Extracting the $\cos 2\phi$ Asymmetry in the Drell-Yan process using Deep Neural Networks

Dinupa Nawarathne

New Mexico State University
Representing the E906/SeaQuest Collaboration

NPML2023 - August 24, 2023

Table of Contents

Structure of the Proton

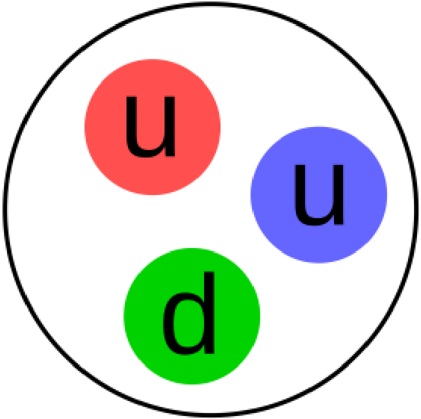


Figure 1: Three valance quarks in proton.

- Protons;
  - Spin 1/2 fermions
  - Composed of three valance quarks: two up (u) quarks and one down (d) quark
  - Quarks are bound together by gluons
  - Properties: mass, charge, spin etc. $\rightarrow$ 3 valance quarks ?

Structure of the Proton

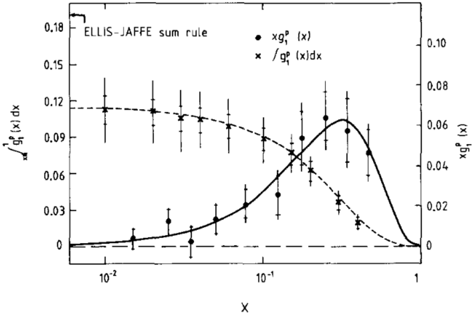


Figure 2: EMC result for the proton spin.[1]

- European Muon Collaboration (EMC):
  - 1st measurement of the total spin of the proton → “Spin Crisis”
  - Quarks contributes only $14 \pm 9 \pm 21\%$ of the proton spin
  - What contributes to the proton spin ?

[1] J. Ashman *et al.*, *Phys. Lett. B* **206**, ed. by V. W. Hughes, C. Cavata, 364 (1988).

Structure of the Proton

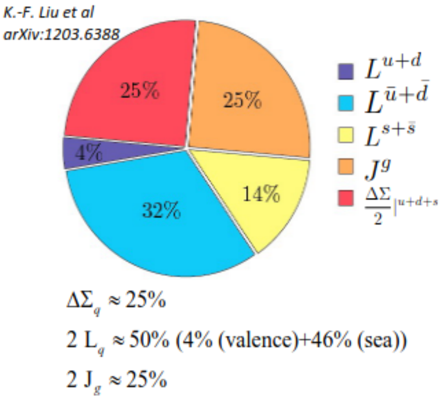


Figure 3: Spin Decomposition according to lattice QCD.[2]

$$\frac{1}{2} = \frac{1}{2}\Delta\Sigma + \Delta G + L_q + L_{\bar{q}} + L_G$$

$\frac{1}{2}\Delta\Sigma$ - contribution of the intrinsic spin of the quarks and anti-quarks

ΔG - contribution of the intrinsic spin of the gluons

L_q, $L_{\bar{q}}$, L_G - contributions of the orbital angular momentum of the valence quarks, sea quarks, and gluons, respectively

[2] K.-F. Liu, *AAPPS Bull.* **32**, 8, arXiv: 2112.08416 (hep-lat) (2022).

Structure of the Proton

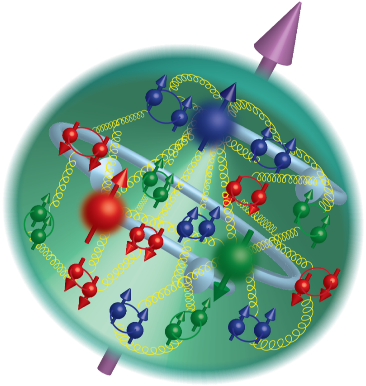


Figure 4: Dynamic structure of the proton.

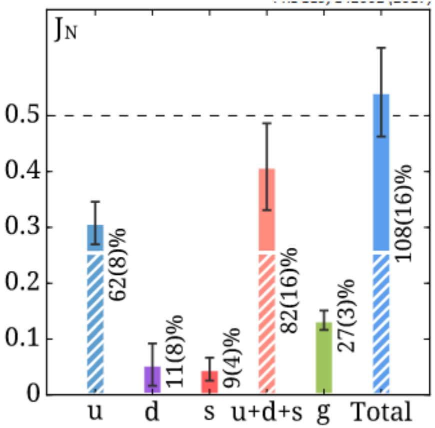


Figure 5: proton spin decomposition in terms of the angular momentum.[3]

[3] K.-F. Liu, *AAPPS Bull.* **32**, 8, arXiv: 2112.08416 (hep-lat) (2022).

Structure of the Proton

- Transverse momentum distributions (TMDs): distributions of the hadron's quark or gluon momenta that are perpendicular to the momentum transfer between the beam and the hadron

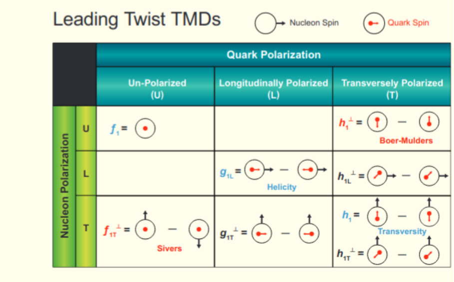


Figure 6: TMDs classification according to the polarization of the quarks and nucleon.[4]

- Provide information on the confined motion of quarks and gluons inside the hadron and complement the information on the hadron structure.
- Boer-Mulders (BM) function: transverse-polarization asymmetry of quarks within an unpolarized hadron

[4] A. Accardi *et al.*, *Eur. Phys. J. A* **52**, ed. by A. Deshpande *et al.*, 268, arXiv: 1212.1701 (nucl-ex) (2016).

Drell-Yan Process

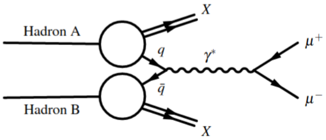


Figure 8: Drell-Yan process.[6]

$$\frac{d\sigma}{d\Omega} \propto 1 + \lambda \cos^2\theta + \mu \sin 2\theta \cos\phi + \frac{1}{2}\nu \sin^2\theta \cos 2\phi$$

- Drell-Yan process → probing the internal structure of hadrons
- Extraction of ν parameter → BM function

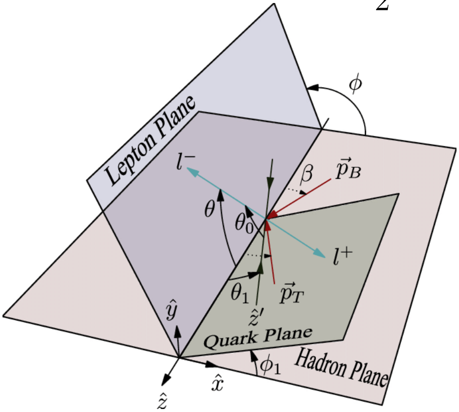


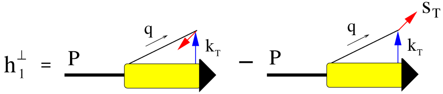
Figure 7: Diagram of the Collins-Soper frame.[5]

[5] J.-C. Peng *et al.*, *Phys. Lett. B* **789**, 356–359, arXiv: 1808.04398 (hep-ph) (2019).

[6] K. Nagai, PhD thesis, Tokyo Inst. Tech, 2017.

Boer-Mulders Function ($h_1^\perp$)

$$h_1^{\perp[C]}(x, k_T^2)\epsilon_T^{ij}k_{T_j} = \frac{M}{2}F.T.\langle P|\bar{\psi}(0)\mathcal{L}_c(0,\varepsilon)\gamma^i\gamma^\dagger\gamma_5\psi(\varepsilon)|P\rangle$$



- Describes the net polarization of quarks inside an unpolarized proton

- Quarks can be polarized on average even inside an unpolarized proton, as long as they are not moving exactly along the proton direction.
- If $h_1^\perp \neq 0 \rightarrow$ then it reflects the presence of a handedness inside the proton $P \cdot (k_T \times s_T)$
- $h_1^\perp \rightarrow$ quark distribution that quantifies a particular spin-orbit correlation

Evidence for Non-zero BM Function

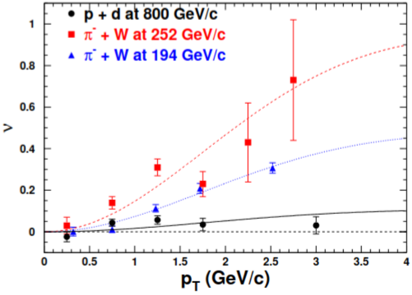


Figure 9: Parameter ν vs. p_T for Drell-Yan process in E866 experiment.[7]

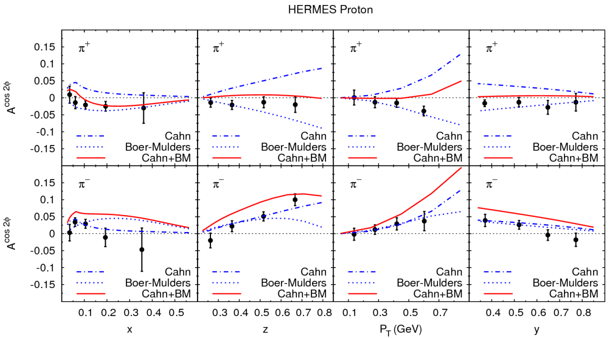


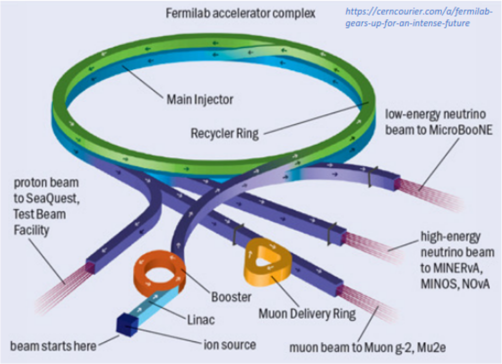
Figure 10: HERMES proton-target data for SIDIS process.[8]

[7] L. Y. Zhu *et al.*, *Phys. Rev. Lett.* **99**, 082301, arXiv: hep-ex/0609005 (2007).

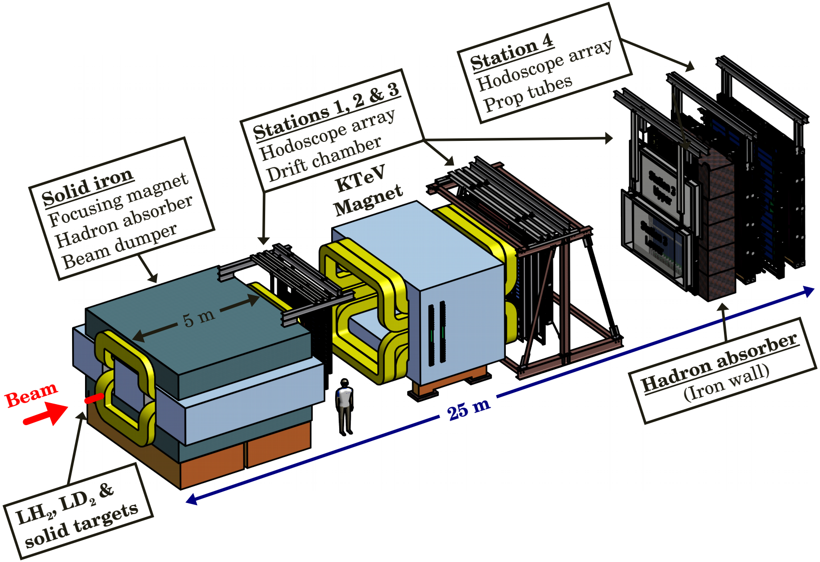
[8] V. Barone *et al.*, *Phys. Rev. D* **81**, 114026, arXiv: 0912.5194 (hep-ph) (2010).

SeaQuest/E906 Experiment

- Fixed target Drell-Yan experiment at Fermilab
- Use 120GeV beam energy from the main injector
- Measure the antiquark structure of the nucleon
- Provides unique access to the vanishing sea quark density at high x
- Data collection was concluded in 2017



SeaQuest/E906 Experiment



[9] C. A. Aidala *et al.*, *Nucl. Instrum. Meth. A* **930**, 49–63, arXiv: 1706.09990 (physics.ins-det) (2019).

Figure 11: SeaQuest/E906 spectrometer.[9]

SeaQuest/E906 Experiment

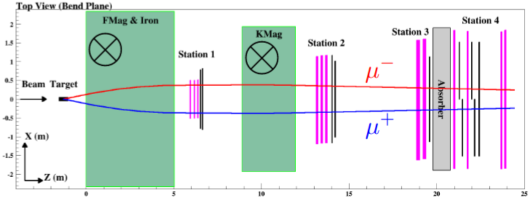


Figure 12: Top view of the detector and muon pair track.[10]

Trigger Name	Target	Hodoscope side	P_t cut	Prescale factor
FPGA-1	$\mu^+\mu^-$	TB or BT	—	1
FPGA-2	$\mu^+\mu^-$	TT or BB	—	10000
FPGA-3	$\mu^+\mu^+$ or $\mu^-\mu^-$	TB or BT	—	123
FPGA-4	μ^+ or μ^-	T or B	—	25461
FPGA-5	μ^+ or μ^-	T or B	> 3 GeV	2427
NIM-1	μ^+ or μ^-	T or B	—	31991
NIM-3	—	—	—	125

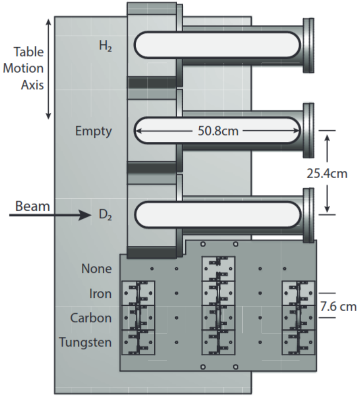
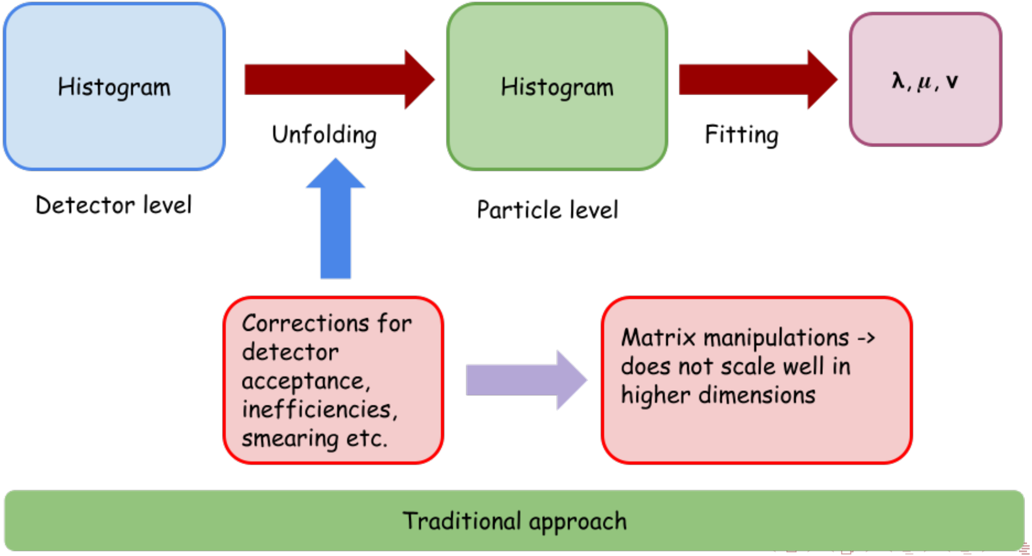


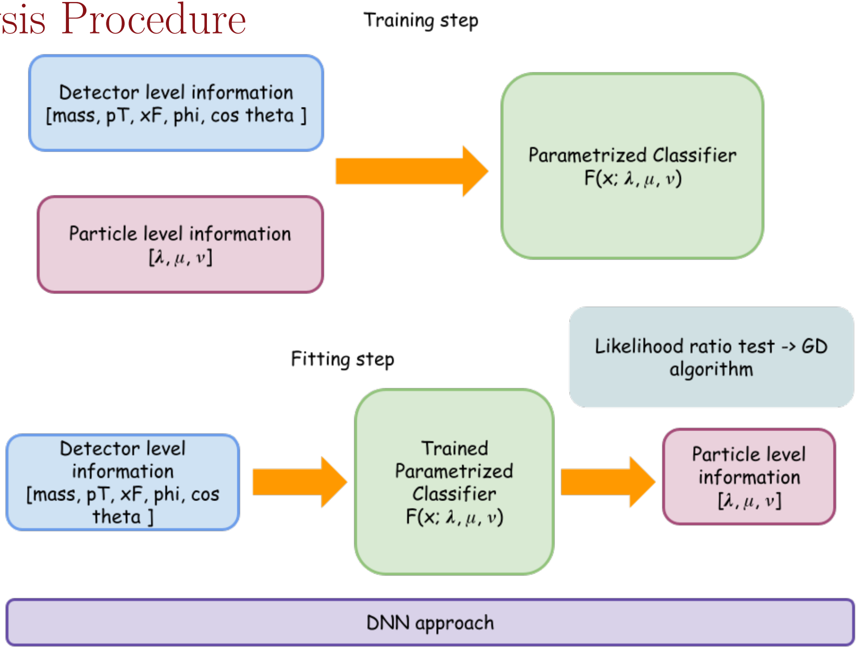
Figure 13: Targets used in the SeaQuest experiment.

[10] K. Nagai, PhD thesis, Tokyo Inst. Tech, 2017.

Analysis Procedure



Analysis Procedure



Likelihood Ratio Test

- The likelihood ratio test is a highly effective method for assessing the goodness of fit.
- Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a distribution with a parameter θ. Suppose that we have observed $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$. We define the the likelihood function as the joint probability of the observed samples as a function of θ;

$$\mathcal{L}(x_1, x_2, \ldots, x_n; \theta) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n; \theta)$$

- To decide between two simple hypotheses $H_0 : \theta = \theta_0$ and $H_1 : \theta = \theta_1$, we define the likelihood ratio:

$$\lambda(x_1, x_2, \ldots, x_n) = \frac{\mathcal{L}(x_1, x_2, \ldots, x_n; \theta_0)}{\mathcal{L}(x_1, x_2, \ldots, x_n; \theta_1)}$$

Likelihood Ratio Test

- To perform a likelihood ratio test, we choose a constant c. We reject H_0 if $\lambda < c$ and accept it if $\lambda \geq c$. The value of c can be chosen based on the desired significance level α.
- Neural networks excel at approximating non-linear functions, making them ideal for use as higher dimensional likelihood functions.
- Our goal is to train the neural network to classify samples accurately. Specifically, we aim to classify samples $\omega_{0i} \in \Omega_0$ as $y = 0$ and $\omega_{1i} \in \Omega_1$ as $y = 1$, regardless of the parameter θ.
- Subsequently, we can utilize the trained neural network to estimate any unknown parameter θ_{unknown} by employing the gradient descent algorithm.[11]

[11] A. Andreassen, B. Nachman, *Phys. Rev. D* **101**, 091901, arXiv: 1907.08209 (hep-ph) (2020).

SeaQuest MC Data Generation

- We generated the Monte Carlo (MC) data using the PYTHIA generator.
- The generated events were then passed through the E906 detector simulation (using GEANT4) to obtain the reconstructed detector information.
- We sample the values of λ, μ and ν uniformly from the ranges of (0.5, 1.5), (−0.5, 0.5), and (−0.5, 0.5), respectively.[12]

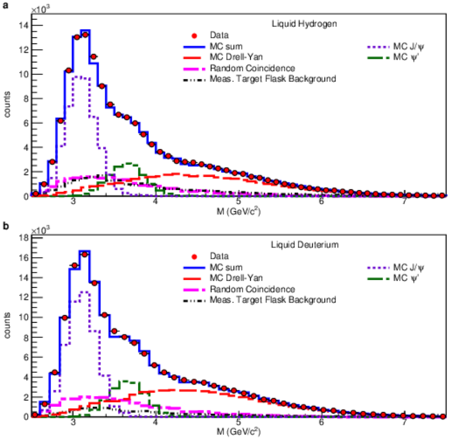


Figure 14: SeaQuest MC data is in good agreement with real data.[13]

[12] L. Y. Zhu *et al.*, *Phys. Rev. Lett.* **99**, 082301, arXiv: hep-ex/0609005 (2007).

[13] J. Dove *et al.*, *Nature* **590**, [Erratum: Nature 604, E26 (2022)], 561–565, arXiv: 2103.04024 (hep-ph) (2021).

Deep Neural Network Architecture

- The neural network consists of five hidden linear layers, each containing 64 nodes. The ReLU function is used to activate the hidden layers, along with batch normalization layers. The final output is passed through a Sigmoid activation function.
- During the training step, we use the following input features for the neural network: mass, p_T, x_F, ϕ, $\cos\theta$, λ, μ, and ν.
- The neural network was trained for 200 epochs, employing early stopping with a patience of 20, to minimize the binary cross-entropy loss.
- During the fitting step, we freeze all the weights and biases of the trained neural network. Then, we employ the gradient descent algorithm to determine the optimal values of λ, μ, and ν by minimizing the loss.

Fitting Step

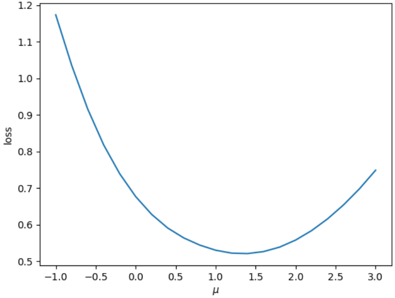


Figure 15: In the fitting step, the loss reaches its minimum at the optimal value.

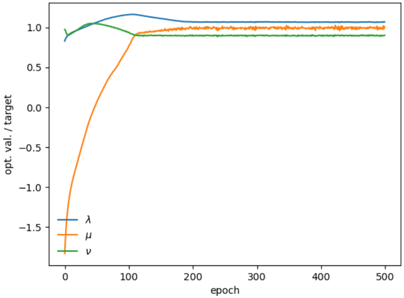


Figure 16: During the fitting step, all three parameters reach the optimal value.

Testing DNN approach

Combination	Coefficient	Injected	Fitted
1	λ	0.84	0.876 ± 0.208
	μ	0.26	0.234 ± 0.054
	ν	-0.34	-0.299 ± 0.052
2	λ	1.33	1.134 ± 0.151
	μ	0.17	0.146 ± 0.050
	ν	-0.34	-0.281 ± 0.043
3	λ	1.12	1.242 ± 0.181
	μ	-0.27	-0.211 ± 0.088
	ν	-0.24	-0.236 ± 0.071

Table 1: Table showing the mean and standard deviation of fitted values of λ, μ, and ν using the gradient descent algorithm with different model initialization.

- Independent events from training data $\rightarrow$ to reduce the biases
- 3 test sample $\rightarrow$ extract the injected parameters within a ± 1.5 standard deviation (σ) interval

Combinatorial Background

- Understanding the background from the experiment data is really important
- Full background simulations $\rightarrow$ computationally expensive
- Variational autoencoders;
  - Generative model $\rightarrow$ can generate new events based on the trained events
  - Computation is fast $\rightarrow$ use of GPUs
  - Can use a higher-dimension inputs
  - Control over the reconstruction error can be achieved using KL divergence
  - Trained VAE $\rightarrow$ background subtracted events $\rightarrow$ un-binned method

Combinatorial Background

- We use an event-mixing method to estimate the combinatoric background from the SeaQuest data.[14]
- To remove the combinatorial background, we employ histogram subtraction → binned method, which does not scale well in higher dimensions

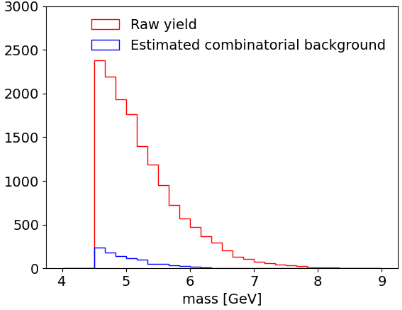


Figure 17: Mix and un-mixed events.

- Our goal is to utilize Variational Autoencoders (VAEs) for generating background-subtracted distributions.[15]
- VAE generated distribution → fitting algorithm → λ, μ, ν

[14] S. F. Pate *et al.*, arXiv: 2302.04152 (hep-ex) (Feb. 2023).

[15] D. P. Kingma, M. Welling, *arXiv preprint arXiv:1312.6114* (2013).

VAEs

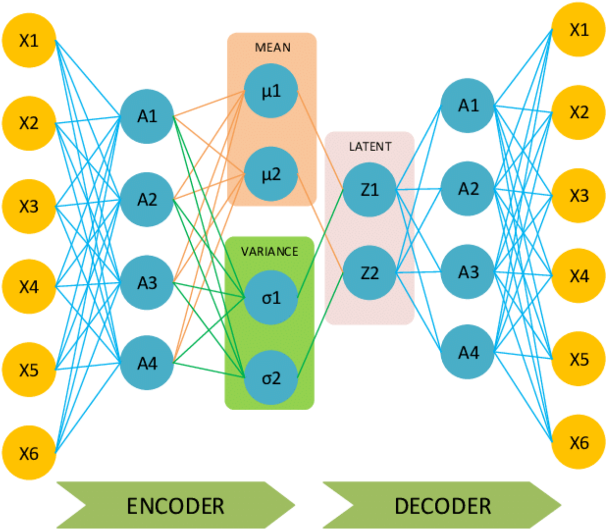
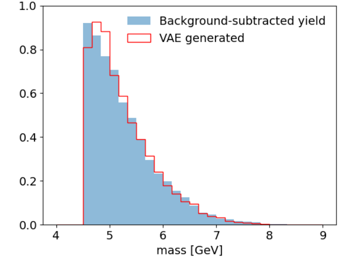
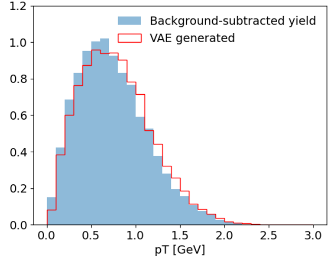


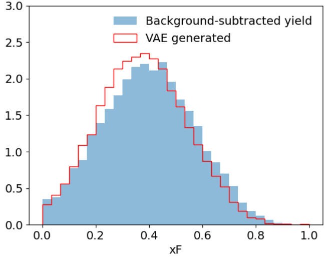
Figure 18: Example of VAE network.

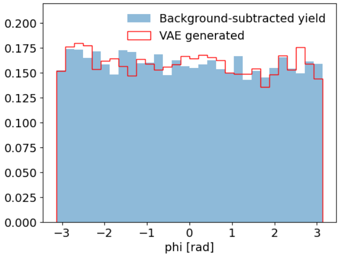
- Enforce the latent space to be Gaussian-like
- Generate noise vector in latent space → decode to generate sample
- Inputs: mass, p_T, x_F, ϕ, $\cos\theta$
- Both the encoder and decoder have 3 hidden layers, each containing 64 nodes activated by the ReLU activation function.
- The latent dimension is 3.

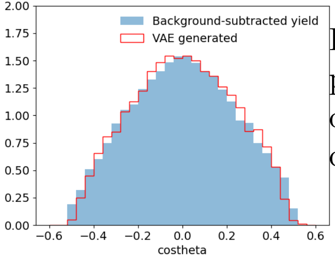
VAE Generated Events











Early result → enhance the prediction accuracy using diffusion models/ conditional VAEs.

Future Trajectory

- Systematic study of the fitting algorithm to better understand the phase space variables.
- Increase the precision of the fitting algorithm using VAE/GAN.[16]
- Increase the accuracy of the prediction for background-subtracted events with Diffusion/CVAE models.

[16] S. Diefenbacher *et al.*, *JINST* **15**, P11004, arXiv: 2009.03796 (hep-ph) (2020).

Summary

- Spin of the proton $\rightarrow$ intrinsic property $\rightarrow$ explain the structure of the proton
- BM function $\rightarrow$ transverse-polarization asymmetry of quarks within an unpolarized hadron
- Neural networks $\rightarrow$ multi-dimensional likelihood functions $\rightarrow$ likelihood ratio test to extract the optimal parameters for the Drell-Yan angular distribution.
- VAEs $\rightarrow$ generate distributions with background removed.
- Our plan is to use this high-dimensional fitting algorithm with VAE generated events to extract the Drell-Yan angular coefficients from the E906/SeaQuest data with higher accuracy.
- Acknowledgement: This work was funded by the DOE office of Science, Medium-Energy Nuclear Physics Program.