Generative modelling for CLAS analysis

Marco Spreafico

on behalf of A(i)DAPT Working Group

A(i)DAPT

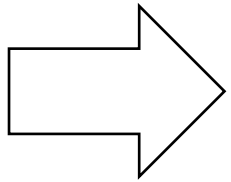
AI for Data Analysis and PreservaTion

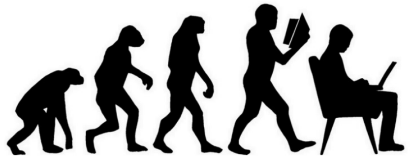
A(i)DAPT motivation

NP/HEP experiments are affected by the following problems:

- Data are affected by detector effects
- Multidimensional problems reduced in lower-dimensional spaces for simplicity
- Large datasets are difficult to manipulate and preserve

Can AI support NP/HEP experiments and allow to extract physics form data in a more efficient way?





A(i)DAPT

AI for Data Analysis and PreservaTion

Develop AI-supported procedures to:

- Unfold detector effects
- Accurately fit data in multiD space
- Preserve data in a more compact format

Collaborative effort:

- ML experts (ODU, JLab)
- Experimentalists (JLab Hall-B)
- Theorists (JPAC, JAM)

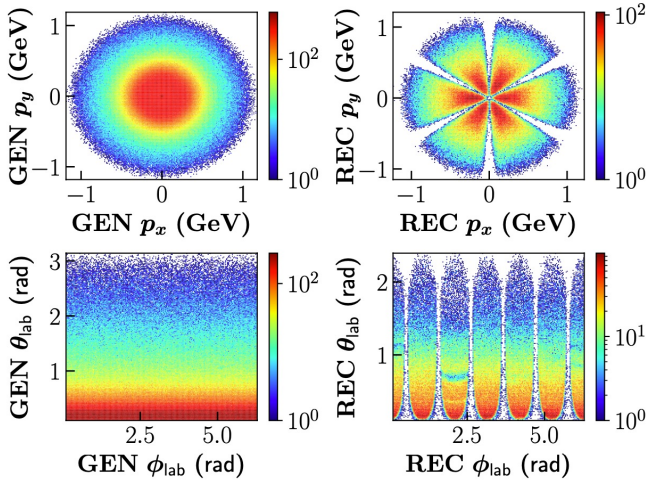
A(i)DAPT motivation

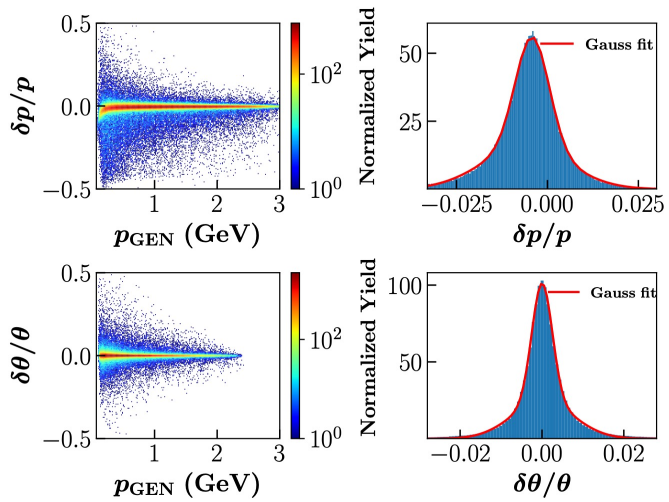
DETECTOR EFFECTS

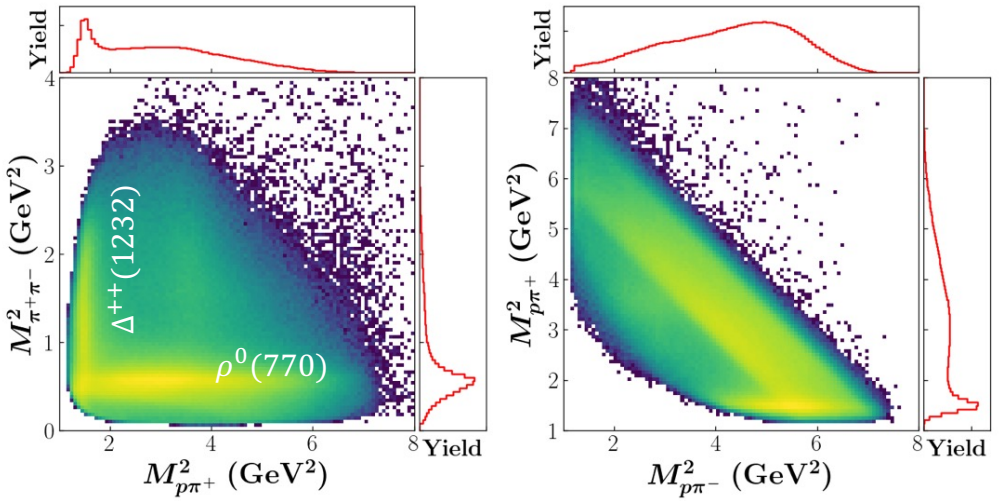
- **Acceptance**: measurements only access a limited portion of phase space
- **Resolution**: limited resolution may lead to interesting effects being washed out

DIMENSIONALITY

When reducing a high-dimensionality problem to a lower dimensionality, resonances may become less clear



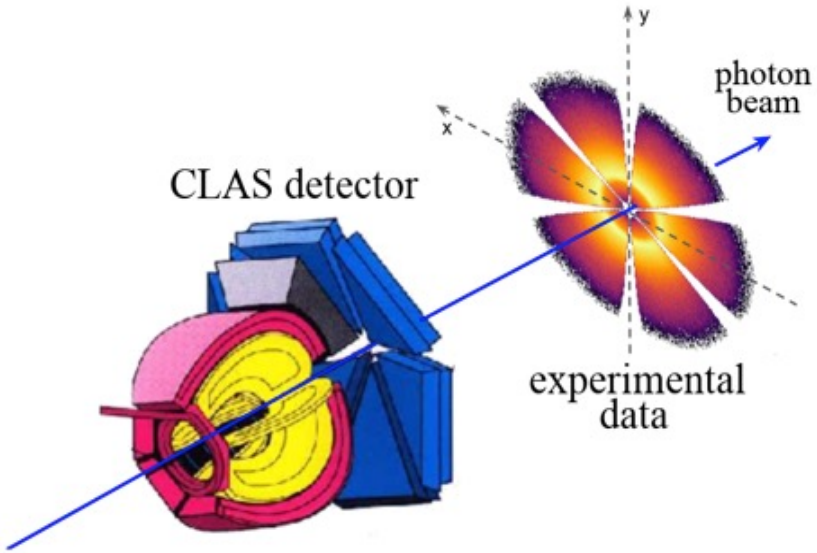




AI could provide a new way to look at data and to extract observables and physics interpretation

Exclusive 2π photoproduction

M. Battaglieri *et al.* (CLAS Collaboration) Phys. Rev. Lett. 102, 102001
M. Battaglieri *et al.* (CLAS Collaboration) Phys. Rev. D 80, 072005



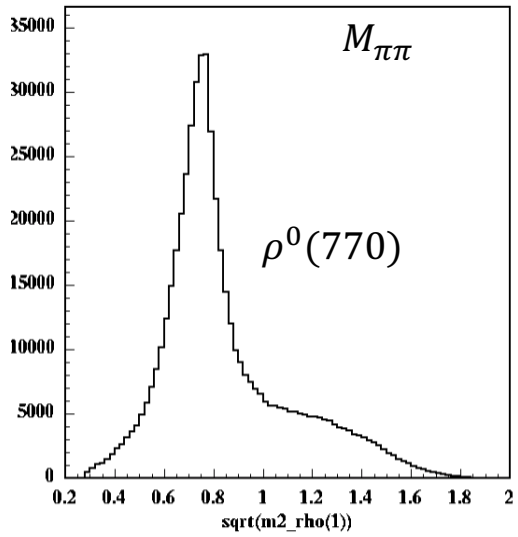
CLAS g11 kinematics

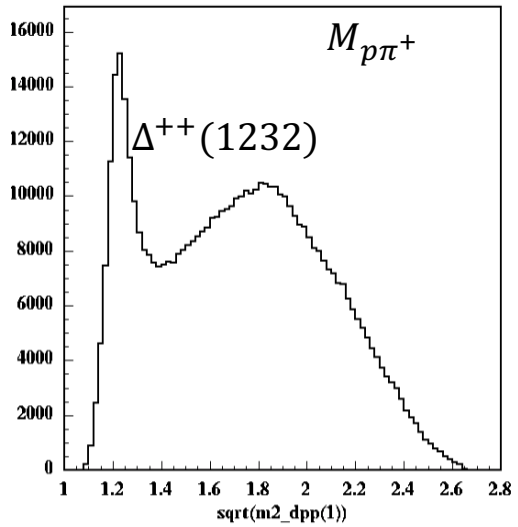
Focus on $\gamma p \to p\pi^+(\pi^-)$

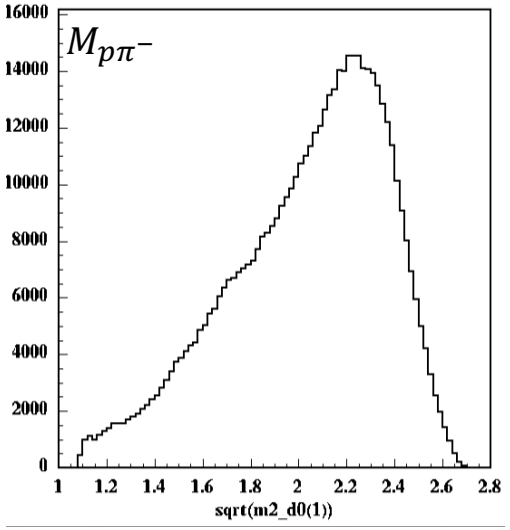
- $E_\gamma = (3-4)$ GeV

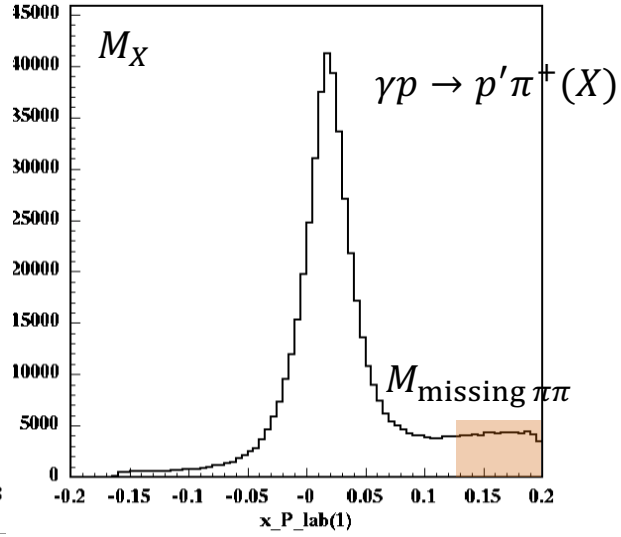
ρ^0 electroproduction: $\gamma p \to p\rho^0$
Δ^{++} resonance excitation: $\gamma p \to \Delta^{++}\pi^-$

- Missing π^- measured as missing momentum
- Multi π background from $\gamma p \to p\omega \to p\pi^+\pi^-\pi^0$

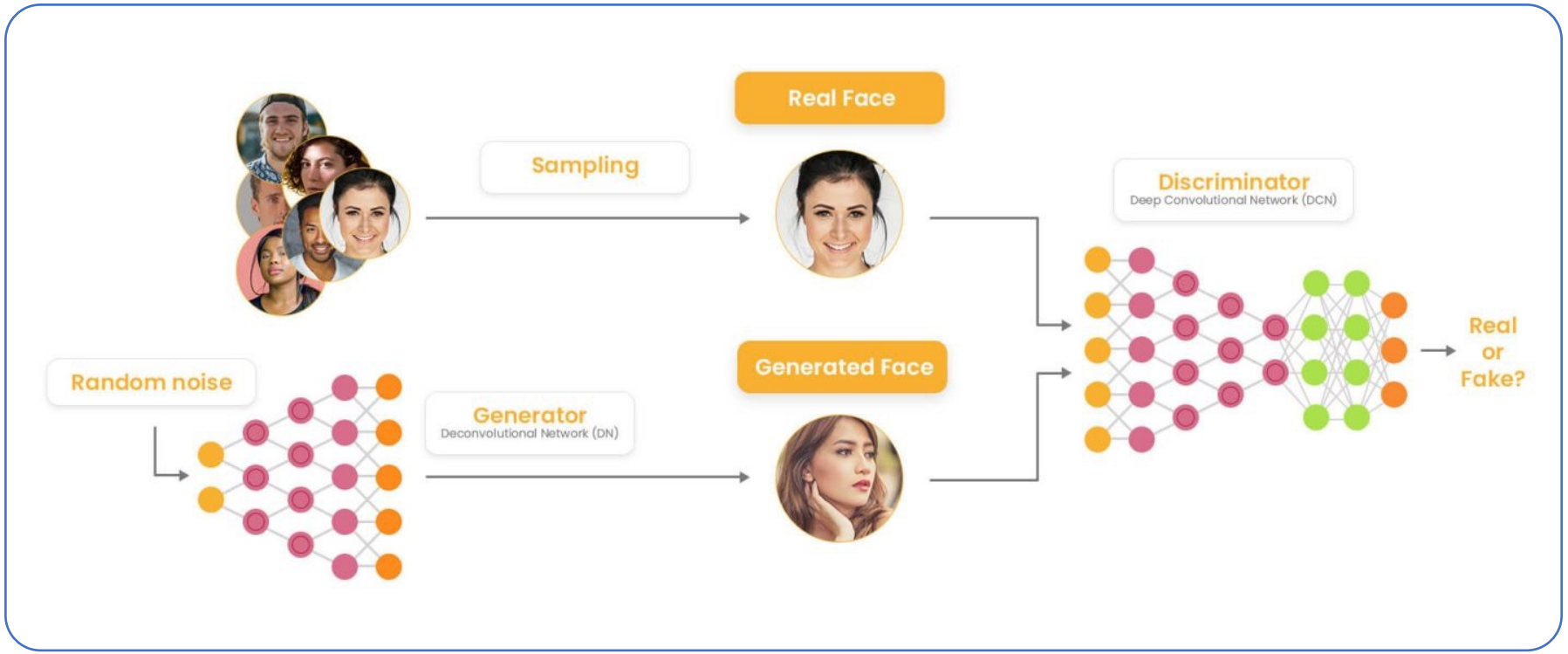








Generative Adversarial Networks (GAN)



Generator

Produce synthetic data that mimic real data

- Can retain high dimensional correlations
- Can produce realistic pseudo-data quickly

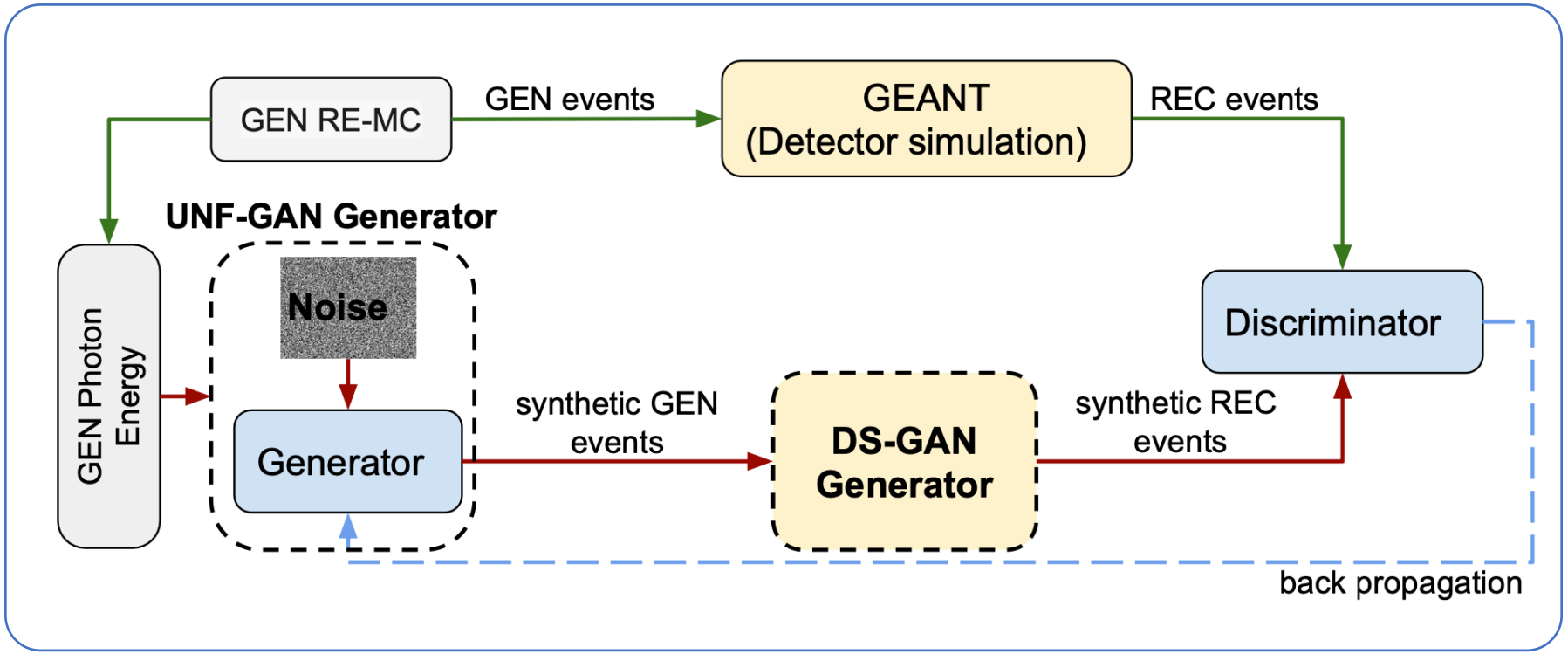
VS

Discriminator

Distinguish between synthetic and real data
Forces generator to generate synthetic data indistinguishable from real data

Credit: T.Alghamdi et al, □□□□□□□□D 108, 094030

2π photoproduction closure test

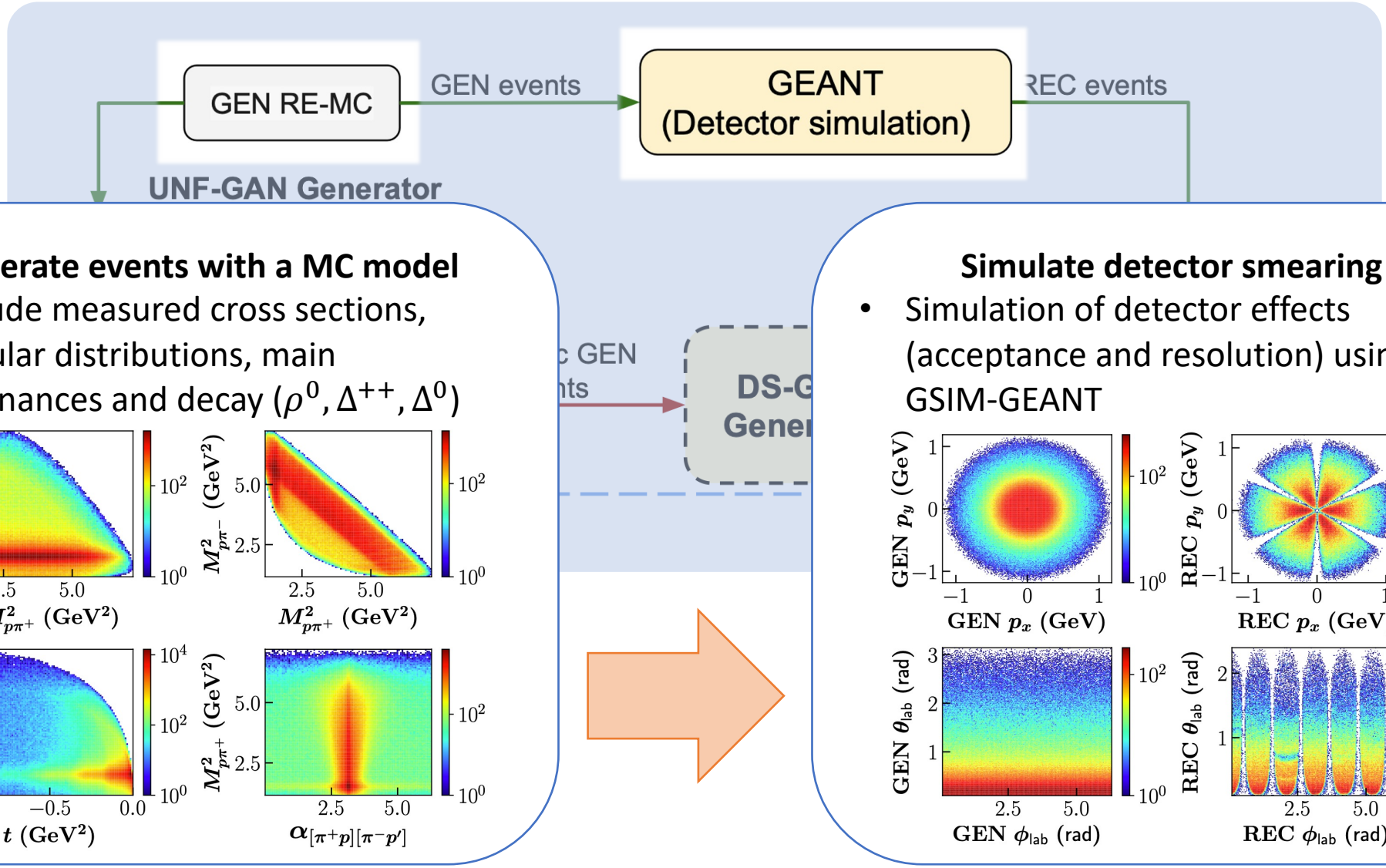


CLOSURE TEST

Can an AI-based model learn underlying features of (pseudo-)data while unfolding detector effects?

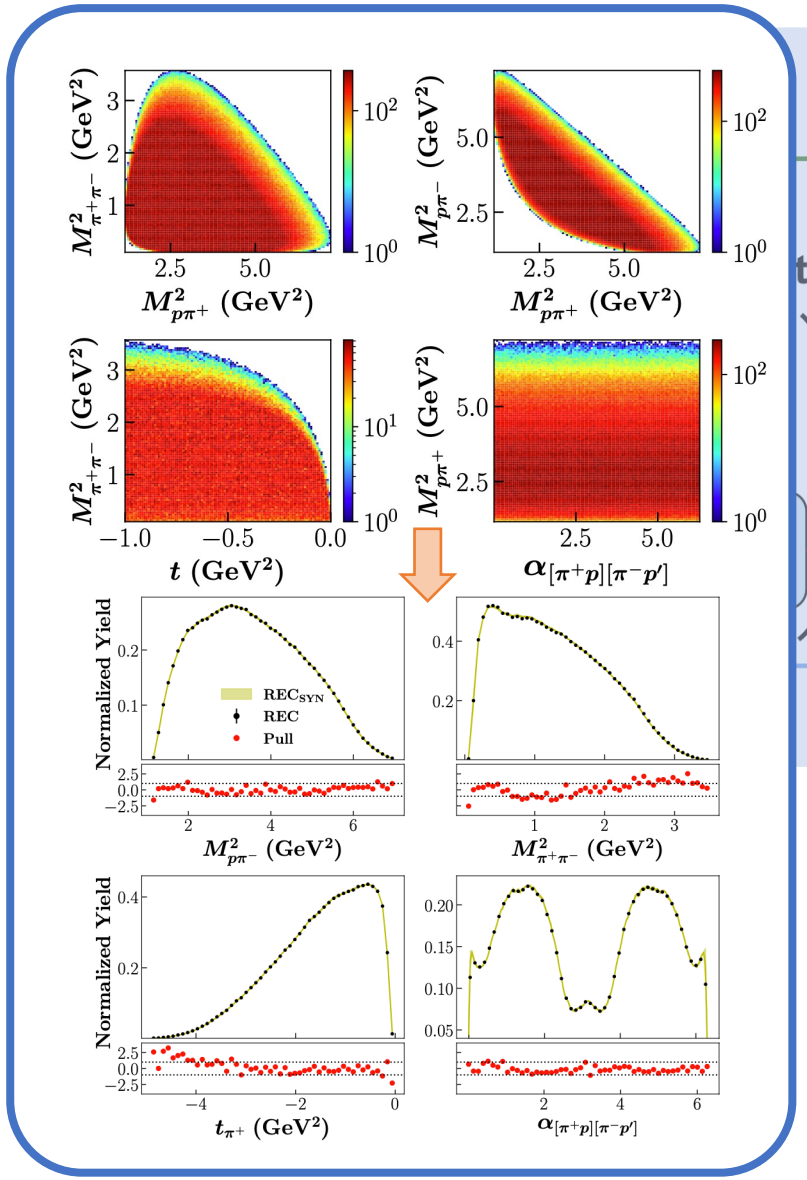
Credit: T.Alghamdi et al, □□□□□□□D 108, 094030

2π photoproduction closure test



Credit: T.Alghamdi et al, [redacted] D 108, 094030

2π photoproduction closure test



GEN events

(De

tor

synthetic GEN events

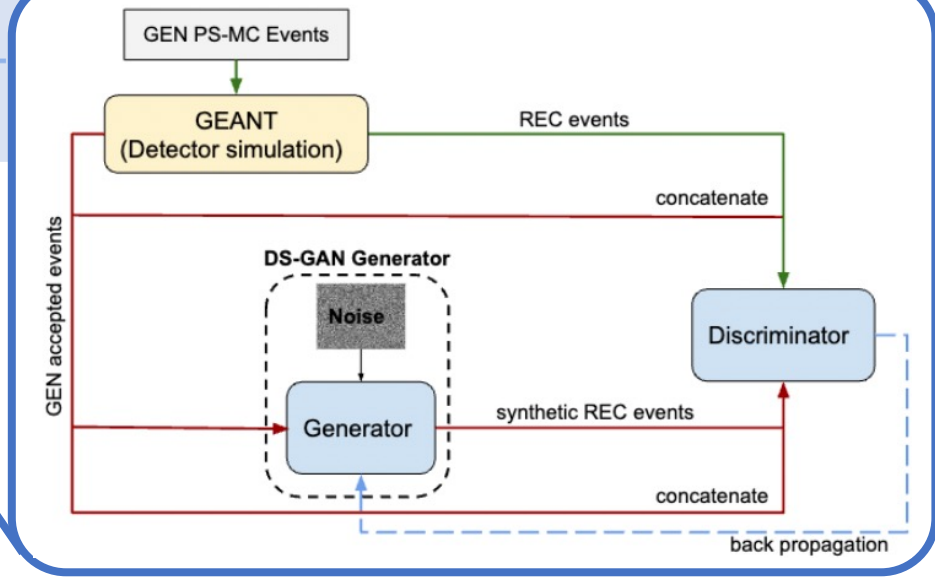
DS-GAN Generator

synthetic REC events

Discriminator

Detector Simulation GAN (DS-GAN)

- AI-based detector proxy
- Train over phase space data to avoid model dependance
- Uncertainty evaluated training 10 independent GANs

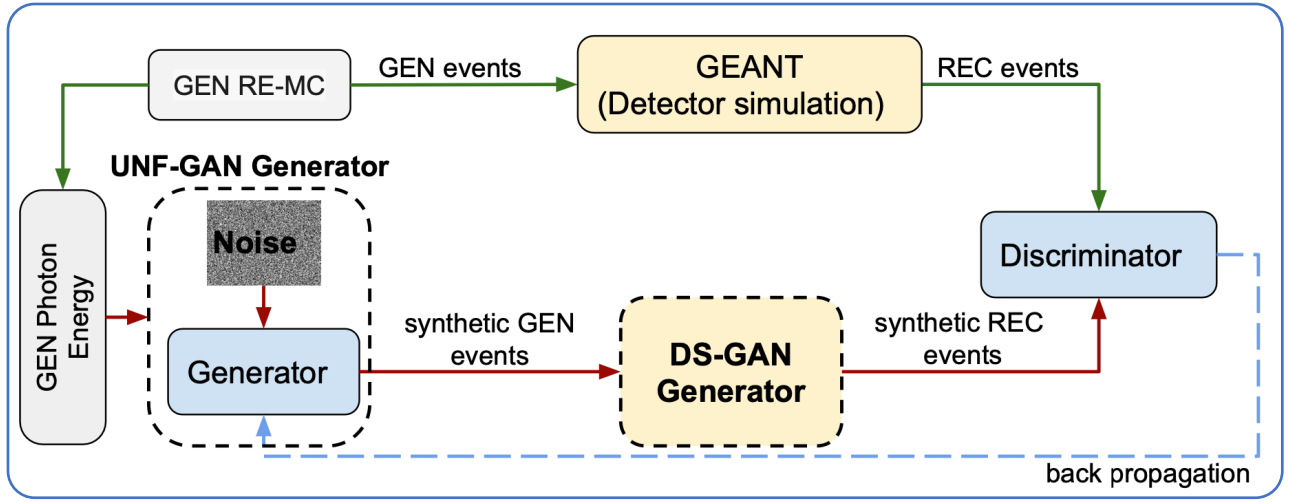


Credit: T.Alghamdi et al, PRD 108, 094030

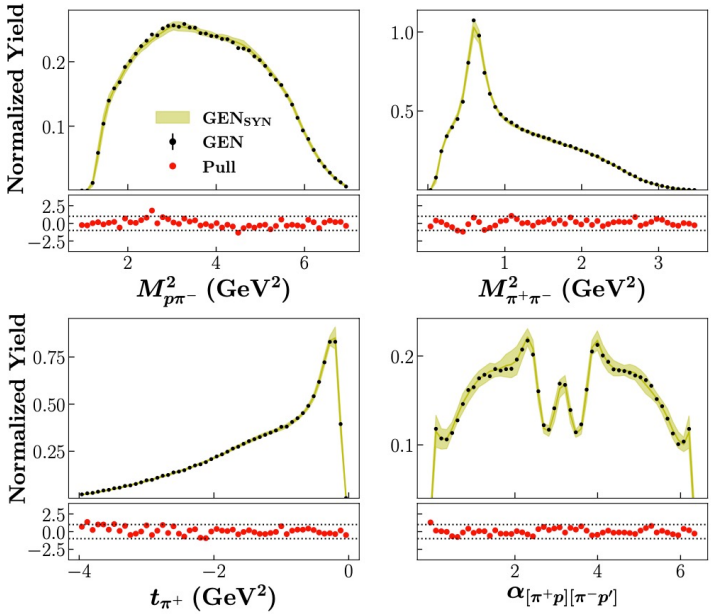
2π photoproduction closure test

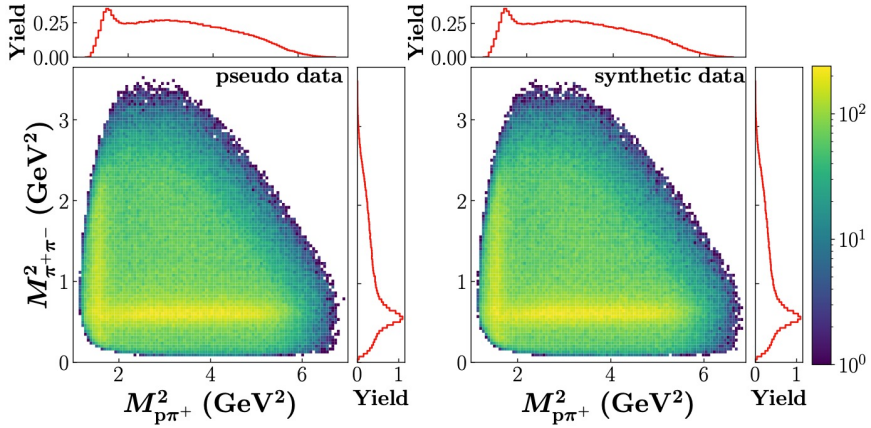
Training of the UNF-GAN with pseudo-data

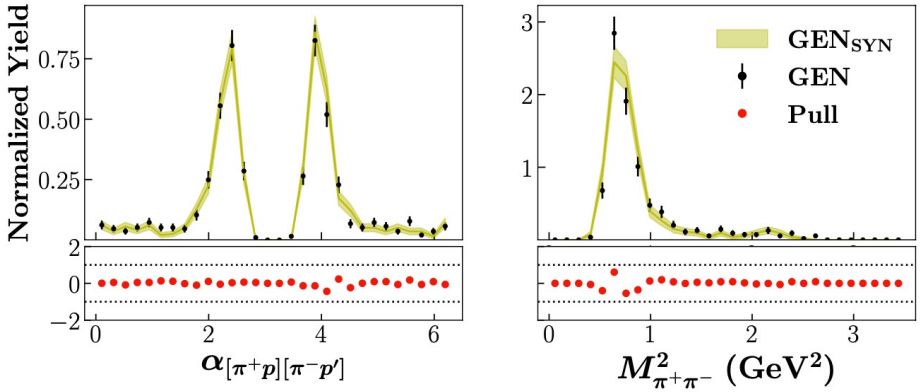
- Trained on MC pseudo-data
- Generated synthetic vertex-level data
- Detector effects applied with DS-GAN
- Uncertainty estimated with pull quantification



1D → **2D** → **4D**





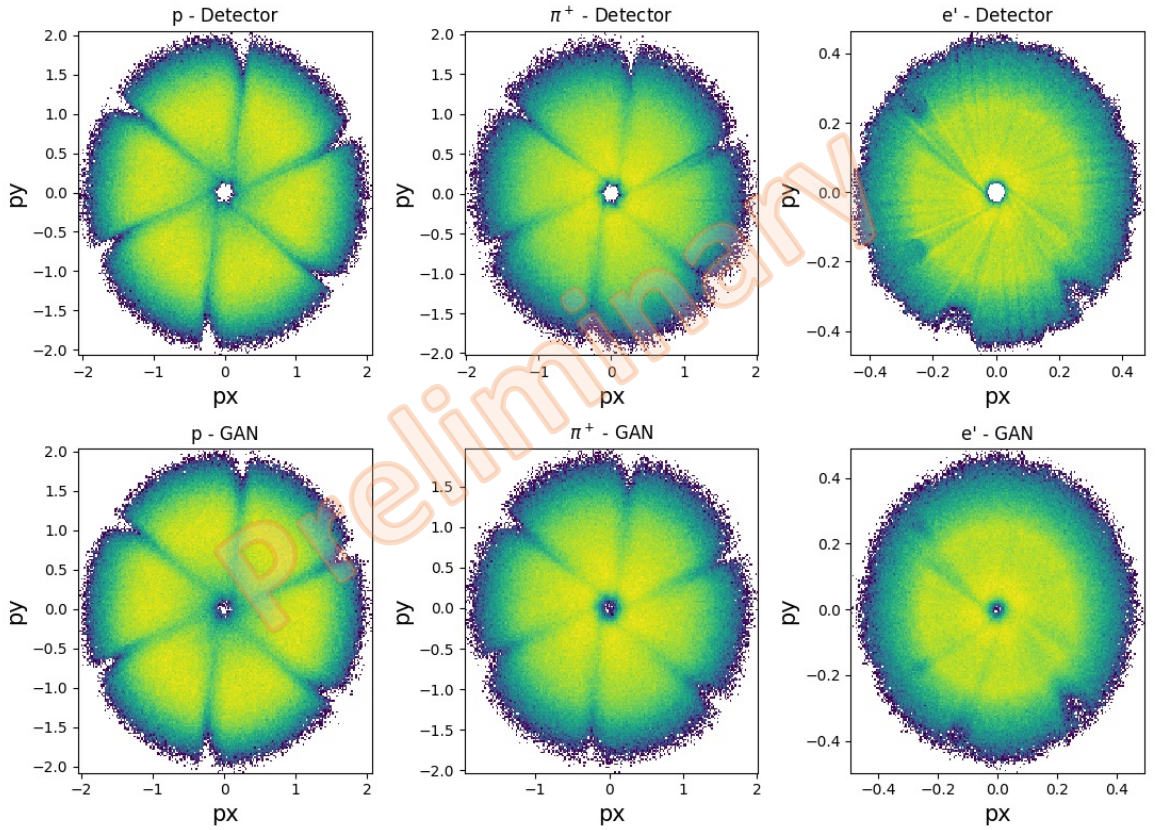


UNF-GAN can reproduce detector level pseudo-data within $\pm 1\sigma$
Correlations are preserved both in 2D and in 4D distributions within $\pm 1\sigma$

Work in progress

The next step to achieve A(i)DAPT goals:

- **Application to real data**
  - Train UNF-GAN using CLAS g11 data
  - Assess GAN capability learn real data features
- **Application to CLAS12 detector and physics**
  - Train DS-GAN on CLAS12 pseudo-data
  - Apply UNF-GAN to electroproduction data
- **Extrapolation of scattering amplitudes**
  - Extract amplitudes from differential cross-sections exploiting theoretical constraints



Our goal is to develop a new tool accessible to everyone and that can be used to improve any analysis

Summary

A(I)DAPT program aims to demonstrate a novel way to extract and interpret physics observables

- We aim at creating AI-powered algorithms to address NP/HEP challenges:
  - Unfold detector effects
  - Preserve data in an alternative and efficient way

- Performed closure test on 2π photoproduction :
  - GAN can mimic realistic pseudo-data
  - GAN synthetic data retain multi-dimensional correlations

- Proven algorithm robustness:
  - It can reproduce different detector layouts (CLAS, CLAS12)
  - It can simulate different processes (photoproduction, electroproduction)

We are working on:
- Application on real data (CLAS and CLAS12 2π data)
- Use of AI to extract scattering amplitudes from cross-section data

Thank you!