

June 15th, 2026

Improving Detector Systematic Uncertainties Through Data-Driven Machine Learning

Harry Hausner
NPML 2026 — UC Irvine



U.S. DEPARTMENT
of **ENERGY**

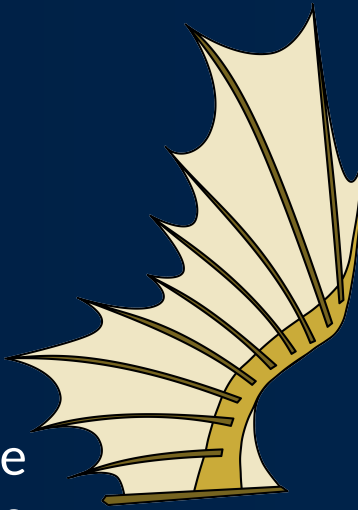
Fermi National Accelerator Laboratory is managed by
FermiForward for the U.S. Department of Energy Office of Science



ICARUS

The Detector

- 760 tons (476 ton active mass)
- 360 photomultiplier tubes
- Cosmic Ray Tagger system surrounds the cryostats to flag cosmogenic interactions
- On the surface with 3m concrete overburden
- Two cryostats, each with two LArTPCs
 - 3 Wire Planes per LArTPC
 - Wires $0^\circ, \pm 60^\circ$ from horizontal
 - 500 V/cm drift electric field
 - Shared cathode between TPC modules
- Far detector for the SBN program



I
C
A
R
U
S

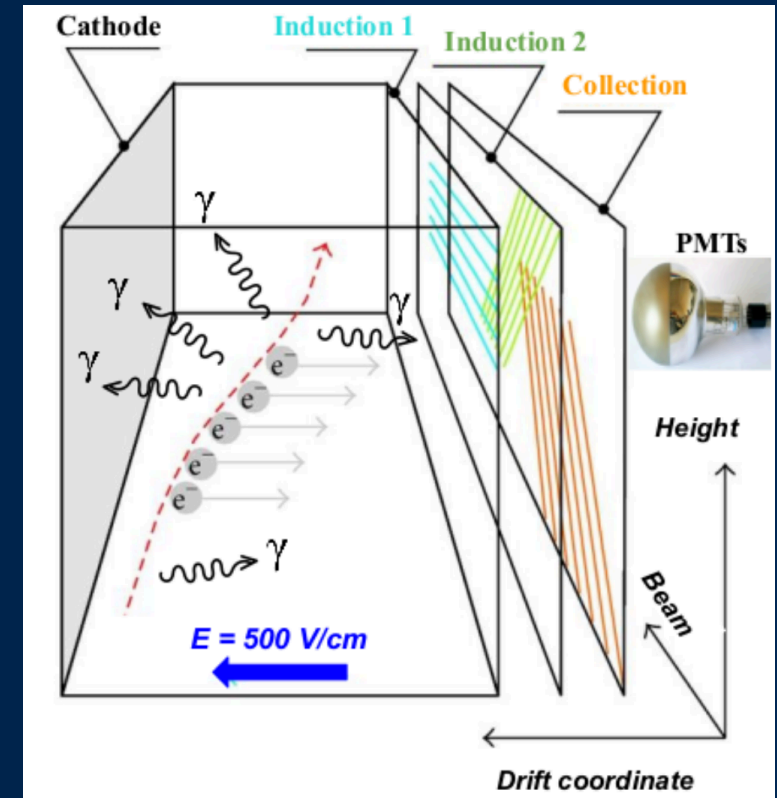




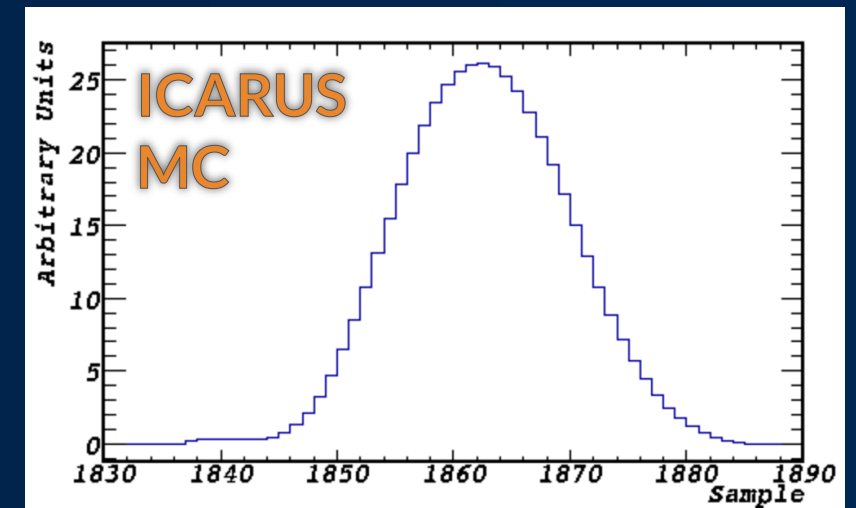
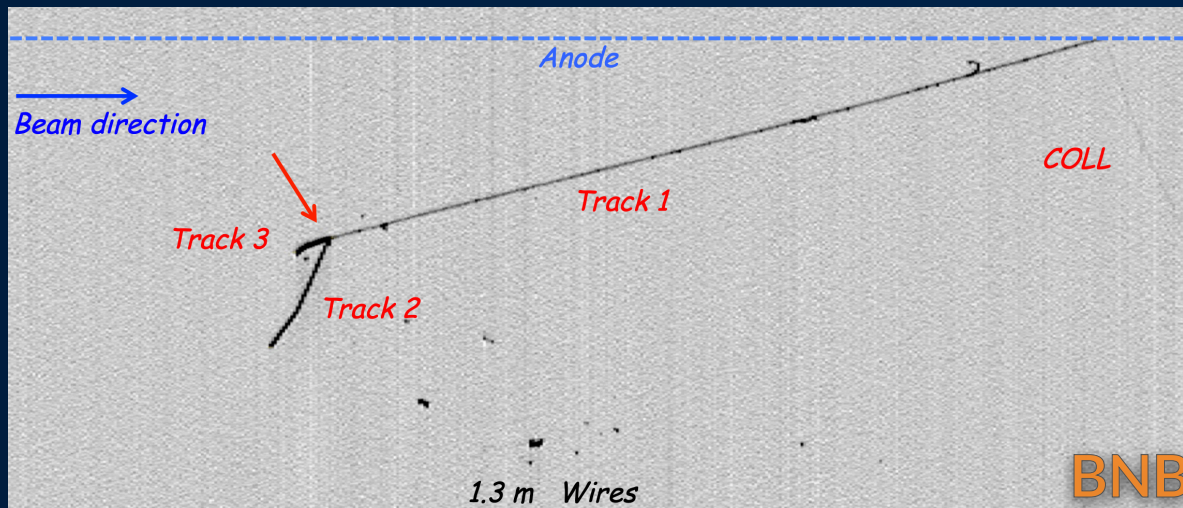
Wire Based TPC

How a signal is made

- In a wire based LArTPC charged activity ionizes the argon and the ionization is drifted across a wire plane, inducing a current on the wires
- The current on the wire is processed into a gaussian-like measure of energy deposited call a “hit”
- These create beautiful 3D reconstructions, but understanding how the drifted ionization produces hits is highly non-trivial



doi:10.1140/epjc/s10052-023-11610-y





Detector Uncertainties

Why do we need to care?

- Detector modeling is a large source of systematic uncertainty in LArTPCs
- For example, in the recent ICARUS ν_μ disappearance search the detector systematic uncertainty was the one of the top 3 sources of uncertainty along with flux and interaction systematics
- For the joint fit with SBND as part of the SBN program we expect detector systematics to be dominant and need to reduce their contribution to below 3% to match our design goals
- For this we need better techniques for modeling the detector impacts

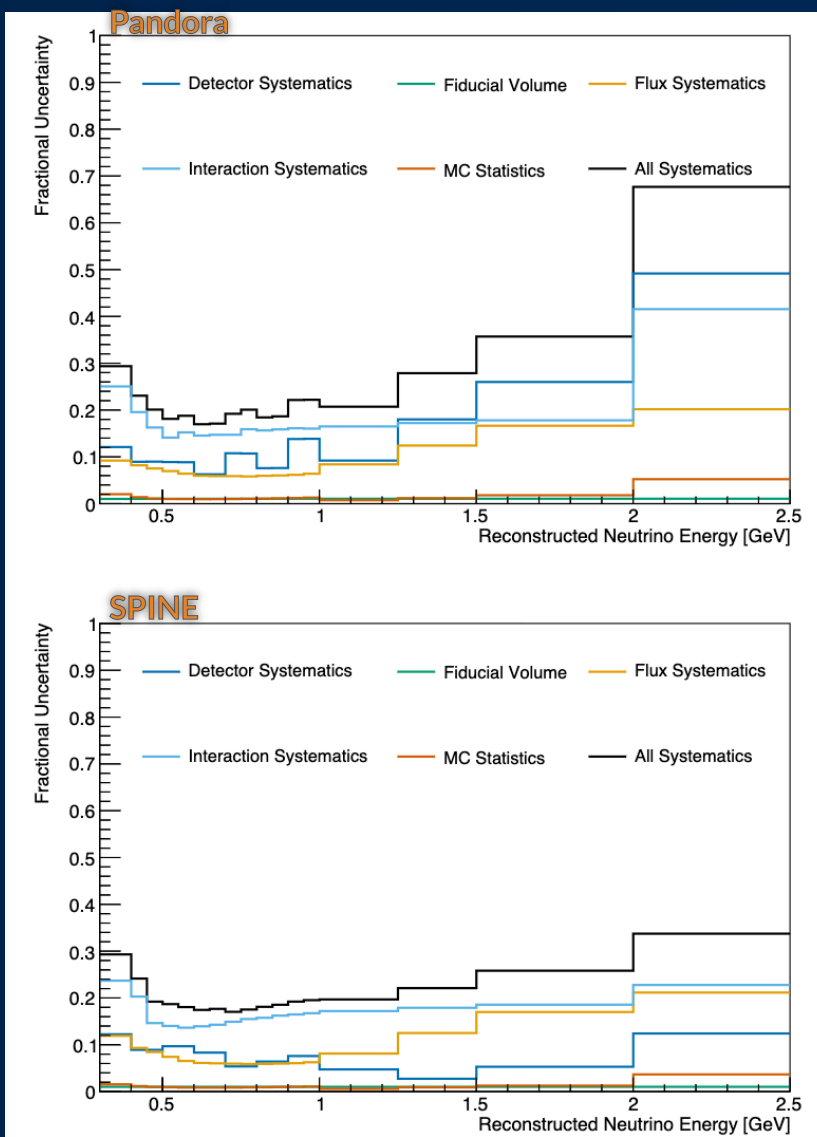


FIG. 21: Fractional size of systematic uncertainty as a function of reconstructed neutrino energy for Pandora (top) and SPINE (bottom).

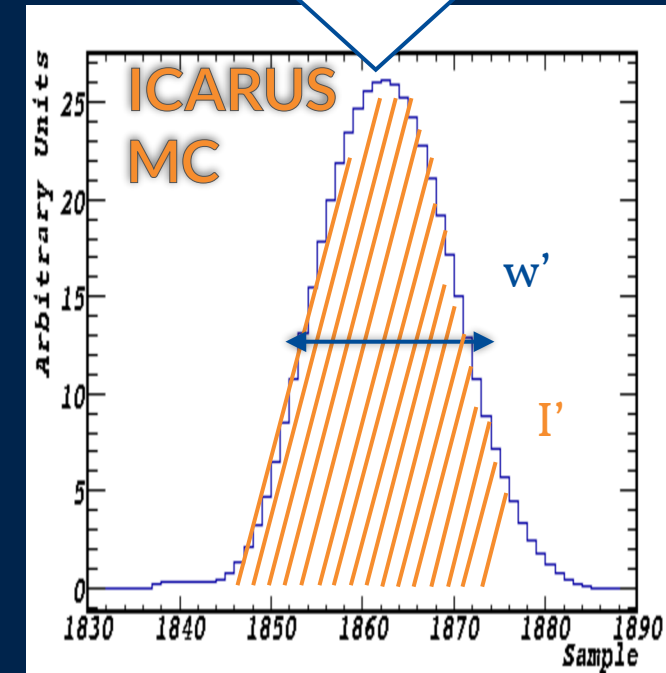
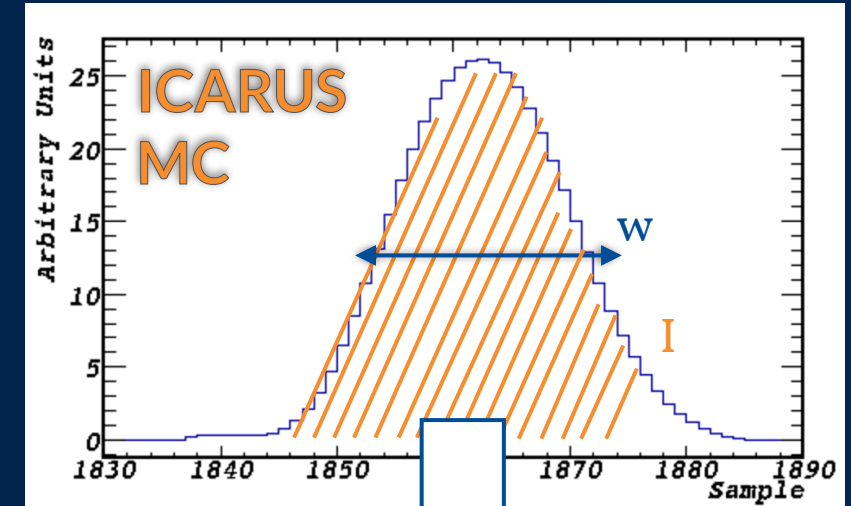
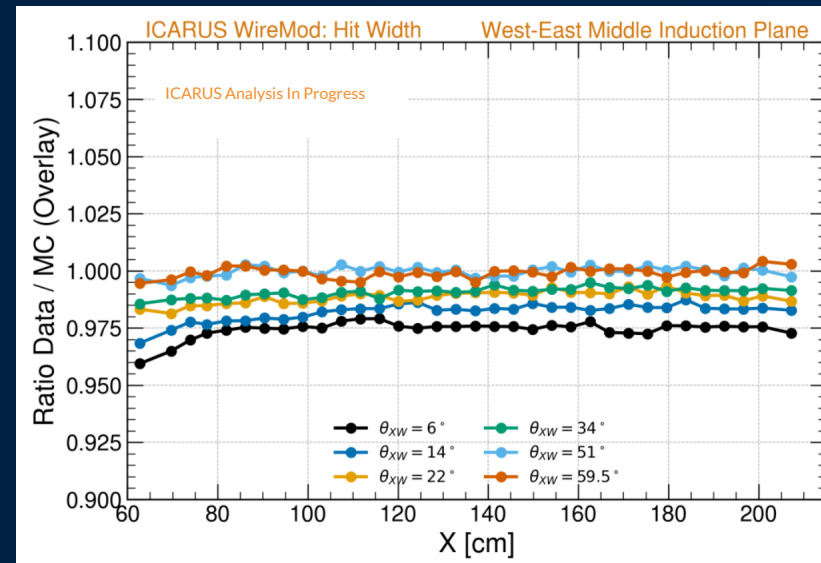
arXiv:2603.22557v1



Assessing Uncertainties with Hit Shapes

WireMod

- Most factors in the detector modeling express themselves in uncertainty on the shape of the hits
 - Recombination, diffusion, etc
- One non-ML, catch-all systematic treatment for these uncertainties from MicroBooNE is called **WireMod**
 - Treat the different TPC systematics as a shaping effect on the hits
- Compare data and MC hits to get ratios of average hit integrals and widths binned in different hit quantities
- Scale hits by these ratios

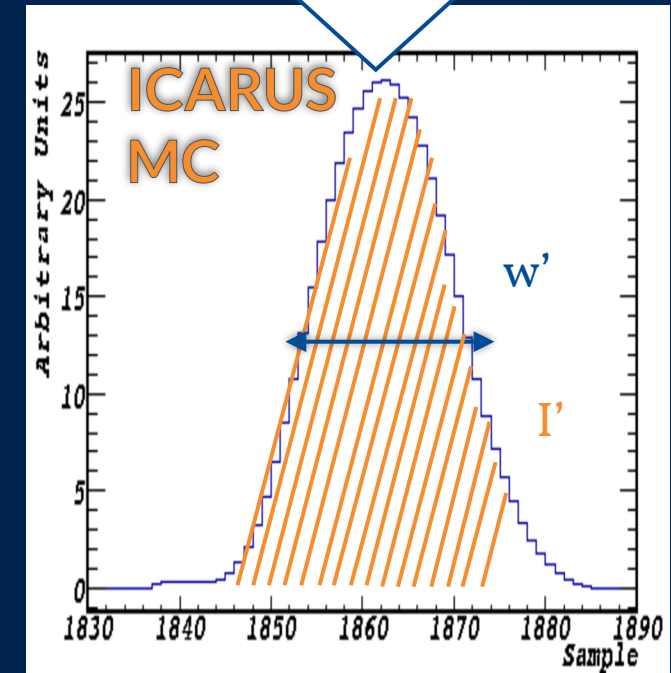
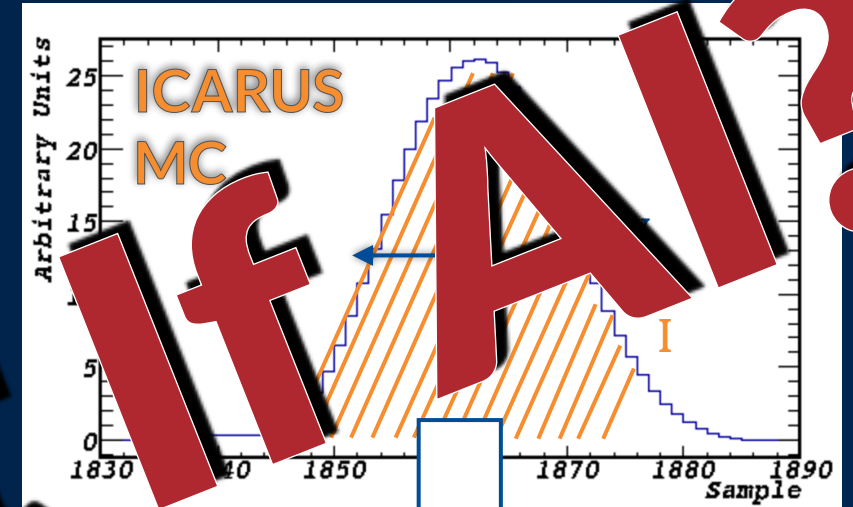
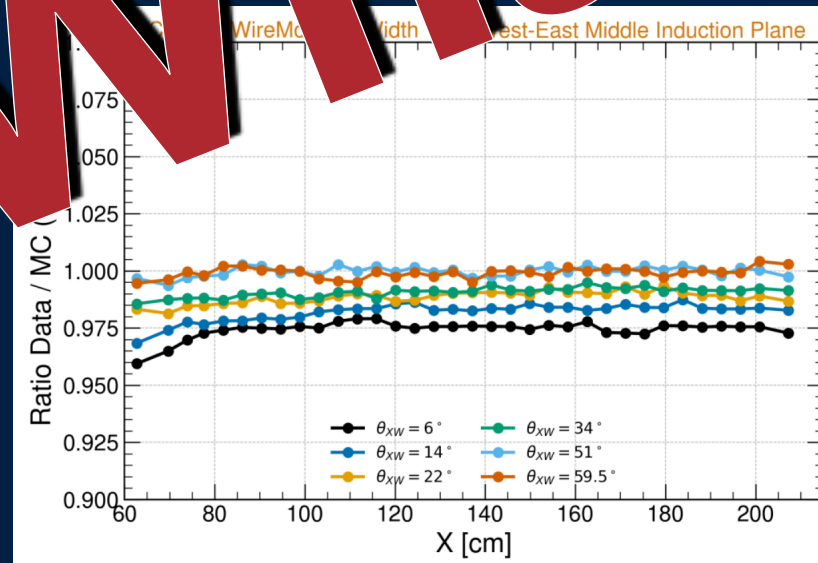




Assessing Uncertainties with Hit Shapes

WireMod

- Most factors in the detector modeling express themselves in uncertainty on the shape of the hits
 - Recombination, diffusion, etc
- One non-ML, catch-all systematic treatment for these uncertainties from MicroBooNE is called **WireMod**
 - Treat the different TPC systematics as a shaping effect on the hits
- Compare data and MC hits to get ratios of average hit integrals and widths, binned in different quantities
 - Shaped by these ratios

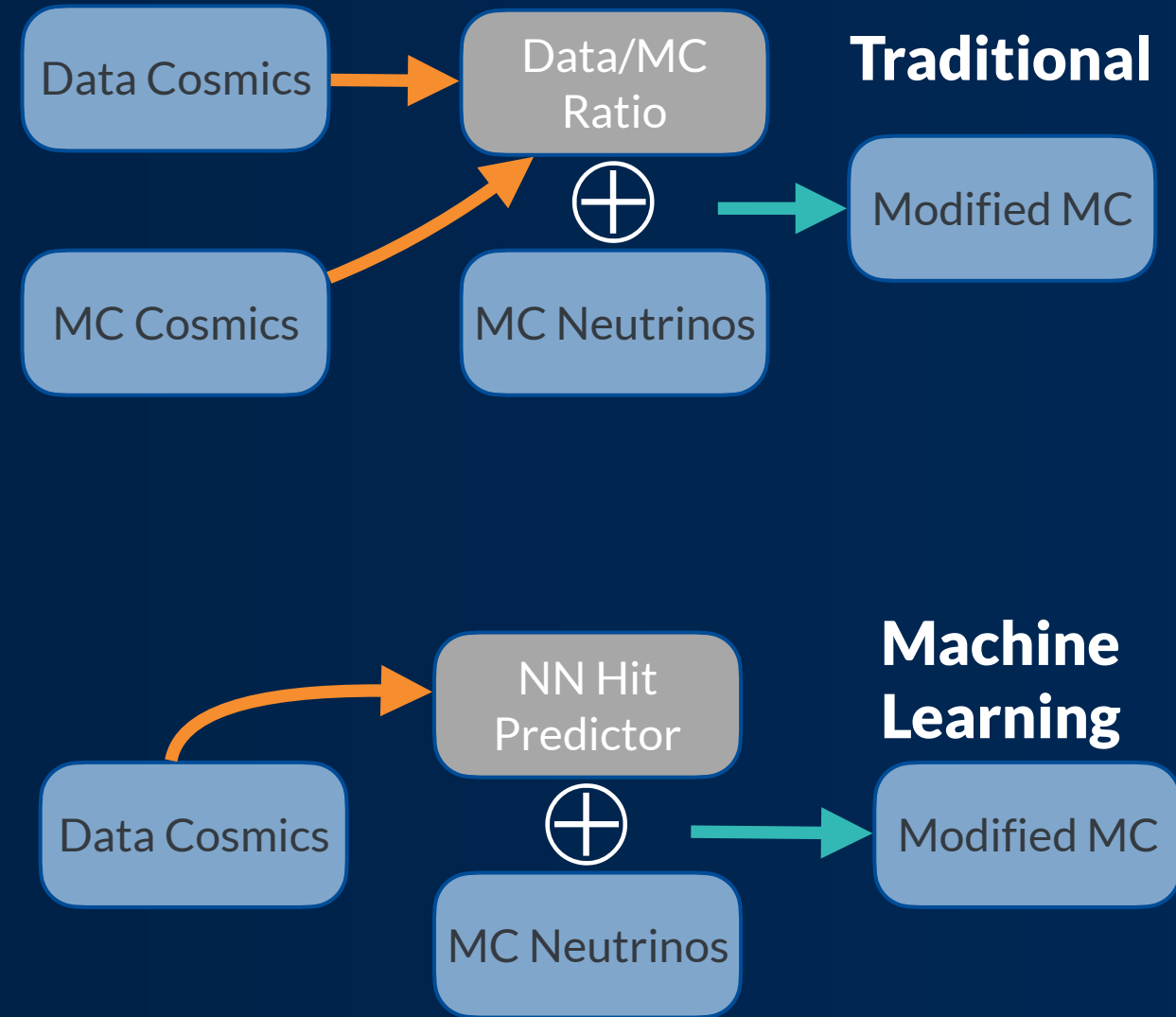




WireMod? No, WAireMLod!

There are actual good reasons...

- The classic WireMod approach requires simulating many cosmic muons, which is expensive
- WireMod also relies on shifting by averages and does not capture effects in the tails of the distributions
- If we use a neural network we can also map hit features to a new integral/width, but with non-linear effects
- Training the NN on data side-steps extra simulation
 - Cosmic data is cheap and plentiful!

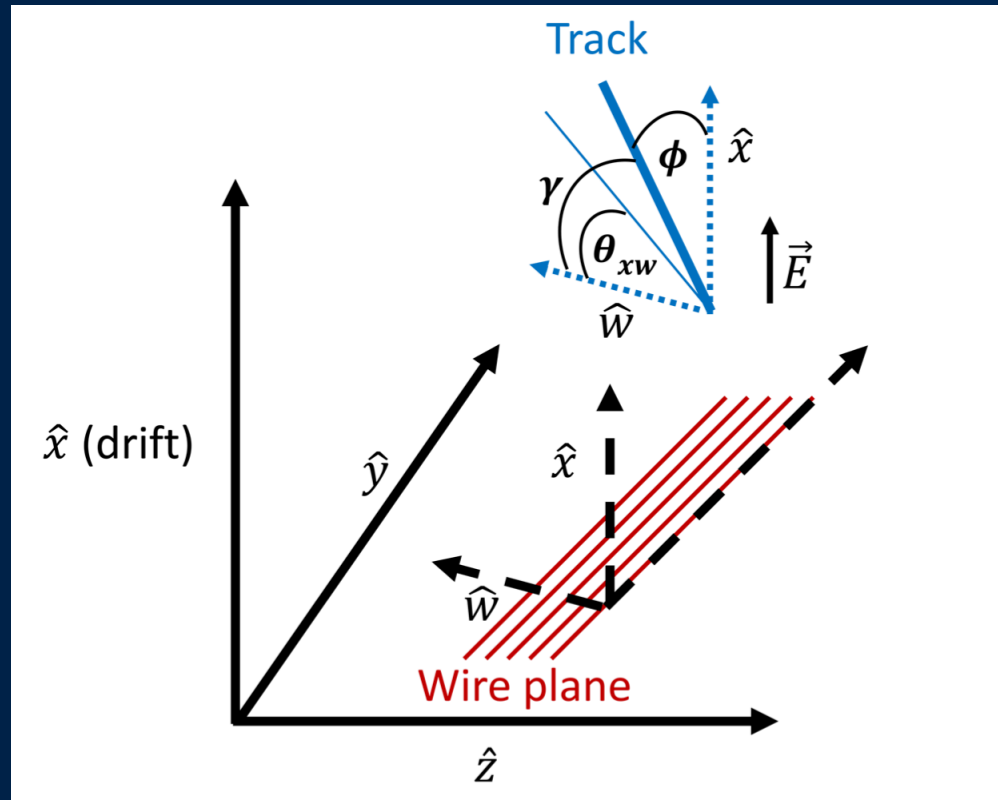




The Model

Features & Targets

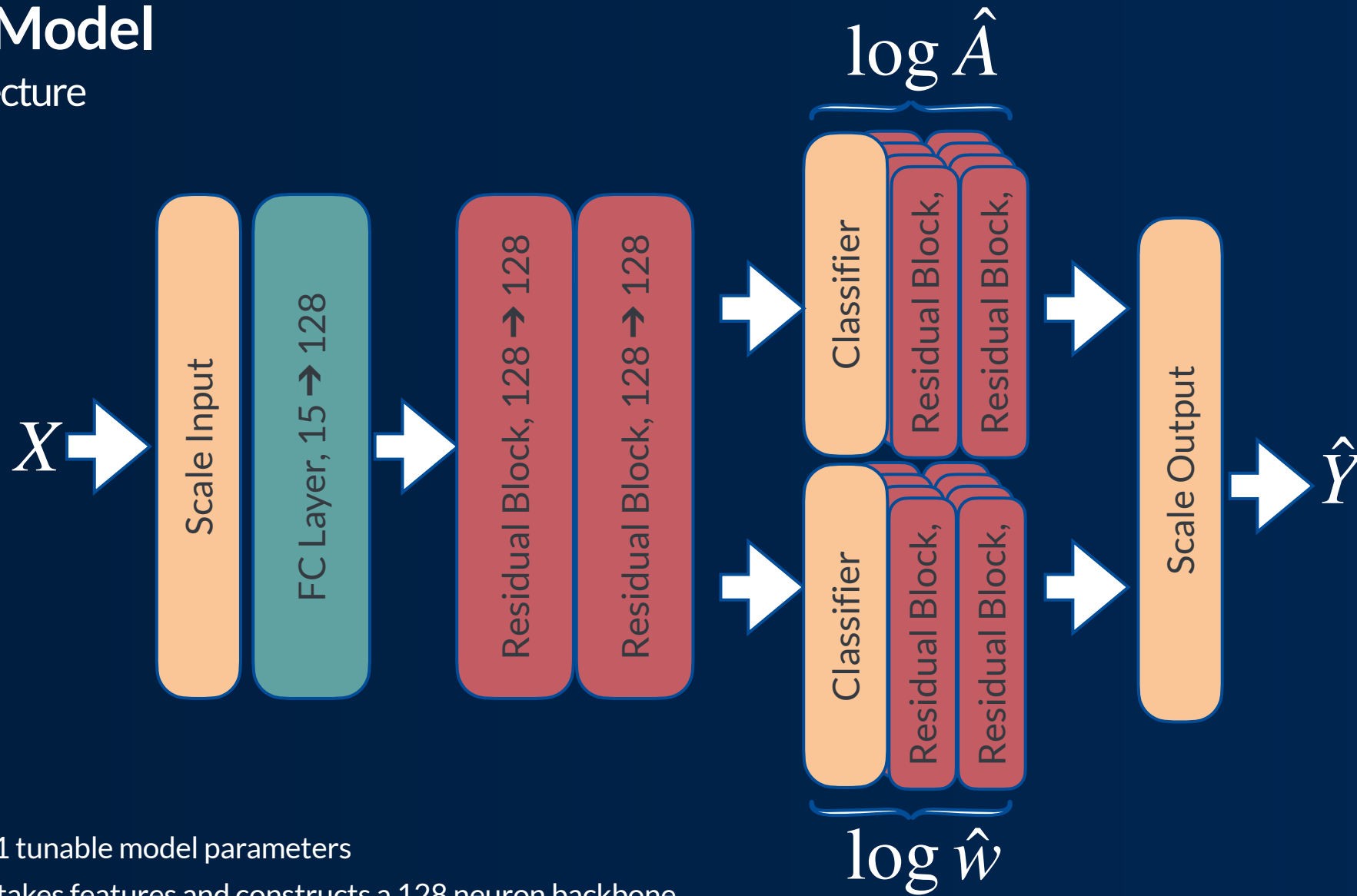
- 15 Input features per hit
 - Position (x, y, z)
 - Particle direction $(\hat{v}_x, \hat{v}_y, \hat{v}_z)$ and direction relative to the target wire $(\hat{v}'_y, \hat{v}'_z)$
 - Angle of the particle relative to the wire $(\sin \theta_{xw}, \cos \theta_{xw})$ and the angle of the wires relative to the horizontal $(\sin \theta_{\text{wire}}, \cos \theta_{\text{wire}})$
 - Readout channel number and which TPC module (c, TPC)
 - Charge deposited per cm $\frac{dQ}{dx}$
- Target the integral and width of the gaussian fit to the hit (I, w)





The Model

Architecture

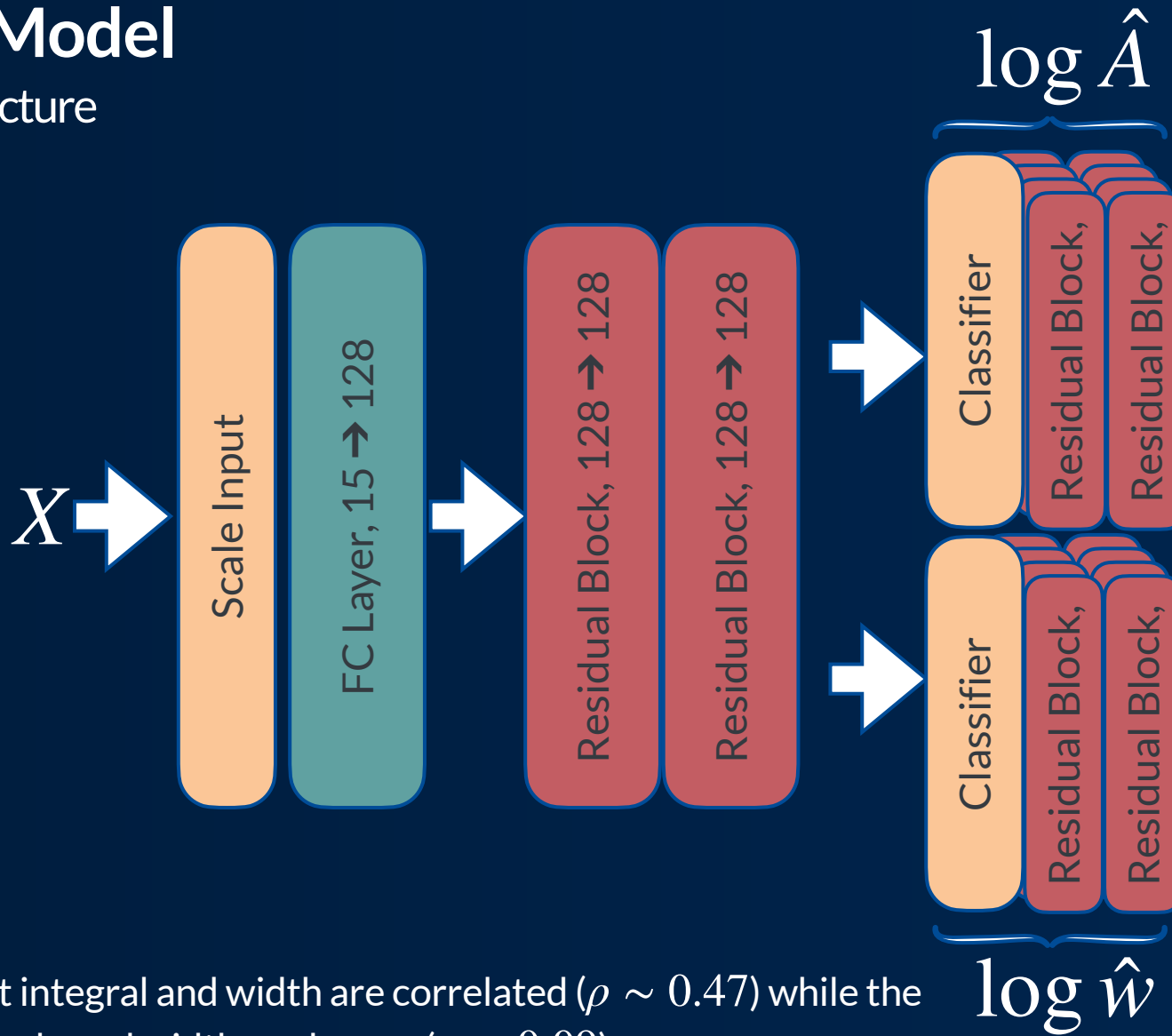


- 440851 tunable model parameters
- Model takes features and constructs a 128 neuron backbone
- Feed into two classifiers, which then each use a mixture of 4 experts to predict $\log \hat{A}$ and $\log \hat{w}$
- Combine into outputs with $\hat{I} = \sqrt{2\pi\hat{A}\hat{w}}$

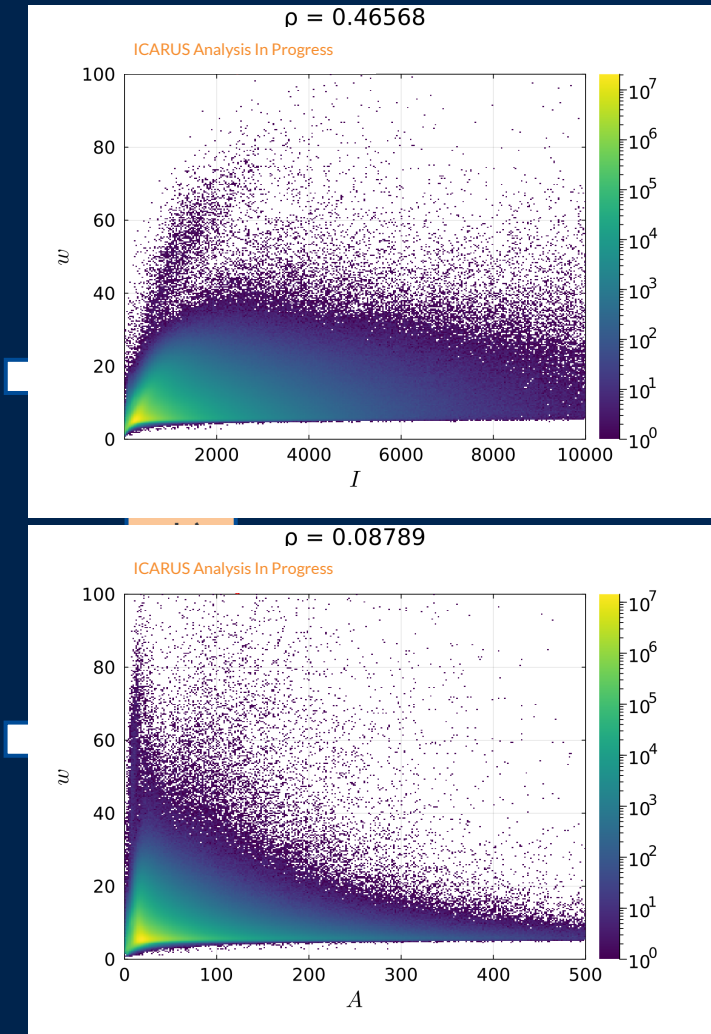


The Model

Architecture



- The hit integral and width are correlated ($\rho \sim 0.47$) while the amplitude and width are less so ($\rho \sim 0.09$)
- Training with separate attention heads for A and w should hopefully let the model better learn these features





Training

- Train using 1 billion hits from ICARUS off-beam cosmic muons
 - Corresponds to 788049 muon tracks
- Loss made of several terms

$$l_{\text{Gaus}} = \sqrt{\frac{1}{2\sqrt{\pi}} \int_{-\infty}^{\infty} dx \left(G(x; I, w) - G(x; \hat{I}, \hat{w}) \right)^2} = \sqrt{\frac{I^2}{w} + \frac{\hat{I}^2}{\hat{w}} - \frac{2\sqrt{2}I\hat{I}}{\sqrt{w^2 + \hat{w}^2}}} \quad \leftarrow \text{Physics}$$

$$\left. \begin{aligned} l_{\text{Wasserstein},w} &= \frac{1}{N} \sum_{i=1}^N (\log w_i - \log \hat{w}_i)^2, \text{ where } w_i \leq w_{i+1}, \hat{w}_i \leq \hat{w}_{i+1} \\ l_{\text{Wasserstein},I} &= \frac{1}{N} \sum_{i=1}^N (\log I_i - \log \hat{I}_i)^2, \text{ where } I_i \leq I_{i+1}, \hat{I}_i \leq \hat{I}_{i+1} \end{aligned} \right\} \quad \leftarrow \text{Distribution Tails}$$

$$\left. \begin{aligned} l_{\text{H},\text{sparse}} &= - \left\langle \sum_{i=1}^4 s_i^w \log s_i^w \right\rangle - \left\langle \sum_{i=1}^4 s_i^A \log s_i^A \right\rangle \\ l_{\text{H},\text{ballance}} &= \sum_{i=1}^4 \langle s_i^w \rangle \log \langle s_i^w \rangle + \sum_{i=1}^4 \langle s_i^A \rangle \log \langle s_i^A \rangle \end{aligned} \right\} \quad \leftarrow \text{Model Capacity}$$

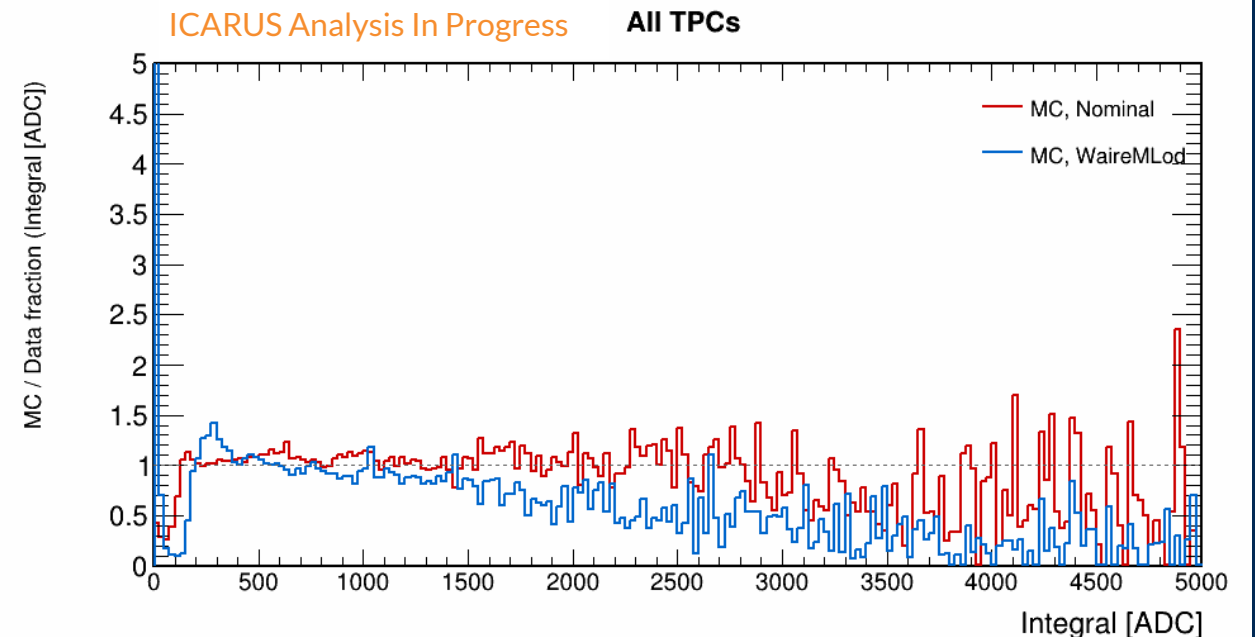
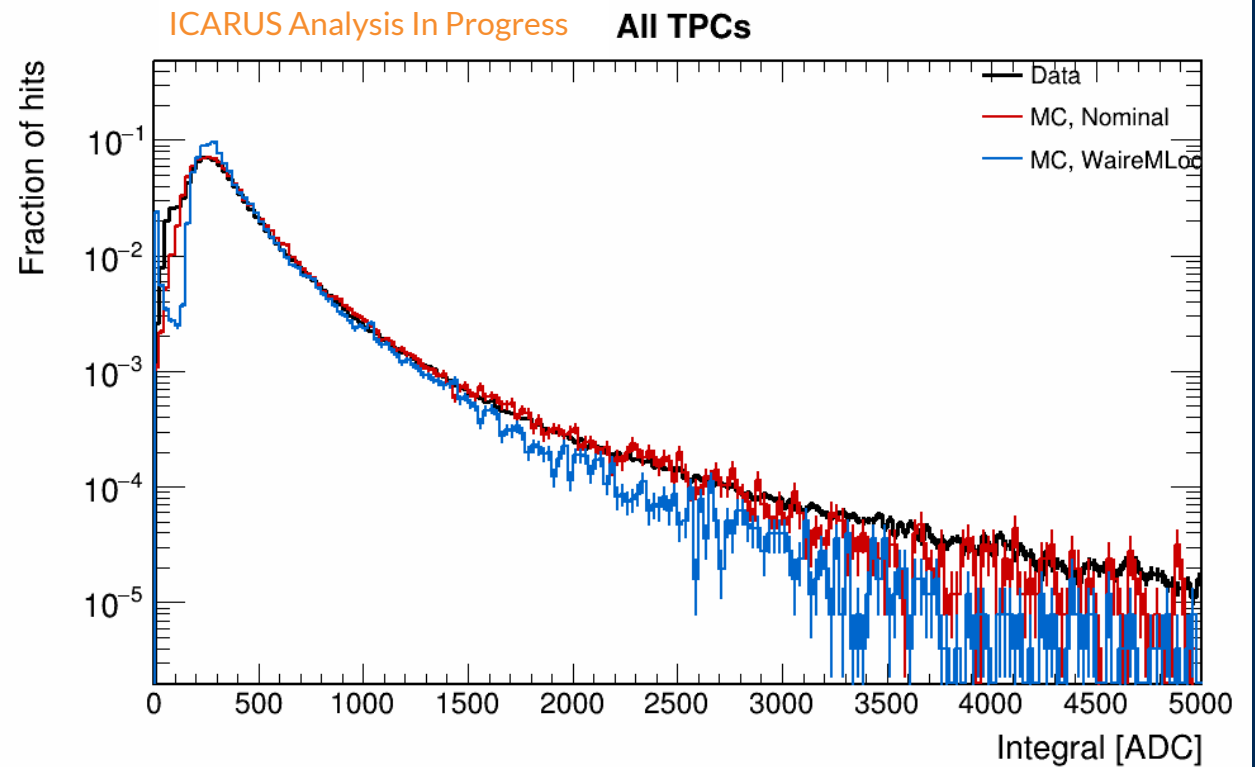
$$l_{\text{total}} = l_{\text{Gaus}} + \alpha_1 \times l_{\text{Wass},w} + \alpha_2 \times l_{\text{Wass},I} + \alpha_3 \times l_{\text{H},\text{sparse}} + \alpha_4 \times l_{\text{H},\text{ballance}}$$



Closure Tests

Integral

- We apply the model to MC hits using truth level equivalents to the reconstructed parameters
 - eg, true particle direction, true deposited charge, etc
- Compare the distribution of integrals and widths in the baseline MC and those inferred by the model to data hits
- Small hits are underperforming
- Shifting too many hits to the MPV, which may mean the model is not learning the spread of the distribution well enough
- That the integrals, for which the model was best performing, are mostly reasonable is a good sign that the application to MC features is working as intended

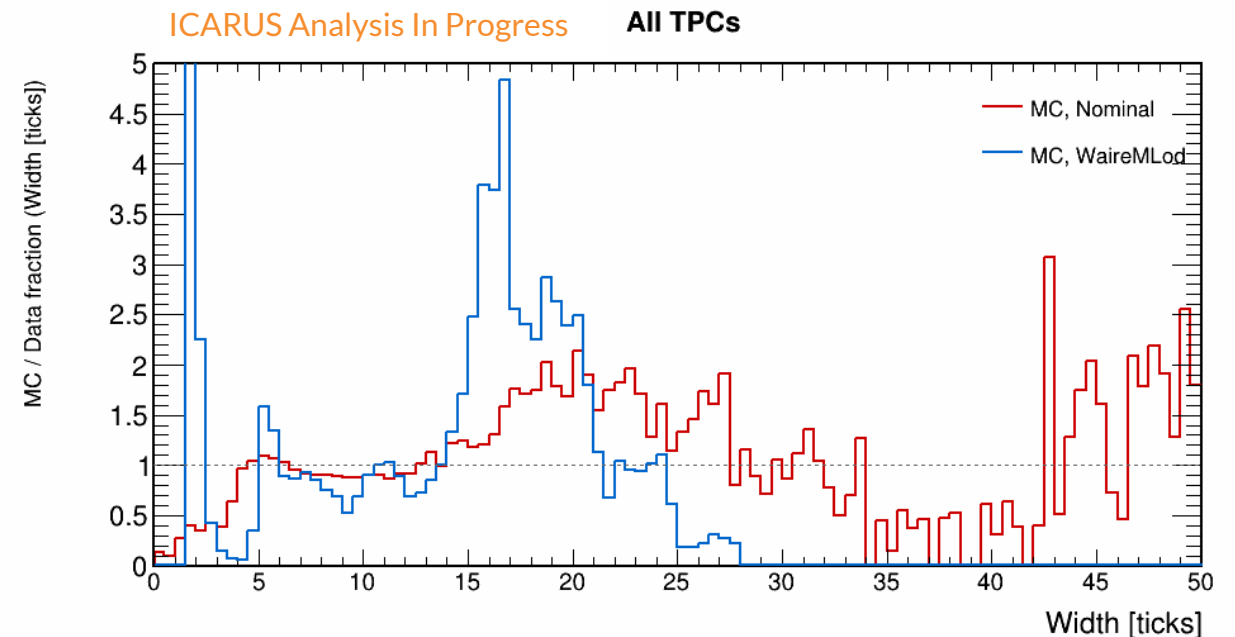
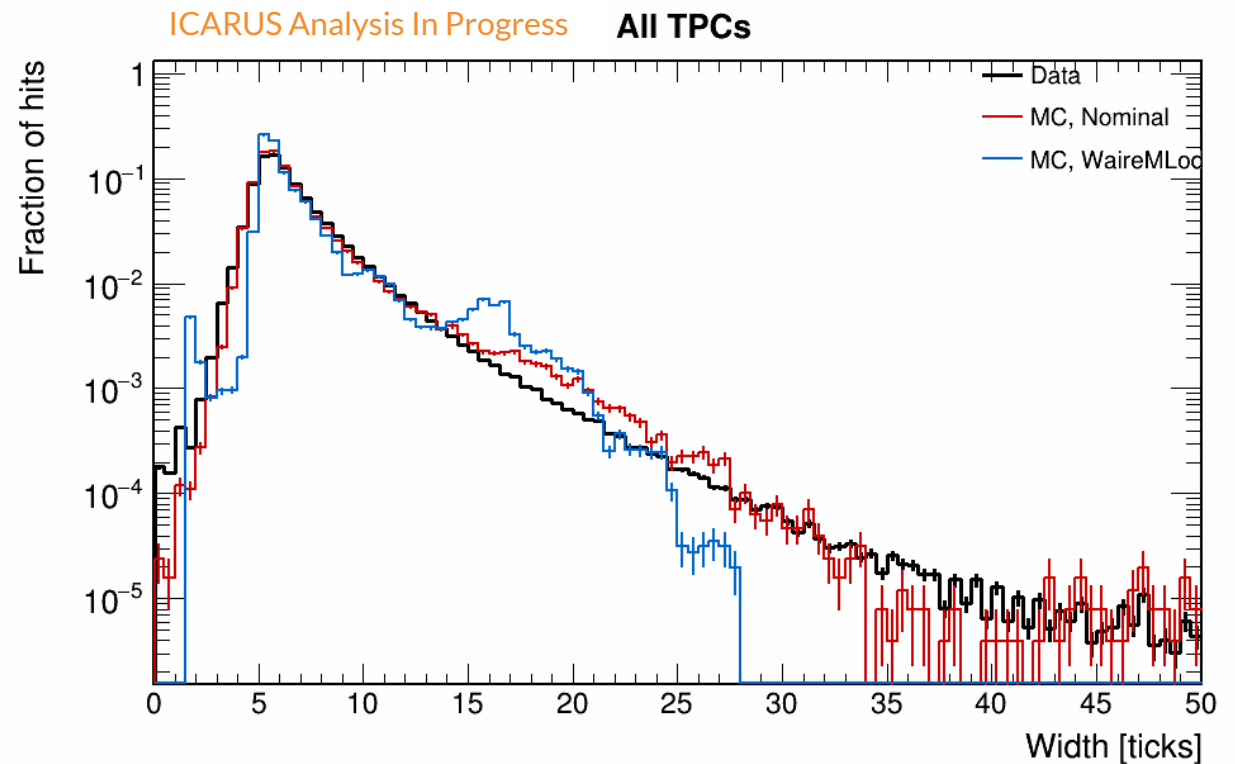




Closure Tests

Width

- Small hits are underperforming
- Shifting too many hits to the MPV, which may mean the model is not learning the spread of the distribution well enough
- The cut off in hit widths is likely due to limitations in the training data set where very narrow and extremely wide hits are rare
- Spikes at the high/low end could be boundary effects
- Not yet clear why the model pulls width towards ~18 ticks



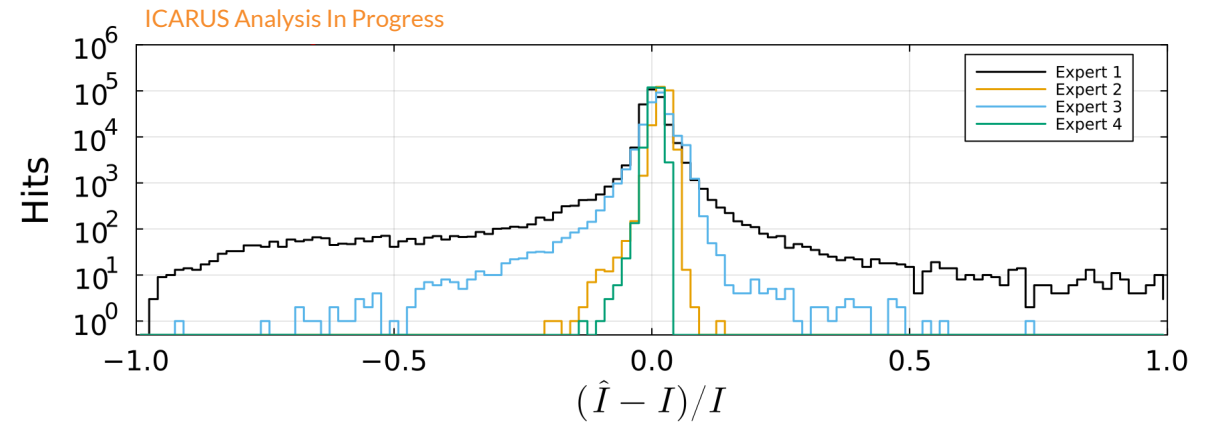


Evaluation

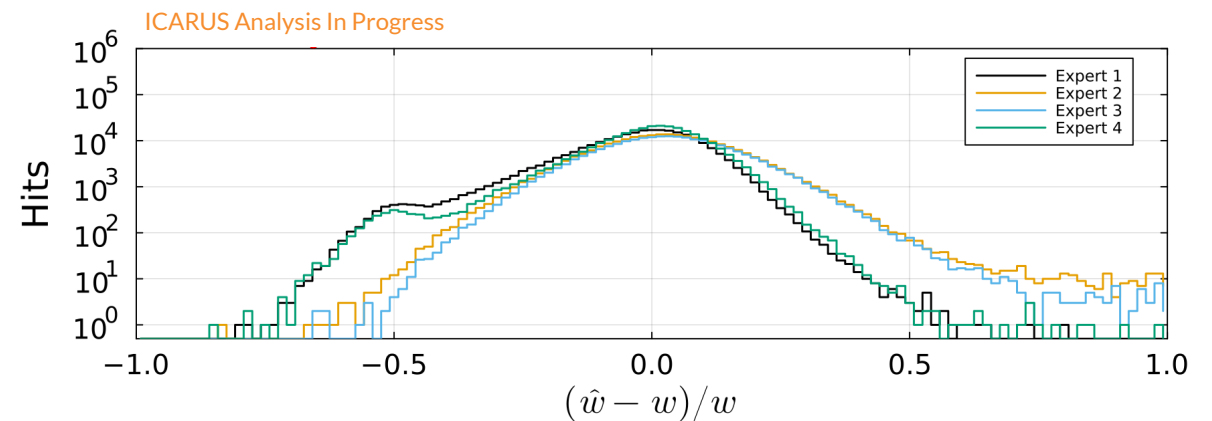
Model Accuracy

- On a validation sample of 4 million hits we get mean integral and width biases of 0.3% and 0.7%, with a resolution in the error of 4.3% and 28% respectively
- We can look at how the model performs by each expert along the A-head and the w-head
- A-experts 2 & 4 perform significantly better, shrinking the resolution down to 1.1% and 0.8%
- w-experts all perform roughly equivalently

Integral error by dominant expert



Width error by dominant expert

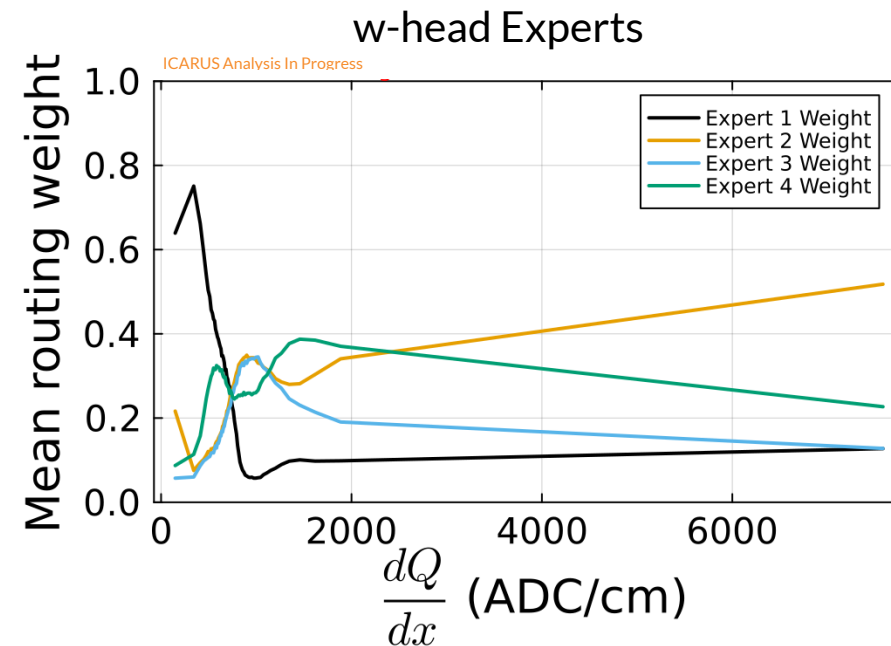
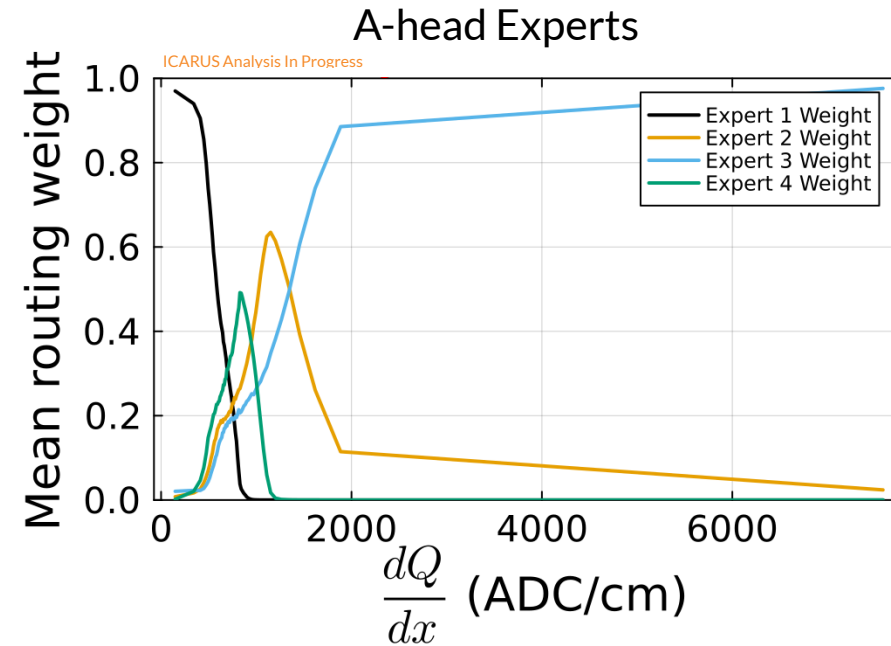




Expert Specialization

dQ/dx

- The A-head experts split the dQ/dx parameter space into low-/mid-/high-range energy deposition regimes
 - The most accurate of these experts being in the mid-range
- The w-head has a low-range expert, but other wise is agnostic to the stopping power
- Since deposited energy is strongly correlated to the size of the hit this split makes physical sense

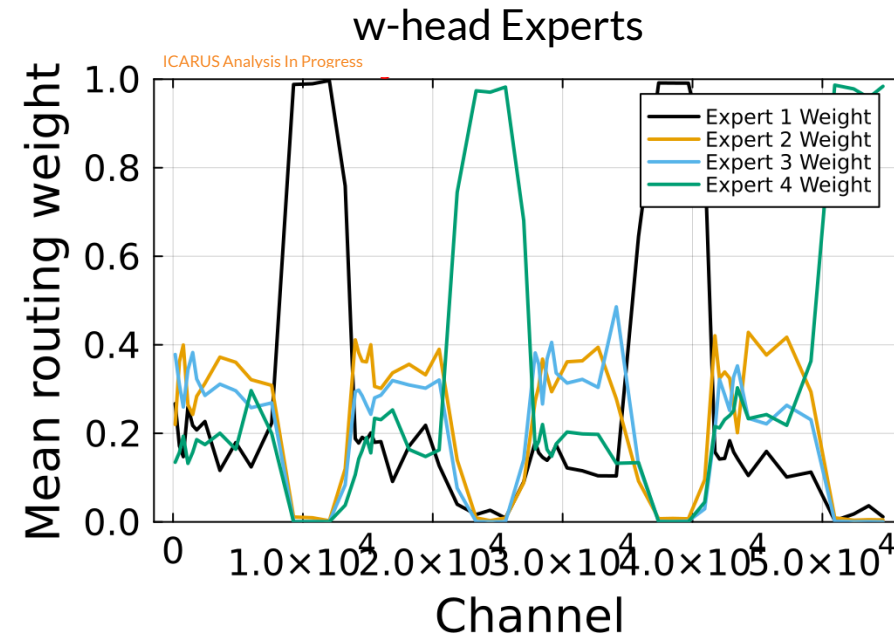
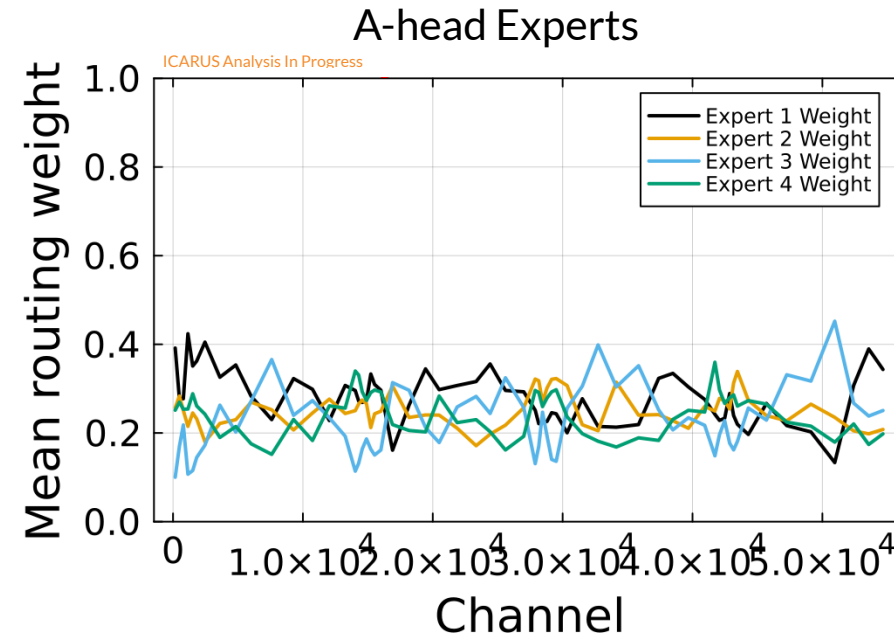




Expert Specialization

Channel Number

- The w-head experts 1 & 4 specialize strongly by the readout channel, splitting the collection plane hits
 - Expert 1 gets EE & WE collection planes while expert 4 gets EW & WW
 - The orientation of the wires flips between TPCs which may explain the split
- The A-head experts do not specialize by channel at all
- Such sharp routing is possibly telling us the collection plane hits need special handling

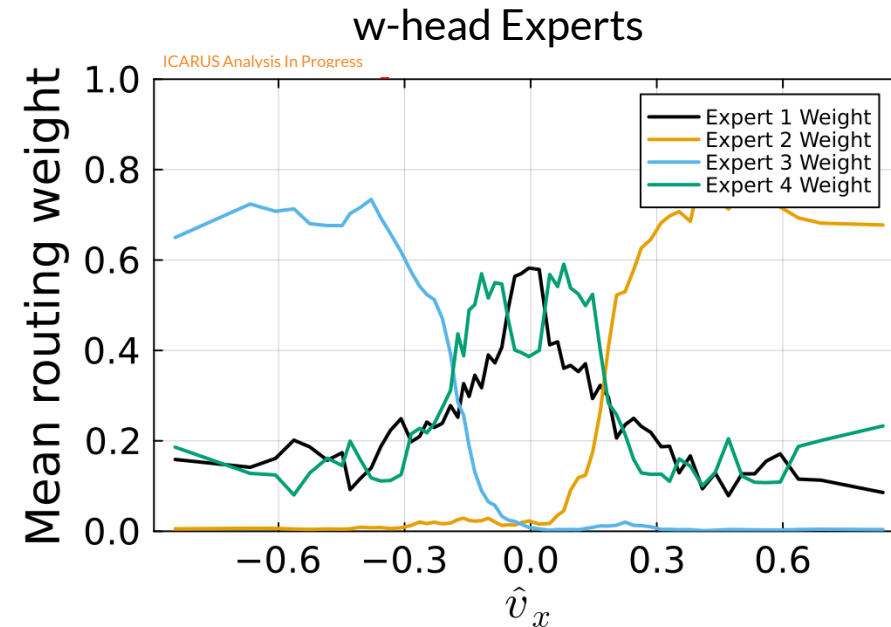
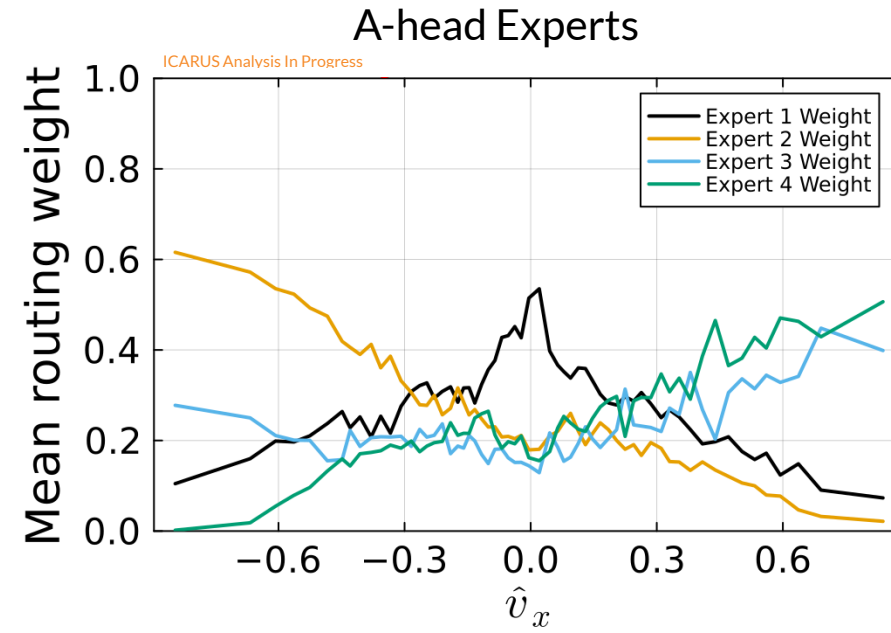




Expert Specialization

Direction Relative to Drift

- Both head's experts prefer to hand off the hits from tracks perpendicular to the wire planes to one expert
 - The w-head experts in the $\hat{v}_x \sim 0$ regime are also the collection plane experts
- The parallel and anti-parallel to the drift get split by both sets of experts
- Could be angular dependence in recombination





Future Improvements

What is to be done

- While the model can encode the hit integral well in the mid- dQ/dx range, improvements need to be made on accuracy at small hits
- Width accuracy is poor at most ranges and needs targeted improvements
- Integrals and widths at the high- and low-tails are performing weakly
- The mixture of experts seem to work well for the hit integrals, but all experts perform similarly for the widths
 - Experiment with more experts for the widths
 - Drop experts for the width head?
- Wasserstein distance is not balancing the distributions as we would like



Conclusions

- We believe that machine learning could provide a new and powerful tool in generating data-like MC samples
- We take a first look at training a multilayer perceptron on off-beam cosmic muons from ICARUS in hopes of improving detector systematics
- In training the model we are able to separate TPC hits into different regimes we expect to be sensitive to different systematic effects
- Model specialization needs tuning as the mixture of experts are specializing along multiple axes at once
- Improvements need to be made in the model's ability to reconstruct the hit width before we can consider it reliable
- We are hopeful this method will enable us to better constrain the systematic uncertainties in ICARUS and improve our sensitivities for future analyses



Fermi *FORWARD*



U.S. DEPARTMENT
of ENERGY