

Neutrino-Nucleus Interactions from Deep Learning

NPML 2025

Krzysztof M. Graczyk

with Beata Kowal & Luis Bonilla & NuWro Team

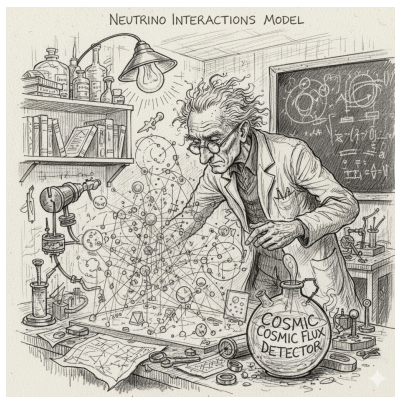


Institute for Theoretical Physics
Poland



Uniwersytet
Wrocławski

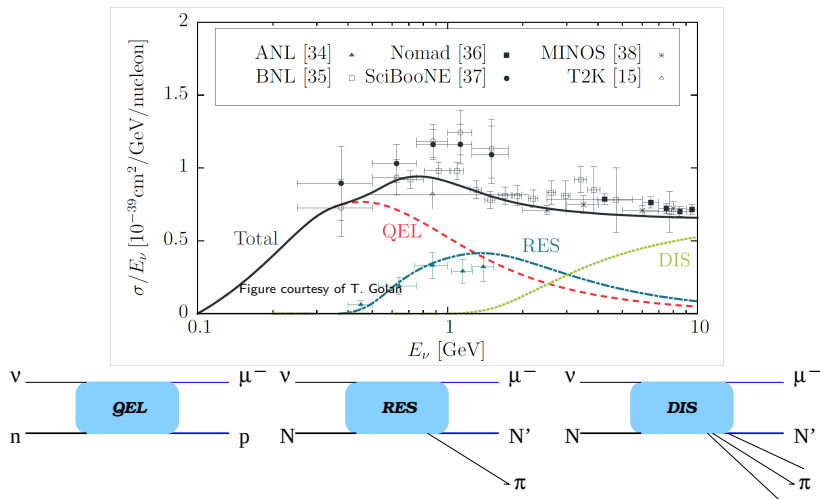
October 29, 2025



Based on:

- GEN ν A: *PRD* 112 (2025) 013007; arXiv:2508.12987
- e A: *PRC* 110 (2024) 025501; *PRL* 135 (2025) 052502; arXiv:2508.00996

- Neutrino oscillation studies
- Goal: predict ν -nucleus interactions with percent-level precision
- **AI-driven models for $(\bar{\nu})\nu$ -nucleus cross sections and collision simulation**
- Build a system that automatically and objectively **updates its physics knowledge as new data and theoretical constraints arrive**



- Neutrino energy is reconstructed mainly from the analysis of quasielastic (QE) scattering events.

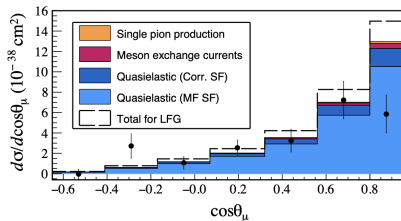
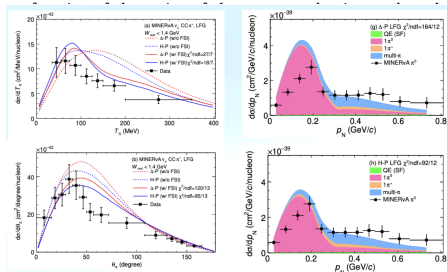


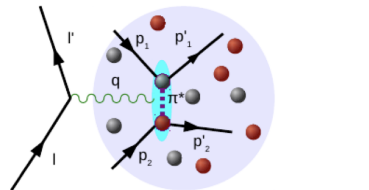
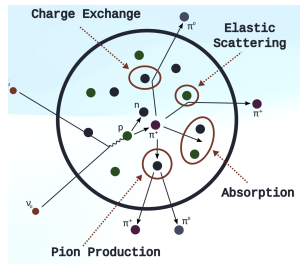
FIG. 2. MicroBooNE CC1p0π differential cross section as



Data from: B. Eberly et al. (MINERvA), Phys. Rev. D92 092008 (2015), arXiv:1406.6415 [hep-ex].

Data from: D. Coploue et al. (MINERvA), Phys. Rev. D 102 072007 (2020), arXiv:2002.05812 [hep-ex].

Figures from Banerjee et al., Phys. Rev. D 109 (2024) 073004.



Figures courtesy of J. Sobczyk.

- Simulate ν -nucleus interactions in realistic conditions:

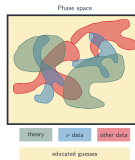
→ Monte Carlo generator of neutrino interactions



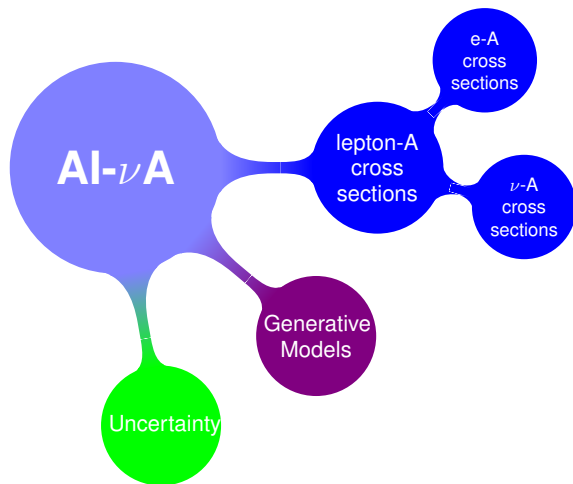
- Build Monte Carlo generators of neutrino interactions

INGREDIENTS:

RECIPE:

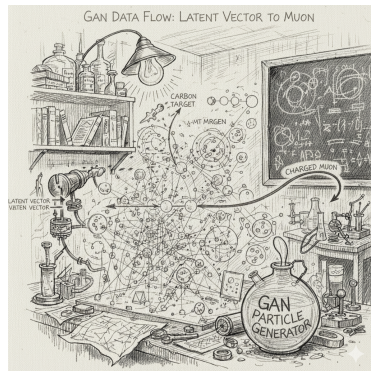


- *The development of Monte Carlo generators can resemble creating Frankenstein's monster: a patchwork of models stitched into a barely consistent whole.*



Use generalization abilities of deep neural networks to make physics

- ~ 1 GeV (anti)neutrino collides with a nucleus:
 - Charged-current (CC) and neutral-current (NC) interactions
 - **Multi-particle final states**
 - Broad energy distribution of initial neutrinos in the beam
 - Candidate models: NF, VAE, **GAN**, Transformers,...
- Start with a simplified, working framework, Bonilla, Graczyk, *et al.* **Phys. Rev. D** 112 (2025) 013007
 - Generative adversarial networks (GANs)
 - CC scattering of muon neutrinos on a carbon target
 - Generate only final muon kinematics for quasielastic (not shown) and **inclusive** contributions
 - Fixed neutrino energy



- Training data from NuWro generator
- 3.88M events uniformly distributed in $E_\nu \in [300 \text{ MeV}, 10 \text{ GeV}]$
- For given E_μ , generate scattering angle θ and muon energy

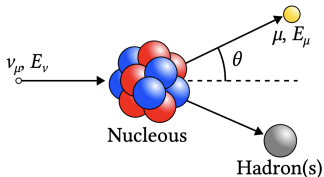
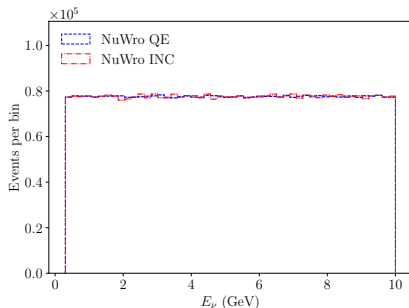


Figure courtesy of Luis.

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



Proxy variables:

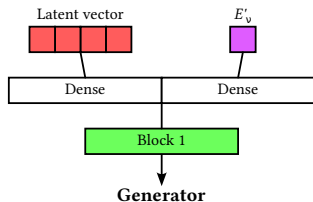
$$E'_\mu = 2\sqrt{1 - \frac{E_\mu - m_\mu}{\Delta E}} - 1,$$

$$\Delta E = E_\nu^- - m_\mu,$$

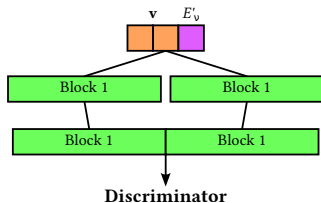
$$\theta' = 2\sqrt{\frac{\theta}{\pi}} - 1,$$

where $E_\nu^- = E_\nu - 9.8 \text{ MeV}$.

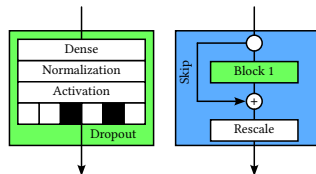
GAN model of ν_μ scattering off ^{12}C : networks



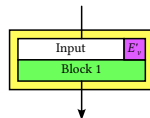
... and three following blocks



... and three following blocks

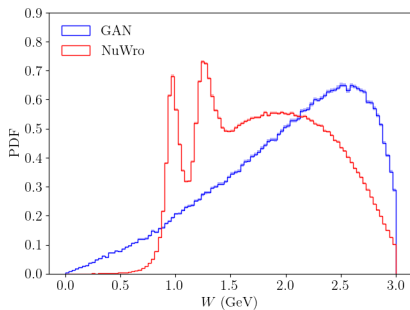
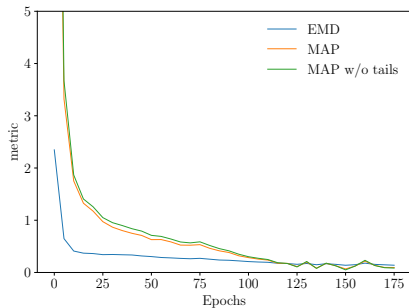
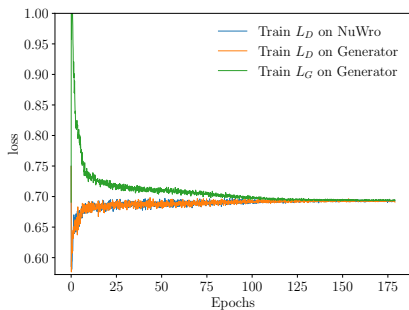


- For each block in the inclusive model, attach the neutrino energy as input.



- Add Gaussian noise layers to mitigate mode collapse.

Training of the GAN on inclusive data



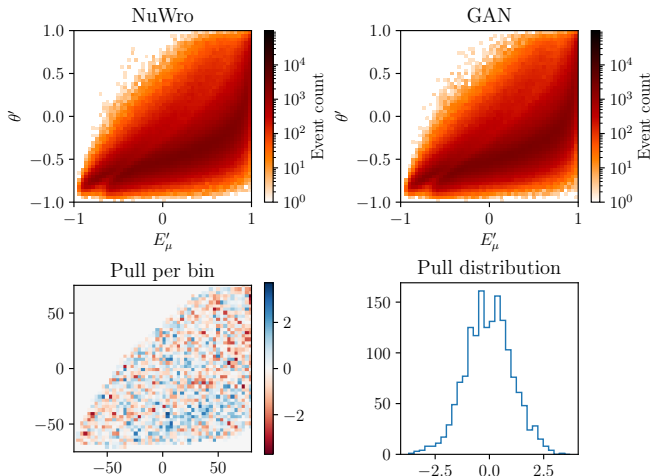
- Mean Averaged Pull (MAP):

$$\text{MAP} = \frac{1}{K} \sum_{i=1}^K |\text{pull}_i|,$$

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

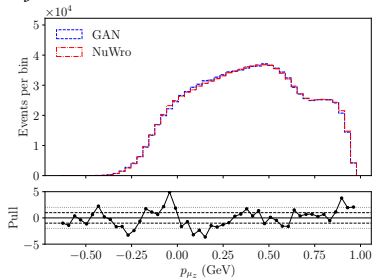
- Earth mover's distance (EMD)

Testing on a dataset uniformly distributed in E_ν

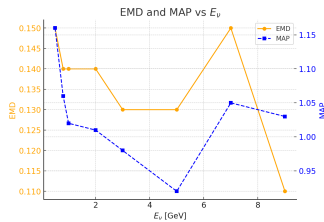
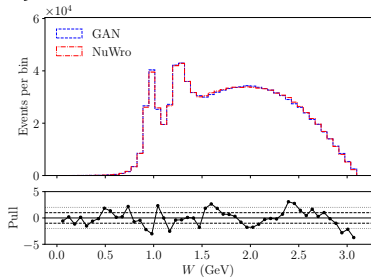


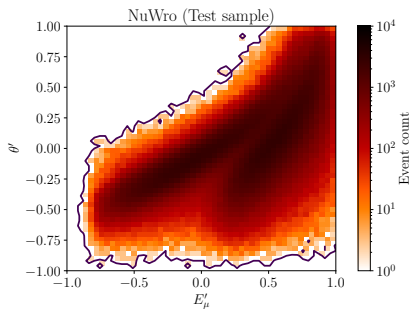
EMD = 0.13, MAP = 0.89 (for NuWro samples: EMD $\sim$ 0.03-0.04, MAP $\sim$ 0.8)

$E_\nu = 1 \text{ GeV}$



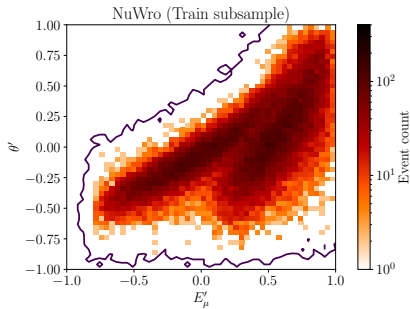
$E_\nu = 5 \text{ GeV}$





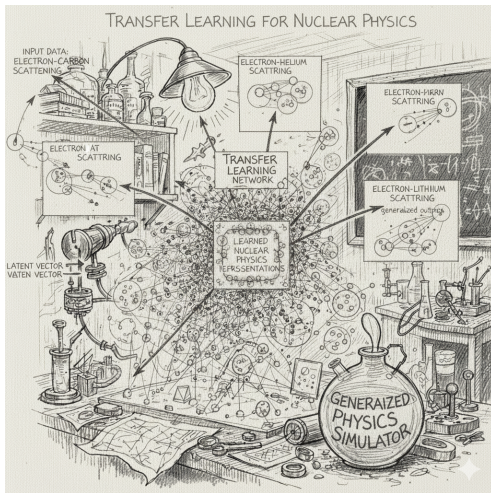
$E_\nu = 1 \text{ GeV}$

- The model extrapolates well.

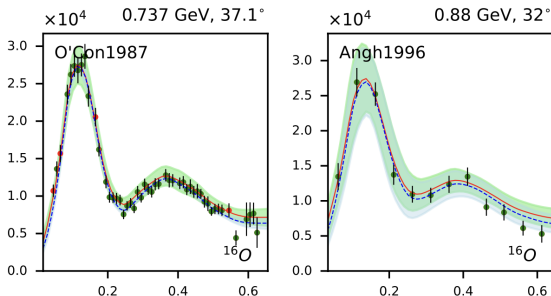


$E_\nu \in [0.95, 1.05] \text{ GeV}$

Teaching networks nuclear physics: transfer learning



- Neutrino energy ~ 1 GeV
- **Similarities between e and ν interactions with nuclear targets**
 - Scarce ν -A data $\leftrightarrow$ informative e -A data
 - Deficiencies in ν -A theory $\leftrightarrow$ well-understood e -A scattering



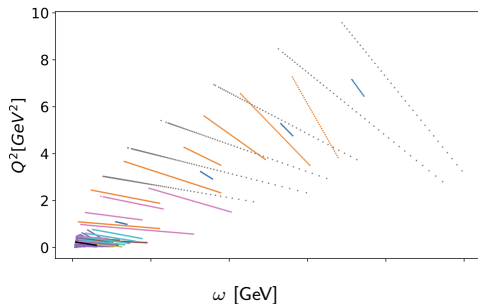
Quasielastic and resonance peaks are clearly distinguished in e -A; not so in ν -A.

Can we teach neural networks nuclear physics?

Step I: teach DNNs nuclear physics

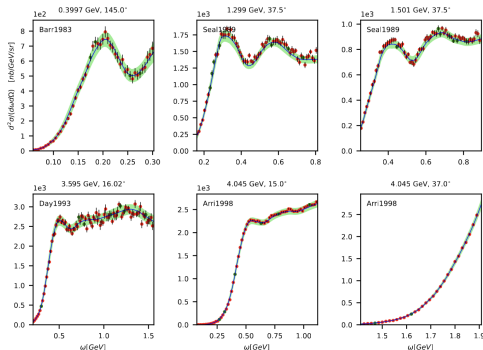
Kowal, Graczyk, *et al.*, *Phys. Rev. C* 110 (2024) 025501

- Inclusive e -C cross-section data
- Broad kinematics: **quasielastic, pion production**, onset of **deep-inelastic scattering**
- **Exclude lowest ω (elastic and discrete nuclear states)**



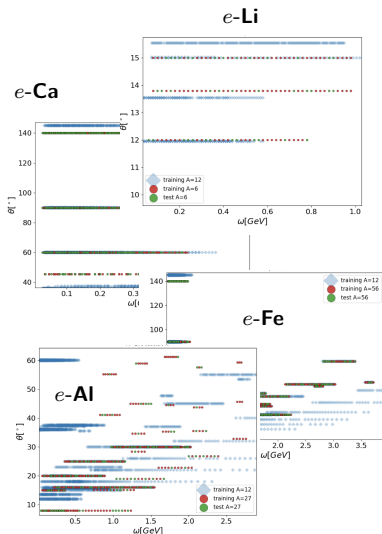


- Predict inclusive e -C cross sections
 - 10 blocks; each block has 300 units and batch normalization
 - **Bootstrap approach:** ensemble of 50 networks
 - **Training to test: 9:1**
- <https://github.com/bekowal/CarbonElectronNeuralNetwork>



Goals

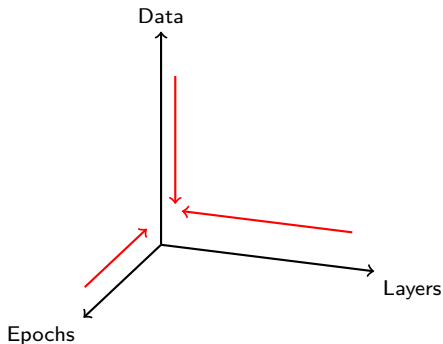
- Use DNN generalization to build an e -A cross-section model
- Transfer nuclear-physics knowledge from one process to another
- **Do DNNs capture universal relationships?**



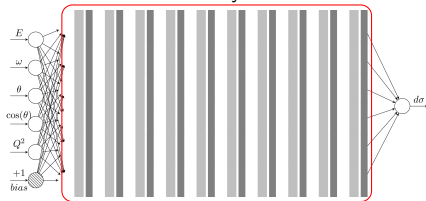
Step II: Have DNNs learned nuclear physics?

Graczyk, Kowal, *et al.* *Phys. Rev. Lett.* 135 (2025) 052502

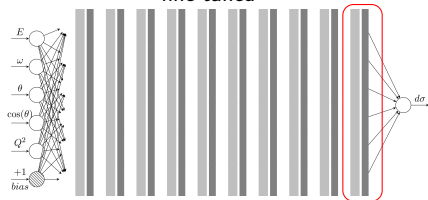
- **Representation learning**: does a DNN encode nuclear-physics relationships?
- Consider e scattering off ${}^6\text{Li}$, ${}^{16}\text{O}$, ${}^{27}\text{Al}$, ${}^{40}\text{Ca}$, ${}^{56}\text{Fe}$, and ${}^3\text{He}$
- **Transfer physics knowledge** from carbon to other targets
- Test strategy:
 - Slight fine-tuning of the carbon DNN
 - Reduce new information
 - How deep should re-optimization go? Layer-wise analysis



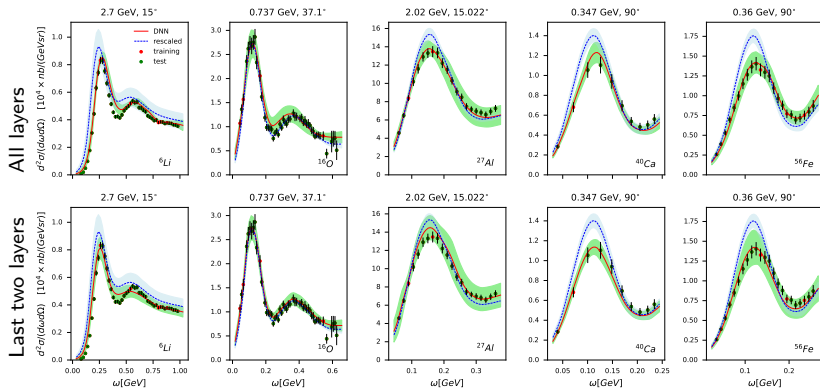
First scenario: all layers fine-tuned



Second scenario: last weight layers fine-tuned

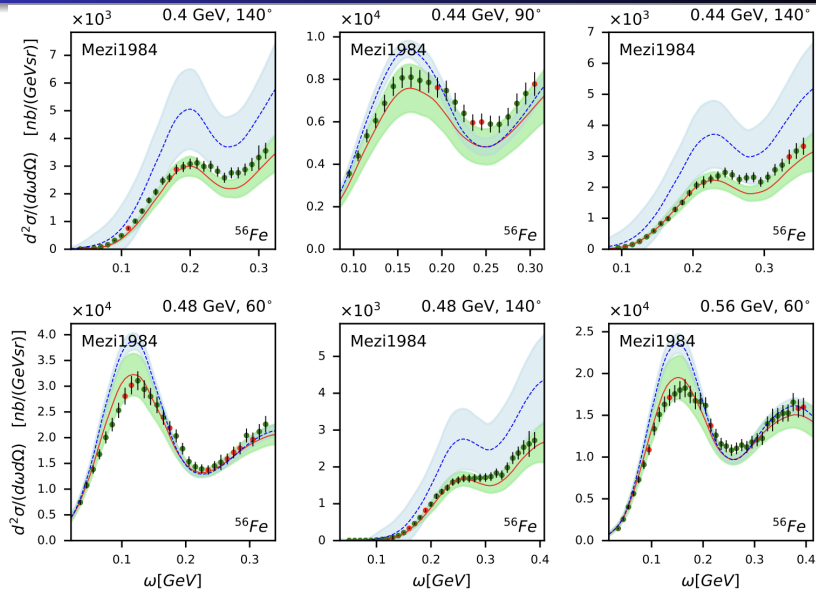


- Use individual fine-tuning for each target.
- Minimize training data to test transfer learning: **training:test = 1:9.**

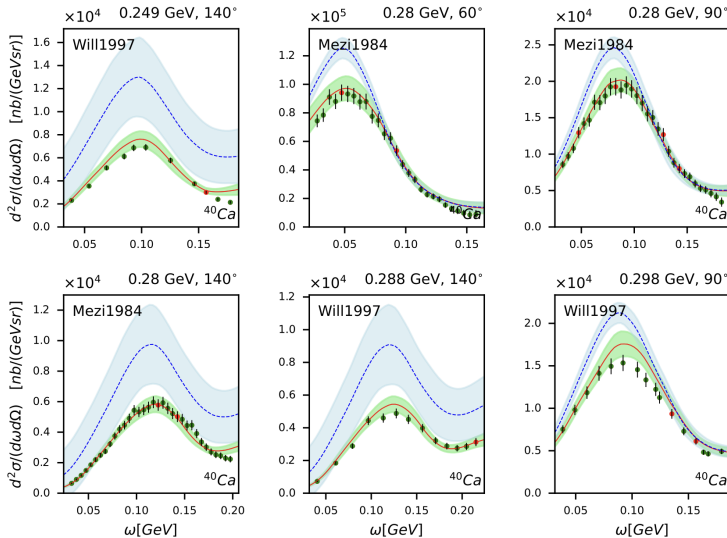


- Kinematics correspond to DUNE, Hyper-K, and T2K configurations.
- Models optimized for 450-1200 epochs; the carbon model for $\sim 40,000$ epochs.

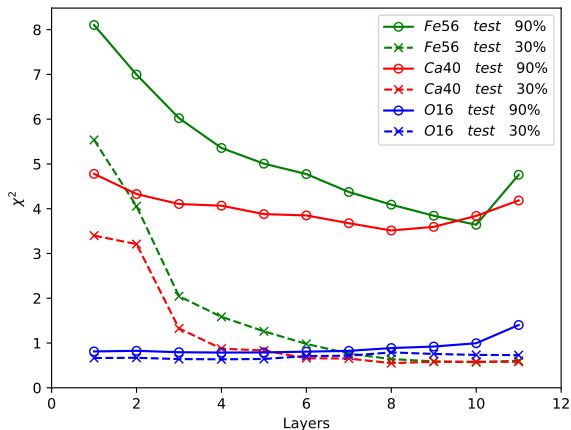
Iron, all layers fine-tuned, training to test = 1:9



- Systematic normalization is taken into account.



$$\chi^2/N \text{ (test)}$$

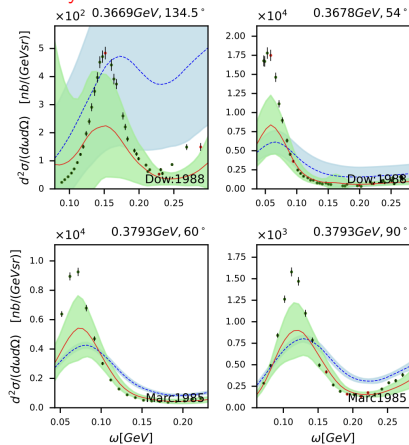


- Fe: deepest optimization ~6-10 layers
- Ca: deep optimization ~4-8 layers
- O: shallow optimization 1-2 layers

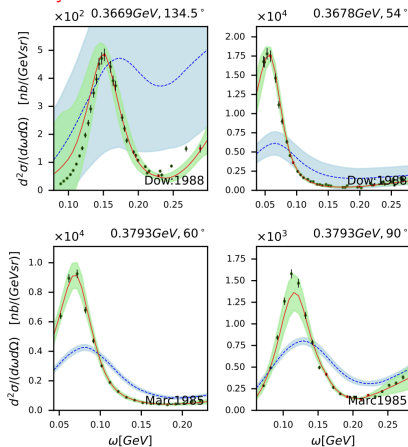
- Layer-wise analysis helps reveal similarities between nuclear targets.

^3He : deep vs. shallow optimization, training to test = 1:9

Two layers

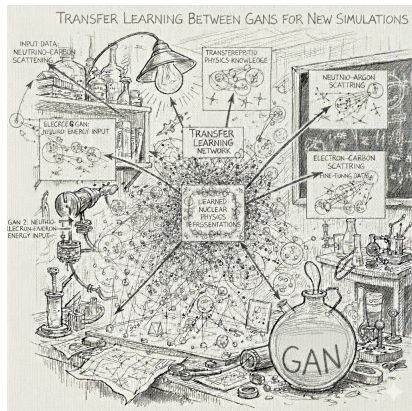
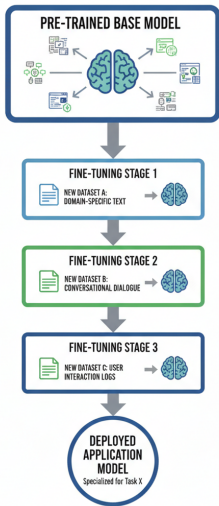


All layers

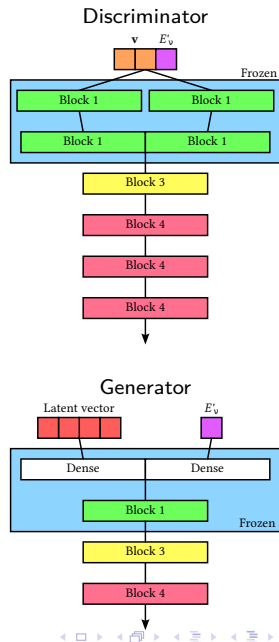


- The method works for ^3He , but full fine-tuning is needed.

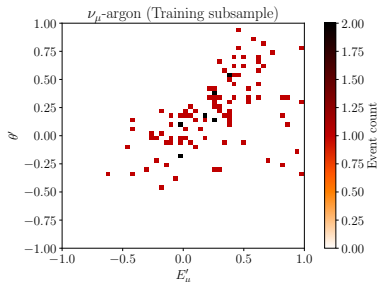
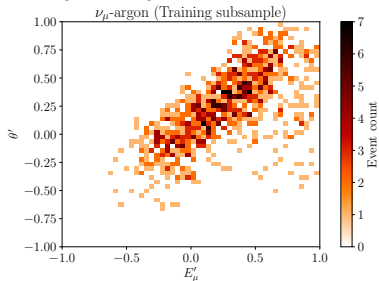
Adapting ν_μ -C GAN via transfer learning



- Bonilla, Graczyk *et al.*,
arXiv:2508.12987
- From a pre-trained ν_μ -C model (4M events; **carbon spectral function**) to:
 - (A) ν_μ -Ar scattering
Argon spectral function; Banerjee *et al.*, *Phys. Rev. D* 109 (2024) 073004
 - (B) $\bar{\nu}_\mu$ -C scattering
Carbon spectral function
 - (C) ν_μ -C scattering with an alternative interaction model
NUWRO v25.03; local Fermi gas;
 $M_A^{\text{QE}} = 1.350 \text{ GeV}$, $M_A^{\text{RES}} = 1.034 \text{ GeV}$; new SPP model

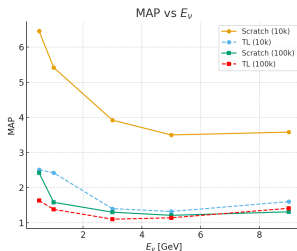
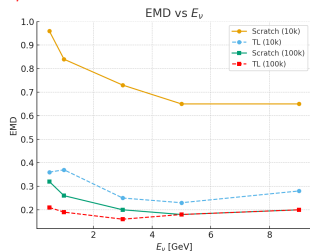


$E_\nu \in [450, 550] \text{ MeV}$

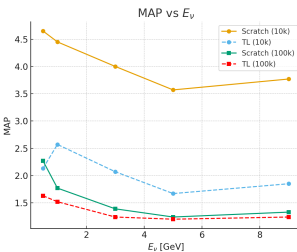
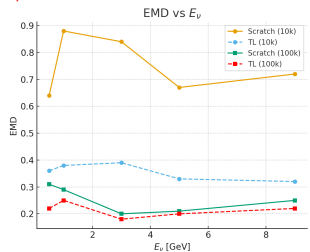


- Compare models trained from scratch and via transfer learning (TL).
- Use 100,000 and 10,000 samples to train/tune the models.

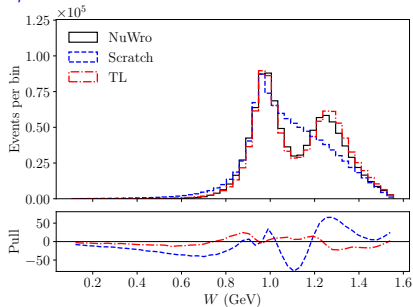
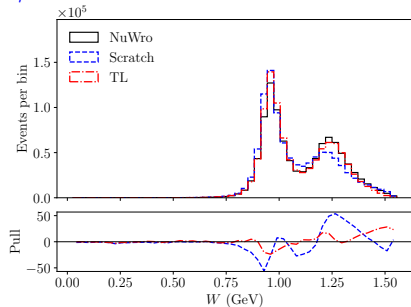
ν_μ -Ar scattering



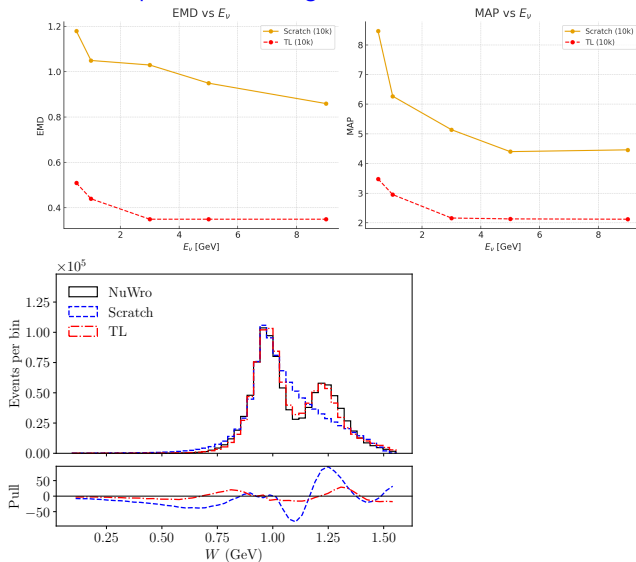
$\bar{\nu}_\mu$ -C scattering



- TL works even with a smaller training set.
- With more data, from-scratch models approach TL performance.

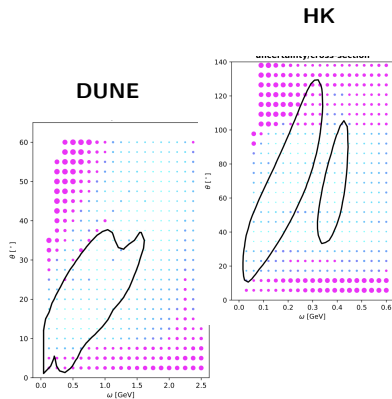
$\nu_\mu\text{-Ar}$  $\bar{\nu}_\mu\text{-C}$ 

10,000 samples in the training dataset



$E_\nu = 1$ GeV

- GANs can simulate QE and inclusive (anti)neutrino scattering.
- Deep neural networks can extract universal properties of the nuclear medium.
- Transfer learning can help to model related processes even with limited information.
- From the e-C model, we can learn how well we know nuclear physics in the kinematic domain probed by Dune and HK,
Kowal, Graczyk, *et al.*,
arXiv:2408.09936

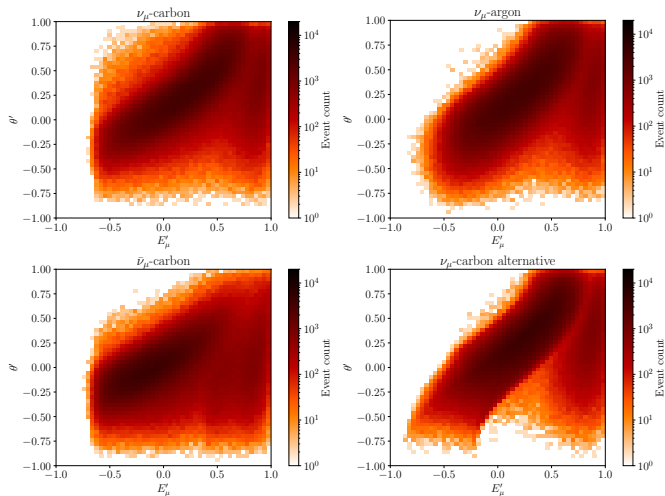


Pink points > 100%

DUNE: 10% for $\omega \in [0.5, 1.5]$ GeV and for $\theta \in [15^\circ, 30^\circ]$

HK: $\sim 20\%$ in wide range

Differences between data for all scenarios, $E_\mu = 500$ MeV



Generating 1 million events on a single-core CPU takes

- approximately 12 minutes with NuWro (full simulation),
- and around 28 seconds with GAN.
- When GAN predictions are made on a GPU (such as RTX 4080), the time is reduced to approximately 3 seconds.