Explainable machine learning of parton shower mechanisms

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

[a] Lawrence Berkeley National Laboratory, Nuclear Science Division

[b] Los Alamos National Laboratory, Theoretical Division

[c] Stony Brook University, C. N. Yang Institute for Theoretical Physics

ICHEP 2022

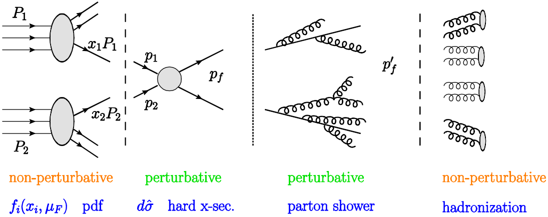
YSL, D. Neill, M. Płoskoń, F. Ringer, Phys. Lett. B 829, 137055 (2022) [arXiv:2012.06582 [hep-ph]]

Introduction

- Parton shower (and fragmentation function) most challenging uncertgainties in QCD modeling and measurement
- Parton showers and shower-like evolution are being pushed towards perturbative accuracy

  (e.g. D. Neill, F. Ringer, N. Sato arXiv:2008.09532, D. Neill, arXiv:2010.02934; M. Ebert, I. Stewart, Y. Zhao, PRD 99, 034505 (2019))

- Today's challenges for parton shower:
  - Perturbative accuracy
  - How to interface with non-perturbative effects
  - Modification in (hot/cold) nuclear environment



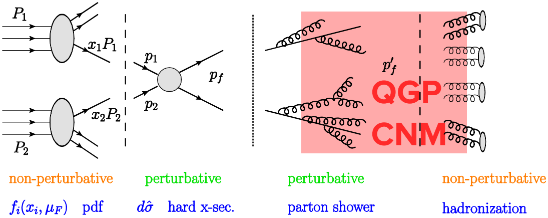
Adapted from O. Biebel, Phys. Rep. 3, 165 (2001)

- Machine learning parton showers on experimental data can potentially yield insights difficult to arrive at from first principles

Introduction

- Parton shower (and fragmentation function) most challenging uncertgainties in QCD modeling and measurement
- Parton showers and shower-like evolution are being pushed towards perturbative accuracy
  (e.g. D. Neill, F. Ringer, N. Sato arXiv:2008.09532, D. Neill, arXiv:2010.02934; M. Ebert, I. Stewart, Y. Zhao, PRD 99, 034505 (2019))

- Today's challenges for parton shower:
  - Perturbative accuracy
  - How to interface with non-perturbative effects
  - Modification in (hot/cold) nuclear environment



Adapted from O. Biebel, Phys. Rep. 3, 165 (2001)

- Machine learning parton showers on experimental data can potentially yield insights difficult to arrive at from first principles

White box machine learning

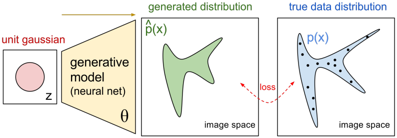
$\Rightarrow$ **Prediction/Generation**

$\Rightarrow$ Prediction/Generation
\+ **Underlying mechanism**

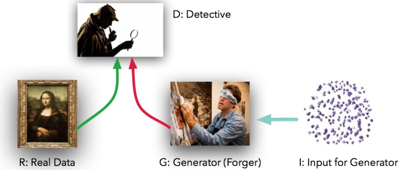
- For us physicists, ultimate goal is to understand the mechanism
- "White box" because we aim for **algorithmic transparency** — not a post-hoc explaination of the individual ("local") decision
- Algorithmic transparency cannot be easily added, unlike post-hoc methods

Generative Adversarial Networks

- Emulate the natural distribution by minimizing an overarching "loss" between generate data to real data
- Ensures a model is trainable using only data



- Two neural networks: the generator ("forger") and discriminator ("detective")
- Simultaneously optimize both, causing both to be in competition with each other (Nash equilibrium)

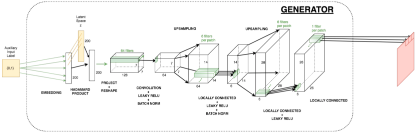


https://medium.com/@devnag/generative-adversarial-networks-gans-in-50-lines-of-code-p

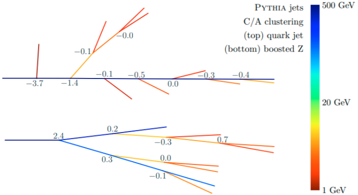
- Why?
  - The generator can non-deterministically produce splitting as it sees fit
  - Only require that no analysis exists that can (easily) distinguish the generator output from "reality" (the discriminator mostly fails)

Past Results

- LA-GAN, L. de Oliveira, M. Paganini, B. Nachman (Comput. Softw. Big Sci. 1, 4, 2017; arXiv:1701.05927)
  - Convolutional neural network – black box
  - “Jet images” (not individual partons/particles)



- JUNIPR, A. Andreassen, I. Feige, C. Frye, M. Schwartz EPJC 79 102, 2019; arXiv:1804.09720)
  - Not actual showers, but jet clustering hierachy



- **All past results are black boxes ⇒ no possibility to extract knowledge**
- **Lack of physics-motivated NN ⇒ no access to the parton evolution**

Our Contributions

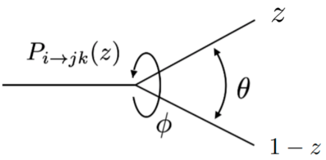
1. First NN parton shower with physics-motivated architecture
   - Fully randomized splitting of individual partons
   - Recurrents splitting using the same underlying kernel
   - Efficient parallel execution on a GPU
     - Mean number of splittings: ≈ 90
     - 20k–30k showers tree running in parallel (per 24 GB RAM)
     - Execution time $\approx 95 \pm 4$ μs/full shower
2. First non-black-box ML parton shower
   - First NN shower where the internal/per-step splitting z and θ can be plotted

$\Rightarrow$ Possibly the first non-black-box GAN?

Shower Implementation

- Time-ordered $1 \to 2$ DGLAP shower
- Shower from initial p_T down to Λ_{QCD}
- Probability of advancing time Δt per split:

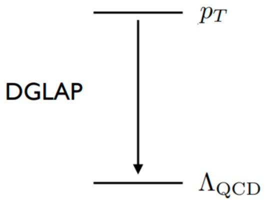
$$p(\Delta t) = \exp\left(-\Delta t \sum_{i \in \text{flavor}} \int_{\epsilon}^{1-\epsilon} dz\, P_i(z)\right) \quad (1)$$



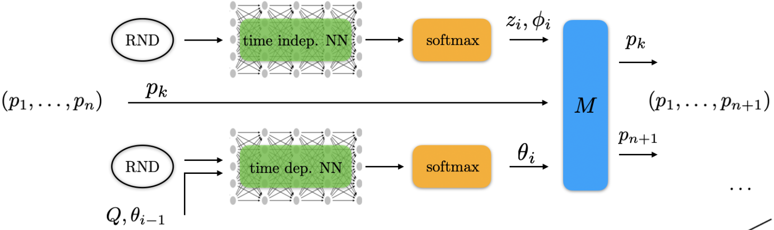
- z is sampled from DGLAP $P_{i\to jk}(z)$ (currently i, j, k are gluons)

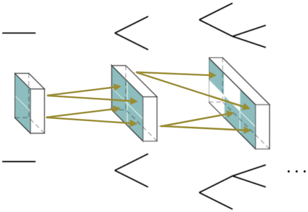
- QCD evolution of time vs. θ:

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

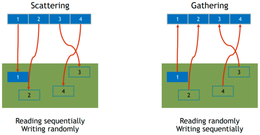


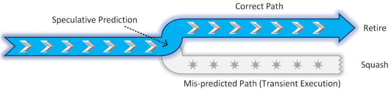
Implementation as NN (Generator)





- 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 in parallel by scatter-gather
- Termination by speculative execution



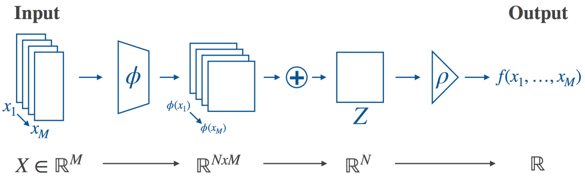


(S. Constable, IEEE SecDev 2020)

(S. Jones, GTC 2017)

Implementation as NN (Discriminator, Training)

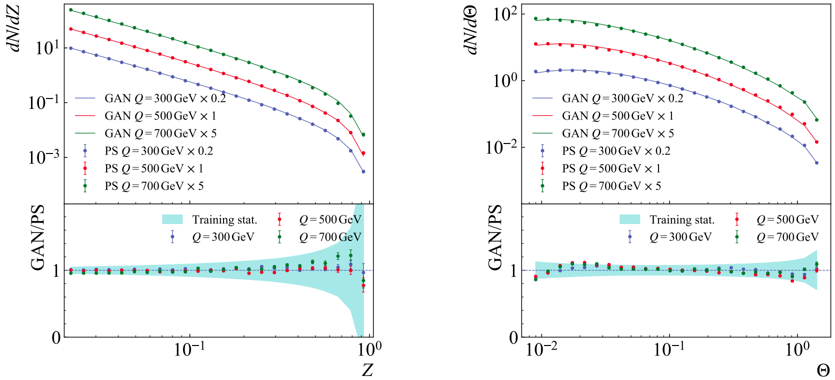
- Deep Sets, M. Zaheer et al., arXiv:1703.06114
- General form of permutation-invariant function
- **No access to the intermediate shower**



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

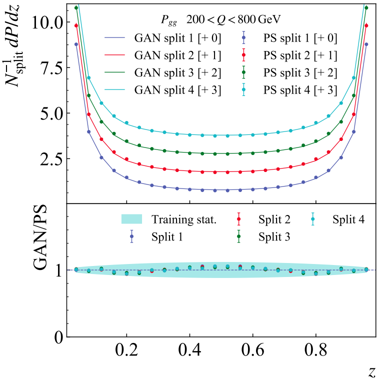
- “Conditional GAN”: original parton + noise as input (M. Mirza, S. Osindero, “Conditional generative adversarial nets”, arXiv:1411.1784 [cs.LG])
- Initialization: generator is pre-trained to have reasonable z and θ distribution
- Modified “vanilla”/DCGAN training:
  - Asynchronous generator/discriminator updates (update generator only if confident in the discriminator)
  - Test that an optimization step really improved
- Result shown with $200 < Q < 800$ GeV, $\epsilon = 0.02$ and after 500 epochs

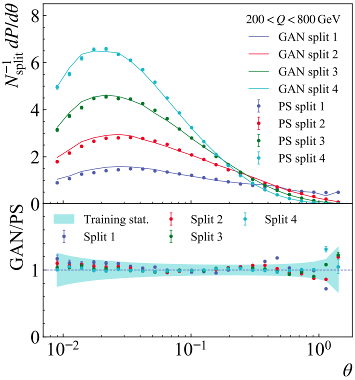
Final Z, Θ spectrum



- PS: Original DGLAP parton shower
- Final distribution agrees exceedingly well with PS
- About 2–3 orders of magnitude agreement

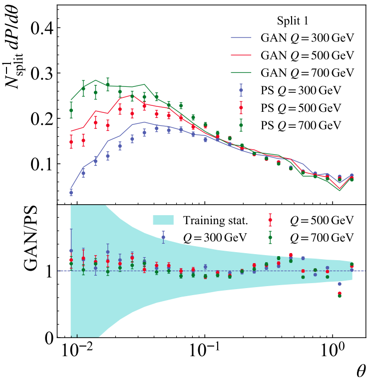
(Internal) z, θ spectrum

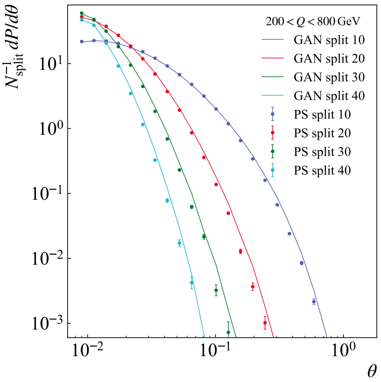




- z follows the DGLAP shape within $z_{\mathrm{cutoff}} = 0.03$
- θ for the first 4 steps also follow parton shower
- At large θ the function becomes too stiff for the size of the neural network we are using

(Internal) Q dependent θ spectrum





- Even Q dependent behavior is modeled correctly in the neural network
- Large splits are also reasonable

Summary

- A first feasibility demonstration that white-box (x)AI can successfully learn the underlying physics of the parton shower
- Constrained GAN is capable of learning the DGLAP parton shower without seeing the individual splittings
- Work is starting point to eventually train NN directly on experimental data, extracting physics from full event information
- Future direction:
  - Inclusion of fragmentation into hadrons
  - Study of collective effects, hot medium effects in heavy-ion and cold nuclear effects at electron–ion collider