Lawrence Berkeley National Laboratory
LBL Publications

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

Permalink
https://escholarship.org/uc/item/04c5j5ks

Authors
Lai, Yue Shi
Neill, Duff
Płoskoń, Mateusz
et al.

Publication Date
2020-12-11

Peer reviewed

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

Yue Shi Lai,[1, *] Duff Neill,[2, †] Mateusz Płoskoń,[1, ‡] and Felix Ringer[1, §]

[1] *Nuclear Science Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*
[2] *Theoretical Division, MS B283, Los Alamos National Laboratory, Los Alamos, NM 87545, USA*
(Dated: December 15, 2020)

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. 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.

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] a probabilistic model was introduced based on jet clustering and in Ref. [33] a convolutional autoencoder within a shower was used which qualitatively reproduces jet observables. See also Refs. [34–36] 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 [37–39]. Starting with highly energetic quarks or gluons which are produced in hard-scattering events, parton showers simulate the parton branching processes that occur during the evolution from the hard scale to the infrared which is followed by the hadronization step. While the general concept of parton showers is well established, important questions about the perturbative accuracy [40–46], nonperturbative effects [47–50] and the modification in the nuclear environment [51–65], remain a challenge.

In this work, we propose an explainable or White Box AI approach [66, 67] 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 parton shower, which not only reproduces the final distribution of particles but also learns the under-
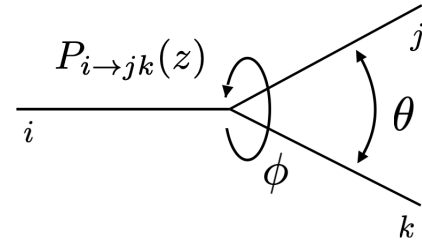

FIG. 1. Parton splitting process $i \to jk$ with longitudinal momentum fraction z, relative splitting angle of the two daughter partons θ and azimuthal angle ϕ.

lying showering mechanism using the complete event information. GANs consist of two competing neural networks, the generator and discriminator. The design of our generator network allows to not only describe the final distribution of particles of the shower but the different layers also give access to the underlying physics encoded in the parton branching processes. 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 like the Altarelli-Parisi splitting function as well as Monte Carlo time dependent variables such as the splitting angle. We use 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 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
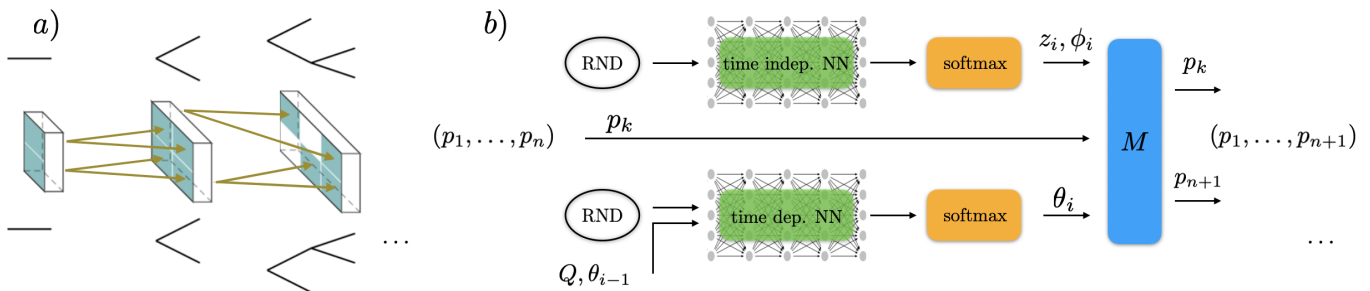
FIG. 2. Schematic illustrations of the generator network: *a)* Parallelized data structure of the random splitting trees executed on the GPU. *b)* Flow diagram of the *i*th splitting process ($n \rightarrow n+1$ partons) of a randomly chosen parton with momentum p_k. The time dependent and independent 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 to the module M which determines the four-vectors of the two daughter partons from the variables of the $1 \rightarrow 2$ splitting process and the parent momentum p_k.

data. In particular, in our work, the necessary permutation invariance is achieved by using so-called deep sets which were developed in Refs. [68–70].

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 [71]. In heavy-ion collisions, the presence of quark-gluon plasma (QGP) [72–80] 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 [51–59, 61–63, 65] and experimental [81–85] 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.

The parton shower. The parton shower we use for training the GAN is designed to solve the DGLAP evolution equations, see Refs. [50, 86]. 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 $1 \rightarrow 2$ branching processes according to the DGLAP evolution equations. There are three variables that describe a DGLAP splitting process $i \rightarrow jk$ as illustrated in Fig. 1. First, the large light cone momentum fraction z of the daughter partons relative to the parent is de-

termined 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

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

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

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

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 $\epsilon = 0.03$ 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 $1 \rightarrow 2$ 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 supplemental material. After the shower terminates, we record the final momentum fractions Z of the partons relative to the initial momentum scale Q as well
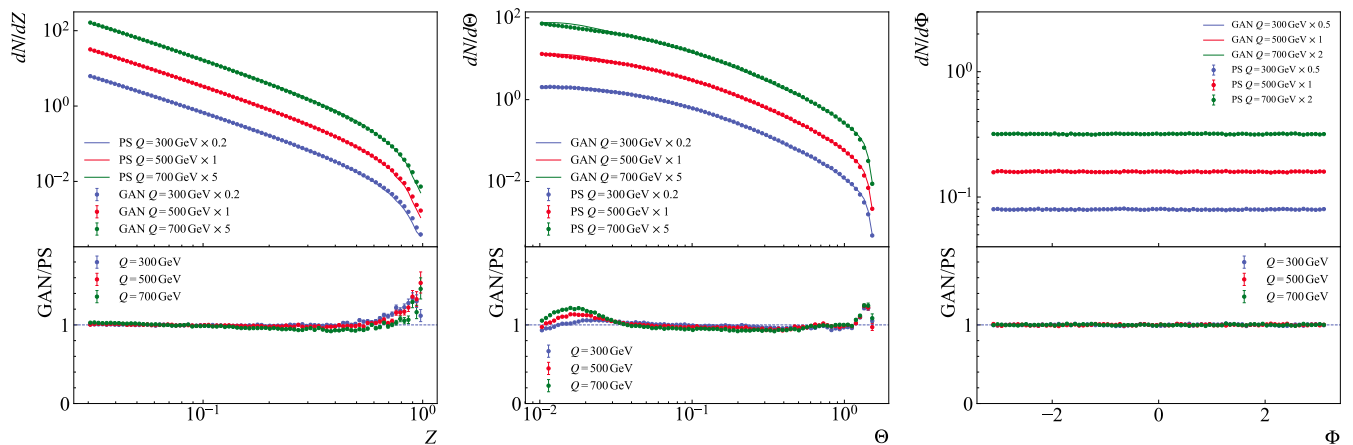
FIG. 3. Comparison of the parton shower and GAN in terms of the final distribution of particles. The three panels show the momentum fraction Z, the polar angle Θ and the azimuthal angle Φ (from left to right) for $Q = 300, 500, 700$ GeV.

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 test 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.

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. The required permutation invariance is built into the discriminator network by using so-called deep sets which were developed in Refs. [68–70]. 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. 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, 87]. 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 Fig. 2. We use the exponential linear unit (ELU) [88] as the activation function, to avoid step functions in the resulting z and θ distributions. 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 supplemental material.

Numerical results. We first verify that the GAN can reproduce the final distribution of particles and we then

---

[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.
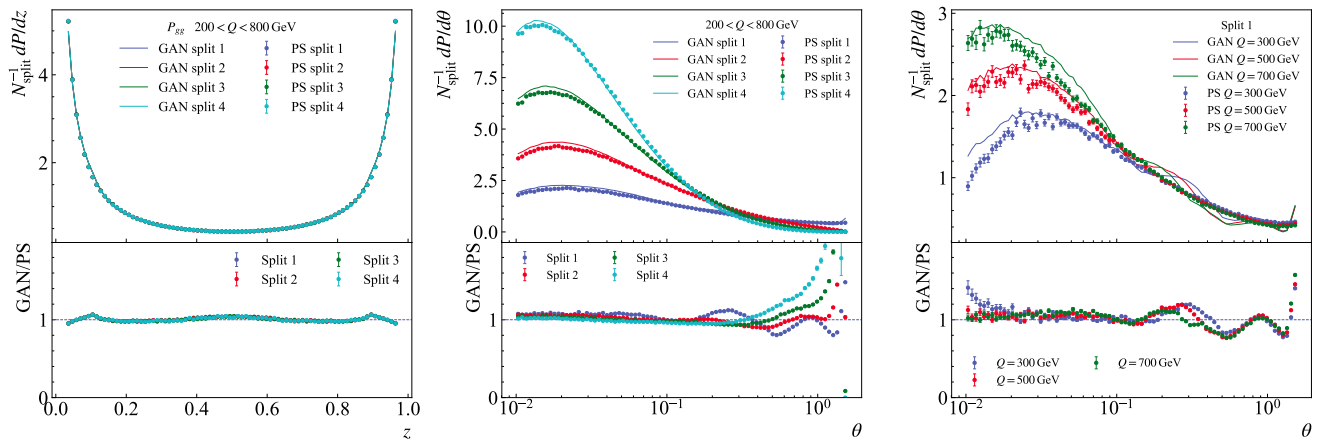
FIG. 4. Comparison of the momentum fraction z, i.e. the Altarelli-Parisi splitting function $P_{g \to gg}(z)$ (left) and the relative splitting angle θ (middle) of the first four splittings from the parton shower and the GAN for $Q = 200 - 800$ GeV. In addition, we show the θ distribution for three different values of Q for the first splitting (right).

consider the underlying physics by sampling from the different layers of the network. 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. 3, where 3.5×10^8 events from the GAN after 700 training epochs is compared to 3.5×10^7 parton shower events. We observe very good agreement for all three distributions. The good 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 different layers of the network, 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. 4. The distribution of the momentum fraction z is shown in the left panel for the $g \to gg$ splitting process. We observe very 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. 3) 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. 4. 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. 4. Even though the GAN is optimized to reproduce only the Q-integrated distribution, the Q-dependence of the 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. 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.

Conclusions. 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 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 we used deep sets which yield a permutation invariant representation of the training data of vari-

able 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. 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. We expect our results to be particularly relevant for future studies of nonperturbative physics, collective effects, and the modification of the vacuum parton shower in heavy-ion collisions or electron-nucleus collisions at the future Electron-Ion Collider.

---

* ylai@lbl.gov
† duff.neill@gmail.com
‡ mploskon@lbl.gov
§ fmringer@lbl.gov

[1] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman, and A. Schwartzman, "Jet-images — deep learning edition," *JHEP* **07** (2016) 069, arXiv:1511.05190 [hep-ph].

[2] P. T. Komiske, E. M. Metodiev, and M. D. Schwartz, "Deep learning in color: towards automated quark/gluon jet discrimination," *JHEP* **01** (2017) 110, arXiv:1612.01551 [hep-ph].

[3] G. Kasieczka, T. Plehn, M. Russell, and T. Schell, "Deep-learning Top Taggers or The End of QCD?," *JHEP* **05** (2017) 006, arXiv:1701.08784 [hep-ph].

[4] E. M. Metodiev, B. Nachman, and J. Thaler, "Classification without labels: Learning from mixed samples in high energy physics," *JHEP* **10** (2017) 174, arXiv:1708.02949 [hep-ph].

[5] C. Englert, P. Galler, P. Harris, and M. Spannowsky, "Machine Learning Uncertainties with Adversarial Neural Networks," *Eur. Phys. J. C* **79** no. 1, (2019) 4, arXiv:1807.08763 [hep-ph].

[6] B. Hashemi, N. Amin, K. Datta, D. Olivito, and M. Pierini, "LHC analysis-specific datasets with Generative Adversarial Networks," arXiv:1901.05282 [hep-ex].

[7] S. Otten, S. Caron, W. de Swart, M. van Beekveld, L. Hendriks, C. van Leeuwen, D. Podareanu, R. Ruiz de Austri, and R. Verheyen, "Event Generation and Statistical Sampling for Physics with Deep Generative Models and a Density Information Buffer,"
arXiv:1901.00875 [hep-ph].

[8] A. Butter, T. Plehn, and R. Winterhalder, "How to GAN LHC Events," *SciPost Phys.* **7** no. 6, (2019) 075, arXiv:1907.03764 [hep-ph].

[9] R. Di Sipio, M. Faucci Giannelli, S. Ketabchi Haghighat, and S. Palazzo, "DijetGAN: A Generative-Adversarial Network Approach for the Simulation of QCD Dijet Events at the LHC," *JHEP* **08** (2019) 110, arXiv:1903.02433 [hep-ex].

[10] S. Farrell, W. Bhimji, T. Kurth, M. Mustafa, D. Bard, Z. Lukic, B. Nachman, and H. Patton, "Next Generation Generative Neural Networks for HEP," *EPJ Web Conf.* **214** (2019) 09005.

[11] Y. Alanazi, N. Sato, T. Liu, W. Melnitchouk, M. P. Kuchera, E. Pritchard, M. Robertson, R. Strauss, L. Velasco, and Y. Li, "Simulation of electron-proton scattering events by a Feature-Augmented and Transformed Generative Adversarial Network (FAT-GAN)," arXiv:2001.11103 [hep-ph].

[12] L.-G. Pang, K. Zhou, N. Su, H. Petersen, H. Stöcker, and X.-N. Wang, "An equation-of-state-meter of quantum chromodynamics transition from deep learning," *Nature Commun.* **9** no. 1, (2018) 210, arXiv:1612.04262 [hep-ph].

[13] P. T. Komiske, E. M. Metodiev, B. Nachman, and M. D. Schwartz, "Pileup Mitigation with Machine Learning (PUMML)," *JHEP* **12** (2017) 051, arXiv:1707.08600 [hep-ph].

[14] **NNPDF** Collaboration, R. D. Ball *et al.*, "Parton distributions from high-precision collider data," *Eur. Phys. J. C* **77** no. 10, (2017) 663, arXiv:1706.00428 [hep-ph].

[15] M. Paganini, L. de Oliveira, and B. Nachman, "Accelerating Science with Generative Adversarial Networks: An Application to 3D Particle Showers in Multilayer Calorimeters," *Phys. Rev. Lett.* **120** no. 4, (2018) 042003, arXiv:1705.02355 [hep-ex].

[16] K. Datta and A. J. Larkoski, "Novel Jet Observables from Machine Learning," *JHEP* **03** (2018) 086, arXiv:1710.01305 [hep-ph].

[17] A. J. Larkoski, I. Moult, and B. Nachman, "Jet Substructure at the Large Hadron Collider: A Review of Recent Advances in Theory and Machine Learning," *Phys. Rept.* **841** (2020) 1–63, arXiv:1709.04464 [hep-ph].

[18] Y.-T. Chien and R. Kunnawalkam Elayavalli, "Probing heavy ion collisions using quark and gluon jet substructure," arXiv:1803.03589 [hep-ph].

[19] J. H. Collins, K. Howe, and B. Nachman, "Anomaly Detection for Resonant New Physics with Machine Learning," *Phys. Rev. Lett.* **121** no. 24, (2018) 241803, arXiv:1805.02664 [hep-ph].

[20] K. Zhou, G. Endrődi, L.-G. Pang, and H. Stöcker, "Regressive and generative neural networks for scalar field theory," *Phys. Rev. D* **100** no. 1, (2019) 011501, arXiv:1810.12879 [hep-lat].

[21] Y. S. Lai, "Automated Discovery of Jet Substructure Analyses," arXiv:1810.00835 [nucl-th].

[22] P. T. Komiske, E. M. Metodiev, and J. Thaler, "Energy Flow Networks: Deep Sets for Particle Jets," *JHEP* **01** (2019) 121, arXiv:1810.05165 [hep-ph].

[23] Y.-L. Du, K. Zhou, J. Steinheimer, L.-G. Pang, A. Motornenko, H.-S. Zong, X.-N. Wang, and H. Stöcker, "Identifying the nature of the QCD

[24] L.-G. Pang, K. Zhou, and X.-N. Wang, "Interpretable deep learning for nuclear deformation in heavy ion collisions," `arXiv:1906.06429 [nucl-th]`.

[25] A. Andreassen, P. T. Komiske, E. M. Metodiev, B. Nachman, and J. Thaler, "OmniFold: A Method to Simultaneously Unfold All Observables," *Phys. Rev. Lett.* **124** no. 18, (2020) 182001, `arXiv:1911.09107 [hep-ph]`.

[26] S. Carrazza and F. A. Dreyer, "Jet grooming through reinforcement learning," *Phys. Rev. D* **100** no. 1, (2019) 014014, `arXiv:1903.09644 [hep-ph]`.

[27] G. Kasieczka, S. Marzani, G. Soyez, and G. Stagnitto, "Towards Machine Learning Analytics for Jet Substructure," `arXiv:2007.04319 [hep-ph]`.

[28] L. Li, Y.-Y. Li, T. Liu, and S.-J. Xu, "Learning Physics at Future e^-e^+ Colliders with Machine," `arXiv:2004.15013 [hep-ph]`.

[29] G. Kanwar, M. S. Albergo, D. Boyda, K. Cranmer, D. C. Hackett, S. Racanière, D. J. Rezende, and P. E. Shanahan, "Equivariant flow-based sampling for lattice gauge theory," *Phys. Rev. Lett.* **125** no. 12, (2020) 121601, `arXiv:2003.06413 [hep-lat]`.

[30] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proceedings of NIPS'14*, pp. 2672–2680. Cambridge, MA, USA, 2014. `http://dl.acm.org/citation.cfm?id=2969033.2969125`.

[31] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *CoRR* **abs/1511.06434** (2016) .

[32] A. Andreassen, I. Feige, C. Frye, and M. D. Schwartz, "JUNIPR: a Framework for Unsupervised Machine Learning in Particle Physics," *Eur. Phys. J. C* **79** no. 2, (2019) 102, `arXiv:1804.09720 [hep-ph]`.

[33] J. Monk, "Deep Learning as a Parton Shower," *JHEP* **12** (2018) 021, `arXiv:1807.03685 [hep-ph]`.

[34] A. Bogatskiy, B. Anderson, J. T. Offermann, M. Roussi, D. W. Miller, and R. Kondor, "Lorentz Group Equivariant Neural Network for Particle Physics," `arXiv:2006.04780 [hep-ph]`.

[35] A. J. Larkoski and T. Melia, "Covariantizing Phase Space," `arXiv:2008.06508 [hep-ph]`.

[36] T. Faucett, J. Thaler, and D. Whiteson, "Mapping Machine-Learned Physics into a Human-Readable Space," `arXiv:2010.11998 [hep-ph]`.

[37] T. Sjostrand, S. Mrenna, and P. Z. Skands, "A Brief Introduction to PYTHIA 8.1," *Comput. Phys. Commun.* **178** (2008) 852–867, `arXiv:0710.3820 [hep-ph]`.

[38] M. Bahr *et al.*, "Herwig++ Physics and Manual," *Eur. Phys. J. C* **58** (2008) 639–707, `arXiv:0803.0883 [hep-ph]`.

[39] T. Gleisberg, S. Hoeche, F. Krauss, M. Schonherr, S. Schumann, F. Siegert, and J. Winter, "Event generation with SHERPA 1.1," *JHEP* **02** (2009) 007, `arXiv:0811.4622 [hep-ph]`.

[40] Z. Nagy and D. E. Soper, "Parton shower evolution with subleading color," *JHEP* **06** (2012) 044, `arXiv:1202.4496 [hep-ph]`.

[41] S. Höche and S. Prestel, "The midpoint between dipole and parton showers," *Eur. Phys. J. C* **75** no. 9, (2015) 461, `arXiv:1506.05057 [hep-ph]`.

[42] S. Alioli, C. W. Bauer, C. Berggren, F. J. Tackmann, and J. R. Walsh, "Drell-Yan production at NNLL'+NNLO matched to parton showers," *Phys. Rev. D* **92** no. 9, (2015) 094020, `arXiv:1508.01475 [hep-ph]`.

[43] M. Dasgupta, F. A. Dreyer, K. Hamilton, P. F. Monni, and G. P. Salam, "Logarithmic accuracy of parton showers: a fixed-order study," *JHEP* **09** (2018) 033, `arXiv:1805.09327 [hep-ph]`. [Erratum: JHEP 03, 083 (2020)].

[44] G. Bewick, S. Ferrario Ravasio, P. Richardson, and M. H. Seymour, "Logarithmic accuracy of angular-ordered parton showers," *JHEP* **04** (2020) 019, `arXiv:1904.11866 [hep-ph]`.

[45] M. Dasgupta, F. A. Dreyer, K. Hamilton, P. F. Monni, G. P. Salam, and G. Soyez, "Parton showers beyond leading logarithmic accuracy," *Phys. Rev. Lett.* **125** no. 5, (2020) 052002, `arXiv:2002.11114 [hep-ph]`.

[46] J. R. Forshaw, J. Holguin, and S. Plätzer, "Building a consistent parton shower," `arXiv:2003.06400 [hep-ph]`.

[47] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand, "Parton Fragmentation and String Dynamics," *Phys. Rept.* **97** (1983) 31–145.

[48] G. Marchesini and B. Webber, "Monte Carlo Simulation of General Hard Processes with Coherent QCD Radiation," *Nucl. Phys. B* **310** (1988) 461–526.

[49] A. Metz and A. Vossen, "Parton Fragmentation Functions," *Prog. Part. Nucl. Phys.* **91** (2016) 136–202, `arXiv:1607.02521 [hep-ex]`.

[50] D. Neill, F. Ringer, and N. Sato, "Calculating the energy loss of leading jets," in *10th International Conference on Hard and Electromagnetic Probes of High-Energy Nuclear Collisions: Hard Probes 2020* . 8, 2020. `arXiv:2008.09532 [hep-ph]`.

[51] M. Gyulassy and X.-n. Wang, "Multiple collisions and induced gluon Bremsstrahlung in QCD," *Nucl. Phys. B* **420** (1994) 583–614, `arXiv:nucl-th/9306003`.

[52] R. Baier, Y. L. Dokshitzer, A. H. Mueller, S. Peigne, and D. Schiff, "Radiative energy loss and p(T) broadening of high-energy partons in nuclei," *Nucl. Phys. B* **484** (1997) 265–282, `arXiv:hep-ph/9608322`.

[53] B. Zakharov, "Fully quantum treatment of the Landau-Pomeranchuk-Migdal effect in QED and QCD," *JETP Lett.* **63** (1996) 952–957, `arXiv:hep-ph/9607440`.

[54] M. Gyulassy, P. Levai, and I. Vitev, "Reaction operator approach to nonAbelian energy loss," *Nucl. Phys. B* **594** (2001) 371–419, `arXiv:nucl-th/0006010`.

[55] X.-N. Wang and X.-f. Guo, "Multiple parton scattering in nuclei: Parton energy loss," *Nucl. Phys. A* **696** (2001) 788–832, `arXiv:hep-ph/0102230`.

[56] P. B. Arnold, G. D. Moore, and L. G. Yaffe, "Photon and gluon emission in relativistic plasmas," *JHEP* **06** (2002) 030, `arXiv:hep-ph/0204343`.

[57] J.-w. Qiu and I. Vitev, "Coherent QCD multiple scattering in proton-nucleus collisions," *Phys. Lett. B* **632** (2006) 507–511, `arXiv:hep-ph/0405068`.

[58] H. Liu, K. Rajagopal, and U. A. Wiedemann, "Calculating the jet quenching parameter from AdS/CFT," *Phys. Rev. Lett.* **97** (2006) 182301, `arXiv:hep-ph/0605178`.

[59] N. Armesto *et al.*, "Comparison of Jet Quenching Formalisms for a Quark-Gluon Plasma 'Brick'," *Phys. Rev. C* **86** (2012) 064904, arXiv:1106.1106 [hep-ph].

[60] Y. Mehtar-Tani, J. G. Milhano, and K. Tywoniuk, "Jet physics in heavy-ion collisions," *Int. J. Mod. Phys. A* **28** (2013) 1340013, arXiv:1302.2579 [hep-ph].

[61] **JET** Collaboration, K. M. Burke *et al.*, "Extracting the jet transport coefficient from jet quenching in high-energy heavy-ion collisions," *Phys. Rev. C* **90** no. 1, (2014) 014909, arXiv:1312.5003 [nucl-th].

[62] J.-W. Qiu, F. Ringer, N. Sato, and P. Zurita, "Factorization of jet cross sections in heavy-ion collisions," *Phys. Rev. Lett.* **122** no. 25, (2019) 252301, arXiv:1903.01993 [hep-ph].

[63] J. Putschke *et al.*, "The JETSCAPE framework," arXiv:1903.07706 [nucl-th].

[64] P. Caucal, E. Iancu, and G. Soyez, "Deciphering the z_g distribution in ultrarelativistic heavy ion collisions," *JHEP* **10** (2019) 273, arXiv:1907.04866 [hep-ph].

[65] V. Vaidya and X. Yao, "Transverse Momentum Broadening of a Jet in Quark-Gluon Plasma: An Open Quantum System EFT," arXiv:2004.11403 [hep-ph].

[66] F. de Avila Belbute-Peres, K. Smith, K. Allen, J. Tenenbaum, and J. Z. Kolter, "End-to-end differentiable physics for learning and control," in *Advances in Neural Information Processing Systems*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, eds., vol. 31, pp. 7178–7189. Curran Associates, Inc., 2018. https://proceedings.neurips.cc/paper/2018/file/842424a1d0595b76ec4fa03c46e8d755-Paper.pdf.

[67] A. Koul, S. Greydanus, and A. Fern, "Learning finite state representations of recurrent policy networks," *CoRR* **abs/1811.12530** (2018) , arXiv:1811.12530. http://arxiv.org/abs/1811.12530.

[68] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Póczos, R. Salakhutdinov, and A. J. Smola, "Deep sets," *CoRR* **abs/1703.06114** (2017) , arXiv:1703.06114. http://arxiv.org/abs/1703.06114.

[69] E. Wagstaff, F. B. Fuchs, M. Engelcke, I. Posner, and M. A. Osborne, "On the limitations of representing functions on sets," *CoRR* **abs/1901.09006** (2019) , arXiv:1901.09006. http://arxiv.org/abs/1901.09006.

[70] B. Bloem-Reddy and Y. Teh, "Probabilistic symmetries and invariant neural networks," *Journal of Machine Learning Research* **21** no. 90, (2020) 1–61. http://jmlr.org/papers/v21/19-322.html.

[71] A. Accardi *et al.*, "Electron Ion Collider: The Next QCD Frontier: Understanding the glue that binds us all," *Eur. Phys. J. A* **52** no. 9, (2016) 268, arXiv:1212.1701 [nucl-ex].

[72] J. D. Bjorken, "Highly relativistic nucleus-nucleus collisions: The central rapidity region," *Phys. Rev. D* **27** (Jan, 1983) 140–151.

[73] **BRAHMS** Collaboration, I. Arsene *et al.*, "Quark gluon plasma and color glass condensate at RHIC? The Perspective from the BRAHMS experiment," *Nucl. Phys. A* **757** (2005) 1–27, arXiv:nucl-ex/0410020.

[74] **PHENIX** Collaboration, K. Adcox *et al.*, "Formation of dense partonic matter in relativistic nucleus-nucleus collisions at RHIC: Experimental evaluation by the PHENIX collaboration," *Nucl. Phys. A* **757** (2005) 184–283, arXiv:nucl-ex/0410003.

[75] **PHOBOS** Collaboration, B. Back *et al.*, "The PHOBOS perspective on discoveries at RHIC," *Nucl. Phys. A* **757** (2005) 28–101, arXiv:nucl-ex/0410022.

[76] **STAR** Collaboration, J. Adams *et al.*, "Experimental and theoretical challenges in the search for the quark gluon plasma: The STAR Collaboration's critical assessment of the evidence from RHIC collisions," *Nucl. Phys. A* **757** (2005) 102–183, arXiv:nucl-ex/0501009.

[77] B. V. Jacak and B. Muller, "The exploration of hot nuclear matter," *Science* **337** (2012) 310–314.

[78] B. Müller, J. Schukraft, and B. Wysłouch, "First Results from Pb+Pb Collisions at the LHC," *Annu. Rev. Nucl. Part. S.* **62** no. 1, (2012) 361–386.

[79] P. Braun-Munzinger, V. Koch, T. Schäfer, and J. Stachel, "Properties of hot and dense matter from relativistic heavy ion collisions," *Phys. Rept.* **621** (2016) 76–126.

[80] W. Busza, K. Rajagopal, and W. van der Schee, "Heavy Ion Collisions: The Big Picture, and the Big Questions," *Ann. Rev. Nucl. Part. Sci.* **68** (2018) 339–376.

[81] **PHENIX** Collaboration, A. Adare *et al.*, "Heavy Quark Production in $p + p$ and Energy Loss and Flow of Heavy Quarks in Au+Au Collisions at $\sqrt{s_{NN}} = 200$ GeV," *Phys. Rev. C* **84** (2011) 044905, arXiv:1005.1627 [nucl-ex].

[82] **CMS** Collaboration, A. M. Sirunyan *et al.*, "Measurement of prompt and nonprompt charmonium suppression in PbPb collisions at 5.02 TeV," *Eur. Phys. J. C* **78** no. 6, (2018) 509, arXiv:1712.08959 [nucl-ex].

[83] **STAR** Collaboration, L. Adamczyk *et al.*, "Measurements of jet quenching with semi-inclusive hadron+jet distributions in Au+Au collisions at $\sqrt{s_{NN}}$ = 200 GeV," *Phys. Rev. C* **96** no. 2, (2017) 024905, arXiv:1702.01108 [nucl-ex].

[84] **ALICE** Collaboration, S. Acharya *et al.*, "Measurements of inclusive jet spectra in pp and central Pb-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV," *Phys. Rev. C* **101** no. 3, (2020) 034911, arXiv:1909.09718 [nucl-ex].

[85] **ATLAS** Collaboration, M. Aaboud *et al.*, "Measurement of the nuclear modification factor for inclusive jets in Pb+Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV with the ATLAS detector," *Phys. Lett. B* **790** (2019) 108–128, arXiv:1805.05635 [nucl-ex].

[86] M. Dasgupta, F. Dreyer, G. P. Salam, and G. Soyez, "Small-radius jets to all orders in QCD," *JHEP* **04** (2015) 039, arXiv:1411.5182 [hep-ph].

[87] M. J. Dolan and A. Ore, "Equivariant Energy Flow Networks for Jet Tagging," arXiv:2012.00964 [hep-ph].

[88] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (elus)," *CoRR* **abs/1511.07289** (2016) , arXiv:1511.07289.

[89] D. R. Cox, "The regression analysis of binary sequences," *Journal of the Royal Statistical Society: Series B (Methodological)* **20** no. 2, (1958) 215–232.

[90] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *CoRR* **abs/1412.6980** (2015) , arXiv:1412.6980.

Supplemental material

We first discuss the splitting kinematics of the DGLAP parton branching process. In particular, we focus on setting up the full event kinematics in spherical coordinates. We then present more details of the GAN setup.

The angles of the daughter partons relative to the parent direction

The shower is designed to conserve the momentum in the plane orthogonal to the direction of the parent parton that splits, and also conserve the energy or the light-cone momentum components parallel to the parent direction, the two being equivalent up to power-corrections in the small splitting angle limit. Formally this does not conserve the total global transverse momentum relative to the initiating parton of the cascade, on the order of $5 \sim 10\%$ of the total energy of the jet, and necessarily builds up a total non-zero invariant mass of the final state, but does preserve the angular structure of the shower, and the distribution of energy implied by the DGLAP evolution equations. More sophisticated momentum conservation schemes exist, preserving more of the structure of the distribution of partons in phase-space. This is necessary for the resummation of logarithms beyond leading logarithmic order, but such complications are unnecessary for our proof-of-concept.

We consider the DGLAP $1 \to 2$ parton splitting as illustrated in Fig. 1. The splitting process is characterized in terms of the longitudinal momentum fractions z and $1 - z$ of the two daughter partons, their relative opening angle θ and their orientation in azimuth ϕ. In order to determine the spherical coordinates of the two daughter partons, we start by calculating their angle with respect to the parent direction, which we denote by $\theta_{1,2p}$. The angles $\theta_{1,2p}$ are illustrated in Fig. 5, and we have $\theta = \theta_{1p} + \theta_{2p}$. The two angles can be determined from the relative splitting angle θ which is related to the Monte Carlo time and the momentum fraction z. We consider the splitting of a parent parton with momentum l^μ (in the $-z$ direction) to two daughter partons with momentum q^μ and $l^\mu - q^\mu$. Both partons after the splitting are on-shell $q^2 = (l - q)^2 = 0$. Using light cone coordinates, we have

$$|\vec{q}| = q^0 = \frac{1}{2}(q^- + q^+) = \frac{1}{2}(zl^- + (1-z)l^+) \approx \frac{1}{2}zl^- . \tag{3}$$

where we used

$$q^+ = \frac{l^+}{l^-}(l^- - q^-) , \tag{4}$$

which follows from $(l - q)^2 = 0$. In addition, we have $l^2 = l^+ l^-$ and $q^- = zl^-$. The approximation in Eq. (3) holds for $l^+ \ll l^-$. Similarly, we find

$$|\vec{l} - \vec{q}| = l^0 - q^0 = \frac{1}{2}(l^+ + l^- - (q^+ + q^-)) = \frac{1}{2}((1-z)l^- + zl^+) \approx \frac{1}{2}(1-z)l^- . \tag{5}$$

In order to write the angle θ_{1p} of the daughter parton with momentum q^μ in terms of the splitting angle θ and the momentum fraction z, we consider

$$\cos\theta_{1p} = \frac{\vec{q} \cdot \vec{l}}{|\vec{q}||\vec{l}|} . \tag{6}$$

We rewrite the expression in terms of the momenta q^μ and $l^\mu - q^\mu$ as

$$
\begin{aligned}
\cos\theta_{1p} &= \frac{\vec{q} \cdot (\vec{l} - \vec{q}) + |\vec{q}|^2}{|\vec{q}||(\vec{l} - \vec{q}) + \vec{q}|} \\
&= \frac{|\vec{q}||\vec{l} - \vec{q}|\cos\theta + |\vec{q}|^2}{|\vec{q}|\sqrt{|\vec{l} - \vec{q}|^2 + |\vec{q}|^2 + 2|\vec{l} - \vec{q}||\vec{q}|\cos\theta}} \\
&= \frac{(1 - z)\cos\theta_{1p} + z}{\sqrt{(1 - z)^2 + z^2 + 2z(1 - z)\cos\theta}} .
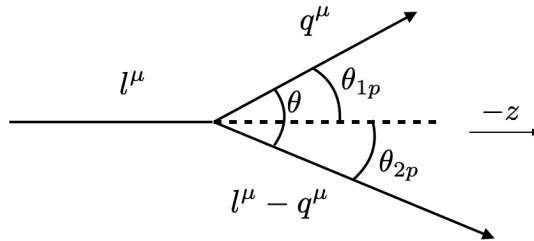\end{aligned}
\tag{7}
$$

FIG. 5. Illustration of the DGLAP $1 \to 2$ parton branching process with the relative opening angle θ of the two daughter partons and their angles relative to the parent direction $\theta_{1,2p}$.

The last line is obtained by inserting the expressions for $|\vec{q}|$ and $|\vec{l} - \vec{q}|$ which were obtained above. We thus find the following expression for the angle between the parent direction and the daughter parton with momentum q^μ:

$$\theta_{1p} = \arccos\left(\frac{z + (1-z)\cos\theta}{\sqrt{1 - 2z(1-z)(1-\cos\theta)}}\right). \tag{8}$$

Then the angle of the other daughter parton is given as $\theta_{2p} = \theta - \theta_{1p}$.

The direction of the two daughter partons in absolute spherical coordinates

Given the direction of the parent parton in absolute spherical coordinates $(\tilde{\Theta}_p, \tilde{\Phi}_p)$ and the kinematics of the $1 \to 2$ splitting (the azimuthal direction and the angles θ_{ip} derived above), we can now determine the spherical coordinates of the two daughter partons $(\tilde{\Theta}_{di}, \tilde{\Phi}_{di})$, $i = 1, 2$. We start with the vector pointing in the direction of the parent parton. In spherical coordinates, we have

$$\vec{r}_p = \begin{pmatrix} \sin\tilde{\Theta}_p \cos\tilde{\Phi}_p \\ \sin\tilde{\Theta}_p \sin\tilde{\Phi}_p \\ \cos\tilde{\Theta}_p \end{pmatrix}. \tag{9}$$

In order to generate the random distribution in azimuth, we construct a random vector $\vec{r}_r$ which is then orthonormalized to get a basis vector in the plane transverse to the parent direction. We use flat sampling for each component $\vec{r}_r^i$ in the range of $[1, -1]$. The normalized random vector transverse to the parent direction can then be written as

$$\vec{r}_A = \frac{1}{N}((\vec{r}_p \cdot \vec{r}_r)\vec{r}_p - \vec{r}_r), \tag{10}$$

where the normalization factor N is given by

$$N = \left(\sum_i \left((\vec{r}_p \cdot \vec{r}_r)^2 \vec{r}_p^i - \vec{r}_r^i\right)\right)^{1/2}. \tag{11}$$

By construction, we thus have

$$\vec{r}_A \cdot \vec{r}_p = 0, \quad \vec{r}_A^2 = 1. \tag{12}$$

We can then construct a second basis vector $\vec{r}_B$ by calculating the cross product

$$\vec{r}_B = \vec{r}_A \times \vec{r}_p, \tag{13}$$

which is normalized and orthogonal to both $\vec{r}_A$ and $\vec{r}_p$. We write the vectors $\vec{r}_{di}$ of the two daughter partons $i = 1, 2$ as a sum of two vectors. The first term is the projection of the daughter's direction onto the direction of the parent parton which is proportional $\sim \cos\theta_{1,2p}$. The second vector is in the transverse plane relative to the parent direction and parametrized in terms of $\vec{r}_{A,B}$ and a random variable ϕ chosen in the range of $[0, 2\pi]$ (flat sampling). The
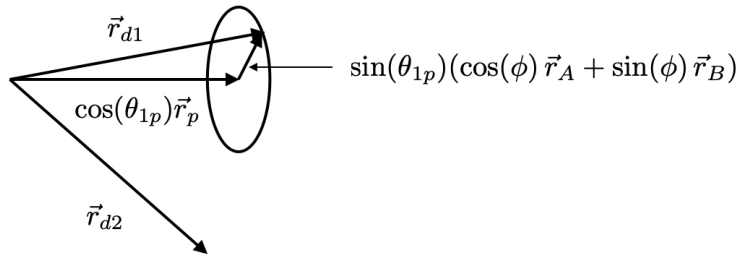
FIG. 6. Illustration of the vectors and angles relevant to determine the direction of the two daughter partons in absolute spherical coordinates.

magnitude of that second vector is given by $\sin\theta_{1,2p}$. For the two daughters $i = 1, 2$, the resulting vector can be written as

$$\vec{r}_{di} = \cos(\theta_{ip})\,\vec{r}_p \pm \sin(\theta_{ip})(\cos(\phi)\,\vec{r}_A + \sin(\phi)\,\vec{r}_B)\,. \tag{14}$$

See Fig. 6 for an illustration of the vectors and angles relevant for setting up the full splitting kinematics of the two daughter partons. We can then write the polar and azimuthal angle of the two daughter partons as

$$
\begin{aligned}
\tilde{\Theta}_{di} &= \arccos(r_{di}^z)\,, \\
\tilde{\Phi}_{di} &= \pi + \arctan\left(\frac{r_{di}^y}{r_{di}^x}\right)\,.
\end{aligned}
\tag{15}
$$

More details of the GAN setup

For practical purposes we split the generator network into a time dependent and a time independent part. Both parts consist of five hidden layers with 50 neurons and an exponential linear unit (ELU) [88] activation to avoid discontinuous steps in the generated distributions. The time dependent network generates the next splitting angle θ'_i taking as input the previous angle θ'_{i-1}, the initial scale Q, and a uniformly distributed $[0, 1)$ random number. The time independent network generates the variables z'_i, ϕ taking as input uniformly distributed $[0, 1)$ random numbers. We also take the momentum of the parent parton as input to the time independent network. Through the training process, the GAN learns that this information is not necessary to generate the variables z', ϕ. To avoid vanishing gradients, the immediate output neuron of the time independent neural network generates a transformed $z'_i = -\log(1/z_i - 1)$, which is then converted to z_i that is bounded by $(0, 1)$. Similarly, the θ'_i from the output neuron of the time dependent network is converted to θ_i which is bounded by $(0, \pi/2)$.

We propagate an event record of current partons, the initial scale Q, and the current θ_i throughout the shower process. A random parton (pure gluon shower) is selected for the splitting process by double indexing: We first sort the current list of showered partons in descending values of Z_i, note its indexing order, and count the number N of partons when their momenta are above the cutoff ϵ and are therefore able to split. A random parton is then chosen from the first N partons, and, using the order of the sorted index, it is mapped to the event record. Since the processing is *de facto* executed in parallel, we calculate the splitting of a parton even if $N = 0$ which is then reversed afterwards. The highest number of branching processes occur for $Q = 800$ GeV. In this case, our implementation on Nvidia Titan RTX reaches an execution time of $95 \pm 4\,\mu$s/event.

The discriminator network consists of a sequence of two deep sets networks [68–70]. The first deep sets network takes the list of partons from the shower as input, and produces the per-event activation. The second deep sets network uses the output of the first network, the per-event activation, as input and produces the statistical activation for the entire batch. We augment observables derived from the deep sets with the 2nd to 5th moment of the whole batch parton momenta, in order to have a fall-back in the first training epochs, until the deep sets are fully trained. The deep sets and moments are combined by a shallow network with one layer of 20 hidden neurons.

We employ a modified training process compared to the original GAN which we summarize here. We use the binary cross entropy as the loss function [89] which is given by

$$L = -\frac{1}{2}E_x[\log D(x)] - \frac{1}{2}E_c[\log(1 - D(G(c)))] \,, \tag{16}$$

where E is the expectation value, D the discriminator, and G the generator. The conditional vector c contains both the initial parton and a sufficient amount of random numbers for the full shower. The ADAM optimizer [90] is used for both the discriminator and generator, where the exponential decay rate for the first and second moment are chosen as $\beta_1 = 0.5$ and $\beta_2 = 0.999$. The learning rate is $\lambda = 5 \times 10^{-4}$ for the discriminator, and $\lambda = 5 \times 10^{-6}$ for the generator. To guard against generator training steps that may inadvertently deteriorate the generator, the discriminator D is trained in each epoch until $D(x) > 0.5$ for parton shower result x, and $D(G) < D(x)$ where G are the partons generated by the GAN. After each generator training step, its finite step size result is tested and reverted in case it resulted in reduced $D(G)$ scores.

The time-dependent and independent networks of the generator are first pre-trained to be in the vicinity of the physical value. We observe that at the beginning of the training, the untrained discriminator allows the generator to deviate further from the pre-trained values. After ~ 500 epochs the discriminator is sufficiently trained to correct the generator, and closure with the parton shower occurs after ~ 700 epochs. This mostly concerns the θ variable whereas z is more robust. We note that this is not a general limitation but allows the training to proceed by a local minimization. An alternative that we plan to explore in the future is a global optimization in combination with a random initialization.