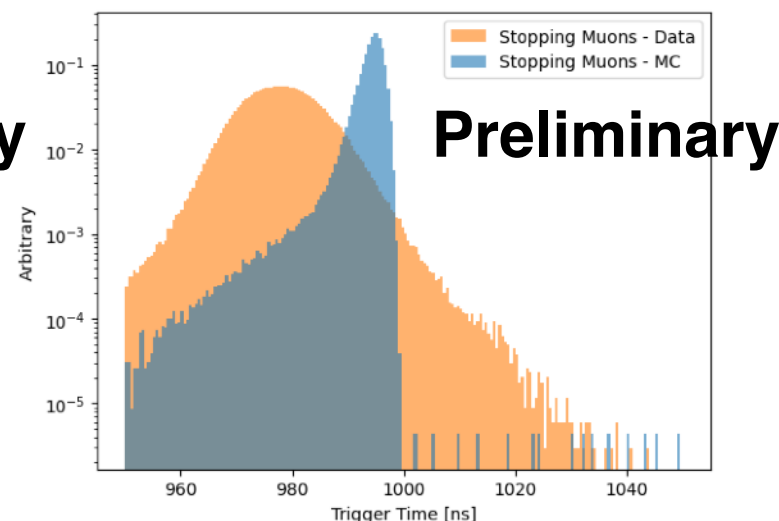
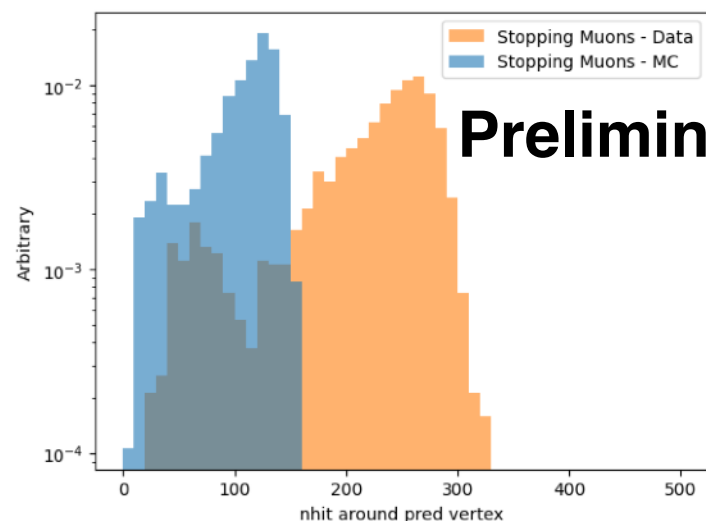
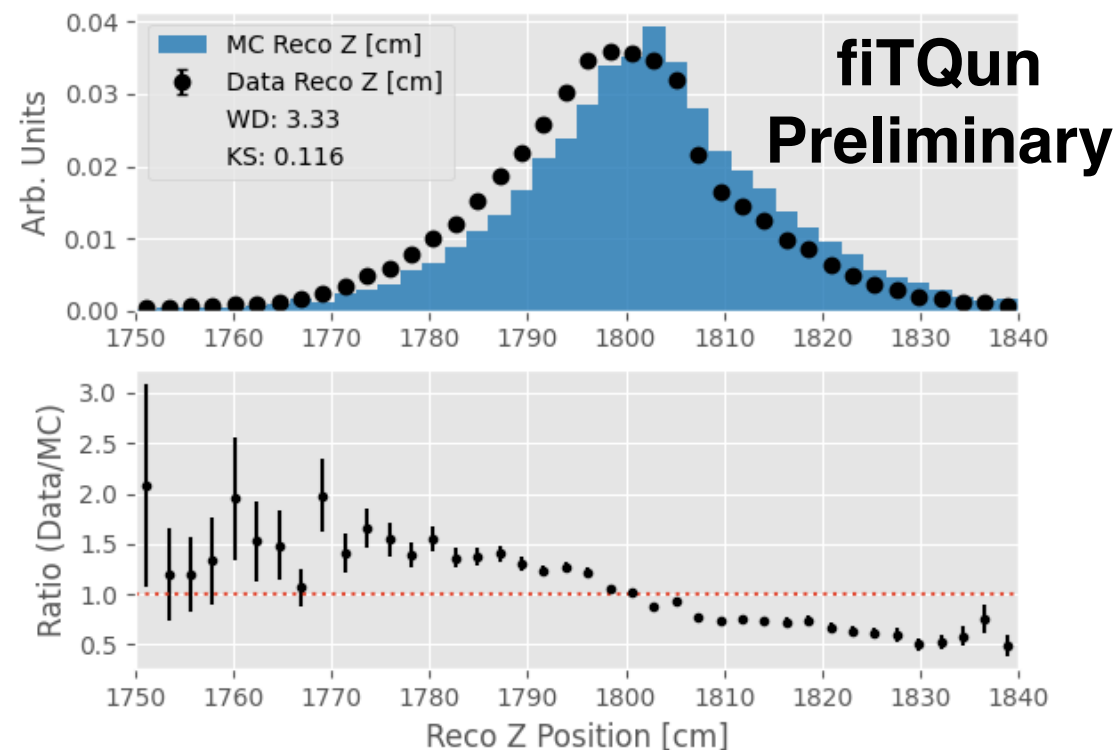
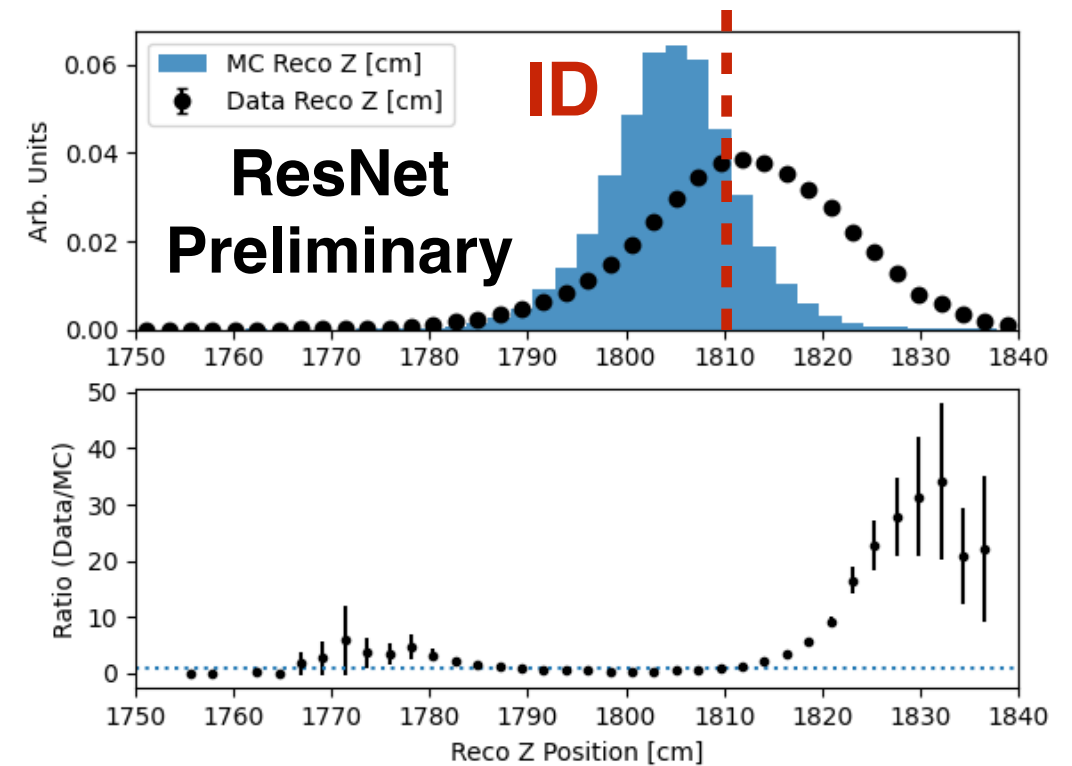
- Good performance in simulation bodes well, but need to check **real data**
 - Use control sample of stopping muons
- High energy muons in the atmosphere make it through the mountain and into SuperK
 - Deposit rest of the energy, and decay to electron
- Gather a sample of this in data, then create one in MC with same kinematic variables
- Compare results of ML output
 - Generally we don't expect perfect agreement
 - Good validation if similar results to classical reconstruction algorithm are found



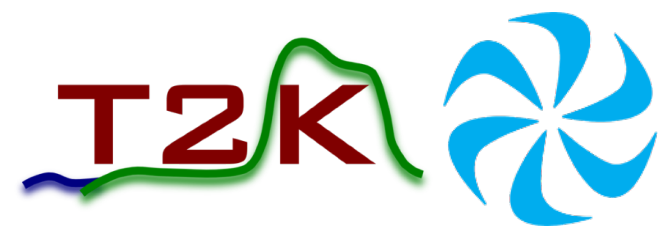
Challenge: Position Reconstruction



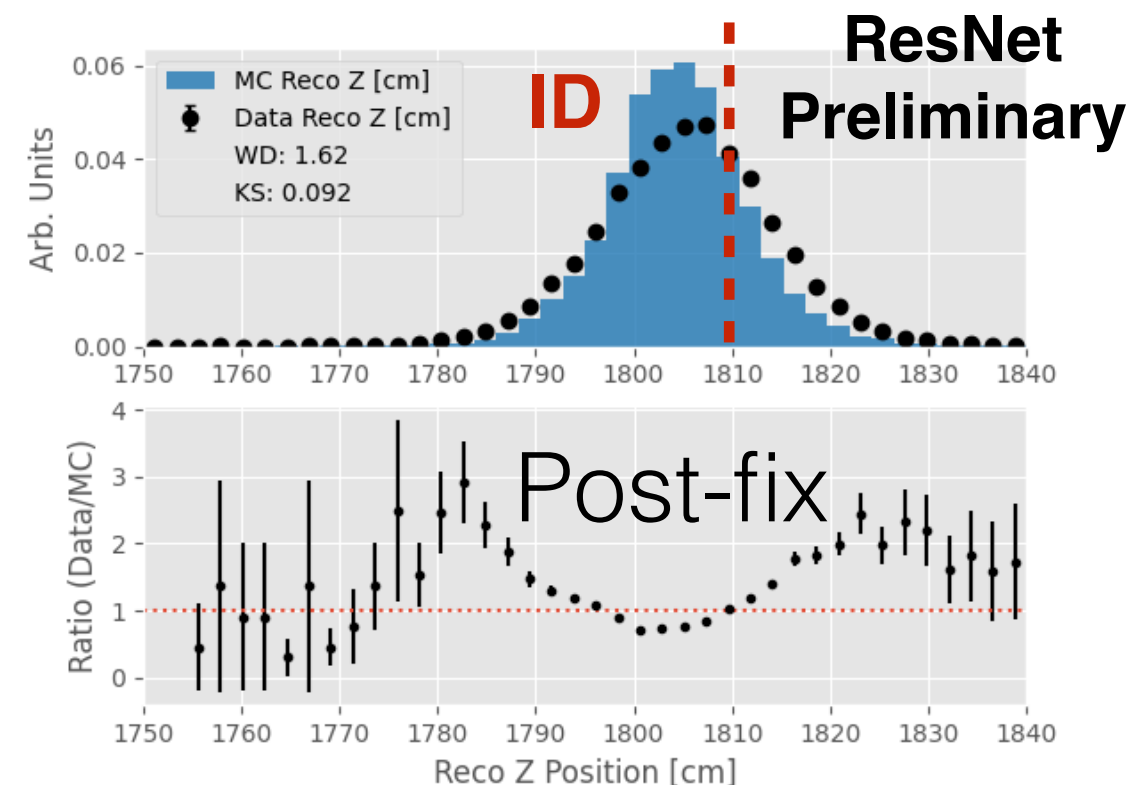
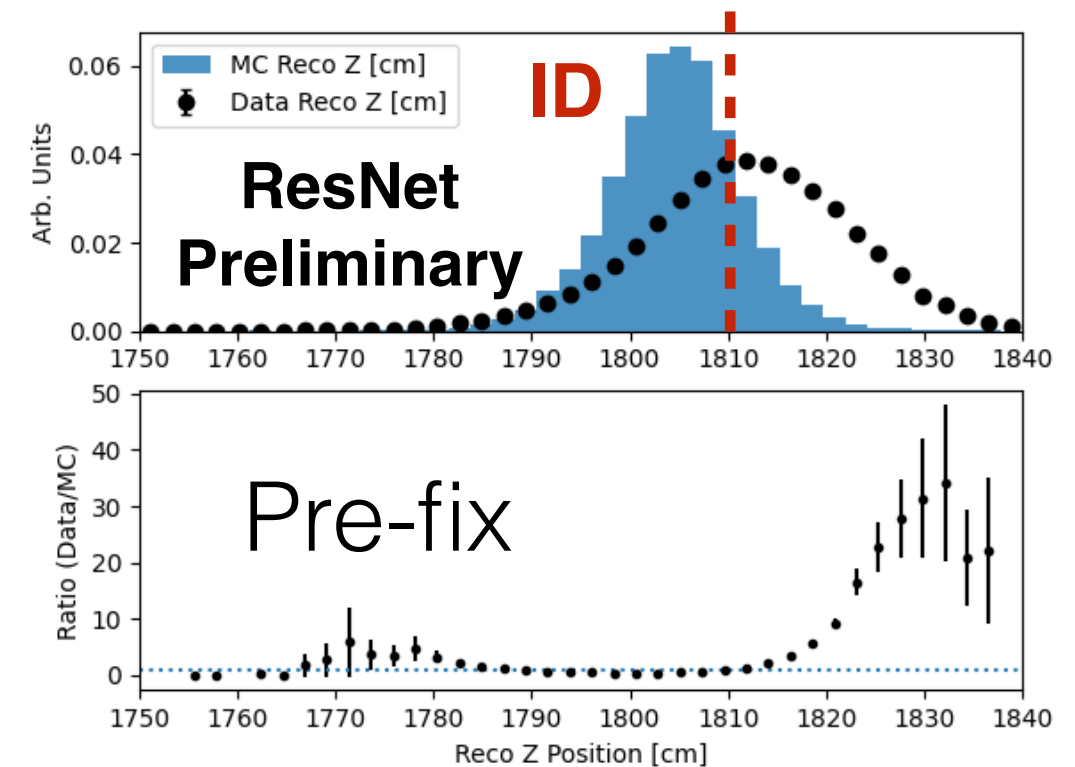
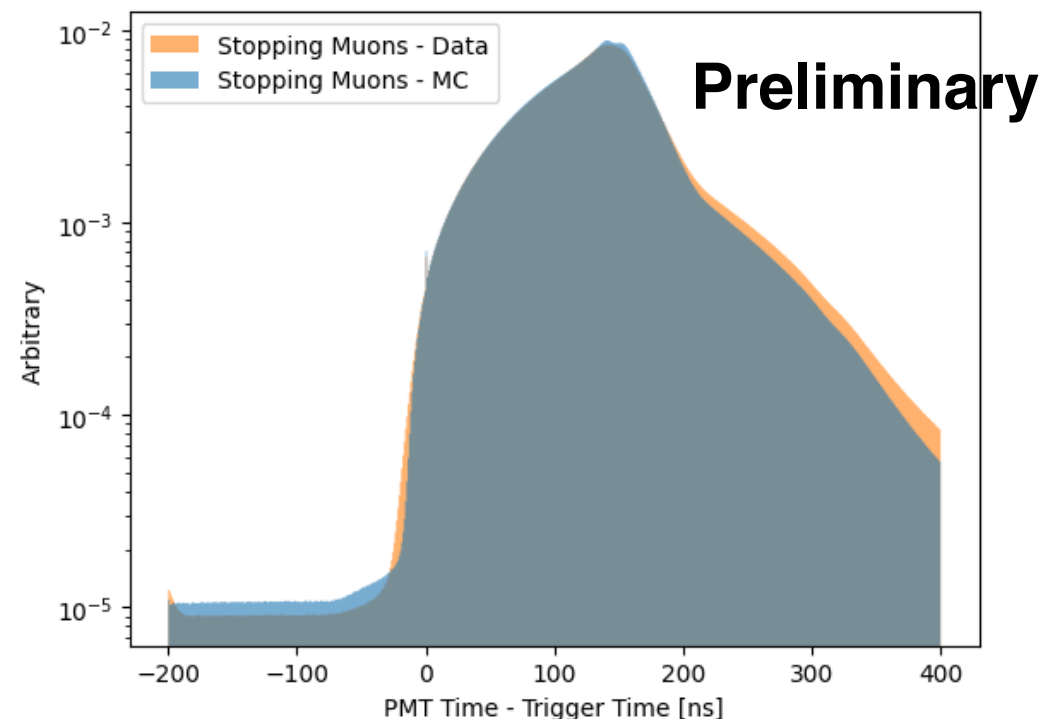
- When running reconstruction using ResNet, every variable had a good data/MC match except position
 - A large bias was found at the boundary, not found in fiTQun
- First hypothesis: light leak between outer and inner detector
 - Through back of PMT
- Second hypothesis: time-to-trigger differences between data and simulation
 - fiTQun unaffected because its time pre-process renders PMT time similar between data and MC



Results on Regression



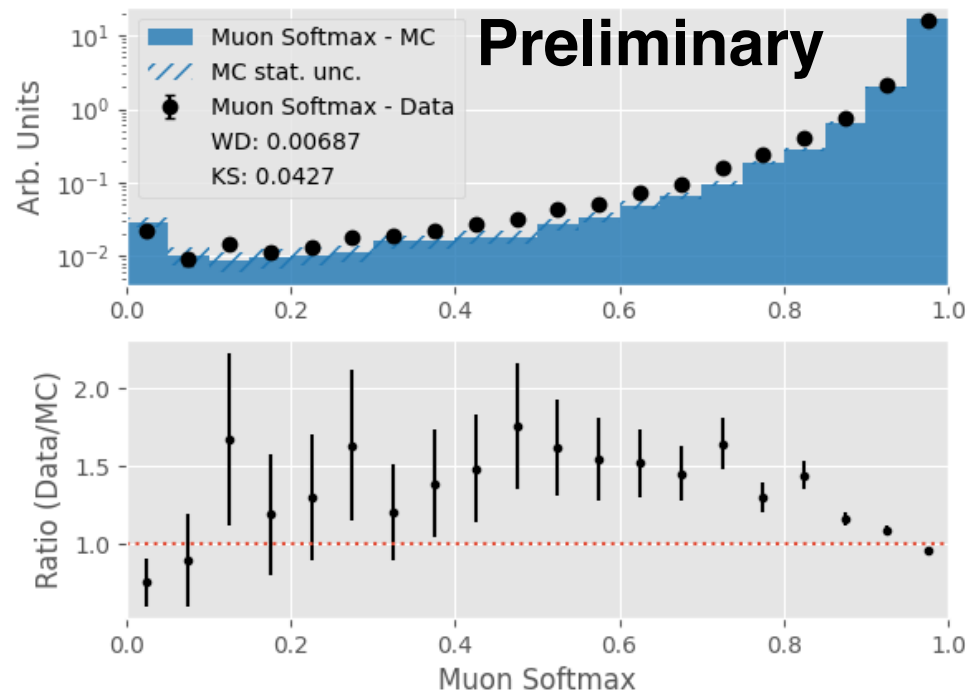
- After thorough investigating, found that effect from light leak was present but minimal
- PMT time was main culprit - small changes in time led to large change in position reconstruction
- We redefined pre-processing for PMT time in ResNet
 - Based on re-emulation of trigger
 - Outputs very similar distributions in data and MC
- Agreement now much better at the boundary



Challenge: Classification labels

ResNet

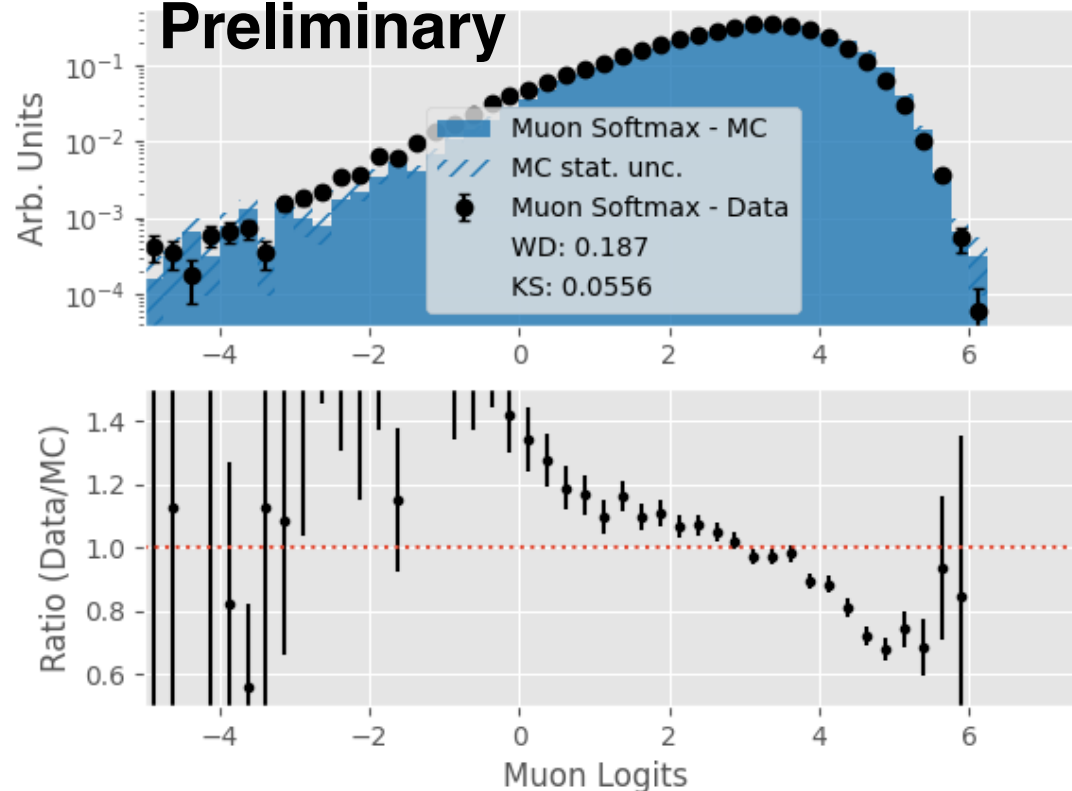
Preliminary



- Training softmax to be 0 or 1 for classification forces network to certainty at either end
- Data is inherently different than MC
 - Dead/noisy PMT. Water quality. Reflectivity. PMT efficiency.
 - All these typically push data to be less certain
- Testing label smoothing regularization to teach network not to always predict extremes

ResNet

Preliminary

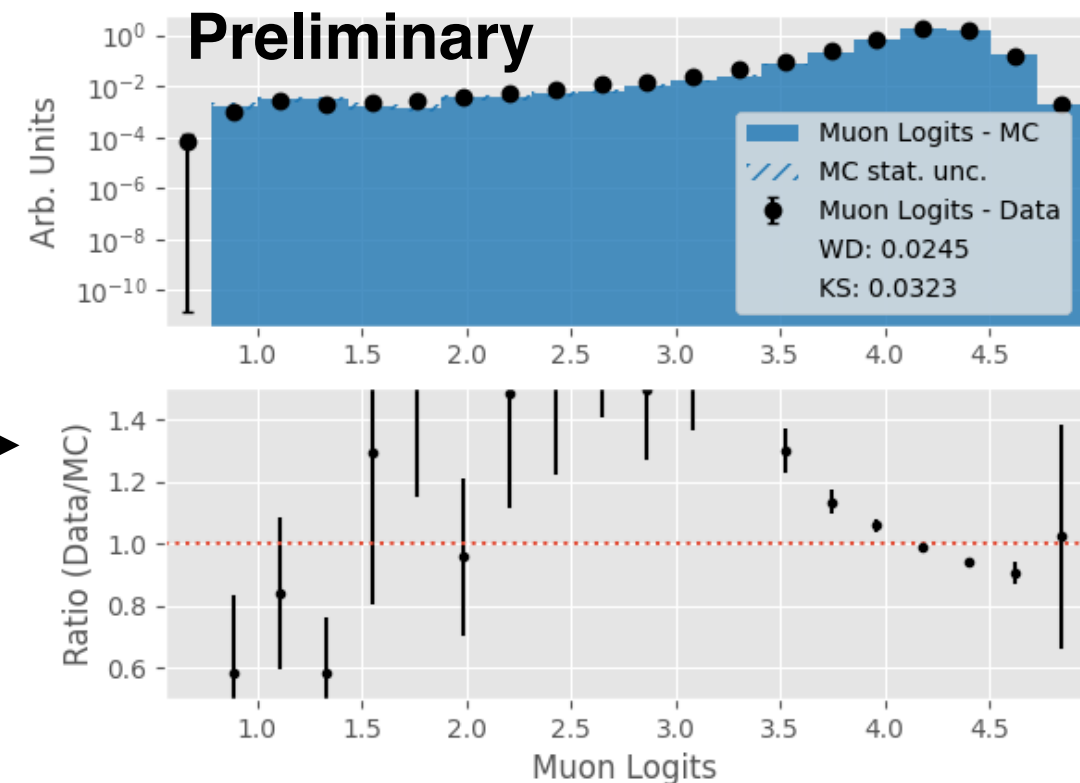


Adding
Label
Smoothing

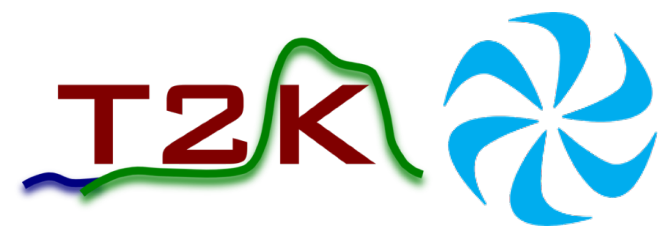


ResNet

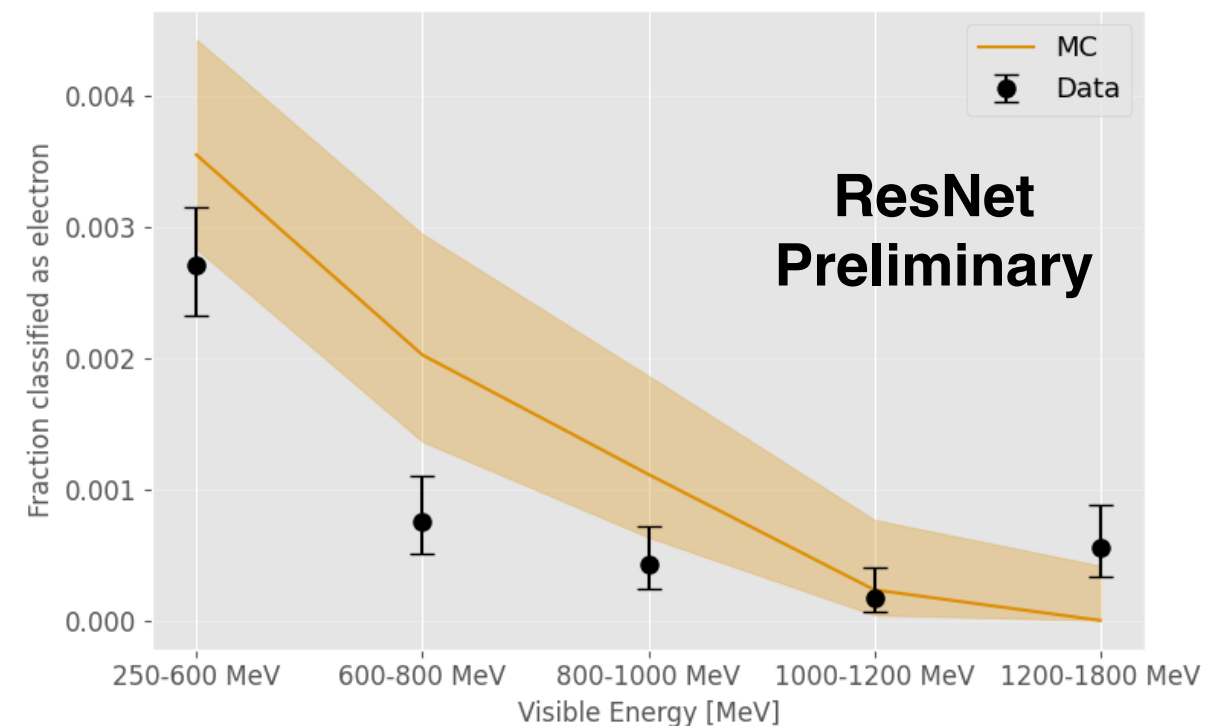
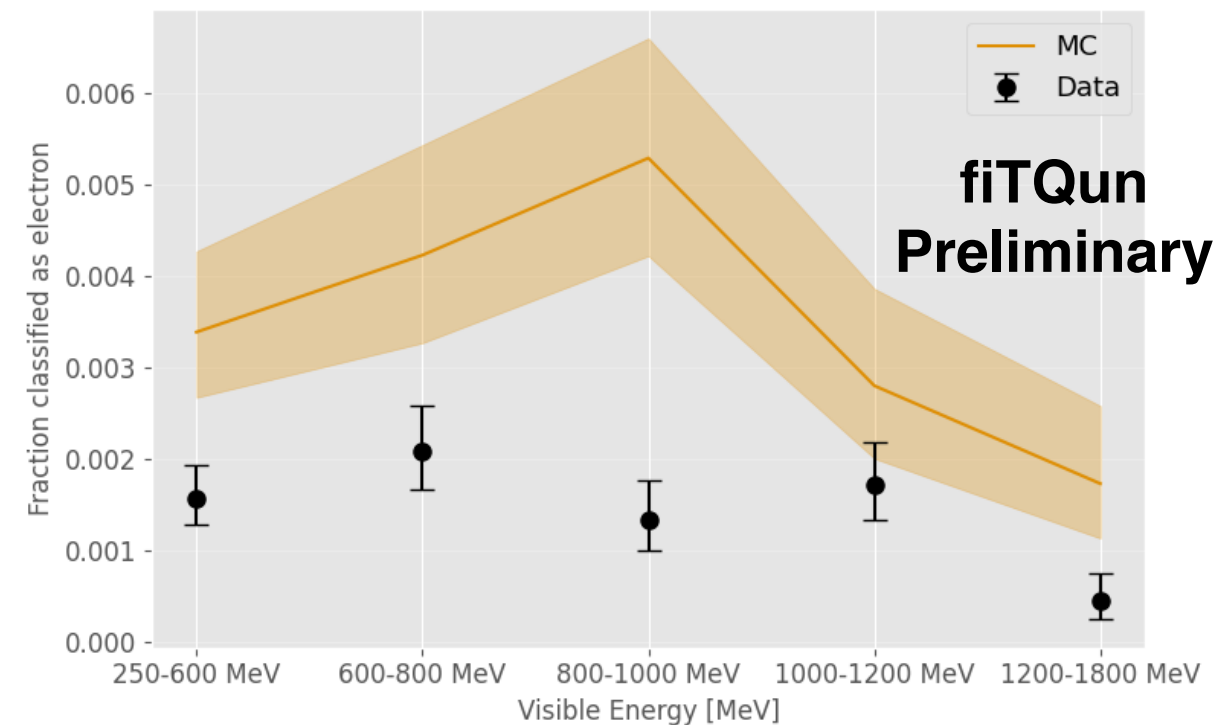
Preliminary



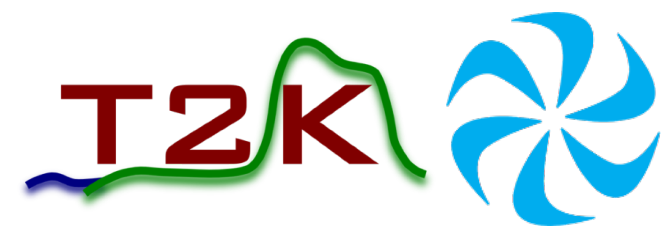
Results on Classification



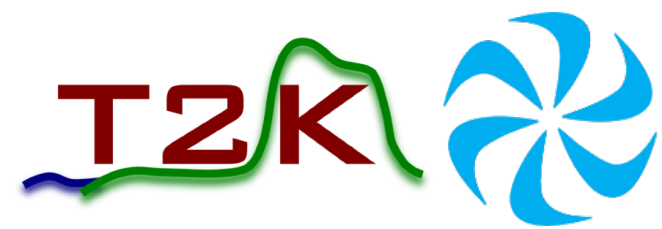
- What ultimately matters in classification is how this affects the final purity
- In stopping muon sample, assume all that pass quality selections are **muons**
- Look at all those that are misclassified as electrons
 - Use fiTQun optimized cuts from T2K
 - Use ResNet cut on ResNet output that returns similar mis-classification rate
- Generally good agreement
 - Both fiTQun and ResNet have generally lower mis-classification in data
 - Thought to be because an algorithm was used on data to choose muon-like events



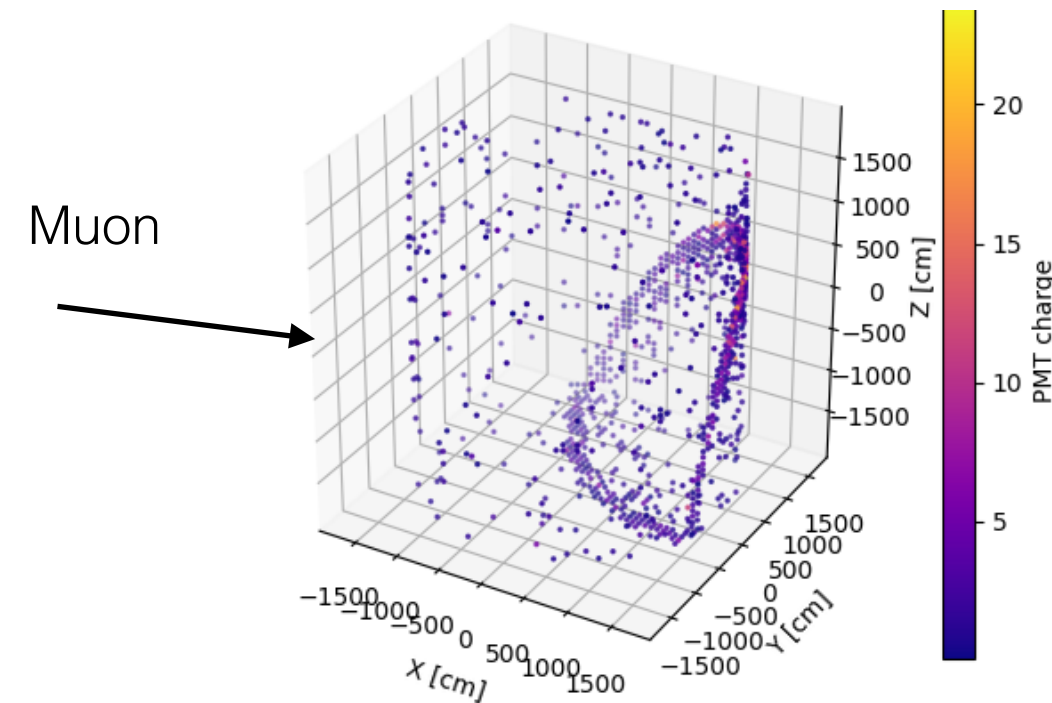
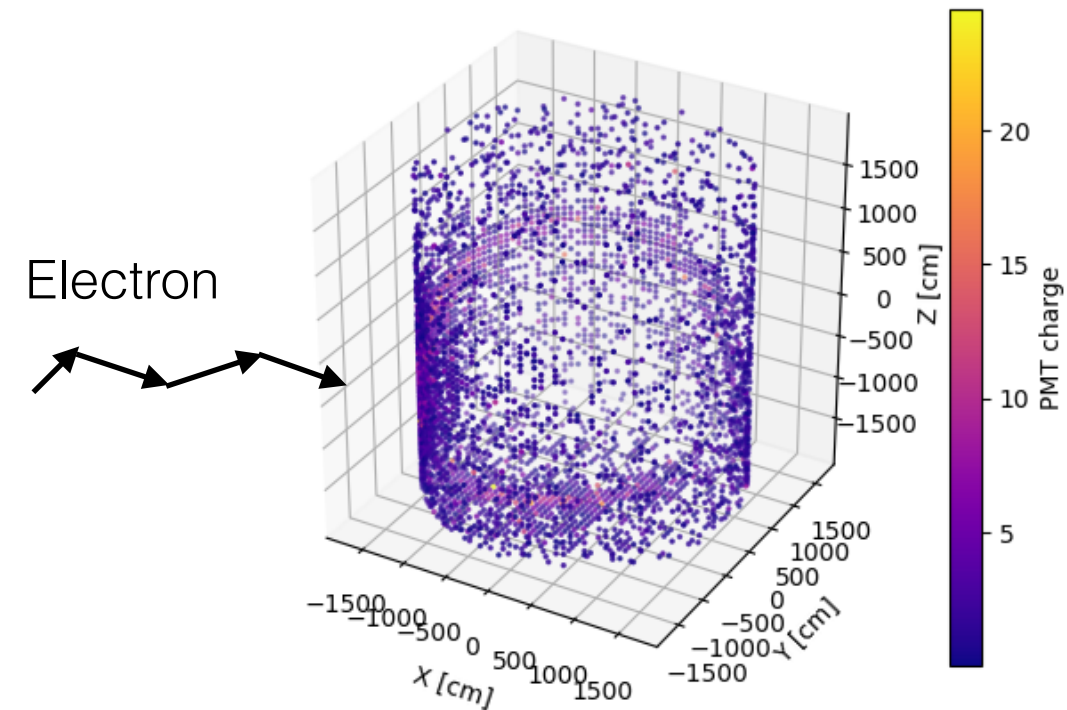
- An algorithm has been trained to do classification and regression using ResNet on T2K/SuperK events
 - Targeting muon neutrino disappearance analysis
- Performance in both regression and classification outperforms current algorithm significantly
 - Time speedup compared to classical algorithm: **500x**
- Some challenges noticed in momentum bias and classification at low energy
- Performance in stopping muon control sample generally well understood
 - Additional studies done on dead PMT effects on momentum reconstruction



Water Cherenkov data



- Classification networks can reconstruct particle PID
 - Electrons will produce **larger rings** due to multiple interactions and showering
 - Muons will often have **thinner rings**
 - Charged pions will have **muon-like rings**, but hadronic interactions can lead to distinct signatures
 - Exactly what is being learned is under investigation
- Regression networks can reconstruct vertex position, direction, particle momentum
- We generate data using official SuperK simulation software, SKDETSIM (v13p90)
 - Generate samples of electrons, muons and π^+
 - **Energy is uniformly sampled** from 0 to 2 GeV above Cherenkov threshold
 - Uniform in vertex position & direction
 - Keep all PMT hits in first timing window (NQISK)
 - Quality cuts added in training & testing
 - ~5M of each particle type for training/testing



- Studies are ongoing to understand the effect of dead PMTs on performance
 - SKDETSIM simulation will have all PMTs work, while real data will have 0-~2% of PMTs dead depending on time since replacement
 - Effects on classification seem minimal until we hit 10% dead PMTs
 - Effects on momentum bias are higher, with a $>1\%$ bias as we hit 2% dead PMTs
- Current solution is to incorporate SK4 dead PMTs into training

