Transformer-based Reweighting of Neutrino Generators

Andrew Cudd, Roger Huang, Callum Wilkinson

6/18/2026

NPML 2026, UC Irvine

BERKELEY LAB

Neutrino Event Generators

Neutrino event generators simulate the possible output products of any particular neutrino-nucleus interaction

- Key to any simulation-based analysis!

Several generators exist, taking different theoretical approaches

NuWro * Wrocław Neutrino Event Generator *

Genie
UNIVERSAL NEUTRINO GENERATOR & GLOBAL FIT

NEUT

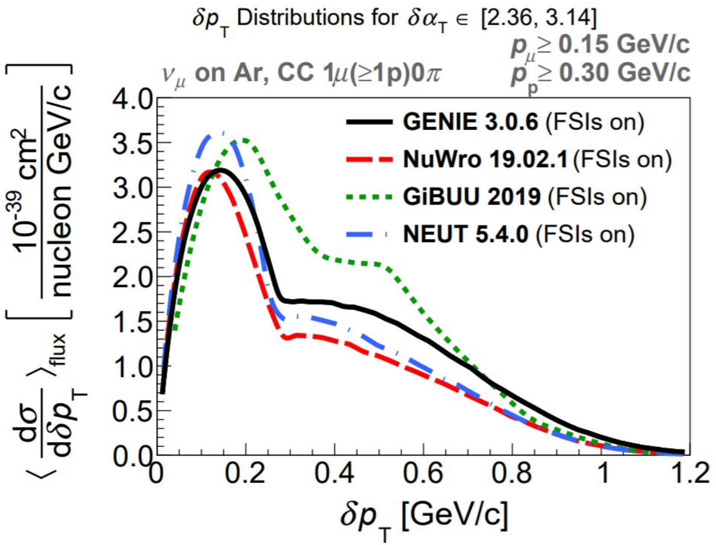
ACHILLES

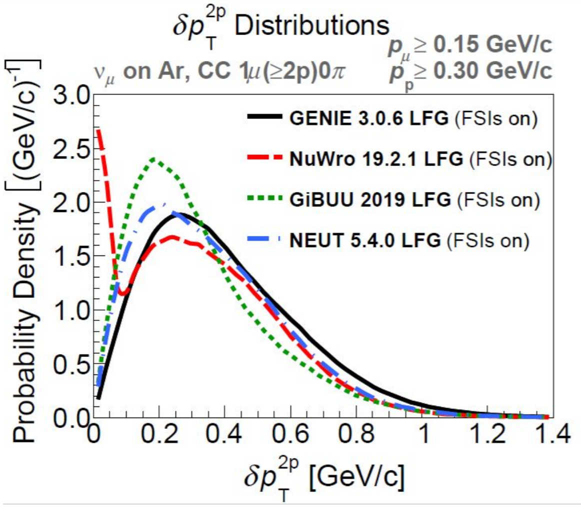
GiBUU
The Giessen Boltzmann-Uehling-Uhlenbeck Project

Neutrino Event Generators

Many approximations and assumptions are used in every existing generator, and they can be tuned for different purposes

Due to our still relatively poor understanding of neutrino-nucleus interactions, the **generators often disagree quite noticeably** with each other

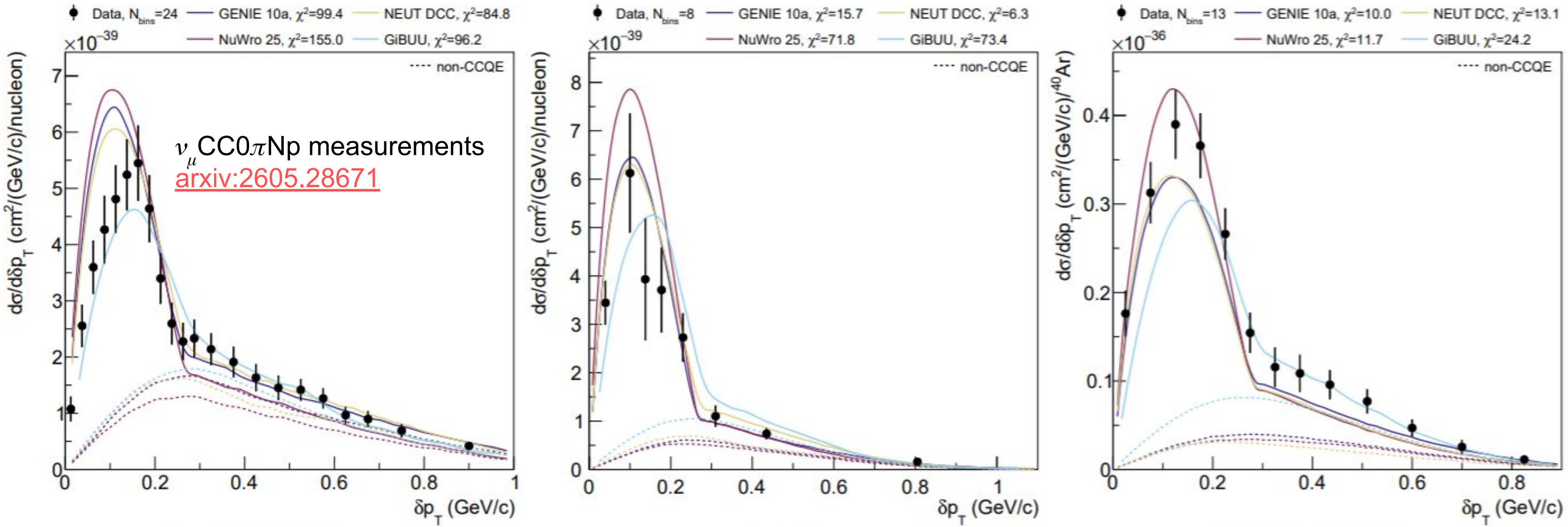




Plots: arXiv:2201.04664

Neutrino Event Generators

No individual generator is a global good fit to all existing neutrino measurements



(a) MINERvA (b) T2K (c) MicroBooNE

Neutrino Event Generators

Since (probably) no generator correctly represents nature, we don't want analyses to be overly dependent on their modeling assumptions

- Check how analysis changes depending on choice of generator

Detector simulation is computationally very expensive

- We cannot generate a full statistics detector-reco-level MC dataset for every generator we want to test against
- Most generators provide ways to reweight their results so that modeling uncertainties can be studied without re-running detector simulation

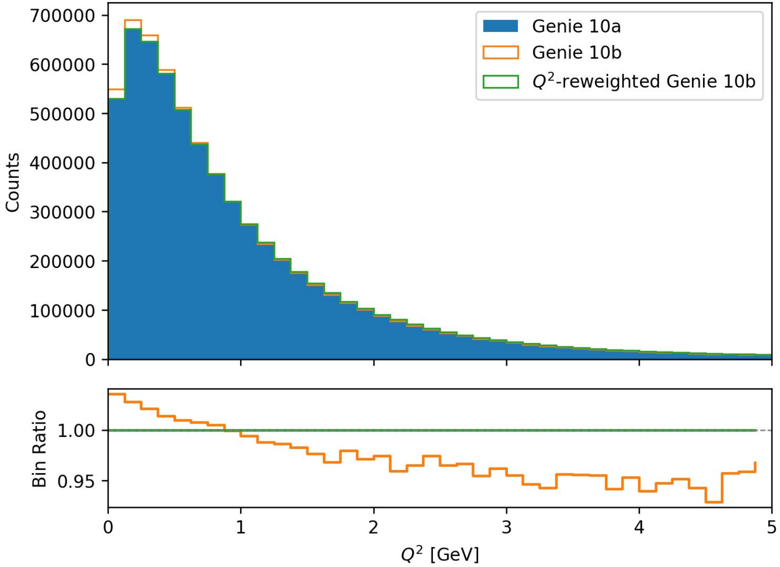
Running just the generators is computationally cheap

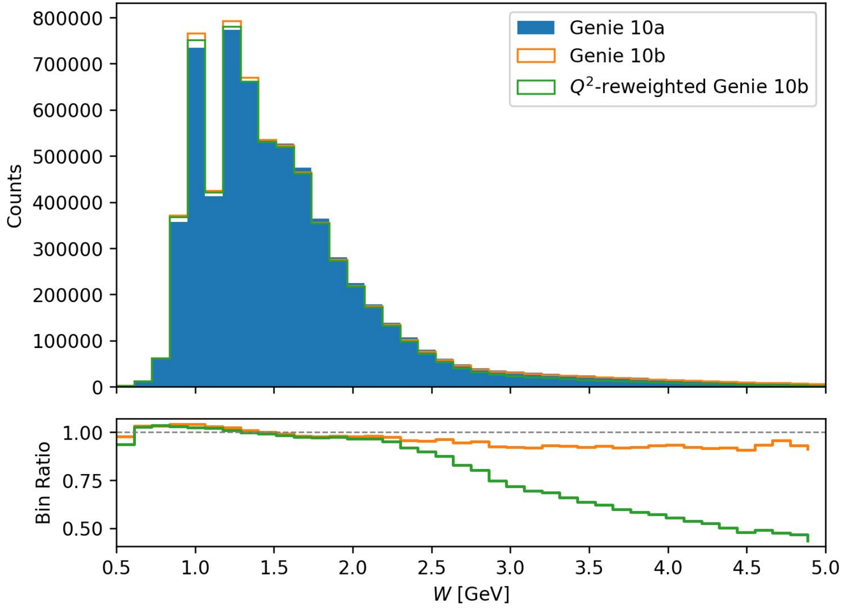
- Ideally, run only one set of events through detector simulation and reweight it to match other generators

Bin-based Reweighting

Traditional reweighting methods were bin-by-bin or with splines

But binned reweighting in one variable (e.g. Q^2) does not generally give valid results for other variables (e.g. W)



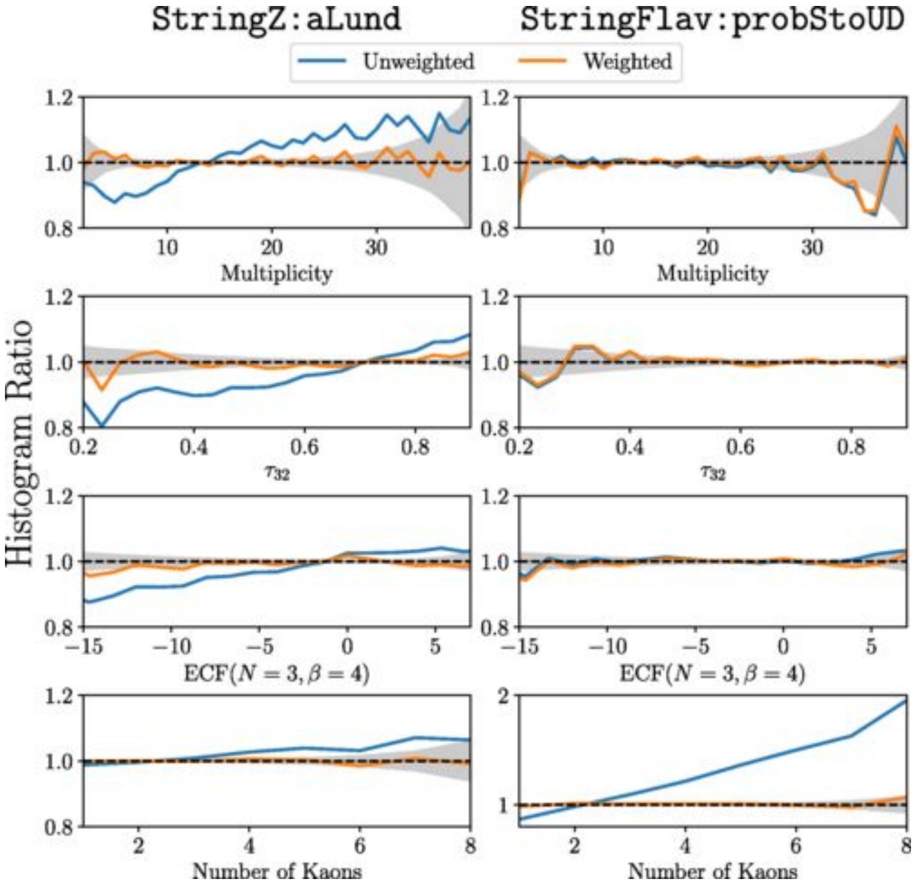


ML-Based Generator Reweighting with the Likelihood Ratio “Trick”

A classifier trained to distinguish between two datasets **learns an approximation to their likelihood ratio***

Reweighting one generator’s output to match another’s is **finding the likelihood ratio** between their distributions

- The reweighting is a function of **all inputs we use to train the classifier**



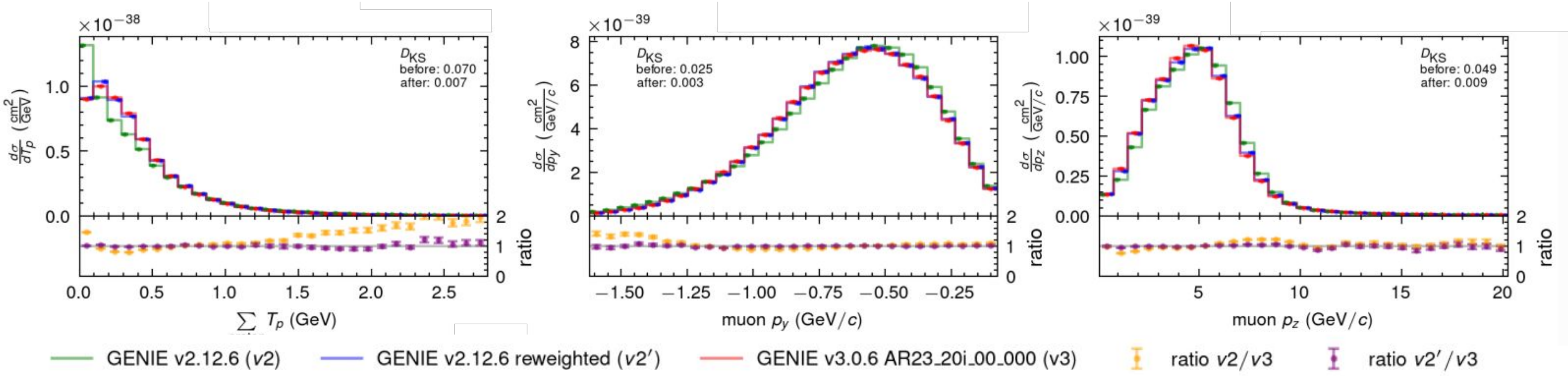
*J. High Energ. Phys. 2024, 136 (2024)

Phys. Rev. D 101, 091901

BDT Reweighting for Neutrino Generators

Prior work using **BDTs for multivariate reweighting** of neutrino generator outputs has already been done in MINERvA (pictured) and DUNE

But BDTs are generally not powerful enough to cover the entire possible phase space of interest from generators



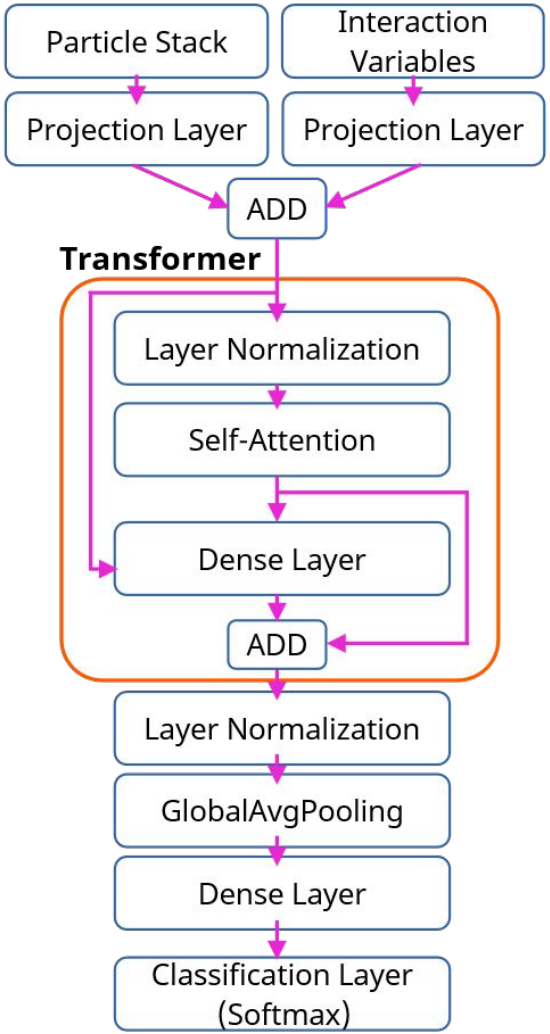
arXiv:2510.07463

Transformer-based Reweighting

Ideal reweighting would be based on the **full particle stack** output of the generators

- All observables of interest are derived from this particle stack
- Reweighting would be **valid for any analysis with any experiment/detector**

Transformer architecture is well-suited for learning the inter-particle relations and handling variable-length particle stacks.



Training Details

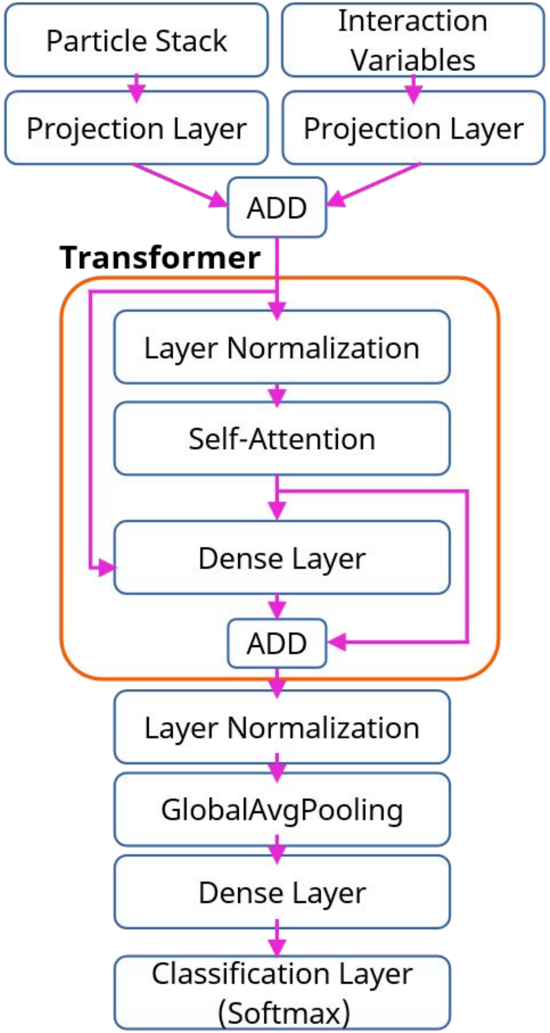
Training data: ~50M Genie10a ν_μ CC events and ~50M Genie10b ν_μ CC events, generated with flat neutrino flux

Model inputs:

- Up to 10 particles (muons, protons, neutrons, pions) per event, with 4-momentum and PID for each
- Event-level variables q_0, q_3, Q^2, W, E_ν
- ~1M model parameters

Training time: ~4 hours on 6x4 A100 GPUs

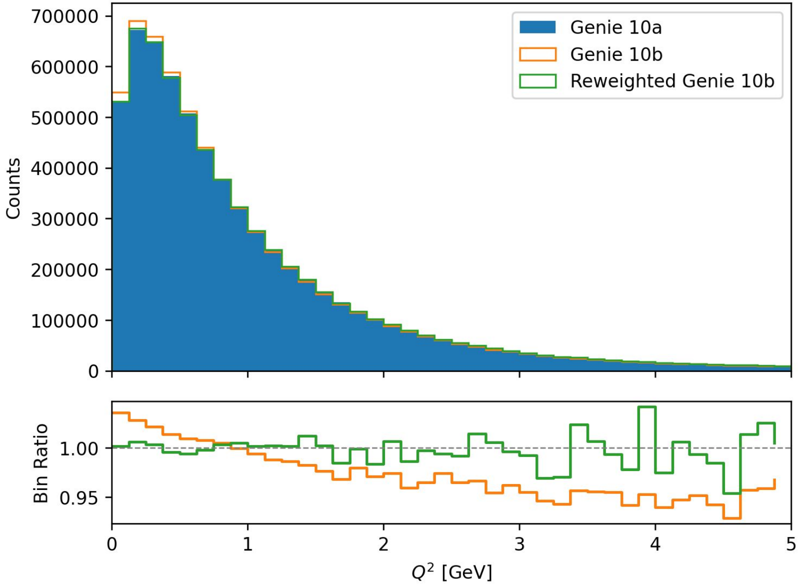
Test data: ~10M Genie10a and Genie10b ν_μ CC events generated with DUNE near detector neutrino flux

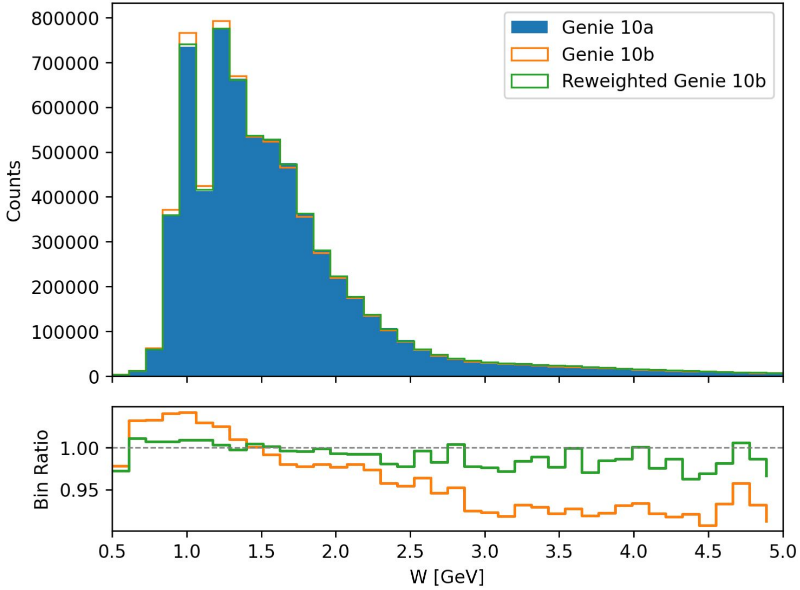


1D Results

Test setup: Reweighting Genie 10b ν_μ CC events to match Genie 10a

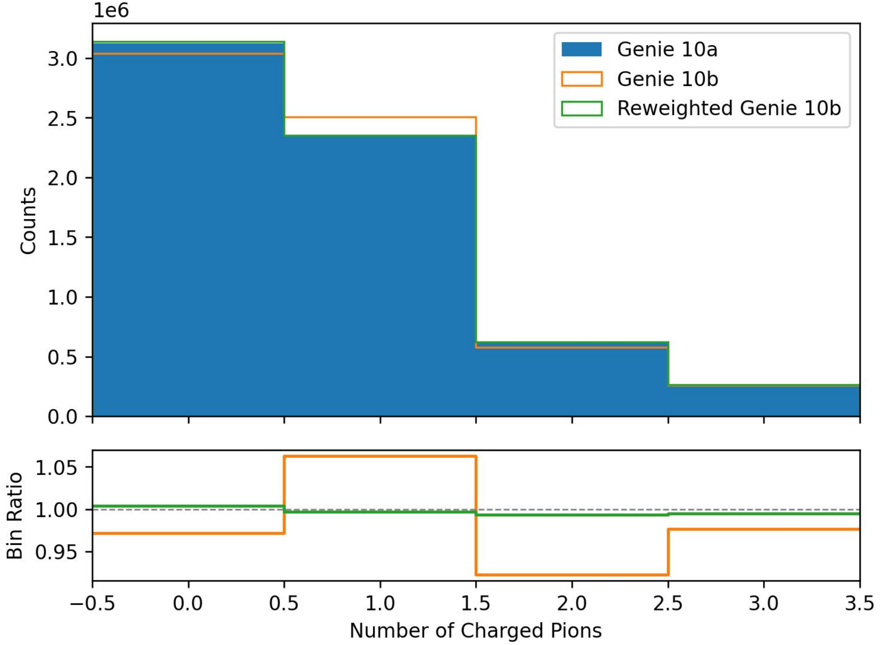
Projecting the results into a few 1D comparisons shows the reweighted Genie 10b outputs look reasonably similar to the Genie 10a outputs

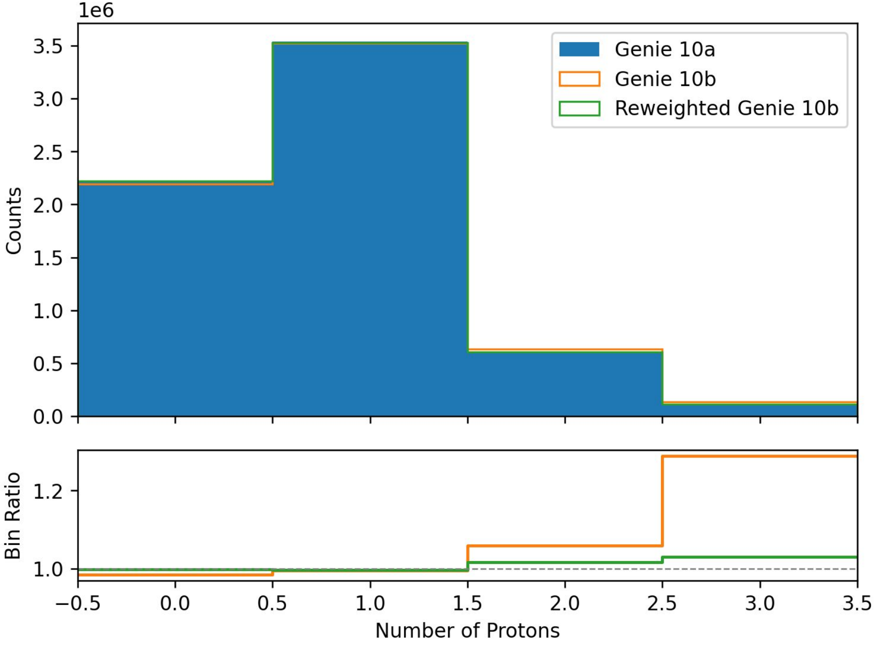




1D Results

The number of $\pi^{\pm}$ and protons over some detector threshold (150 MeV for pions, 450 MeV for protons here) is not an explicit input, but is learned quite well too





Metrics - Effective Sample Size

What result is **good enough**? Ideally checked with an actual analysis, but what if we want to evaluate our general usefulness?

In the limit of infinite statistics, we expect to always be able to distinguish between the reweighted distribution and the target distribution

Real analyses do not have infinite statistics in their MC, so what we actually care about is an **effective sample size**

- A level of statistics where the reweighted distribution and target distribution are identical within statistical uncertainties

Many standard tests exist for this on binned distributions, but we want an **unbinned comparison**, since our method is performing unbinned reweighting

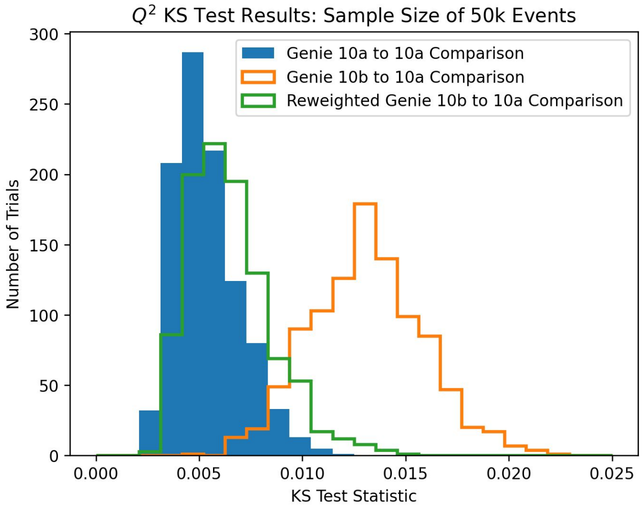
1D Metrics - KS Test

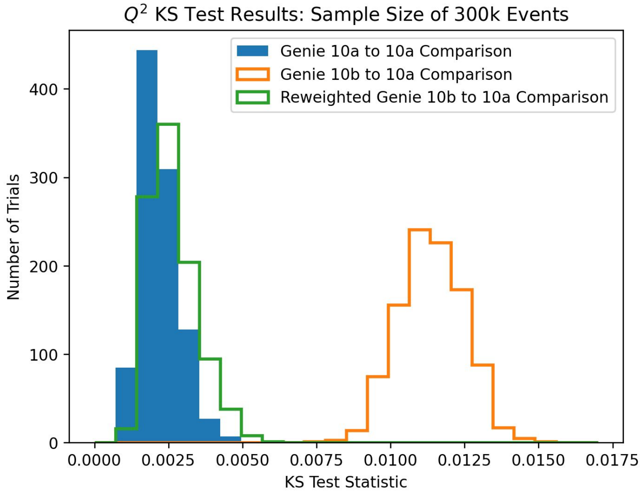
KS test provides a standard unbinned 1D comparison metric

For any given sample size, we can bootstrap resample our events to build a distribution of KS test results given the available statistics

Larger test statistics gives greater separating power as expected

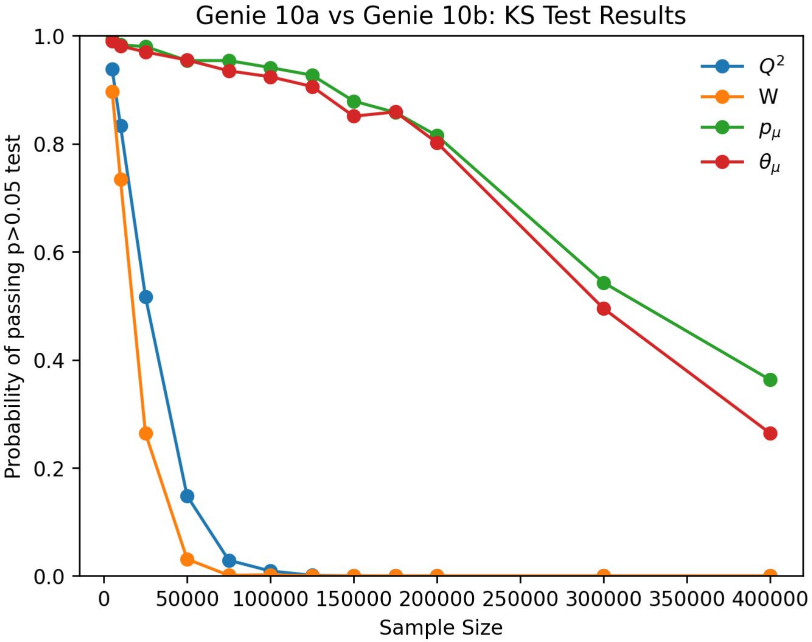
Does comparing our reweighted result against the target distribution (Genie 10a) give **a KS test result compatible with a Genie 10a self-comparison?**



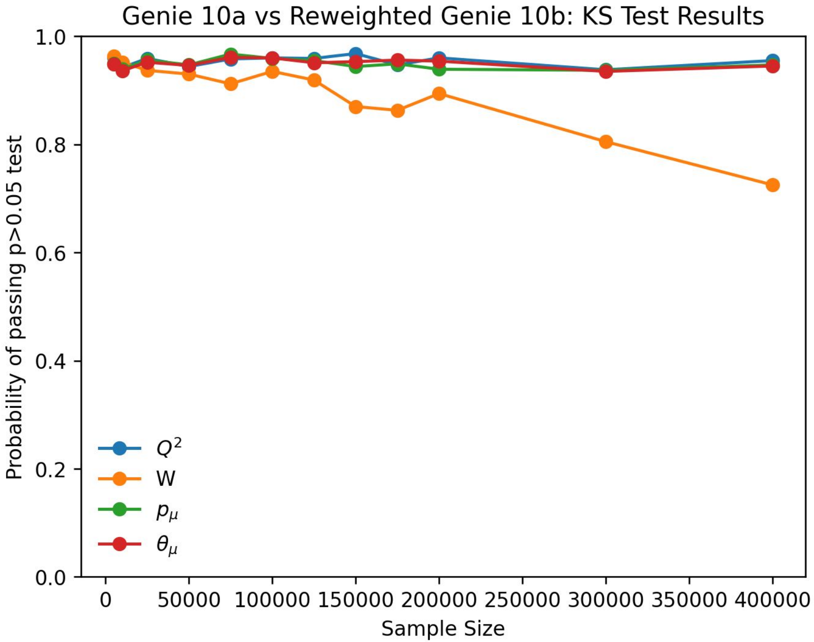


1D Metrics - KS Test Results

Without reweighting: sanity check that the KS test distributions are very distinguishable at small sample sizes



Transformer-based reweighting: reasonable up to fairly large sample sizes



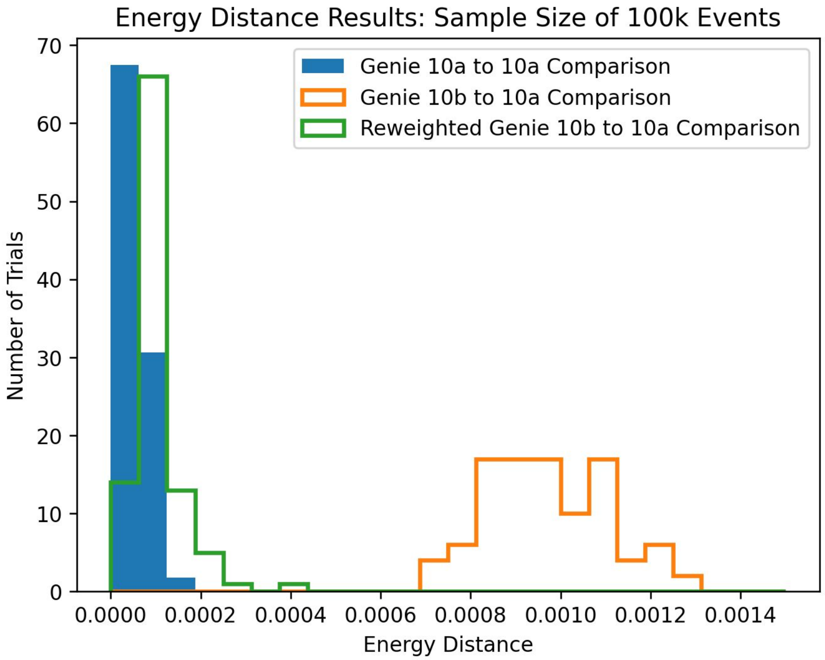
Multidimensional Metrics - Energy Distance

$$D^2(F,G) = 2\,\mathrm{E}\,\|X - Y\| - \mathrm{E}\,\|X - X'\| - \mathrm{E}\,\|Y - Y'\|$$

Ideally want an **unbinned, multidimensional test** to evaluate our reweighted result

Energy distance provides this, for some choice of observables:

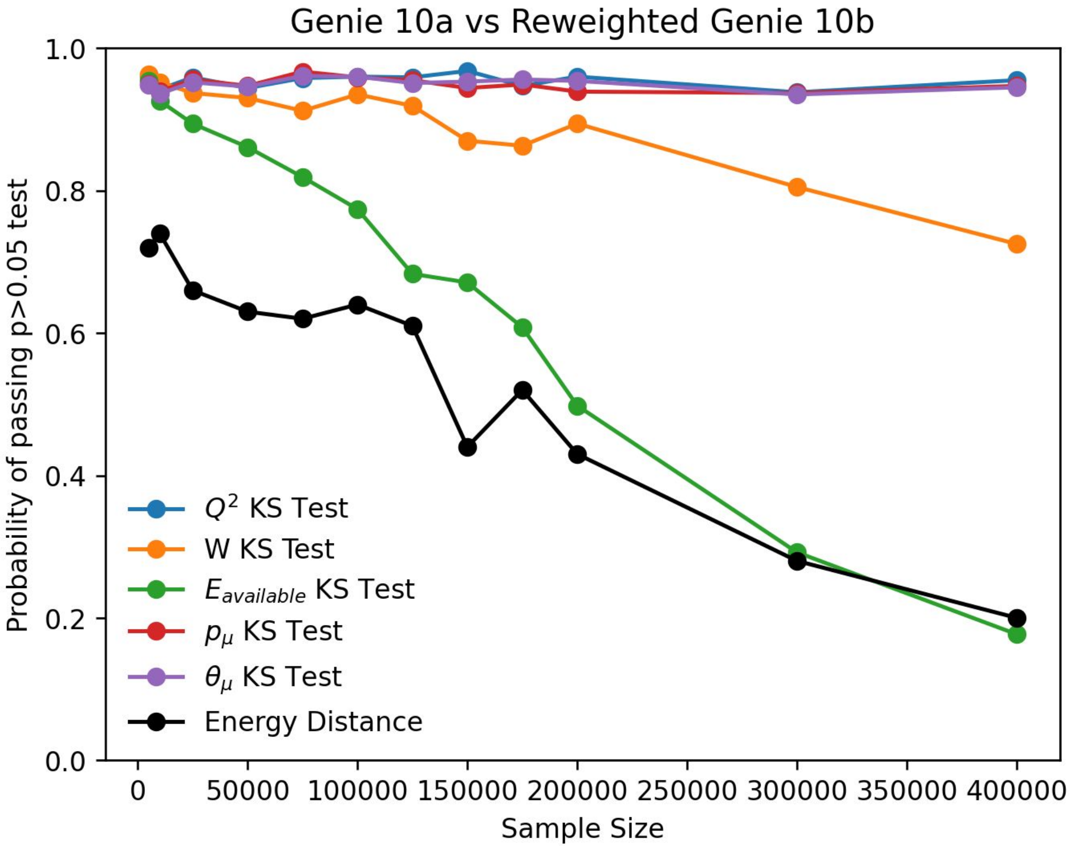
- Q^2, W, $E_{available}$, p_μ, θ_μ, Number of $\pi^\pm$ with p > 150 MeV, Number of protons with p > 450 MeV



Combined Metrics

Energy distance with all observables combined **shows more discriminatory power** than the individual KS tests

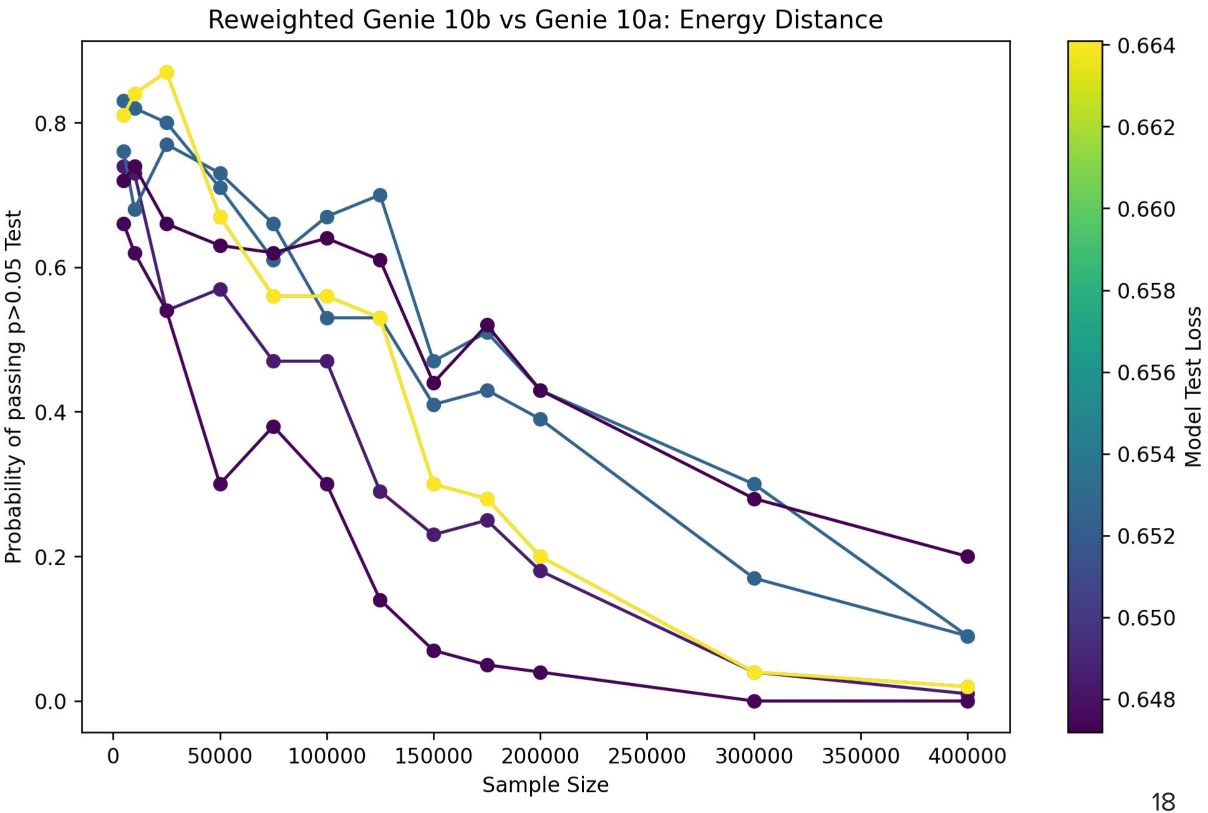
The cutoff for “good enough” will still be arbitrary, but this provides a fair comparison between different reweighting results



Model Comparisons

Modifying some of the model's parameters results in different test losses

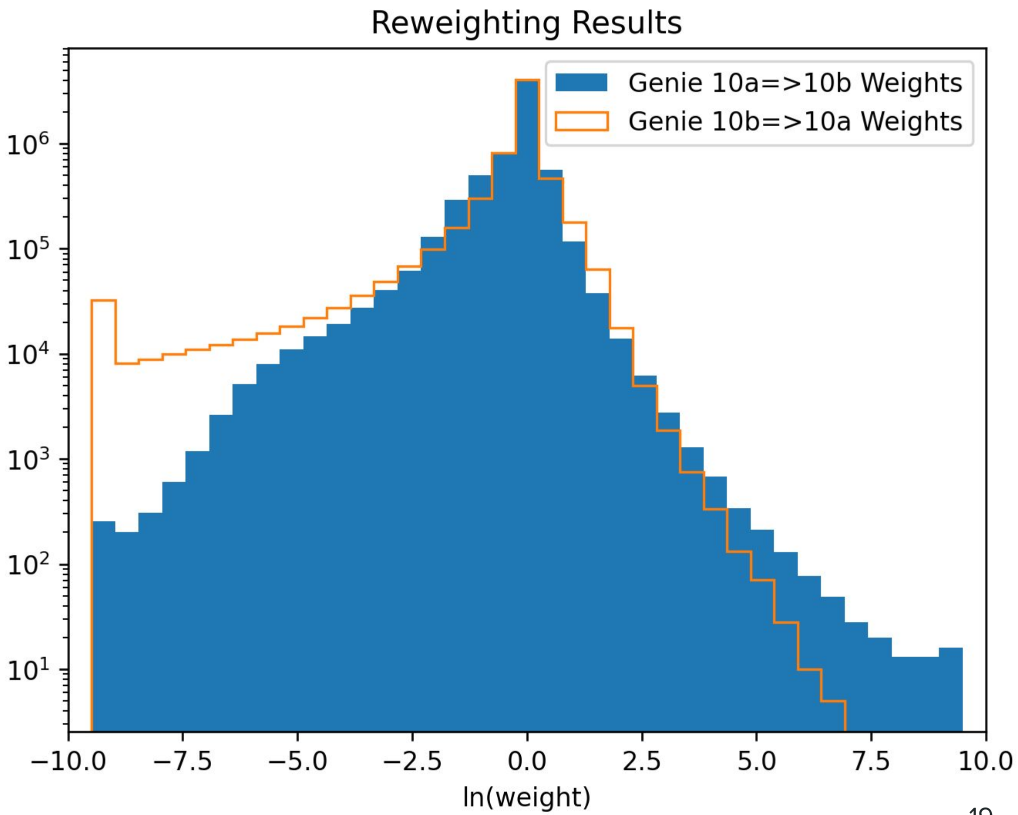
Within some range, performance does not strictly improve with lower test loss



Notes on General Reweighting

Reweighting between models is problematic when they have disparate phase spaces

- **Weights near 0:** source model covers a phase space that's not in the target model
- If target model includes phase space not in the source model, **these will never be covered**
    - Or with very large fluctuations through large weights on the source model

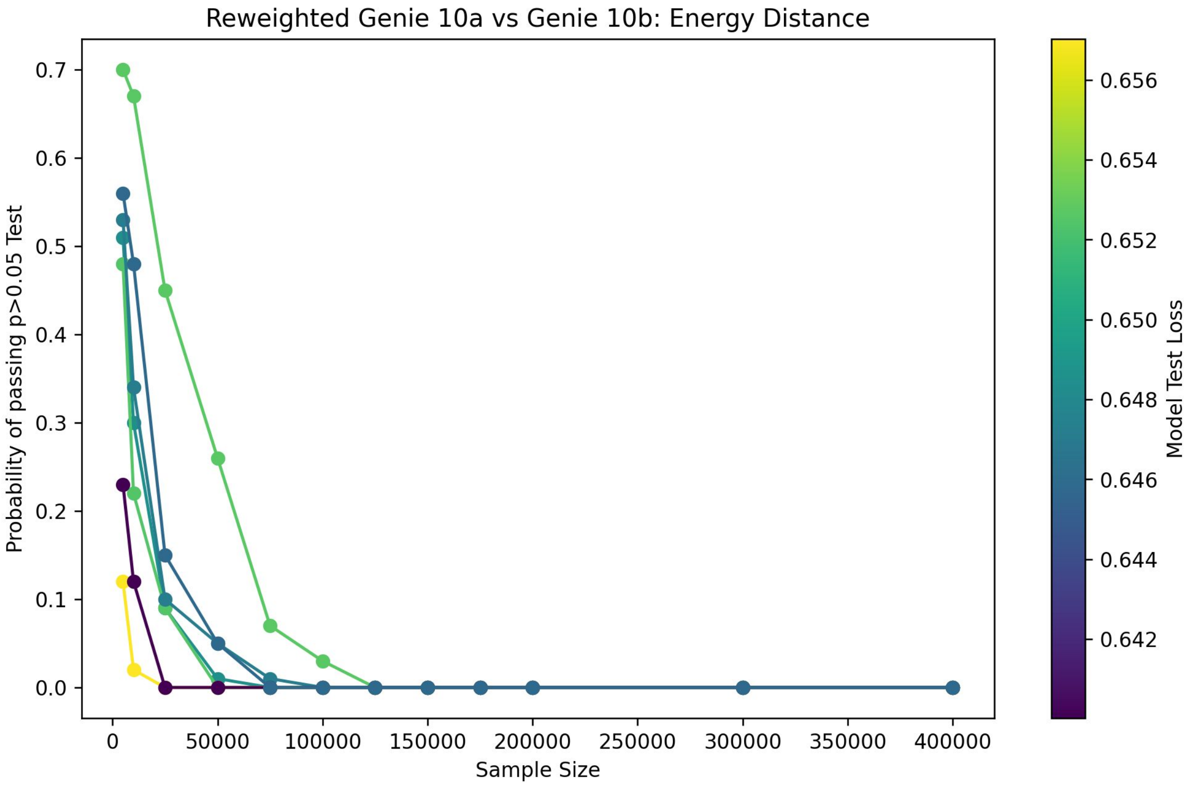


Reverse Direction: Reweighting Genie 10a to 10b

Reweighting from Genie 10a to 10b shows much worse performance

Near-0 weights in the 10b => 10a direction suggest 10b includes classes of events that don’t exist in 10a

- 10a can never be reweighted to cover these events



Summary

Neutrino analyses would benefit from a generator reweighting method that is valid over all observables of interest simultaneously

Modern ML classifiers using the likelihood ratio trick show promise for this

- Such models can be trained once and used for any experiment
- We have demonstrated some unbinned multidimensional metrics that can **evaluate and compare** the quality of their likelihood ratio estimations

This does not yet allow reweighting between generators covering disparate phase spaces, but a surrogate model might