# AI/ML for Semi-Inclusive and Exclusive DIS Measurements

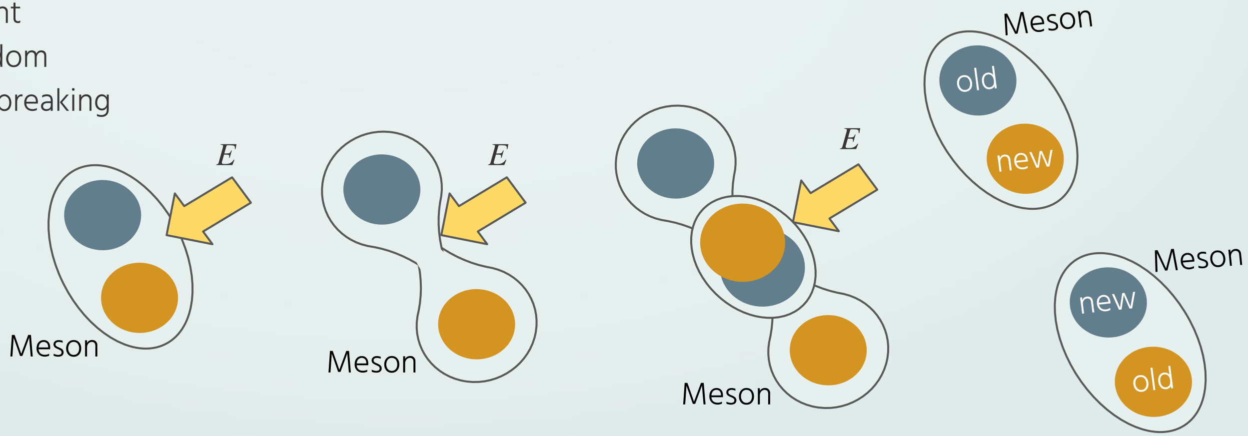
Stjepan Oresic
University of Regina
Supervisor: Dr. Zisis Papandreou

EPIC Collaboration

University of Regina | Faculty of Science

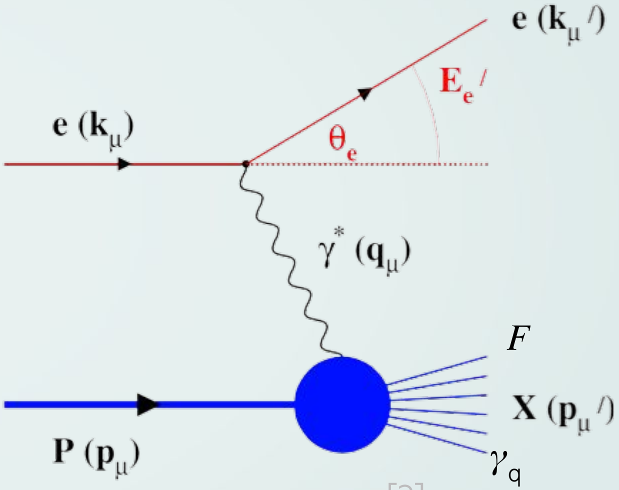
# Quantum Chromodynamics (QCD)

- Theory of the strong interaction between quarks mediated by gluons
- Quantum field theory called a non-abelian gauge theory with symmetry group SU(3)
- Large body of **experimental evidence** for QCD has been gathered over the years
- Three major properties
  - Color confinement
  - Asymptotic freedom
  - Chiral symmetry breaking



# Deep-Inelastic Scattering (DIS)

- QCD framework → structure of hadrons
- Protons are made out of gluons and quarks
    - Proven by probing a proton with a virtual photon at high energies
- **Deep Inelastic Scattering (DIS) processes**
- **Bjorken scaling**
- Unique points in ($x$,$Q^2$) plane → 3 different reconstruction methods




[2]

$$x = \frac{Q^2}{2\vec{P}\bullet\vec{q}} \qquad y = \frac{\vec{p}\bullet\vec{q}}{\vec{p}\bullet\vec{k}}$$

$$Q^2 = -\vec{q}\bullet\vec{q} = -(\vec{k}\bullet\vec{k'})^2$$

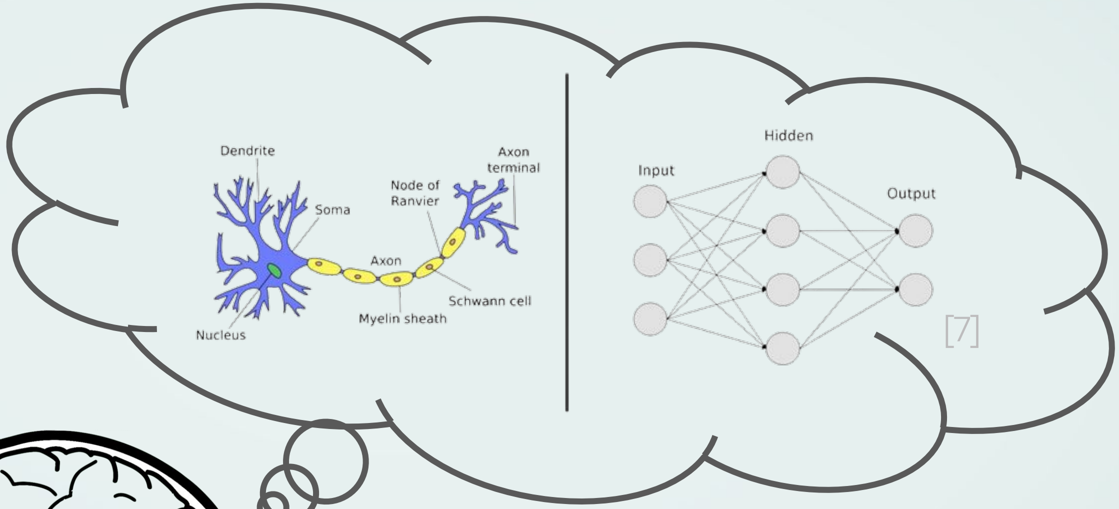
# Deep-Inelastic Scattering (DIS)

- Radiative corrections
    - Large contributions of initial state radiation from electron expected
    - Electron radiates a $\gamma$ before interacting → Energy degraded
- **Choice of the reconstruction method determines the size of systematic uncertainty**
- Which method is the best and is there a workaround?

# DIS RECONSTRUCTION METHODS

| Method | Advantages | Disadvantages |
| --- | --- | --- |
| **Electron Method (EL)** | $Q^2$ very precise | Accuracy of x poor for low y<br><br>Sensitive to QED radiation |
| **Jacques-Blondel Method (JB)** | Jet algorithms not needed (still can be used)<br><br>x more accurate especially at low y | Needs precise jet energy measurements |
| **Double Angle Method (DA)** | Large parts of phase space superior to other methods<br><br>No need for precise jet energy measurements | Suffers from same problems as electron method |

# Neural Networks - General

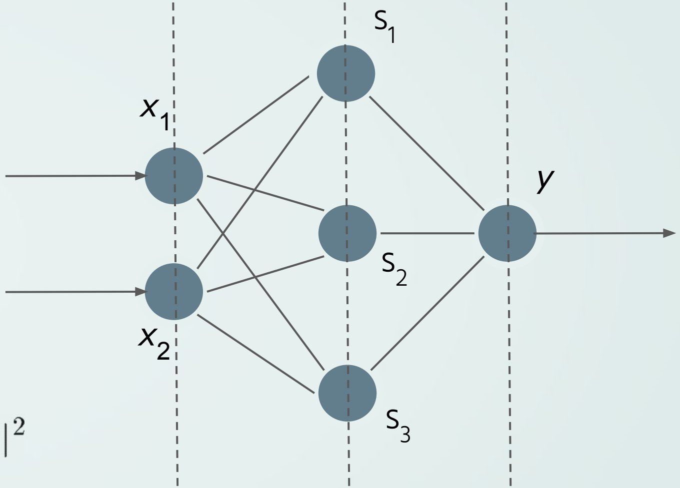



[7]

[6]

- Inspiration comes from the structure of the brain
- If our brain is so powerful and efficient → why not try to use that framework?
- Axon = output of a neuron → transmits the signal to other neurons

# Neural Networks - General




- Mapping from $\mathbb{R}^2$ to $\mathbb{R}^1$ with transformations in between
- How many nodes do we need?
- Function $f$ in the hidden layer applied componentwise
- Supervised learning → **minimize prediction and result difference** (weight matrix $\mathbf{W}$)
- In case we know true value → compute error $E(W,\widehat{W}) = \sum_{i=1}^{p} ||t_k - y_k||^2$
- **Neural network is at its best when this error term is minimized**

$$\begin{bmatrix} s_{1k} \\ s_{2k} \\ s_{3k} \end{bmatrix} = f(\begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \end{bmatrix} \begin{bmatrix} x_{1k} \\ x_{2k} \end{bmatrix}) \longrightarrow \begin{bmatrix} s_{1k} \\ s_{2k} \\ s_{3k} \end{bmatrix} = f(\begin{bmatrix} w_{11}x_{1k} + w_{12}x_{2k} \\ w_{21}x_{1k} + w_{22}x_{2k} \\ w_{31}x_{1k} + w_{32}x_{2k} \end{bmatrix})$$

# Neural Networks - General

- Other half gives:

$$y_k = \widehat{w}_1 s_{1k} + \widehat{w}_2 s_{2k} + \widehat{w}_3 s_{3k}$$

- Combined:

$$y_k = \widehat{w}_1 f(w_{11}x_{1k} + w_{12}x_{2k}) + \widehat{w}_2 f(w_{21}x_{1k} + w_{22}x_{2k}) + \widehat{w}_3 f(w_{31}x_{1k} + w_{32}x_{2k})$$

- Relationship between $\mathbf{y_k}$ and 9 variables
- Gradient descent

$$\widehat{w}_{ij}^{NEW} = \widehat{w}_{ij}^{OLD} - \alpha \frac{\partial E}{\partial w_{ij}^{OLD}}$$

- **Update $\mathbf{W}$ based on new observations**
- **Retrace steps!** → computer does it for us

# Neural Networks - General

- Other half gives:

$$y_k = \widehat{w}_1 s_{1k} + \widehat{w}_2 s_{2k} + \widehat{w}_3 s_{3k}$$

- Combined:

$$y_k = \widehat{w}_1 f(w_{11}x_{1k} + w_{12}x_{2k}) + \widehat{w}_2 f(w_{21}x_{1k} + w_{22}x_{2k}) + \widehat{w}_3 f(w_{31}x_{1k} + w_{32}x_{2k})$$

- Relationship between $\mathbf{y_k}$ and 9 variables
- Gradient descent

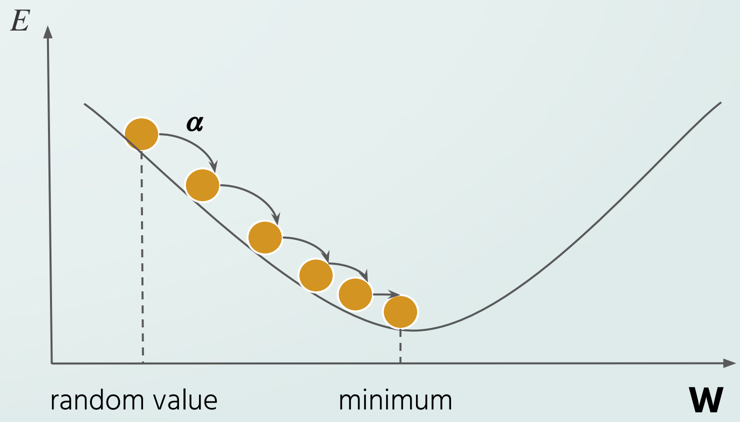
$$\widehat{w}_{ij}^{NEW} = \widehat{w}_{ij}^{OLD} - \alpha \frac{\partial E}{\partial w_{ij}^{OLD}}$$

- **Update $\mathbf{W}$ based on new observations**
- **Retrace steps!** → computer does it for us

# Neural Networks - DIS

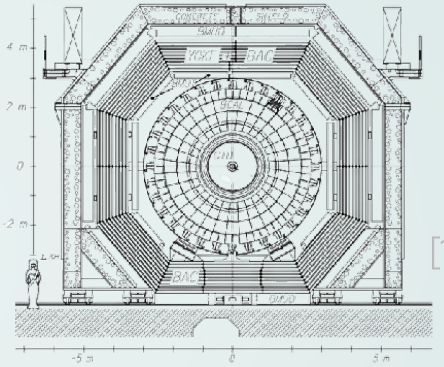
- Weighting classical DIS reconstructions → using all four of the measured quantities as corrections
- We obtain: $$Q^2_{NN} = A_{Q^2}(Q^2_{EL}, Q^2_{JB}, Q^2_{DA}) + L_{Q^2}(A_{Q^2}, E'_e, \theta_e) + H_{Q^2}(A_{Q^2}, P_{T,H}, \delta_H)$$ $$x_{NN} = A_x(x_{EL}, x_{JB}, x_{DA}) + L_x(A_x, Q^2_{NN}, E'_e, \theta_e) + H_x(A_x, P_{T,H}, \delta_H)$$
- Splitting of data (80-20)
- Divide data into bins
- **Compare classical methods to neural network method**
- Parameters:
  - Epochs: 100
  - Batch value: 10 000
  - $\alpha = 10^{-5}$
  - Regularization: $10^{-6}$
  - Momentum: 0.9

**No parameter optimization yet.**

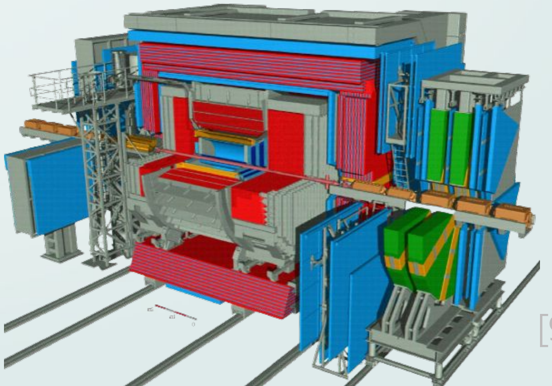


# Experiments and Results - ZEUS

- Neutral current DIS events + simulated data of ZEUS Experiment → **reconstruct four-momentum transferred to the hadronic system**
- **ZEUS detector**
  - 4π solid angle coverage, advanced tracking and Uranium-Scintillator calorimeter, solenoid of 1.43 T
  - Recorded 0.5f b −1 data from e ± p collisions at HERA in 1993-2007 at various beam energies
  - Simulated and real data is available for analyses
- **HERA** → 27.5 GeV electron and 920 GeV proton accelerator
- Monte Carlo simulated e+p DIS events that are provided by the ZEUS collaboration
  - Color Dipole Model (CDM)



[10]



[9]

# Experiments and Results

## Bin Definitions

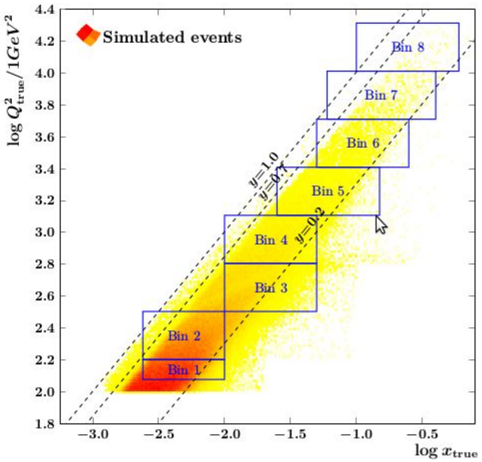



Figure: Distribution of $(x, Q^2)$ for the training set and boundaries of bins.

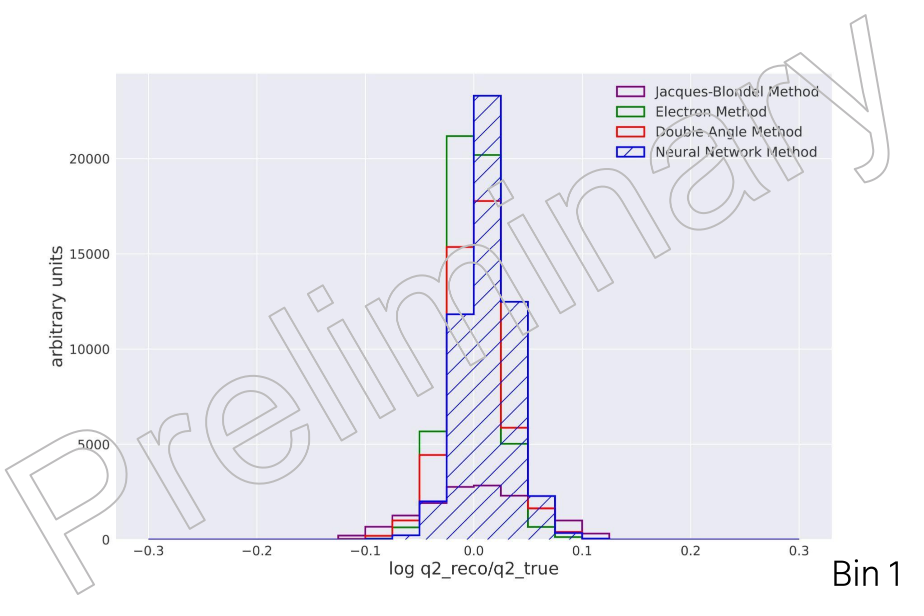
| Bin | $Q^2$ $(GeV^2)$ | $x$ |
|---|---|---|
| 1 | 120 - 160 | 0.0024 - 0.010 |
| 2 | 160 - 320 | 0.0024 - 0.010 |
| 3 | 320 - 640 | 0.01 - 0.05 |
| 4 | 640 - 1280 | 0.01 - 0.05 |
| 5 | 1280 - 2560 | 0.025 - 0.150 |
| 6 | 2560 - 5120 | 0.05 - 0.25 |
| 7 | 5120 - 10240 | 0.06 - 0.40 |
| 8 | 10240 - 20480 | 0.10 - 0.60 |

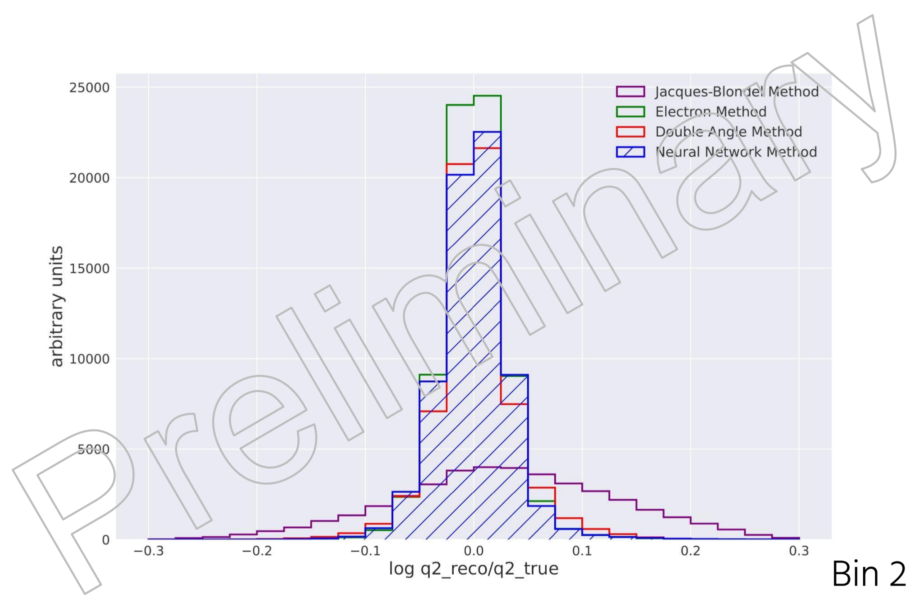
Table: Kinematic bins in $x$ and $Q^2$ used for performance comparisons. The bins were chosen to be close to the bins used in the analyses of hadronic final state in the ZEUS experiment.

[11]

# Experiments and Results

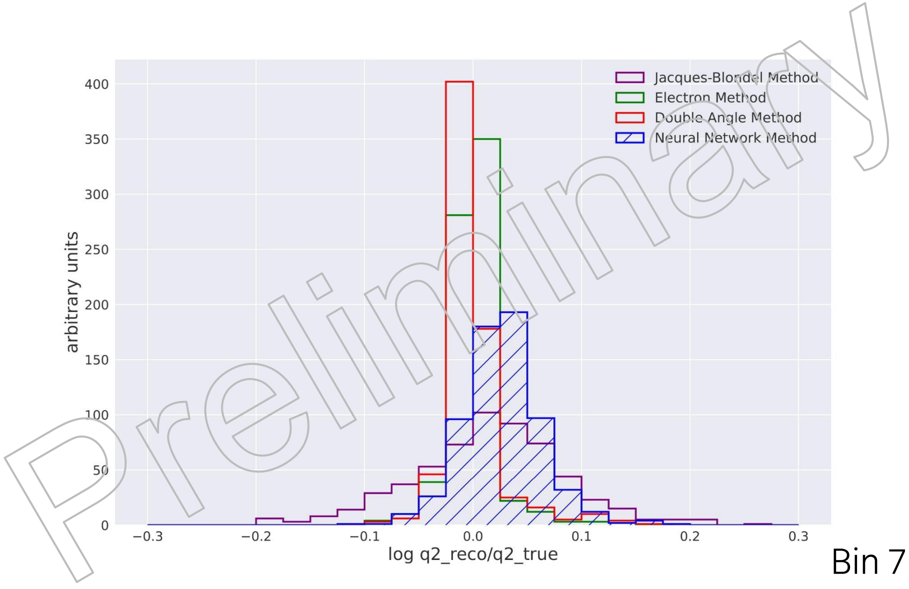
## $Q^2$



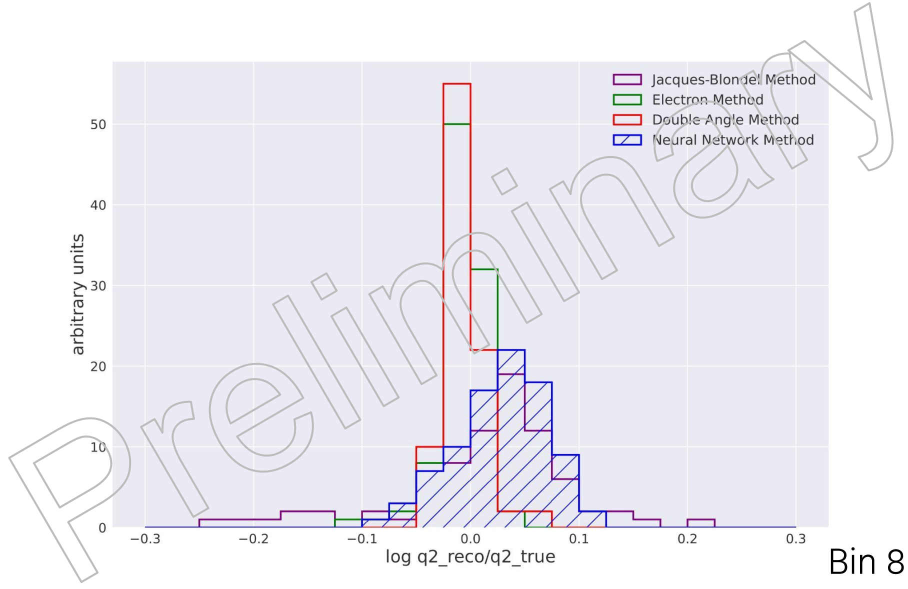



# Experiments and Results

## $Q^2$



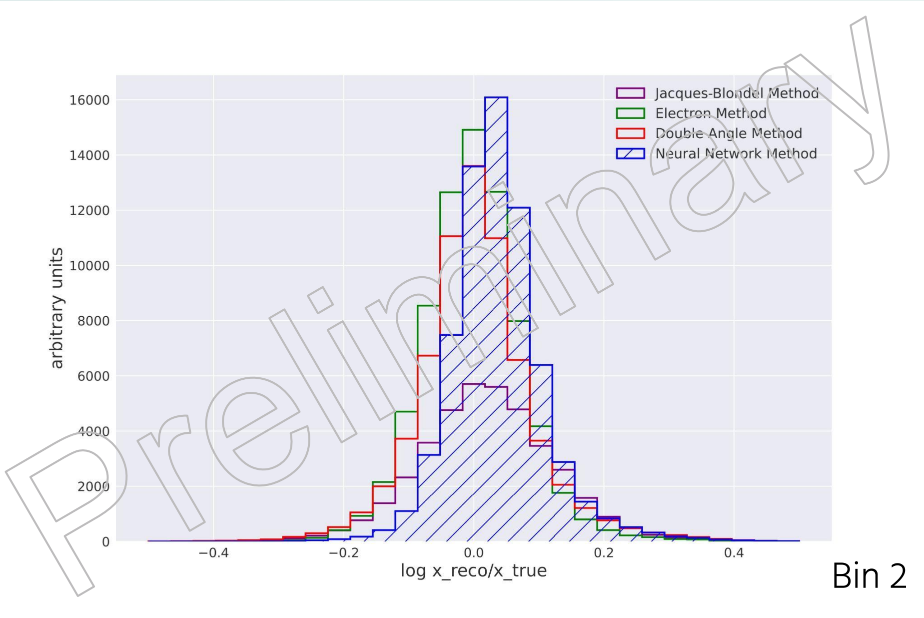

Bin 7

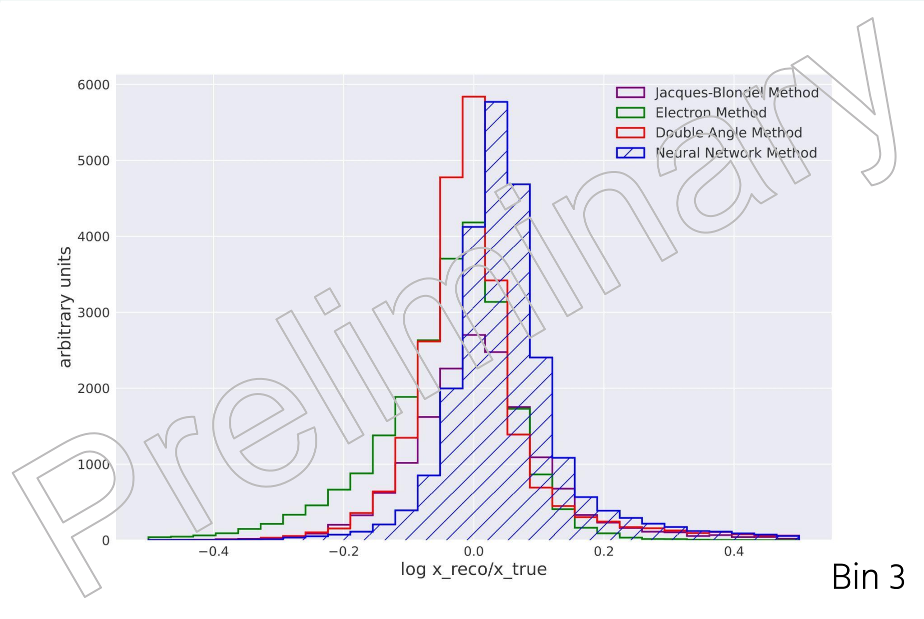



Bin 8

# Experiments and Results

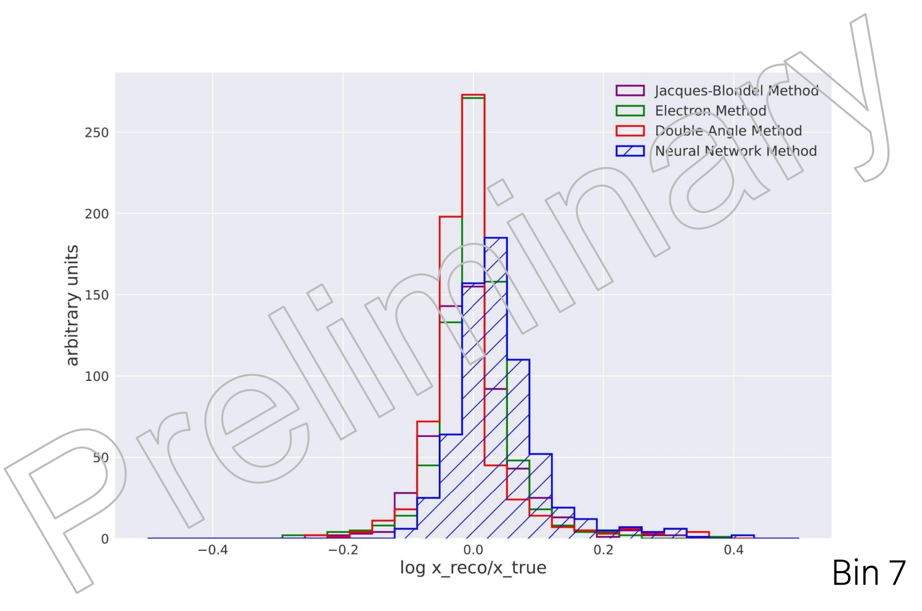
## x



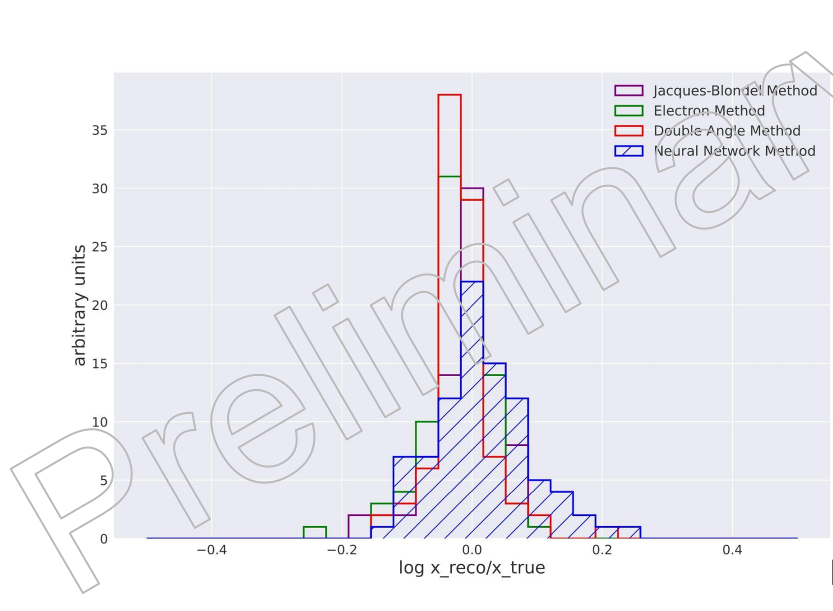



# Experiments and Results

## x




Bin 7

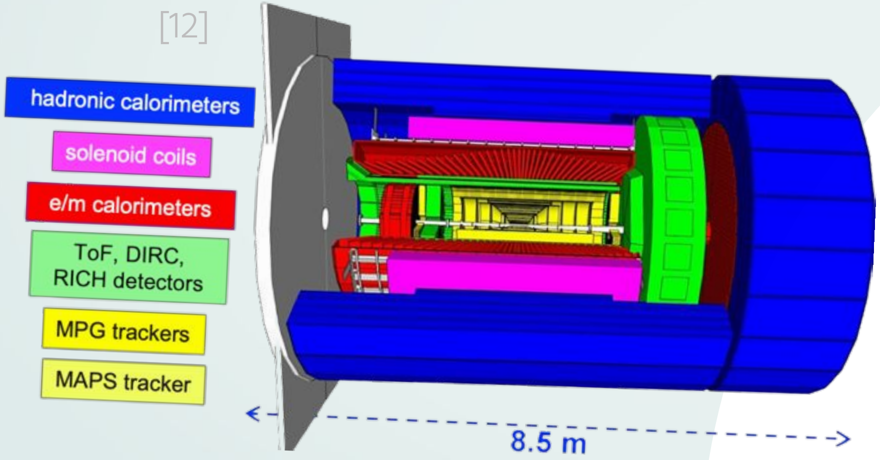



Bin 8

| Bin | Entries | Resolution of log $Q^2$ | Resolution of log $x$ |
|---|---|---|---|
| 1 | 132 702 | NN: 0.026 **EL: 0.022**<br>JB: 0.050 DA: 0.027 | NN: 0.079 **EL: 0.068**<br>JB: 0.085 DA: 0.082 |
| 2 | 166 230 | NN: 0.033 **EL: 0.031**<br>JB: 0.104 DA: 0.037 | NN: 0.081 **EL: 0.077**<br>JB: 0.105 DA: 0.091 |
| 3 | 56 988 | NN: 0.036 EL: 0.028<br>JB: 0.091 **DA: 0.026** | NN: 0.109 EL: 0.115<br>JB: 0.110 **DA: 0.104** |
| 4 | 34 040 | NN: 0.032 **EL: 0.027**<br>JB: 0.093 DA: 0.031 | NN: 0.092 **EL: 0.073**<br>JB: 0.090 DA: 0.080 |
| 5 | 13 602 | NN: 0.034 **EL: 0.026**<br>JB: 0.083 DA: 0.029 | NN: 0.095 **EL: 0.079**<br>JB: 0.091 DA: 0.088 |
| 6 | 4 830 | NN: 0.037 **EL: 0.023**<br>JB: 0.077 DA: 0.026 | NN: 0.082 **EL: 0.068**<br>JB: 0.076 DA: 0.078 |
| 7 | 1 542 | NN: 0.045 **EL: 0.022**<br>JB: 0.071 DA: 0.027 | NN: 0.080 **EL: 0.059**<br>JB: 0.070 DA: 0.072 |
| 8 | 204 | NN: 0.051 EL: 0.022<br>JB: 0.079 **DA: 0.018** | NN: 0.075 EL: 0.054<br>JB: 0.053 **DA: 0.050** |

# Summary and Further Research

[12]




- Neural networks can be used for DIS kinematics reconstruction
- ZEUS Experiment simulation data used for $x$ and $Q^2$ reconstruction
- Appropriate parameter optimization and selection of the training set → **DNNs expected to sufficiently outperform classical reconstruction methods on most of the kinematic range considered**
- Collaborators performed similar analysis with ECCE data
- Results promising and **ready for EIC studies → EPIC simulation campaign**

# References


[1] R. G. Roberts, The Structure of the Proton: Deep Inelastic Scattering, Cambridge University Press, 1993
[2] https://wiki.bnl.gov/eic/index.php/DIS:_What_is_important
[3] S. Bentvelsen, J. Engelen, P. Kooijman, Reconstruction of (x, Q**2) and extraction of structure functions in neutral current scattering at HERA, Workshop on Physics at HERA, 23-42, 1992
[4] C. Schafer, The Neural Network, its Techniques and Applications, 2016
[5] https://media.licdn.com/dms/image/C4E12AQFSvHxzhDZGMQ/article-cover_image-shrink_720_1280/0/1561599110872?e=2147483647&v=beta&t=j3rZZSW0fGZx-aQqwpUer-WDoLbGrO9gGuEe386OdOY
[6] https://toppng.com/uploads/preview/hand-drawn-brain-png-images-your-brain-and-you-what-neuroscience-means-11563243382gmpb27dx1p.png
[7] https://d35fo82fjcw0y8.cloudfront.net/2019/04/08023413/Neural_Network_Brain_Mimic.jpeg
[8] A. Farhat, Reconstructing DIS Kinematics at the EIC Using Deep Learning, XXIX International Workshop on Deep-Inelastic Scattering and Related subjects, in Santiago de Compostela, Galicia, Spain, May 4, 2022
[9] http://www.mit.edu/~hasell/IMAGES/ZEUS_3d.jpg
[10] https://www.researchgate.net/publication/30403552/figure/fig3/AS:668999188746258@1536513244843/Cross-section-of-the-ZEUS-detector.png
[11] M. Diefenthaler et. al., Deeply Learning Deep Inelastic Scattering Kinematics, Eur.Phys.J.C 82 (2022) 11, 1064, 2021
[12] https://wiki.bnl.gov/EPIC/index.php?title=Main_Page

[13] Template created by SlidesGo, including icons by Flaticon, and infographics & images by Freepik

# Neural Networks - Physics Examples

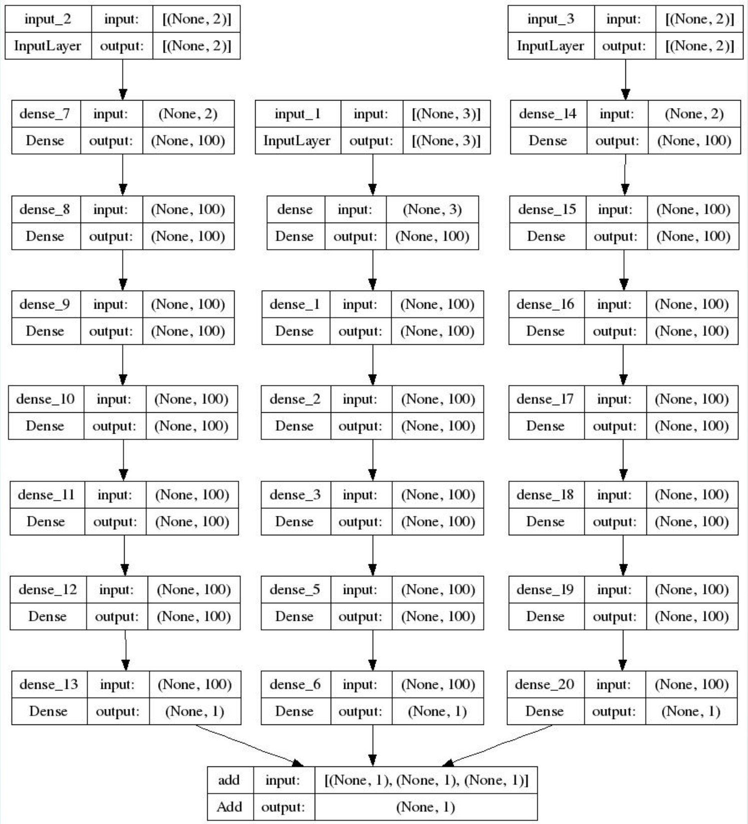
- **Example 1:**
  - Inverse problem in heavy-ion collisions (HIC)
  - High energies → LQCD predicts that the transition between QGP and hadron resonance gas is a smooth crossover
  - Supervised CNN (Du et al., 2020b; Pang et al., 2018), point cloud networks (Steinheimer et al., 2019), and unsupervised AE (Wang et al., 2020) → trained to identify QCD phase transition types using final state hadrons
- **Example 2:**
  - Interaction between bottom and anti-bottom quarks in QGP → modeled as a heavy quark potential → variational function form represented by deep neural networks (Shi et al., 2021)
- **Example 3:**
  - Collision-based experiments → events often categorized by event type for analysis
  - Selection typically computationally expensive in traditional analyses
  - Common task in scintillator detectors in low-energy experiments → discriminate between the neutron and $\gamma$ signals
  - Neural network analysis of pulse shapes → effectively discriminate between these signals (Doucet et al., 2020)

arXiv: 2112.02309

# Neural Networks - Parameters

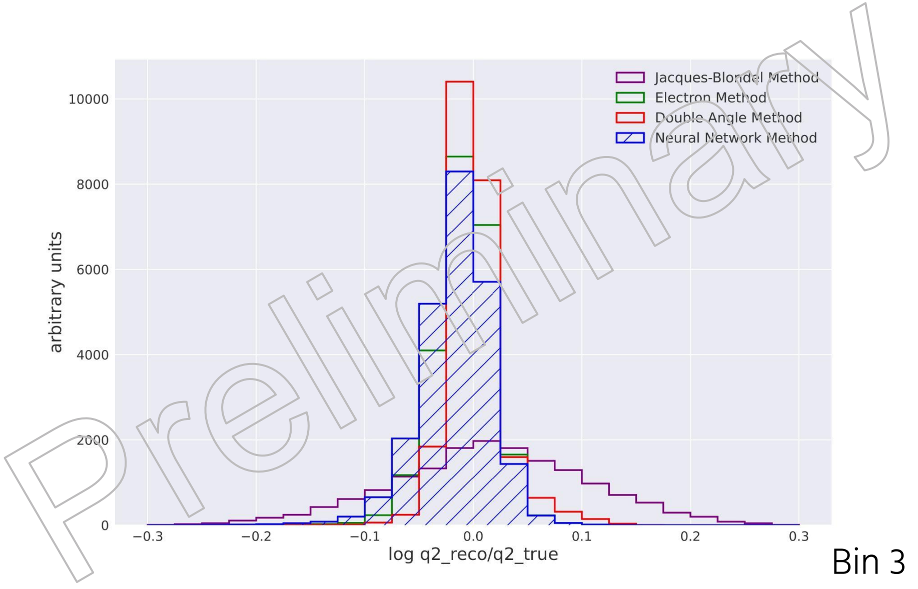
- **Epochs :**
  - Total number of iterations training the data set in one cycle for training the machine learning model
- **Batch value :**
  - Number of samples to go through before updating the model parameters
- **Learning rate :**
  - Measure of how much the weights form the neural network are updated according to the estimated error
- **Regularization parameter :**
  - Parameters that control the loss function, so that it is not over-fitted
- **Momentum :**
  - Adds momentum factor times the weight delta from the previous iteration to back-propagation, adds a boost to the weight change, which makes training faster
  - In case of oscillating weights, momentum dampens oscillations

# Layers in NN for DIS

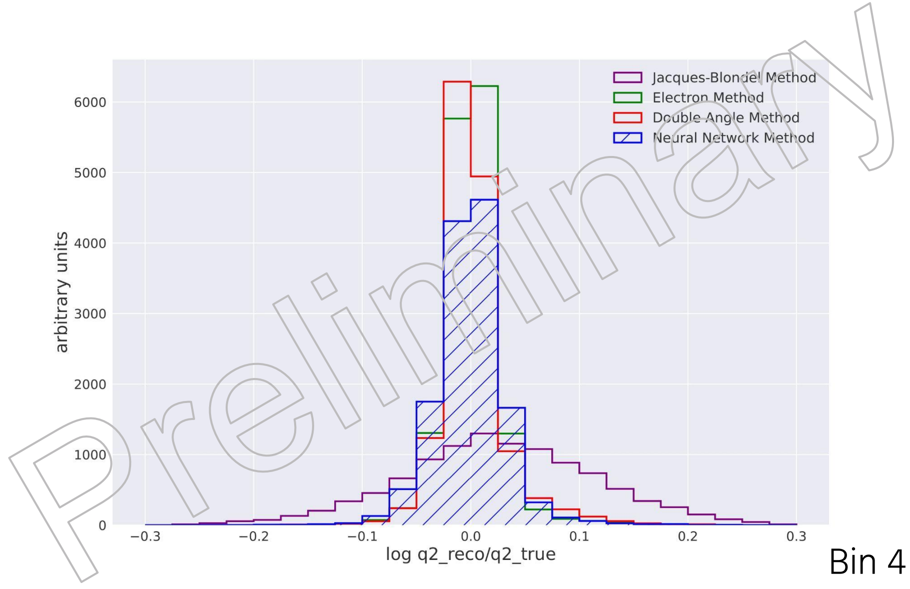


# Experiments and Results

## $Q^2$



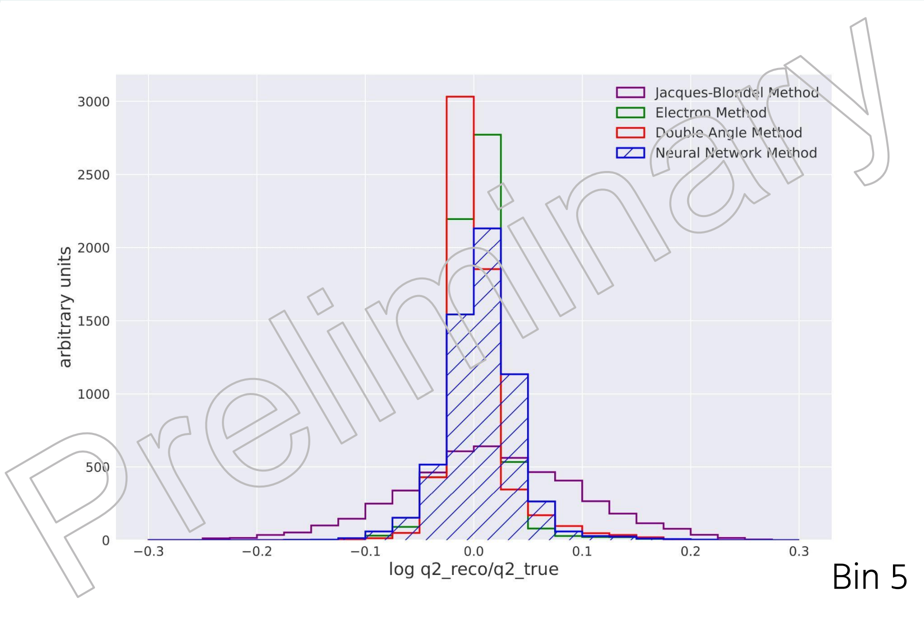

Bin 3



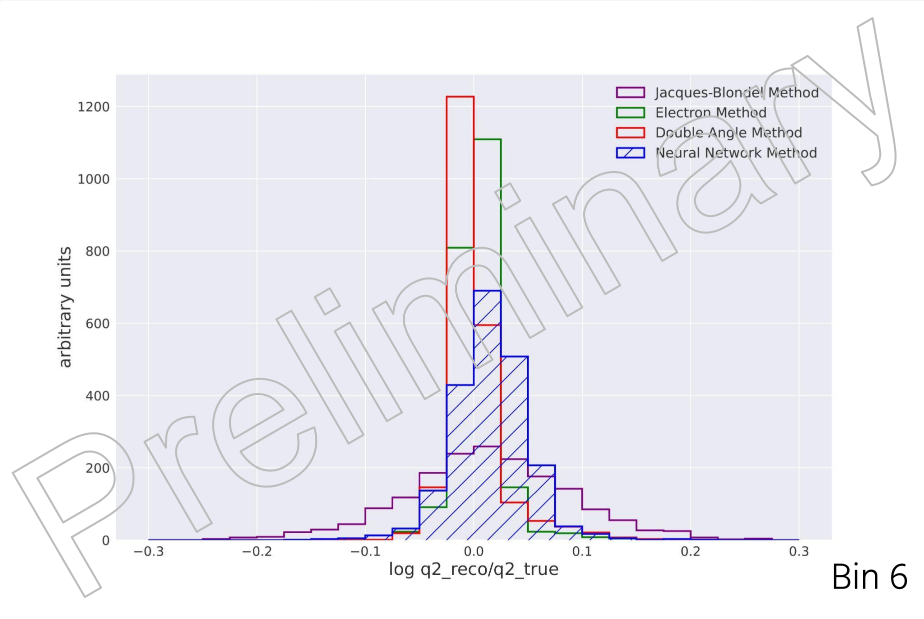

Bin 4

# Experiments and Results

## $Q^2$



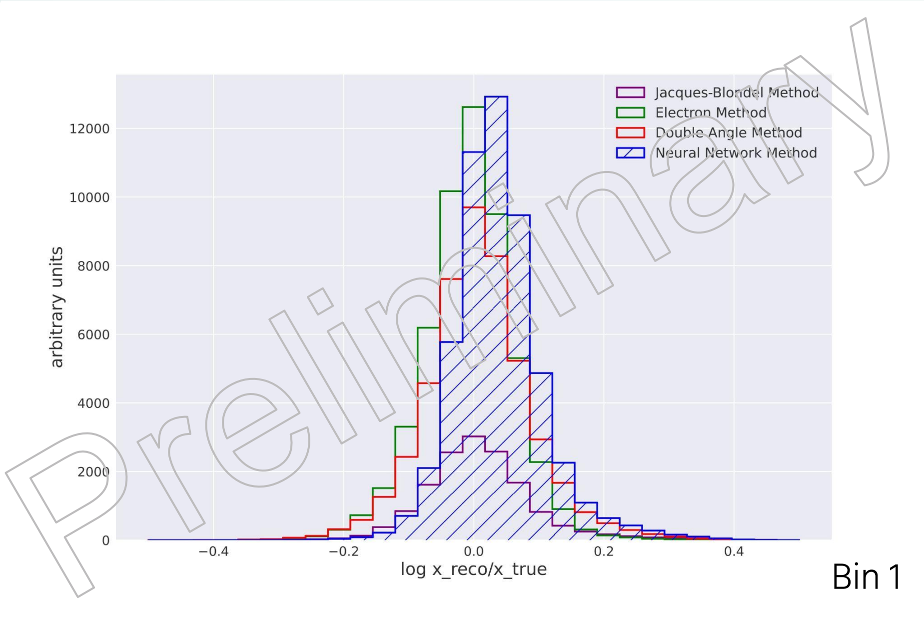
Bin 5

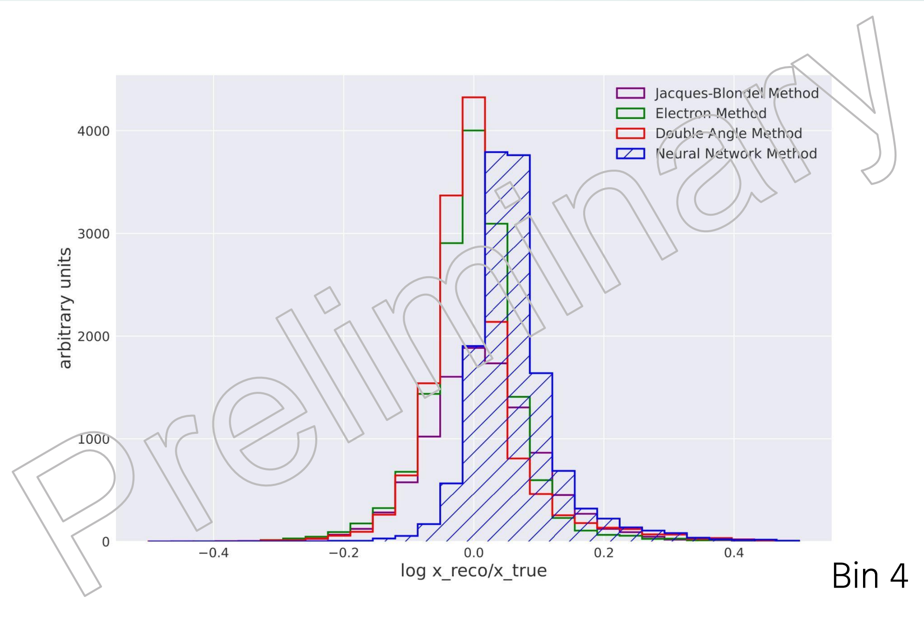


Bin 6

# Experiments and Results

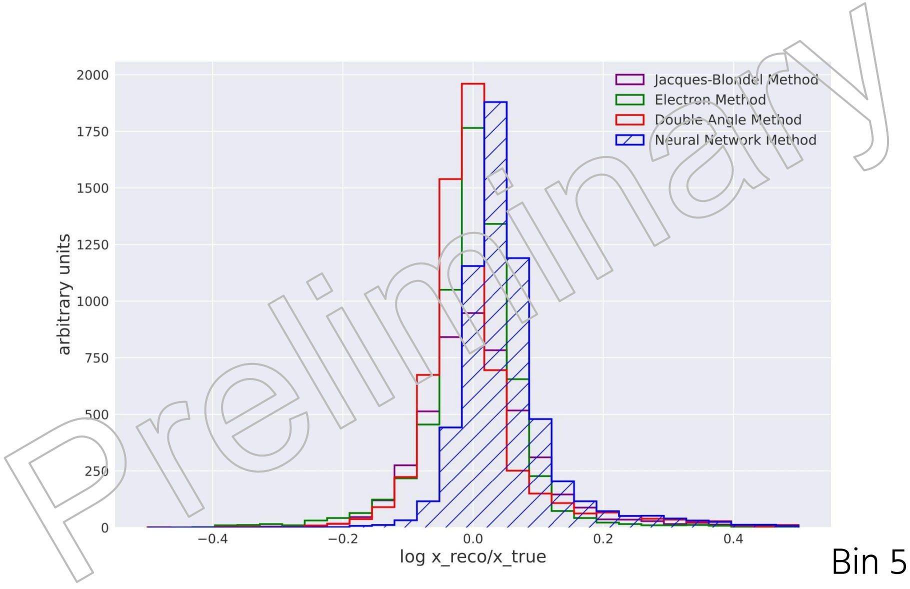
## x



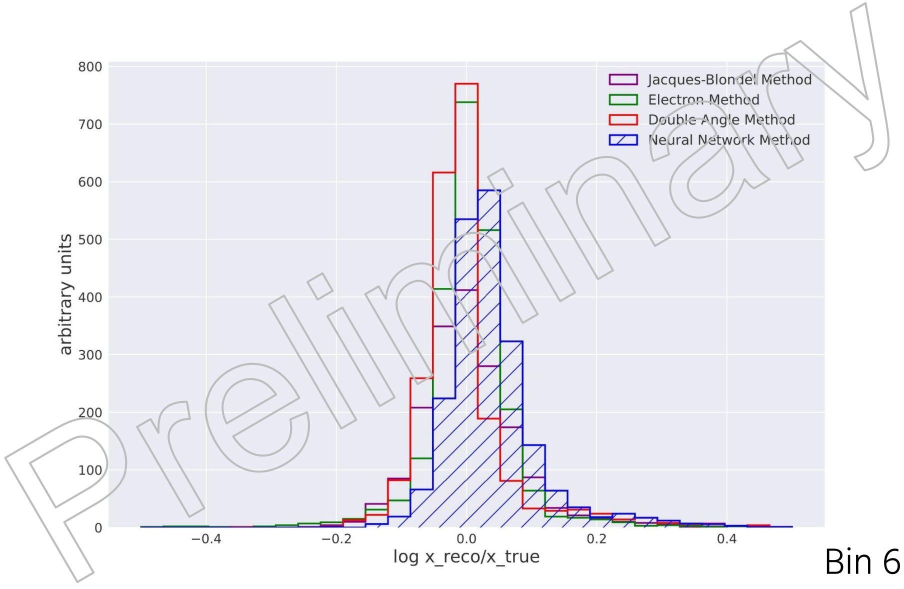



# Experiments and Results

## x

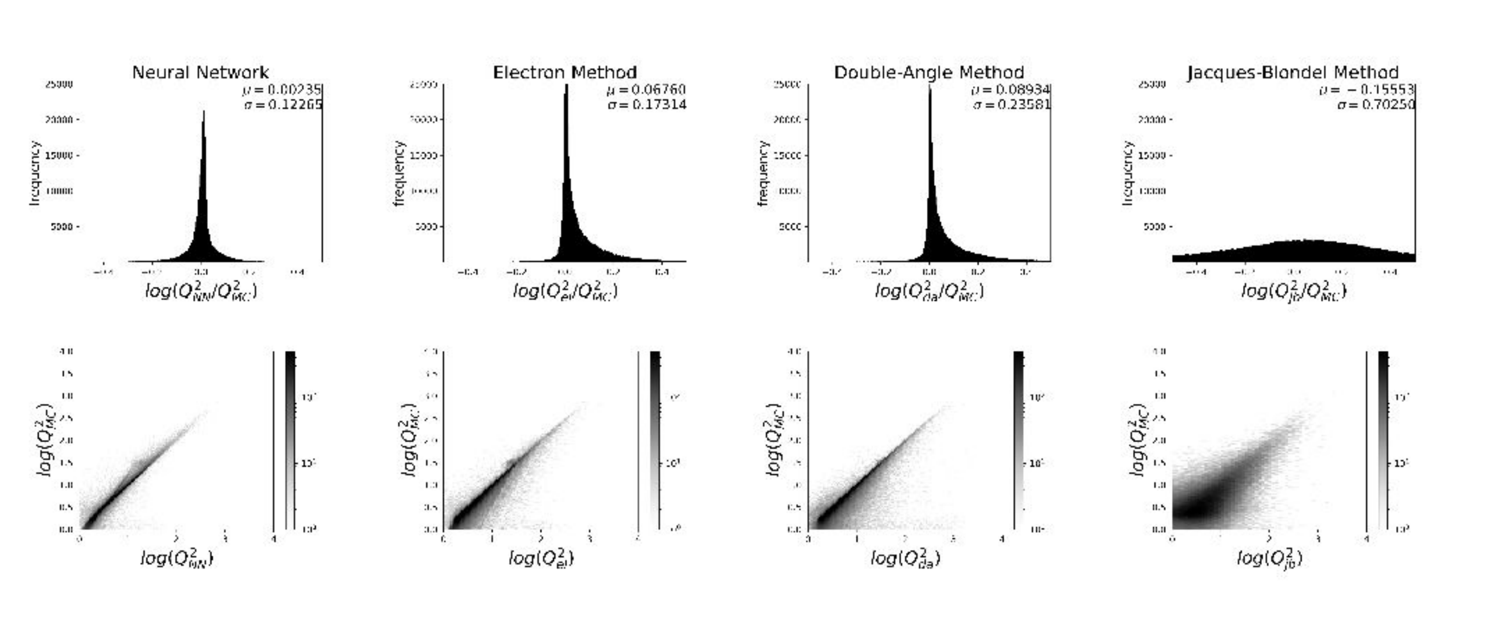



Bin 5



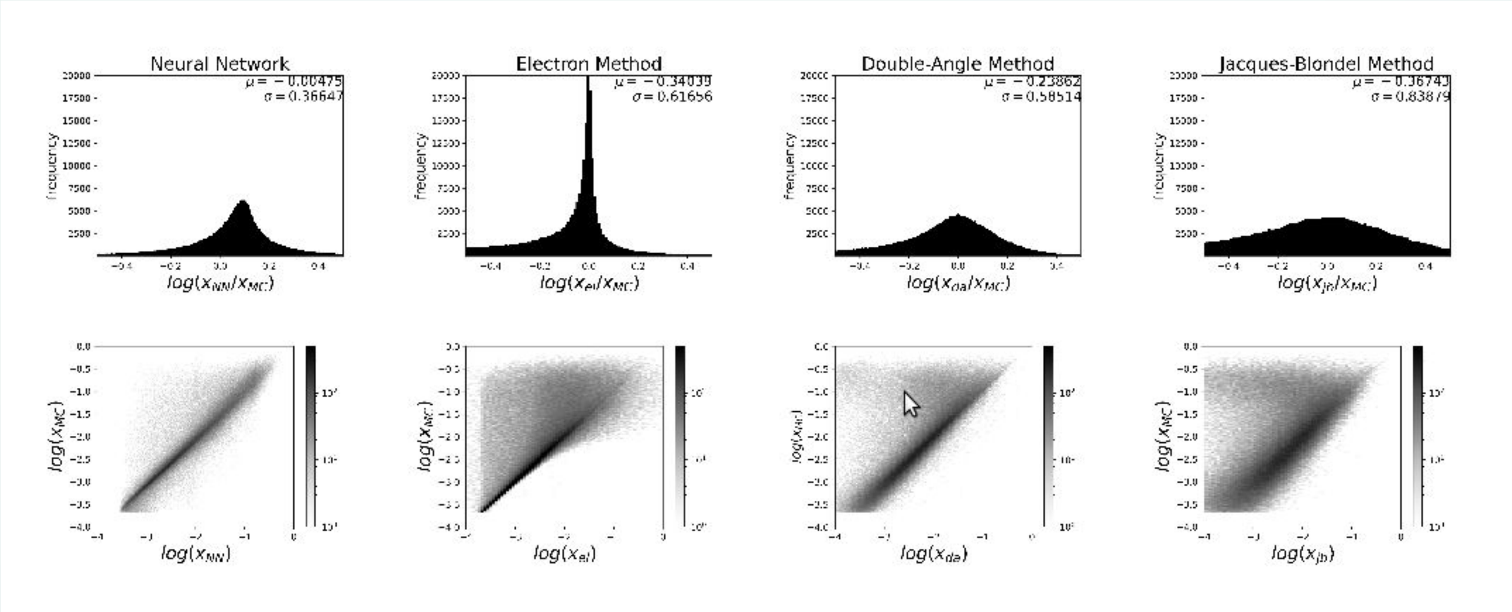

Bin 6

# DNN in ECCE Simulations




[A. Farhat et. al.]

# DNN in ECCE Simulations




[A. Farhat et. al.]

# Monte Carlo Details

- Inclusion of QED and higher order QCD radiative effects using the HERACLES 4.6.6 package with DJANGOH 1.6 interface and the ARIADNE 4.12 and LEPTO 6.5.1 packages for the simulation of the parton cascade
- For both samples the same set of kinematic cuts was applied during the generation
- Same set of PDFs were used, CTEQ5D
- Same hadronisation settings were used to model the hadronisation with the Pythia6 program
- The essential difference between the two samples is the way the higher order corrections are partially modelled with the corresponding algorithms (QCD cascades). Namely, the LEPTO MCEG utilises the parton shower approach, while ARIADNE implements a colordipole model. Accordingly, we label the data-sets produced by the LEPTO generator as “CDM” data sets and those with ARIADNE as “MEPS” data sets
- In the phase-space region at low x and very low inelasticity y, the QED predictions from the Monte Carlo simulations are not reliable because of a limit of higher orders in the calculations
- To ensure optimal electron identification and electron energy resolution, similar to the previous physics analyses, a kinematic cut on y is used

[11]

- Evtake_iwant: error code
- Mc_x: Bjorken x (from initial and final leptons)
- Mc_y: Bjorken y (from initial and final leptons)
- Mc_q2: Bjorken Q2 (from initial and final leptons)
- Mc_x_cr: x (from exchanged photon)
- Mc_q2_cr: Q2 (from exchanged photon)
- Sierror: Sinistra error code (0=OK)
- Sincand: Number of candidates
- Siprob: Electron Probability
- Sipos: CAL+HES+SRTD position
- Sipt: Pt calculated from SiPos
- Sidca: Distance of Closest Approach
- Sitrkp: Momentum of the track
- Siein: Electron energy in Cone
- Sienin: Energy in Cone not from electron
- Zvtx: vertex z (VCTVTX_V(3) or ZTVTXPRM_V(3)); 0 if no vertex
- Sith: Theta calculated from SiPos
- Siecorr: Corrected energy from emEnergyCorrection5.fpp

```
Siecorr[i][0] =
      FCAL: Electron energy corrected using dead material map
      BCAL: Electron energy corrected using dead material map
      RCAL: Electron energy corrected for nonuniformities
Siecorr[i][1] =
      FCAL: same as Siecorr[i][0]
      BCAL: Electron energy corr. for dead material and nonuniformities (new)
      RCAL: Siecorr[i][0] also corrected for dead material
Siecorr[i][2] =
      FCAL: Siecorr[i][1] also corrected for nonuniformities
      BCAL: Siecorr[i][0] also corrected for nonuniformities (old)
      RCAL: same as Siecorr[i][1]
```

- Sizuhmom: Hadronic 4-momentum (Zufos)
- Sicchmom: Hadronic 4-momentum (CorandCut)
- Siccempz: E-Pz from CorAndCut
- Sixel: x Bjorken calculated with electron method
- Siq2el: virtuality Q2 calculated with electron method
- Siyel: inelasticity y calculated with electron method
- Sixjb: x Bjorken calculated with double-angle method based on zufos
- Siq2jb: virtuality Q2 calculated with double-angle method based on zufos
- Siyjb: inelasticity y calculated with double-angle method based on zufos
- Sixda: x Bjorken calculated with double-angle method based on zufos
- Siq2da: virtuality Q2 calculated with double-angle method based on zufos
- Siyda:inelasticity y calculated with double-angle method based on zufos

[11]

# More DIS Variables

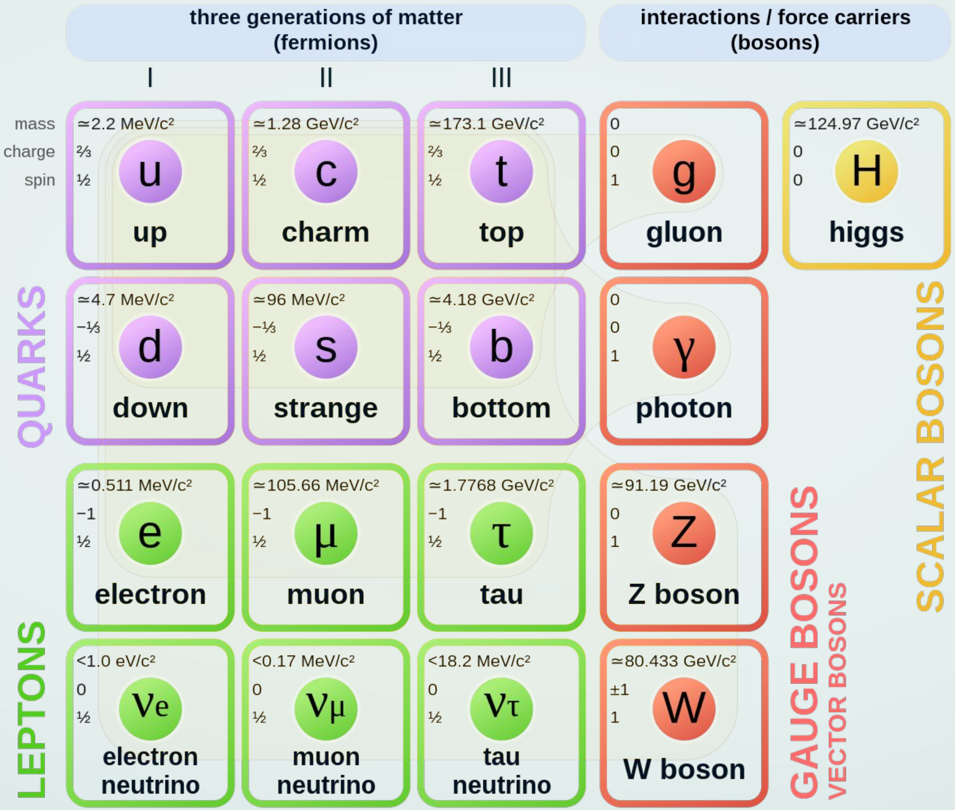
- **Energy-longitudinal momentum balance:** To suppress photoproduction and beam-gas interaction background events and imperfect Monte Carlo simulations of those, restrictions are put on the energy-longitudinal momentum balance. This quantity is defined as:

$$\delta = \delta_l + \delta_H = (E_{l'} - P_{z,l'}) + (E_H - P_{z,H}) = \sum_i (E_i - P_{z,i})$$

  - where the final summation index runs over all energy deposits in the detector
- **Missing transverse energy:** To remove beam-related background and cosmic-ray events, a cut on the missing energy is imposed. P_T,miss/√ET < 2.5 GeV^1/2, where E_T is the total transverse energy in the CAL and P_T,miss is the missing transverse momentum, the transverse component of the vector sum of the hadronic final state and scattered electron momenta

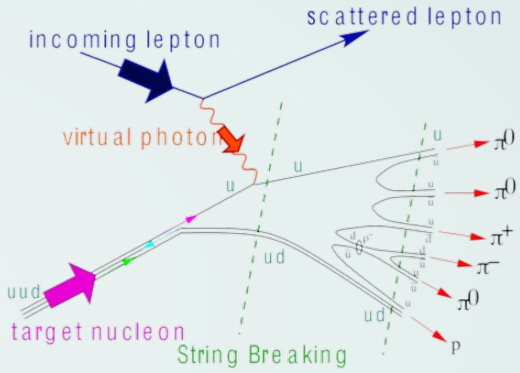
[11]

# Standard Model of Particle Physics

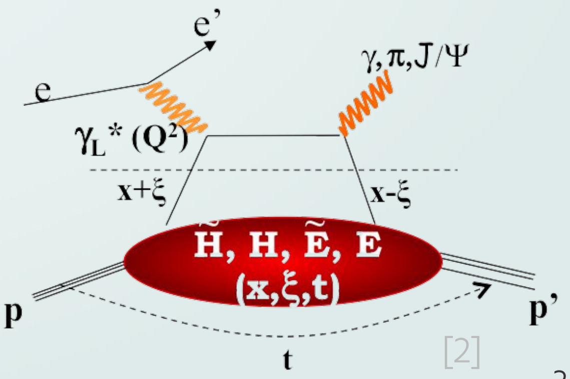


# Deep-Inelastic Scattering (DIS)

- Straightforward to measure $E$, $F$ and $\vartheta \rightarrow \gamma$ more involved
- Quality of detector affects measurement accuracy
- Unique points in $(x,Q^2)$ plane → 6 different reconstruction methods
- **Electron Method**
  - Combining $E$ and $\vartheta$ gives unique point in $(x,Q^2)$
  - For $E$ of a few GeV → large uncertainty in $x$ reconstruction
- **Jacques-Blondel Method**
  - $F$ and $\gamma$
- **Double-Angle Method**
  - Use electron and final state hadron flow
  - Measure $E$ of both → $\vartheta$ and $\gamma$
  - Determination of $\gamma$ gives remarkably good results
    - Angles independent of $E$ fluctuations of jet (can be shown)



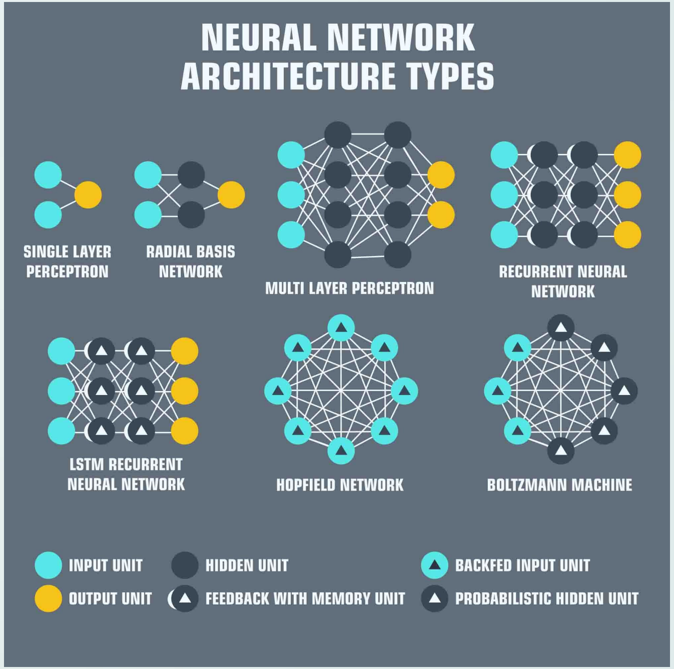

[2]




[2]

# Neural Networks - General

- Powerful mathematical model combining linear algebra, biology and statistics to solve a problem in a unique way
- Takes a given amount of inputs and then calculates a specified number of outputs aimed at targeting the actual result
- Problems such as pattern recognition, linear classification, data fitting and more can all be conquered with a neural network
- **Supervised**
  - Given a collection of input data and a collection of output data
- **Unsupervised**
  - We do not have output data
- Neural networks thrive at **adaptive learning**




[5]

# Neural Networks - General

- Layers are represented as vectors
- All observations → each layer is a matrix
- Output matrix of neural network $\mathbf{T}$, $\mathbf{Y}$ is a prediction of $\mathbf{T}$ based on the inputs
- Supervised learning → **minimize $\mathbf{T}$ and $\mathbf{Y}$ difference** (weight matrix $\mathbf{W}$)
- **But do we really need neural networks for this?**
- $\mathbf{W}$ maps from $\mathbf{X}$ to $\mathbf{Y}$ such that $\mathbf{T}=\mathbf{W}\mathbf{X}+\mathbf{b}$
- We like linear functions → $[\mathbf{W}|\mathbf{b}]$ → $\mathbf{T}=[\mathbf{W}|\mathbf{b}]\,[\mathbf{X}\ \mathbf{1}]^{\mathbf{T}}$ → $\mathbf{T}=\mathbf{W}\mathbf{X}$
- Pseudoinverse $\hat{\mathbf{W}}=\mathbf{T}\mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^{\mathbf{T}}$ → compute $\mathbf{Y}=\hat{\mathbf{W}}\mathbf{X}$
- Compare $\mathbf{Y}$ and $\mathbf{T}$ → **linear algebra method**
- Linear relationship between $\mathbf{T}$ and $\mathbf{X}$