Exploring GANs for the simulation of ν_μ scattering and their adaptability to new data using transfer learning.

Neutryno Physics Machine Learning 2026

José L. Bonilla
joseluis.bonillaramirez@uwr.edu.pl

June 16, 2026

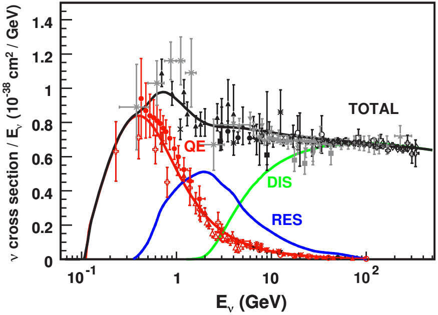
Uniwersytet Wrocławski

Outline

Motivation

Present and future experimental collaborations such as T2K, MicroBooNE, DUNE, and Hyper-Kamiokande goal is to perform precise measurements of the neutrino oscillation parameters.

- Require procedures for E_ν reconstruction from detected final-state particles.
- Rely on Monte Carlo (MC) simulations.
- Limitations due to models assumptions and approximations.
- Fitting MC to data sometimes is difficult.



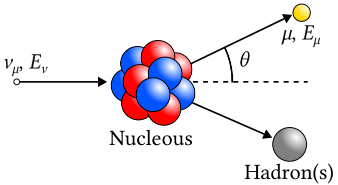
Rev. Mod. Phys. 84 (2012), pp. 1307–1341

Objective

- We focus on the development of alternatives that can learn directly from available experimental data.
- We use *conditional Generative Adversarial Networks* (cGANs) to simulate muon neutrino ν_μ scattering and obtain muon kinematic distributions.
- The resultant model can be “extrapolated” to an unseen experimental data set to further train using transfer learning.
- The idea of transfer learning with electron scattering cross sections has been explored before (Graczyk, Kowal, et al: Phys.Rev.Lett. 135 (2025) 5, 5)

Objective

In particular, we simulate charged current (CC) muon neutrino-carbon scattering[1]. Implemented on `Keras 2.15` (`Tensowflow`) For our first study, we created two networks:



Neutrino-nucleus scattering

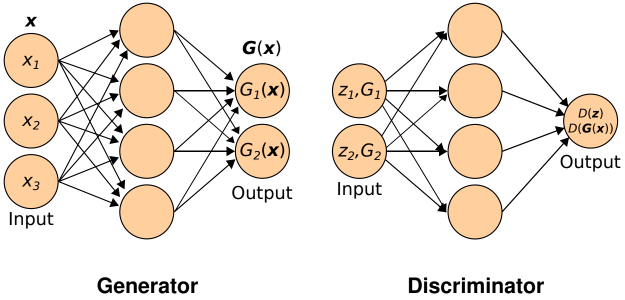
- Quasi-elastic (QE) scattering (not shown today).
- Inclusive (INC) scattering (quasi-elastic and all other process).

[1] All details: Phys.Rev.D 112 (2025) 1, 013007

What are GANs?

GANs enable training an artificial neural network *Generator* that learns to generate data produced by an unknown or poorly understood mechanism, via a second network *Discriminator* loss.

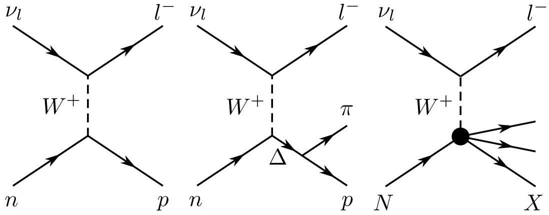
- A GAN model takes an input latent vector $\boldsymbol{x}$ of “noise” drawn from a Gaussian distribution and generates samples that match the features of the training data.



NuWro

- The GAN is trained on data generated by the NuWro `21.09` MC generator to produce scattering events.
- NuWro has been under development since 2004 at the University of Wrocław.
- It is optimized for the energy range characteristic of accelerator-based neutrino sources (hundreds of MeV to tens of GeV).

$\Rightarrow$



Kinematic settings

Input:

- Neutrino energy E_ν.

Output, we fous on μ kinematics only:

- Muon energy E_μ.
- Muon dispersion angle θ w.r.t neutrino beam.

The longitudinal momentum $p_{\mu,z}$, E_μ, and θ are related as follows:

$$p_{\mu,z} = \sqrt{E_\mu^2 - m_\mu^2} \cos\theta, \tag{1}$$

where m_μ is the muon mass.

Kinematic settings

We consider neutrino energies following a **random uniform** distribution from $E_{\nu,min} = 300$ MeV to $E_{\nu,max} = 10$ GeV. For optimization porpoises, re-scale the energy variable:

$$E'_\nu = 2\frac{E_\nu - E_{\nu,min}}{E_{\nu,max} - E_{\nu,min}} - 1,$$

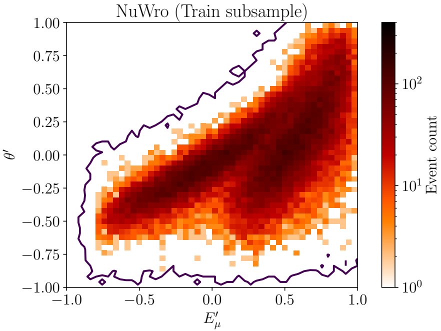
Similarly, for the training of the generator we use a pair (E'_μ, θ'), given by:

$$E'_\mu = 2\sqrt{1 - \frac{E_\mu - m_\mu}{\Delta E}} - 1, \quad \theta' = 2\sqrt{\frac{\theta}{\pi}} - 1,$$

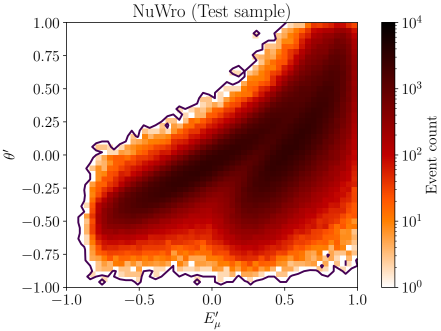
where $\Delta E = E_\nu - 9.8\text{ MeV} - m_\mu$.To train the GAN, we sampled $\sim$ 4M events with NuWro.

Data coverage neutrino-carbon model

Even though we can produce a huge amount of data for training, it is very difficult to cover the whole kinematic region with the training dataset.



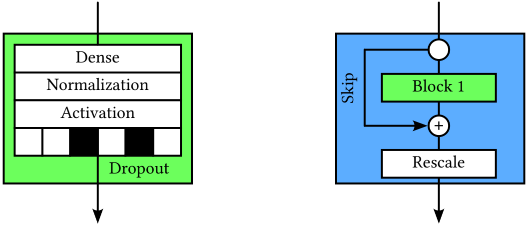
$950 < E_\nu < 1050$ MeV



$E_\nu = 1$ GeV

Network architecture

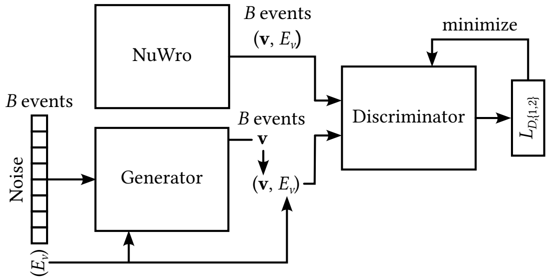
The Generator G and Discriminator D are made with basic building blocks as follows:



Which we called `Block 1` and `Block 2`, respectively. Furthermore, we concatenate the initial E'_ν information to the input of `Block 1` to create `Block 3`, and then `Block 4`. Additionally, in some of the networks we add Gaussian noise in some blocks.

GAN optimization scheme

GANs are optimized using the so-called mini-max framework. A successful model is a result of a balance between two losses: of the discriminator and of the generator.



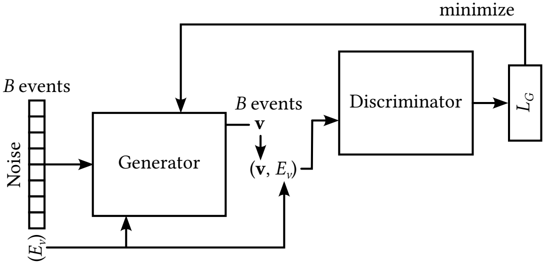
$$L_{D,1}(\mathbf{v}) = -\frac{1}{B}\sum_{j=1}^{B}\log[1 - D(\mathbf{v}_j, E_{\nu,j})]$$

$$L_{D,2}(G) = -\frac{1}{B}\sum_{i=1}^{B}\log[D(G(\mathbf{x}_i, E_{\nu,i}), E_{\nu,i})]$$

The loss used is cross-entropy.

GAN optimization scheme

GANs are optimized using the so-called mini-max framework. A successful model is a result of a balance between two losses: of the discriminator and of the generator.



$$L_G(G) = -\frac{1}{B}\sum_{k=1}^{B} \log[1 - D(G(\mathbf{x}_k, E_{\nu,k}), E_{\nu,k})]$$

The loss term is inverted for G

Models quality

A way of checking the quality of the model is by estimating the *pull*. Given the histograms of events generated by NuWro and the GAN, the i-th component of the pull is

$$\text{pull}_i = \frac{n_{NuWro,i} - n_{gan,i}}{\sqrt{\sigma^2_{NuWro,i} + \sigma^2_{gan,i}}}, \tag{2}$$

where $n_{\text{NuWro},i}$ and $n_{gan,i}$ are the bin contents at the i-th bin, and $\sigma^2_{\text{NuWro},i} = n_{NuWro,i}$ and $\sigma^2_{gan,i} = n_{gan,i}$ are statistical uncertainties. The bin contents of a histogram are expected to follow the Poisson distribution.

Models quality

To check the goodness of the model, we evaluate the mean of the absolute value of pulls

$$\mathrm{MAP} = \frac{1}{K}\sum_{i=1}^{K} |\mathrm{pull}_i| \tag{3}$$

on 2-dimensional (E'_μ, θ') distributions, where K is the number of bins that satisfy $n_{nuw,i} \neq 0$ and $n_{gan,i} \neq 0$, or $n_{nuw,i} > 5$ and $n_{gan,i} > 5$ to remove “tails” with low statistics.

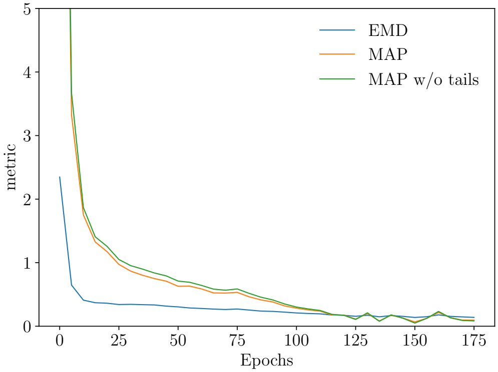
$$\mathbb{E}_{x\sim N(0,1)}[|x|] = \sqrt{2/\pi} \approx 0.80. \tag{4}$$

Models quality

The *Wasserstein distance*, or *Earth Mover's Distance* EMD, is another metric we used to monitor the quality of the models. It measures how different two given histograms are and how much "work" one must do to redistribute one histogram into another. We evaluate this metric to monitor the quality of the training and final models. We use scipy's implementation.

Models quality

- The MAP and the EMD depend on histogram binning.
- 1M events were generated with both NuWro and the GAN, with a E'_μ and θ' binning of 50×50.
- We selected a checkpoint from the training that presents low values of the metrics.



Training performance of GAN

GAN results

In order to compare the production of events of the GAN, 1M NuWro events were produced with energies randomly sampled from a uniform distribution in the 300 MeV-10 GeV range, as well as for different values of neutrino energy within the training range,

$$E_\nu = 500 \text{ MeV}, \ 800 \text{ MeV}, \ 1 \text{ GeV}, \ 2 \text{ GeV}, \ 3 \text{ GeV}, \ 5 \text{ GeV}, \ 7 \text{ GeV}, \ 9 \text{ GeV}.$$

GAN results INC

To test if our GAN model can correctly reproduce the QE and $\Delta(1232)$ resonance peaks, let us now consider event distributions for hadronic invariant mass W

$$W = \sqrt{M_N^2 + 2\omega M_N - Q^2}, \tag{5}$$

where $\omega = E_\nu - E_\mu$, M_N is the average mass of the nucleon, and Q^2 the four-momentum transfer

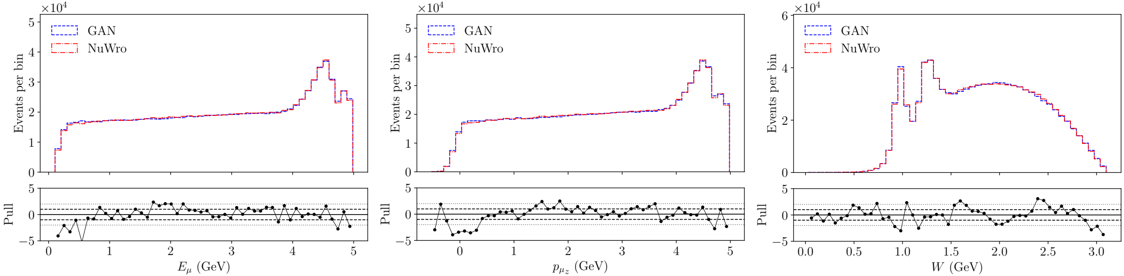
$$Q^2 = 2E_\nu(E_\mu - p_{\mu,z}) - m_\mu^2, \tag{6}$$

so the value of W can be obtained from the muon kinematics alone[2].

[2] If one disregards the Fermi motion of nucleons in the target nucleus and their binding energies effects.

GAN results G-INC 5 GeV

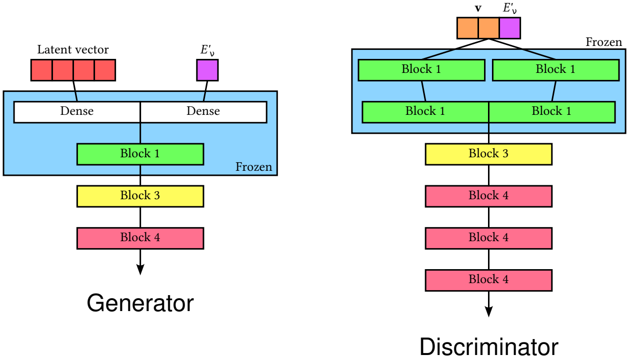
We observe good agreement between the results from NuWro and G-INC for these variables, both in the quasielastic (QE) region and at the peak of the $\Delta(1232)$ resonance. This consistency extends to the onset of deep inelastic (DIS) region at larger W.



Transfer learning

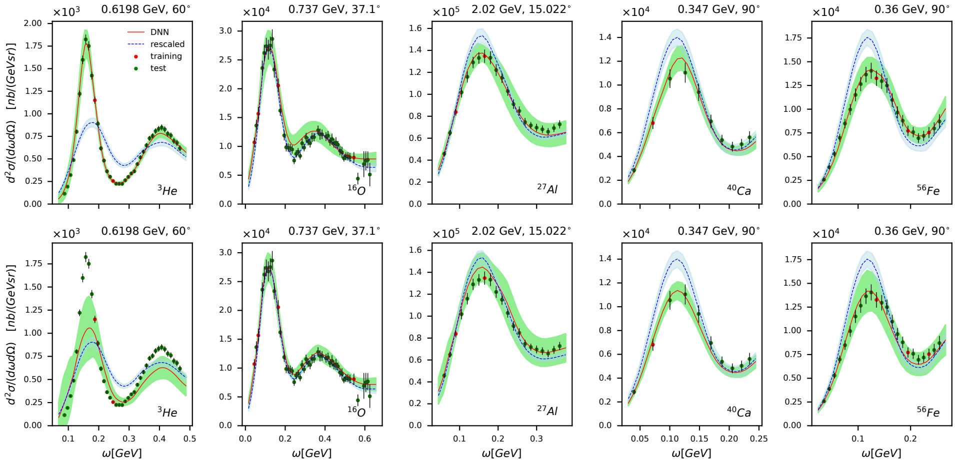
- This technique is useful when data in the target domain is limited, but have a lot of data on a related domain[3].

- We take advantage of the knowledge obtained from the ν_μ-carbon data, and adjust on a different dataset.
- With TL we “freeze” some layers of the base model and let the rest of the model weights update.



[3]Some small tweaks were made to the base network: Phys.Rev.D 113 (2026) 5, 5.

Transfer learning, from e^--carbon scattering $\partial^2\sigma/\partial\omega\partial\Omega$



Phys.Rev.Lett. 135 (2025) 5, 5

Transfer learning, from ν_μ-carbon scattering events

- We adapted the ν_μ-carbon model to 3 different scenarios.
  - neutrino-argon.
  - antineutrino-carbon.
  - alternative neutrino-carbon model (Newer NuWro `25.03`).

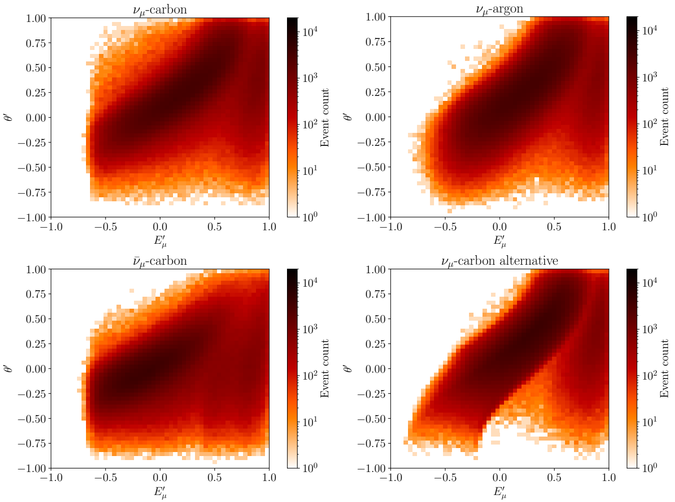
Transfer learning, from ν_μ-carbon scattering events

- We adapted the ν_μ-carbon model to 3 different scenarios.
  - neutrino-argon.
  - antineutrino-carbon.
  - alternative neutrino-carbon model (Newer NuWro `25.03`).
    - ★ Increased axial masses.
    - ★ QE spectral function was changed to Local Fermi Gas.
    - ★ New Delta pion production model

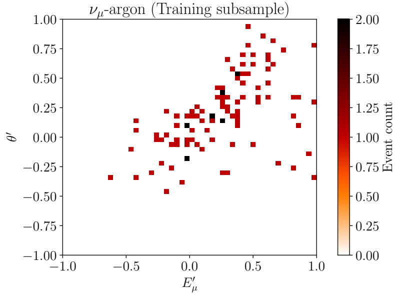
Transfer learning, from ν_μ-carbon scattering events

- We adapted the ν_μ-carbon model to 3 different scenarios.
  - neutrino-argon.
  - antineutrino-carbon.
  - alternative neutrino-carbon model (Newer NuWro `25.03`).
    - Increased axial masses.
    - QE spectral function was changed to Local Fermi Gas.
    - New Delta pion production model
- This time only INC models were considered, with training samples containing 10k and 100k events. Today I'm only showing training on 10k events.
- We also train the models from scratch using the same datasets to compare performance.

Samples domains

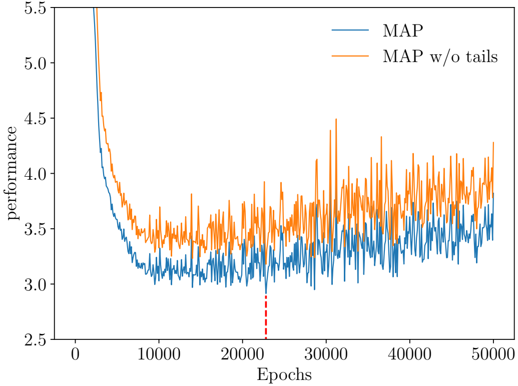


Different sample with alternative underlying physics, cover different but similar domains.

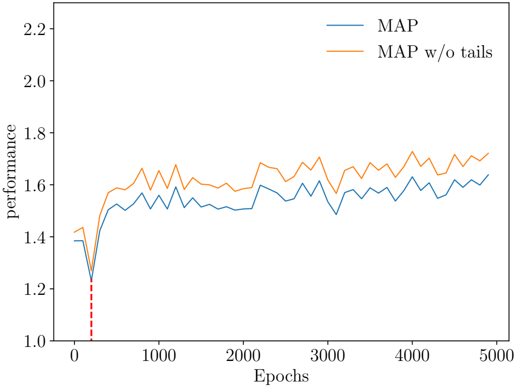


Training data $450 < E_\nu < 550$ MeV

Model convergence

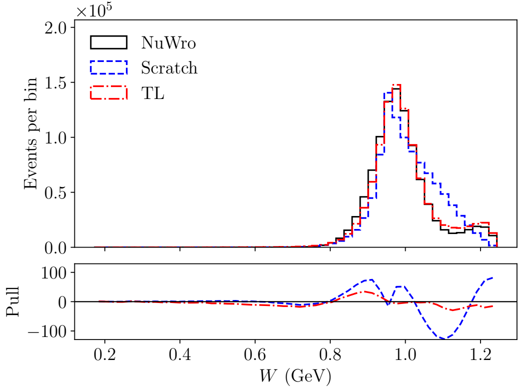


From Scratch

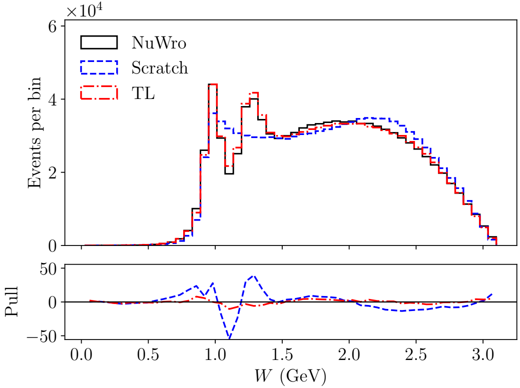


Transfer learning

TL results, from ν_μ-carbon to ν_μ-argon

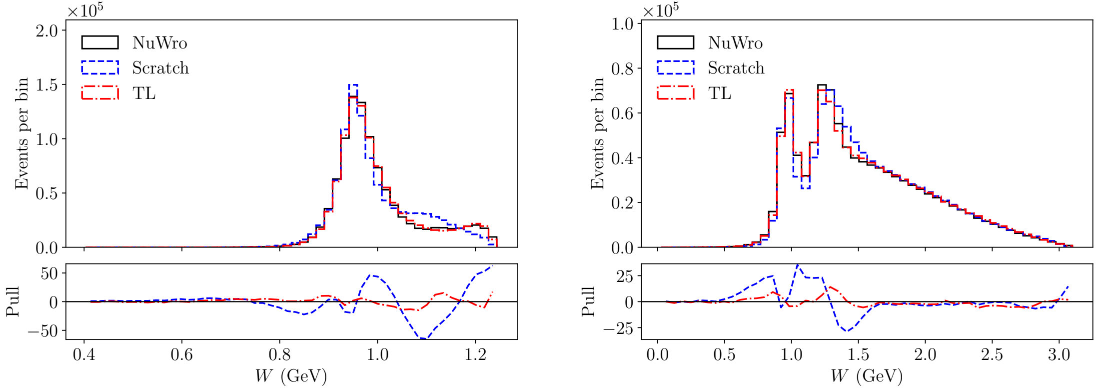


$E_\nu = 500$ MeV



$E_\nu = 5$ GeV

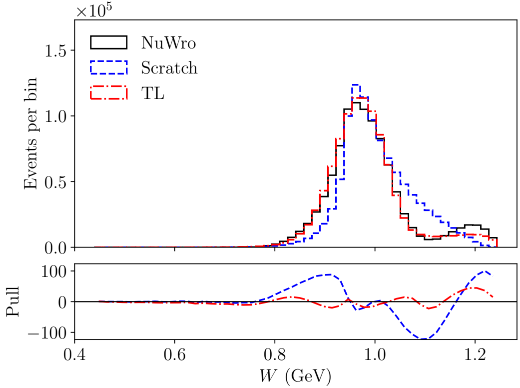
TL results, from ν_μ-carbon to $\bar{\nu}_\mu$-carbon



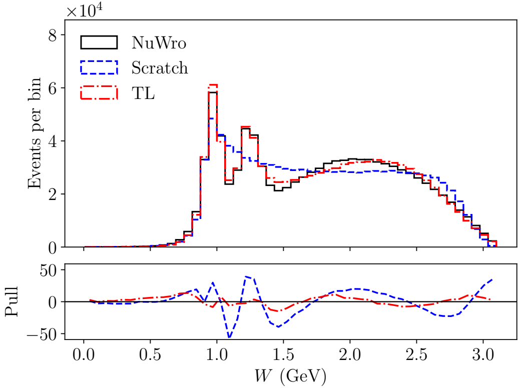
$E_\nu = 500$ MeV

$E_\nu = 5$ GeV

TL results, from ν_μ-carbon to alternative ν_μ-carbon



$E_\nu = 500$ MeV



$E_\nu = 5$ GeV

Summary

- We have discussed the development of GANs models for simulating CC ν_μ-nucleus scattering.
- We consider various kinematic distributions of the charged lepton. The models successfully reproduce the peak structure in data distributions.
- These models can be easily adapted to more realistic scenarios after retraining them on experimental data. They can serve as pre-trained models that can be fine-tuned for specific applications.
- As a bonus, they generate events significantly faster than “classical” generators. The generator model size is < 1 MB!

Thank you!