Nucleon axial form factor from a Bayesian neural-network analysis of neutrino-scattering data*

Luis Alvarez Ruso, Krzysztof M. Graczyk, **Eduardo Saúl Sala**

*arXiv:1805.00905

Introduction

Neutrino interactions with matter:

- Crucial for oscillation experiments.
- Realistic modeling of neutrino interactions with nuclei required.
- Key ingredient for models are the amplitudes and cross sections.

Introduction

Neutrino interactions with matter:

- Crucial for oscillation experiments.
- Realistic modeling of neutrino interactions with nuclei required.
- Key ingredient for models are the amplitudes and cross sections.

Axial form factor as a source of uncertainty:

- Function of Q^2.
- Axial coupling: $g_A = F_A(Q^2 = 0) = 1.2723 \pm 0.0023$.
- Dipole ansatz:

$$F_A^{\text{dipole}}(Q^2) = g_A \left(1 + \frac{Q^2}{M_A^2}\right)^{-2},$$

  - Not theoretically well founded.
  - Electric and magnetic form factors have non dipole shape.

Introduction

Bubble chamber experiments: $\nu_\mu + d \to \mu^- + p + p$

- Experimental Q^2 distribution of observed events.
    - ANL.
    - BNL.
    - FNAL.
- Known electromagnetic form factors from electron scattering data.
- Axial form factor can be extracted.
- Deuteron effects have to be considered.

F_A parametrization

- Which parametrization?
- Specific functional form $\rightarrow$ bias the results of the analysis?
- How many parameters?

F_A parametrization

- Which parametrization?
- Specific functional form $\rightarrow$ bias the results of the analysis?
- How many parameters?
- **Neural networks** $\rightarrow$ model-independent.
- **Bayesian statistics** $\rightarrow$ comparisons between models.

Statistical model

General methods:

- Non-parametric:
  - No particular functional model is assumed.
  - Probabilities are determined only by the data.
  - Large size of the data $\rightarrow$ introduction of many internal parameters.
  - Computationally expensive.

Statistical model

General methods:

- Non-parametric:
  - No particular functional model is assumed.
  - Probabilities are determined only by the data.
  - Large size of the data $\rightarrow$ introduction of many internal parameters.
  - Computationally expensive.
- Parametric:
  - Specific functional form of the model assumed.
  - Easy to find the optimal configuration.
  - Limited ability for an accurate description of the data.

Statistical model

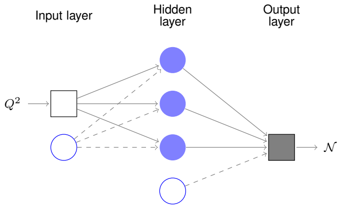
General methods:

- Non-parametric:
  - No particular functional model is assumed.
  - Probabilities are determined only by the data.
  - Large size of the data $\rightarrow$ introduction of many internal parameters.
  - Computationally expensive.
- Parametric:
  - Specific functional form of the model assumed.
  - Easy to find the optimal configuration.
  - Limited ability for an accurate description of the data.
- **Semi-parametric**:
  - Best features from both.
  - Broad class of functions is considered $\rightarrow$ optimal model.
  - One realization are **neural-networks**: Methodology developed for last 30 years.

Neural networks

Feed-forward NN in a multilayer perceptron (MLP) configuration:

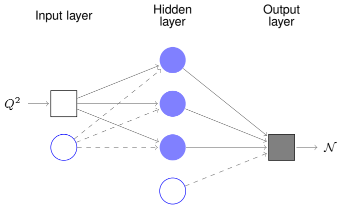
- Nonlinear map $\mathcal{N} : \mathbb{R}^{\text{in}} \mapsto \mathbb{R}^{\text{out}}$

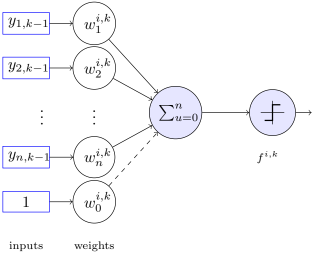


Neural networks

Feed-forward NN in a multilayer perceptron (MLP) configuration:

- Nonlinear map $\mathcal{N} : \mathbb{R}^{\text{in}} \mapsto \mathbb{R}^{\text{out}}$

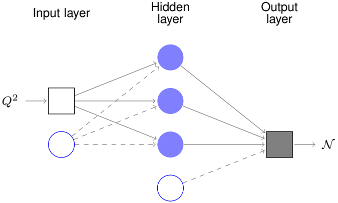


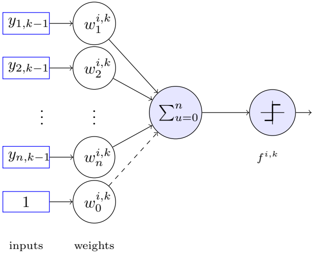


Neural networks

Feed-forward NN in a multilayer perceptron (MLP) configuration:

- Nonlinear map $\mathcal{N} : \mathbb{R}^{\text{in}} \mapsto \mathbb{R}^{\text{out}}$





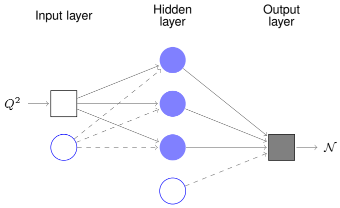
- For every unit:

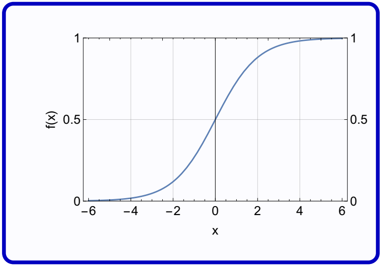
$$y_{i,k} = f^{i,k}\left(\sum_{u\in\text{previous layer}} w_u^{i,k} y_{u,k-1}\right)$$

Neural networks

Feed-forward NN in a multilayer perceptron (MLP) configuration:

- Nonlinear map $\mathcal{N} : \mathbb{R}^{\text{in}} \mapsto \mathbb{R}^{\text{out}}$





- For every unit:

$$y_{i,k} = f^{i,k}\left(\sum_{u \in \text{previous layer}} w_u^{i,k} y_{u,k-1}\right)$$

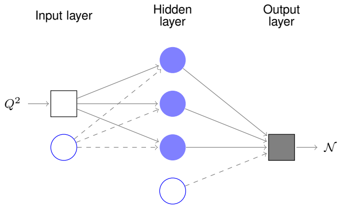
Activation function:

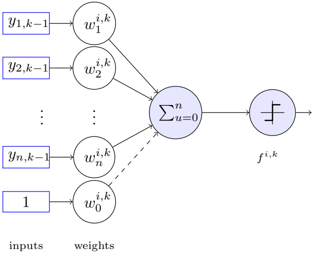
- $f(x) = \frac{1}{1+\exp(-x)}$.

Neural networks

Feed-forward NN in a multilayer perceptron (MLP) configuration:

- Nonlinear map $\mathcal{N} : \mathbb{R}^{\text{in}} \mapsto \mathbb{R}^{\text{out}}$





- For every unit:

$$y_{i,k} = f^{i,k}\left(\sum_{u\in\text{previous layer}} w_u^{i,k} y_{u,k-1}\right)$$

Activation function:

- $f(x) = \frac{1}{1+\exp(-x)}$.
- Bias: $f(x) = 1$.
- Output: $f(x) = x$.

Neural networks

MLP $\mathcal{N} : \mathbb{R} \mapsto \mathbb{R}$ with $M = 3$ and a single hidden layer:

$$\mathcal{N}_M(Q^2; \{w_j\}) = \sum_{n=1}^{M} w_{2M+n}\, f\left(w_n Q^2 + w_{M+n}\right) + w_{3M+1}.$$

- Cybenko's theorem: for large enough M, can map arbitrarily well any continuous function and its derivative.

Neural networks

MLP $\mathcal{N} : \mathbb{R} \mapsto \mathbb{R}$ with $M = 3$ and a single hidden layer:

$$\mathcal{N}_M(Q^2; \{w_j\}) = \sum_{n=1}^{M} w_{2M+n}\, f\left(w_n Q^2 + w_{M+n}\right) + w_{3M+1}.$$

- Cybenko's theorem: for large enough M, can map arbitrarily well any continuous function and its derivative.

Constrains of parametrization

- Q^2-range: $(0, 3)$ GeV^2.
- $F_A(Q^2 = 0) = g_A$.
- $F_A(Q^2)/G_D(Q^2)$ of the order of g_A.

Neural networks

MLP $\mathcal{N} : \mathbb{R} \mapsto \mathbb{R}$ with $M = 3$ and a single hidden layer:

$$\mathcal{N}_M(Q^2; \{w_j\}) = \sum_{n=1}^{M} w_{2M+n}\, f\left(w_n Q^2 + w_{M+n}\right) + w_{3M+1}.$$

- Cybenko's theorem: for large enough M, can map arbitrarily well any continuous function and its derivative.

Re-scale the output $\rightarrow$ normalizing to dipole ansatz:

$$F_A(Q^2) = F_A^{\mathrm{dipole}}(Q^2) \times \mathcal{N}_M(Q^2; \{w_i\})$$

$$F_A^{\mathrm{dipole}}(Q^2) = g_A \left(1 + \frac{Q^2}{M_A^2}\right)^{-2}; \qquad M_A = 1\mathrm{GeV}$$

- Neural-network response $\rightarrow$ deviation of F_A from F_A^{dipole}

Bayesian framework for MLP

General idea:

- How many parameters? $\rightarrow$ number of units in the hidden layer: M.
  - Too many parameters $\rightarrow$ over-fitting.
  - Too simple model $\rightarrow$ under-fitting.

Bayesian framework for MLP

General idea:

- How many parameters? $\rightarrow$ number of units in the hidden layer: M.
    - Too many parameters $\rightarrow$ over-fitting.
    - Too simple model $\rightarrow$ under-fitting.

- Bayes' theorem:

$$\mathcal{P}(\mathcal{N}_M \mid \mathcal{D}) = \frac{\mathcal{P}(\mathcal{D} \mid \mathcal{N}_M)\mathcal{P}(\mathcal{N}_M)}{\mathcal{P}(\mathcal{D})}$$

We use $\mathcal{P}(\mathcal{N}_M \mid \mathcal{D})$ to compare different models.

Bayesian framework for MLP

General idea:

- How many parameters? $\rightarrow$ number of units in the hidden layer: M.
  - Too many parameters $\rightarrow$ over-fitting.
  - Too simple model $\rightarrow$ under-fitting.

- Bayes' theorem:

$$\mathcal{P}(\mathcal{N}_M \mid \mathcal{D}) = \frac{\mathcal{P}(\mathcal{D} \mid \mathcal{N}_M)\mathcal{P}(\mathcal{N}_M)}{\mathcal{P}(\mathcal{D})}$$

  We use $\mathcal{P}(\mathcal{N}_M \mid \mathcal{D})$ to compare different models.

- Assuming all NN configurations are equally suited to describe data.
  $\rightarrow$ Prior for each model: $\mathcal{P}(\mathcal{N}_1) = \mathcal{P}(\mathcal{N}_2) = \cdots = \mathcal{P}(\mathcal{N}_M)$

Bayesian framework for MLP

Model comparison:

$$\mathcal{P}(\mathcal{N}_M \mid \mathcal{D}) = \frac{\mathcal{P}(\mathcal{D} \mid \mathcal{N}_M)\mathcal{P}(\mathcal{N}_M)}{\mathcal{P}(\mathcal{D})}$$

For a given model:

- Posterior $= \frac{\text{Likelihood} \times \text{Prior}}{\text{Evidence}}$

$$\mathcal{P}(\mathbf{w} \mid \mathcal{D}, \mathcal{N}_M) = \frac{\mathcal{P}(\mathcal{D} \mid \mathbf{w}, \mathcal{N}_M)\mathcal{P}(\mathbf{w} \mid \mathcal{N}_M)}{\mathcal{P}(\mathcal{D} \mid \mathcal{N}_M)}$$

- Key to calculate the $\mathcal{P}(\mathcal{N}_M \mid \mathcal{D})$:

$$\mathcal{P}(\mathcal{D} \mid \mathcal{N}_M) = \int \mathcal{P}(\mathcal{D} \mid \mathbf{w}, \mathcal{N}_M)\mathcal{P}(\mathbf{w} \mid \mathcal{N}_M)\, d\mathbf{w}$$

Bayesian framework for MLP

Evaluating the evidence:

- For many problems the posterior has a strong peak at $\mathbf{w}_{\mathrm{MP}}$

$$\underbrace{\mathcal{P}(\mathcal{D} \mid \mathcal{N}_M)}_{\text{Evidence}} \approx \underbrace{\mathcal{P}(\mathcal{D} \mid \mathbf{w}_{\mathrm{MP}}, \mathcal{N}_M)}_{\text{Best fit likelihood}} \underbrace{\mathcal{P}(\mathbf{w}_{\mathrm{MP}} \mid \mathcal{N}_M)\, \Delta\mathbf{w}}_{\text{Occam factor}}$$

Bayesian framework for MLP

Evaluating the evidence:

- For many problems the posterior has a strong peak at $\mathbf{w}_{\mathrm{MP}}$

$$\underbrace{\mathcal{P}(\mathcal{D} \mid \mathcal{N}_M)}_{\text{Evidence}} \approx \underbrace{\mathcal{P}(\mathcal{D} \mid \mathbf{w}_{\mathrm{MP}}, \mathcal{N}_M)}_{\text{Best fit likelihood}} \underbrace{\mathcal{P}(\mathbf{w}_{\mathrm{MP}} \mid \mathcal{N}_M)\, \Delta\mathbf{w}}_{\text{Occam factor}}$$

- Likelihood in terms of χ^2:

$$\mathcal{P}(\mathcal{D} \mid \mathbf{w}, \mathcal{N}) = \frac{1}{N_L} \exp(-\chi^2)$$

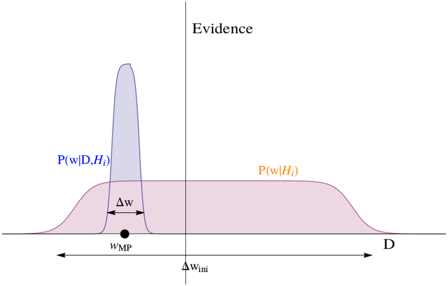
Bayesian framework for MLP

Evaluating the evidence:

- For many problems the posterior has a strong peak at $\mathbf{w}_{\text{MP}}$

$$\underbrace{\mathcal{P}(\mathcal{D} \mid \mathcal{N}_M)}_{\text{Evidence}} \approx \underbrace{\mathcal{P}(\mathcal{D} \mid \mathbf{w}_{\text{MP}}, \mathcal{N}_M)}_{\text{Best fit likelihood}} \underbrace{\mathcal{P}(\mathbf{w}_{\text{MP}} \mid \mathcal{N}_M)\, \Delta \mathbf{w}}_{\text{Occam factor}}$$

- Interpretation of the Occam factor for one parameter:



Occam factor penalizes complex models:

- Δw is the posterior uncertainty in w
- Assume uniform on large interval Δw_{ini}
- $\mathcal{P}(w_{\text{MP}} \mid \mathcal{N}_M) = \frac{1}{\Delta w_{ini}}$

$$\text{Occam factor} = \frac{\Delta w}{\Delta w_{ini}} \rightarrow \frac{V(\text{Posterior})}{V(\text{Prior})}$$

J.C. MacKay, Neural Computation 4, 415 (1992)

Bayesian framework for MLP

Predictive power:

- Choose the right model.
- Avoid overfit and underfit.
- Overestimation and underestimation of uncertainties:

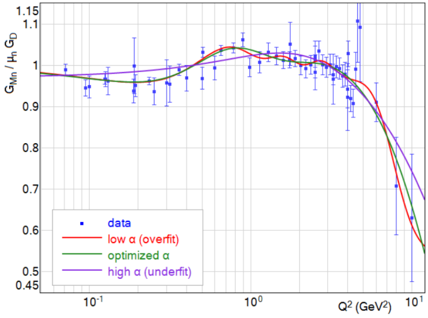
This approach has been used in:

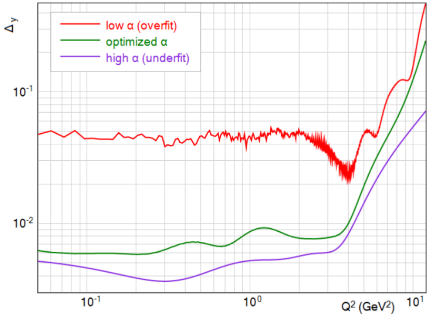
- Parametrization of EM nucleon form factors,
  K. M. Graczyk et al., JHEP 1009 (2010)
- Proton Radius,
  K. M. Graczyk and C. Juszczak, PRC90, 054334 (2014)

Bayesian framework for MLP

Predictive power:

- Choose the right model.
- Avoid overfit and underfit.
- Overestimation and underestimation of uncertainties:





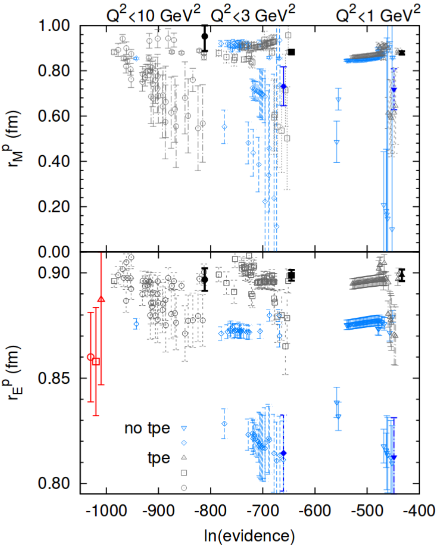
Fit of $G_{Mn}/\mu_n G_D$ data.

K. M. Graczyk, PRC84, 034314 (2011)

Bayesian framework for MLP

Application of this method:
Proton radii determination

K. M. Graczyk and C. Juszczak,
PRC90, 054334 (2014)



Analysis of ANL data

Neutrino-induced CCQE: $\nu_\mu(k) + n(p) \to \mu^-(k') + p(p')$

$$\frac{d\sigma_{\nu n}}{dQ^2} = \frac{G_F^2 m_N^2}{8\pi E_\nu^2}\left[A(Q^2) + B(Q^2)\frac{(s-u)}{m_N^2} + C(Q^2)\frac{(s-u)^2}{m_N^4}\right]$$

- A, B, C are functions of $F_{1,2}^V(Q^2)$ and $F_{A,P}(Q^2)$.
- $F_{1,2}^V(Q^2)$ from electron scattering data. $F_P(Q^2)$ given in terms of $F_A(Q^2)$

Analysis of ANL data

Neutrino-induced CCQE: $\nu_\mu(k) + n(p) \to \mu^-(k') + p(p')$

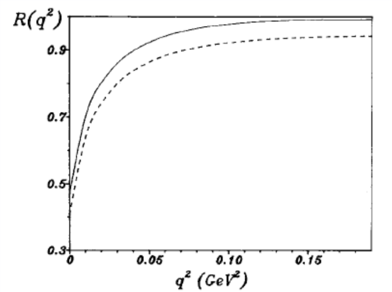
$$\frac{d\sigma_{\nu n}}{dQ^2} = \frac{G_F^2 m_N^2}{8\pi E_\nu^2}\left[A(Q^2) + B(Q^2)\frac{(s-u)}{m_N^2} + C(Q^2)\frac{(s-u)^2}{m_N^4}\right]$$

- $A,$
- F_1 …)

Bubble chamber experiments: $\nu_\mu + d \to \mu^- + p + p$

$$\frac{d\sigma_{\nu d}}{dQ^2} = R(Q^2)\frac{d\sigma_{\nu n}}{dQ^2}$$

Singh, Arenhovel Z. Phys. A324 (1986)



Analysis of ANL data

Neutrino-induced CCQE: $\nu_\mu(k) + n(p) \to \mu^-(k') + p(p')$

$$\frac{d\sigma_{\nu n}}{dQ^2} = \frac{G_F^2 m_N^2}{8\pi E_\nu^2}\left[A(Q^2) + B(Q^2)\frac{(s-u)}{m_N^2} + C(Q^2)\frac{(s-u)^2}{m_N^4}\right]$$

- A, B, C are functions of $F_{1,2}^V(Q^2)$ and $F_{A,P}(Q^2)$.
- $F_{1,2}^V(Q^2)$ from electron scattering data. $F_P(Q^2)$ given in terms of $F_A(Q^2)$
- Events:

$$N^{th} = \int_0^\infty dE_\nu \frac{d\sigma}{dQ^2}(E_\nu, F_A, Q^2)\phi\left(E_\nu\right)$$

- Neutrino flux:

$$\phi\left(E_\nu\right) = p\frac{1}{\sigma(E_\nu, F_A)}\frac{dN}{dE_\nu}$$

Experimental E_ν distribution of observed envents $\to \frac{dN}{dE_\nu}$

Barish et al. PRD19 (1979)

Analysis of ANL data

Neutrino-induced CCQE: $\nu_\mu(k) + n(p) \to \mu^-(k') + p(p')$

$$\frac{d\sigma_{\nu n}}{dQ^2} = \frac{G_F^2 m_N^2}{8\pi E_\nu^2}\left[A(Q^2) + B(Q^2)\frac{(s-u)}{m_N^2} + C(Q^2)\frac{(s-u)^2}{m_N^4}\right]$$

- A, B, C are functions of $F_{1,2}^V(Q^2)$ and $F_{A,P}(Q^2)$.
- $F_{1,2}^V(Q^2)$ from electron scattering data. $F_P(Q^2)$ given in terms of $F_A(Q^2)$
- Events:

$$N^{th} = \int_0^\infty dE_\nu \frac{d\sigma}{dQ^2}(E_\nu, F_A, Q^2)\phi(E_\nu)$$

- Neutrino flux:

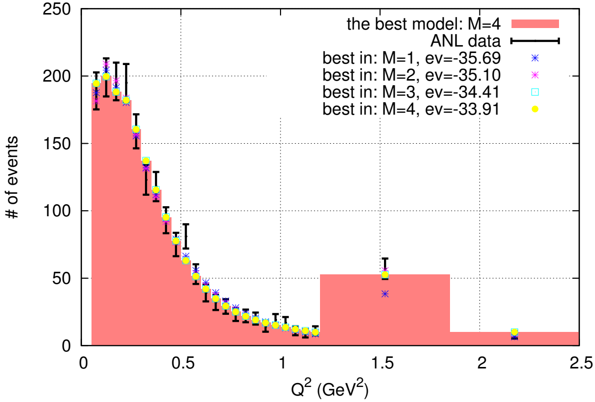
$$\phi(E_\nu) = p\frac{1}{\sigma(E_\nu, F_A)}\frac{dN}{dE_\nu}$$

-

$$\chi^2 = \left(\frac{F_A(0) - g_A}{\Delta g_A}\right)^2 + \sum_{i=k}^{n_{\mathrm{ANL}}} \frac{\left(N_i - N_i^{th}\right)^2}{N_i} + \left(\frac{1-p}{\Delta p}\right)^2 \qquad \Delta p = 20\%$$

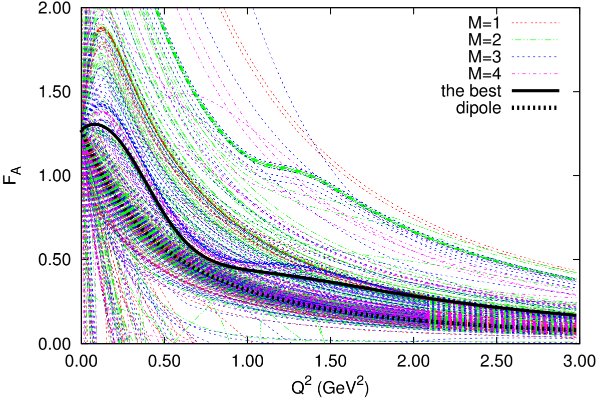
Numerical results

- About 17000 fits have been collected.
- MLPs with: $M = 1, 2, 3$ and 4, hidden units have been trained.
- Best model $\rightarrow$ maximal value of the evidence.
- All the best models within each MLP type reproduce well the ANL data.

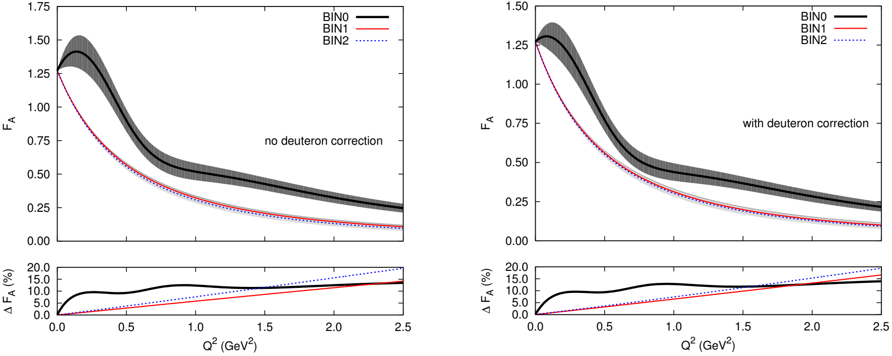


Numerical results

- About 17000 fits have been collected.
- MLPs with: $M = 1, 2, 3$ and 4, hidden units have been trained.
- Best model $\rightarrow$ maximal value of the evidence.
- All the best models within each MLP type reproduce well the ANL data.

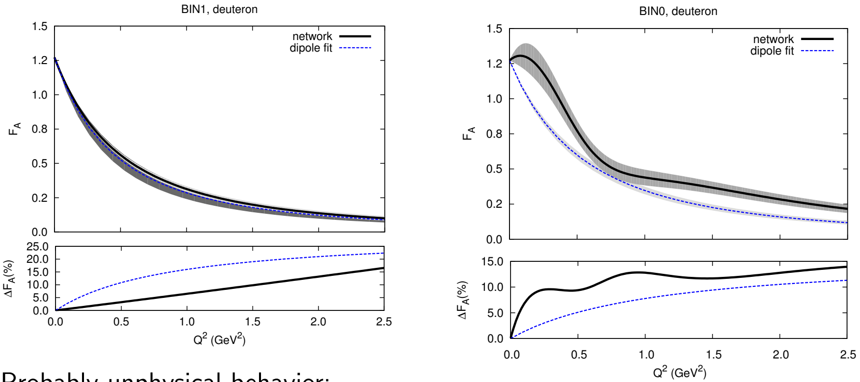


Best fits; $F_A(Q^2)$ functions:



- BIN0: all ANL bins included.
  - $r_A^2 < 0$ **incompatible with previous results.**
- BIN*k*: ANL bins without the first *k* bins.

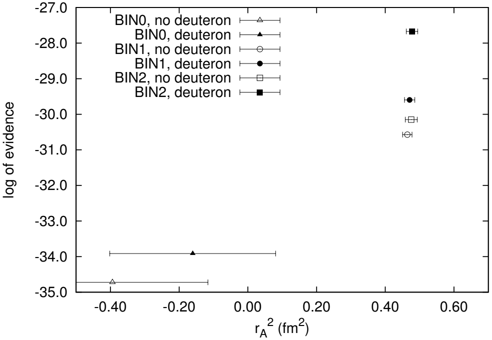
Comparison with dipole



Probably unphysical behavior:

- Improper description of the nuclear corrections.
- Low quality of the measurements at low-Q^2 $\rightarrow$ efficiency?
- Lack of very low-Q^2 data.

Dependence of r_A^2 on the evidence



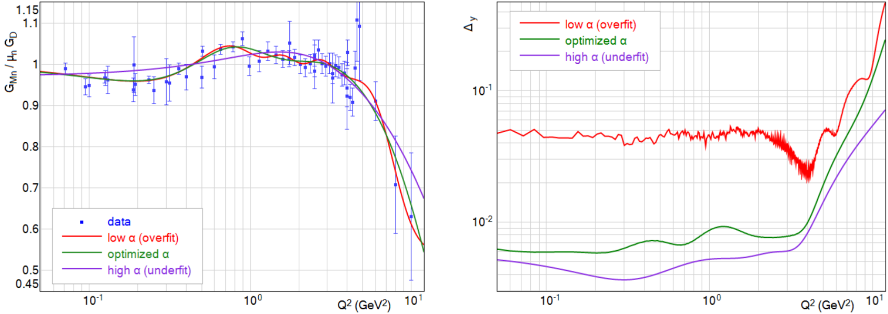
Best fit, for BIN2:
$r_A^2 = 0.478 \pm 0.017$ fm^2

- z-expansion:
  $r_A^2 = 0.46 \pm 0.22$ fm^2
  Meyer et al., PRD93 (2016)
- muon capture by protons:
  $r_A^2 = 0.43 \pm 0.24$ fm^2
  Hill et al., arXiv:1708.08462

- Small errors $\rightarrow$ optimal model.

Predictive power

Avoid overfit and underfit.



Fit of $G_{Mn}/\mu_n G_D$ data.

K. M. Graczyk, PRC84, 034314 (2011)

Conclusions

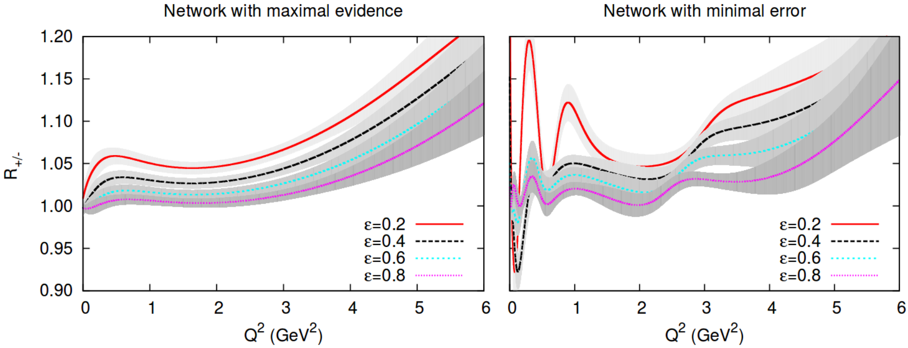
- First Bayesian analysis of the neutrino-deuteron scattering data.
- With the full ANL data set F_A has a local maximum at low Q^2.
- Inclusion of deuteron correction reduces the peak in F_A.
- Removing the lowest Q^2 region a value of r_A^2 consistent with available determinations could be obtained.
- Corrections from the deuteron-structure play a crucial role at low Q^2.
- Experimental errors in this kinematic region could be underestimated.
- Analyses without the low Q^2 data do not show any significant deviation from the dipole shape.
- New more precise measurements of neutrino cross section on hydrogen and deuterium are needed.

Thank for your attention!

Section 7

Backup

Maximal evidence vs minimal error



Fit of $R_{+/-} = \frac{d\sigma(e^+p\to e^+p)}{d\sigma(e^-p\to e^-p)}$.

K. M. Graczyk, C. Juszczak, J. Phys. G42, 034019 (2015)

Likelihood, prior and posterior densities:

- Prior: weights are Gaussian distributed:

$$\mathcal{P}(\{w_j\},\mathcal{N}) = \left(\frac{\alpha}{2\pi}\right)^{W/2} \exp\left(-\alpha\,\frac{1}{2}\sum_{i=1}^{W} w_i^2\right)$$

  α is the width of the prior.

- Likelihood in terms of χ^2:

$$\mathcal{P}(\mathcal{D}\mid\{w_j\},\mathcal{N}) = \frac{1}{N_L}\exp(-\chi^2)$$

- Take into account also

$$\mathcal{P}\left(F_A(0) = g_A \mid \{w_j\},\mathcal{N}\right) \sim \exp(-\chi^2_{g_A})$$

- Modification of the likelihood: $\chi^2 \to \chi^2_{ANL} + \chi^2_{g_A}$

Evidence

- Optimal configuration of $\rho_{MP} = \{\{w_j\}_{MP}, \alpha_{MP}\}$ close to

$$\mathcal{E} = \chi^2 + \alpha \, \frac{1}{2} \sum_{i=1}^{W} w_i^2$$

$\alpha \rightarrow$ plays the role of regularizer to deal with over-fitting.

Evidence for the model:

$$\ln \mathcal{P}\left(\mathcal{D} \middle| \mathcal{N}\right) \approx -\chi^2 - \alpha_{\mathsf{MP}} \frac{1}{2} \sum_{i=1}^{W} \{w_j\}^2_{\mathsf{MP}}$$
$$-\frac{1}{2} \ln \mid A \mid + \frac{W}{2} \ln \alpha_{\mathsf{MP}} - \frac{1}{2} \ln \frac{\gamma}{2} + M \ln(2) + \ln(M!).$$

Normalization factors common to all models are omitted.
$|A|$ denotes the determinant of the Hessian matrix:

$$A_{ij} = \nabla_i \nabla_j \; \chi^2 \Big|_{\{w_k\}=\{w_k\}_{\mathsf{MP}}} + \delta_{ij} \alpha_{\mathsf{MP}} \, .$$

The parameter

$$\gamma = \sum_{i=1}^{W} \frac{\lambda_i}{\alpha + \lambda_i},$$

measures the effective number of weights, whose values are controlled by the data.
The λ_is are eigenvalues of the matrix $\nabla_i \nabla_j \chi^2 \mid_{\vec{w}=\vec{w}_{MP}}$.

Best-fit parametrization for BIN0 with deuteron correction:

$$\begin{aligned}\mathcal{N}(Q^2,\{w_j\}) &= \frac{w_9}{e^{-Q^2 w_1 - w_2}+1} + \frac{w_{10}}{e^{-Q^2 w_3 - w_4}+1} + \frac{w_{11}}{e^{-Q^2 w_5 - w_6}+1} \\ &\quad + \frac{w_{12}}{e^{-Q^2 w_7 - w_8}+1} + w_{13}\,.\end{aligned}$$

- Contribution from four sigmoids (units).
- Given unit, typically, describes one particular feature of the function.
- If the data dependence is trivial then some units may describe the same features and can be similar in the response.

Weights w_{1-13}:

$$\begin{aligned}\{w_j\} = &\{-2.174061, 0.1991515, 2.140942, -0.1947798, -2.174070, \\ &0.1991740, -5.481409, 2.501837, -2.502352, 2.308397, \\ &-2.502347, 3.120895, -0.1638095\}\end{aligned}$$

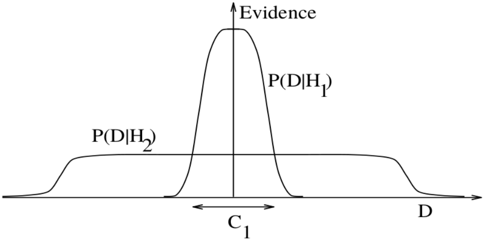
Bayesian framework for MLP

Evaluating the evidence:

- For many problems the posterior has a strong peak at $\mathbf{w}_{\mathrm{MP}}$

$$\underbrace{\mathcal{P}(\mathcal{D} \mid \mathcal{N}_M)}_{\text{Evidence}} \approx \underbrace{\mathcal{P}(\mathcal{D} \mid \mathbf{w}_{\mathrm{MP}}, \mathcal{N}_M)}_{\text{Best fit likelihood}} \underbrace{\mathcal{P}(\mathbf{w}_{\mathrm{MP}} \mid \mathcal{N}_M)\, \Delta \mathbf{w}}_{\text{Occam factor}}$$

- Bayes embodies Occam's razor:



- H_1 limited range of predictions.
- H_2 weaker prediction of C_1 data.

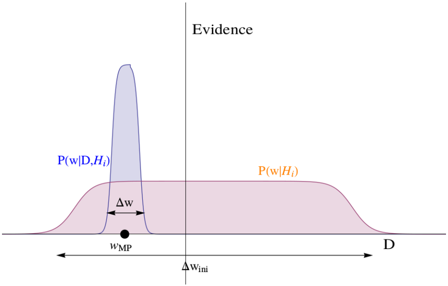
Bayesian framework for MLP

Evaluating the evidence:

- For many problems the posterior has a strong peak at $\mathbf{w}_{\text{MP}}$

$$\underbrace{\mathcal{P}(\mathcal{D} \mid \mathcal{N}_M)}_{\text{Evidence}} \approx \underbrace{\mathcal{P}(\mathcal{D} \mid \mathbf{w}_{\text{MP}}, \mathcal{N}_M)}_{\text{Best fit likelihood}} \underbrace{\mathcal{P}(\mathbf{w}_{\text{MP}} \mid \mathcal{N}_M)\, \Delta \mathbf{w}}_{\text{Occam factor}}$$

- Interpretation of the Occam factor for one parameter:



Occam factor penalizes complex models:

- Δw is the posterior uncertainty in w
- Assume uniform on large interval Δw_{ini}
- $\mathcal{P}(w_{\text{MP}} \mid \mathcal{N}_M) = \frac{1}{\Delta w_{ini}}$

$$\text{Occam factor} = \frac{\Delta w}{\Delta w_{ini}} \rightarrow \frac{V(\text{Posterior})}{V(\text{Prior})}$$

J.C. MacKay, Neural Computation 4, 415 (1992)