


Optimizing the Matrix Element Method with Deep Neural Networks for High-Energy Physics Analyses

Olokunboyo A. Olaiya 

Department of Physics and Astronomy, University of New Hampshire, Durham, USA

Email: olokunboyoabiola@gmail.com

How to cite this paper: Olaiya, O.A. (2025) Optimizing the Matrix Element Method with Deep Neural Networks for High-Energy Physics Analyses. *Journal of High Energy Physics, Gravitation and Cosmology*, 11, 802-814.

<https://doi.org/10.4236/jhepgc.2025.113051>

Received: March 28, 2025

Accepted: July 4, 2025

Published: July 7, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The Matrix Element Method (MEM) is a widely used algorithm in experimental and theoretical high-energy physics (HEP) analyses. The MEM is based on the Lagrangian method to assess the compatibility of experimental events with a hypothetical process. The matrix element is then regressed with the help of a Deep Neural Network (DNN). The integration results can be approximated with DNN, which makes it possible to use the MEM for parameter scans, innovative physics searches, and other uses. The DNN can be trained to nearly mimic the outcomes of the direct numerical integration of the matrix element. It is possible to identify the specific ways in which the parton shower impacts the analysis by using the method to analyze these occurrences using fixed-order calculations. The method can, in theory, be applied to any measurement; however, processes involving intermediate resonances and leading to many-particle final states are anticipated to yield the highest improvement compared to cut-based analysis techniques. This research provides insights to understand the importance of optimizing machine learning strategies and fine-tuning regressions while also exploring the use of symmetry in particle physics with an interest in particle interactions and decays.

Keywords

Deep Neural Network, Matrix Element Method, Regressions

1. Introduction

Development in experimental and theoretical high-energy physics (HEP) analyses now employ the instrumentation of modern machine learning methods, involving deep learning, which are rapidly gaining applicability. Machine learning, a widely

used automated inference procedure in HEP, is often associated with multivariate techniques like Boosted Decision Trees (BDT) [1]. In recent years, the set of methods and tools commonly employed in HEP has expanded significantly due to the deep learning revolution. To fully leverage the information contained in a given event, such as those generated by the Large Hadron Collider (LHC), statistical analysis necessitates enhancements in machine learning analysis techniques [2]. Neural networks and BDT are used in HEP to learn from simulations or data. Nonetheless, difficulties encompass the acquisition and generalization of physical data, employing control samples or regularization strategies, and curtailing reverse engineering possibilities because of indirect training data [3] [4].

In contrast, the Matrix Element Method (MEM) regression leverages our understanding of the Standard Model (SM) through Lagrangian to assess the compatibility of experimental events with a hypothetical process [2]. The inherent understanding of the physics process and detector response renders the internal dynamics of MEM more straightforward to interpret compared to machine learning techniques. Moreover, because there is no training step involved, MEM can be effectively utilized even when the dataset contains very few events, a scenario in which other methods encounter difficulties. The MEM, derived from Tevatron experiments, is widely used in particle physics analyses. It's used at the LHC for searches and measurements of processes like top quark and anti-quark pair production and Higgs boson decay ($t\bar{t}H$) and spin correlation. However, its complexity and computational demands make it more CPU-intensive than other methods [5]. The result of matrix element integration can be approximated with a Deep Neural Network (DNN), which makes it possible to use the MEM for parameter scans, innovative physics searches, and other uses. In the absence of a closed or computationally feasible form, the probabilities given by MoMEMta can be understood as indistinguishable functions of the 4-momentum. Given appropriate assumptions and a sufficiently large width divided across several layers, a neural network can approximate any function. While there's no guarantee that the MEM's output conforms to all of these presumptions, it sufficiently motivated this study. In comparison to matrix element evaluation based on direct integration, this method could produce much quicker results. DNNs often take significantly less computation time than classical MEM integrations, requiring a simulated sample and several hours of training. Advancements in the MEM field, such as GPU acceleration, parallel computing, BDT, neural networks, and normalizing flows, can reduce computation times and avoid integration variable optimizations [2].

In this study, I would like to understand the details of the computational/probabilistic distribution required for MEM, its fitting with DNN, topology using the Drell-Yan weights model, and application to real-life analysis for certain physical processes based on experimental data [2].

2. Methodology

- 1) Regression is a machine learning approach that is used to predict the contin-

uous outcomes of interaction and decays based on the input variables. DNNs are simply types of neural networks having many layers between the output and input layers, which permits learning of difficult patterns in the data [5] [6]. To train the neural network, input features—such as particle momenta, energies, and other pertinent variables—derived from experimental data are fed into the system. By using theoretical calculations or Monte Carlo simulations to acquire the true distribution, the network is trained to minimize the difference between its predicted probability distribution and the true distribution. To improve its predictions, the network iteratively modifies its internal parameters (weights and biases) through techniques like gradient descent and backpropagation [7]. From first-principles computations for high-energy processes in a power series expansion order by order, we can simplify the expansion as in:

$$F(\mathbf{x}) = f_{00}(\mathbf{x}) + \alpha f_{01}(\mathbf{x}) + \alpha^2 \{f_{11}(\mathbf{x}) + f_{02}(\mathbf{x})\} + \dots \quad (1)$$

in which case $\alpha \ll 1$ is a small expansion parameter. Where the functions f_{ij} are related to the real part of a complex number, $f_{ij} \in \mathbb{R}(\mathbf{M}_i^* \mathbf{M}_j)$ and $i+j$ gives the power of the expansion coefficient, that is a generic term such that $\alpha^{i+j} f_{ij}$ ($\mathbf{M}_i \in \mathbb{C}$ with $i \in \{0, 1, 2, \dots\}$ is called the matrix element). The number of closed loops contributing to the Feynman graph is represented by the subscript, which is the total for a particular scattering process used in computing the matrix element. Here, the second-order terms (the order is given by the power of the α) $\{f_{11}(\mathbf{x}) + f_{02}(\mathbf{x})\}$ is the term of interest, where f_{02} is computationally space- and time-expensive than f_{11} . This difference results from the Feynman graphs, where f_{02} has two loops and f_{11} has one loop. To expedite the computation of the functions, constructing a regressor with a representative sample is an easy fix. Speeding up Monte Carlo simulations is highly relevant not only in particle physics but in a very large set of problems addressed by all branches of physics using perturbative expansion like the one in Equation (1) [8]. However, the regressors must generate very highly accurate predictions over the whole function domain in order to reach the precision required for matching with experimental data. The specifications for the regressors, specifically pertaining to the definition of high precision, are as follows:

- 1) High precision: prediction error $< 1\%$ over more than 90% of the domain of the function.
- 2) Speed: prediction time per data point of $< 10^{-4}$ seconds.
- 3) Lightweight: the disk size of the regressors should be a few megabytes at the most for portability.

Compared to the Monte Carlo statistical error, for instance, the prediction error—whose exact definition is provided in Equation (2)—guarantees that propagating the approximation mistake on f_{02} to the entire function F stays a sub-leading source of error.

$$\delta = \frac{f(\mathbf{x})_{\text{predicted}} - f(\mathbf{x})_{\text{true}}}{f(\mathbf{x})_{\text{true}}} \quad (2)$$

where δ gives the difference between the predicted value of the function $f(\mathbf{x})_{predicted}$, and its true value $f(\mathbf{x})_{true}$, normalized by $f(\mathbf{x})_{true}$ [8].

The terms that emerge in the generation of four charged leptons in proton-proton collisions, $pp \rightarrow l_1^+ l_1^- l_2^+ l_2^-$, where each lepton pair is mediated by an electrically neutral electroweak gauge boson, such as a Z boson or a photon (γ), are represented by the functions f_{ij} in Equation (1). This process's leading-order function (f_{00}) equals the square of the scattering amplitude at the tree level. Higher terms have graphs with internal legs that form a closed loop, whereas graphs without such legs are connected with a tree-level amplitude. For instance, adding the second-order term results in a time penalty of 1500 times greater accuracy in predicting the rate of creation of four electrons [8]-[10].

2) Let's talk about asymmetry. The functions in question are maps $f_{ij}^{(n)} \equiv f_{ij} : \mathbb{R}^n \rightarrow \mathbb{R}$, where $n \in \{2, 4, 8\}$ and $i, j \in \{0, 1, 2\}$. The feature space, or domain of the functions, is filled in from a uniform distribution and linearly transferred to the unit hypercube [10]. For instance, the physical phase-space coordinates in the 4-dimension are: the process total energy $\sqrt{s_{12}}$, having a di-boson system scattering angle of θ^* , with masses of the two bosons, m_{34} and m_{56} . Since the bosons are believed to be off their mass shell, these masses are not fixed in general [8] [9]. The 8-dimension case gives additional angles with leptons in the final state, and the 2-dimension case gives fixed masses. As a result, various physics analyses will call for various regressors. The 4D regressors, on the other hand, are more versatile and do not have any fixed physical parameters [8]. The smaller number of necessary functions in the third column of **Table 1** is obtained by leveraging the symmetry properties discussed below derived from physics domain knowledge. For the data used in Ref. [8], it stems from the symmetries manifesting in the scattering process that was simulated. In **Table 2**, the last column indicates whether summing the functions has a physics meaning; in the cases where it does, *i.e.* 2D and 8D, only the single regressor of the sum of the functions is required. The complete functions, $f_{ij}^{(n)}$, for any dimension n , is full. Certain combinations of the external particles that the method specifies allow pairs of functions to be mapped into each other. Thus, in feature space, the second coordinate, x_2 , is subject to a linear transformation both separately and in conjunction with the permutation of the third and fourth coordinates, x_3 and x_4 . For instance, in 4D, the number of independent functions is reduced from 162 to 25 by two permutations, $\pi_{12} := p_1 \leftrightarrow p_2$ and $\pi_{34} := p_3 \leftrightarrow p_4$ as seen in **Table 3**. Here, p_i represents a particle with label i .

Table 1. The dimension determines how many of the resulting functions-technically known as helicity amplitudes-there are [8].

Variable	2D	4D and 8D
$\sqrt{s_{12}}$	$[m_{34} + m_{56}, 1 \text{ TeV}]$	$[m_{34} + m_{56}, 3 \text{ TeV}]$
$\cos \theta^*$	$[-1, 1]$	$[-1, 1]$
m_{34}, m_{56}	$m_Z = 91.1876 \text{ GeV}$	$[50, 130] \text{ GeV}$

Table 2. Symmetry characteristics minimize the quantity of necessary functions [8].

Dimensionality	Total functions	Independent functions	Sum is physical?
2D	18	5	Yes
4D	162	25	No
8D	8	4	Yes

Table 3. Symmetry representation [8].

Permutation	Particle Symmetry	Coordinate Symmetry
π_{12}	$p_1 \leftrightarrow p_2$	$x_2 \rightarrow 1 - x_2$
π_{34}	$p_3 \leftrightarrow p_4$	$x_2 \rightarrow 1 - x_2$ and $x_3 \leftrightarrow x_4$

3) Describing the matrix method. The methodology can, in theory, be applied to any measurement; however, processes involving intermediate resonances and leading to many-particle final states are anticipated to yield the highest improvement in comparison to cut-based analysis techniques. Generally speaking, the MEM allows for the simultaneous determination of multiple unknown parameters (both experimental and theoretical parameters characterizing the detector response and the physics processes being measured) in a single measurement, thereby reducing systematic uncertainties [11]. The possibility that L_{sample} will see a sample of chosen events in the detector is the foundation of the Matrix Element technique. The likelihood is computed as a function of the assumed values for each of the parameters to be monitored and is derived directly from the theoretical prediction for the detector resolution and the differential cross-sections of the relevant processes. The measurement of the parameters is obtained by minimising $-\ln L_{sample}$, where the likelihood L_{sample} for the total event sample is calculated by multiplying the likelihoods of observing each individual event [11] [12]. The majority of analysis techniques in experimental particle physics, in contrast, compare distributions from observed events in the detector with corresponding distributions from simulated events generated theoretically, passed through a detector simulation, and then reconstructed using the same event reconstruction software [13]. The sample likelihood L_{sample} for N measured events to have measured properties x_1, \dots, x_N can be written as:

$$L_{sample}(x_1, \dots, x_N; \vec{\alpha}, \vec{\beta}, \vec{f}) = \prod_{i=1}^N L_{evt}(x_i; \vec{\alpha}, \vec{\beta}, \vec{f}) \tag{3}$$

where the symbol $\vec{\alpha}$ denotes assumed values of the physics parameters to be measured, $\vec{\beta}$ stands for parameters describing the detector response that are to be determined, and \vec{f} is defined below. The likelihood $L_{evt}(x_i; \vec{\alpha}, \vec{\beta}, \vec{f})$ to observe event x_i under the assumption of parameter values $\vec{\alpha}, \vec{\beta}$ and \vec{f} is given as the linear combination:

$$L_{evt}(x_i; \vec{\alpha}, \vec{\beta}, \vec{f}) = \sum_{Processes, P} f_P L_P(x_i; \vec{\alpha}, \vec{\beta}) \quad (4)$$

where the sum is over all individual processes P that could have led to the observed event x_i , $L_P(x_i; \vec{\alpha}, \vec{\beta})$ is the likelihood to observe this event under the assumption that it was produced via process P , and f_P denotes the fraction of events from process P in the entire event sample, with $\sum_P f_P = 1$. In total, the physics parameters $\vec{\alpha}$, the detector response described by $\vec{\beta}$, and the event fractions \vec{f} are to be determined simultaneously from the minimization of $-\ln L_{sample}$ [12].

The MEM is used to calculate $P(x|\alpha)$, which is the likelihood of observing an experimental event in light of theoretical hypothesis α . Here, x stands for the 4-momenta of any number of particles that have been detected in their final condition. The aim is to differentiate the parton level particles y generated at the contact point prior to hadronization and their detection from the experimentally observed particles x . α can be used to denote various models or any set of parameters (such as the mass of a resonance) [2]. The probability that a hard scattering will result in a partonic final state y for hadron colliders is proportional to the differential cross section, which is calculated as

$$d\sigma(q_1, q_2, y) = \frac{(2\pi)^4 |\mathbb{M}(q_1, q_2, y)|^2}{q_1 q_2 s} d\Phi(y) \quad (5)$$

where q_1 and q_2 are the initial state parton momentum fractions and s is the center-of-mass energy. $d\Phi(y)$ is the n-body phase space of the final state y , while $|\mathbb{M}(q_1, q_2, y)|^2$ denotes the matrix element for the given process α .

The parton distribution functions (PDF) $f_a(q)$ (for each parton q of flavor a), efficiency $\epsilon(y)$ to reconstruct and select the hadronic state y , and transfer function $T(x|y)$ normalized with respect to x are all involved in the transmission of the parton-level 4-momenta y to the experimentally observed ones x . The latter parameterizes the hadronization, the parton shower, and the detector response (which results in a blurring of the momenta of the detected particles due to its limited resolution) [2] [13]. A convolution of the differential cross section with the transfer function and a total over the initial states yields the probability $P(x|\alpha)$:

$$P(x|\alpha) = \frac{1}{\sigma_\alpha^{obs}} \int_{q_1, q_2} \sum_{\alpha_1, \alpha_2} \int_y dq_1 dq_2 f_{\alpha_1}(q_1) \times f_{\alpha_2}(q_2) \frac{(2\pi)^4 |\mathbb{M}(q_1, q_2, y)|^2}{q_1 q_2 s} T(x|y) \epsilon(y) \quad (6)$$

where σ_α^{obs} stands for the visible/observed cross-section and ensures that the probability is normalized. It is often computed *a posteriori* with $\sigma_\alpha^{obs} = \sigma_\alpha \langle \epsilon \rangle$, where $\sigma_\alpha = \int d\sigma_\alpha(y)$ is the total cross-section and $\langle \epsilon \rangle$ is the expectation of the selection efficiency. Define the MEM weights as $W(x|\alpha) = \sigma_\alpha^{obs} \times P(x|\alpha)$. In addition, the weights can span several orders of magnitude which is why most of the

time we will use the event information defined as $I_\alpha = -\log_{10}(P(x|\alpha))$ or $I'_\alpha = -\log_{10}(W(x|\alpha))$ which only differs by a constant term. Several complex processes are involved in the transfer function, and in order to make its integration easier, a number of assumptions are typically made. It is possible to factorize the transfer function for the various particles since each particle's detection and measurement in the final state are independent, in first order. By factorizing the various parts of the measured 4-momentum, this argument can be advanced even further [2]. Thus, the transfer function is written as

$$T(x|y) = \prod_{i=1}^n T_i(x^i|y^i), \tag{7}$$

$$T(x^i|y^i) = T_i^E(x^i|y^i) T_i^\eta(x^i|y^i) T_i^\phi(x^i|y^i)$$

where the index i refers to the final-state particles. The resolutions in η and ϕ are specified as delta functions and are typically very narrow. Conversely, the energy resolution is contingent upon the particle characteristics and must replicate the behavior of the simulation detector effects, which are usually Gaussian in the case of rapid simulations [2] [11] [13].

4) Drell-Yan process and weights. In this process, a virtual photon, or Z boson, is created when a quark from one hadron and an antiquark from another hadron annihilate. After that, this virtual particle decays into two leptons with opposing charges (e.g., electrons and muons). This method is important because it yields important information regarding parton distribution functions (PDFs). These PDFs characterize the distribution of momentum among the partons (quarks and gluons) of an incoming high-energy nucleon. Deep inelastic scattering (DIS) is strongly associated with the Drell-Yan mechanism. The Feynman diagram for the Drell-Yan process can be obtained by rotating the DIS Feynman diagram by 90 degrees [14].

5) DNN decisions and fitting with MEM. Fully linked layers with linear activations for the output layer and leaky *ReLU* activations for the hidden layers make up the DNN architectures used in Ref. [8], see **Figure 1**. The Leaky *ReLU* performs better than other activation functions like ReLU, softplus, and ELU [9] [15] [16] in a comparison study of activation functions in “Deep neural networks and skip connections,” which has demonstrated time affordability. In order to evaluate the training stage's progress, the coefficient of determination, or R^2 , can be computed using,

$$1 - R^2 = \frac{\sum_i (f(\mathbf{x}_i)_{predicted} - f(\mathbf{x}_i)_{true})^2}{N\sigma^2} \tag{8}$$

where σ^2 is the true values' variance over the N validation points that are identified by the index i . Keep in mind that the mean-squared error, or the selected objective function, is proportional to $1 - R^2$. While the mean-squared error alone does not allow us to compare different models self-consistently, the constant of proportionality, σ^2 , standardizes it in a way that makes it a meaningful figure of merit during the training phase [8] [10] [13].

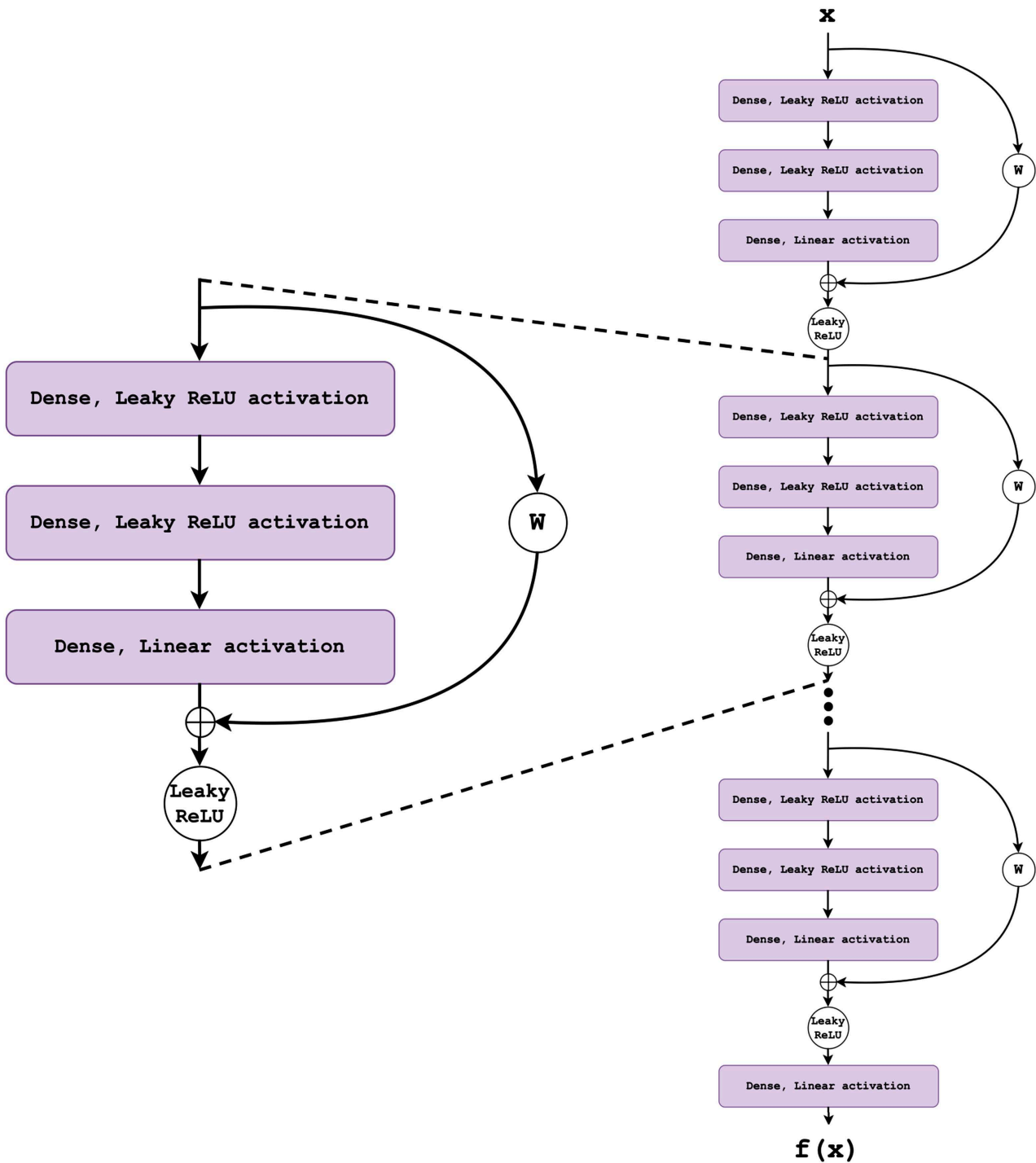


Figure 1. The building block with skip connections in a DNN. The activation of *Leaky ReLU* is totally associated with the first two levels. After being added to the input via the skip link and having linear activation, the last layer is converted using a non-linear *Leaky ReLU* function. If the input and output dimensions for the block and transformation gate I are different, then the weight matrix W can be trained. To create the neural network, the blocks are layered one after the other [8].

Depending on how sophisticated the process is, calculating an event’s weight with MoMEMta can take a few seconds to several minutes. It quickly becomes prohibitive to repeat this for every hypothesis α and every event. In real-world

analyses with several hypotheses and perhaps extra model parameters, this frequently complicates the method's use. To put it briefly, Ref. [8] study's method processes simulated samples using MoMEMta and generates event information I' for various hypotheses. The outcome is then utilized to train a DNN together with MoMEMta inputs, which are the 4-momenta of every observable particle. The 4-momenta of all the detected particles and the missing transverse energy (P_T and ϕ) are the inputs of MoMEMta; as such, these are the inputs we wish to supply to the DNN [2]. Particles produced may be boosted in one or both directions, depending on the longitudinal difference in momentum between the initial partons that collide in the detector. The network's capacity to explain the interesting portion of the matrix element will be hampered by the need to learn about the longitudinal boost in addition to the function in Equation (6). The situation is greatly improved by using the P_T , η , and ϕ angles for each particle as inputs since the $\Delta\eta$ between two particles is well approximated and independent of the longitudinal boost [2].

3. Proof of Concept and Application to $t\bar{t}$ Production

Let's use the MEM in this study to produce two opposite sign leptons and two jets (bjets) initiated by b quarks as detected particles as a proof of concept. The primary contributions to this topology make it intriguing—Drell-Yan $Z \rightarrow l^+l^-$ production with additional jets, and top quark pair production $t\bar{t}$ having leptonic decays of the W bosons from top quarks—are very dissimilar in the way they are treated in the integration. While there are no missing particles in the former case, the later involves undiscovered neutrinos whose degrees of freedom must be taken into consideration. Considering the resonant

$H \rightarrow Z(\rightarrow l^+l^-)A(\rightarrow b\bar{b})$ process that results in the context of Two Higgs Doublet Models (2HDM), and has been studied by the ATLAS and CMS collaborations [17] [18]. The unknown masses of the H and A bosons will demonstrate the effectiveness of our method when the parameter space is multi-dimensional, which is precisely the situation where the classical integration is impracticable. The multiple resonances in this process also pose an interesting challenge from the integration point of view [2] [16]. A synopsis of the events for which the weights have been calculated indicates that, in order to prevent the network from becoming unbalanced, only 500 K events for the $t\bar{t}$ process have been employed in the training [2]. For both M_A and M_H , the $H \rightarrow ZA$ samples and weights are divided into 23 mass configurations up to 1 TeV. For each event $t\bar{t}$, Drell-Yan, and $H \rightarrow ZA$, the event information I'_α will be assessed 23 times using various mass parameters specific to the $H \rightarrow ZA$ instance. The data given in Ref. [2] also includes the number of faulty weights and the number that were recomputed using additional iterations. **Figure 2** displays the I'_{DY} distributions for each of the three sample types. The weights from the MEM computed with MoMEMta and the ones from the DNN agree quite well. The model with six layers and 200 neurons, with *RELU* and *SELU* [9] [15] [16] activation functions for the hidden and output layers, respectively. Utilizing the Breit-Wigner resonances in a

variable change allows for a reasonable computation time reduction (approximately 3.2 times slower than for the Drell-Yan weights) [1] [8]. **Figure 3** displays the corresponding $I'_{t\bar{t}}$ distributions. Although there is less of a discrepancy in the weight distribution between the Drell-Yan and $t\bar{t}$ samples, the Drell-Yan events have a longer tail. The distinct mass configurations M_H and M_A that make up this sample are the source of the double peak observed in the $H \rightarrow ZA$ scenario. Low weights are associated with high (pseudo)scalar masses, but low masses are more in line with the theory of $t\bar{t}$. Overall, there is good agreement between the weights from the DNN and those obtained classically [2]. Considering the application's next-leading order (NLO) of MEM in $t\bar{t}$ production, we can examine the impact of missing higher-order effects in the likelihood computation as an example. With POWHEG [19] [20], a single top-quark event is produced, and PYTHIA8 [21] is used to sprinkle it down [12]. It is possible to identify the specific ways in which the parton shower impacts the analysis by using the MEM to analyze these occurrences using fixed-order calculations. The extraction of the top-quark mass using the MEM from showering events produced using POWHEG and PYTHIA8 is depicted in the left plot of **Figure 4**. The likelihood function is exclusively computed in the Born approximation (blue) and with NLO adjustments (red) in the study displayed in **Figure 4**. To derive the estimator for the top-quark mass \hat{m}_t as the position of the minimum and the related statistical uncertainty $\Delta\hat{m}_t$ as the breadth of the parabola, a parabola is fitted to the negative logarithm of the likelihood function (*Log-Likelihood*). By factorizing in two different factorization and renormalization scales in the likelihood calculation, the theoretical uncertainty resulting from missing higher orders is assessed. The missing higher-order adjustments in the corresponding likelihood calculations are the reason for the biases in the recovered estimators $\hat{m}_t^{Born} = 150.19$ GeV and $\hat{m}_t^{NLO} = 163.75$ GeV with regard to the input value for the top-quark mass $m_t^{true} = 173.2$ GeV. When there is a discrepancy between the probability density used to calculate likelihood and the actual distribution of the events, the MEM is known to provide biased estimators. Hence, there is a need for the calibration of the method in introducing its associated uncertainties.

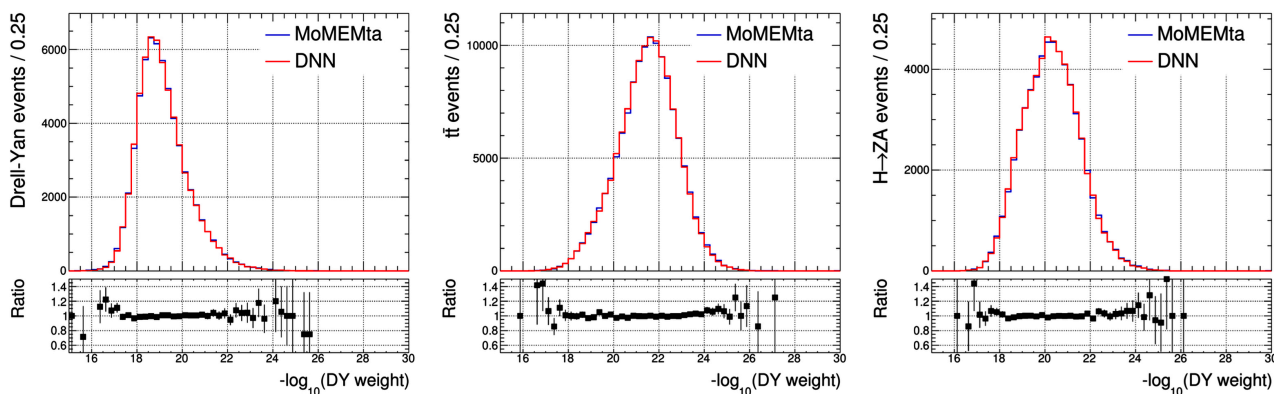


Figure 2. Distributions of the event information I'_{DY} from both MoMEMta and the DNN for the three samples Drell-Yan (left), $t\bar{t}$ (middle) and $H \rightarrow ZA$ (right) events [2].

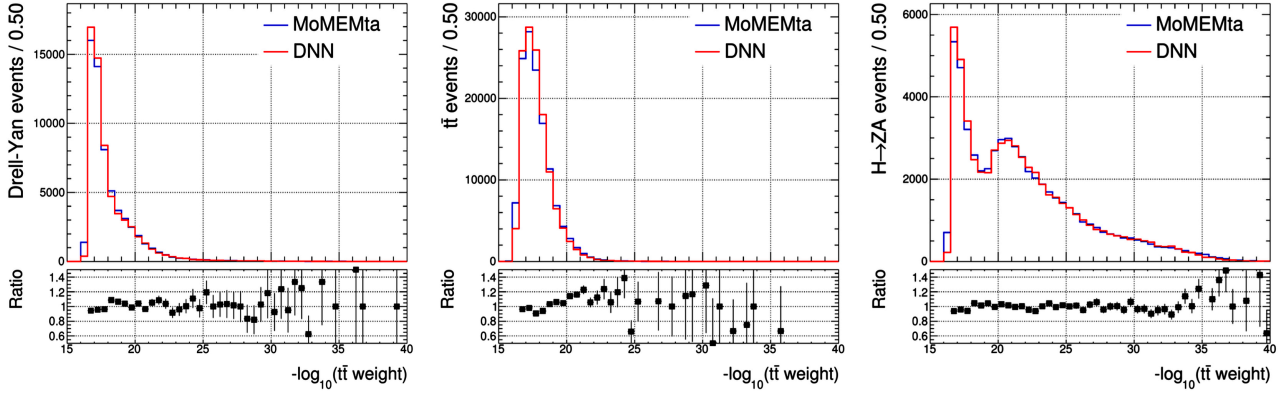


Figure 3. Distributions of the event information I'_{tt} from both MoMEMta and the DNN for the three samples: Drell-Yan (left), $t\bar{t}$ (middle) and $H \rightarrow ZA$ (right) events [2].

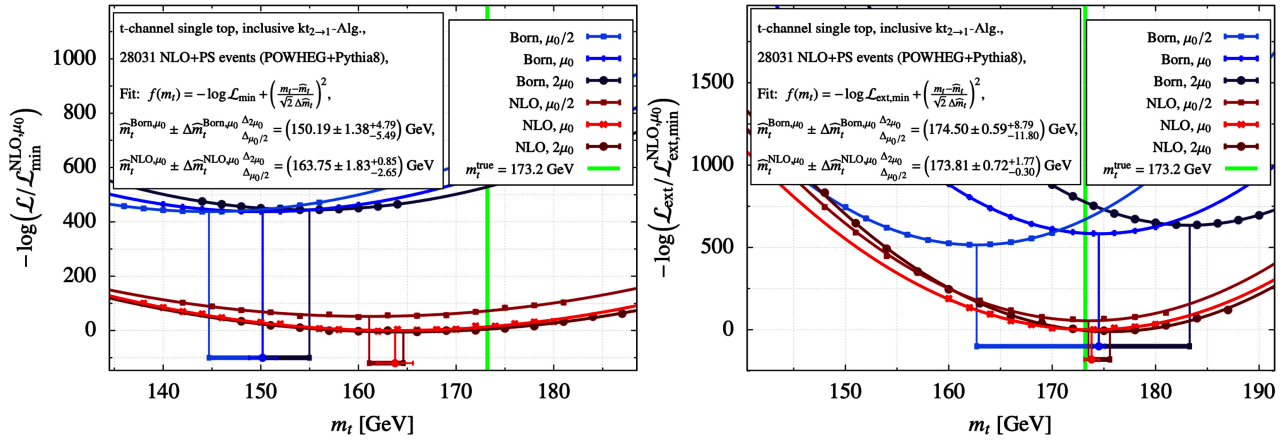


Figure 4. **Left:** Top-quark mass extraction via the MEM at leading order (LO) (blue) and NLO (red) from single top-quark events generated with POWHEG + PYTHIA8. **Right:** Same but with extended likelihoods functions [12].

4. Conclusion

This study proposes the utilization of a Deep Neural Network (DNN) to approximate the integral of the Matrix Element Method (MEM) with the objective of expediting MEM computations. It is inferred from the few examined and documented example processes that, with the assistance of specialized tools like MoMEMta, a DNN can be effectively trained to closely replicate the outcomes of the direct numerical integration of the matrix element. Although the weights in this regression employing the DNN inherently contain errors, these are unlikely to significantly impact performance across various applications. The adoption of accelerated weight computations presents a plethora of opportunities, such as systematic research, probability scans, parameter scans, and, more broadly, the application of MEM to a wider array of physics analyses, including the discovery of novel physics phenomena.

Acknowledgements

I wish to express my sincere gratitude to Professor Elena Long for her invaluable

introduction to “Special Topics in Particle Physics”. The research by Simone Alioli *et al.* served as a significant impetus for this study.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- [1] Nguyen, N.T.T., Kenyon, G.T. and Yoon, B. (2020) A Regression Algorithm for Accelerated Lattice QCD That Exploits Sparse Inference on the D-Wave Quantum Annealer. *Scientific Reports*, **10**, Article No. 10915. <https://doi.org/10.1038/s41598-020-67769-x>
- [2] Bury, F. and Delaere, C. (2021) Matrix Element Regression with Deep Neural Networks—Breaking the CPU Barrier. *Journal of High Energy Physics*, **2021**, Article No. 20. [https://doi.org/10.1007/jhep04\(2021\)020](https://doi.org/10.1007/jhep04(2021)020)
- [3] Dery, L.M., Nachman, B., Rubbo, F. and Schwartzman, A. (2017) Weakly Supervised Classification in High Energy Physics. *Journal of High Energy Physics*, **2017**, Article No. 145. [https://doi.org/10.1007/jhep05\(2017\)145](https://doi.org/10.1007/jhep05(2017)145)
- [4] Kuusela, M., Vatanen, T., Malmi, E., Raiko, T., Aaltonen, T. and Nagai, Y. (2012) Semi-Supervised Anomaly Detection—Towards Model-Independent Searches of New Physics. *Journal of Physics. Conference Series*, **368**, Article ID: 012032. <https://doi.org/10.1088/1742-6596/368/1/012032>
- [5] Brusino, N. (2016) Search for the Standard Model Higgs Boson Decaying into $b\bar{b}$ Produced in Association with Top Quarks Decaying Hadronically in Pp Collisions at $\sqrt{s} = 8$ TeV with the ATLAS Detector. *4th Annual Large Hadron Collider Physics*, Lund, 13-18 June 2016, Sweden, 1-4. <https://doi.org/10.22323/1.276.0192>
- [6] Dalitz, R.H. and Goldstein, G.R. (1999) Test of Analysis Method for Top—Antitop Production and Decay Events. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, **455**, 2803-2834. <https://doi.org/10.1098/rspa.1999.0428>
- [7] Leigh, M., Raine, J.A., Zoch, K. and Golling, T. (2023) ν -Flows: Conditional Neutrino Regression. *SciPost Physics*, **14**, 159. <https://doi.org/10.21468/scipostphys.14.6.159>
- [8] Bishara, F., Paul, A. and Dy, J. (2024) High-Precision Regressors for Particle Physics. *Scientific Reports*, **14**, Article No. 5294. <https://doi.org/10.1038/s41598-024-52941-4>
- [9] Nair, V. and Hinton, G.E. (2010) Rectified Linear Units Improve Restricted Boltzmann Machines. *Proceedings of the 27th International Conference on International Conference on Machine Learning*, Haifa, 21-24 June 2010, 807-814.
- [10] Grazzini, M., Kallweit, S. and Wiesemann, M. (2018) Fully Differential NNLO Computations with MATRIX. *The European Physical Journal C*, **78**, Article No. 537. <https://doi.org/10.1140/epjc/s10052-018-5771-7>
- [11] Fiedler, F., Grohsjean, A., Haefner, P. and Schieferdecker, P. (2010) The Matrix Element Method and Its Application to Measurements of the Top Quark Mass. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **624**, 203-218. <https://doi.org/10.1016/j.nima.2010.09.024>
- [12] Martini, T., Kraus, M., Peitzsch, S. and Uwer, P. (2020) The Matrix Element Method as a Tool for Precision and Accuracy. *Proceedings of European Physical Society Conference on High Energy Physics—PoS (EPS-HEP2019)*, Belgium, 10-17 July 2019, 1-9. <https://doi.org/10.22323/1.364.0673>

- [13] Fiedler, F. (2010) Precision Measurements of the Top Quark Mass. <https://api.semanticscholar.org/CorpusID:118787406>
- [14] Chen, L., Gao, J. and Liang, Z. (2014) Nuclear Dependencies of Azimuthal Asymmetries in the Drell-Yan Process. *Physical Review C*, **89**, Article ID: 035204. <https://doi.org/10.1103/physrevc.89.035204>
- [15] Sun, Y., Wang, X. and Tang, X. (2015) Deeply Learned Face Representations Are Sparse, Selective, and Robust. 2015 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, 7-12 June 2015, 2892-2900. <https://doi.org/10.1109/cvpr.2015.7298907>
- [16] Clevert, D., Unterthiner, T. and Hochreiter, S. (2015) Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs). <https://api.semanticscholar.org/CorpusID:5273326>
- [17] Saggio, A. (2020) Search for 2HDM Neutral Higgs Bosons through the Process $\rightarrow ZA \rightarrow l^+l^- b\bar{b}$ with the CMS Detector. *Proceedings of European Physical Society Conference on High Energy Physics—PoS (EPS-HEP2019)*, Belgium, 10-17 July 2019, 1-5. <https://doi.org/10.22323/1.364.0361>
- [18] Petrucciani, G. (2013) Searches for a Heavy Higgs Boson Decaying into WW, ZZ in Final States with Hadronic Jets and Neutrinos. In: Petrucciani, G., Ed., *The Search for the Higgs Boson at CMS*, Scuola Normale, 147-163. https://doi.org/10.1007/978-88-7642-482-3_7
- [19] Alioli, S., Nason, P., Oleari, C. and Re, E. (2010) A General Framework for Implementing NLO Calculations in Shower Monte Carlo Programs: The POWHEG Box. *Journal of High Energy Physics*, **2010**, Article No. 43. [https://doi.org/10.1007/jhep06\(2010\)043](https://doi.org/10.1007/jhep06(2010)043)
- [20] Alioli, S., Nason, P., Oleari, C. and Re, E. (2009) NLO Single-Top Production Matched with Shower in POWHEG: s - and t -Channel Contributions. *Journal of High Energy Physics*, **2009**, 111. <https://doi.org/10.1088/1126-6708/2009/09/111>
- [21] Sjöstrand, T., Ask, S., Christiansen, J.R., Corke, R., Desai, N., Ilten, P., *et al.* (2015) An Introduction to PYTHIA 8.2. *Computer Physics Communications*, **191**, 159-177. <https://doi.org/10.1016/j.cpc.2015.01.024>