Two-Photon Exchange Effect Studied with Neural Networks

Krzysztof M. Graczyk*
Institute of Theoretical Physics, University of Wrocław, pl. M. Borna 9, 50-204, Wrocław, Poland
(Dated: June 8, 2011)

The novel approach to the extraction of the two-photon exchange (TPE) correction from the elastic *ep* scattering data is presented. The Bayesian framework for the neural networks is adapted. As the result the empirical fits of the TPE correction and electromagnetic form-factors are obtained. They are given by the one multidimensional function approximated by the feed forward neural network. In order to get a model independent approximation a large number of the different network architectures is considered. The Bayesian algorithm for choosing the best model is applied. A strong dependence of the TPE fit on the choice of parameterization is observed.

PACS numbers: 13.40.Gp, 25.30.Bf, 14.20.Dh, 84.35.+i
Keywords: proton form-factors, two-photon exchange correction, artificial neural networks

Study of the elastic *ep* scattering gives an opportunity to explore the structure of the proton. From the *ep* cross section data the magnetic (G_M), and electric (G_E) proton form-factors are obtained, via the Longitudinal-Transverse (LT) separation [1]. It is convenient to consider, in the data analysis, the reduced cross section which, in the one-photon exchange (OPE) approximation, reads

$$\sigma_{1\gamma,R}(Q^2,\epsilon) = \tau G_M^2(Q^2) + \epsilon G_E^2(Q^2), \quad (1)$$
$$\tau = Q^2/4M^2, \qquad \epsilon = \left(1+2(1+\tau)\tan^2(\theta/2)\right)^{-1},$$

where Q^2 and θ are the four-momentum transfer and scattering angle respectively.

The form-factor ratio $\mathcal{R}_{1\gamma}(Q^2) = \mu_p G_E(Q^2)/G_M(Q^2)$ (μ_p – the magnetic moment of the proton) can be extracted from the so-called polarization transfer (PT) measurements [2]. It turned out that there exists the systematic discrepancy between the form-factor ratio data obtained via the LT separation and the PT measurements. It seems that taking into account the TPE correction, the one which is not included in the classical treatment of the radiative corrections, cancels this discrepancy [3, 4]. Moreover it is claimed that the TPE correction to $\mu_p G_E/G_M$ ratio extracted from the PT measurements is negligible [3, 6]. But taking into account the TPE contribution in the LT separation: $\sigma_{1\gamma+2\gamma,R}(Q^2,\epsilon) \to \sigma_{1\gamma,R}(Q^2,\epsilon) + \Delta C_{2\gamma}(Q^2,\epsilon)$, affects significantly the values of the proton form-factors. For review of the TPE physics see [5] while the recent discussions of the TPE effect can be found in: [6, 7].

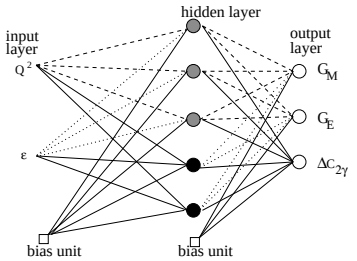

FIG. 1: Network of 2-(3-2)-3 type: 2 input units, one layer of hidden units, 2 output units. Notice that the FF sector (grey filled units and dashed connections) in contrast to TPE sector (black units and solid connections) is connected only with Q^2. Dotted lines denote the "switched off" connections. Every solid/dashed line represents one weight parameter.

The dominant TPE contribution, $\Delta C_{2\gamma}(Q^2,\epsilon)$, is given by the interference between the OPE and TPE amplitudes. Hence for the e^+p scattering $\Delta C_{2\gamma}(e^+p) \to -\Delta C_{2\gamma}(e^-p)$. Therefore the magnitude of the TPE term can be evaluated by measuring the ratio of the e^+p to e^-p elastic cross sections [8]: $\mathcal{R}_{+/-} = 1 - 2\Delta C_{2\gamma}/\sigma_{1\gamma+2\gamma,R}$. A deviation of this function from unity indicates the importance of the TPE effect.

The direct prediction of the proton form-factors and TPE correction is a difficult task. One has to deal with the problems of quantum chromodynamics in the non-perturbative regime. The successful approaches are rather phenomenological, and contain plenty of internal parameters, which are fixed to reproduce the experimental data (for review see [1, 5]).

On the other hand, the existing elastic polarized and unpolarized e^-p and e^+p scattering data, covers kinematical region broad enough to reconstruct the Q^2 dependence of the form-factors. Combining the cross section data with the PT measurements and the e^-p/e^+p ratio data allows to get an information about the TPE contribution. The aim of this paper is to find the parameterization of the form factors and TPE contribution relying only on the knowledge of the experimental data (without additional model dependent assumptions).

It is interesting that only three complex form factors are required to describe the scattering amplitude for the elastic unpolarized/polarized *ep* [3] scattering. They depend on Q^2 and ϵ. Hence six real functions have to be determined from the data. In this analysis we assume

*Electronic address: kgraczyk@ift.uni.wroc.pl

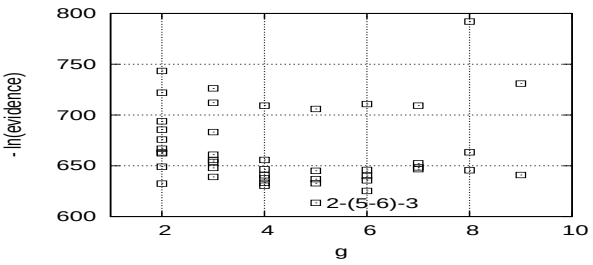

FIG. 2: Log of evidence.

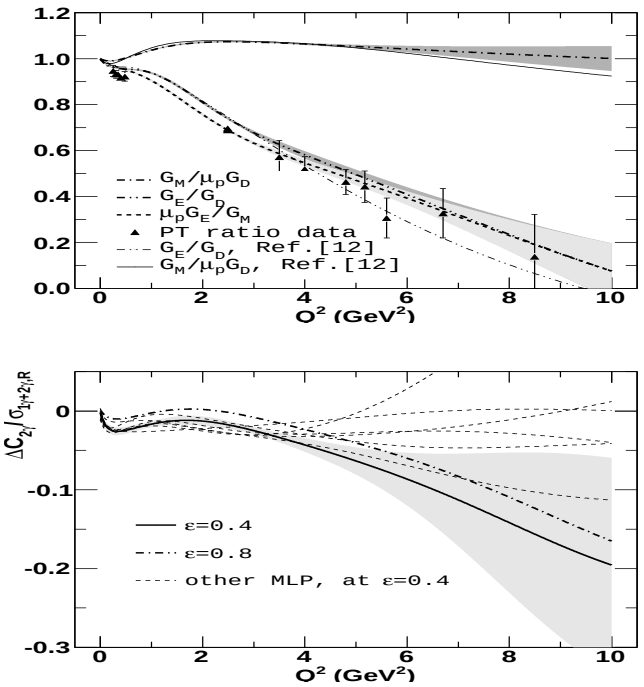

FIG. 3: Top panel: G_E/G_D, $G_M/\mu_p G_D$ ($G_D = 1/(1 + Q^2/0.71)^2$) and ratio $\mu_p G_E/G_M$. The predictions of the proton form-factors of Ref. [12] are also shown. The PT $\mu_p G_E/G_M$ data are taken from [2, 6, 11]. Shaded areas denote 1σ uncertainty. Bottom panel: the Q^2-dependence of the $\Delta C_{2\gamma}/\sigma_{1\gamma+2\gamma,R}$ at ϵ =0.4 and 0.8. The shaded area denotes 1σ uncertainty computed for the fit at $\epsilon = 0.4$. The dotted lines denote the TPE term predicted at $\epsilon = 0.4$ by the networks which have lower than the best fit evidence values.

that the PT ratio data is not affected by TPE. Then one can show that only three unknown functions have to be found: two proton form-factors and $\Delta C_{2\gamma}$ TPE function. In this analysis we consider also three different data types: the cross section, PT, and e^+p/e^-p ratio data. Using at least three different data types appeared to be necessary because lack of model assumptions about TPE term.

In order to approximate the form-factors and TPE function one has to assume particular empirical parameterization. However, it is obvious that the choice of the functional form of the parameterization has an impact on the fit and its uncertainties. In particular it is the case of the TPE contribution. This problem has not been discussed in the previous analyses.

In this approach fitting the data means the construction of the statistical model with ability to predict the from factors and TPE term. The Bayesian statistics allows us to compare different models. Indeed we consider as many different data parameterizations as possible and the best model is indicated by the objective Bayesian procedure. In practise one has to evaluate the probability distribution in the space of all functional parameterizations of the form-factors and the TPE contribution. The best model maximizes this probability. This idea has been already developed to approximate the electromagnetic form-factors [9] and, here, it is adapted to study the TPE effect.

In order to construct the statistical model, the space of G_M, G_E and $\Delta C_{2\gamma}$ functions is spanned by the artificial neural networks (ANN). Given ANN of architecture $\mathcal{A}_i$ represents the class of functions, which has an ability to represent the data. At the beginning of the analysis it is assumed that all possible models are equally likely. Then with the help of Bayes' theorem the posterior probability for given model (network) $\mathcal{P}(\mathcal{A}_i|\mathcal{D})$ is computed. $\mathcal{D}$ is the experimental data. In order to choose the most suitable approximation of the data, it is enough to evaluate the evidence $\mathcal{P}(\mathcal{D}|\mathcal{A}_i)$ – the probabilistic measure of goodness of the fit (Sec. 3.1 of [9]). The logarithm of evidence is given by two main contributions: the misfit of the approximate data (the experimental error function at the minimum) and the Occam factor. The latter penalizes complex models. The most optimal model has the highest value of the evidence.

It is obvious that the magnetic and electric form-factors as well as the TPE correction function are correlated. All of them should be determined by the same underlaying fundamental model. Therefore, one can imagine that there exists a multidimensional function, defined by the set of parameters, which simultaneously describes all G_M, G_E and $\Delta C_{2\gamma}$. In order to approximate such multidimensional map, we consider the particular type of ANN, the feed-forward neural network in the so-called multi-layer perceptron (MLP) configuration. In this analysis MLP maps 2-dimensional input space ($\vec{in} = (Q^2, \epsilon)^T$) to output space, spanned by three functions $\vec{out} = (G_M, G_E, \Delta C_{2\gamma})^T$. We consider MLP networks which consist of three layers of units: input, hidden layer of units and output. Using MLP with only one hidden layer is motivated by the Kolmogorov function superposition theorem. It states that every continuous function can be approximated by one-hidden layer feed forward neural network.

Each single neuron (unit) of network calculates its output value as an activation function f_{act} of the weighted sum of its inputs $f_{act}\left(\sum_i w_i \mu_i\right)$, where w_i denotes the i-

th weight parameter, while μ_i represents the output value of the unit from the previous layer. In this analysis the activation functions of the hidden units are given by the sigmoid function while for the output units the linear activation functions are applied. The process of finding the optimal values of the weights is called the training of the network.

Notice that MLPs with larger number of units (with many weights) have better ability to represent the data. However, usually too complex parameterizations exactly resemble the data and the generality of the description is lost (the data is over-fitted). Moreover the complex parameterizations lead to larger uncertainties then the simple models. On the other hand too simple parameterizations are not capable to code all the important information hidden in the measurements. Simple functions tend to underestimate the uncertainties. A task of finding the optimal statistical model which represents the data accurately enough, but does not overfit the data is known in the statistic as bias-variance trade off problem. In the previous global analyses of the *ep* data the degree of the complexity of the form-factor and TPE parameterizations was chosen with the help of phenomenological arguments and common sense. In this work we wish to replace the common sense with the objective Bayesian algorithm, while the phenomenological constraints are reduced to the necessary minimum.

The G_M and G_E depend only on Q^2. This property is achieved by the particular choice of the architecture of MLP, namely some of the connections are erased. As the result the network is divided into two sectors. One, called FF sector, is disconnected with ϵ input and the second, called TPE sector, is connected with both input values. The form-factors and TPE correction are still determined by the large subset of common weights. An example of 2-(3-2)-3 network ($\mathcal{A}_{3,2}$) is drawn in Fig. 1. It consists of: 2 input units, 5 units in hidden layer (3 units belong to FF sector, and 2 units belong to the TPE sector), and 3 output units.

In this work we consider the networks of type 2-(g-t)-3, where $4 \leq g + t = M \leq 12$. In the preliminary stage of the analysis, it has been observed that the networks with either g=1, or t=1 have been not able to approximate the data well (similarly as the networks with $M < 4$). Therefore we consider only models with $g, t > 1$. Finally we discuss 45 different ANN architectures. For every network $\mathcal{A}_{g,t}$ type the evidence is computed (see Fig. 2). The network 2-(5-6)-3 is obtained as the most optimal model. It has the highest evidence value.

Every map $\mathcal{A}_{g,t}$, which has been discussed in the model comparison, is defined by set of weight parameters which maximizes the posterior probability:

$$\mathcal{P}\left(\vec{w}|\ \mathcal{D}, \{\mathcal{I}\}, \mathcal{A}_{g,t}\right) = \frac{\mathcal{P}\left(\mathcal{D}|\ \vec{w}, \{\mathcal{I}\}, \mathcal{A}_{g,t}\right) \mathcal{P}\left(\vec{w}|\ \{\mathcal{I}\}, \mathcal{A}_M\right)}{\mathcal{P}\left(\mathcal{D}|\ \{\mathcal{I}\}, \mathcal{A}_M\right)}. \tag{2}$$

$\mathcal{P}\left(\mathcal{D}|\ \vec{w}, \{\mathcal{I}\}, \mathcal{A}_M\right)$ is the likelihood function of the data, while $\mathcal{P}\left(\vec{w}|\ \{\mathcal{I}\}, \mathcal{A}_M\right)$ denotes the prior probability. $\{I\}$ is the set of initial constraints.

The data likelihood function is defined by: $\mathcal{P}(\mathcal{D}|\vec{w}, \{\mathcal{I}\}, \mathcal{A}_{gt}) \sim \exp\left(-S_{ex}(\mathcal{D}, \vec{w})\right)$, where $S_{ex}(\mathcal{D}, \vec{w}) = \chi^2_\sigma + \chi^2_{PT} + \chi^2_{+/-} + \chi^2_{G_M} + \chi_{G_E}$ is the total error function. By $\chi^2_{\sigma,PT,+/-}$ we denote the error functions of: the cross section (27 sets, similarly as in [10]), PT (14 sets) and e^+p/e^-p ratio (3 sets) data. In the case of the cross section data, similarly as in [10], the systematic normalization uncertainties are taken into account: for every data set a normalization parameter is introduced and is fixed during the training. The selection of the PT ratio data is the same as in [10], but the two data sets are replaced with their recent updates [2]. Eventually we include also the latest PT measurements of the form-factor ratio [6, 11]. Since the presence of the PT ratio data is required to properly extract the TPE contribution, we consider only the cross section points below $Q^2 = 10$ GeV^2. Above this limit the PT data is not available. Eventually $\chi^2_{G_{M/E}}$ denotes the error function introduced to take into account the two artificial form-factor points (see discussion below).

We distinguish the ANN and physical initial constraints. The ANN constraints are introduced in order to face the over-fitting problem. Indeed, defining the prior probability as it follows: $\mathcal{P}\left(\vec{w}|\ \{\mathcal{I}\}, \mathcal{A}_{gt}\right) \sim \exp\left[-(\alpha/2)\sum_{i \in all\ weights} w_i^2\right]$ prevents from getting the over-fitted parameterizations. This kind of the prior assumption does not affect the final results [9]. Notice that the regularization parameter, α, is also established in the optimal way (see: [9], Sec. 3.1). The physical constraints are motivated by the general properties of the form-factors and the TPE term. We assume that: at $Q^2 = 0$, $G_M/\mu_p = G_E = 1$ and $\Delta C_{2\gamma}(\epsilon = 1) = 0$. In practice, three artificial data points are added to the experimental data sets, namely $(G_M(0)/\mu_p = 1, \Delta G_M(0) = \Delta)$, $(G_E(0) = 1, \Delta G_M(0) = \Delta)$ and $(\mathcal{R}_{+/-}(0,1) = 1, \Delta\mathcal{R}_{+/-}(0,1) = \Delta)$, where $\Delta = 0.01$.

The optimal values of the weight parameters are established by applying the quick-prop gradient descent training algorithm. In reality for every network $\mathcal{A}_{g,t}$ 10^3 networks, with the randomly chosen initial values of weights, have been trained. Among them, the parameterization with the highest evidence was taken for the further models comparison.

In Fig. 3 (top panel) we plot the form-factor ratio $\mathcal{R}_{1\gamma}$ computed with the network $\mathcal{A}_{5,6}$ (our best fit). The shaded areas denote 1σ uncertainty computed from the covariance matrix of the fit. Our predictions of the form-factors are compared to the results of Ref. [12] where the TPE function has been postulated based on the phenomenological arguments. The discrepancies between our fits and those of Ref. [12] appear above Q^2=4 GeV^2.

In the bottom panel of Fig. 3 the Q^2 dependence

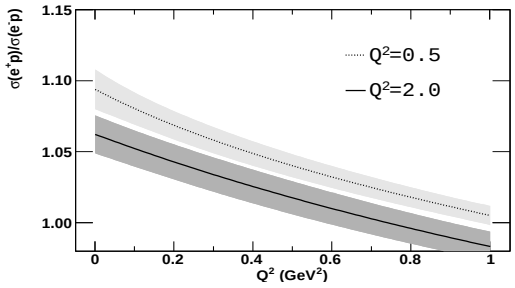

FIG. 4: Ratio $\mathcal{R}_{+/-}$ predicted by the network 2-(5-6)-3. The grey areas denote 1σ uncertainty.

of the ratio $\Delta C_{2\gamma}/\sigma_{1\gamma+2\gamma}$ is presented. We see that at $Q^2 \sim 0.2$ GeV^2 the TPE correcting term has a local minimum and it becomes an decreasing function of Q^2 above 2 GeV^2. With growing Q^2 fit uncertainty also enlarges. Indeed, above $Q^2 = 6$ GeV^2 the number of experimental points is limited and the data is not accurate enough to get an exact approximation. It is interesting to mention that for large ϵ (above 0.8) and Q^2 around 1.5 GeV^2 the TPE correction is positive.

In the bottom panel of Fig. 3 we plot also the TPE contribution (dotted lines) predicted by the models: $\mathcal{A}_{4,2}$, $\mathcal{A}_{4,3}$, $\mathcal{A}_{6,2}$, $\mathcal{A}_{6,3}$, $\mathcal{A}_{6,4}$, and $\mathcal{A}_{5,7}$. They are characterized by lower evidence values than $\mathcal{A}_{5,6}$ model, but they could be acceptable due to the χ^2 method (their χ^2_{min} values are much lower then number of points). The difference between these fits and prediction of $\mathcal{A}_{5,6}$ is spectacular. It demonstrates that the model comparison is crucial for the proper choice of the TPE parameterization.

Having in mind the forthcoming measurements of the elastic e^-p and e^+p scattering [8] in Fig. 4, we plot our predictions of the ratio $\mathcal{R}_{+/-}$. Even thought, we have not assumed the linearity of TPE term in ϵ, the final fit behaves like linear function of ϵ, as it has been observed in plenty of previous global analyses [5].

The obtained TPE function has particular analytical form (see [14]), which can be written as the Taylor series in ϵ. If one neglects higher than linear ϵ terms then the TPE correction is the sum of two contributions, which play the particular role in the LT separation. The one corrects the magnetic form-factor and it appears to be negative. The other modifies the electric form-factor and it is the positive function of Q^2.

The aim of this work was to extract from the elastic *ep* scattering data the proton form-factors and TPE function. It was done by adapting the Bayesian statistical methods developed for the feed forward neural networks. The phenomenological constraints were limited to the necessary minimum. The final form-factors and TPE parameterizations were chosen among the large set of possible models. Therefore the obtained empirical parameterizations are model independent and unbiased.

The analytical form of the fits and covariance matrix can be taken from [14]. All numerical computations have been done with the use of the C++ library developed by K.M.G.

Acknowledgements

The work has been supported by the Polish Ministry of Science Grant, project number: N N202 368439.

We thank to C. Giunti for inspiring discussions in the early stage of the project. We thank also J. Zmuda and J. Nowak for reading the manuscript.

[1] C. F. Perdrisat, V. Punjabi and M. Vanderhaeghen, Prog. Part. Nucl. Phys. **59** (2007) 694.

[2] G. Ron *et al.*, *Low Q^2 measurements of the proton form factor ratio mu_pG_E/G_M*, arXiv:1103.5784 [nucl-ex]; A. J. R. Puckett *et al.*, *Reanalysis of Proton Form Factor Ratio Data at $\mathbf{Q^2}$ = 4.0, 4.8, and 5.6 GeV^2*, arXiv:1102.5737 [nucl-ex].

[3] P. A. M. Guichon and M. Vanderhaeghen, Phys. Rev. Lett. **91** (2003) 142303.

[4] P. G. Blunden, W. Melnitchouk and J. A. Tjon, Phys. Rev. Lett. **91** (2003) 142304.

[5] J. Arrington, P. G. Blunden and W. Melnitchouk, *Review of two-photon exchange in electron scattering*, arXiv:1105.0951 [nucl-th].

[6] M. Meziane *et al.* [GEp2gamma Collaboration], Phys. Rev. Lett. **106** (2011) 132501.

[7] J. Guttmann, N. Kivel, M. Meziane and M. Vanderhaeghen, arXiv:1012.0564 [hep-ph]; D. Borisyuk and A. Kobushkin, Phys. Rev. D **83** (2011) 057501; I. A. Qattan and A. Alsaad, Phys. Rev. C **83** (2011) 054307.

[8] J. Mar *et al.*, Phys. Rev. Lett. **21** (1968) 482; A. Browman, F. Liu, and C. Schaerf, Phys. Rev. 139 B1079 (1965); D. Yount, J. Pine, Phys. Rev. 128, (1962) 1842.

[9] K. M. Graczyk, P. Plonski and R. Sulej, JHEP **1009** (2010) 053.

[10] W. M. Alberico, S. M. Bilenky, C. Giunti and K. M. Graczyk, Phys. Rev. C **79** (2009) 065204.

[11] A. J. R. Puckett *et al.*, Phys. Rev. Lett. **104** (2010) 242301.

[12] J. Arrington, W. Melnitchouk and J. A. Tjon, Phys. Rev. C **76** (2007) 035205.

[13] J. Arrington *et al.*, VEPP-3 experiment, arXiv:nucl-ex/0408020; JLab experiment E04-116.

[14] `http://www.ift.uni.wroc.pl/~kgraczyk/nn.html`