Reply to Report 1 for 2308.00027v1

Title: "Returning CP-observables to the Frames They Belong"

Authors: Jona Ackerschott, Rahool Kumar Barman, Dorival Gonçalves, Theo Heimel, Tilman Plehn

We would like to thank the referee for carefully reading our manuscript and providing detailed and valuable comments. The version of the manuscript that we resubmit addresses the aspects that the report brought to our consideration. Please find below the comments from the referee and our answers.

1. **Referee's comment**: *Unfolding has been extensively used in HEP analyses and there are many (non NN-based) unfolding algorithms available, including iterative method, singular value decomposition, bin-by-bin correlation etc. Many of these methods already exist in a popular analysis code such as Root. Why does one needs cINN (or NN)? What would be advantages of using NN over existing unfolding methods? Is it clear that conventional unfolding methods might miss anything important that NN might capture? There are some comparison studies such as* `https://arxiv.org/pdf/2104.03036.pdf` *but it should be straightforward for authors to make a quick comparison for Top Yukawa CP phase.*

   Author response: Machine learning (ML) and neural network (NN) based unfolding techniques, like our cINN approach, offer several key advantages over classical unfolding methods. These include the potential to unfold the full phase space in a bin-independent fashion rather than being limited to single observables. To be precise, our cINN-based unfolding method allows for the unfolding of single events and does not require prior reconstruction of target observables. The main reason for using ML unfolding for CP-observables lies in the last point. Classical methods necessitate the prior reconstruction of CP-sensitive observables before unfolding can occur. Consequently, the improved sensitivity to CP-violation is unattainable in classical unfolding techniques. This is because classical unfolding is still dependent of the inherent sensitivity reached by prior classical reconstruction.

   To further clarify this point, we added the following comments to page 3: "Machine Learning provides generally unbinned techniques and allows for a simultaneous reconstruction of the full phase-space, hence removing the need for a prior reconstruction of any targeted observable [58]. Probabilistic techniques specifically can render the

unfolding algorithm valid event by event, resulting in its viability even for low-statistic measurements [60].".

2. **Referee's comment**: *There are already several suggestions on using ML for unfolding. Especially the following references advertise multi-dimensional unfolding as authors study in their paper:*

   - `https://arxiv.org/pdf/1911.09107.pdf`
   - `https://arxiv.org/pdf/2203.16722.pdf`.

   I understand that authors are focusing more on physics side (CP phase) and authors simply could have used conventional unfolding methods or existing ML-based unfolding. Are there particular reasons why cINN might be more suitable than these for CP phase study?

   Author response: We used the cINN here as an example of an existing ML unfolding technique (introduced in [1]). The main advantage of the cINN is that it allows for the unfolding of single events, such that it is already accurate for low event statistics. In a real application, this is an advantage over other ML-approaches (and classical approaches) together with the proposed iterative improvement [2].

3. **Referee's comment**: *I might have missed discussion but do authors perform unfolding with unbinned data or histograms? It seems that authors are doing unbinned analysis from the following sentence in conclusion "Modern machine learning makes it possible to unfold high-dimensional distributions, covering all correlations without binning." It may be good to make this clear in earlier sections, if not mentioned.*

   Author response: We perform unbinned unfolding which is inherent to the cINN method. For more details, please see the answers to questions 1 and 2.

4. **Referee's comment**: *I see PDF information for event generation but what scale is chosen?*

   Author response: We choose the dynamic factorization scale μ_F that comes default in MadGraph, where μ_F is determined by the transverse mass $\mu_F = m_T = \sqrt{E^2 - p_z^2}$ of the k_T clustered $2 \to 2$ event topology. In scenarios with multiple Feynman diagrams contributing to the process, the $2 \to 2$ clustering may proceed in different ways leading to different values of m_T. In such cases, the transverse mass of the diagram with the highest weight is typically used to set the scale [3, 4]. Analogously, the renormalization scale μ_R is defined by the transverse mass leading to the choice $\mu_R = \mu_F$. We have included the following comments on Page 4 to clarify this further:

   'We use `MadGraph5_aMC@NLO` [3] with `NNPDF2.3QED` [5] to generate signal events at leading order with $\sqrt{s} = 14$ TeV, considering the default dynamic factorization and renormalization scales. The process to determine the factorization scale μ_F involves k_T

clustering the events into a $2 \to 2$ topology. μ_F is then defined based on the transverse mass of the clustered system. The renormalization scale μ_R is set equal to μ_F.'

5. **Referee's comment**: *It seems that the (red) error bar in the right panel of Fig. 7 comes from statistical uncertainty. Wouldn't there be some systematic uncertainties coming from NN and from unfolding procedure? Is it obvious that such systematic uncertainties are much smaller than statistical uncertainty and therefore negligible?*

   Author response: The referee correctly pointed out that there may be additional sources of systematic uncertainty. One of the main sources of systematic uncertainties in our analysis is missing information at the detector level, due to which, the network is unable to recover the true parton level information fully. Secondly, as pointed out by the referee, imperfections in the NN architecture, training, unfolding procedures, and loss functions also contribute to the overall systematic uncertainty. This typically happens since the specific choice of cINN architecture and loss function does not guarantee convergence to the true global minimum. In practice, the training attains a local minimum, which can be somewhat different from the true global minimum. Thirdly, model-dependency related errors can also add-on to the systematic uncertainty.

   We would like to note that we do not anticipate these systematic uncertainties to be smaller than the statistical uncertainty. However, the main focus of our analysis is to understand whether the cumulative impact of these systematic uncertainties on the true sensitivity is less than the impact of systematic errors in the classical reconstruction. In this sense, we do not ignore or dismiss these errors, but rather treat them in our analysis. We investigate whether the cINN based unfolding method provides a more accurate representation of the underlying new physics despite these systematic uncertainties when compared to the classical reconstruction technique. It is observed that the cINN approach is able to more closely approximate the parton-level phase space, while taking into account the NN-based systematic uncertainties as well as model-related errors, which illustrates its potential advantage.

   We added a discussion on this to the new draft in Sec. 3.3 (third-last paragraph): "Here we can see that both approaches suffer from systematic uncertainties. In the classical approach these uncertainties predominantly originate in simplifying assumptions about reconstructed objects. The cINN approach gathers uncertainties mostly from model dependence and additionally from imperfections in the neural network architecture, training and loss function. Note that missing information on reco-level is also a source of systematic uncertainty for both approaches.

   As mentioned earlier, and as can be seen in the left panel of Fig. 7, generative unfold-

ing leads to a major improvement over classical reconstruction in terms of systematic bias. If this improvement actually leads to an improved sensitivity, it is encoded in the difference between the two unfolded kinematic distributions, shown in solid lines. To quantify this, we calculate the reduced χ^2 values for θ_{CS}, $\Delta\eta_{t_\ell t_h}$, b_4, and $\Delta\Phi_{t_\ell t_h}$ between the SM and $\alpha = \pi/4$ hypotheses, using the Poisson errors of the bin counts. The reduced χ^2 values are computed with ∼120k events and 64 bins, for three scenarios: parton-level truth (blue), classical reconstruction from Ref. [29] (green), and the cINN-based generative model trained on SM events (red). A higher χ^2 value indicates a greater sensitivity to new physics."

6. **Referee's comment**: *It is quite interesting to notice that classical reconstruction methods and cINN method are comparable for θ_{CS} and b_4. Do we know why?*

   Author response:

   While the bottleneck for the cINN method is model dependence, such that the quality of the reconstruction can be generally expected to be somewhat consistent across observables, for the classical reconstruction it is not so clear. We would expect that better detector-level proxies lead to better sensitivity, but we cannot conclude that a bad reconstruction necessarily means bad sensitivity. This is meant in the sense that we might accidentally create a sensitive observable from a bad proxy.

   In the case that our consistency expectation for the cINN observables is roughly correct, the comparable sensitivities for θ_{CS} and b_4 could be explained by the classical reconstruction having accidentally better sensitivity here than for the other two observables. The consistency of the cINN and the accidentally/surprisingly good sensitivity of θ_{CS} and b_4 is mentioned in the paper.

We have fully addressed all the referee comments, and we hope that with these clarifications and associated changes made to the manuscript, the paper can be accepted for publication in SciPost.

References

[1] M. Bellagente, A. Butter, G. Kasieczka, T. Plehn, A. Rousselot, R. Winterhalder, L. Ardizzone, and U. Köthe, *Invertible Networks or Partons to Detector and Back Again*, SciPost Phys. **9** (2020) 074, arXiv:2006.06685 [hep-ph].

[2] M. Backes, A. Butter, M. Dunford, and B. Malaescu, *An unfolding method based on*

conditional Invertible Neural Networks (cINN) using iterative training, arXiv:2212.08674 [hep-ph].

[3] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, *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, arXiv:1405.0301 [hep-ph].

[4] V. Hirschi and O. Mattelaer, *Automated event generation for loop-induced processes*, JHEP **10** (2015) 146, arXiv:1507.00020 [hep-ph].

[5] NNPDF, R. D. Ball, V. Bertone, S. Carrazza, L. Del Debbio, S. Forte, A. Guffanti, N. P. Hartland, and J. Rojo, *Parton distributions with QED corrections*, Nucl. Phys. **B877** (2013) 290, arXiv:1308.0598 [hep-ph].