Snowmass Letter of Interest

Matrix Element Method in the Machine Learning Era

Philip Chang[1], **Matthew Feickert**[2], **and Mark Neubauer**[2]

[1]University of California at San Diego, La Jolla, CA 92093, USA
[2]University of Illinois at Urbana-Champaign, Champaign, IL 61820, USA

1 Introduction

The Matrix Element (ME) Method [1–4] is a powerful technique which can be utilized for measurements of physical model parameters and direct searches for new phenomena. It has been used extensively by collider experiments at the Tevatron for Standard Model (SM) measurements and Higgs boson searches [5–10] and at the LHC for measurements in the Higgs and top quark sectors of the SM [11–17]. The ME method is based on *ab initio* calculation of the probability density function $\mathscr{P}$ of an event with observed final-state particle momenta $\mathbf{x}$ to be due to a physics process ξ with theory parameters $\boldsymbol{\alpha}$. One can compute $\mathscr{P}_\xi(\mathbf{x}|\boldsymbol{\alpha})$ by means of the factorization theorem from the corresponding partonic cross-sections of the hard scattering process involving parton momenta $\mathbf{y}$ and is given by

$$\mathscr{P}_\xi(\mathbf{x}|\boldsymbol{\alpha}) = \frac{1}{\sigma_\xi^{\text{fiducial}}(\boldsymbol{\alpha})} \int d\Phi(\mathbf{y}_{\text{final}})\, dx_1\, dx_2\, \frac{f(x_1)f(x_2)}{2sx_1x_2}\, |\mathscr{M}_\xi(\mathbf{y}|\boldsymbol{\alpha})|^2\, \delta^4(\mathbf{y}_{\text{initial}} - \mathbf{y}_{\text{final}})\, W(\mathbf{x},\mathbf{y}) \tag{1}$$

where and x_i and $\mathbf{y}_{\text{initial}}$ are related by $y_{\text{initial},i} \equiv \frac{\sqrt{s}}{2}(x_i, 0, 0, \pm x_i)$, $f(x_i)$ are the parton distribution functions, $\sqrt{s}$ is the collider center-of-mass energy, $\sigma_\xi^{\text{fiducial}}(\boldsymbol{\alpha})$ is the total cross section for the process ξ (with $\boldsymbol{\alpha}$) times the detector acceptance, $d\Phi(\mathbf{y})$ is the phase space density factor, $\mathscr{M}_\xi(\mathbf{y}|\boldsymbol{\alpha})$ is the matrix element (typically at leading-order (LO)), and $W(\mathbf{x},\mathbf{y})$ is the probability density (aka "transfer function") that a selected event $\mathbf{y}$ ends up as a measured event $\mathbf{x}$.

One can use calculations of Equation 1 in a number of ways to search for new phenomena at particle colliders. For measurement of model parameters $\boldsymbol{\alpha}$, one would maximize the likelihood function for observed events $\mathscr{L}(\boldsymbol{\alpha})$ given by

$$\mathscr{L}(\boldsymbol{\alpha}) = \prod_i \sum_k f_k \mathscr{P}_{\xi_k}(\mathbf{x}_i|\boldsymbol{\alpha}) \tag{2}$$

where f_k are the fractions of (non-interfering) processes contributing to the data. For new particle searches, one can (using Bayes' Theorem [18]) compute for a hypothesized signal S the probability $p(S|\mathbf{x})$ given by

$$p(S|\mathbf{x}) = \frac{\sum_i \beta_{S_i} \mathscr{P}_{S_i}(\mathbf{x}|\boldsymbol{\alpha}_{S_i})}{\sum_i \beta_{S_i} \mathscr{P}(\mathbf{x}|\boldsymbol{\alpha}_{S_i}) + \sum_j \beta_{B_j} \mathscr{P}(\mathbf{x}|\boldsymbol{\alpha}_{B_j})} \tag{3}$$

where, S_i and B_j, denote all signal and background processes relevant to the considered phase space and β are the *a priori* expected process fractions. According to the Neyman-Pearson Lemma [19], Equation 3 is the optimal discriminant function for S in the presence of B and can be used to extract a signal fraction in the data.

2 Applications of the Matrix Element Method

2.1 Advantages over Training-based Methods

As a multivariate analysis approach, the ME method brings in several unique and desirable features, most notably it (1) does not require training data being an *ab initio* calculation of event probabilities, (2) incorporates all available kinematic information of a hypothesized process, including all correlations, and (3) has a clear physical meaning in terms transition probabilities within the framework of quantum field theory.

2.2 Limitations with Current Techniques

One drawback to the ME method is that it has traditionally relied on LO matrix elements, although nothing in principle limits the ME method to LO calculations. Techniques that accommodate initial-state QCD radiation within the LO ME framework using transverse boosting and dedicated transfer functions to integrate over the transverse momentum of initial-state partons have been developed [20]. Another challenge is development of the transfer functions which rely on tediously hand-crafted fits to full simulated Monte-Carlo events.

The most serious difficulty in the ME method, and the one which has limited its applicability to searches for beyond-the-SM physics and precision measurements at collider experiments, is that it is very *computationally intensive*. If this limitation could be overcome, as argued in [21], then it would enable more widespread use of ME methods for analysis of LHC data. This could be particularly important for extending the new physics reach of the HL-LHC which will be dominated by increases in integrated luminosity rather than center-of-mass collision energy.

Accurate evaluation of Equation 1 is computationally challenging primarily for two reasons: (1) it involves high-dimensional integration over a large number of events, signal and background hypotheses, and systematic variations and (2) it involves sharply-peaked integrands[1] over a large domain in phase space. In reference to point (1), the matrix element $\mathscr{M}_\xi(\mathbf{y}|\boldsymbol{\alpha})$ in the method involves all partons in the $n \to m$ process, so when the 4-momentum of particles are not completely measured experimentally (e.g. neutrinos), one must integrate over the missing information which increases the dimensionality of the integration. In reference to point (2), a clever technique to re-map the phase space in order to reduce the sharpness of integrate in that space in an automated way (`MADWEIGHT` [22]) is often used in conjunction with a matrix element calculation package (`MADGRAPH_aMCNLO` [23]). In practice, evaluation of definite integrals by the ME approach invokes techniques such as importance sampling (see `VEGAS` [24, 25] and `FOAM` [26]) or recursive stratified sampling (see `MISER` [27]) Monte Carlo integration.

2.3 Sustainable Matrix Element Method using Deep Learning

Despite the attractive features of the ME method and promise of further optimization and parallelization of the evaluation of Equation 1, the computational burden of the ME technique will continue to limit its range of applicability for practical data analysis without new and innovative approaches. This is especially true when one considers the process of producing a physics publication which involves many selection, sample and systematic iterations for which ME calculations are required. The primary idea put forward in this Letter of Interest is to utilize modern *machine learning techniques to dramatically speed up the numerical evaluation of Equation* 1 and therefore broaden the applicability of the ME method to the benefit of the HL-LHC and future colliders physics program. There has been recent advances in this area evaluating `MadGraph` matrix elements within machine learning frameworks [28], that serves as proof of principle that these techniques are feasible.

3 Example Analysis Flow

The expensive full ME calculations should be done as infrequently as possible; ideally once for DNN training and once more for a final pass before publication, with the DNNs utilized as a good approximation in between. A future analysis flow using the ME method with DNNs might look something like the following: One performs a large number of ME calculations using a traditional numerical integration technique like `VEGAS` or `FOAM` on a large CPU resource like an HPC, Cloud or the Grid, ideally exploiting hardware acceleration of GPUs or FPGAs. The DNN training data is generated from the phase space sampling in performing the full integration in this initial pass, and DNNs are trained either *in situ* or *a posteriori*. The accuracy of the DDN-based ME calculation can be assessed through this procedure. As the analysis develops through selection and/or sample changes, systematic treatment, etc., the DNN-based ME calculations are used in place of the full ME calculations to make the analysis nimble and to preserve the ME calculations.

4 Outlook and Opportunities

In this letter of interest, we outlined a number of the advantages and challenges of the MEM technique as a tool in new physics searches and measurements at colliders. The case for improvements in the MEM technique is strong for the HL-LHC as sensitivity gains will come from increases in luminosity rather than energy requiring full exploitation of the kinematic information from collisions, e.g. to search for new physics from virtual effects from high-scale physics in detailed analyses of event kinematics. Many of these arguments also hold at proposed future high energy colliders (e.g. FCC-ee/hh), particularly the physics interpretability inherent in the MEM technique. Given these challenges and opportunities, we propose some open questions which warrant exploration as part of the Snowmass process:

- **Applications:** What are the key applications of the MEM technique for physics at the HL-LHC and future colliders?
- **Precision**: How well can the MEM calculations be approximated by machine learning techniques? How general are these approximations across the physics processes required in the key applications above?
- **Accuracy**: How can we make the MEM calculations more accurate without a large increase the computational requirements? E.g. can we go beyond LO and employ calculations of kinematics at NLO with new (possibly ML-based) techniques?

[1] a consequence of imposing energy/momentum conservation in the processes.

- **Efficiency**: Can the computational costs of approximating the MEM calculations with machine learning techniques be further reduced with new or revised approaches, such as networks inspired by known physics?
- **Software and Accelerators**: What is the value added of implementing a common software project for ME calculations — E.g. in the spirit of `MoMEMta` [29]. Can the MEM perform well in languages with a rich ML ecosystem such as scientific python? This work has already begun in the SCAILFIN project, with initial investigations into building tools based on the flexible `JAX` machine learning framework [30] to exploit automatic differentiation alongside hardware acceleration [31].
- **Portability and Sustainability**: If accurate models of MEM calculations can be constructed, say in DNNs trained to learn the ME integration with sufficient generality, these could be shared widely and used broadly in multiple analyses. These trained DNNs can be thought of as preserving the essence of MEM calculations in a way that allows for subsequent *fast forward-execution*, from minutes to milliseconds per event, even opening up the possibility of utilizing MEM-based inference in detector triggering. In this way, DNNs can enable MEM to be both *nimble* and *sustainable*, neither of which is true today, and consequently broaden the application of MEM techniques to HEP research.
- **MEMaaS**: What might be gained by employing shared software and computing infrastructure that physicists could use to generate and share approximate MEM models through a common API?

Snowmass 2021 provides an opportunity for the HEP community to explore the role of the MEM technique in present and future collider experiments and how we can leverage the rapid developments in machine learning and data science to address the challenge of computational tractability to broaden its applicability in HEP.

References

1. K. Kondo, "Dynamical Likelihood Method for Reconstruction of Events With Missing Momentum. 1: Method and Toy Models", *J. Phys. Soc. Jap.* **57** (1988) 4126–4140, doi:10.1143/JPSJ.57.4126.
2. F. Fiedler, A. Grohsjean, P. Haefner, and P. Schieferdecker, "The Matrix Element Method and its Application in Measurements of the Top Quark Mass", *Nucl. Instrum. Meth.* **A624** (2010) 203–218, doi:10.1016/j.nima.2010.09.024, arXiv:1003.1316.
3. I. Volobouev, "Matrix Element Method in HEP: Transfer Functions, Efficiencies, and Likelihood Normalization", *ArXiv e-prints* (January, 2011) arXiv:1101.2259.
4. F. Elahi and A. Martin, "Using the modified matrix element method to constrain $L_\mu - L_\tau$ interactions", *Phys. Rev.* **D96** (2017), no. 1, 015021, doi:10.1103/PhysRevD.96.015021, arXiv:1705.02563.
5. D0 Collaboration, "A precision measurement of the mass of the top quark", *Nature* **429** (2004) 638–642, doi:10.1038/nature02589, arXiv:hep-ex/0406031.
6. CDF Collaboration, "Precision measurement of the top quark mass from dilepton events at CDF II", *Phys. Rev.* **D75** (2007) 031105, doi:10.1103/PhysRevD.75.031105, arXiv:hep-ex/0612060.
7. CDF Collaboration, "First Measurement of *ZZ* Production in panti-p Collisions at $\sqrt{s}$ = 1.96-TeV", *Phys. Rev. Lett.* **100** (2008) 201801, doi:10.1103/PhysRevLett.100.201801, arXiv:0801.4806.
8. CDF Collaboration, "Inclusive Search for Standard Model Higgs Boson Production in the WW Decay Channel using the CDF II Detector", *Phys. Rev. Lett.* **104** (2010) 061803, doi:10.1103/PhysRevLett.104.061803, arXiv:1001.4468.
9. D0 Collaboration, "Observation of Single Top Quark Production", *Phys. Rev. Lett.* **103** (2009) 092001, doi:10.1103/PhysRevLett.103.092001, arXiv:0903.0850.
10. CDF Collaboration, "First Observation of Electroweak Single Top Quark Production", *Phys. Rev. Lett.* **103** (2009) 092002, doi:10.1103/PhysRevLett.103.092002, arXiv:0903.0885.
11. CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Phys. Lett.* **B716** (2012) 30–61, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.
12. CMS Collaboration, "Measurement of the properties of a Higgs boson in the four-lepton final state", *Phys. Rev.* **D89** (2014), no. 9, 092007, doi:10.1103/PhysRevD.89.092007, arXiv:1312.5353.
13. ATLAS Collaboration, "Measurements of Higgs boson production and couplings in the four-lepton channel in pp collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector", *Phys. Rev.* **D91** (2015), no. 1, 012006, doi:10.1103/PhysRevD.91.012006, arXiv:1408.5191.
14. CMS Collaboration, "Measurement of spin correlations in $t\bar{t}$ production using the matrix element method in the muon+jets final state in pp collisions at $\sqrt{s} = 8$ TeV", *Phys. Lett.* **B758** (2016) 321–346, doi:10.1016/j.physletb.2016.05.005, arXiv:1511.06170.

15. CMS Collaboration, "Search for a Standard Model Higgs Boson Produced in Association with a Top-Quark Pair and Decaying to Bottom Quarks Using a Matrix Element Method", *Eur. Phys. J.* **C75** (2015), no. 6, 251, doi:10.1140/epjc/s10052-015-3454-1, arXiv:1502.02485.
16. ATLAS Collaboration, "Search for the Standard Model Higgs boson produced in association with top quarks and decaying into $b\bar{b}$ in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector", *Eur. Phys. J.* **C75** (2015), no. 7, 349, doi:10.1140/epjc/s10052-015-3543-1, arXiv:1503.05066.
17. ATLAS Collaboration, "Evidence for single top-quark production in the s-channel in proton-proton collisions at $\sqrt{s}$ =8 TeV with the ATLAS detector using the Matrix Element Method", *Phys. Lett.* **B756** (2016) 228–246, doi:10.1016/j.physletb.2016.03.017, arXiv:1511.05980.
18. M. Bayes and M. Price, "An essay towards solving a problem in the doctrine of chances. by the late rev. mr. bayes, f. r. s. communicated by mr. price, in a letter to john canton, a. m. f. r. s.", *Philosophical Transactions* **53** (1763) 370–418, doi:10.1098/rstl.1763.0053, arXiv:http://rstl.royalsocietypublishing.org/content/53/370.full.pdf+html.
19. J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses", *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **231** (1933), no. 694-706, 289–337, doi:10.1098/rsta.1933.0009, arXiv:http://rsta.royalsocietypublishing.org/content/231/694-706/289.full.pdf.
20. J. Alwall, A. Freitas, and O. Mattelaer, "The Matrix Element Method and QCD Radiation", *Phys. Rev.* **D83** (2011) 074010, doi:10.1103/PhysRevD.83.074010, arXiv:1010.2263.
21. P. Chang, S. Gleyzer, M. Neubauer, and D. Zhong, "Sustainable matrix element method through deep learning", doi:10.5281/zenodo.4008241.
22. P. Artoisenet, V. Lemaitre, F. Maltoni, and O. Mattelaer, "Automation of the matrix element reweighting method", *JHEP* **12** (2010) 068, doi:10.1007/JHEP12(2010)068, arXiv:1007.3300.
23. J. Alwall et al., "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", *JHEP* **07** (2014) 079, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.
24. G. P. Lepage, "A new algorithm for adaptive multidimensional integration", *Journal of Computational Physics* **27** (1978), no. 2, 192 – 203, doi:http://dx.doi.org/10.1016/0021-9991(78)90004-9.
25. T. Ohl, "Vegas revisited: Adaptive Monte Carlo integration beyond factorization", *Comput. Phys. Commun.* **120** (1999) 13–19, doi:10.1016/S0010-4655(99)00209-X, arXiv:hep-ph/9806432.
26. S. Jadach, "Foam: A general-purpose cellular monte carlo event generator", *Computer Physics Communications* **152** (2003), no. 1, 55 – 100, doi:http://dx.doi.org/10.1016/S0010-4655(02)00755-5.
27. W. H. Press and G. R. Farrar, "RECURSIVE STRATIFIED SAMPLING FOR MULTIDIMENSIONAL MONTE CARLO INTEGRATION", *Submitted to: Comp.in Phys.* (1989).
28. F. Bury and C. Delaere, "Matrix Element Regression with Deep Neural Networks – breaking the CPU barrier", arXiv:2008.10949.
29. "MoMEMta: Modular Matrix Element Implementation". https://github.com/MoMEMta.
30. J. Bradburyet al., "JAX: composable transformations of Python+NumPy programs", 2018. http://github.com/google/jax.
31. M. Feickert, "PyMELA". https://github.com/scailfin/pyMELA.