Explainable Machine Learning of parton showers

Yue Shi Lai

LBNL NSD

NSD Staff Meeting, Nov. 17 2020

The project

- Project with YSL, Mateusz Ploskon, Felix Ringer
- Physics goals:
  - Probe parts of theory calculation inaccessible by normal experimental means
  - Automated analysis of LHC (possibly RHIC) heavy-ion
  - Testing next-generation parton showers (e.g. non-pertubative kernels)
- CS/ML goals:
  - GAN with a human understandable generator
  - Physics-“inspired”/constrained GAN

Learning the underlying physics of high-energy collisions – an explainable AI approach

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: November 17, 2020)

We propose an explainable/white box AI approach to study the underlying physics of high-energy collisions using full event information. As a proof of concept, we present an approach based on a Generative Adversarial Network (GAN) which learns the underlying physics from the final output of a DGLAP-based parton shower. We train the network directly on the four-vectors of the final set of particles that are produced by the shower. This is achieved through so-called deep sets which leads to a permutation invariant data representation. We demonstrate for the first time that the network is not only able to learn the final distribution of particles, but it also learns the underlying parton branching mechanism, i.e. the Altarelli-Parisi splitting function, the ordering variable of the shower, and the scaling behavior. While the proof of concept reported in this work is focused on perturbative physics of the parton shower, we foresee various applications of our framework to physics in high-energy collisions which is 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 [1–26]. For example, in Refs. [1–5] jet tagging techniques were developed which often outperform traditional techniques. In Refs. [6–11] Generative Adversarial Networks (GANs) [27, 28], 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. [29] a probabilistic model was introduced based on jet clustering and in Ref. [30] a convolutional autoencoder within a shower was used which qualitatively reproduces jet observables. See also Refs. [31–33] 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. In particular, general purpose parton showers play an important role in our understanding of high energy collider experiments [34–36]. Starting with highly energetic quarks or gluons which are produced in hard-scattering events, parton showers take into account the parton branching processes that occur during the evolution from the hard scale to the infrared followed by the hadronization. While the general concept of parton showers is well established, important questions about the perturbative accuracy [37–43], nonperturbative effects [44–47] and the modification in a nuclear environment [48–60], remain a challenge.

In this work, we propose an explainable/white box AI [61, 62] approach to learn the underlying physics of high energy particle collisions. As a proof of concept, we present results of a GAN trained on the output of a parton shower which not only reproduces the final distribution of particles but also learns the underlying showering mechanism using full event information. GANs consist of two competing neural networks, the generator and discriminator. Due to the design of our generator network, we are not only able to 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, besides the final distribution of particles, demonstrate that the network can learn the Altarelli-Parisi splitting function $P_{i\to jk}(z)$, splitting angles of individual branching processes and the dependence of the shower on the energy scale Q, see Fig. 1. This 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 such that we can also incorporate nonperturbative physics in the future. In order to use full event information, we use a data representation which is directly given by the four-vectors of the final state particles. In order to avoid that the network picks up an unphysical ordering of the list of four-vectors during the training process, we instead use sets to represent the data. In our work, the required per-

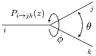
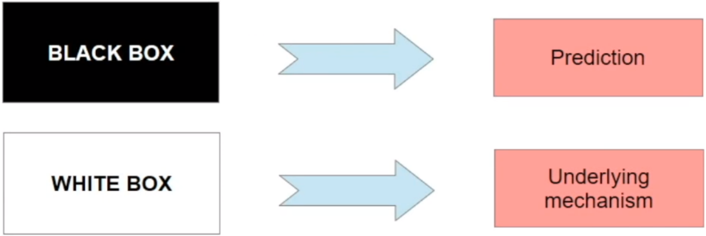


FIG. 1. Parton splitting process $i \to jk$ with momentum fraction z, relative splitting angle of the two daughter partons θ and azimuthal angle ϕ.

White box machine learning

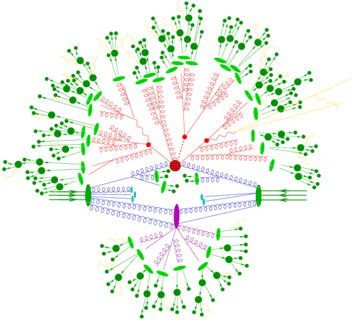


https://www.youtube.com/watch?v=6wjTV4Odraw

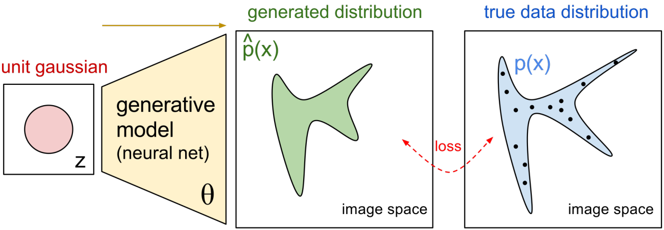
- Prediction only useful for analysis cuts, etc.
- For us physicists, ultimate goal is to understand the mechanism

Why parton shower?

- This talk is narrowly about ML for parton shower
- Expect parton shower to be most relevant for energy loss in QGP and cold nuclear matter
- Next generation parton shower have better accuracy and interface to nonperturbative physics
    - Fragmentation function at NLL accuracy, D. Neill, F. Ringer, N. Sato arXiv:2008.09532, D. Neill, arXiv:2010.02934
    - Evolution of TMDPDF, M. Ebert, I. Stewart, Y. Zhao, PRD 99, 034505 (2019)

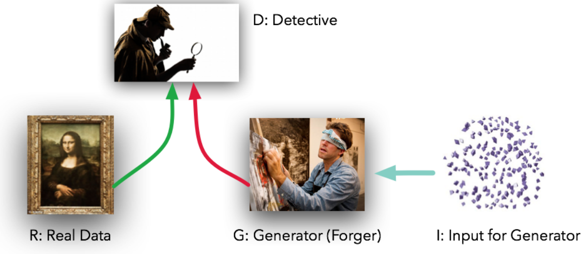


Unsupervised learning



- Minimize an overarching “loss” between generate data to real data
- Ensures a model is trainable using only data
  - ⇒ Once fragmented ⇒ automated data analysis for pp and PbPb
- So far most work with unsupervised learning in NP/HEP for simulation

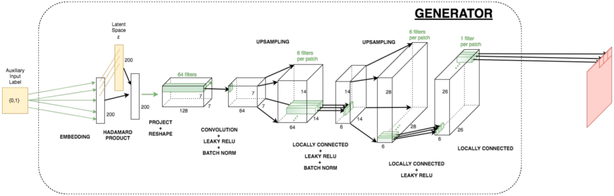
Generative Adversarial Networks

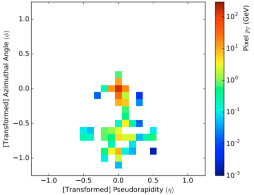


https://medium.com/@devnag/generative-adversarial-networks-gans-in-50-lines-of-code-pytorch-e81b79659e3f

- Two neural networks: the generator (“forger”) and discriminator (“detective”)
- Simultaneously optimize both, causing both to be in competition with each other (Nash equilibrium)
- Why?
  - The generator can non-deterministically produce splitting as it sees fit
  - Only require that no analysis exists that can (easily) distinguish the generator output from “reality” (the discriminator mostly fails)

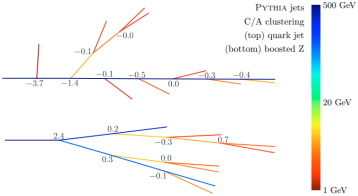
What we don't want to do

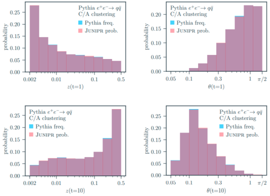




- LA-GAN, L. de Oliveira, M. Paganini, B. Nachman (Comput. Softw. Big Sci. 1, 4, 2017; arXiv:1701.05927)
- Black box of convolutional neural network
- "Jet images" (not individual partons/particles)

What we don't want to do





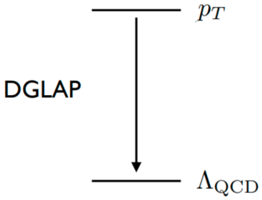
- JUNIPR, A. Andreassen, I. Feige, C. Frye, M. Schwartz EPJC 79 102, 2019; arXiv:1804.09720)
- Impose a jet clustering hierachy instead of parton shower
- No access to evolution equation vs. what was put in by hand (jet clustering)
- Still a black box

Our contributions

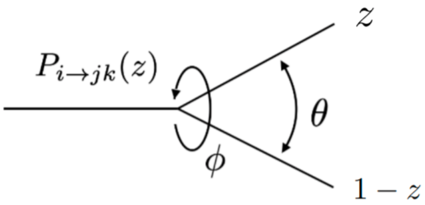
- Dynamic NN simulating a random splitting history, and still efficiently executed in parallel on a GPU
  - Full shower $Q = 800$ GeV: tree depth ≈ 90, execution time $\approx 95 \pm 4$ µs/full shower
  - 20k–30k showers tree running in parallel and all different
- GAN training of a non-black-box
- First non-black-box ML parton shower
  - First NN shower where the internal/per-step splitting z and θ are plotted

Shower implementation

- Time-ordered $1 \to 2$ shower by Duff Neill (LANL)
- Main idea is to shower from p_T down to Λ_{QCD}



- $1 \to 2$ means a split given by (z, θ)



Shower implementation

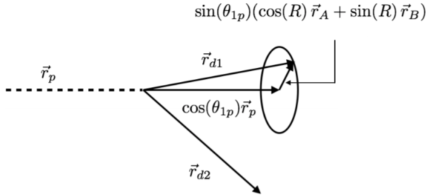
- For "manual" calculation, probability of Δt per split is

$$p(\Delta t) = \exp\left(-\Delta t \sum_{i \in \text{flavor}} \int_{\epsilon}^{1-\epsilon} dz\, P_i(z)\right) \tag{1}$$

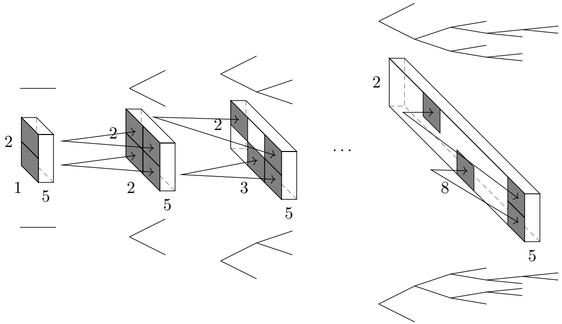
- Time vs. θ is

$$t(Q, \theta) = \int_{Q \tan(\pi/2)}^{Q \tan(\theta/2)} \frac{dt'}{t'} \frac{\alpha_S(t')}{\pi} \tag{2}$$

- z is sampled from DGLAP $P_{i \to jk}(z)$ (currently i, j, k are gluons)
- Manually calculate unique solution of θ_{1p}, $\theta_{2p} = \theta - \theta_{1p}$ that satisfies (θ, z)

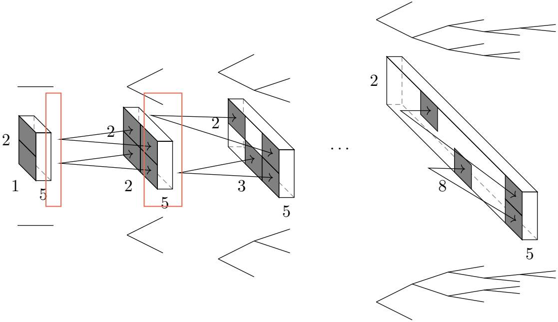


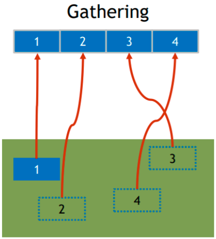
Implementation as NN



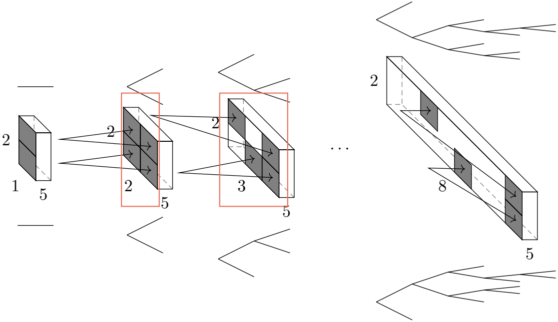
- Basic structure is similar to an actual theory model
  - Operating on $p^+ = \frac{1}{\sqrt{2}}(p^0 + p^z)$ and unit 3-momentum vector
  - Recursively splits by the same neural network kernel
  - Batched random splitting by scatter-gather
  - All splits are speculative, undo the split if it turns out $\theta < \theta_{\text{final}}$ (or $t > t_{\text{final}}$)

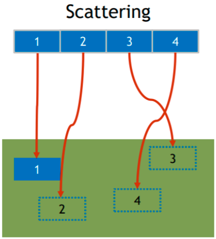
Implementation as NN



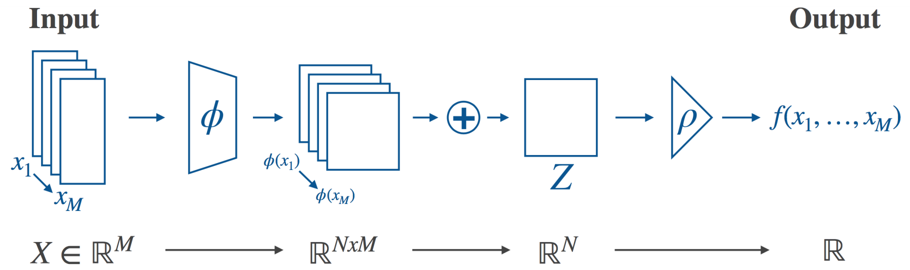


Implementation as NN





Discriminator



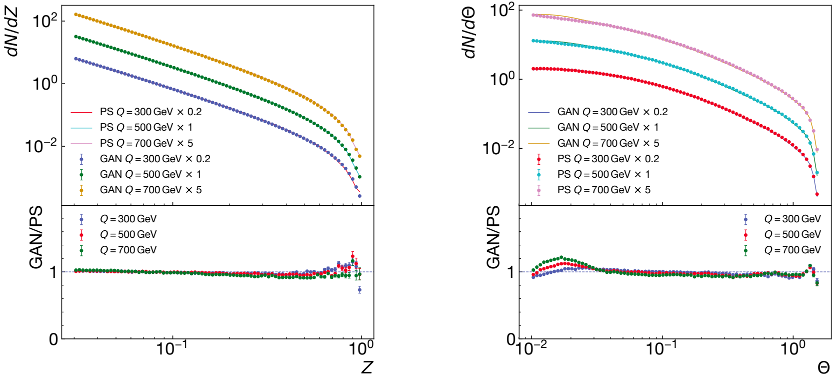
https://www.inference.vc/content/images/2019/02/Architecture.png

- Deep Sets, M. Zaheer et al., arXiv:1703.06114
- General form of permutation-invariant function
- Deep Sets is run twice:
  - $\approx$ 100 showered partons $\Rightarrow$ per-event observables
  - $\approx$ 20000 events $\Rightarrow$ ensemble observables

More GAN details

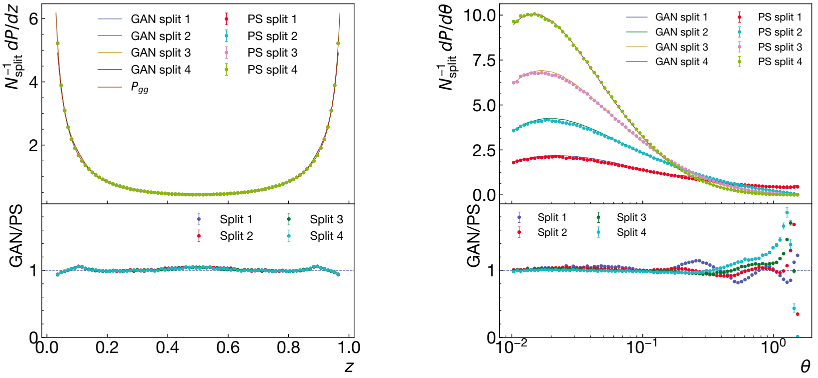
- "Conditional GAN": original parton + noise as input (M. Mirza, S. Osindero, "Conditional generative adversarial nets", arXiv:1411.1784 [cs.LG])
- Initialization: generator is pre-trained to have reasonable z and θ distribution
- Modified "vanilla"/DCGAN training:
  - Asynchronous generator/discriminator updates (update generator only if confident in the discriminator)
  - Test that an optimization step really improved

Final Z, Θ spectrum



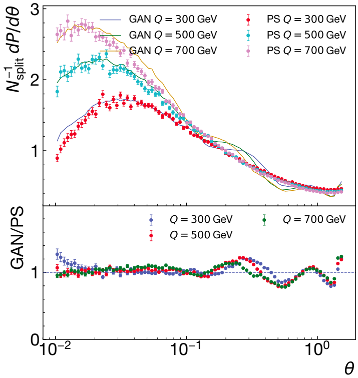
- Final distribution after 400 training epochs, and check back at the point of best fit
- Final distribution agrees exceedingly well with PS
- About 2–3 orders of magnitude agreement

z, θ spectrum



- z follows the DGLAP shape within $z_{\mathrm{cutoff}} = 0.03$
- θ for the first 4 steps also follow parton shower
- At large θ the function becomes too stiff for the size of the neural network we are using

Q dependent θ spectrum



- Even Q dependent behavior is modeled correctly in the neural network

Summary

- A first feasibility demonstration of automated deduction of parton shower by unsupervised, white-box machine learning
- Many extensions
  - Automated data analysis by adding fragmentation
  - Non-perturbative physics
- First proof-of-concept paper is being prepared