# MLEG
# Status and plans

Astrid N. Hiller Blin
**astrid.hiller-blin@itp.uni-tuebingen.de**
Universität Regensburg
Eberhard Karls Universität Tübingen

JPAC Jefferson Lab
Thomas Jefferson National Accelerator Facility

**MC4EIC**
November 17, 2022

# Outline

Event generators in modern particle physics

Machine learning event generators:
Goals and methods
GAN

Electron-proton scattering:
FAT conditional GAN
Unfolding
Training on data

Horizon

# Outline

**<u>Event generators in modern particle physics</u>**

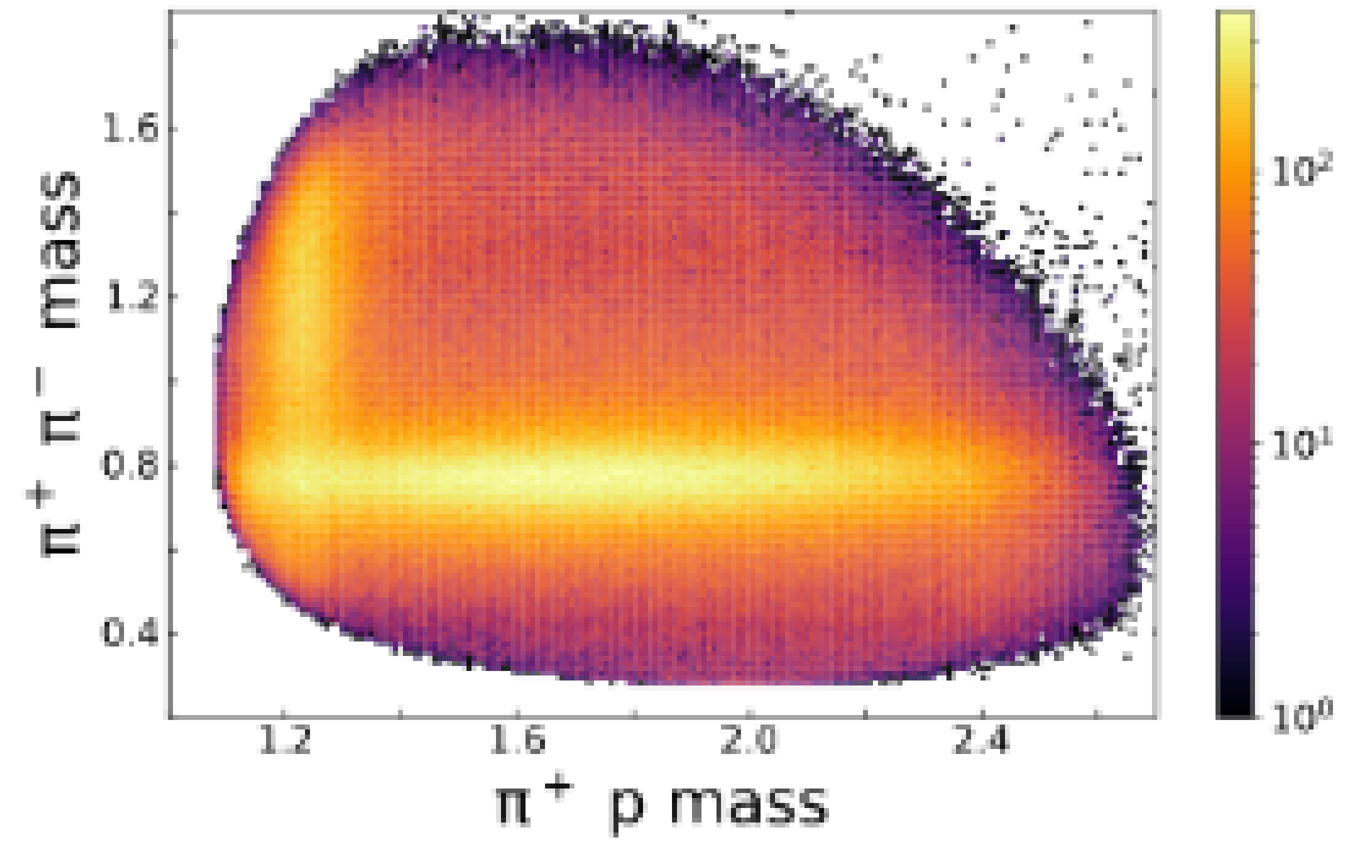
Machine learning event generators:
Goals and methods
GANs

Electron-proton scattering:
FAT conditional GAN
Unfolding
Training on data

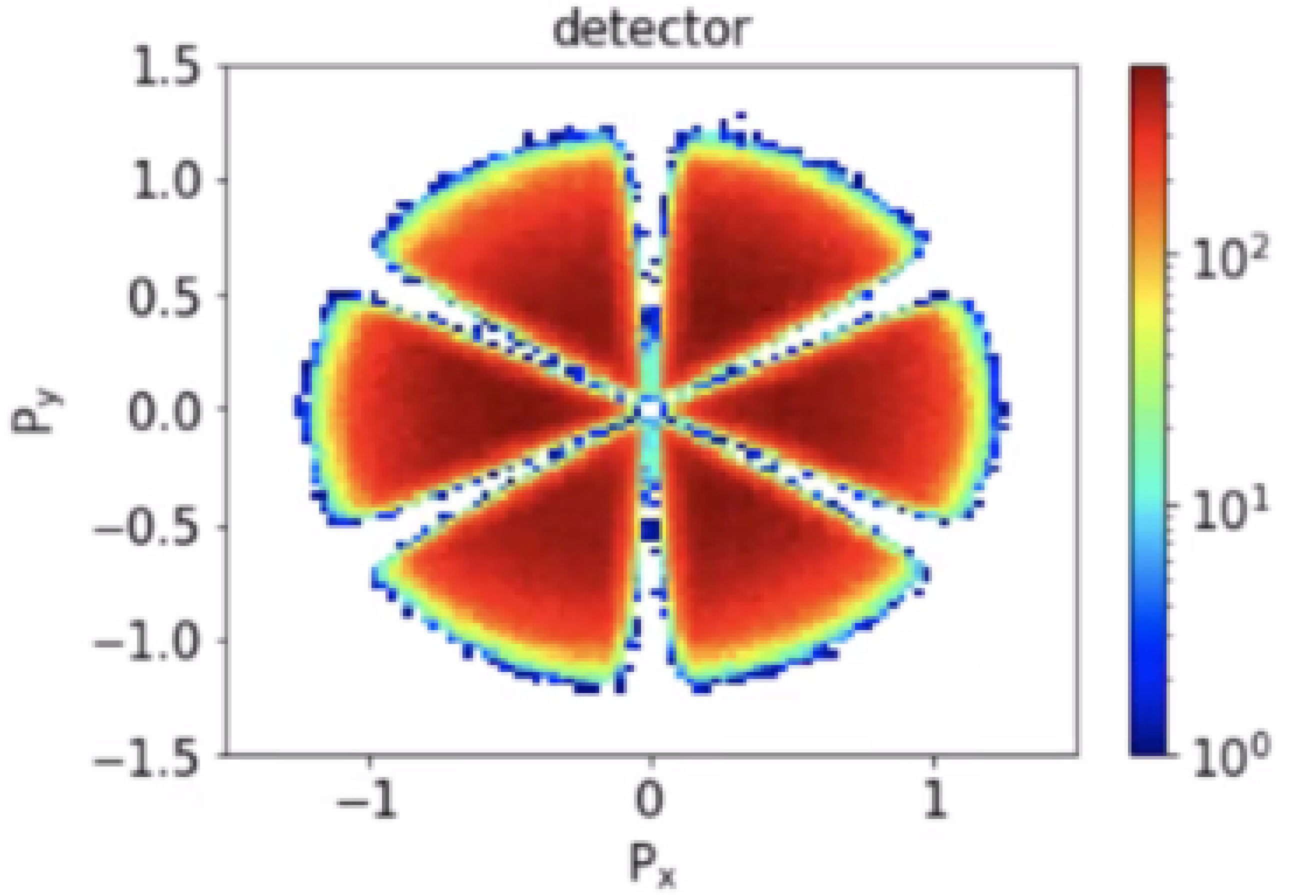
Horizon

# Challenges of modern particle experiments

**Physics:** high kinematic dimensionality, correlations lost in integration, multi-particle final states, resonance peaks.




**Experimental apparatus:** energy losses, decays, new particle production, multiple rescattering.
**Detectors:** limited acceptance, resolution, particle misidentification.




**Simulations for experimental preparation and data analysis are indispensable.**

# Monte Carlo Event Generators (MCEG)

MCEG have been crucial to our physics knowledge to date.

They combine **high-precision experimental data** and the principles of **fundamental theory**, perturbative QCD.

The evolution towards the hadronic regime is well described with **factorization and hadronization phenomenology**.

**Detector simulators** are included to allow for direct experimental simulation.

Some caveats:

– event generation **can be slow** and once generated leads to **large data repositories**;

– phenomenology inputs introduce bias;

– limited physics scope: e.g. spin asymmetries in deep inelastic scattering have not yet been achieved;

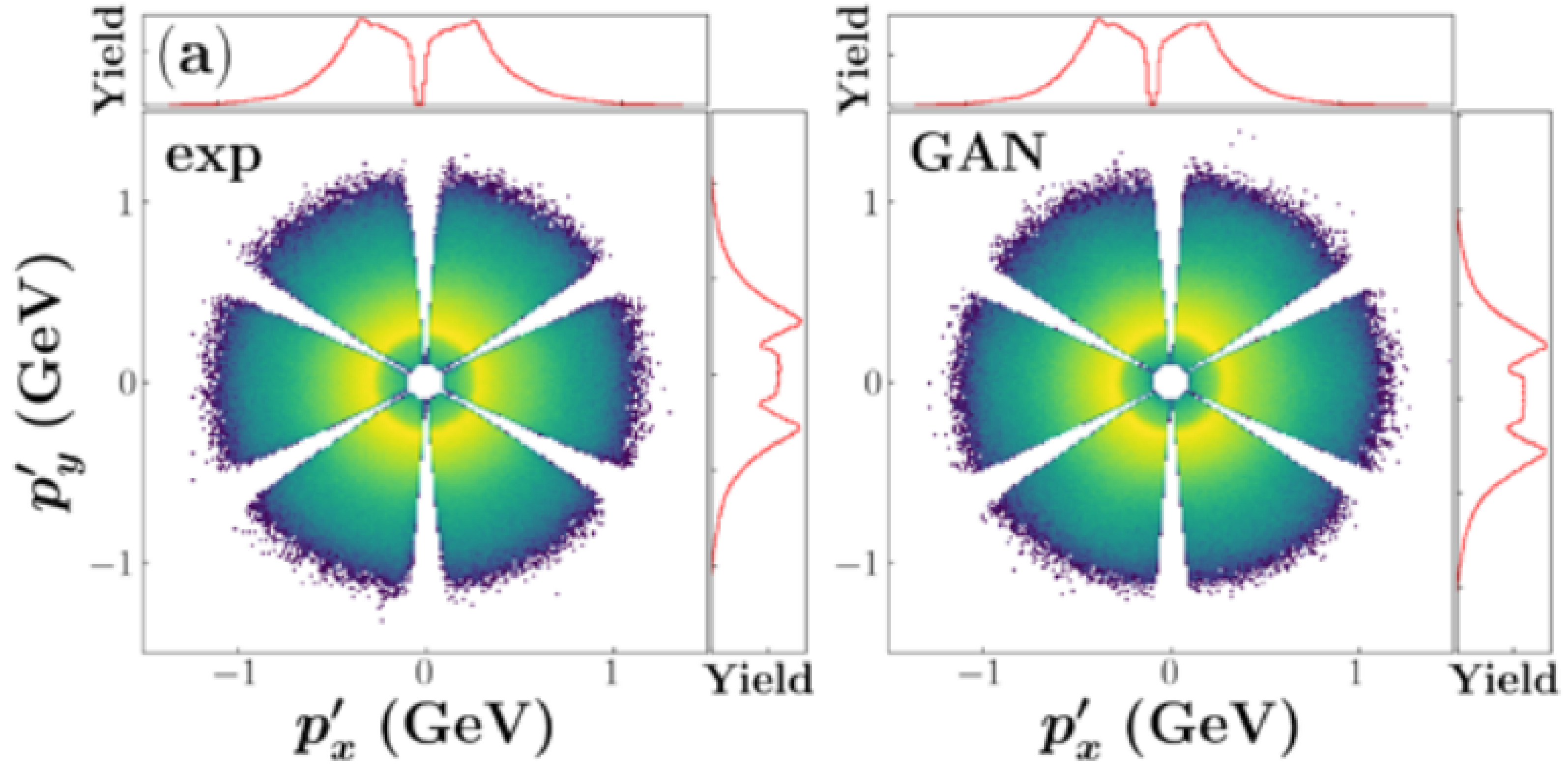
– inverting the Markov chain can be done statistically, but with approximations in practice.
*Bellagente et al., SciPost Phys. 8 (2020) 070*

# Machine Learning Event Generators (MLEG)

MLEG are a strong **complementary** tool for experimental simulation.
*EIC Yellow Report: Abdul Khalek et al., Nucl. Phys. A 1026 (2022) 12247*

When trained on MCEG data, they serve as compact and fast generators:
**after training, the ML file of only a few MB can faithfully generate millions of events per second.**
Advantages for data distribution and storage.
*Alanazi et al., IJCAI 2021, Survey Track, 4286; 2106.00643 [hep-ph]*

When trained on experimental data, one can extract physics with **minimum bias**
and approaches to **unfold vertex-level physics from detector-level events** are underway.




**For the first time in the history of particle experiments,**
**we are in the unique position to incorporate ML tools from the beginning of experimental design.**

# Outline

Event generators in modern particle physics

**<u>Machine learning event generators:</u>**
**<u>Goals and methods</u>**
GANs

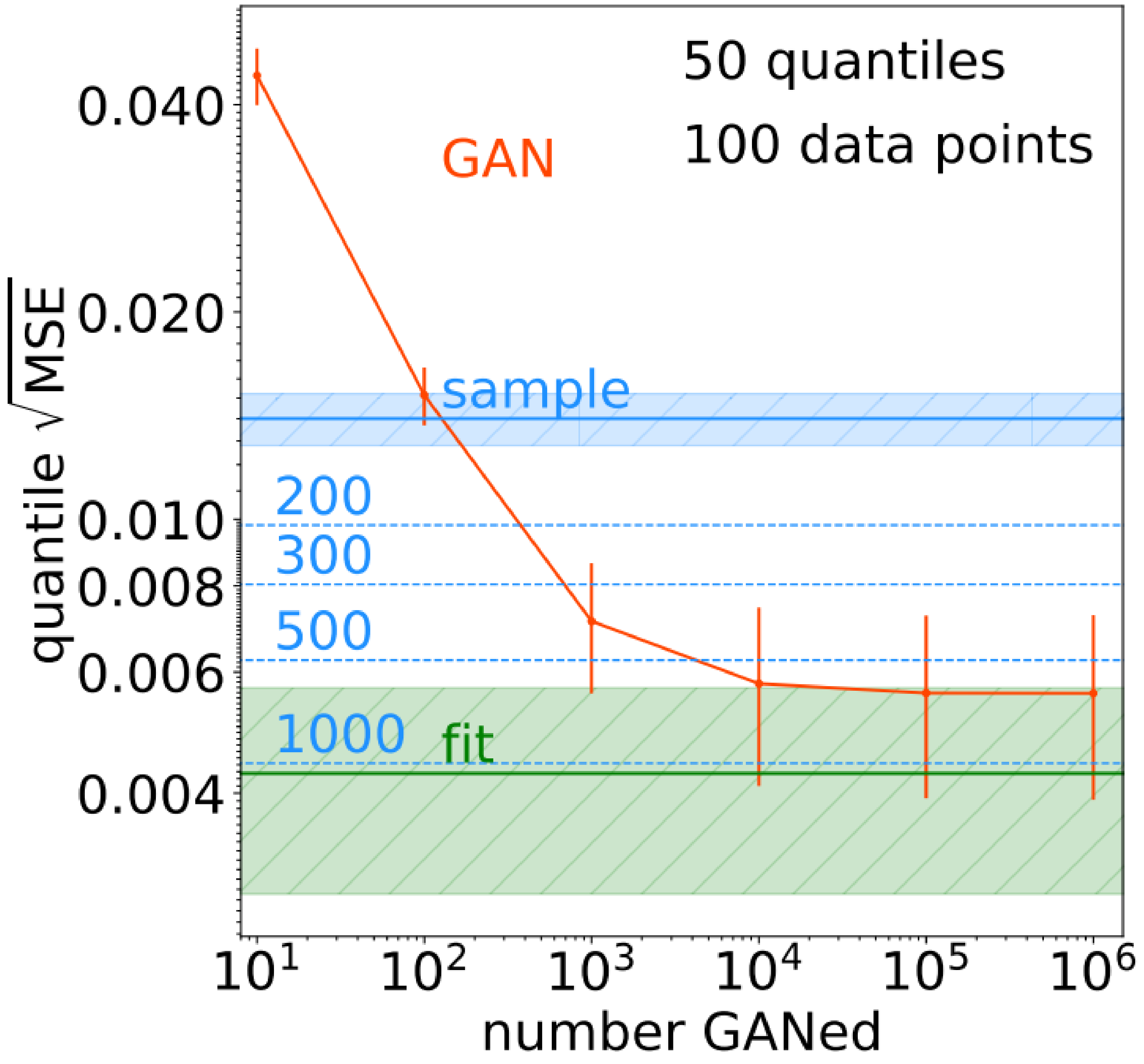
Electron-proton scattering:
FAT conditional GAN
Unfolding
Training on data

Horizon

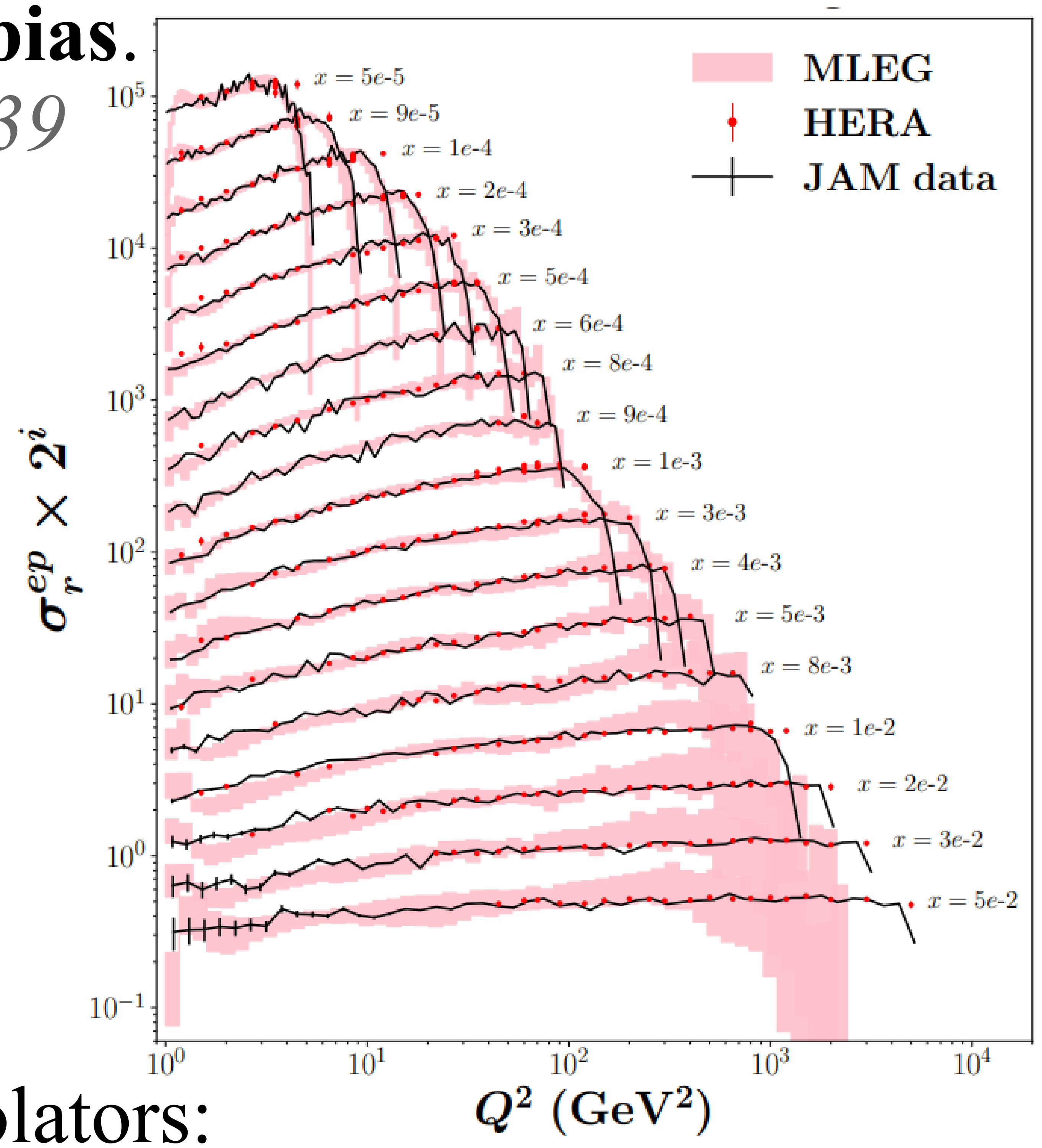
# The fitting procedure

Machine learning is ultimately a fitting procedure, which is very powerful in high-dimensional space.

It can **NOT** outperform functional fits.
In most posed problems the theory function is unknown:
**ML learns the underlying law without bias.**
*Butter et al., SciPost Phys. 10 (2021) 139*




Data from **multiple experiments** can be used in the training process.




Once trained, they serve as excellent interpolators:
**data can be re-generated with high statistics** without the need for binning/histogramming.
*Y. Alanazi et al., Phys. Rev. D 106 (2022) 096002*

# Different methods on the market

Many MLEG approaches exist, also in the context of LHC physics.

Variational auto-encoders (VAE)
*Otten et al., 1901.00875 [hep-ph]*
*Howard et al., 2101.08944 [hep-ph]*

Normalizing flow
*Gao et al., Phys. Rev. D 101 (2020) 076002*

Decision trees
*Darulis et al., 2207.11254*

**Generative adversarial networks (GAN)**
Currently dominate the market of event generation, especially in our field.
*Goodfellow et al., NIPS 2014*

# Outline

Event generators in modern particle physics

Machine learning event generators:
Goals and methods
**GAN**

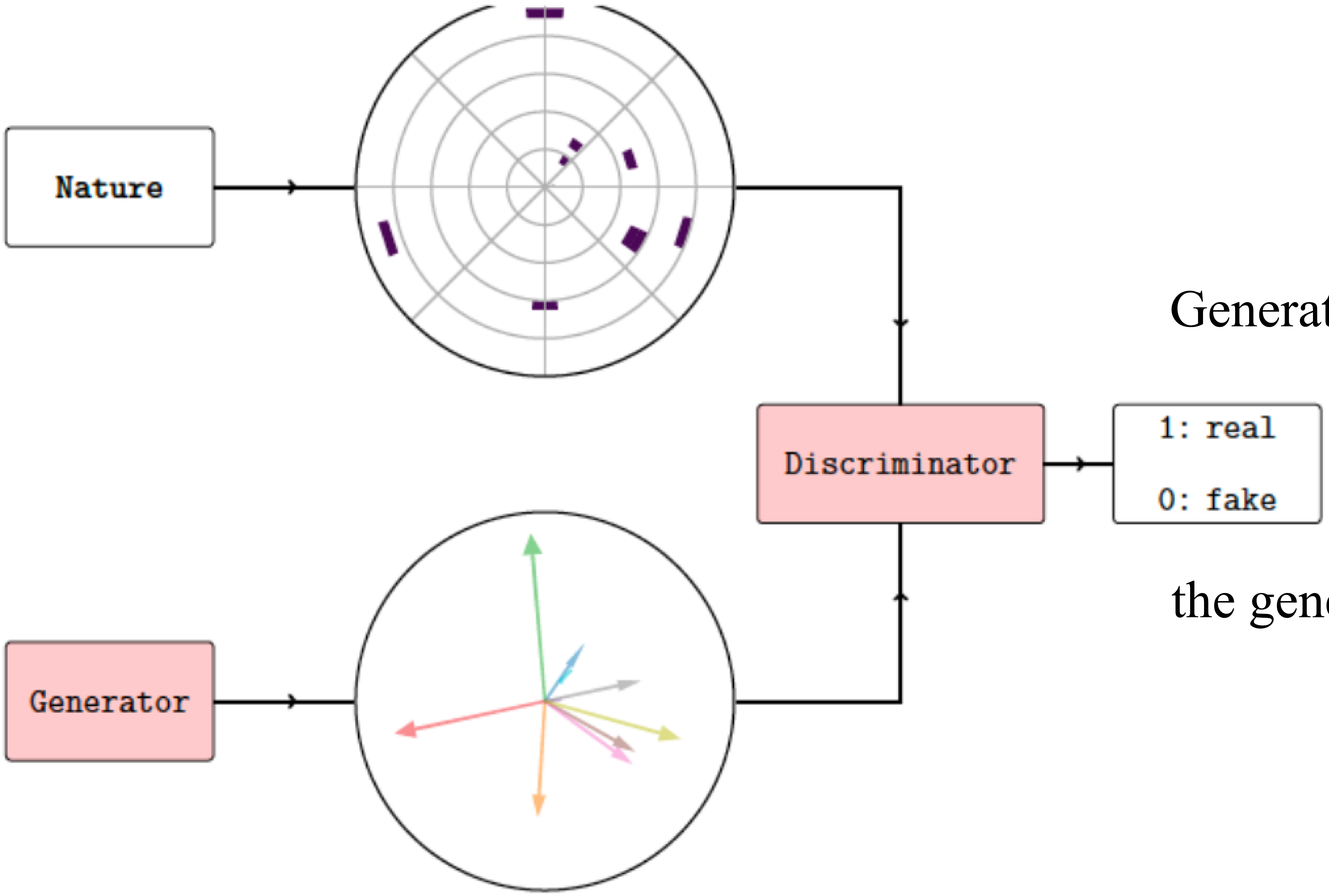
Electron-proton scattering:
FAT conditional GAN
Unfolding
Training on data

Horizon

# Generative adversarial networks (GAN)

**Discriminator** trained to discern real (nature) from fake (generated) events.

**Generator** trained to create events so close to nature that they fool the discriminator.




Generator and discriminator are **trained** simultaneously, competing **adversarially**.

After training convergence, the generator can be used as a **compact data simulator**.

# Panoply of good work

Calorimeter simulation, particle showers.
*Paganini et al., 1705.02355, 1712.10321*

Location-aware jet image training.
*Oliveira et al., 1701.05927*

Detector simulation.
*Musella and Pandolfi, 1805.00850*

Learning sharp local features of phase space, tails and peaks.
*Butter et al., 1907.03764*

.
.
.

# Outline

Event generators in modern particle physics

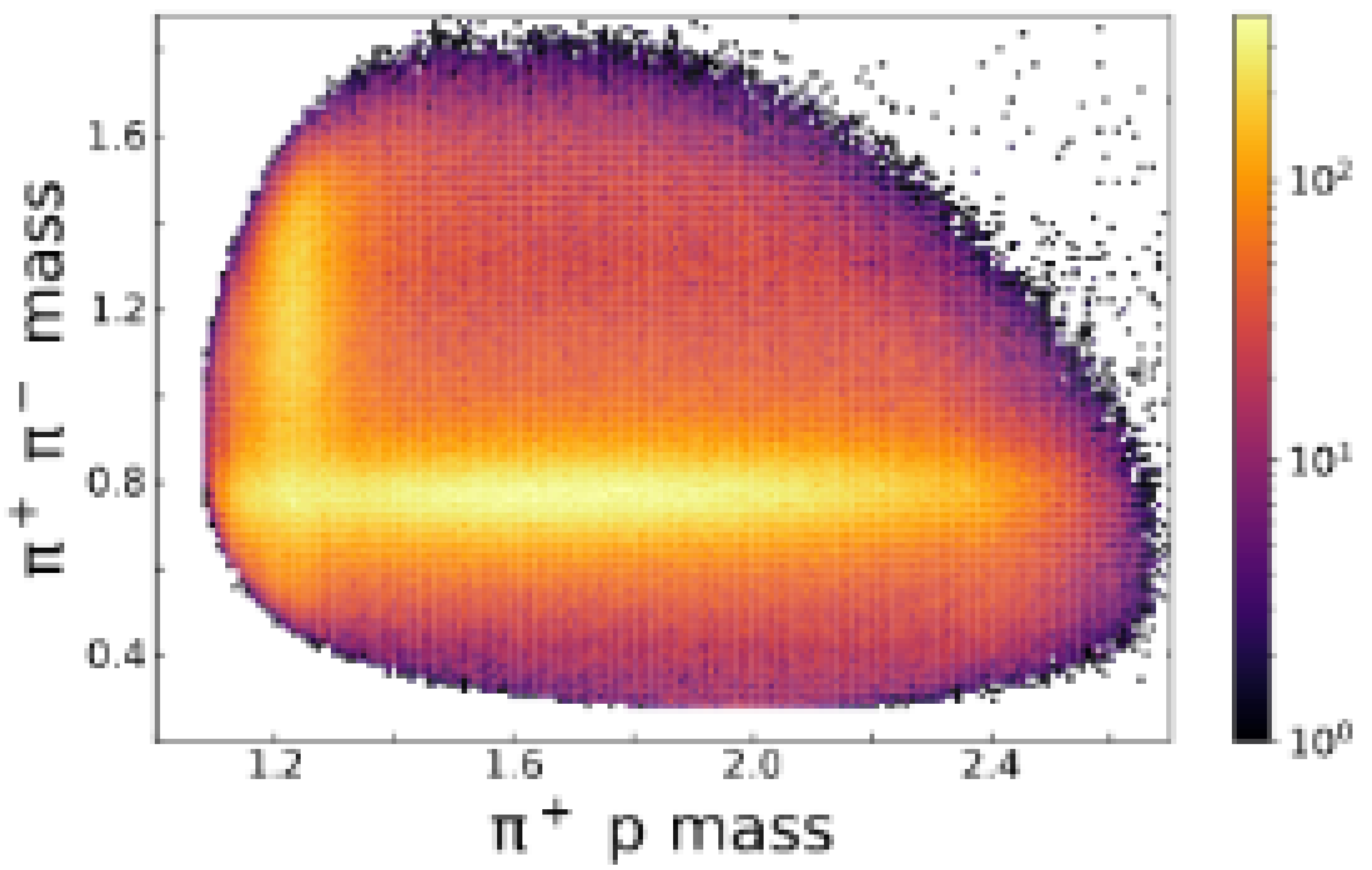
Machine learning event generators:
Goals and methods
GANs

<u>**Electron-proton scattering:**</u>
<u>**FAT conditional GAN**</u>
Unfolding
Training on data

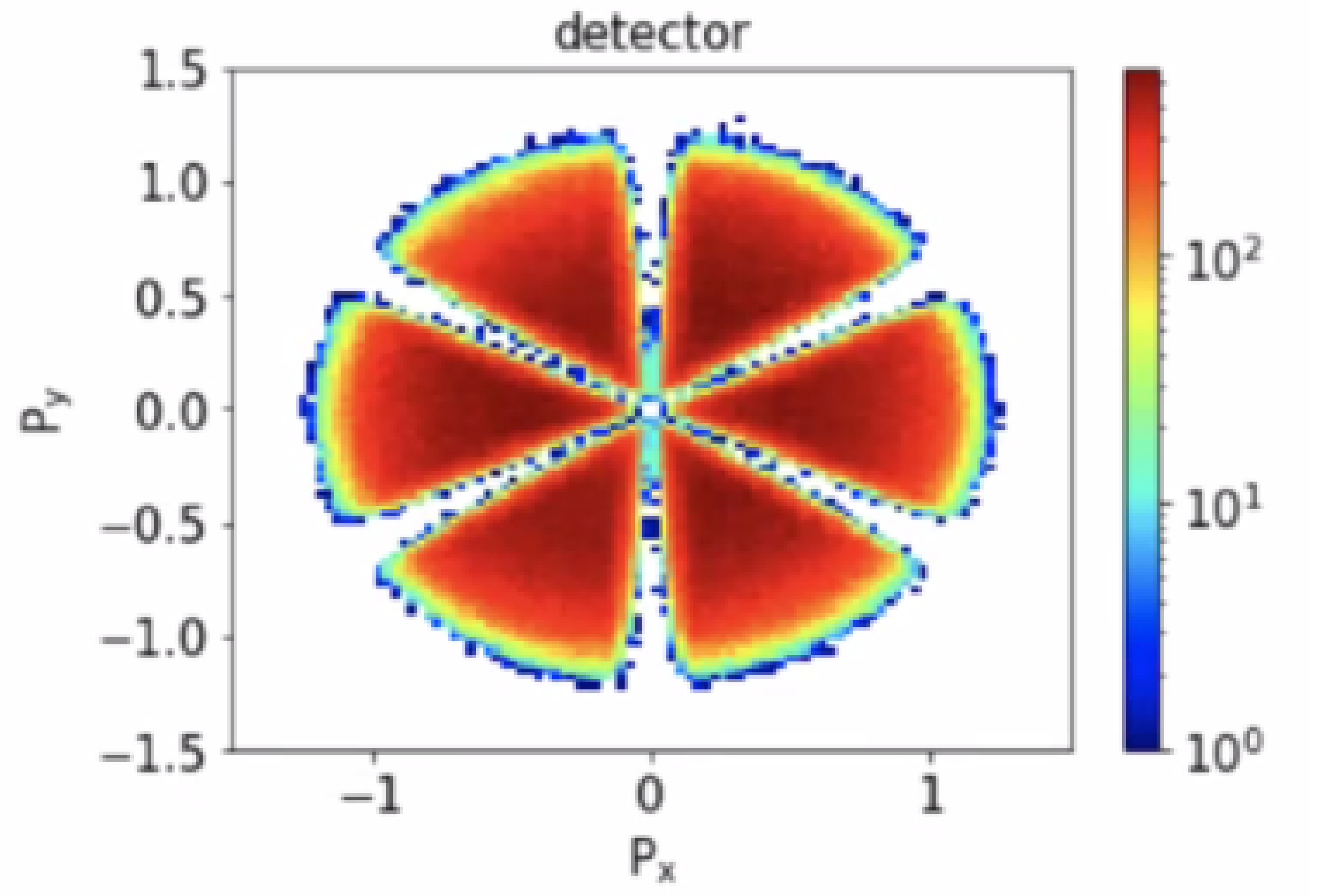
Horizon

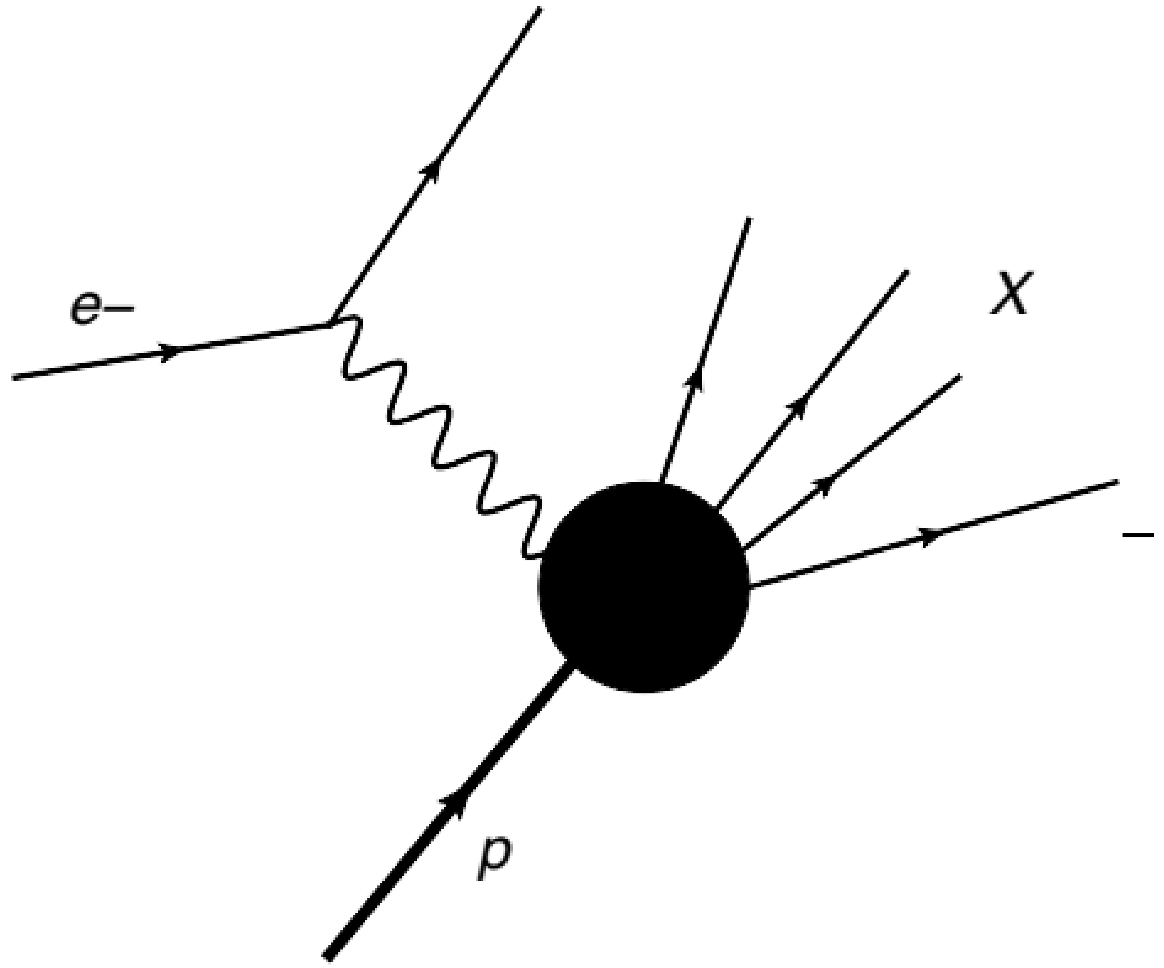
# Challenges and approaches

**Sharp structures** (resonances, limits of phase space, holes in detector acceptance),
**physics laws** (momentum and energy conservation, correlations),
**high dimensionality** lead to




**non-convergence** and vanishing gradients,
parameter oscillation or **instability**,
mode collapse and **overfitting**.

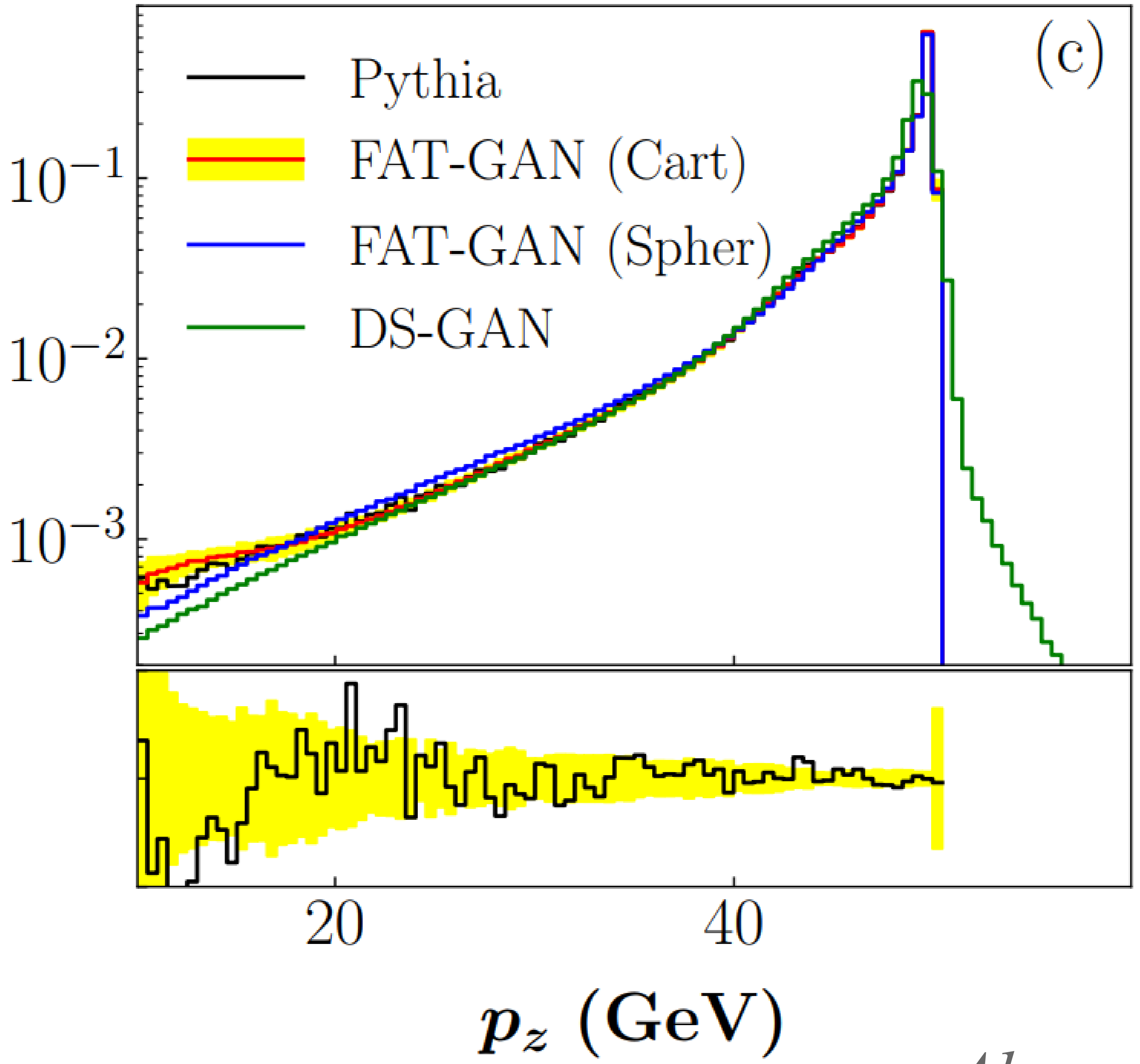


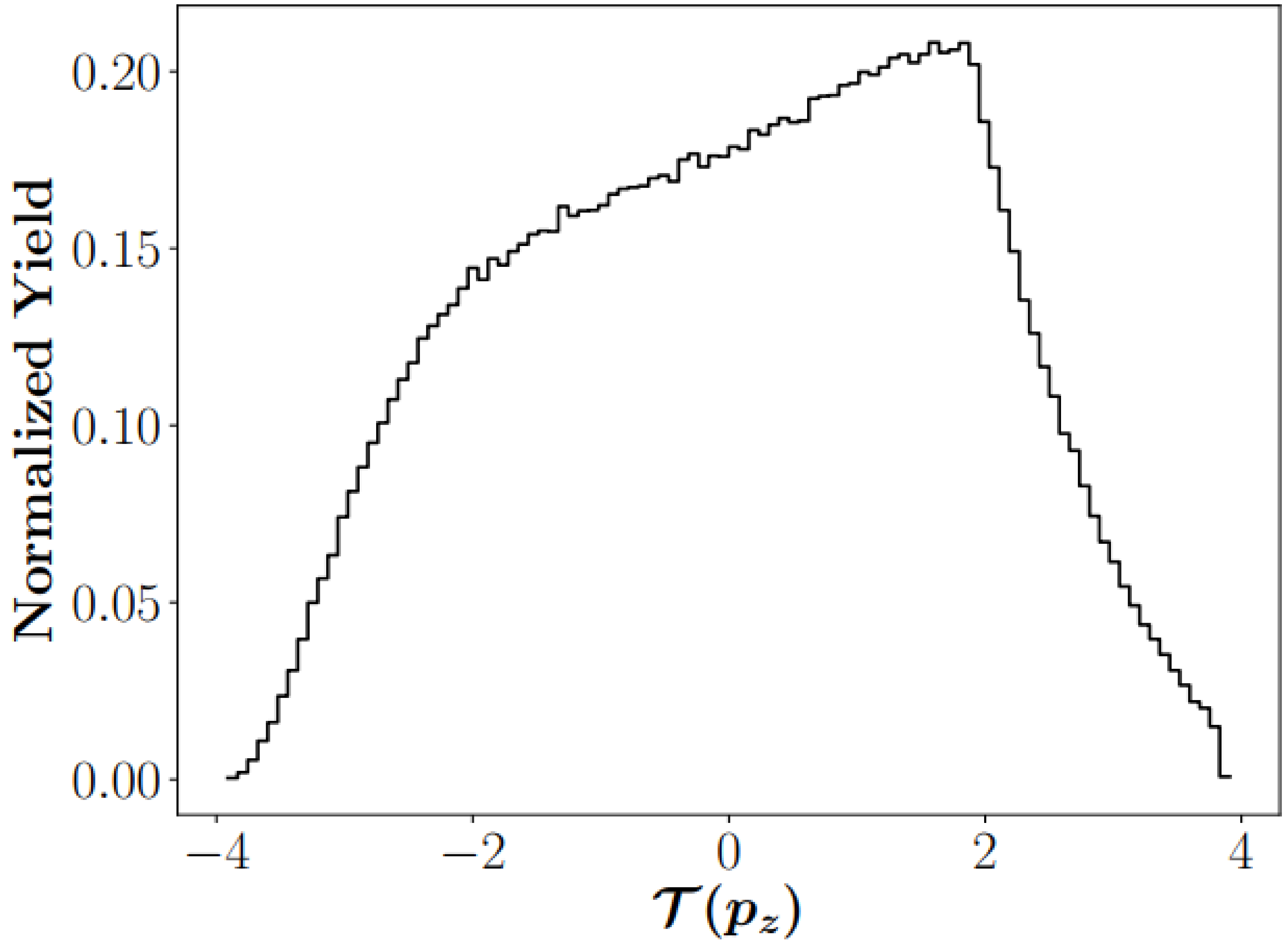




Improved approaches:
– **penalizing samples** too close to boundary;
- **continuous** probabilitstic output of what is fake and what is true;
– feature **augmentation** and **transformation**;
– **conditional** GAN.

# Feature augmentation and transformation (FAT)

The corners of phase space are difficult to learn – the ML does not know that it has to obey physics laws.



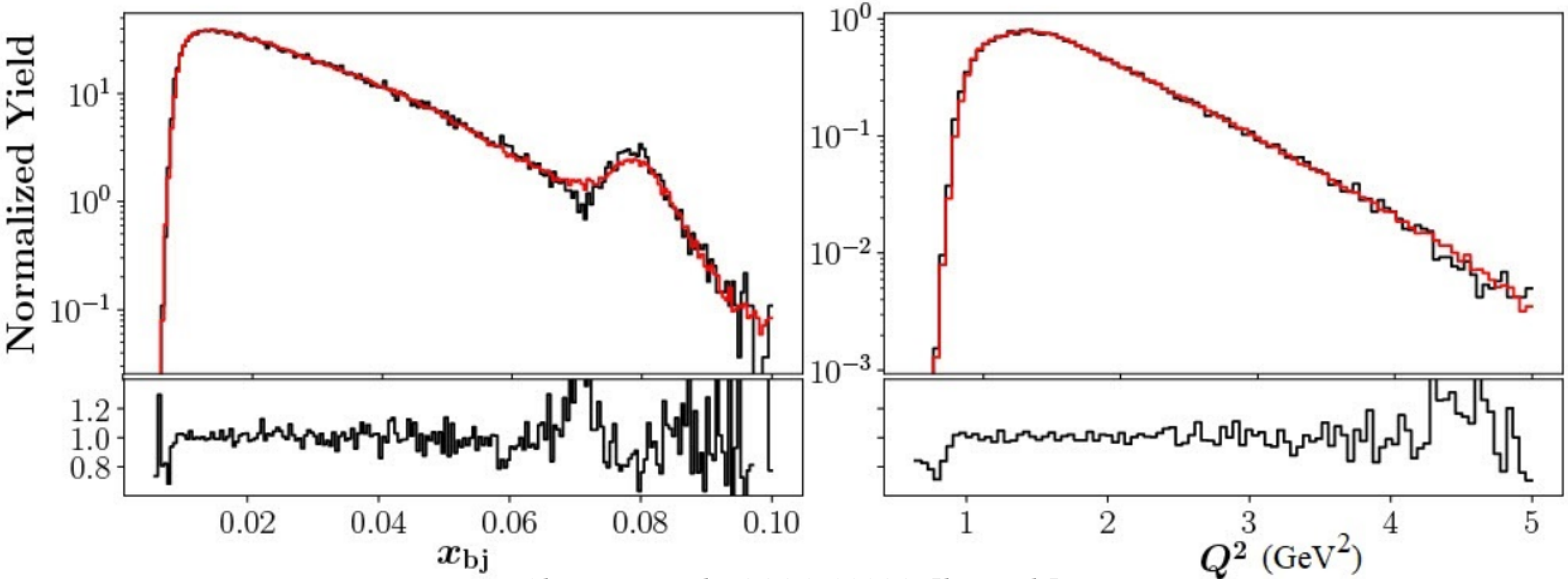

*Alanazi et al., 2001.11103 [hep-ph]*

Transformed features $\mathcal{T}(p_z) = \log\left(\frac{E_b - p_z}{1\text{GeV}}\right)$

Augmented features $p_x,\, p_y,\, \mathcal{T}(p_z),\, p_T,\, E,\, p_z/p_T$

# Conservation of physics correlations

Even features untrained on are well generated: physics correlations are learned and kept in the generated data!
**Physics validation** shows that one can extract the same physics from MLEG as from original data.

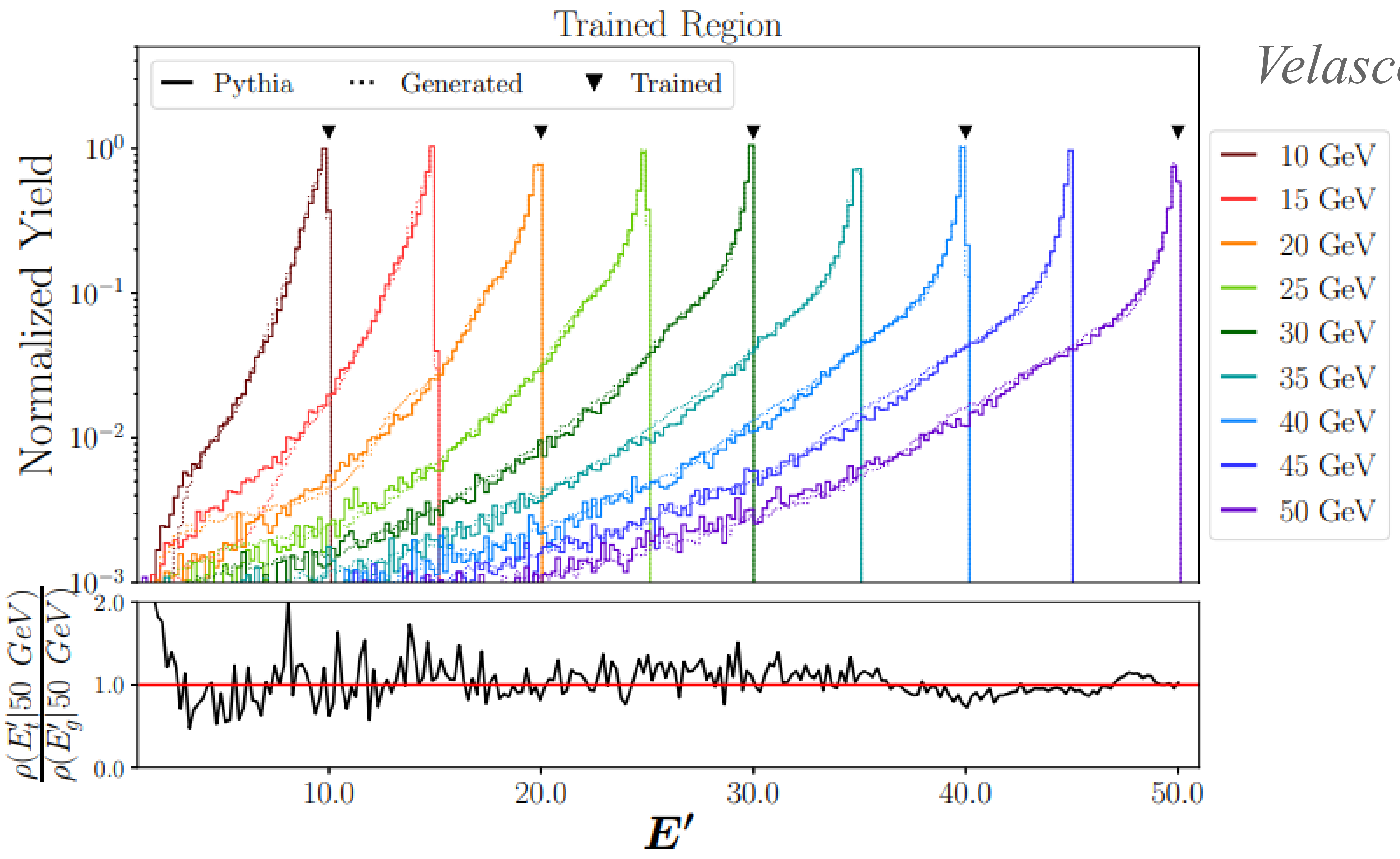



*Alanazi et al., 2001.11103 [hep-ph]*

# Conditional FAT GAN

One wants to avoid training such that only one experiment (one energy setup) can be reproduced.

The cFAT-GAN receives:
– usual **input features**;
– **additional conditions** to each training sample, e.g. beam energy.




*Velasco et al., ICMLA (2020) 372*

The cFAT-GAN has excellent agreement with data **interpolation** region, even for those beam energies **NOT trained on**.

# Outline

Event generators in modern particle physics

Machine learning event generators:
Goals and methods
GANs

Electron-proton scattering:
FAT conditional GAN
**Unfolding**
Training on data

Horizon

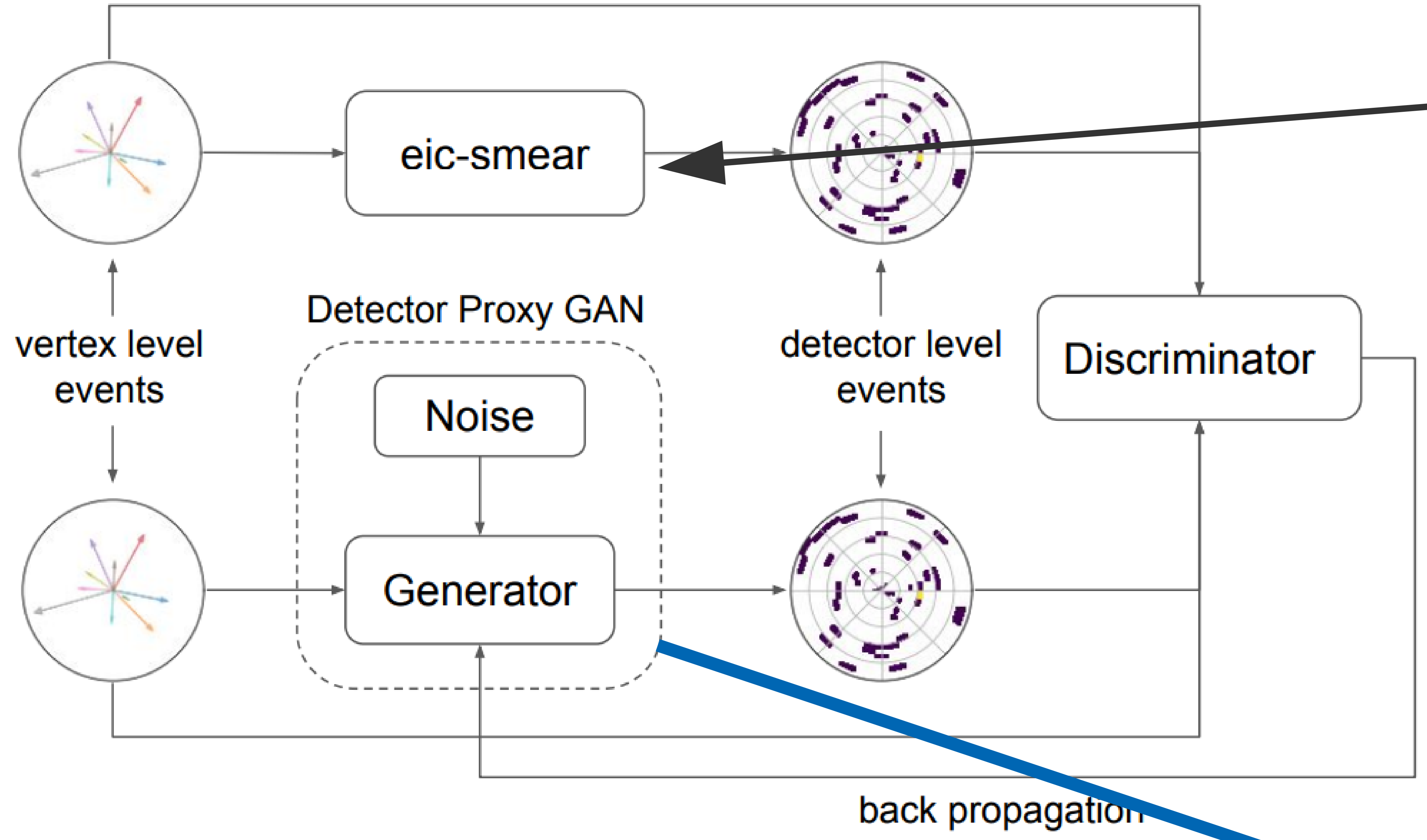
# From vertex to detector and back again

Experimental apparatus effects distort nature into detector-level events.

Unfolding from detector to vertex-level events is a physics reconstruction strategy.
*Datta et al., 1806.00433 [physics.data-an]*
*Andreassen et al., Phys. Rev. Lett. 124 (2020) 182001*

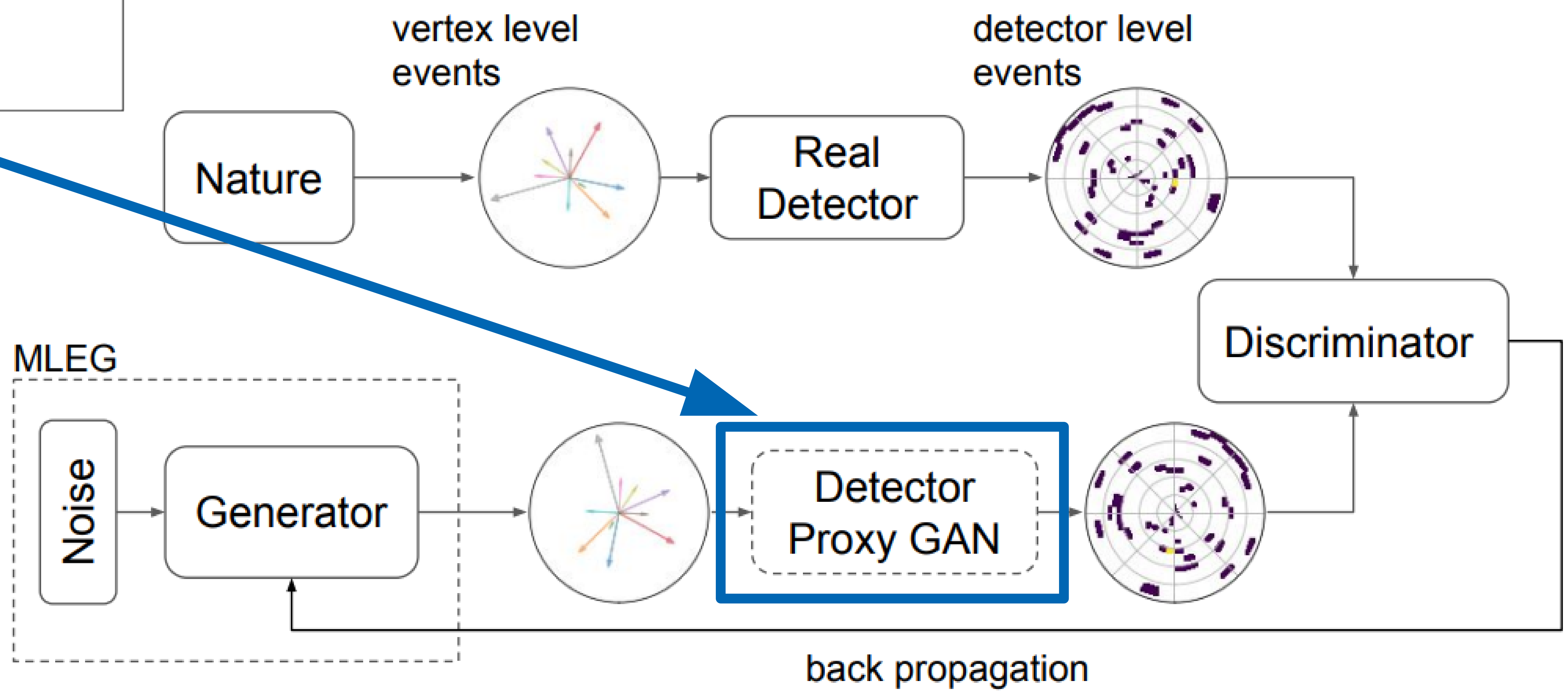
**Conditional GAN** can invert from detector to vertex level by conditioning the training on detector input.
*Bellagente et al., SciPost Phys. 8 (2020) 070*
*Y. Alanazi et al., Phys. Rev. D 106 (2022) 096002*

# Unsupervised vertex learning




Smearing routine for the EIC – mimics detector effects.

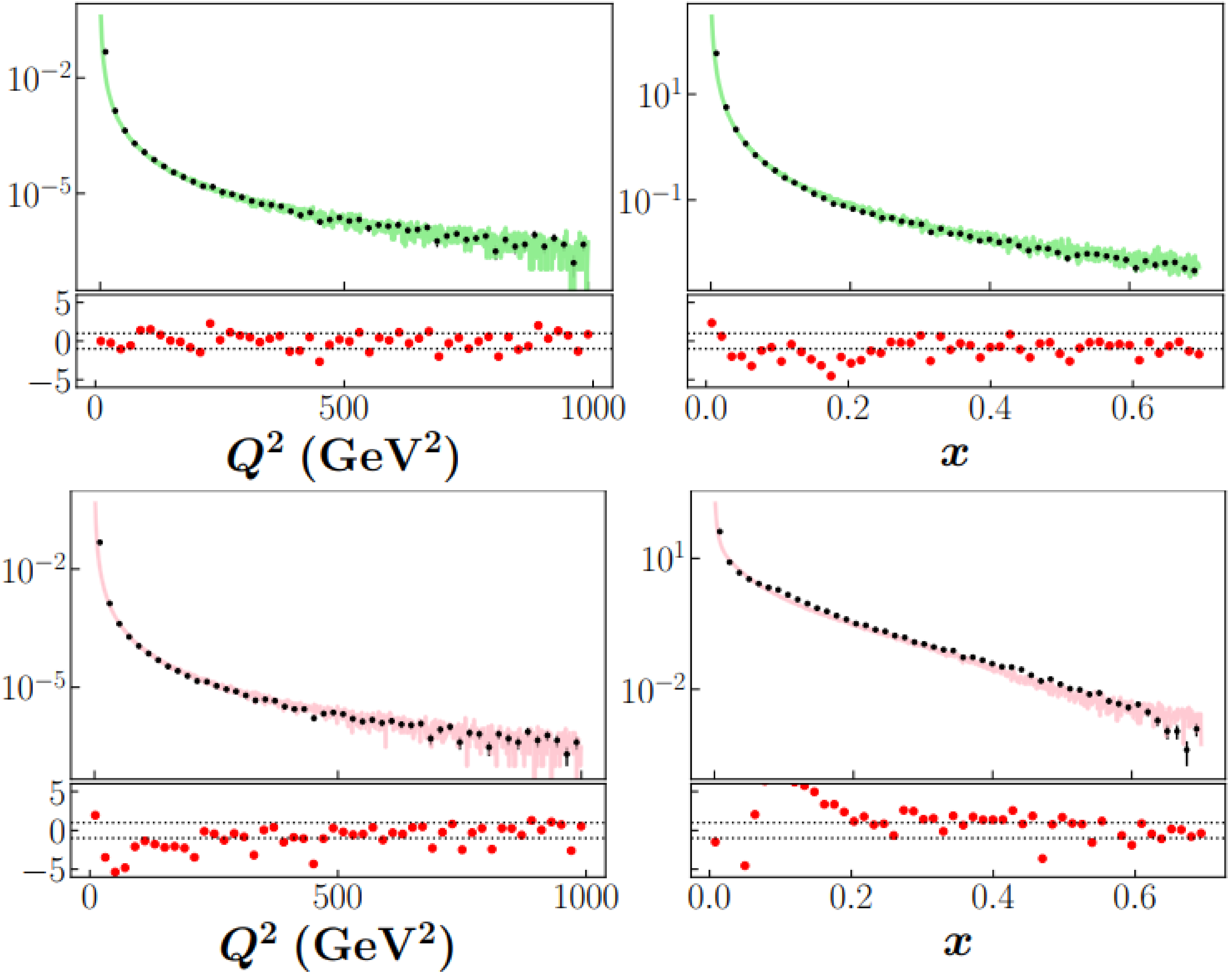
*Y. Alanazi et al., Phys. Rev. D 106 (2022) 096002*




The detector proxy GAN is trained to transform vertex into detector events.

It is then built into the full architecture.

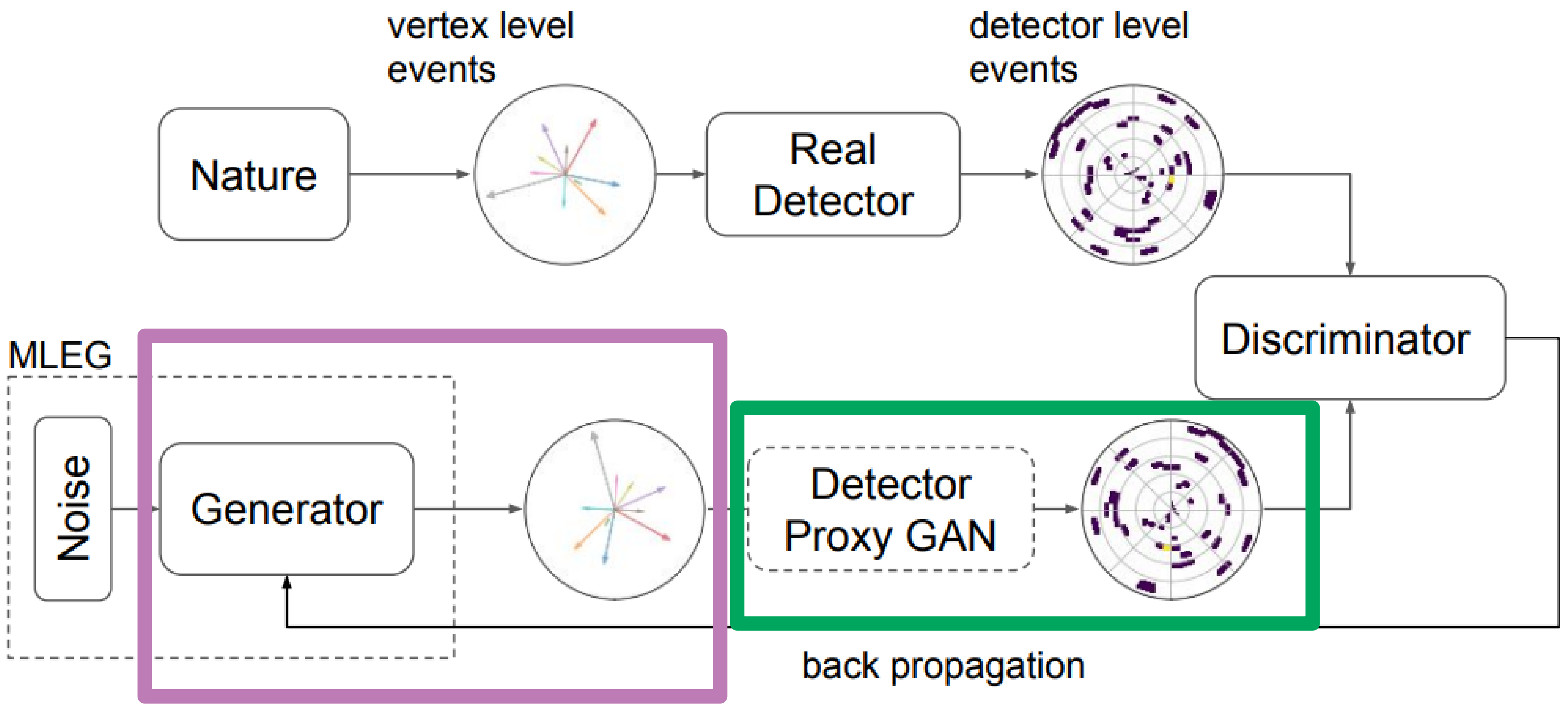
# Closure tests and uncertainty quantification




Training supervised
on **detector-level** samples

Folding detector-level GAN output

$$\text{Pull} = \frac{E_{\text{GAN}} - E_{\text{Data}}}{\sqrt{V_{\text{GAN}} - V_{\text{Data}}}}$$




Training **unsupervised**
on **vertex-level** samples

Unfolded vertex-level GAN ouptut

# Outline

Event generators in modern particle physics

Machine learning event generators:
Goals and methods
GANs

Electron-proton scattering:
FAT conditional GAN
Unfolding
**<u>Training on data</u>**

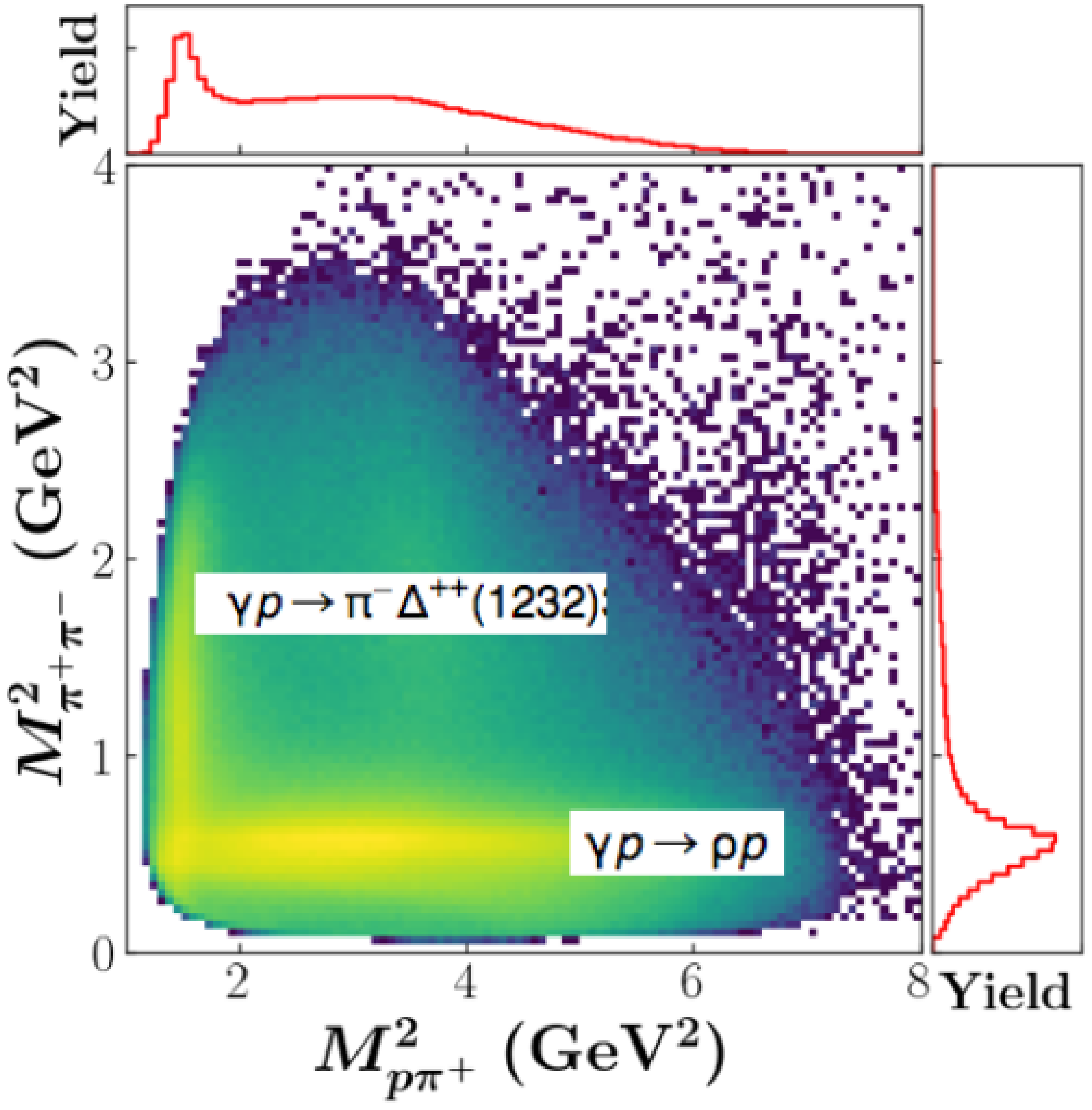
Horizon

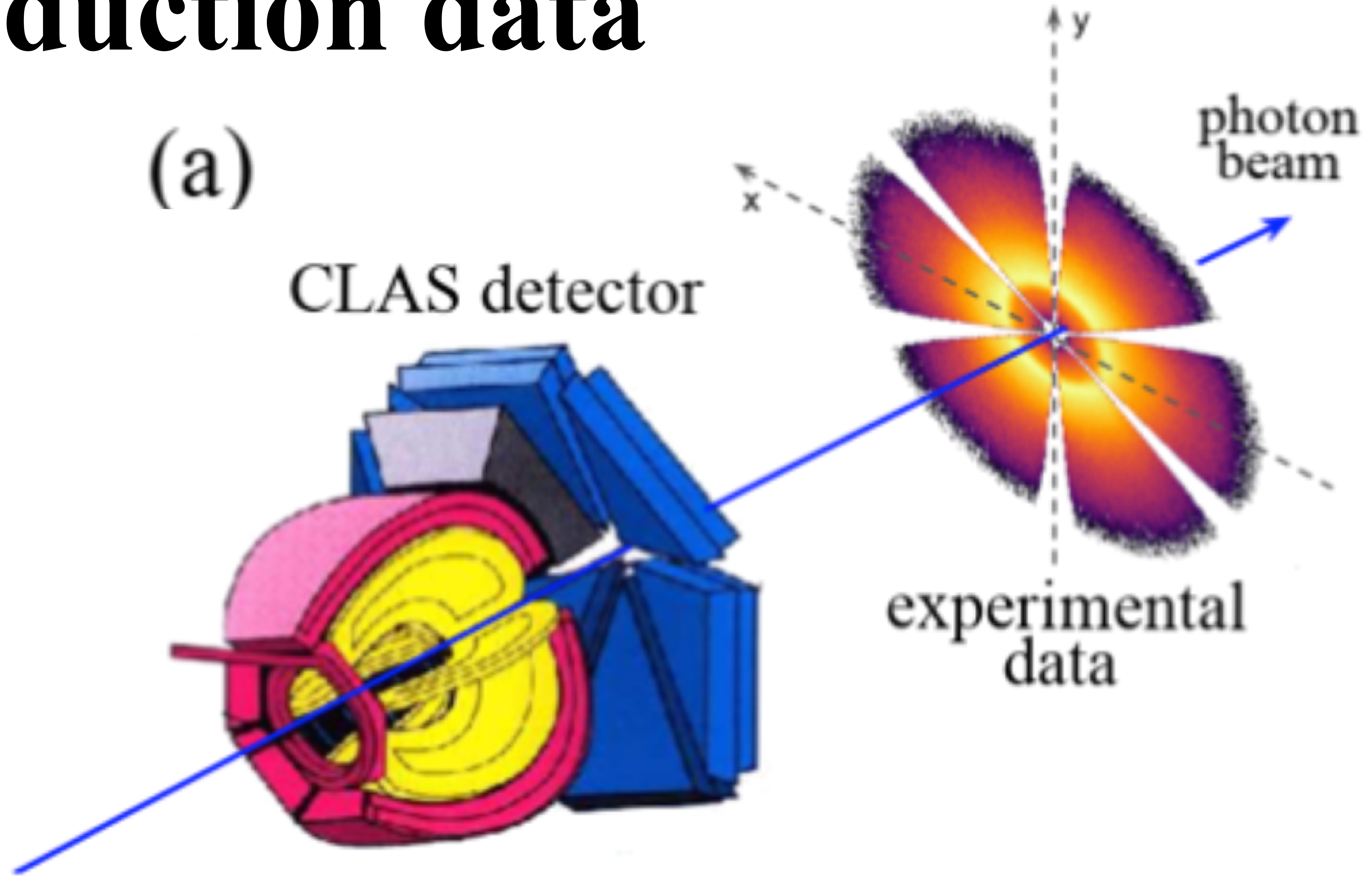
# Two-pion photoproduction data

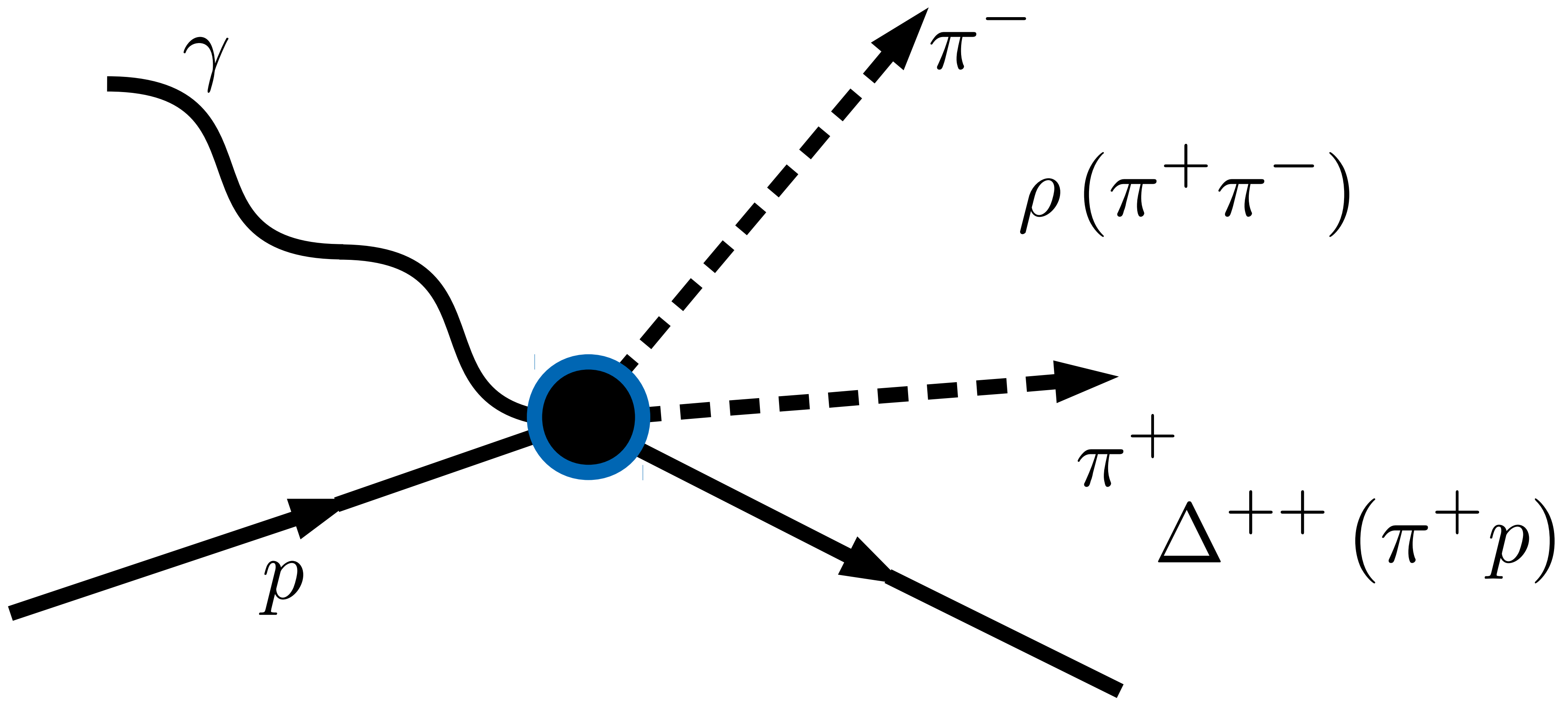
Narrow peaks, holes, steep edges, …

Correlations between variables lost in integrated observables.

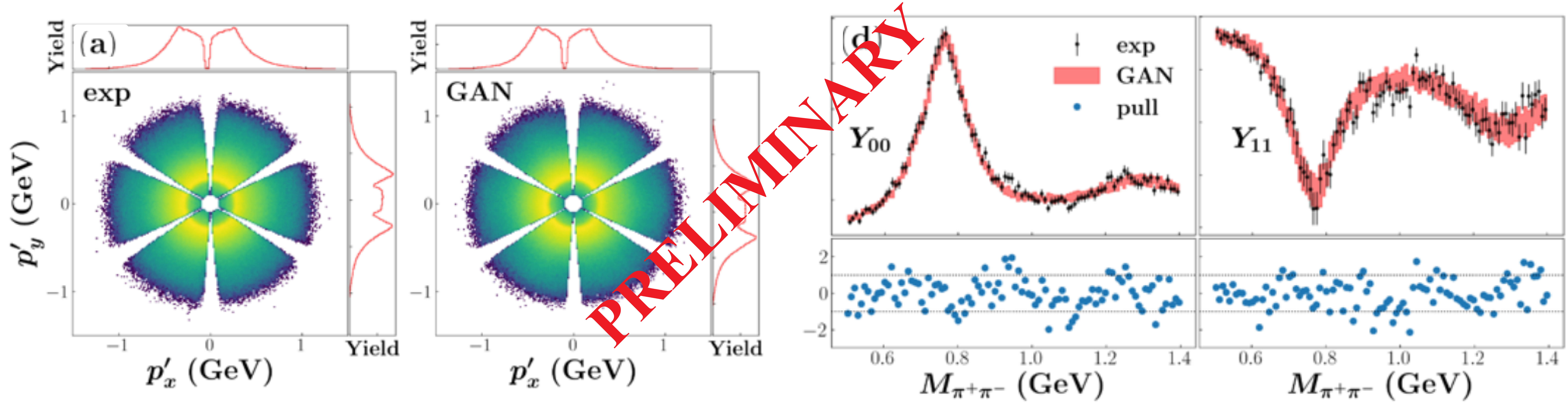
*M. Battaglieri et al. (CLAS Collaboration)*
*Phys. Rev. D 80 (2009) 072005*



# Physics extraction




The GAN mimics the data despite the experimental distortions.

Observables extracted from GAN and data are compatible – **same physics can be extracted!**

**The GAN serves as compact data generator ideal for data storage and distribution.**

# Outline

Event generators in modern particle physics

Machine learning event generators:
Goals and methods
GANs

Applications:
FAT conditional GAN
Unfolding
Training on data

**Horizon**

# Summary and future

Machine learning is a powerful complementary tool for event generation:

– data compression by training on Monte Carlo and real data, **capturing the original correlations;**
– possibility of unfolding **vertex-level physics** from detected data.

Currently available and ongoing works:

– generation of data with **sharp edges** from phase space, detector holes, resonances, …;
– **interpolation** and combination of experiments with different setups;
– **closure tests** of vertex-level extraction, **uncertainty quantification** metrics and **physics validation**.

In future, these benchmarks will allow for:

– **EIC preparation and fast simulators**;
– interpolation and **minimum-bias** physics extraction.