(FAT-GAN) architecture for simulation of electron-proton scattering events

[1]Yasir Alanazi, [2]N Sato, [2]Tianbo Liu, [2]W Melnitchouk, [3]Michelle P Kuchera, [3]Evan Pritchard, [3]Michael Robertson, [3]Ryan Strauss, [4]Luisa Velasco, [1]Yaohang Li

[1]Department of Computer Science, Old Dominion University, Norfolk, Virginia 23529

[2]Theory Center, Jefferson Lab, Newport News, Virginia 23606

[3]Department of Physics, Davidson College, Davidson, North Carolina 28035

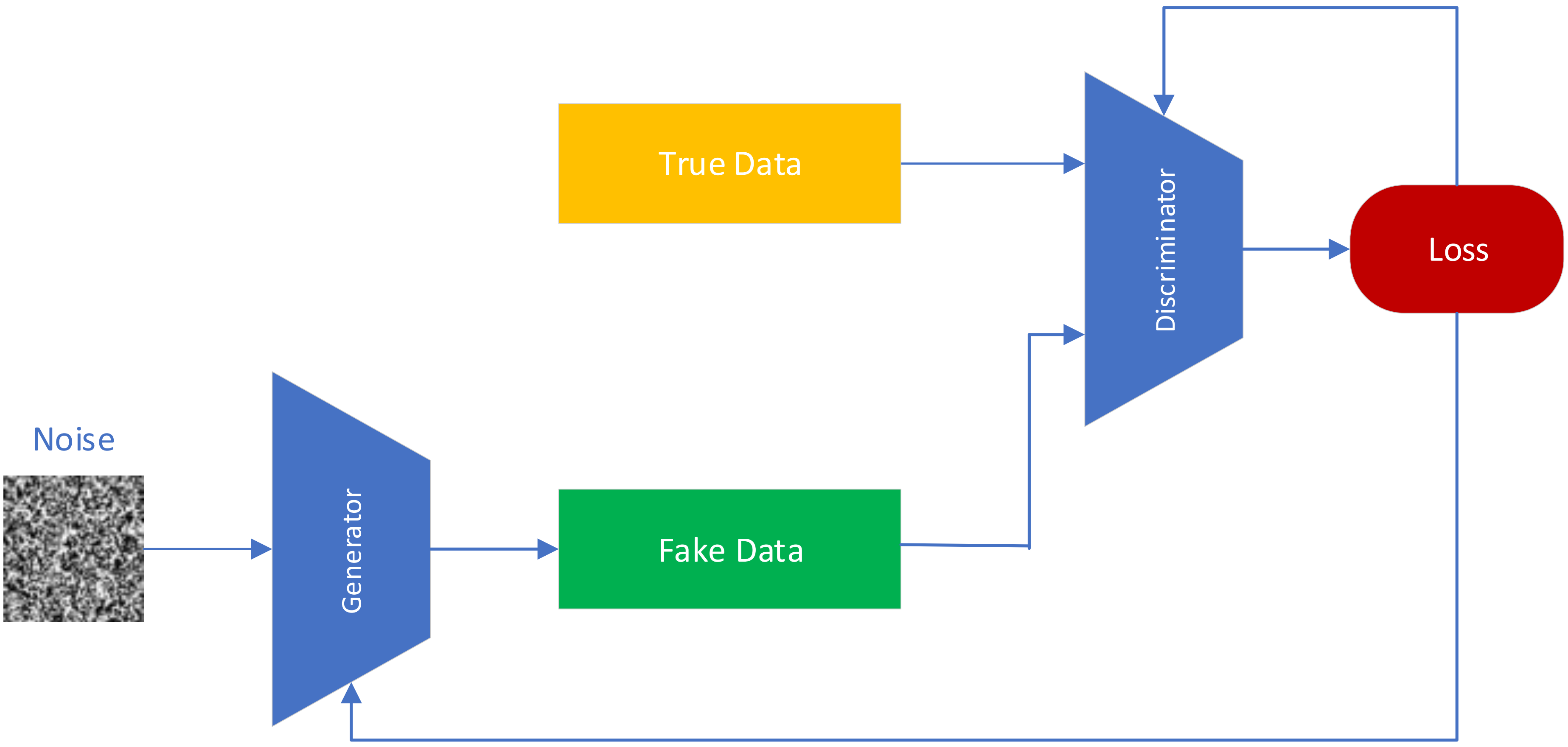
[4]Department of Mathematics and Computer Science, Davidson College, Davidson, North Carolina 28035 5Department of Physics, University of Dallas, Irving, Texas 75062

Generative Adversarial Network (GAN)

- **Generative Adversarial Networks (GAN)**
  - Introduced by Ian Goodfellow et al. in 2014
  - Deep neural network architectures comprised of two nets
    - A Generator
    - A Discriminator
  - Both nets are trying to optimize a different and opposing loss function in a zero-sum game
- **Potential of GAN**
  - Can be trained to mimic any distribution of data
  - Create worlds eerily similar to our own in any domain

"The most interesting idea in the last 10 years in machine learning" – Yann LeCun

General Architecture of GAN

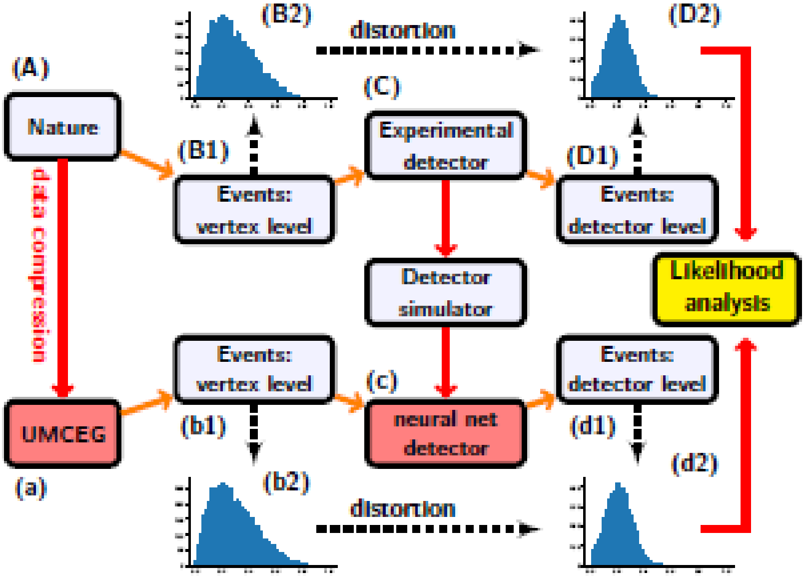


The Power of GAN

- **Can be trained to mimic any distribution of data**
- **Applications**
  - Artificial Arts
  - Virtual Reality
  - New Characters
  - Artificial Music

GAN-based Event Generators

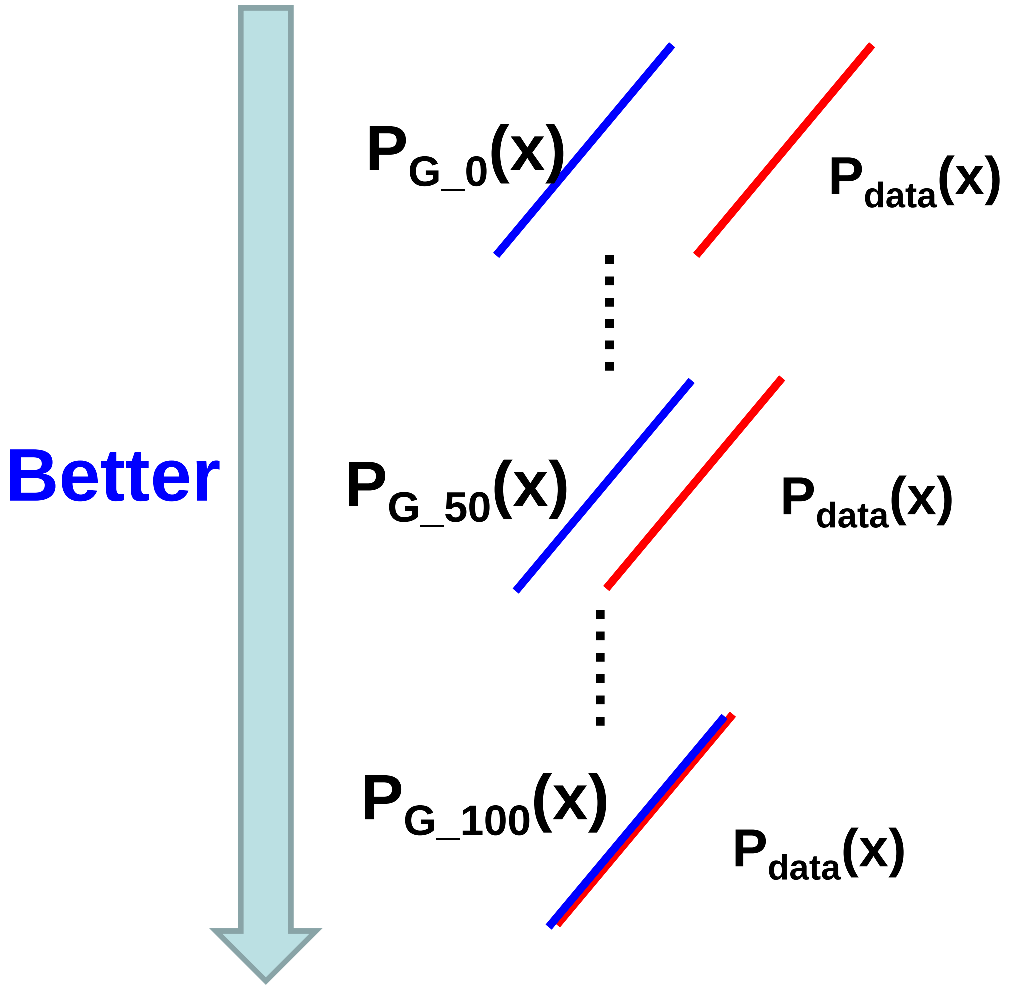
- **Learning from real electron-proton scattering data**
  - Capture rich underlying distributions over data
    - Difficult to model using explicit parameters
- **Faithfully reproducing particle reaction events**
  - No assumptions on femtometer-scale physics theory
- **Overcome the limitations of MCEGs**



Problems in Training a GAN

- **Training a GAN is notoriously difficult**
  - Perfect Discriminator
  - Mode Collapse
  - Non-convergence
  - Imbalance Generator and Discriminator Training
  - Model parameter oscillation
  - Destabilization
  - Vanishing gradient

Too Perfect Discriminator



$JSD(P_{G\_0} \| P_{data}) = \log 2$

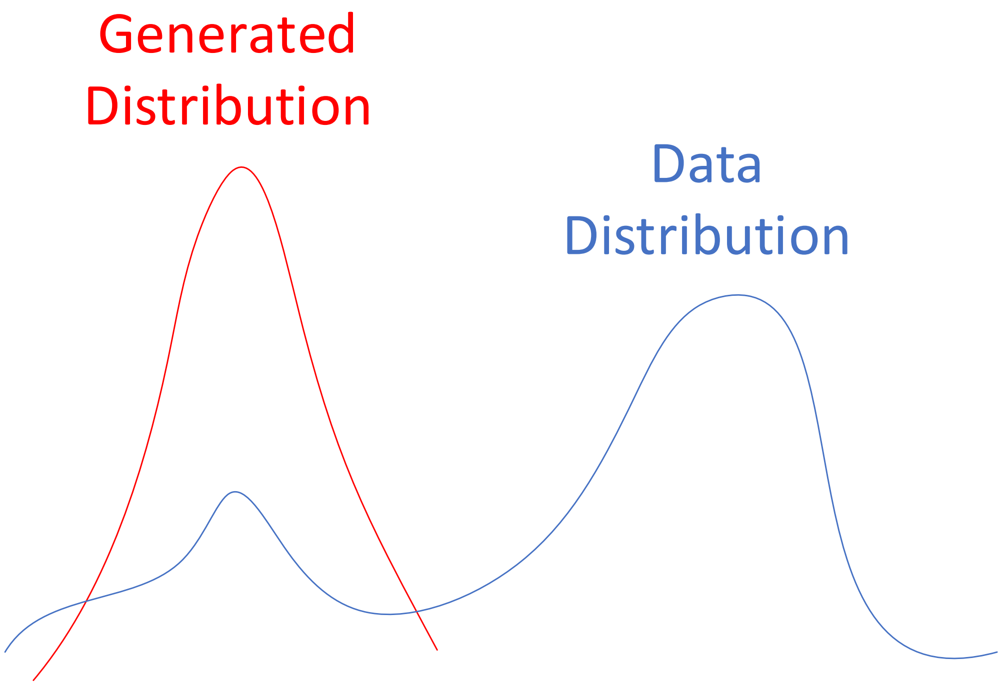
$JSD(P_{G\_50} \| P_{data}) = \log 2$

Not really better ……

$JSD(P_{G\_100} \| P_{data}) = 0$

Mode Collapse

- **Generating the same sample**



Additional GAN Challenges for Physical Event Generators

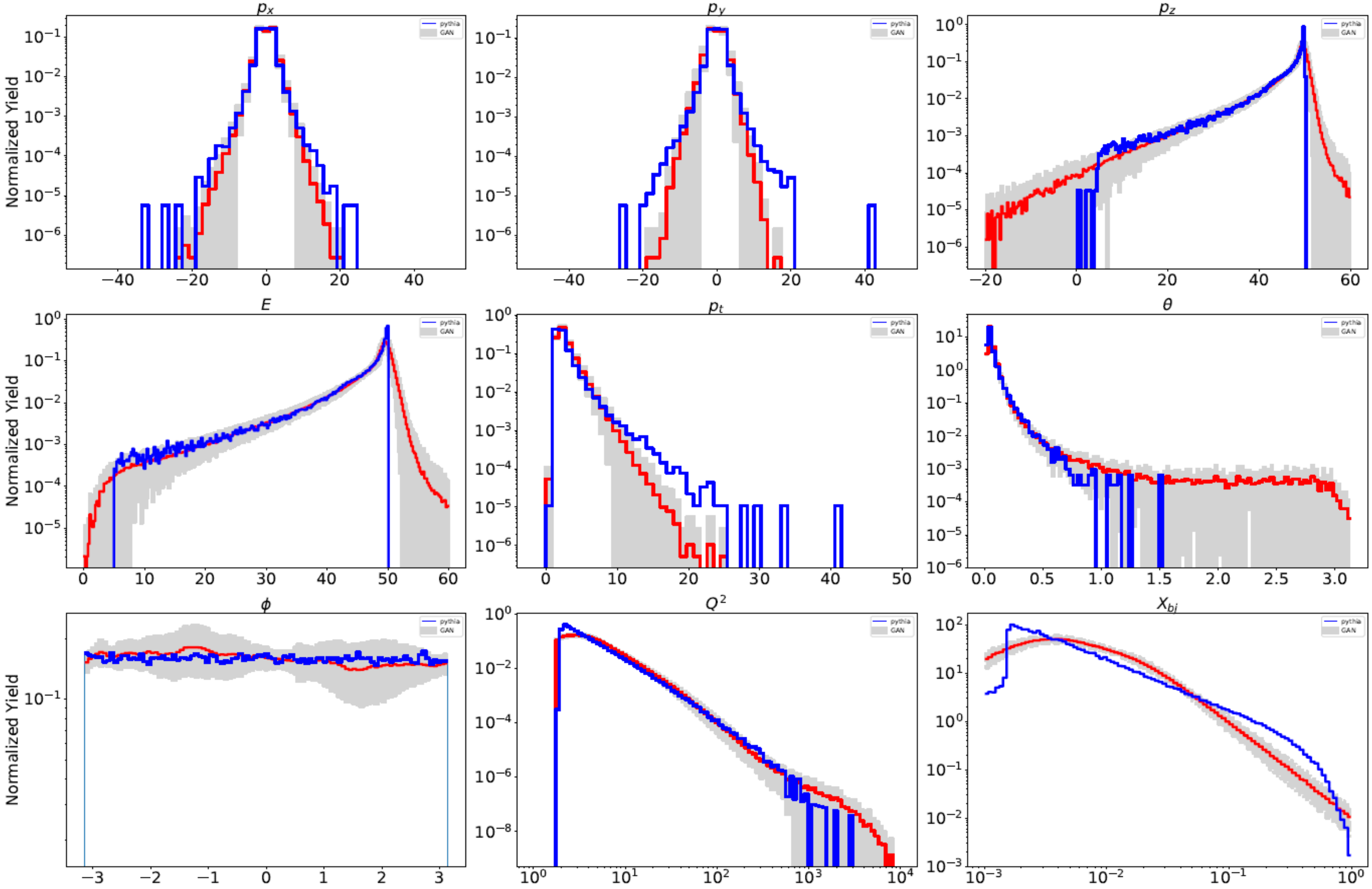
- Precise Event Feature Distributions
  - Replicate the nature of particle reactions faithfully
- Obeying the fundamental Physics Laws
  - Energy Conservation
  - Momentum Conservation

GAN for Inclusive Events

- **Only learns how to generate specific kinds of particles**
  - Instead of the whole spectrum of particles in one event
- **Single electron**
- **Electron + Prion**

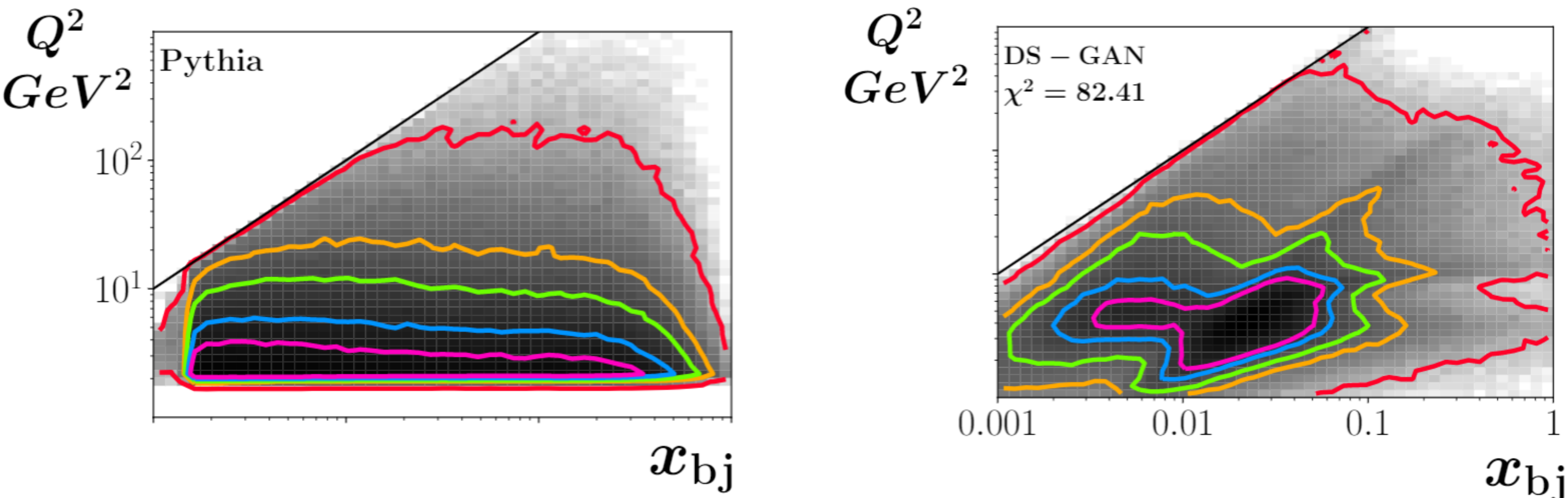
Direct Simulation GAN

- **Directly learn from electron three-momentum vector (px, py, pz)**

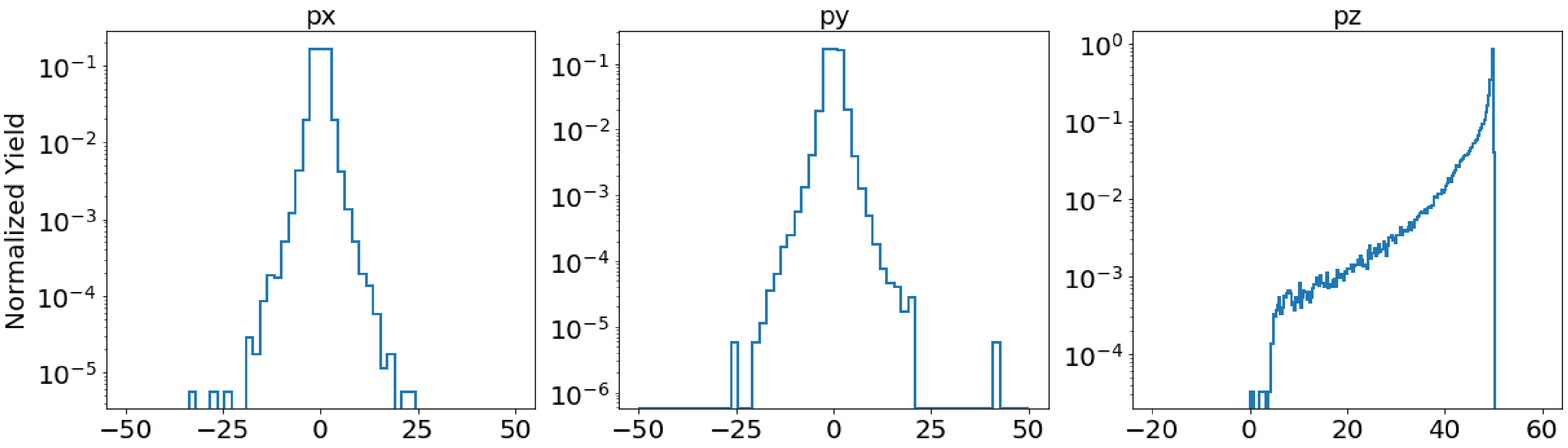


Direct Simulation GAN (cont.)

- **Inter-correlation between physical quantities**



Momenta Distributions of Electrons



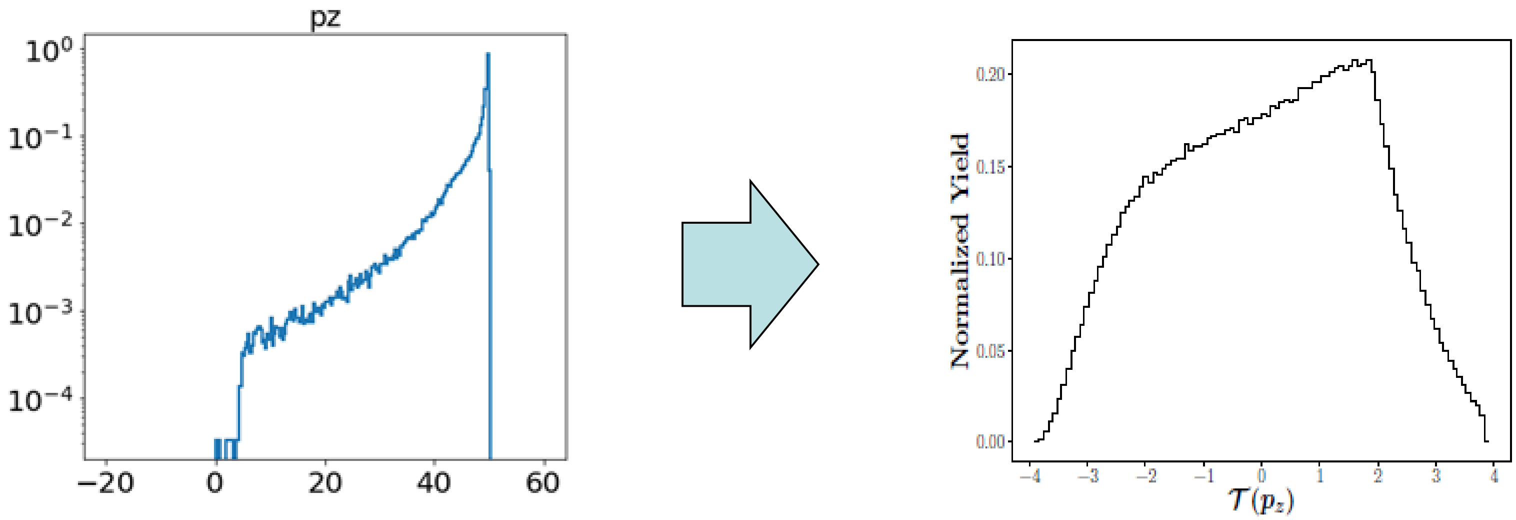
Feature-Augmented Transformed GAN (FAT-GAN)

- **FAT-GAN**
  - Features Transformation
    - Select generated features
      - Not necessarily meaningful physics properties
    - Easier to be generated by the generator
  - Features Augmentation
    - Expand feature space
    - Improve sensitivity of the discriminator
  - Maximum Mean Discrepancy (MMD)
    - Improve Distribution Match
  - Wasserstein Loss
    - Reduce the chance of mode collapse
    - Enhance GAN convergence

Features Transformation

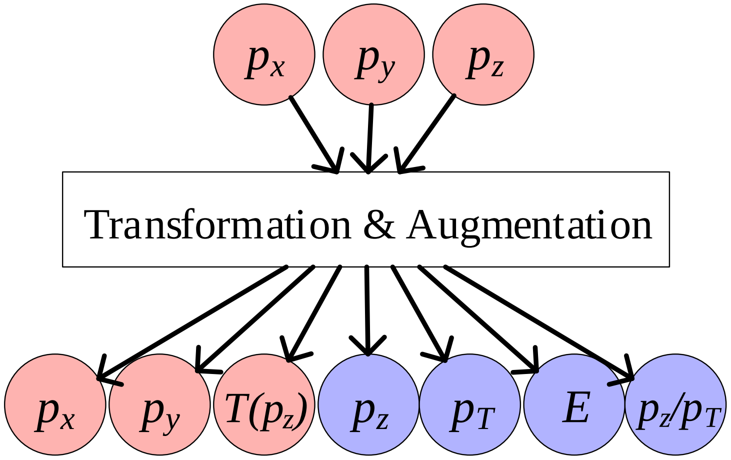
$$\mathcal{T}(p_z) = \log(E_{\mathrm{b}} - p_z)$$

- **Conversion to eliminate sharp edges**
- **Guarantee no generation of non-physical electrons**



Features Augmentation

- **Augment the Feature Space to improve the Sensitivity of Discriminator**



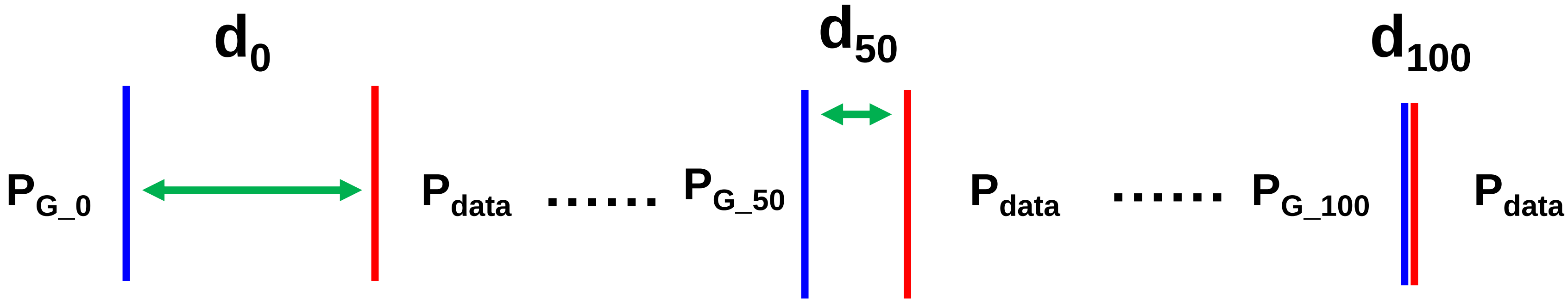
Wasserstein GAN

- **Discriminator**
  - No longer a direct critic of telling the fake samples apart from the real ones
  - Tries to bring $\boldsymbol{x_G}$ distribution closer to $\boldsymbol{x_{true}}$ distribution

- **Wasserstein Loss**

$$\boldsymbol{D_{loss} = E\big(D(x_{true})\big) - E\big(D(x_G)\big)}$$
$$\boldsymbol{G_{loss} = -E(D(x_G))}$$

Wasserstein Distance vs. JSD



$JS(P_{G\_0}, P_{data}) = \log 2$ $JS(P_{G\_50}, P_{data}) = \log 2$ $JS(P_{G\_100}, P_{data}) = 0$

$W(P_{G\_0}, P_{data})=d_0$ $W(P_{G\_50}, P_{data})=d_{50}$ $W(P_{G\_100}, P_{data})=0$

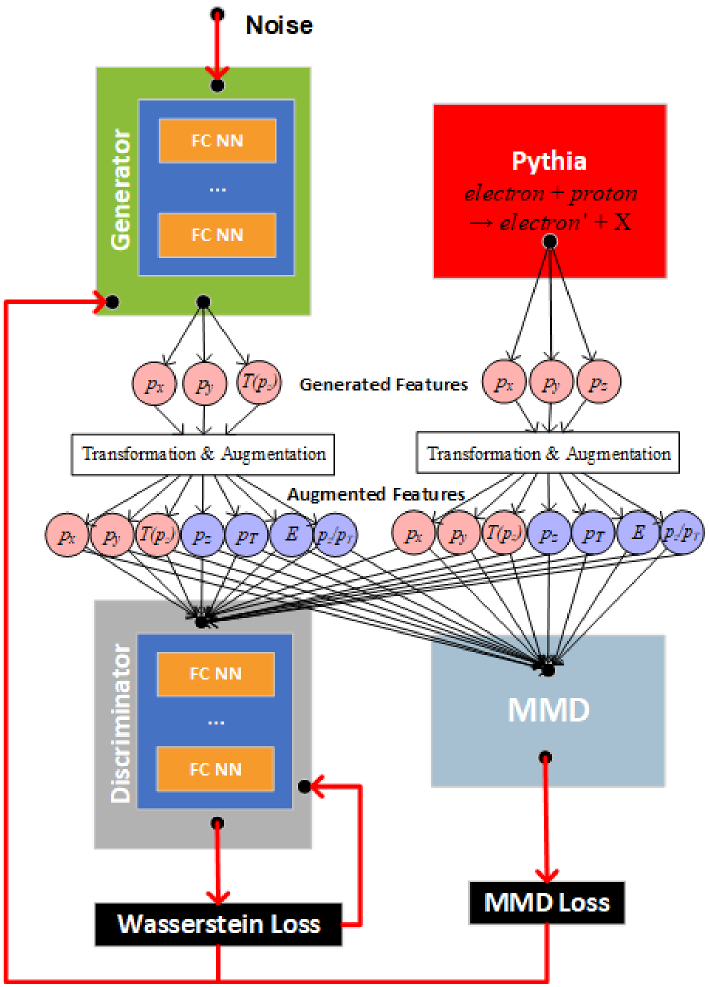
Maximum Mean Discrepancy (MMD)

- MMD
  - A kernel-based two-sample test to compare two distributions
  - Determine if the two sets of samples are drawn from different distributions

$$\begin{aligned}\mathrm{MMD}^2(\boldsymbol{p}, \widetilde{\boldsymbol{p}}) = & \ \mathbb{E}_{\boldsymbol{p}_a, \boldsymbol{p}_{a'} \sim P_{\boldsymbol{p}}}[k(\boldsymbol{p}_a, \boldsymbol{p}_{a'})] \\ & + \mathbb{E}_{\boldsymbol{p}_b, \boldsymbol{p}_{b'} \sim P_{\widetilde{\boldsymbol{p}}}}[k(\boldsymbol{p}_b, \boldsymbol{p}_{b'})] \\ & - 2\, \mathbb{E}_{\boldsymbol{p}_a \sim P_{\boldsymbol{p}}, \boldsymbol{p}_b \sim P_{\widetilde{\boldsymbol{p}}}}[k(\boldsymbol{p}_a, \boldsymbol{p}_b)]\end{aligned}$$

  - $k(\boldsymbol{p}_a, \boldsymbol{p}_b)$ is a Gaussian kernel

FAT-GAN Architecture



- Discriminator Loss

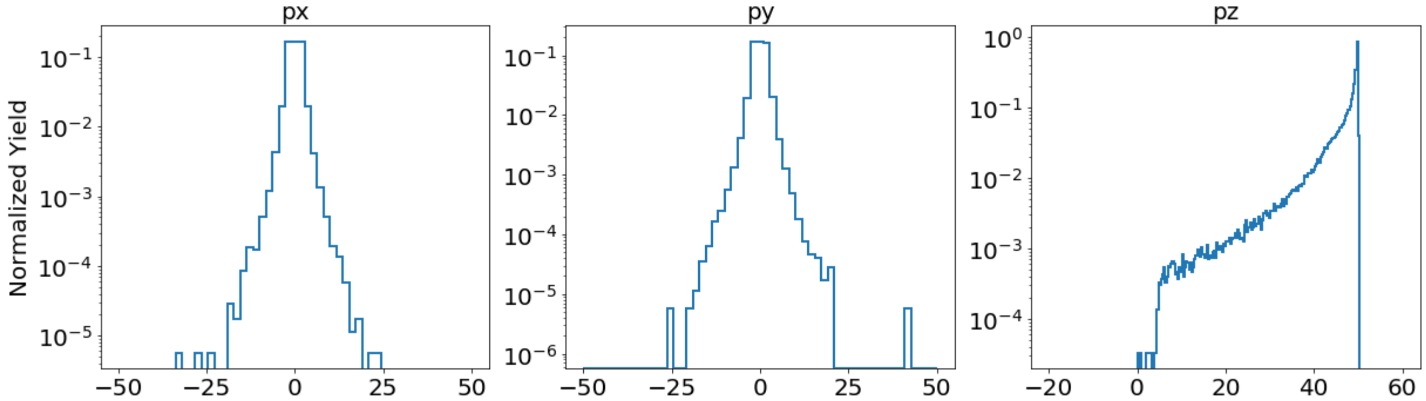
$$L_D = \left(\mathbb{E}[D(\widetilde{\boldsymbol{p}})] - \mathbb{E}[D(\boldsymbol{p})]\right) + \lambda\, \mathbb{E}_{\widehat{\boldsymbol{p}} \sim P_{\widehat{\boldsymbol{p}}}}\left[(\|\nabla_{\widehat{\boldsymbol{p}}} D(\widehat{\boldsymbol{p}})\|_2 - 1)^2\right]$$

- Generator Loss

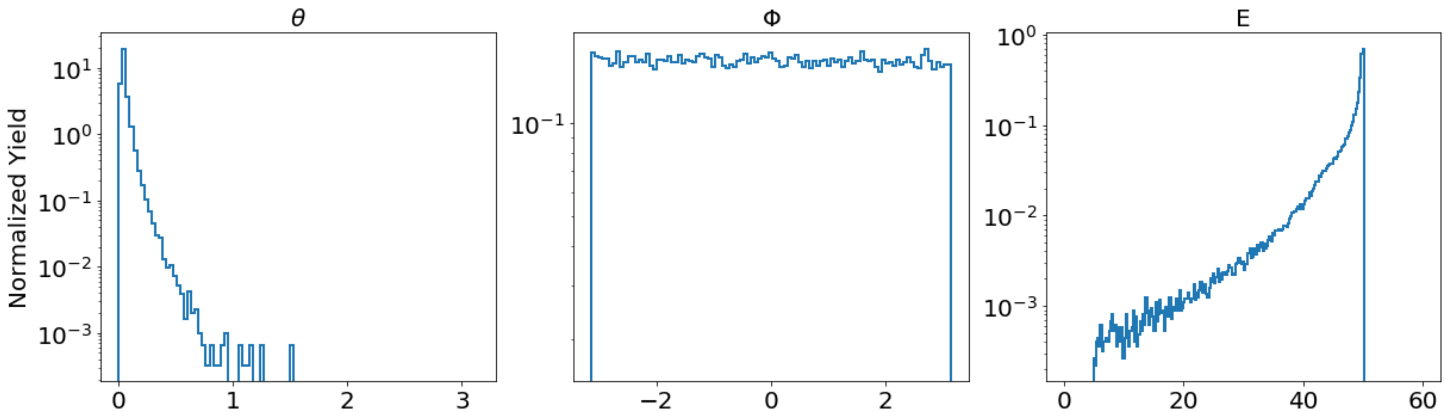
$$L_G = -\mathbb{E}[D(\widetilde{\boldsymbol{p}})] + \eta\, \mathrm{MMD}^2(\boldsymbol{p}, \widetilde{\boldsymbol{p}})$$

Comparison between Representations in Cartesian Coordinates and Spherical Coordinates

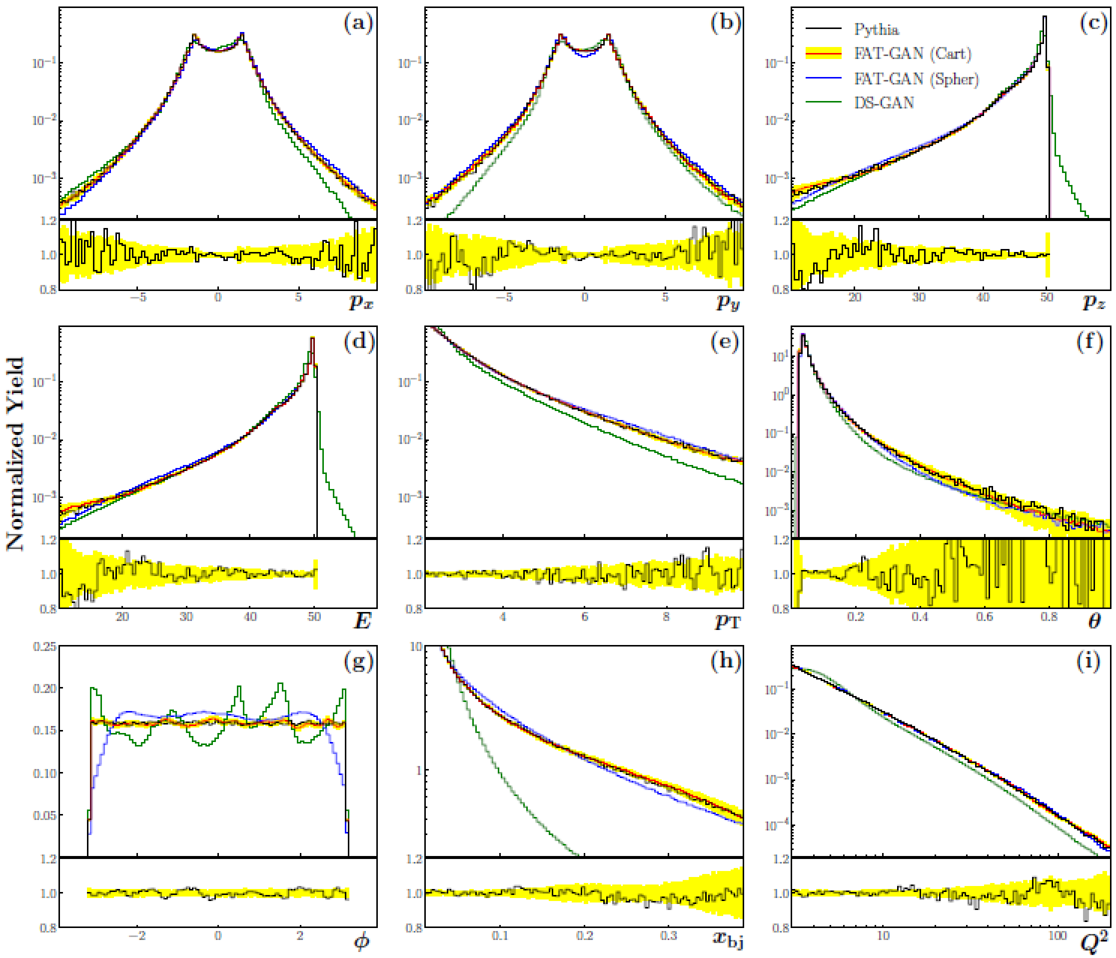
- **Representation in Cartesian Coordinates [FAT-GAN (Cart)]**



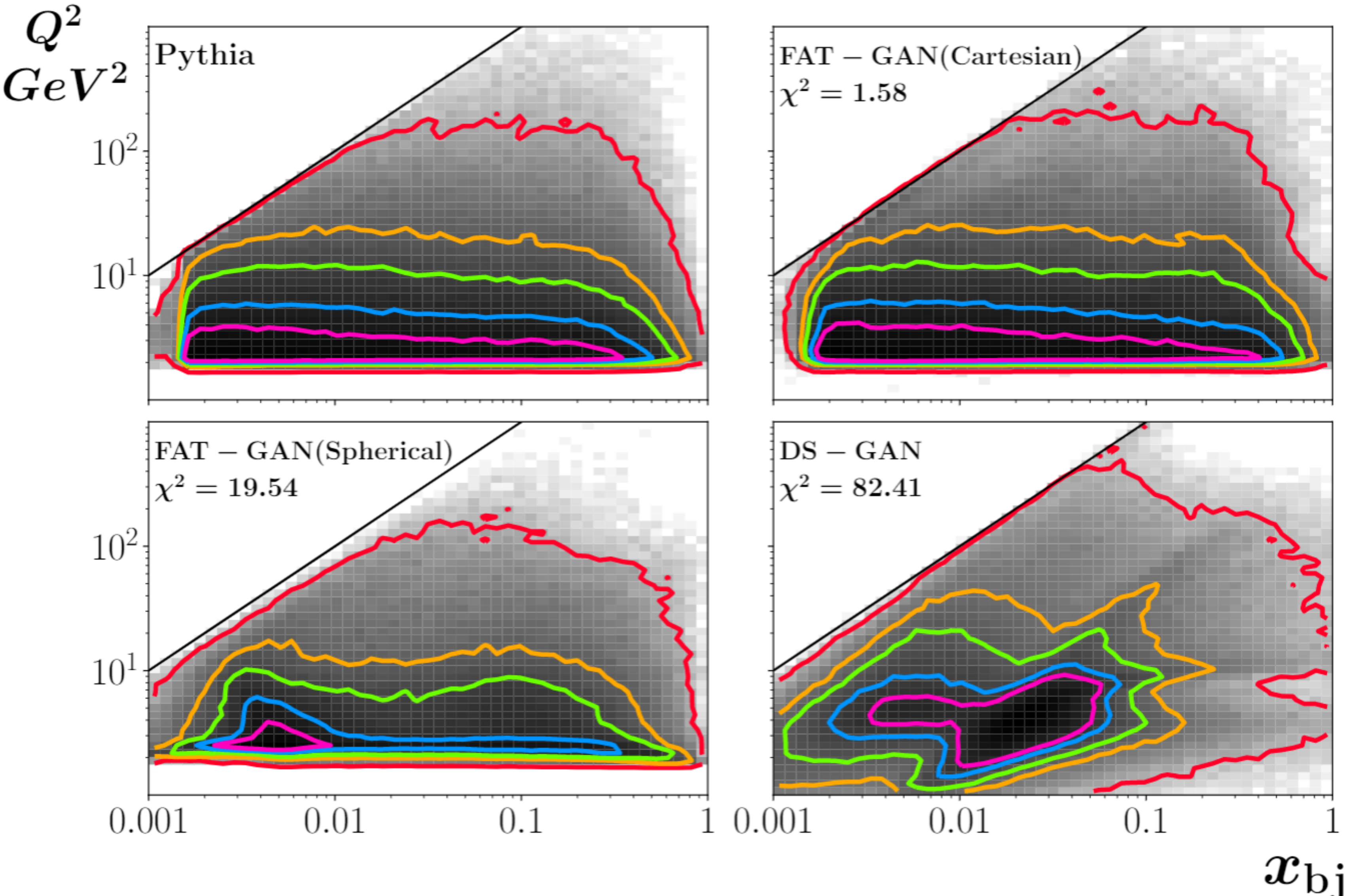
- **Representation in Spherical Coordinates [FAT-GAN (Spher)]**



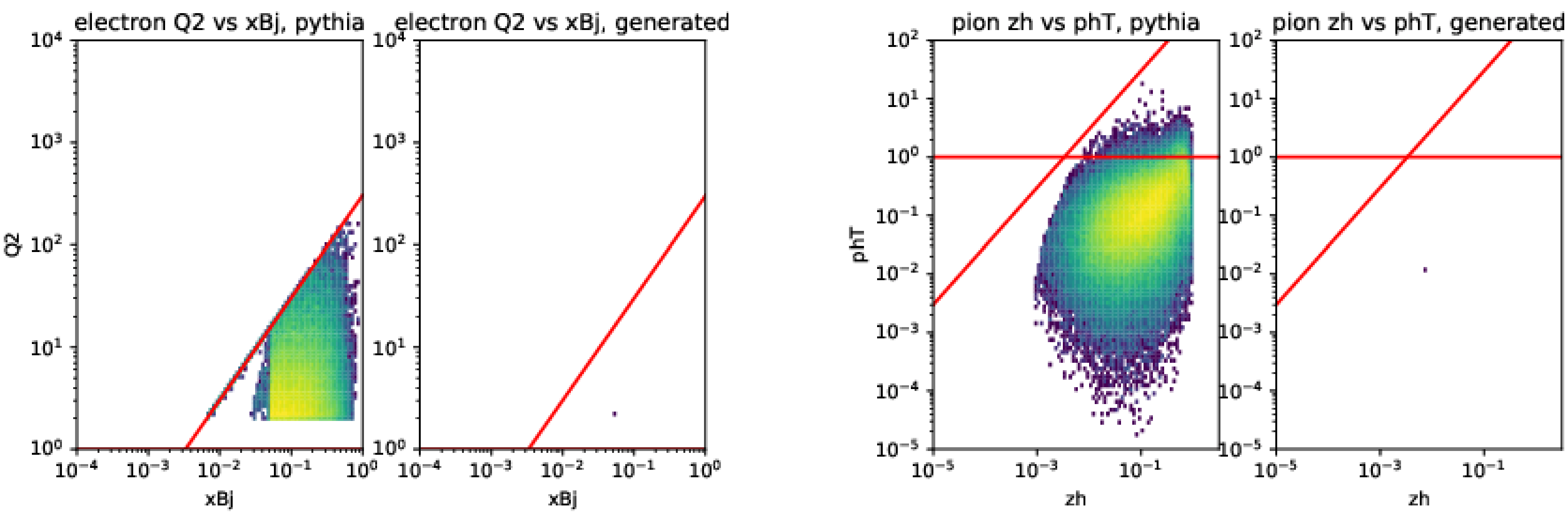
Distributions of Generated Physical Properties



Features Inter-correlations



FAT-GAN for Two Particles (Electron + Prion)



Summary

- **GAN-based Event Generator**
  - Challenge: Complicated patterns in physics property distributions
  - FAT-GAN Features are the KEY
    - Not necessary using meaningful physics properties as generated features
    - Use features that are easy to generate by the generator
    - Augment feature space to make the discriminator sensitive
- **Success so far**
  - We can model the inclusive electron with high precision
  - We can model the inclusive electron and pion with high precision

FAT-GAN Paper

Yasir Alanazi, N. Sato, Tianbo Liu, W. Melnitchouk, Michelle P. Kuchera, Evan Pritchard, Michael Robertson, Ryan Strauss, Luisa Velasco, Yaohang Li, "Simulation of electron-proton scattering events by a Feature-Augmented and Transformed Generative Adversarial Network (FAT-GAN)," arXiv:2001.11103, 2020.

Codes: https://github.com/JeffersonLab/FAT-GAN

Acknowledgements

We thank Jianwei Qiu for helpful discussions. This work was supported by the LDRD project No. LDRD19-13 and No. LDRD20-18.