A(I)DAPT
AI for Data Analysis and PreservaTion

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

AI4EIC
October 11, 2022

Outline

A(I)DAPT – goals and methods

Ongoing applications:

Deep inelastic scattering generator – unfolding closure test

CLAS two-pion photoproduction – the non-trivial resonance region

Future extensions

Outline

A(I)DAPT – goals and methods

Ongoing applications:

Deep inelastic scattering generator – unfolding closure test

CLAS two-pion photoproduction – the non-trivial resonance region

Future extensions

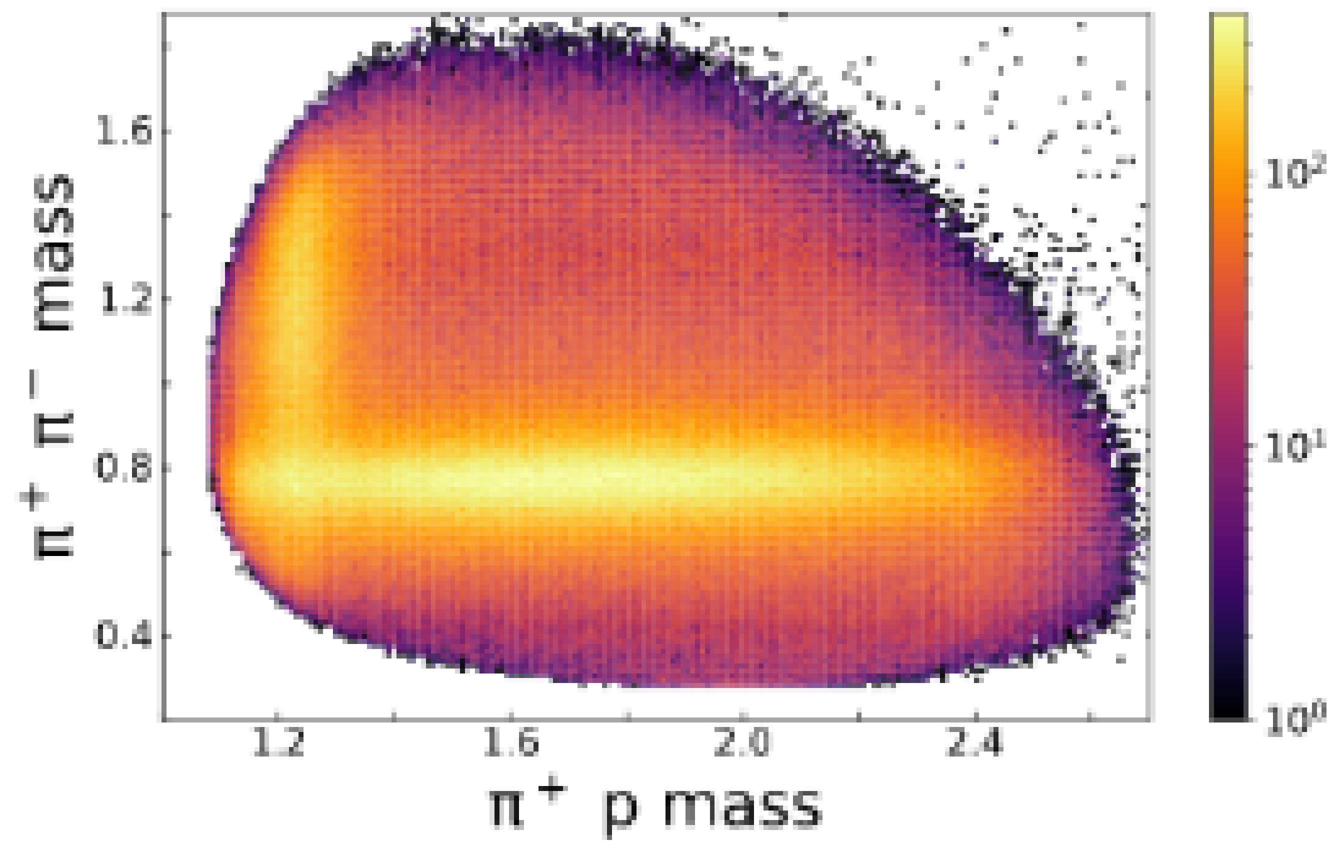
Goals and challenges of modern data analysis

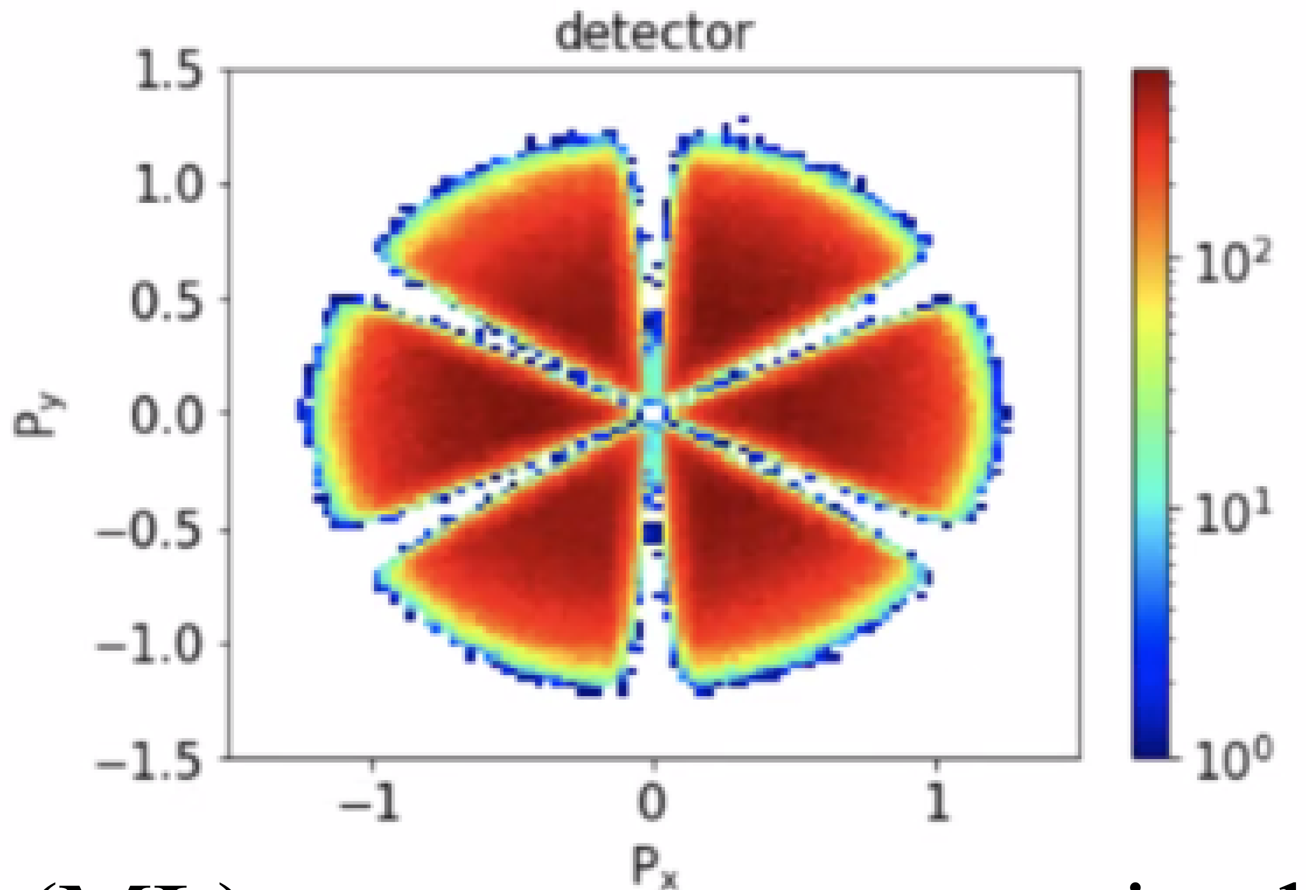
The data analysis of state-of-the-art particle experiments **(JLab, EIC, …)** is challenged by:

– **high dimensionality**, multi-particle final states, multi-variable correlations lost in integration

– **model dependence**, theory assumptions needed in e.g. Monte Carlo (MC) generators

– **experimental apparatus effects**, limited acceptance and resolution (peaks may remain unseen, kinematics washed out to unphysical regions, …).





One way to address these challenges is with machine learning (ML) event generators trained on experimental data:

Y. Alanazi et al., 2106.00643, AI conference journal IJCAI, 2021

– after training, they are **extremely fast** (millions of events per second)

– they serve as **data compression tools** (order of MB vs. experimental data's GB/TB)

– any statistics can be generated: **minimum-bias interpolators**

– they can be trained to generate vertex-level **events free of (unfolded from) detector-effects.**

A(I)DAPT in a nutshell

Collaboration of experimental physicists, theorists, and computer scientists,
centered around Jefferson Lab physics.

Y. Alanazi, T. Alghamdi, P. Ambrozewicz, G. Costantini, A. Hiller Blin, E. Isupov, T. Jeske, Y. Li, L. Marsicano, W. Melnitchouk, V. Mokeev, N. Sato, A. Szczepaniak, T. Viducic

Goals:

Creation of ML event generators for **data compression** and as powerful **interpolation tools**.

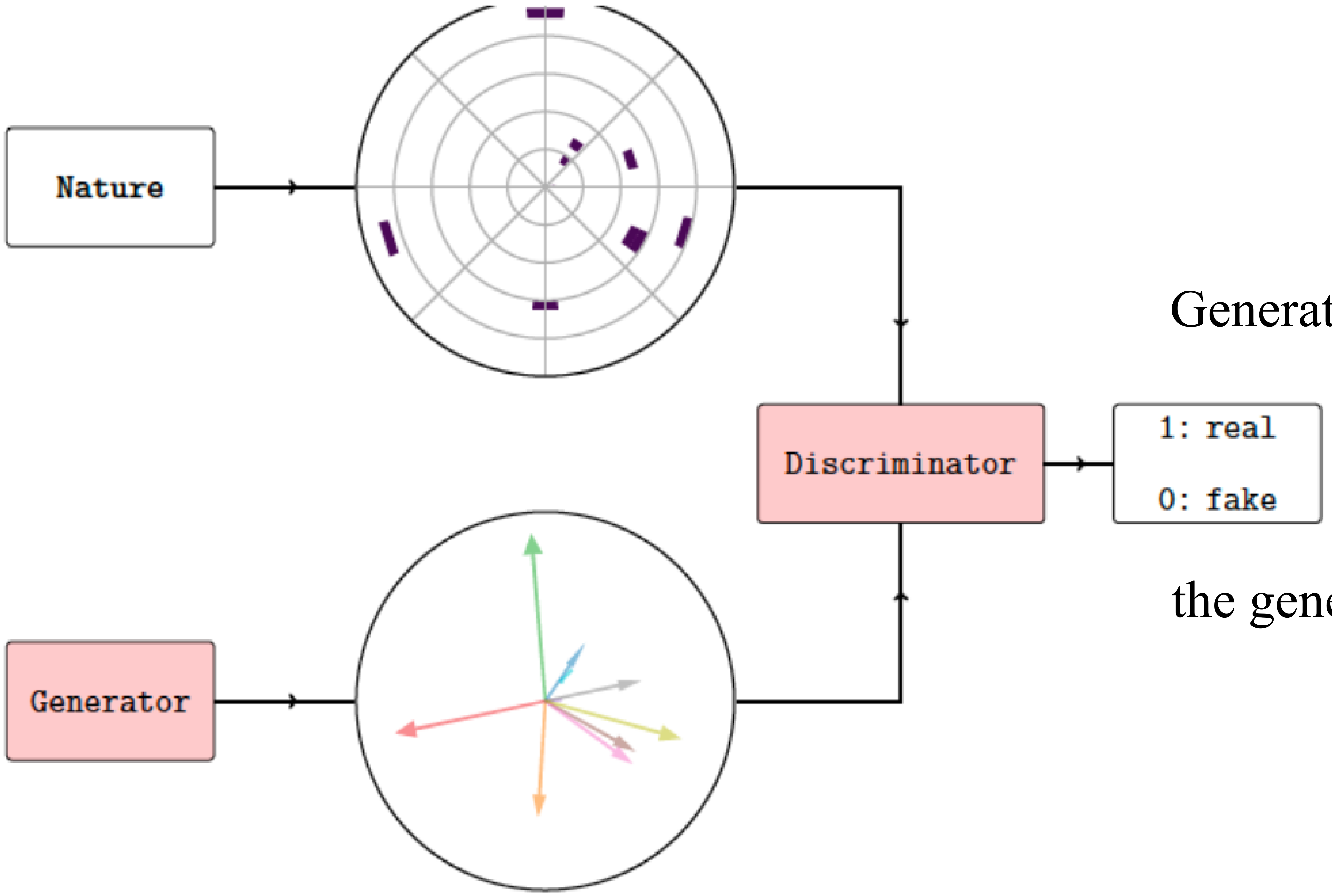
Unfolding of detector effects in order to get the true vertex-level data.

Closure tests, physics validation (kinematics/final-state universality; physics extraction from real/generated data), **uncertainty quantification** (statistical bootstraps, systematics in training data sets and reference frames).

Generative adversarial networks (GANs)

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

Generator is 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**.

Outline

A(I)DAPT – goals and methods

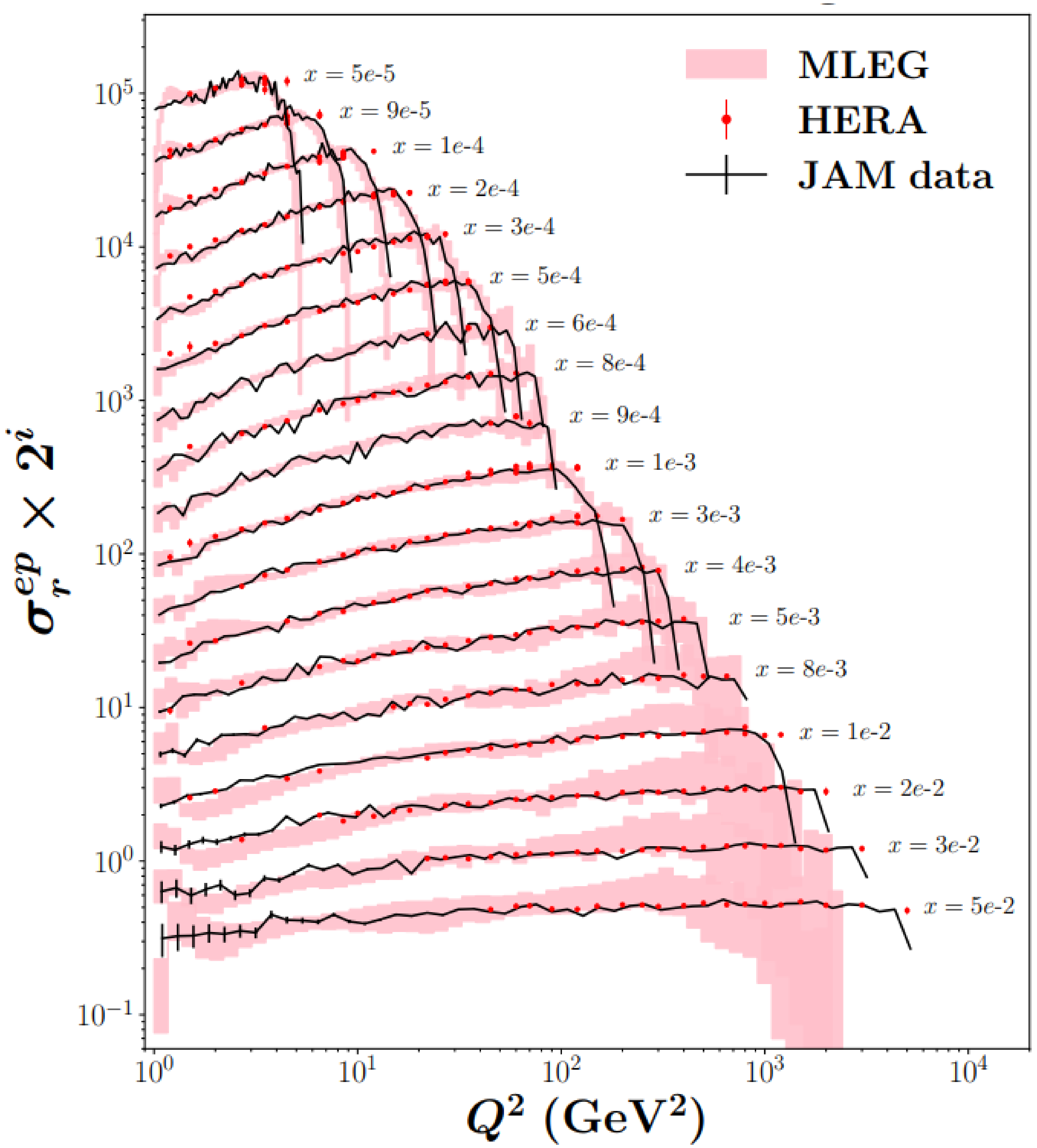
Ongoing applications:

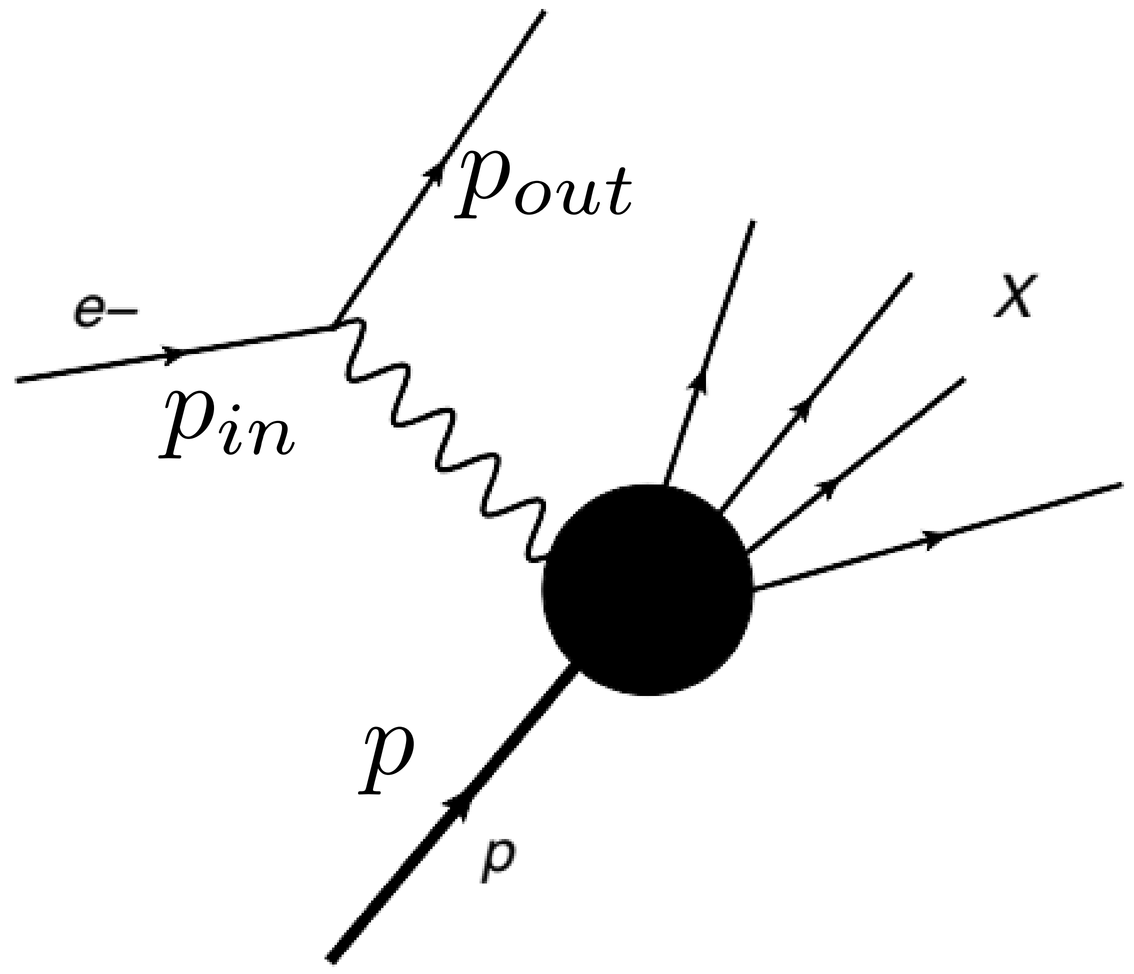
Deep inelastic scattering generator – unfolding closure test

CLAS two-pion photoproduction – the non-trivial resonance region

Future extensions

Training on variables of deep inelastic scattering (DIS)

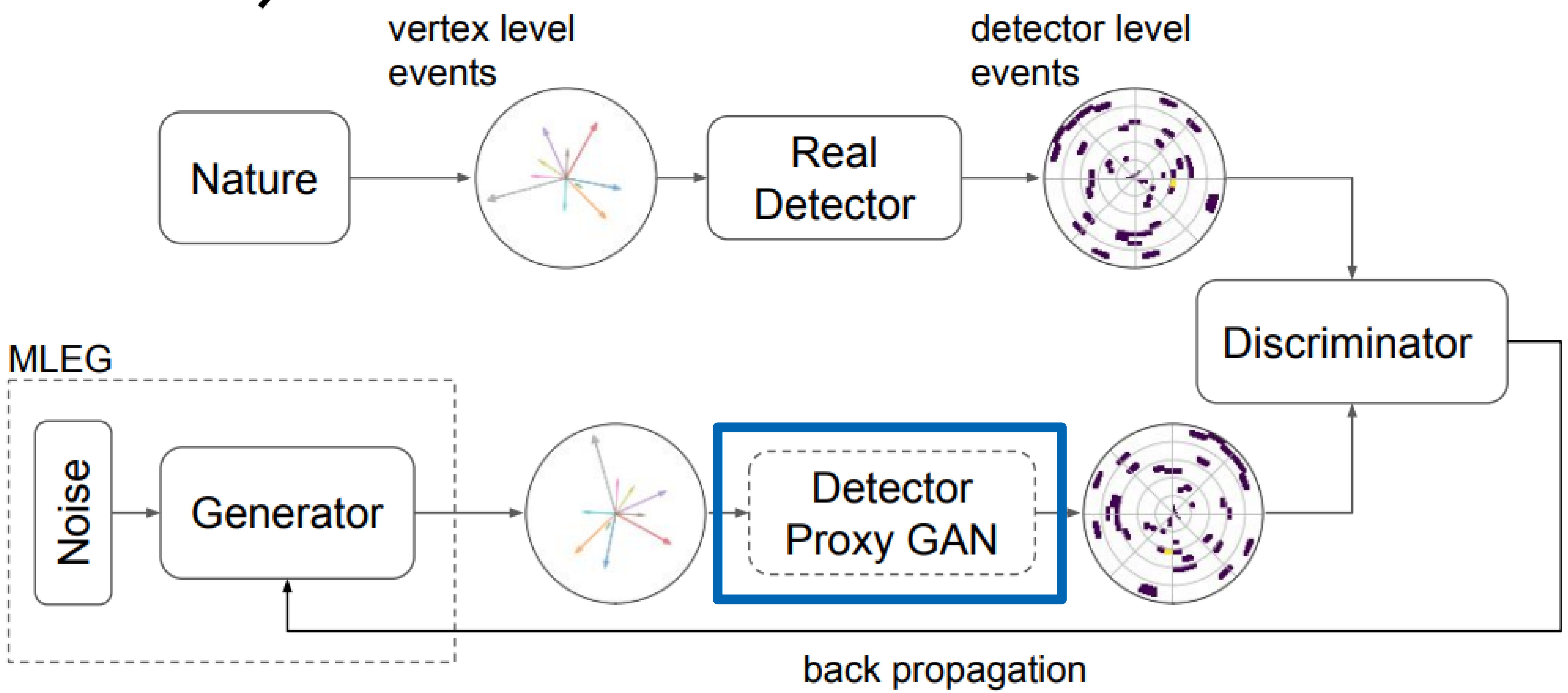




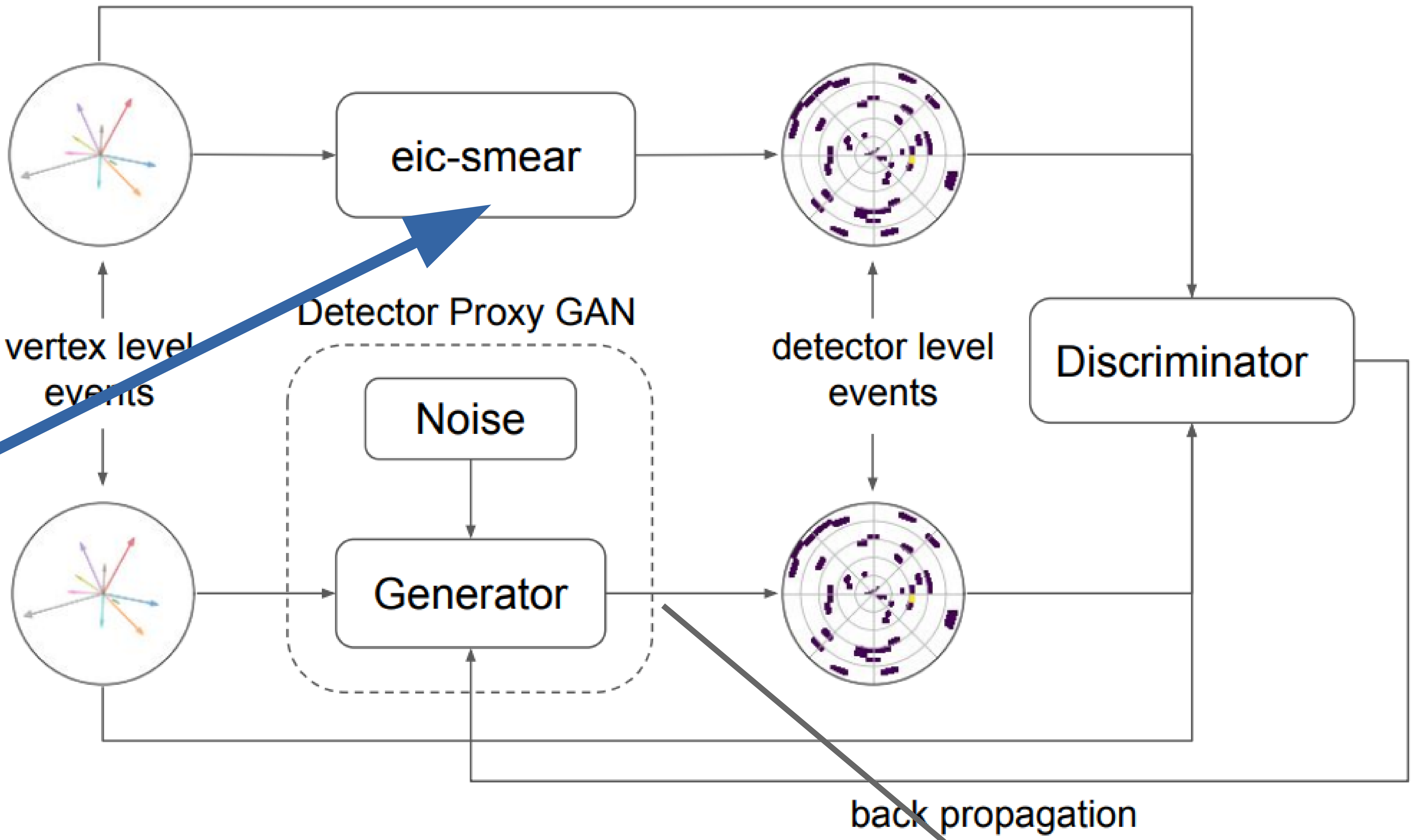
$$Q^2 = -(p_{in} - p_{out})^2$$

$$x = \frac{Q^2}{2p \cdot (p_{in} - p_{out})}$$

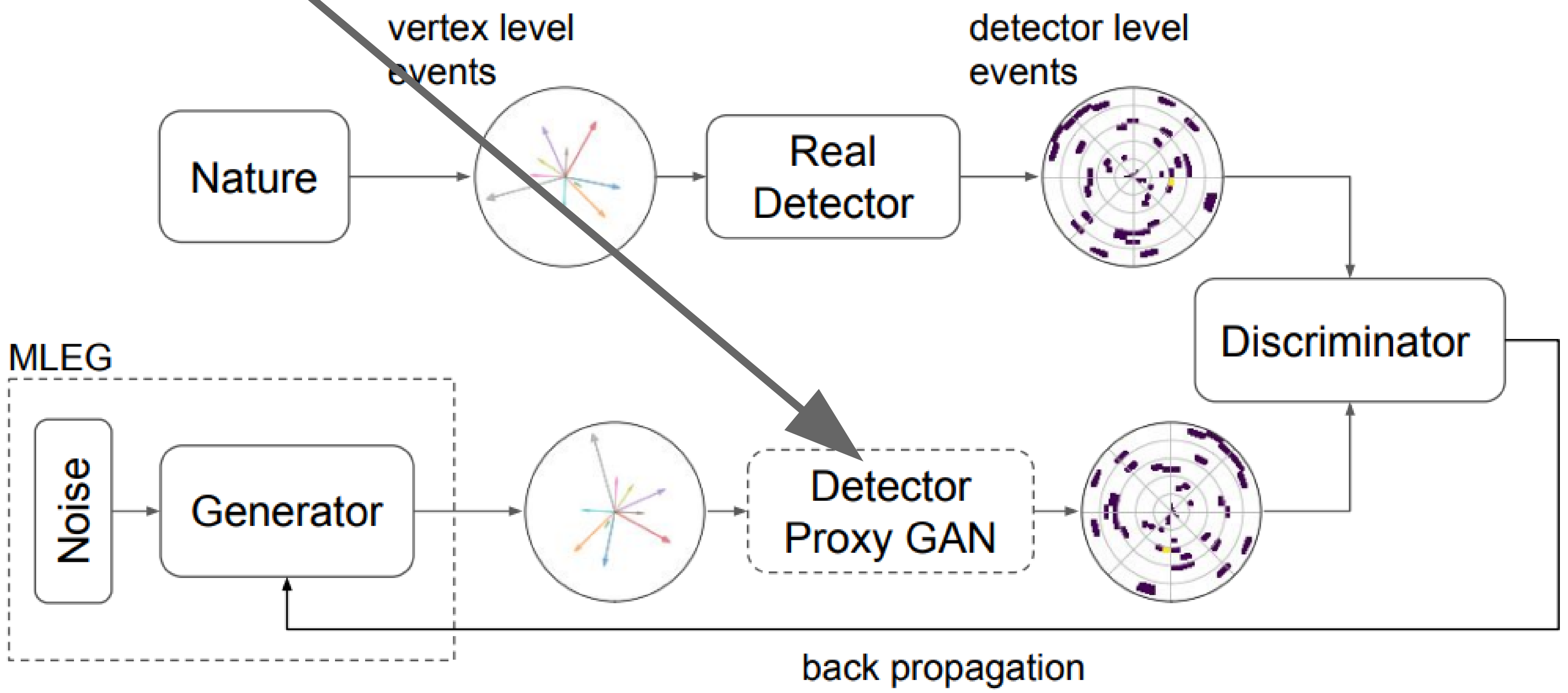
Y. Alanazi et al., 2008.03151



Unfolding – training the detector proxy GAN



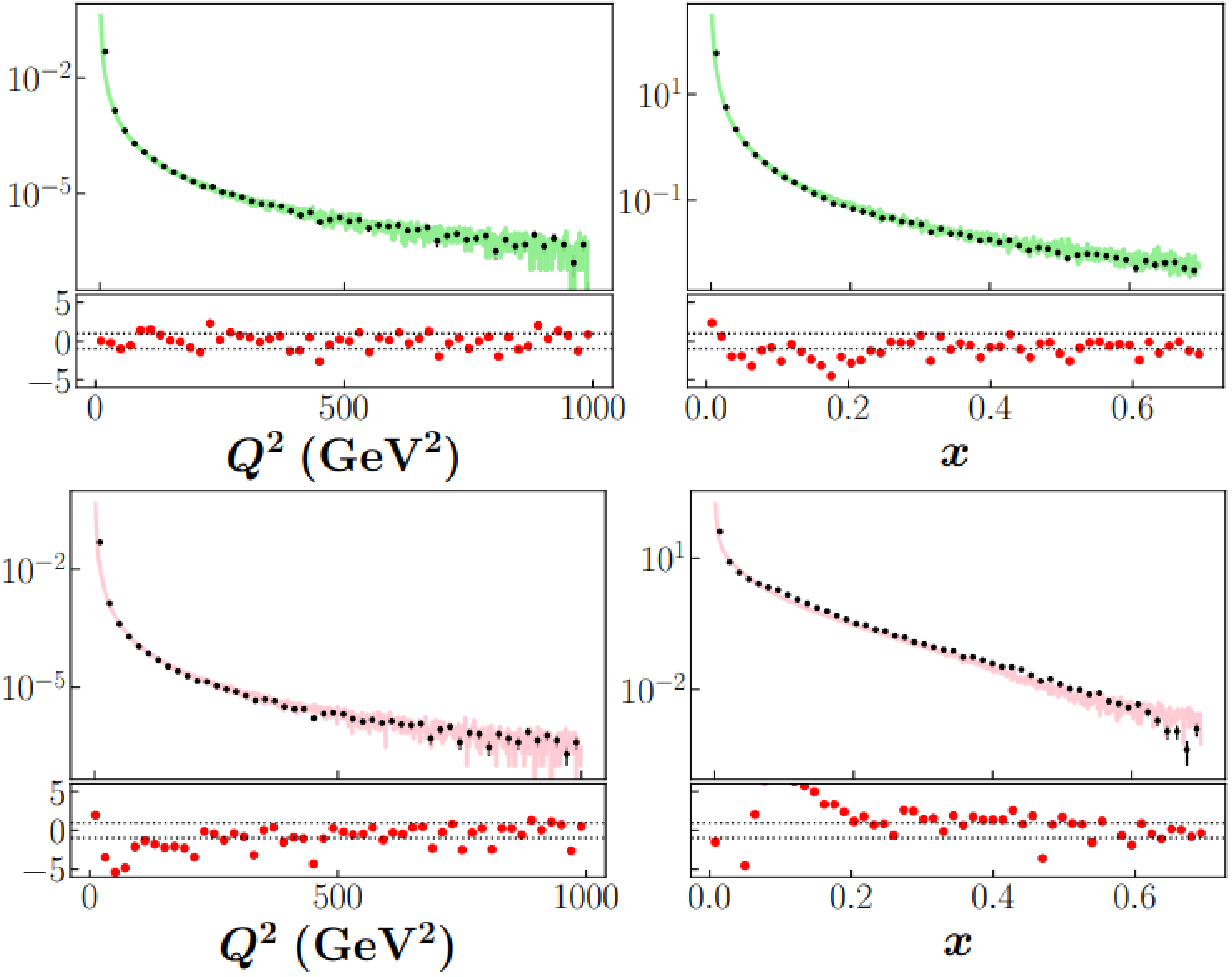
Smearing routine for the EIC – mimics detector effects.



The detector proxy GAN is trained to act as simulator transforming vertex into detector events.

It is then built into the full architecture.

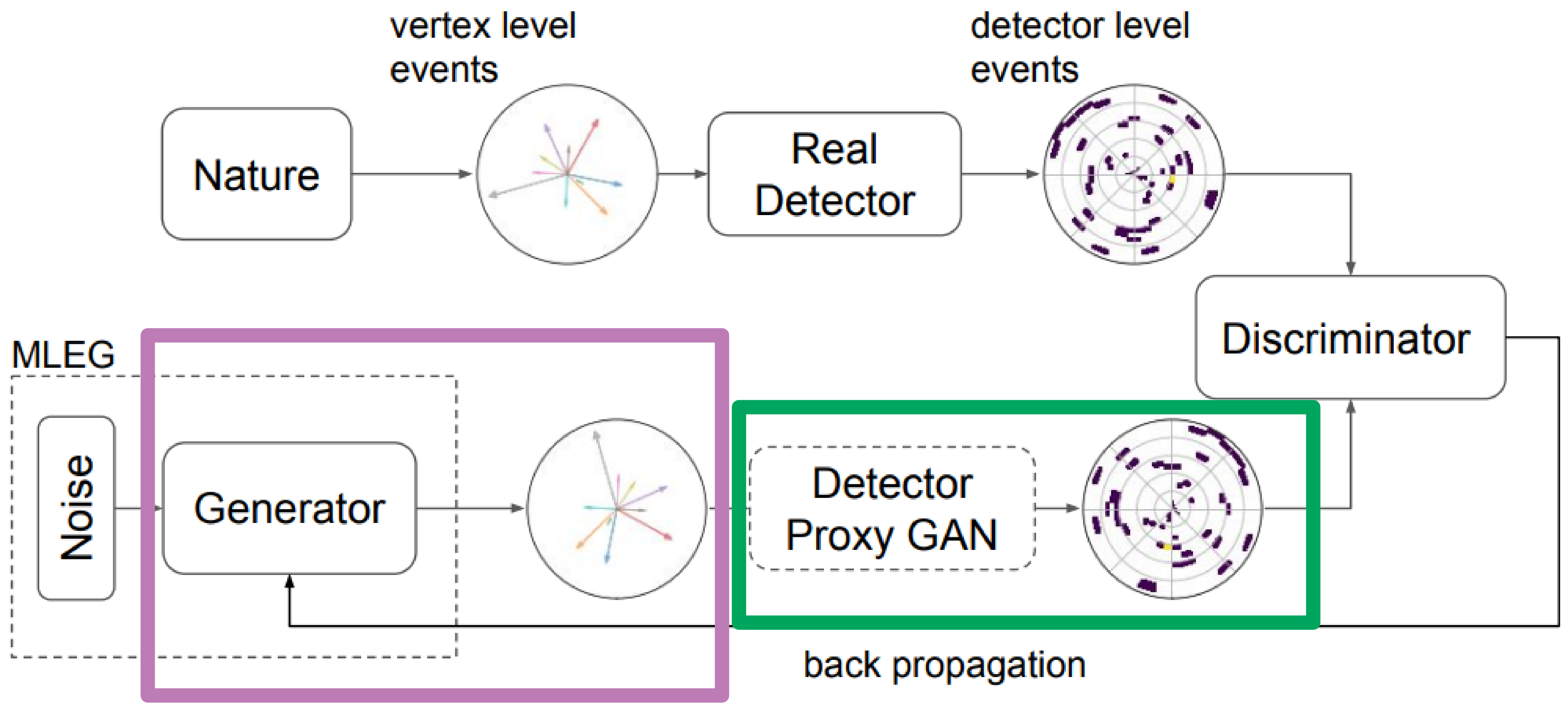
Closure tests



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

Recap on unfolding closure test applied to DIS

Y. Alanazi et al., 2008.03151

Experimental apparatus effects distort nature into detector-level events.

Limited acceptance and resolution:
– need to constrain to fiducial volumes when training;
– unfolding is a mitigation strategy to reconstruct vertex-level events.

Our closure tests demonstrated the validity of our folding GAN architecture to obtain unfolded events:
controlled uncertainty quantification!

Outline

A(I)DAPT – goals and methods

Ongoing applications:

Deep inelastic scattering generator – unfolding closure test

CLAS two-pion photoproduction – the non-trivial resonance region

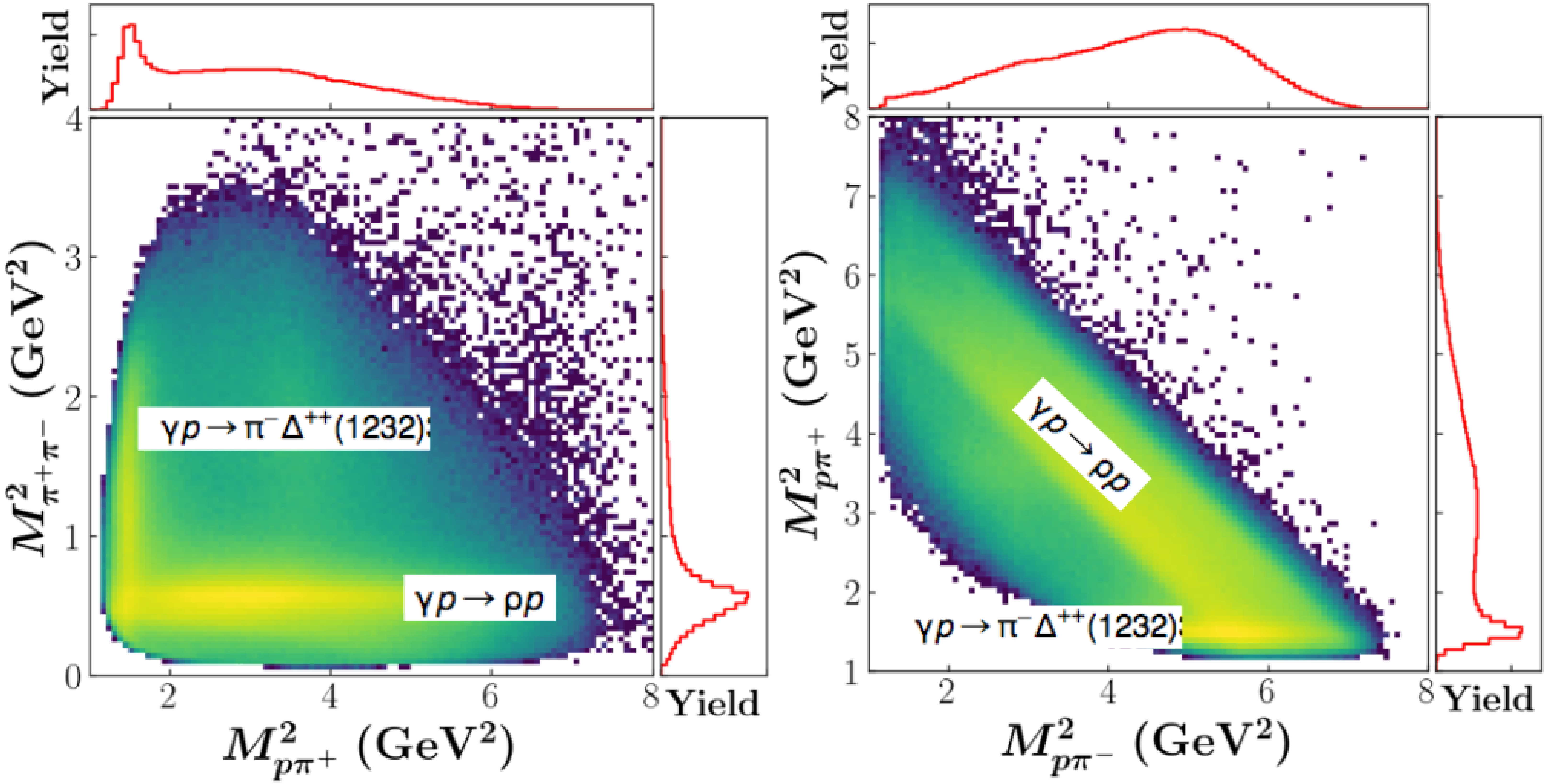
Future extensions

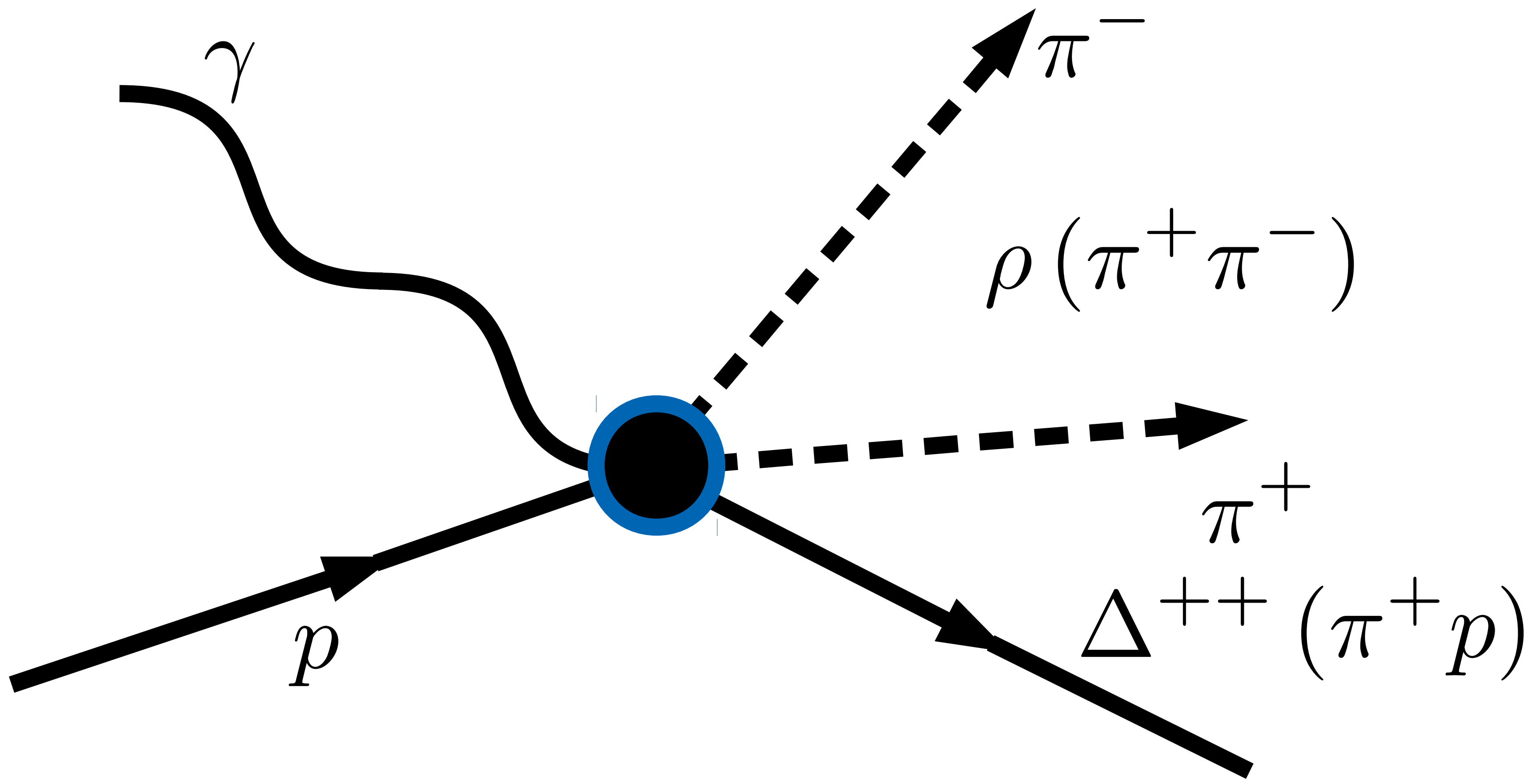
Exclusive two-pion photoproduction

The CLAS data with photon beam energies of 3-4 GeV display highly non-trivial distributions: narrow peaks, holes, steep edges, …

Correlations between variables lost in integrated observables – ML for event-by-event physics interpretation.

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

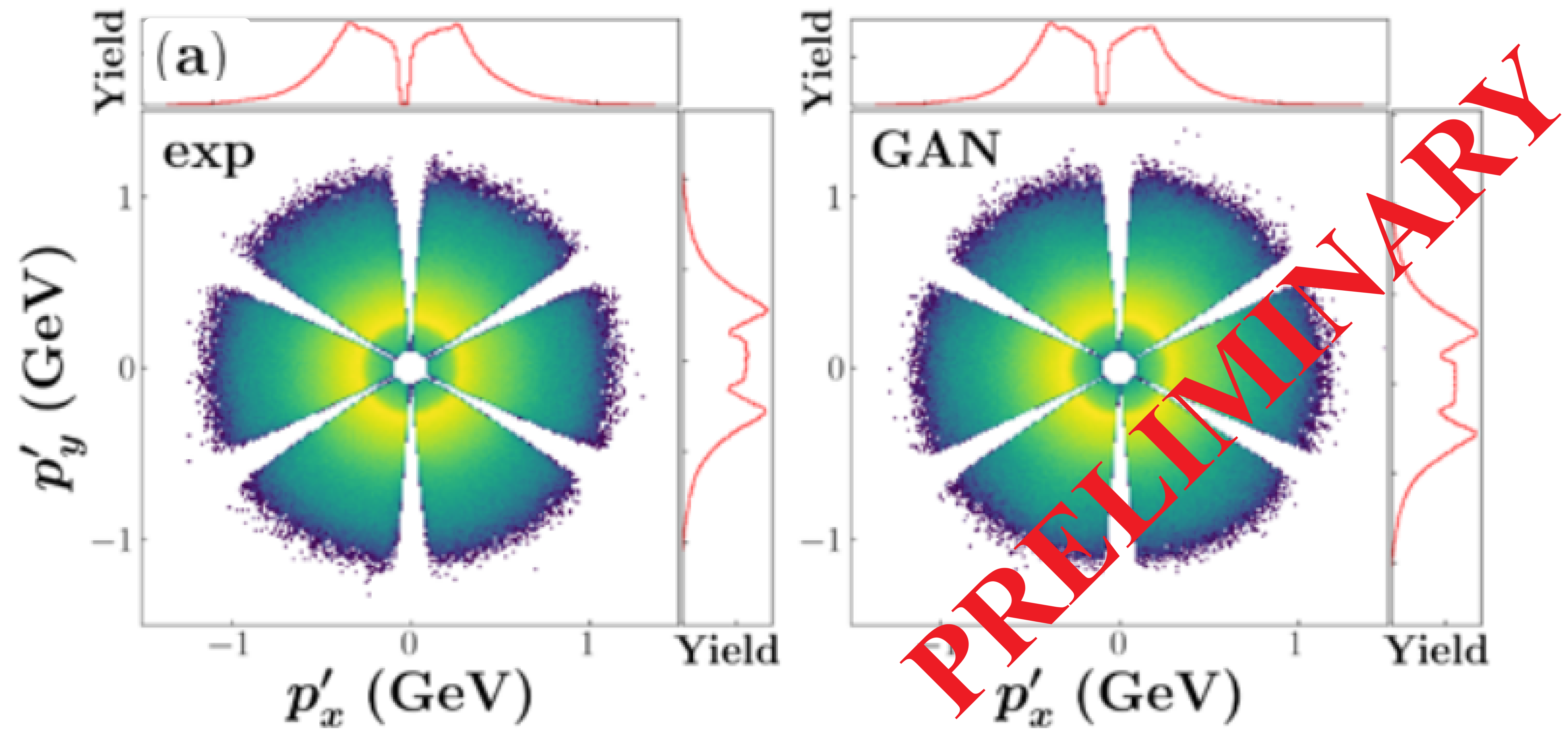


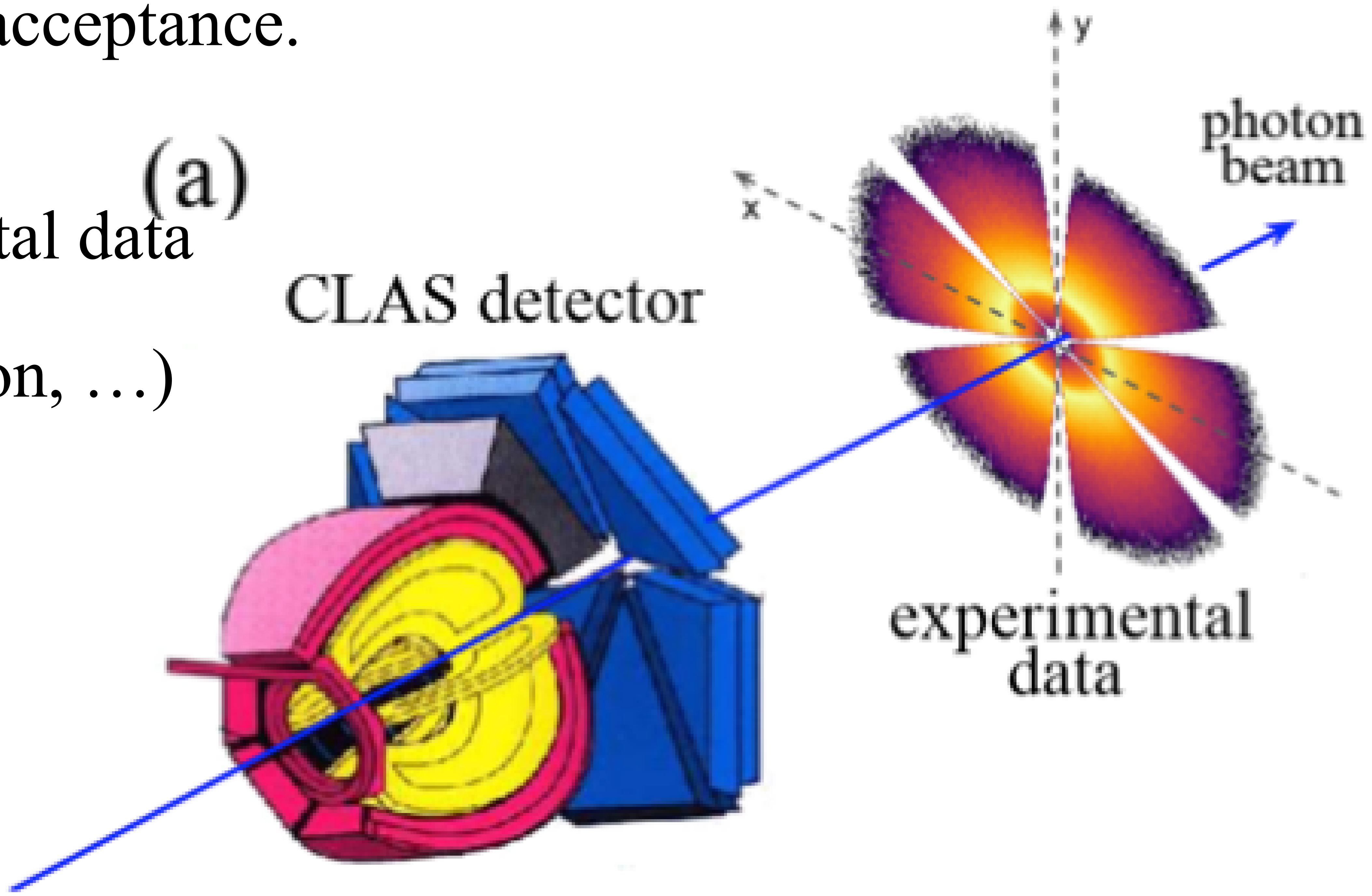


Experimental perks

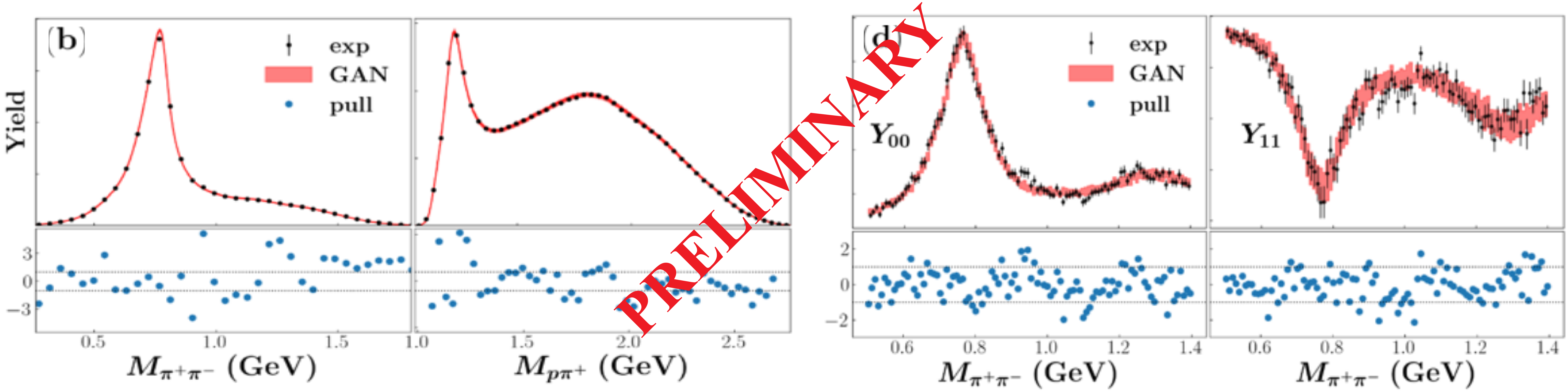
Distortions appear mainly due to the detector acceptance.

Aim is to train a GAN that is:
– equivalent to (displays same correlations as) experimental data
– feature augmented and transformed (FAT)
to obey physical laws (energy and momentum conservation, …)





Physics extraction



The GAN mimics the data despite the experimental distortions.

Even Y_{LM} moments extracted from GAN and data are compatible – same physics can be extracted!

Outline

A(I)DAPT – goals and methods

Ongoing applications:

Deep inelastic scattering generator – unfolding closure test

CLAS two-pion photoproduction – the non-trivial resonance region

Future extensions

Summary and future

We created GAN event generators for:

DIS — closure test of an unfolding architecture for the first time;
CLAS data — captured intricate detector acceptance behavior and multi-particle correlations.

We are currently analysing:

– unfolding of CLAS data to extract physics at vertex level;
– information from hidden layers;
– uncertainty quantification metrics and physics validation.

In future, these benchmarks will allow for

– powerful minimum-bias interpolation tools;
– physics extraction and amplitude analysis from synthetic vertex-level data.