Studying total proton–proton cross section collision at large hadron collider using gene expression programming

A Radi1,2
1Department of Physics, College of Sciences, Sultan Qaboos University, Al Khoudh, Muscat 123, Oman
2Department of Physics, Faculty of Sciences, Ain Shams University, Abbassia, Cairo 11566, Egypt
E-mail: Amr.radi@cern.ch

Abstract. New technique is presented for modeling total cross section of proton-proton (p-p) collision from low to ultra-high energy regions using gene expression programming (GEP). GEP, as a machine learning technique is usually used for modeling physical phenomena by discovering a new function $\sigma_T(\sqrt{s})$. In case of modeling the p-p interactions at the Large Hadron Collider (LHC), GEP is used to simulate and predict the total cross-section which is a function of total center-of-mass from low to high energy $\sqrt{s}$. The discovered function shows a good match as compared with the other models. The predicted values of total cross section are in good agreement with Particle Data Group (PDG).

1. Introduction
The study of the interactions between elementary particles requires high energy collisions as in LHC [1-3]. LHC possibly will yield the way in the direction of our awareness of particle physics beyond the Standard Model. The proton-proton interaction is one of the fundamental interactions in high-energy physics. In order to fully exploit the enormous physics potential, it is important to have a complete understanding of the reaction mechanism.

Recently, different modeling methods, based on soft computing systems, include the application of evolution algorithms [4-7]. GEP [8] can be used as an alternative tool for the simulation of these interactions [5-8]. GEP is a global optimization algorithm and an automatic programming technique that had been applied in particle physics [9-10]. It is a recently developed evolutionary computation method for function discovery and data analysis. GEP uses population of individuals, selects them according to fitness, and produces genetic variation using one or more genetic operators [11].

In this paper, we have discovered the functions of the total cross section $\sigma_T(\sqrt{s})$ of p-p collision from low to ultra-high energy regions using the GEP technique. This paper has five sections. Section 2, gives a review to the basics of the GEP technique. Section 3 explains how GEP is used to model the p-p collision. Finally, the results and conclusion are provided in sections 4 and 5 respectively.

2. An overview of gene expression algorithms
GEP stores the individuals as symbolic strings (linear) which are then presented as an expression tree for evaluation [9]. GEP, provides the possible solutions to a given optimization problem, using the Darwinian principle of survival of the fittest. It uses biologically inspired operations like replication, recombination and mutation. In GEP, the chromosome or genome composed of a linear, symbolic string of fixed length, each consists of one or more genes. Each gene can be represented by an algebraic expression and are
composed of a head and a tail. The head contains symbols that represent both functions \( F = \{\text{sqr}, -, +, *, /, \text{ or other operators}\} \) and terminals \( T = \{x, y, z, \text{ random constants}\} \), whereas the tail contains only terminals. As an example, the algebraic expression: \( \text{sqr}(x-y)+x*z \) can be represented as an Expression Tree (ET) as shown in figure 1.

![Expression Tree](image)

**Figure 1.** An example of gene ET.

Here \( \{Q, -, +, *\} \) represents the head \( h \), “Q” represents the square root function, and \( \{x, y, x, z\} \) represents the tail \( t \).

Each gene individual in the population is assigned a fitness value, which quantifies how well it performs in solving the problem. The fitness value is computed by a problem dependent fitness function. The fitness function that can be applied for evaluating performance of generated solution are measured with the summed square of residuals (SSE), R-square is the ratio of the sum of squares of the regression (SSR) and MSE is the mean square error or the residual mean square [12].

### 3. The proposed GEP for the p-p collision

Total cross sections are proposed with two individual GEP models. The model is trained/predicated using experimental data to simulate the p-p collision that calculates the total cross sections \( \sigma_T(\sqrt{s}) \).

GEP has the potential to discover a new model, to show that the data sets are subdivided into two sets (training and predication). GEP, as configured in table 1, runs until the fitness function is reduced to an acceptable level (MSE = 0.0005). GEP discovers a new model from the training set. The predicated set is used to examine the generalization capabilities of the model.

| GEP Parameters       | \( \sigma_T(\sqrt{s}) \) |
|----------------------|--------------------------|
| Generations          | 800                      |
| Populations          | 30000                    |
| Function set         | \{*, /, +, ln, sqrt\}     |
| Terminal Set         | \{constant, \( \sqrt{s} \)\} |
| Fitness function     | MSE                      |
| Selection method     | Elits and rank           |
| Mutation rate        | 0.01                     |
| Crossover rate       | 0.9                      |

Table 1. GEP optimal parameters that were obtained.
4. Results
The discovered function is tested using the experimental data of the total cross sections $\sigma_T$ with the center-of-mass energy $\sqrt{s}$. The training data considers both the experimental observations at the LHC in the range (10 GeV – 8 TeV) and the Particle Data Group for p-p collision experiments [13-17].

The discovered function has been tested to associate the input patterns to the target output patterns using the error function. The final discovered function for describing the total cross sections $\sigma_T$ with the center-of-mass energy is given by

$$\sigma_T(\sqrt{s}) = ((\text{sqrt}((\sqrt{s} + \exp(7))) + 7.5) + \sin((\text{sqrt}((\sqrt{s}))) + 1/((\sqrt{s}/2))))$$  \hspace{1cm} (1)

The model is in excellent agreement with the extrapolation from low to ultra-high energy regions as in figure 2. In figure 3, the values of the total cross-sections $\sigma_T(\sqrt{s})$ is compared with results of the center-of-mass energy $\sqrt{s}$ between 10 GeV and the higher energies 50000 GeV from cosmic rays [17].

![Figure 2](image1.png)  \hspace{0.5cm} ![Figure 3](image2.png)

**Figure 2.** The progress of the total cross sections $\sigma_T$ function of the centre-of-mass energy $\sqrt{s}$ from low to ultra-high energy regions.

**Figure 3.** The progress of the total cross sections $\sigma_T$ function of the centre-of-mass energy $\sqrt{s}$ between 10-50000 GeV.

In table 2, the elastic scattering cross-section was subtracted from the total cross section to obtain a value for the inelastic cross-section which can then be compared/verified with the measurements of the CMS [13], ATLAS [14], ALICE [15] and TOTAM [16] experiments.

| Experiment | Inelastic cross-section (mb) | Elastic cross-section (mb) | Total cross-section calculated (mb) | Total cross-section predicted (mb) using GEPT ($\sqrt{s}$) |
|------------|------------------------------|----------------------------|------------------------------------|---------------------------------|
| $\sigma$ (TOTOM) | 73.5 (± 0.6, +1.1) | 24.8 (± 1.4) | 98.3 (± 3.0) | 98.1 |
| $\sigma$ (CMS) | 68.0 (± 8.4) | | | |
| $\sigma$ (ATLAS) | 69.4 (± 9.1) | | | |
| $\sigma$ (ALICE) | 72.7 (± 6.2) | | | |
| $\sqrt{s} = 8$ TeV | | | 102.9(± 3.0) | 103.1 |

Table 2. Results of the TOTEM measurements at the LHC energy of $\sqrt{s} = 7$ & 8 TeV [16].
Moreover, the discovered function $\sigma_T(\sqrt{s})$ predicts the total cross sections $\sigma_T$ at $\sqrt{s} = 7$ TeV which is 98.1 mb. This value is the same as experimental of TOTOM (98.3 mb) [16]. The predicted values of $\sigma_T(\sqrt{s})$ at LHC are 103.1 mb and 127.3 mb at $\sqrt{s} = 8$ TeV, and $\sqrt{s} = 14$ TeV, respectively. Those results similar to Nakamura and Block [17, 19, 20].

5. Conclusions

The current work presents GEP as a new technique for constructing functions of the total cross sections, $\sigma_T(\sqrt{s})$ of p-p collision. The discovered functions show good match to the experimental data. Moreover, they are capable of predicting experimental data for $\sigma_T(\sqrt{s})$ that are not used in the training session. Consequently, the predicted values of $\sigma_T(\sqrt{s})$ at the LHC energy in terms of the same parameters are in good agreement with the cosmic-ray experimental data from Particle Data Group. GEP has become one of important research areas in the field of high energy physics.

References

[1] Khachatryan V et al 2014 CMS Collaboration Eur. Phys. J. C 74 3053.
[2] Aaboud M et al 2016 ATLAS Collaboration Phys. Lett. B 761 158-178.
[3] TOTAM Collaboration and ATLAS Collaboration 2016 arXiv:1609.08847v3 [hep-ph].
[4] Eiben A, Smith A, Introduction to Evolutionary Algorithms (Springer, Berlin, 2003).
[5] El-Bakry S and Radi A 2007 Int. J. Mod. Phys. C 18 351.
[6] El-dahshan E, Radi A, El-Bakry M 2009 Int. J. Mod. Phys. C 20 1817.
[7] Whiteson S, Whiteson D 2009 Eng. Appl. Artif. Intel. 22 1203.
[8] Ferreira C 2006 Gene Expression Programming: Mathematical Modeling by an Artificial Intelligence, 2nd Edition, Springer-Verlag Germany.
[9] Teodorescu L, High energy physics data analysis with gene expression programming. In 2005 IEEE Nuclear Science Symposium Conference Record, Vol. 1, pp. 143-147, (2005).
[10] Teodorescu L and Sherwood D 2008 Computer Physics Communications 178 (6) 409–419.
[11] Ferreira C 2001 “Gene expression programming in problem solving” WSC6 tutorial.
[12] Cohen J, Cohen P, West S and Aiken L 2003 applied multiple regression/correlation analysis for the behavioral sciences. Mahwah, NJ: Erlbaum.
[13] CMS-Collaboration 2011 Performance Analysis Note, CMS-PAS-FWD-11-001.
[14] Aad G et al 2011 arXiv: 1104. 0326 [hep-ex] (2011).
[15] Poghosyan M G et al 2011 arXiv:1109.4510 [hep-ex].
[16] TOTEM Collaboration2011 EPL 96 21002.
[17] Nakamura K et al 2010 J. Phys. G 37 075021.
[18] Cudell J et al 2002 Phys. Rev. Lett. 89 201801.
[19] Block M M 2011 arXiv: 1109, 2940 [hep-ph].
[20] Block M M and Halzen F 2005 hep-ph/0506031.