Physics Letters B 829 (2022) 137055

Contents lists available at ScienceDirect

Physics Letters B

www.elsevier.com/locate/physletb

Explainable machine learning of the underlying physics of high-energy particle collisions

Yue Shi Lai [a,*], Duff Neill [b], Mateusz Płoskoń [a], Felix Ringer [a]

[a] Nuclear Science Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[b] Theoretical Division, MS B283, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

ARTICLE INFO

Article history:
Received 24 September 2021
Received in revised form 19 March 2022
Accepted 22 March 2022
Available online 15 April 2022
Editor: G.F. Giudice

ABSTRACT

We present an implementation of an explainable and physics-aware machine learning model capable of inferring the underlying physics of high-energy particle collisions using the information encoded in the energy-momentum four-vectors of the final state particles. We demonstrate the proof-of-concept of our White Box AI approach using a Generative Adversarial Network (GAN) which learns from a DGLAP-based parton shower Monte Carlo event generator. The constrained generator network architecture mimics the structure of a parton shower exhibiting similarities with Recurrent Neural Networks (RNNs). We show, for the first time, that our approach leads to a network that is able to learn not only the final distribution of particles, but also the underlying parton branching mechanism, i.e. the Altarelli-Parisi splitting function, the ordering variable of the shower, and the scaling behavior. While the current work is focused on perturbative physics of the parton shower, we foresee a broad range of applications of our framework to areas that are currently difficult to address from first principles in QCD. Examples include nonperturbative and collective effects, factorization breaking and the modification of the parton shower in heavy-ion, and electron-nucleus collisions.

1. Introduction

In recent years machine learning techniques have lead to range of new developments in nuclear and high-energy physics [1–29]. For example, in Refs. [1–5] jet tagging techniques were developed which often outperform traditional techniques. In Refs. [6–11] Generative Adversarial Networks (GANs) [30,31], a form of unsupervised machine learning, were used to simulate event distributions in high-energy particle collisions. There have also been efforts to infer physics information from data. In Ref. [32,33] a probabilistic model was introduced based on jet clustering and in Ref. [34] a convolutional autoencoder within a shower was used which qualitatively reproduces jet observables. See also Refs. [35–37] for recent work on physics-aware learning.

The underlying physics information of high-energy particle collisions is encoded in hard-scattering processes, the subsequent parton shower and the hadronization mechanism. These steps are modeled by general purpose parton showers used in Monte Carlo event generators which play an important role in our understanding of high-energy collider experiments [38–40]. Starting with highly energetic quarks or gluons which are produced in hard-scattering events, parton showers simulate parton branching processes (soft and collinear emissions) that occur during the evolution from the hard scale to the infrared which is followed by the hadronization step. Parton showers solve renormalization group equations and resum large logarithmic corrections to all orders which arise due to the sensitivity to both hard and soft physics. Since parton showers produce a fully exclusive final state, they are essential tools to improve our understanding of high-energy particle collisions. While the general concept of parton showers is well established, important questions about the perturbative accuracy [41–47], nonperturbative effects [48–51] and the modification in the nuclear environment [52–66], remain a challenge.

In this work, we propose an explainable or White Box AI approach [67,68] to learn the underlying physics of high-energy particle collisions. As a proof of concept, we present results of a GAN trained on the final output of a gluon-only parton shower, which not only reproduces the final distribution of particles but also learns the underlying showering mechanism using the complete event information. We therefore aim at achieving algorithmic transparency [69–71], where the generation of intermediate splitting momentum fraction and angular ordering can be fully understood by human physicists. The technical implementation is a neural network architecture using recurrent, interpretable representation. Unlike traditional post-hoc explainability, a fully func-

* Corresponding author.
E-mail address: ylai@lbl.gov (Y.S. Lai).

https://doi.org/10.1016/j.physletb.2022.137055

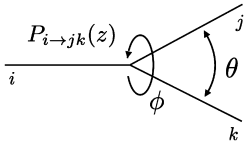

Fig. 1. Parton splitting process $i \to jk$ with longitudinal momentum fraction z, relative splitting angle of the two daughter partons θ and azimuthal angle ϕ.

tional proxy model can be created, and it is possible to run the generator network with proxy model replacing the neural network. This differs from post-hoc methods where the explaination does not offer the path towards a replacement model.

GANs consist of two competing neural networks, the generator and discriminator. In our setup, the generator network is structured in analogy to a Recurrent Neural Network (RNN). Partons are generated through $1 \to 2$ splittings which are taken as input for the next iteration. This "constrained GAN" approach thus mimics the structure of the parton shower and can give access to the underlying physics in the parton branching mechanism. More specifically, we demonstrate that the network can learn the Altarelli-Parisi splitting function $P_{i\to jk}(z)$, the splitting angles of individual branching processes and the dependence of the shower on the energy scale Q, see Fig. 1. This is achieved by separating the GAN into two components such that it can learn both self-similar/fractal aspects of the shower such as the Altarelli-Parisi splitting function as well as Monte Carlo evolution time dependent variables such as the splitting angle. We employ a network architecture that is sufficiently general, and as a result, capable of incorporating nonperturbative physics in the future. In order to use the complete information of each event, we use a data representation which is directly given by the four-vectors of the final state particles. To avoid sensitivity to the unphysical ordering of the list of four-vectors during the training process, we use sets to represent the data. In particular, in our work, the necessary permutation invariance is achieved by using so-called deep sets which were developed in Refs. [72–74].

With the framework introduced in this work, we can access the underlying physics mechanisms effectively departing from the typical black-box paradigm for neural networks. Moreover, we expect that eventually the GAN can be trained directly on experimental data (i.e. measured four-vectors of detected particles). Generally, GANs are ideally suited for such applications due to their generalizability and robustness when exposed to imperfect data sets. We expect that our approach will be particularly relevant for studies of heavy-ion collisions at RHIC and the LHC as well as electron-nucleus collisions at the future Electron-Ion Collider [75]. In heavy-ion collisions, the presence of quark-gluon plasma (QGP) [76–84] leads to modifications of highly energetic jets as compared to the proton-proton baseline. These phenomena are typically referred to as *jet quenching*. Significant theoretical [52–60,62–64,66] and experimental [85–89] efforts have been made to better understand the physics of this process. Using the novel techniques proposed in this work, we will eventually be able to analyze the properties of the medium modified parton shower using, for the first time, the complete event information.

2. The parton shower

The parton shower we use for training the GAN is designed to solve the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) evolution equations, see Refs. [51,90]. In addition, we set up the full event kinematics in spherical coordinates such that we can use the final distribution of partons generated by the shower as input to the adversarial training process. We start with a highly energetic parton which originates from a hard-scattering event at the scale Q. The parton shower cascade is obtained through recursive $1 \to 2$ branching processes according to the DGLAP evolution equations. There are three variables that describe a DGLAP splitting process $i \to jk$ as illustrated in Fig. 1. First, the large light cone momentum fraction z of the daughter partons relative to the parent is determined by sampling from the Altarelli-Parisi splitting functions. Second, the orientation of the two daughter partons, the azimuthal angle ϕ, is obtained by sampling from a flat distribution in the range $[-\pi, \pi]$. Third, the splitting angle θ which is the relative opening angle of the two daughter partons, is determined as follows: First, sample a Monte Carlo time step Δt from the no-emission Sudakov factor

$$\exp\left[-\Delta t \sum_{i=q,\bar{q},g} \int_{\epsilon}^{1-\epsilon} \mathrm{d}z\, P_i(z) \right], \tag{1}$$

where the P_i denote the final state summed Altarelli-Parisi splitting functions for (anti-)quarks and gluons. Then advance the shower time $t \to t + \Delta t$ and solve for the splitting angle θ in

$$t(Q, \theta) = \int_{Q\tan(\pi/2)}^{Q\tan(\theta/2)} \frac{\mathrm{d}t'}{t'} \frac{\alpha_s(t')}{\pi}. \tag{2}$$

We evolve the shower from the hard scale Q down to the hadronization scale which we choose as 1 GeV. We note that the DGLAP shower described here has two cutoff parameters. First, the angular cutoff on the splitting angle θ which is introduced by the hadronization scale and which determines the end of the shower. Second, we introduce the cutoff ϵ on the momentum fraction z, see Eq. (1). For our numerical results we choose $\epsilon = 0.02$ which avoids the singular endpoints. The generated spectrum is accurate in the range $\epsilon < z < 1 - \epsilon$, and emitted partons that violate these bounds are not evolved further in the shower.

From the parent direction and the variables (z, θ, ϕ) of a given $1 \to 2$ splitting, we set up the full event kinematics and determine the absolute position of the two daughter partons in spherical coordinates $(\tilde{\Theta}, \tilde{\Phi})$. The relevant kinematic relations are summarized in the supplementary material. After the shower terminates, we record the final momentum fractions Z of the partons relative to the initial momentum scale Q as well as their corresponding spherical coordinates (Θ, Φ).[1] Together with the on-shell condition they fully specify the exclusive final state distribution of all particles which are produced by the shower. We note that the variables z, ϕ are independent of the shower time t (self-similar or fractal variables), whereas the splitting angle θ is determined from the ordering variable of the shower and it also depends on the scale Q. Therefore, we treat θ differently from the other two variables in the generator network, as discussed below. The shower described here provides an ideal testing ground to explore the use of explainable machine learning that aims to extract the structure of the parton shower, and thus the underlying physics, from the final distribution of particles in the event. We leave the investigation of other shower algorithms and nonperturbative effects for future work.

3. Data representation and setup of the GAN

To avoid any loss of information, we choose to train the GAN directly on sets which contain the event-by-event particle four-vectors produced by the shower introduced above. The required

[1] Note that we use the variables (z, θ, ϕ) to describe an individual $1 \to 2$ splitting processes as shown in Fig. 1, $(\tilde{\Theta}, \tilde{\Phi})$ are the spherical coordinates of partons at intermediate stages of the shower and (Z, Θ, Φ) denote the final distributions of the momentum fraction and angles of the partons after the shower terminates.

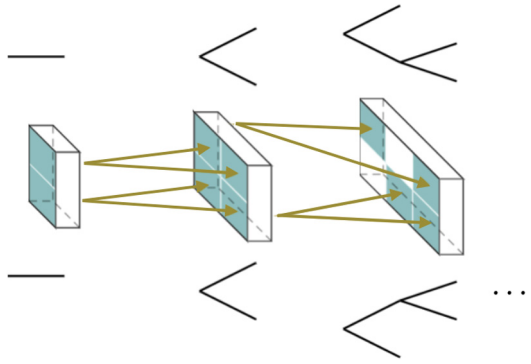

Fig. 2. Schematic illustrations of the data structure of the generator network: Parallelized execution of the random splitting trees on the GPU.

permutation invariance is built into the discriminator network by using so-called deep sets which were developed in Refs. [72–74]. Several equivariant layers are followed by a permutation invariant layer which ensures that the discriminator network is insensitive to the ordering of the input. Since the number of particles that are produced per event fluctuates, the sets of four-vectors have variable length. Deep sets are ideally suited to handle input with different lengths. To accommodate the variable length of the training data we allow the deep sets to contain up to 200 four-vectors which is sufficient for the energy Q that we consider here.

The discriminator network consists of a sequence of two deep sets. The first deep sets network takes the list of parton momenta $p_1, \ldots, p_M$ from the shower as input. A permutation invariant deep set f is obtained as

$$f(p_1, \ldots, p_M) = \rho\left(\sum_{i=1}^{M} \Phi(p_i)\right). \tag{3}$$

Here Φ, ρ denote fully connected neural networks with 3, 2 layers, respectively, with 50 neurons per layer. We use the Leaky ReLU activation function with $\alpha = 0.2$ [91]. The summation operation in Eq. (3) makes the final result permutation invariant. The output layer of these per-event deep sets has dimension 10. The second deep sets network uses the output of the first network from multiple events as input and produces the statistical activation for the entire batch of events. The second layer of deep sets allows us to avoid the mode collapse of the GAN [92].

We note that it is also possible to train the network on a set of observables where Infrared-Collinear safety is built in directly [22,93]. We plan to explore the impact of different data representations in future work which will be particularly relevant once we include nonperturbative effects in the shower.

The generator network mimics the structure of a parton shower. It sequentially produces partons and learns to map n to $n+1$ partons. To simplify the training process, the generator is separated into a Monte Carlo time-dependent and time-independent part. The time-independent part is designed to learn the Altarelli-Parisi splitting function $P_{i\to jk}$ and the azimuthal angle ϕ which are the same for every branching process and independent of Q. Whereas the other part of the network depends on the Monte Carlo time t and on the energy Q, i.e. it changes at every step of the shower and produces emissions which are ordered in the splitting angle θ, see Eq. (2). Both parts of the generator consist of neural networks with 5 hidden layers and 50 neurons, which is illustrated schematically in Figs. 2, 3. We use the exponential linear unit (ELU) [94] as the activation function, to avoid step functions in the resulting z and θ distributions. The separation into two networks simplifies the training process since, for example, the network does not need to consider a Monte Carlo time dependent splitting function. However, we expect that with sufficient computing resources this separation could be removed. We note that the two shower cutoffs discussed above are also explicitly included in the generator network. However, in general, we expect that the cutoffs can be chosen as trainable parameters as well.

Using the shower setup described above, we generate training data for different energies in the range of $Q = 200$–800 GeV. As a proof of concept, we study a pure gluon shower where the gluon that splits is chosen at random. The training process of the GAN is a modified version of the original GAN approach. More details are given in the supplementary material.

4. Numerical results

We first verify that the GAN can reproduce the final distribution of particles and we then consider the underlying physics by sampling from the different units of the RNN. To quantify the agreement between the shower and the GAN, we consider three kinematic variables (Z, Θ, Φ) which characterize the final distribution of particles. The result of the GAN and the parton shower is shown in the three panels of Fig. 4, where 3.5×10^8 events from the GAN after 500 training epochs is compared to 3.5×10^7 parton shower events. We show two types of errors in Fig. 4 (and similarly Fig. 5). The statistical uncertainty of the GAN is negligible and not shown. The markers in the upper and lower panels represent the generated parton shower (PS) distributions and their statistical uncertainty. In addition, we provide a band in the lower panel which represents the statistical uncertainty of 1 standard deviation within the training sample. We observe good agreement which is consistent with the statistical uncertainty of the training sample. The agreement over several orders of magnitude is highly nontrivial even without considering the underlying physics. As expected for a DGLAP shower, the distribution of the parton momentum fractions rises steeply toward small-Z (left panel). The distribution of the polar angle Θ peaks in the direction of the initial parton and Φ is flat which is consistent with the flat sampling of ϕ for each individual splitting.

Having confirmed that the GAN can reproduce the final output of the parton shower, we are now going to analyze the individual splitting processes to verify that the network has also correctly learned the underlying physics. The ability of the GAN to extract information about parton branching mechanism is the main novelty of our work. By sampling from the different units of the RNN architecture, we study the distribution of the variables (z, θ) that characterize the individual splitting processes. As representative examples, we show the results for the first four splittings in the left and middle panel of Fig. 5. The distribution of the momentum fraction z is shown in the left panel for the $g \to gg$ splitting process. We observe good agreement with the Altarelli-Parisi splitting function $P_{g\to gg}$ for all four splittings. In particular, we note that the splitting function diverges for $z \to 1$. Instead, the final Z-distribution (left panel in Fig. 4) falls off steeply toward $Z \to 1$ as expected for a QCD fragmentation spectrum. The strikingly different behavior of the two distributions near the end point clearly demonstrates that the GAN has in fact learned the underlying physics mechanism. Next we consider the Monte Carlo time-dependent θ distribution which is shown in the middle panel of Fig. 5. We observe that it is correctly reproduced by the GAN besides small fluctuations in the tail. The distributions peak at small values of θ. As expected for the ordering variable of the shower, the distributions become more narrow for splittings that occur at later Monte Carlo time. Here, θ is the only variable that depends on the scale Q. We investigate its Q dependence by considering the first splitting of the shower which is shown in the right panel of Fig. 5. Even though the GAN is optimized to reproduce only the Q-integrated distribution, the Q-dependence of the

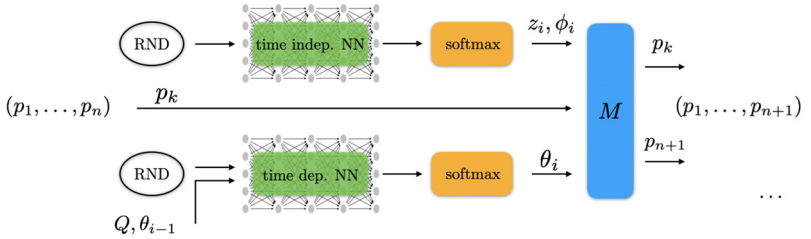

Fig. 3. Flow diagram of the ith splitting process ($n \to n+1$ partons) of a randomly chosen parton with momentum p_k. The Monte Carlo time dependent and independent networks are shown which take as input random numbers (RND) as well as Q, θ_{i-1} in the time dependent case. The output of the two neural networks is passed through a softmax function to the module M which determines the four-vectors of the two daughter partons from the variables of the $1 \to 2$ splitting process and the parent momentum p_k.

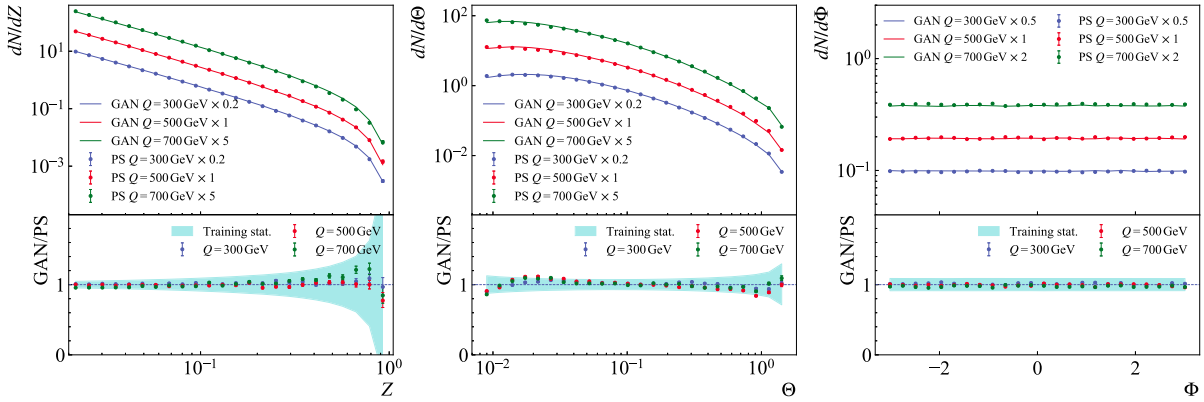

Fig. 4. Comparison of the parton shower and GAN in terms of the final distribution of particles. The three panels show the momentum fraction Z, the polar angle Θ and the azimuthal angle Φ (from left to right) for $Q = 300, 500, 700$ GeV.

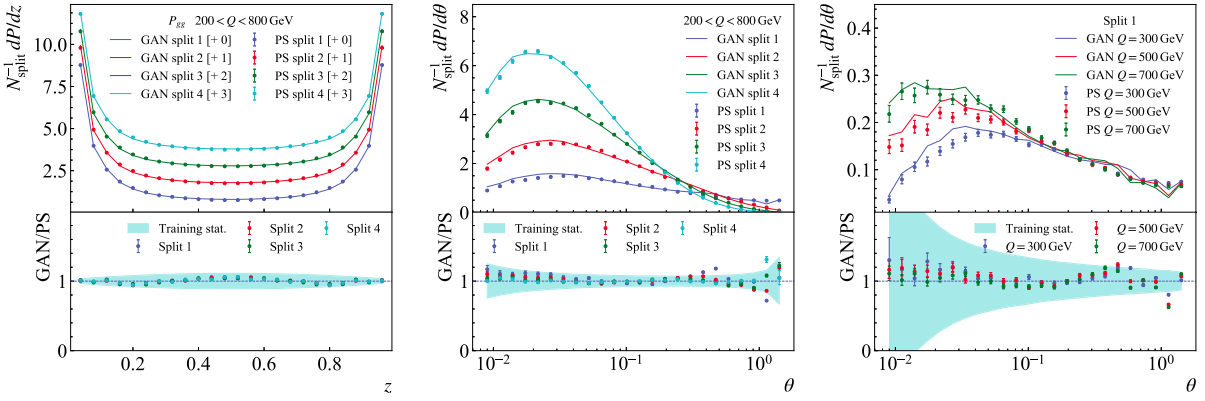

Fig. 5. Comparison of the momentum fraction z, i.e. the Altarelli-Parisi splitting function $P_{g\to gg}(z)$ (left) and the relative splitting angle θ (middle) of the first four splittings from the parton shower and the GAN for $Q = 200-800$ GeV. In addition, we show the θ distribution for three different values of Q for the first splitting (right).

shower is nevertheless well described by the network. We attribute the remaining numerical differences to the finite number of neurons in combination with the activation function and their ability to approximate a steep multi-differential distribution. This can be mitigated by extending the size of the neural network and increasing the size of the training sample. We note that the underlying multidimensional probability distributions are non-linear, contain (integrable) endpoint divergences and span over several orders of magnitude which is nontrivial to capture accurately by the GAN. For the kinematics we consider here, the shower produces on average 60-70 splittings with a maximum of 200. In Fig. 6 we show the z and θ distributions for the splittings 10, 20, 30 and 40. We find good agreement between the parton shower and GAN, even though the θ distributions are steeply falling. The accurate modeling of these splittings which occur at late Mote Carlo times is necessary to reliably pin down the underlying physics mechanisms which is demonstrated here for the first time. Lastly, we find that the distribution of the azimuthal angle ϕ (not shown) also agrees with the parton shower result and we thus conclude that the GAN has in fact accurately learned the underlying physics of the parton shower.

5. Conclusions and outlook

In this letter we proposed an explainable machine learning - a White Box AI - framework which successfully learns the underlying physics of a parton shower - a hallmark of modeling high-energy particle collisions. As a proof of concept, we demonstrated that constrained Generative Adversarial Networks (GANs) using the full event information are capable of learning the parton cascade as described by a parton shower implementing DGLAP evolution equations. As input to the adversarial training process

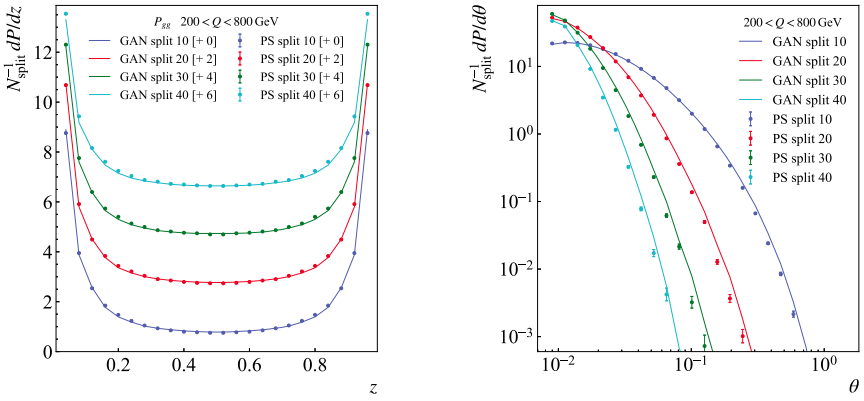

Fig. 6. The momentum fraction z and the relative splitting angle θ of the 10th to 40th splittings for the same kinematics as in Fig. 5.

we used deep sets which yield a permutation invariant representation of the training data of variable length. We found that not only the final distribution of partons in the event can be described by the network but also the physics of individual splittings processes are correctly learned by the GAN. This is achieved by using a constrained GAN, where the generator network is structured like a recursive neural network (RNN). The parton branching mechanism can be learned from the different units of the network. We consider our work as a starting point of a long-term effort with the goal to eventually train networks directly on experimental data designed for extracting the underlying physics using full event information registered in the detectors. We note that the precision of our approach in falsifying theoretical modeling is limited by the systematic experimental biases which we plan to explore in subsequent publications. An important future direction is the inclusion of nonperturbative effects in our framework. The nonperturbative transition from parton to hadron level can be included as an additional unit in the generator network which converts N partons to M hadrons after the perturbative shower terminates. This additional unit can either be treated as a black box or additional constraints from QCD can be included and the missing physics can be learned directly from data. Furthermore, we expect that our results to be particularly relevant for future studies of collective effects, and the modification of the vacuum parton shower in heavy-ion collisions or electron-nucleus collisions at the future Electron-Ion Collider.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to thank Barbara Jacak, James Mulligan, Stefan Prestel, Nobuo Sato and Feng Yuan for helpful discussions. YSL, MP and FR are supported by the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 and the LDRD Program of Lawrence Berkeley National Laboratory. DN is supported by the U.S. Department of Energy under Contract No. DE-AC52-06NA25396 at LANL and through the LANL/LDRD Program.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.physletb.2022.137055.

References

[1] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman, A. Schwartzman, J. High Energy Phys. 07 (2016) 069.
[2] P.T. Komiske, E.M. Metodiev, M.D. Schwartz, J. High Energy Phys. 01 (2017) 110.
[3] G. Kasieczka, T. Plehn, M. Russell, T. Schell, J. High Energy Phys. 05 (2017) 006.
[4] E.M. Metodiev, B. Nachman, J. Thaler, J. High Energy Phys. 10 (2017) 174.
[5] C. Englert, P. Galler, P. Harris, M. Spannowsky, Eur. Phys. J. C 79 (2019) 4.
[6] B. Hashemi, N. Amin, K. Datta, D. Olivito, M. Pierini, arXiv:1901.05282 [hep-ex], 2019.
[7] S. Otten, et al., arXiv:1901.00875 [hep-ph], 2019.
[8] A. Butter, T. Plehn, R. Winterhalder, SciPost Phys. 7 (2019) 075.
[9] R. Di Sipio, M. Faucci Giannelli, S. Ketabchi Haghighat, S. Palazzo, J. High Energy Phys. 08 (2019) 110.
[10] S. Farrell, et al., EPJ Web Conf. 214 (2019) 09005.
[11] Y. Alanazi, et al., arXiv:2001.11103 [hep-ph], 2020.
[12] L.-G. Pang, et al., Nat. Commun. 9 (2018) 210.
[13] P.T. Komiske, E.M. Metodiev, B. Nachman, M.D. Schwartz, J. High Energy Phys. 12 (2017) 051.
[14] R.D. Ball, et al., Eur. Phys. J. C 77 (2017) 663.
[15] M. Paganini, L. de Oliveira, B. Nachman, Phys. Rev. Lett. 120 (2018) 042003.
[16] K. Datta, A.J. Larkoski, J. High Energy Phys. 03 (2018) 086.
[17] A.J. Larkoski, I. Moult, B. Nachman, Phys. Rep. 841 (2020) 1.
[18] Y.-T. Chien, R. Kunnawalkam Elayavalli, arXiv:1803.03589 [hep-ph], 2018.
[19] J.H. Collins, K. Howe, B. Nachman, Phys. Rev. Lett. 121 (2018) 241803.
[20] K. Zhou, G. Endrődi, L.-G. Pang, H. Stöcker, Phys. Rev. D 100 (2019) 011501.
[21] Y.S. Lai, arXiv:1810.00835 [nucl-th], 2018.
[22] P.T. Komiske, E.M. Metodiev, J. Thaler, J. High Energy Phys. 01 (2019) 121.
[23] Y.-L. Du, et al., Eur. Phys. J. C 80 (2020) 516.
[24] L.-G. Pang, K. Zhou, X.-N. Wang, arXiv:1906.06429 [nucl-th], 2019.
[25] A. Andreassen, P.T. Komiske, E.M. Metodiev, B. Nachman, J. Thaler, Phys. Rev. Lett. 124 (2020) 182001.
[26] S. Carrazza, F.A. Dreyer, Phys. Rev. D 100 (2019) 014014.
[27] G. Kasieczka, S. Marzani, G. Soyez, G. Stagnitto, arXiv:2007.04319 [hep-ph], 2020.
[28] L. Li, Y.-Y. Li, T. Liu, S.-J. Xu, arXiv:2004.15013 [hep-ph], 2020.
[29] G. Kanwar, et al., Phys. Rev. Lett. 125 (2020) 121601.
[30] I.J. Goodfellow, et al., in: Proceedings of NIPS'14, Cambridge, MA, USA, 2014, pp. 2672–2680.
[31] A. Radford, L. Metz, S. Chintala, CoRR, arXiv:1511.06434 [abs], 2016.
[32] A. Andreassen, I. Feige, C. Frye, M.D. Schwartz, Eur. Phys. J. C 79 (2019) 102.
[33] A. Andreassen, I. Feige, C. Frye, M.D. Schwartz, Phys. Rev. Lett. 123 (2019) 182001.
[34] J. Monk, J. High Energy Phys. 12 (2018) 021.
[35] A. Bogatskiy, et al., arXiv:2006.04780 [hep-ph], 2020.
[36] A.J. Larkoski, T. Melia, arXiv:2008.06508 [hep-ph], 2020.
[37] T. Faucett, J. Thaler, D. Whiteson, arXiv:2010.11998 [hep-ph], 2020.
[38] T. Sjostrand, S. Mrenna, P.Z. Skands, Comput. Phys. Commun. 178 (2008) 852.
[39] M. Bahr, et al., Eur. Phys. J. C 58 (2008) 639.
[40] T. Gleisberg, et al., J. High Energy Phys. 02 (2009) 007.
[41] Z. Nagy, D.E. Soper, J. High Energy Phys. 06 (2012) 044.
[42] S. Höche, S. Prestel, Eur. Phys. J. C 75 (2015) 461.
[43] S. Alioli, C.W. Bauer, C. Berggren, F.J. Tackmann, J.R. Walsh, Phys. Rev. D 92 (2015) 094020.
[44] M. Dasgupta, F.A. Dreyer, K. Hamilton, P.F. Monni, G.P. Salam, J. High Energy Phys. 09 (2018) 033, Erratum: J. High Energy Phys. 03 (2020) 083.
[45] G. Bewick, S. Ferrario Ravasio, P. Richardson, M.H. Seymour, J. High Energy Phys. 04 (2020) 019.
[46] M. Dasgupta, et al., Phys. Rev. Lett. 125 (2020) 052002.

[47] J.R. Forshaw, J. Holguin, S. Plätzer, arXiv:2003.06400 [hep-ph], 2020.
[48] B. Andersson, G. Gustafson, G. Ingelman, T. Sjostrand, Phys. Rep. 97 (1983) 31.
[49] G. Marchesini, B. Webber, Nucl. Phys. B 310 (1988) 461.
[50] A. Metz, A. Vossen, Prog. Part. Nucl. Phys. 91 (2016) 136.
[51] D. Neill, F. Ringer, N. Sato, in: 10th International Conference on Hard and Electromagnetic Probes of High-Energy Nuclear Collisions: Hard Probes 2020, 2020.
[52] M. Gyulassy, X.-n. Wang, Nucl. Phys. B 420 (1994) 583.
[53] R. Baier, Y.L. Dokshitzer, A.H. Mueller, S. Peigne, D. Schiff, Nucl. Phys. B 484 (1997) 265.
[54] B. Zakharov, JETP Lett. 63 (1996) 952.
[55] M. Gyulassy, P. Levai, I. Vitev, Nucl. Phys. B 594 (2001) 371.
[56] X.-N. Wang, X.-f. Guo, Nucl. Phys. A 696 (2001) 788.
[57] P.B. Arnold, G.D. Moore, L.G. Yaffe, J. High Energy Phys. 06 (2002) 030.
[58] J.-w. Qiu, I. Vitev, Phys. Lett. B 632 (2006) 507.
[59] H. Liu, K. Rajagopal, U.A. Wiedemann, Phys. Rev. Lett. 97 (2006) 182301.
[60] N. Armesto, et al., Phys. Rev. C 86 (2012) 064904.
[61] Y. Mehtar-Tani, J.G. Milhano, K. Tywoniuk, Int. J. Mod. Phys. A 28 (2013) 1340013.
[62] K.M. Burke, et al., Phys. Rev. C 90 (2014) 014909.
[63] J.-W. Qiu, F. Ringer, N. Sato, P. Zurita, Phys. Rev. Lett. 122 (2019) 252301.
[64] J. Putschke, et al., arXiv:1903.07706 [nucl-th], 2019.
[65] P. Caucal, E. Iancu, G. Soyez, J. High Energy Phys. 10 (2019) 273.
[66] V. Vaidya, X. Yao, arXiv:2004.11403 [hep-ph], 2020.
[67] F. de Avila Belbute-Peres, K. Smith, K. Allen, J. Tenenbaum, J.Z. Kolter, in: S. Bengio, et al. (Eds.), Advances in Neural Information Processing Systems, vol. 31, Curran Associates, Inc., 2018, pp. 7178–7189.
[68] A. Koul, S. Greydanus, A. Fern, CoRR, arXiv:1811.12530 [abs], 2018.
[69] Z.C. Lipton, CoRR, arXiv:1606.03490 [abs], 2016.
[70] A.B. Arrieta, et al., CoRR, arXiv:1910.10045 [abs], 2019.
[71] F. Fan, J. Xiong, G. Wang, CoRR, arXiv:2001.02522 [abs], 2020.
[72] M. Zaheer, et al., CoRR, arXiv:1703.06114 [abs], 2017.
[73] E. Wagstaff, F.B. Fuchs, M. Engelcke, I. Posner, M.A. Osborne, CoRR, arXiv:1901.09006 [abs], 2019.
[74] B. Bloem-Reddy, Y. Teh, J. Mach. Learn. Res. 21 (2020) 1.
[75] A. Accardi, et al., Eur. Phys. J. A 52 (2016) 268.
[76] J.D. Bjorken, Phys. Rev. D 27 (1983) 140.
[77] I. Arsene, et al., Nucl. Phys. A 757 (2005) 1.
[78] K. Adcox, et al., Nucl. Phys. A 757 (2005) 184.
[79] B. Back, et al., Nucl. Phys. A 757 (2005) 28.
[80] J. Adams, et al., Nucl. Phys. A 757 (2005) 102.
[81] B.V. Jacak, B. Muller, Science 337 (2012) 310.
[82] B. Müller, J. Schukraft, B. Wysłouch, Annu. Rev. Nucl. Part. Sci. (2012) 62–361.
[83] P. Braun-Munzinger, V. Koch, T. Schäfer, J. Stachel, Phys. Rep. 621 (2016) 76.
[84] W. Busza, K. Rajagopal, W. van der Schee, Annu. Rev. Nucl. Part. Sci. 68 (2018) 339.
[85] A. Adare, et al., Phys. Rev. C 84 (2011) 044905.
[86] A.M. Sirunyan, et al., Eur. Phys. J. C 78 (2018) 509.
[87] L. Adamczyk, et al., Phys. Rev. C 96 (2017) 024905.
[88] S. Acharya, et al., Phys. Rev. C 101 (2020) 034911.
[89] M. Aaboud, et al., Phys. Lett. B 790 (2019) 108.
[90] M. Dasgupta, F. Dreyer, G.P. Salam, G. Soyez, J. High Energy Phys. 04 (2015) 039.
[91] A.L. Maas, A.Y. Hannun, A.Y. Ng, in: ICML Workshop on Deep Learning for Audio, Speech and Language Processing, 2013.
[92] T. Salimans, et al., in: D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, R. Garnett (Eds.), Advances in Neural Information Processing Systems, vol. 29, Curran Associates, Inc., 2016.
[93] M.J. Dolan, A. Ore, arXiv:2012.00964 [hep-ph], 2020.
[94] D.-A. Clevert, T. Unterthiner, S. Hochreiter, CoRR, arXiv:1511.07289 [abs], 2016.