The A(i)DAPT program
AI for Data Analysis and Preservation

M.Battaglieri (INFN)

on behalf of A(i)DAPT Working Group

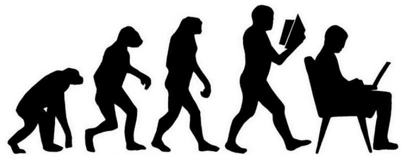
A(i)DAPT

AI for Data Analysis and PreservaTion

A(i)DAPT

- Data collected by NP/HEP experiments are (always) affected by the detector's effects
- Before starting physics analysis the detector's effects unfolding are required
- Traditional observables may not be adequate to extract physics in multidimensional space (multi-particles in the final state)
- At High-Intensity frontiers, data sets are large and difficult to manipulate/preserve

Shall AI support NP/HEP experiments to extract physics from data in a more efficient way?



A(i)DAPT
AI for Data Analysis and PreservaTion

Develop AI-supported procedures to:

- Accurately fit data in multiD space
- Unfold detector effects
- Compare *synthetic* (AI-generated) to experimental data
- Quantify the uncertainty (UQ)

Collaborative effort (regular meeting)

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

The A(i)DAPT road map

- **Deploy an AI Generative Model to reproduce NP/HEP data**
  - Detector effects unfolding: smearing
  - Detector effects unfolding: acceptance
  - Extract few dimensions cross-section (PDF) (e.g. inclusive electron scattering - MC)
  - Extend the closure test to cross-sections in a mutiD phase-space (e.g. 2-pion photoproduction - MC)
  - Validate the analysis procedure extracting cross-section from data (e.g. high energy CLAS-g11 2-pion data)
  - Combine data of the same final state taken in different kinematics (e.g. low energy CLAS-g11 2-pion data)
  - Combine data from different final states (e.g. CLAS-g11 3-pion/ω data)
  - Extract cross-section and amplitudes in a 2-body reaction (e.g. ππ scattering - MC)
  - Extract amplitudes from a multi-particle exclusive channel (e.g. CLAS-g11 2-pion data)
  - Extract amplitudes in multi- coupled-channel analysis (e.g. CLAS-g11 2-pion + 3-pion/ω data)
  - Connect NN features to different physics processes (e.g. baryon and meson resonances in CLAS-g11 2-pion data)
  - …
- **Extract physics out of our data**

This talk

- Deploy an AI Generative Model to reproduce NP/HEP data
- Detector effects unfolding: smearing
- Detector effects unfolding: acceptance
- Extract few dimensions cross-section (PDF) (e.g. inclusive electron scattering - MC)
- Extend the closure test to cross-sections in a mutiD phase-space (e.g. 2-pion photoproduction - MC)
- Validate the analysis procedure extracting cross-section from data (e.g. high energy CLAS-g11 2-pion data)
- Combine data of the same final state taken in different kinematics (e.g. low energy CLAS-g11 2-pion data)
- Combine data from different final states (e.g. CLAS-g11 3-pion/ω data)
- Extract cross-section and amplitudes in a 2-body reaction (e.g. ππ scattering - MC)
- Extract amplitudes from a multi-particle exclusive channel (e.g. CLAS-g11 2-pion data)
- Extract amplitudes in multi- coupled-channel analysis (e.g. CLAS-g11 2-pion + 3-pion/ω data)
- Connect NN features to different physics processes (e.g. baryon and meson resonances in CLAS-g11 2-pion data)
- …
- Extract physics out of data

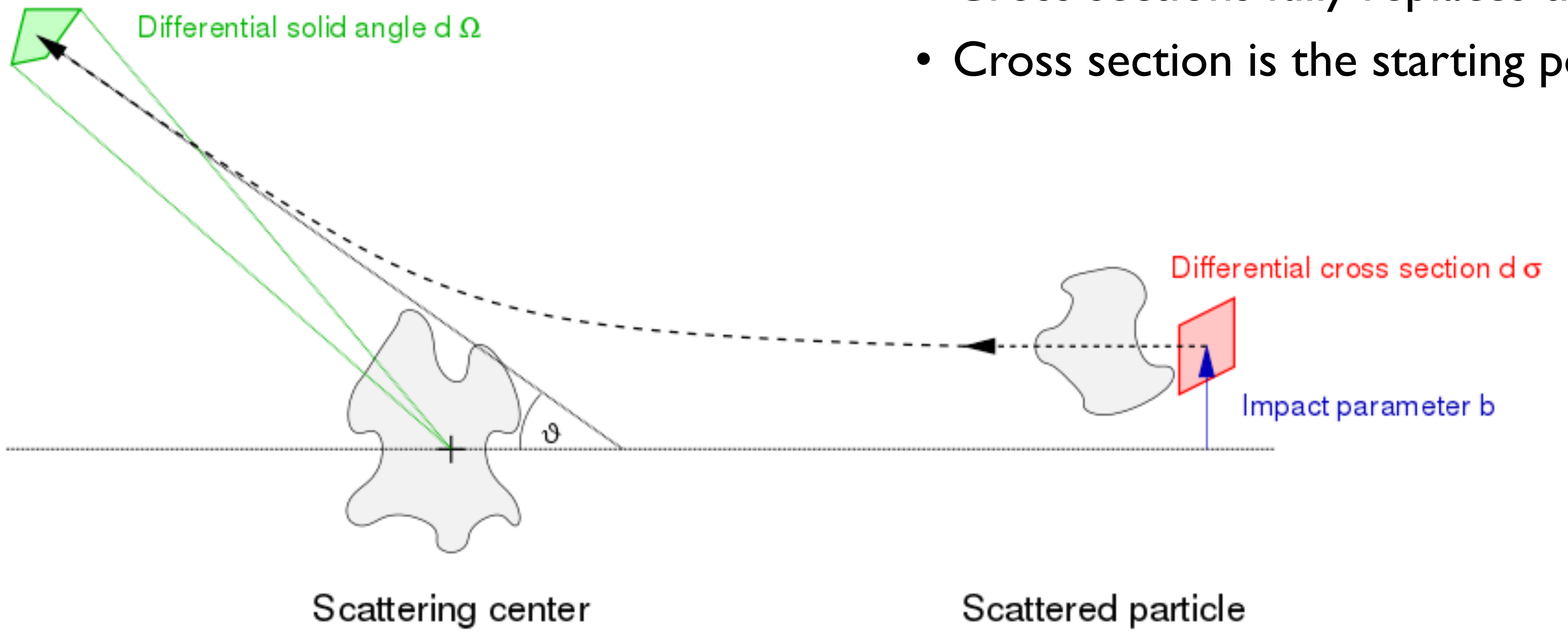
This talk

- Deploy an AI Generative Model to reproduce NP/HEP data ✓
- Detector effects unfolding: smearing ✓
- Detector effects unfolding: acceptance **in progress**
- Extract few dimensions cross-section (PDF) (e.g. inclusive electron scattering - MC) ✓
- Extend the closure test to cross-sections in a mutiD phase-space (e.g. 2-pion photoproduction - MC) ✓
- Validate the analysis procedure extracting cross-section from data (e.g. high energy CLAS-g11 2-pion data)
- Combine data of the same final state taken in different kinematics (e.g. low energy CLAS-g11 2-pion data)
- Combine data from different final states (e.g. CLAS-g11 3-pion/ω data)
- Extract cross-section and amplitudes in a 2-body reaction (e.g. ππ scattering - MC) **in progress**
- Extract amplitudes from a multi-particle exclusive channel (e.g. CLAS-g11 2-pion data)
- Extract amplitudes in multi- coupled-channel analysis (e.g. CLAS-g11 2-pion + 3-pion/ω data)
- Connect NN features to different physics processes (e.g. baryon and meson resonances in CLAS-g11 2-pion data)
- …
- Extract physics out of data

The cross section in particle physics

$$\frac{d\sigma}{d\Omega} = (2\pi)^4 m_i m_f \frac{p_f}{p_i} |T_{fi}|^2$$

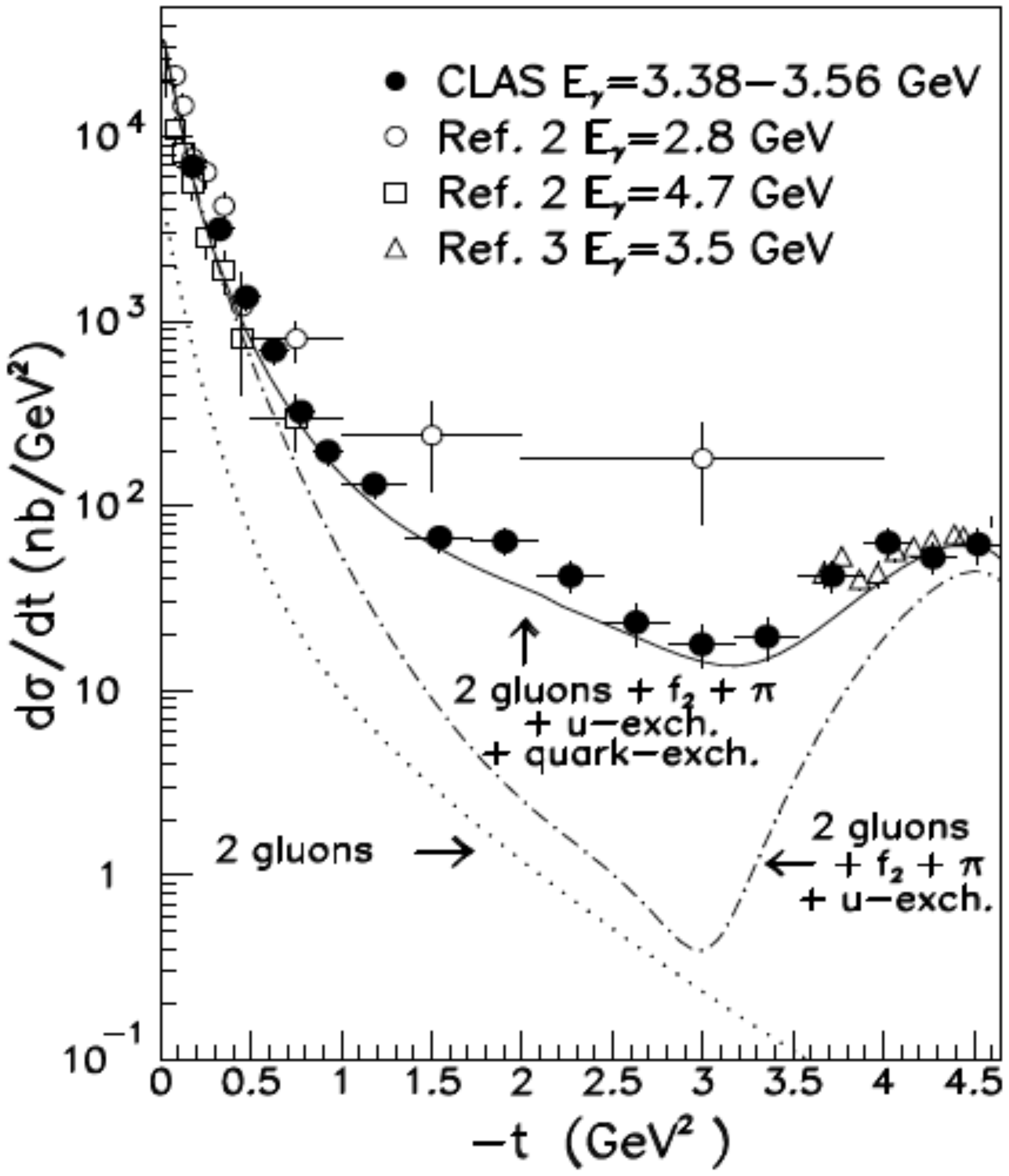
- The *cross section* is related to the transition probability between an initial to a final state
- In case of scattering, cross sections provides information about the elementary interaction
- Cross section is expressed as squared sum of scattering amplitudes (complex functions) depending on the kinematic Lorentz-invariant of the problem and embedding the interaction properties
- It is derived by measuring the momentum distributions of reaction particle (at different CM energy)
- Correlations between particles in the final state reflects the underlying dynamics
- Cross sections fully replaces the 4-mom data sample in a compact and efficient way
- Cross section is the starting point for any higher level physics analysis



- Traditional approach: particles (4-momenta) measured into the detector, extract the relevant observables, extract physics mechanisms
- Cross section **preserves** this information as replacement for the original particle-by-particle scattering information

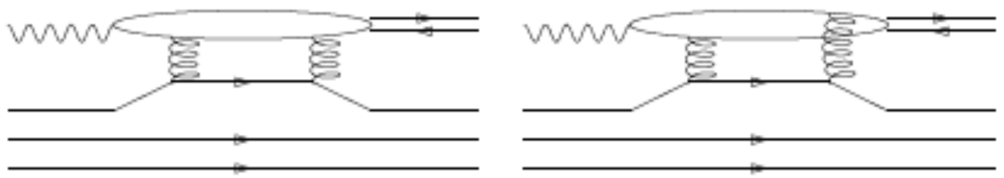
Exclusive reactions: 2 → 2

M. Battaglieri *et al.* (CLAS Collaboration) Phys. Rev. Lett. 90, 022002

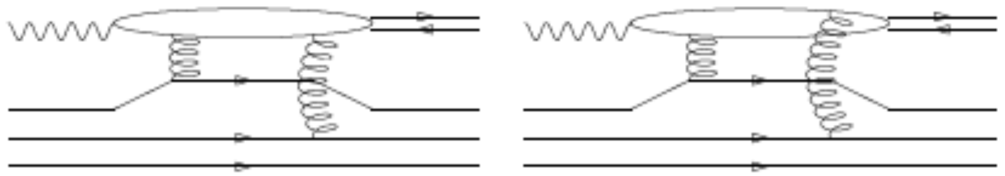


JLab-CLAS g6 ω photo production at large momentum transfer

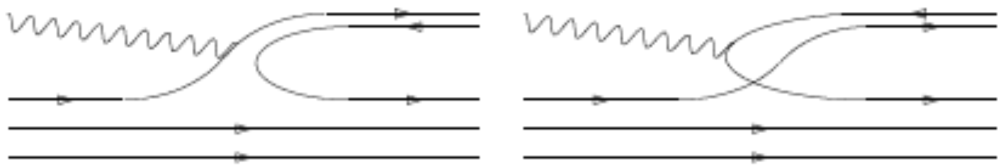
$\gamma\, p \rightarrow p\, \omega$



2-gluon exchange



Correlations



quark exchange

2 → 2 scattering (no polarisation)

- Initial state: known
- Final state: 2 x 3
- Parameters: (2 x 3) - 4 = 2
- Possible choice: -*t* and *ϕ*
- *the* physics depends only on one variable (-t)

- It worked (and still works!) well if limited to channels with a single variable
- Xsec, Polarization observables, angular distribution, decay matrix, …

Exclusive reactions: 2 → 3

M. Battaglieri *et al.* (CLAS Collaboration) Phys. Rev. Lett. 90, 022002

2 → 3 scattering (no polarization)
- Initial state: known
- Final state: 3 x 3
- Parameters: (3 x 3) - 4 = 5 (E_γ fixed)
- Possible choice: $M^2_{\pi\pi}$, $M^2_{p\pi}$ θ_π, α, ϕ

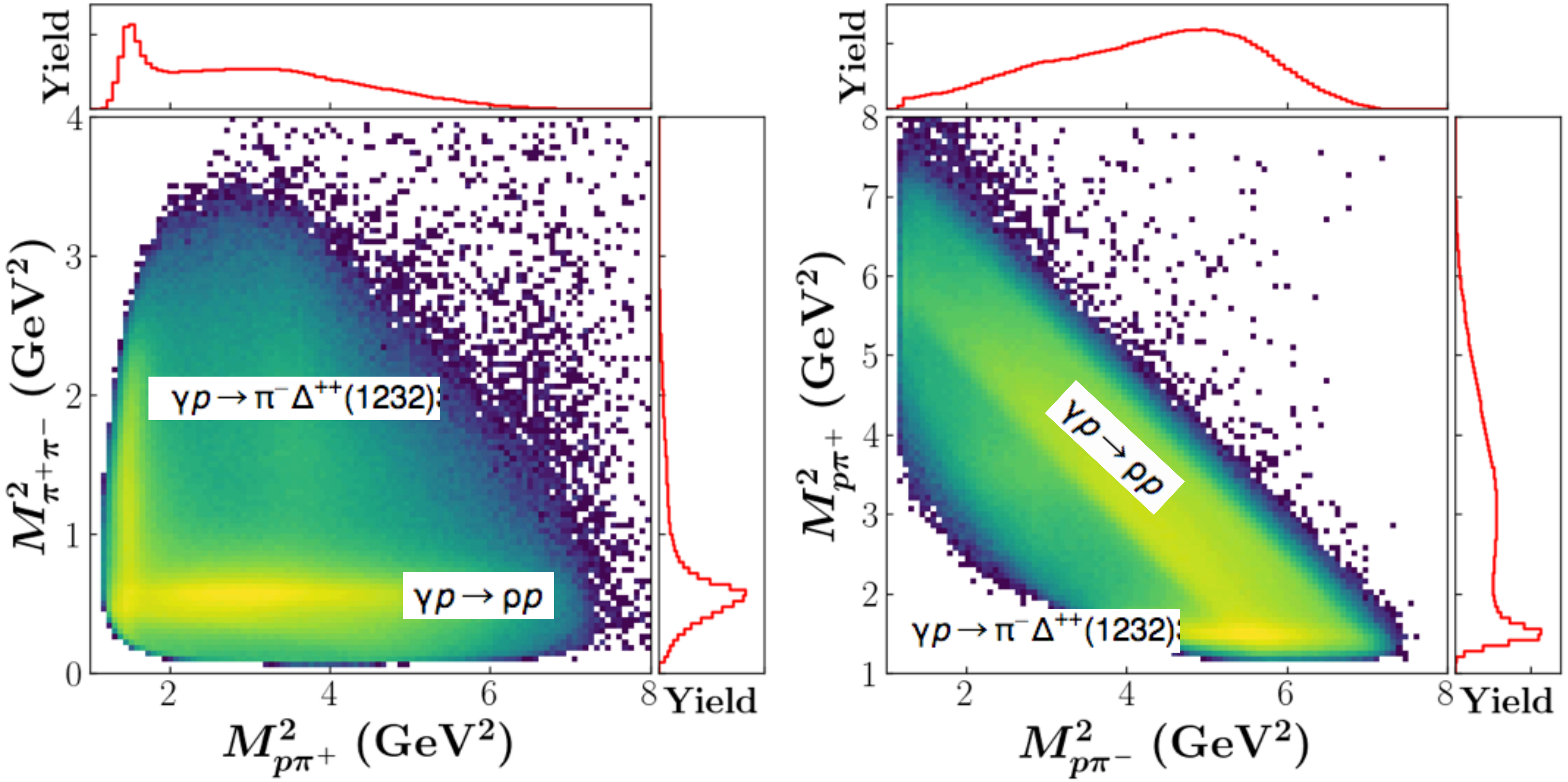
- It does not work (in practice) when you have several independent variables: multi-particle final states (spectroscopy) or multi-variable correlations (SIDIS)
- In the integration to reduce to 1-dim all correlations are lost

AI may provide a new way to look at data and extract observables and physics interpretation (on event by event base)

CLAS g11 2π photo production

- E_γ = (3.0 - 3.8) GeV
  $\gamma\, p \to p\, \pi^+\, \pi^-$ exclusive reaction
  - data set analyses so far $\gamma\, p \to p\, \pi^+ (\pi^-)$ + small contamination of $\gamma\, p \to p\, \pi^+$ (more than a missing π^-)
  - complicated dynamic for the overlap of (pπ) to form Δ baryon resonances and (ππ) to form meson resonances

$$\frac{d\sigma\,(\gamma\, p \to p\, \pi^+\pi^-)}{dM_{\pi\pi}\, dM_{p\pi}\, d\cos(\theta_\pi)\, d\alpha\, d\phi}$$



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

Detector unfolding

- **Detector effects make measured observables (detector-level) DIFFERENT from 'true' observables (vertex-level)**
  I. **Acceptance**: any measurements only access a limited region of the phase space.

  How to recover the unmeasured region?
  - Interpolation: holes in the phase space
  - Extrapolation: border of the accessible phase space

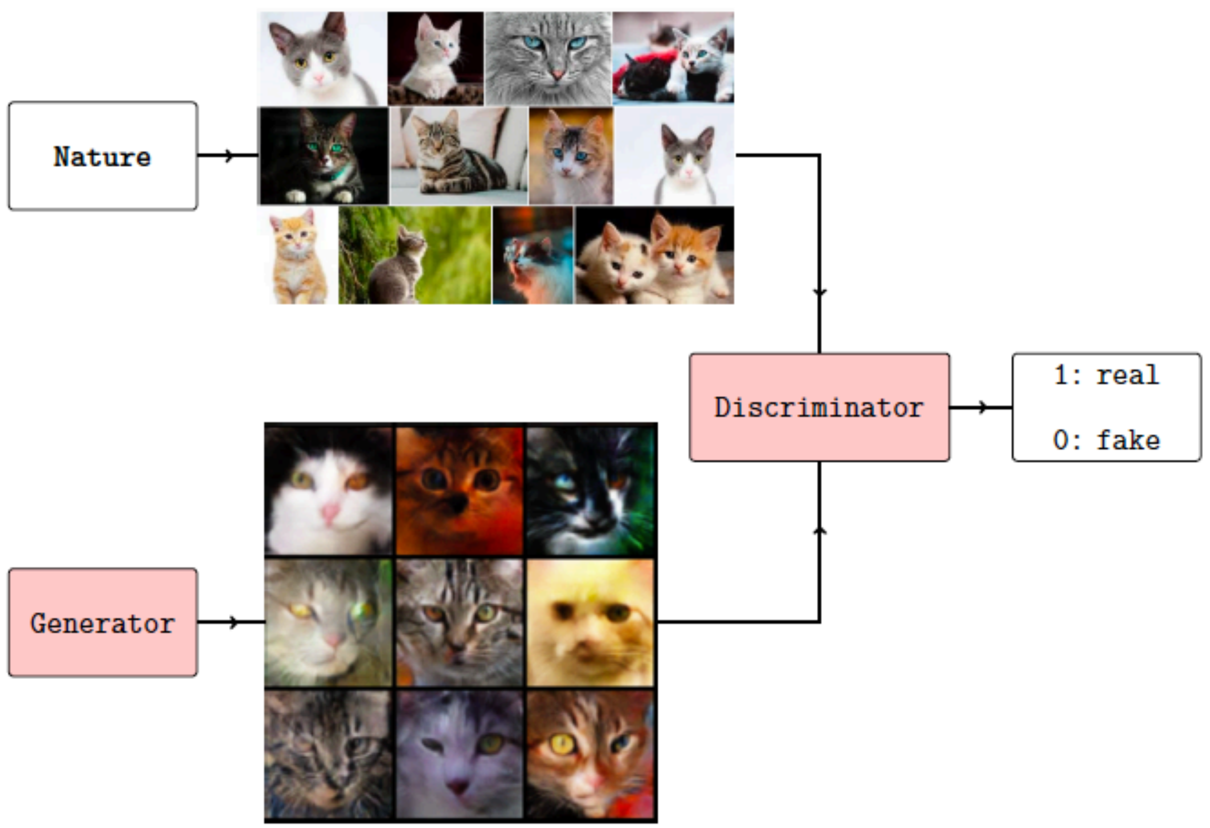
  II. **Resolution**: any measurement has an experimental resolution that may hide or washout the effect searched for
  - A spike could be not resolved, the measurement may extend in an unphysical region (e.g. negative squared missing mass)
- **For both effects, one needs to quantify the systematic errors introduced to the vertex-level observables**

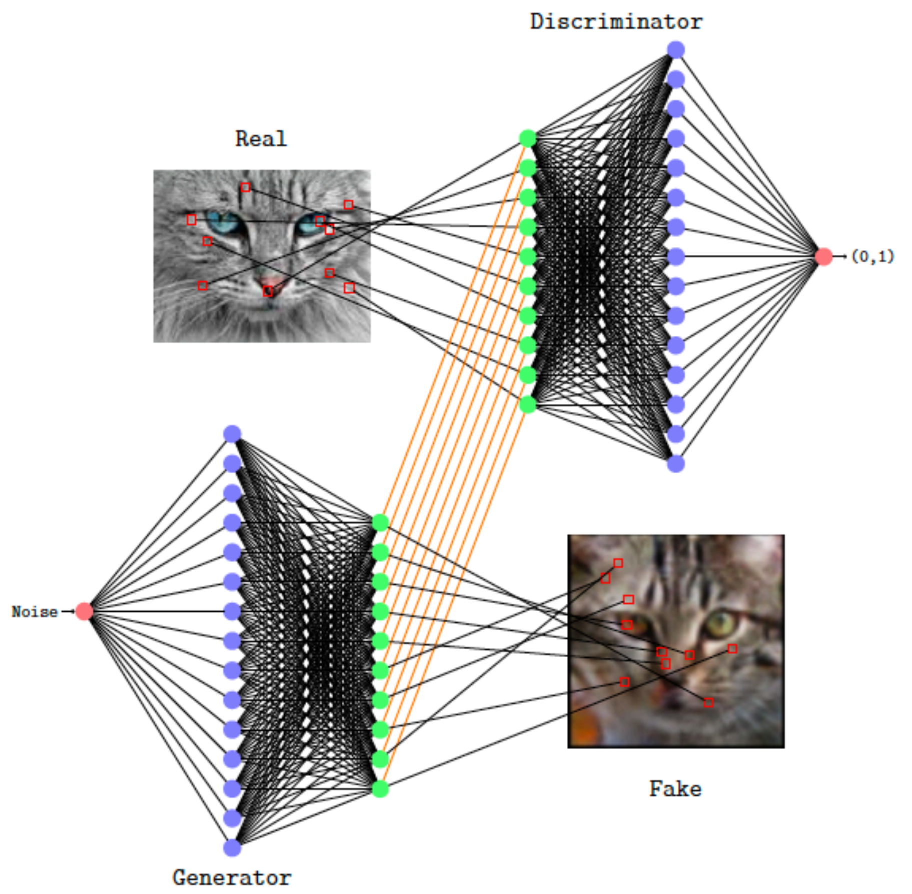
- **Mitigation strategy:**
  - Acceptance: '*fiducial volumes*' to exclude unmeasured or poorly-measured regions verifying the training convergence
  - Resolution: closure test with a reasonable model of the detector using a detector proxy (parametric or GEANT-based)

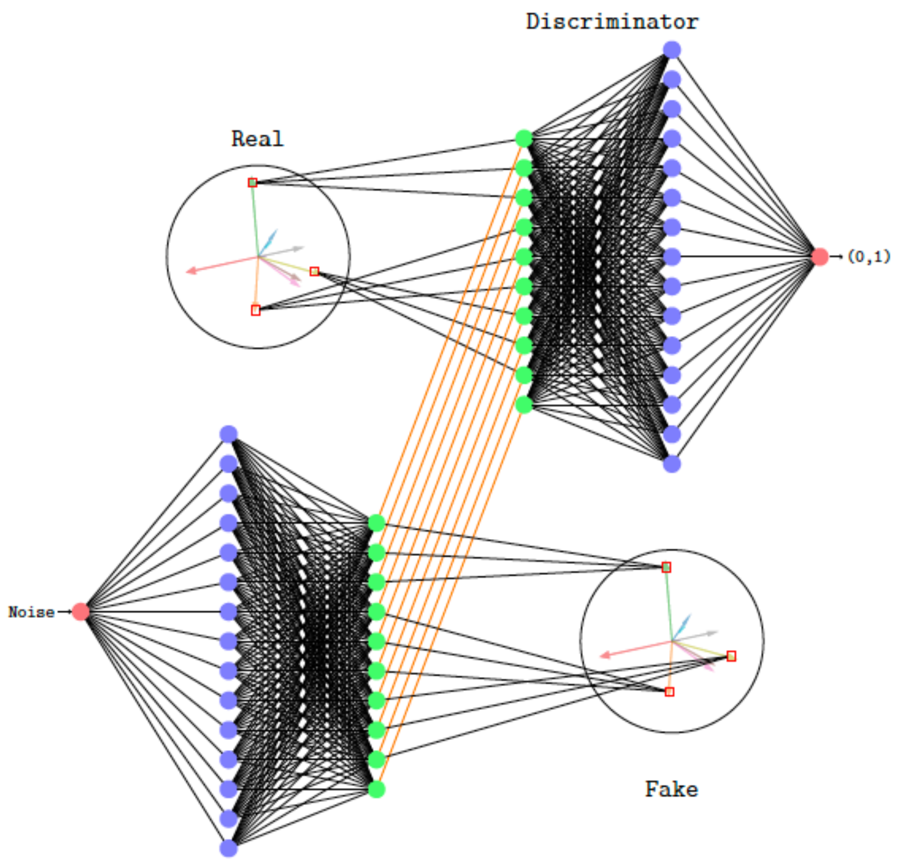
https://doi.org/10.24963/ijcai.2021/588

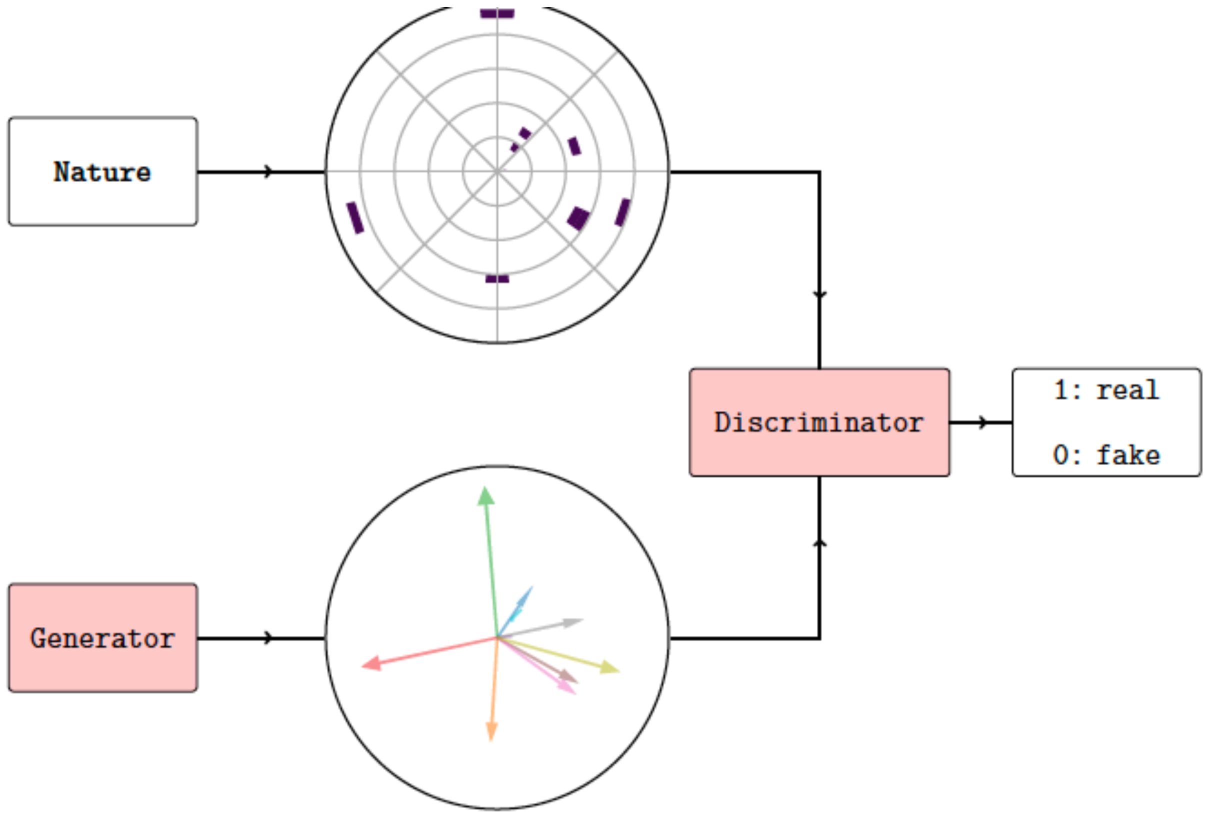
Generative Adversarial Network (GANs)

- The colored boxes are built using NNs
- Discriminator is trained to output "real" for Nature samples
- Generator is trained to fool the discriminator
- The Generator can be used as data compression tool
- Typical size for the Generator: O(MB) - to be compared to NP/HEP experiments data set O(GB/TB)
- Simple to distribute instead of events stored on tapes







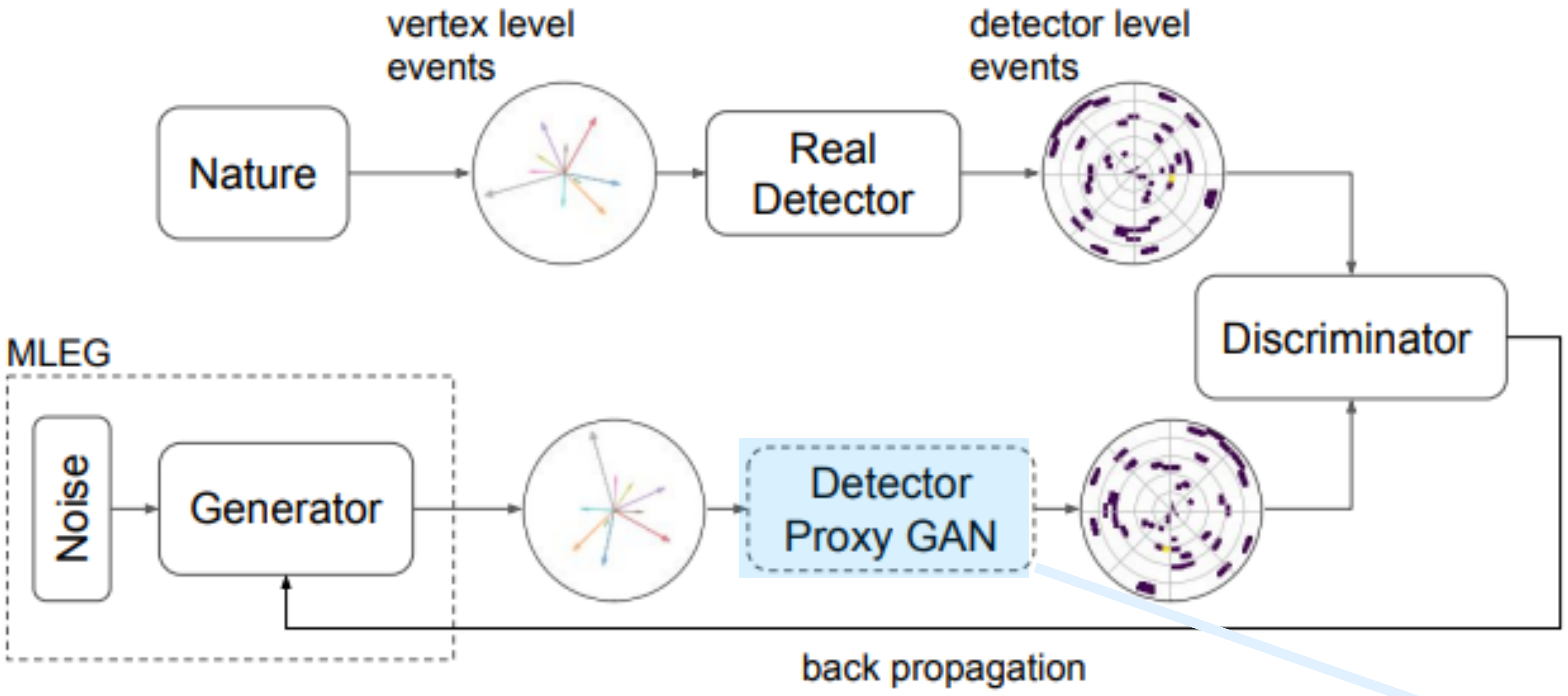


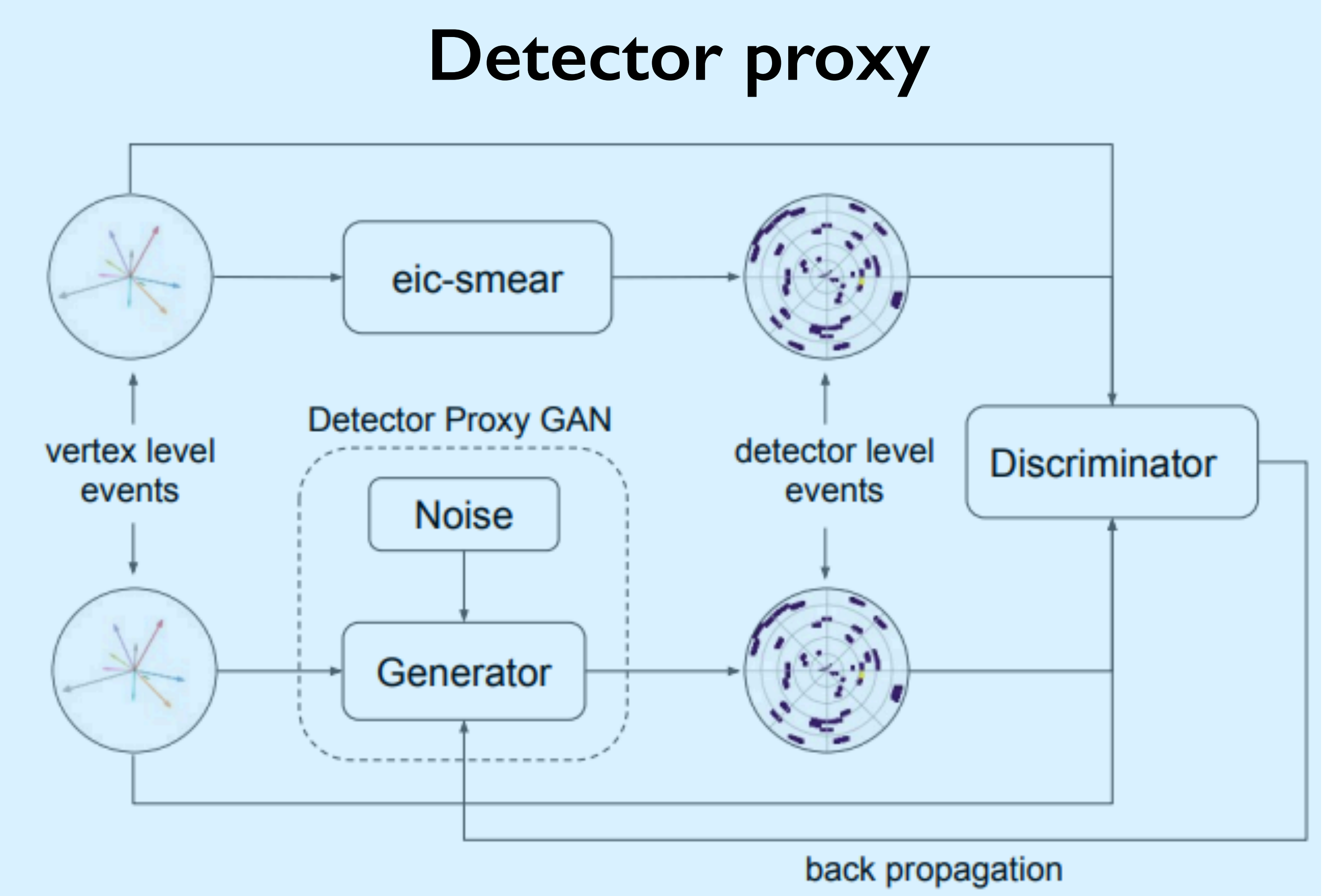
Credit: Y. Li, N.Sato

A simple study case: DIS scattering

Y. Alanazi, P. Ambrozewicz, M. Battaglieri, A.N. Hiller Blin, M.P. Kuchera, Y. Li, T. Liu, R.E. McClellan, W. Melnitchouk, E. Pritchard, M. Robertson, N. Sato, R. Strauss, and L. Velasco Phys. Rev. D **106**, 096002

ML Event Generator GAN scheme





- 100-d white noise entered at 0, unit standard dev.
- Generator: 5 hidden layers / 512 neurone per layer, ReLU activation function. Last layer connected to 2 neurons output to generate v_1 and v_2 variables
- Discriminator: same NN architecture as for the generator
- Detector proxy: similar architecture
- Least Squares GAN (LSGAN)
- Trained adversarially for 100000 epochs (pass through the training data set)
- Adam's optimizer

- *eic-smear*: parametric smearing routine for the Electron Ion Collider detectors (no GEANT-based simulations)
- Parameters tuned to reproduce ZEUS/H1 detectors
- Full 4π acceptance

A simple study case: DIS scattering

Y. Alanazi, P. Ambrozewicz, M. Battaglieri, A.N. Hiller Blin, M.P. Kuchera, Y. Li, T. Liu, R.E. McClellan, W. Melnitchouk, E. Pritchard, M. Robertson, N. Sato, R. Strauss, and L. Velasco Phys. Rev. D **106**, 096002

I) GAN training w/o detector effects

Pseudo-data sample (JAM)

- Inclusive electron DIS generated at E_{CM}=318.2 GeV (HERA kinematics)
- 2-dim differential cross section $d\sigma/dxdQ^2$
- Lorentz boosted from CM to Lab (+ uniform azimuthal angle)
- To reduce violation of momentum conservation on the edge of the phase space due to smearing effects, electron momentum is replaced by new variables:

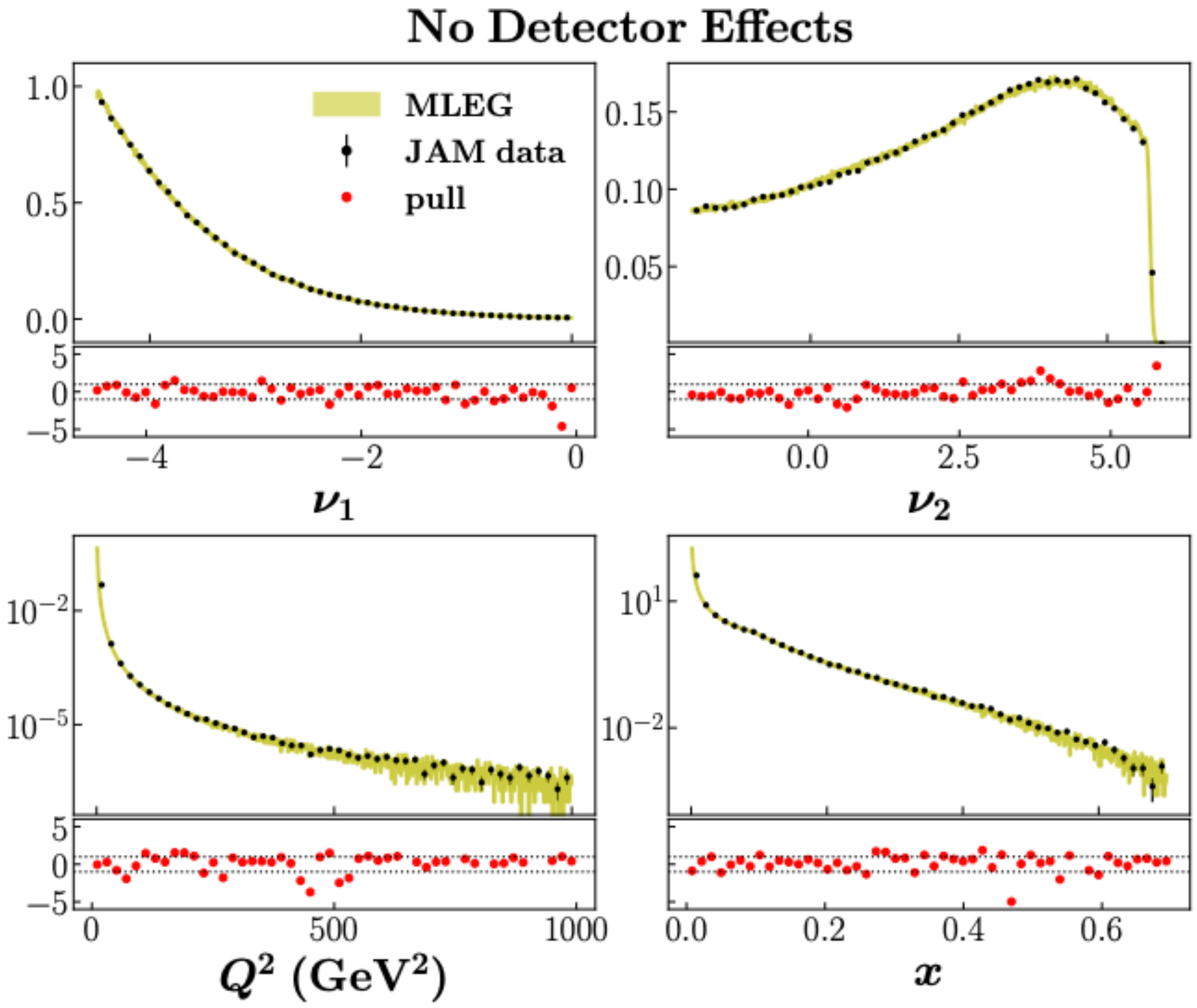
$$\nu_1 = \ln\left((k'_0 - k'_z)/1\,\mathrm{GeV}\right),$$
$$\nu_2 = \ln\left((2E_e - k'_0 - k'_z)/1\,\mathrm{GeV}\right),$$

Uncertainty Quantification via *pull* calculation

- Metric: *pull*

$$\text{pull} = \frac{\mathrm{E}[\mathcal{P}(\mathcal{O}|\text{bin})]_{\text{GAN}} - \mathrm{E}[\mathcal{P}(\mathcal{O}|\text{bin})]_{\text{JAM}}}{\sqrt{\mathrm{V}[\mathcal{P}(\mathcal{O}|\text{bin})]_{\text{GAN}} + \mathrm{V}[\mathcal{P}(\mathcal{O}|\text{bin})]_{\text{JAM}}}}$$

- *Bootstrap with* 10 independently trained GANs



A simple study case: DIS scattering

Y. Alanazi, P. Ambrozewicz, M. Battaglieri, A.N. Hiller Blin, M.P. Kuchera, Y. Li, T. Liu, R.E. McClellan, W. Melnitchouk, E. Pritchard, M. Robertson, N. Sato, R. Strauss, and L. Velasco Phys. Rev. D **106**, 096002

I) GAN training w/o detector effects

Pseudo-data sample (JAM)

- Inclusive electron DIS generated at E_{CM}=318.2 GeV (HERA kinematics)
- 2-dim differential cross section dσ/dxdQ2
- Lorentz boosted from CM to Lab (+ uniform azimuthal angle)
- To reduce violation of momentum conservation on the edge of the phase space due to smearing effect, electron momentum is replaced by new variables:

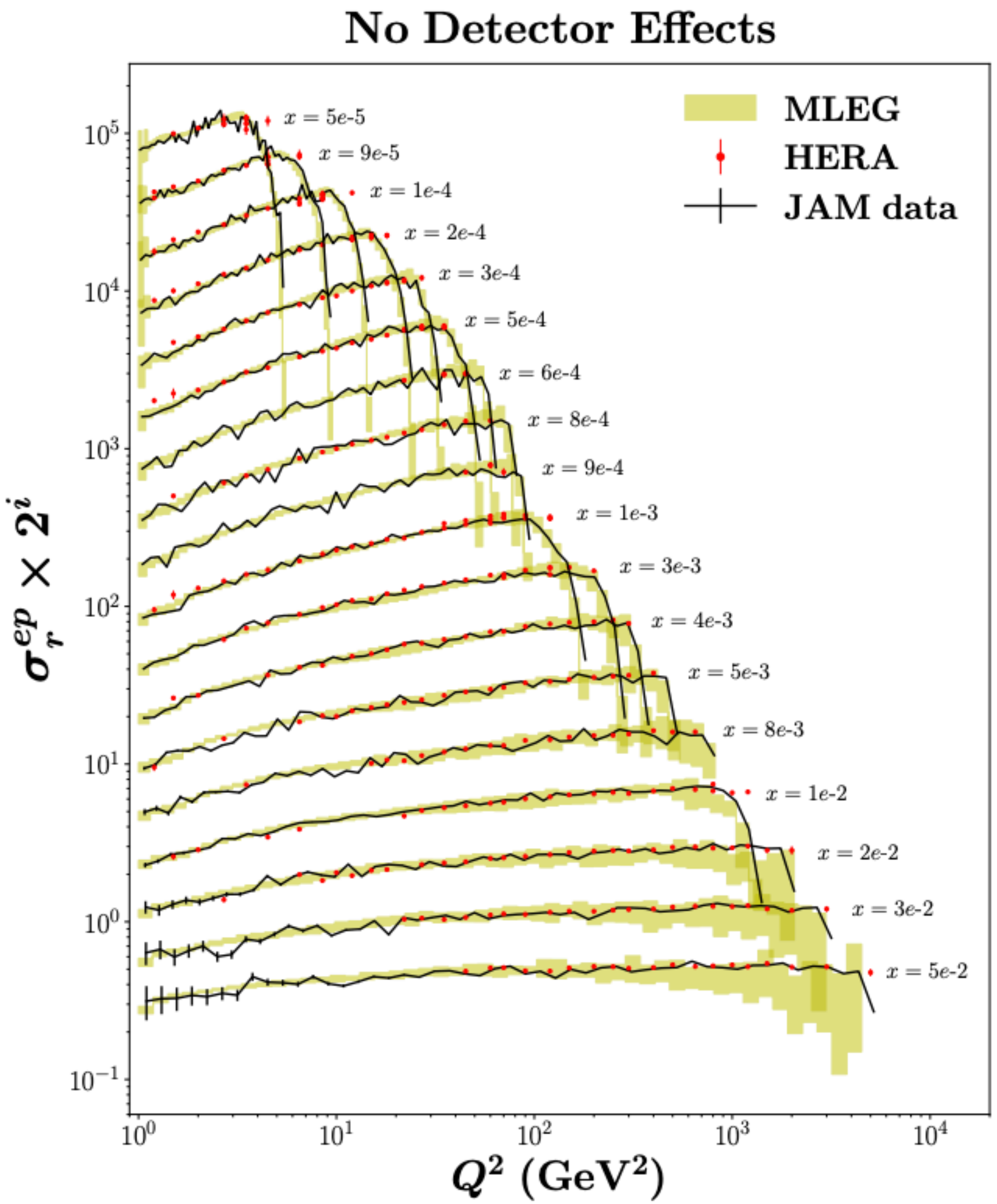
$$\nu_1 = \ln\left((k'_0 - k'_z)/1\,\mathrm{GeV}\right),$$
$$\nu_2 = \ln\left((2E_e - k'_0 - k'_z)/1\,\mathrm{GeV}\right),$$

Uncertainty Quantification via *pull* calculation

- Metric: *pull*

$$\mathrm{pull} = \frac{\mathrm{E}[\mathcal{P}(\mathcal{O}|\mathrm{bin})]_{\mathrm{GAN}} - \mathrm{E}[\mathcal{P}(\mathcal{O}|\mathrm{bin})]_{\mathrm{JAM}}}{\sqrt{\mathrm{V}[\mathcal{P}(\mathcal{O}|\mathrm{bin})]_{\mathrm{GAN}} + \mathrm{V}[\mathcal{P}(\mathcal{O}|\mathrm{bin})]_{\mathrm{JAM}}}}$$

- *Bootstrap with* 10 independently trained GAN

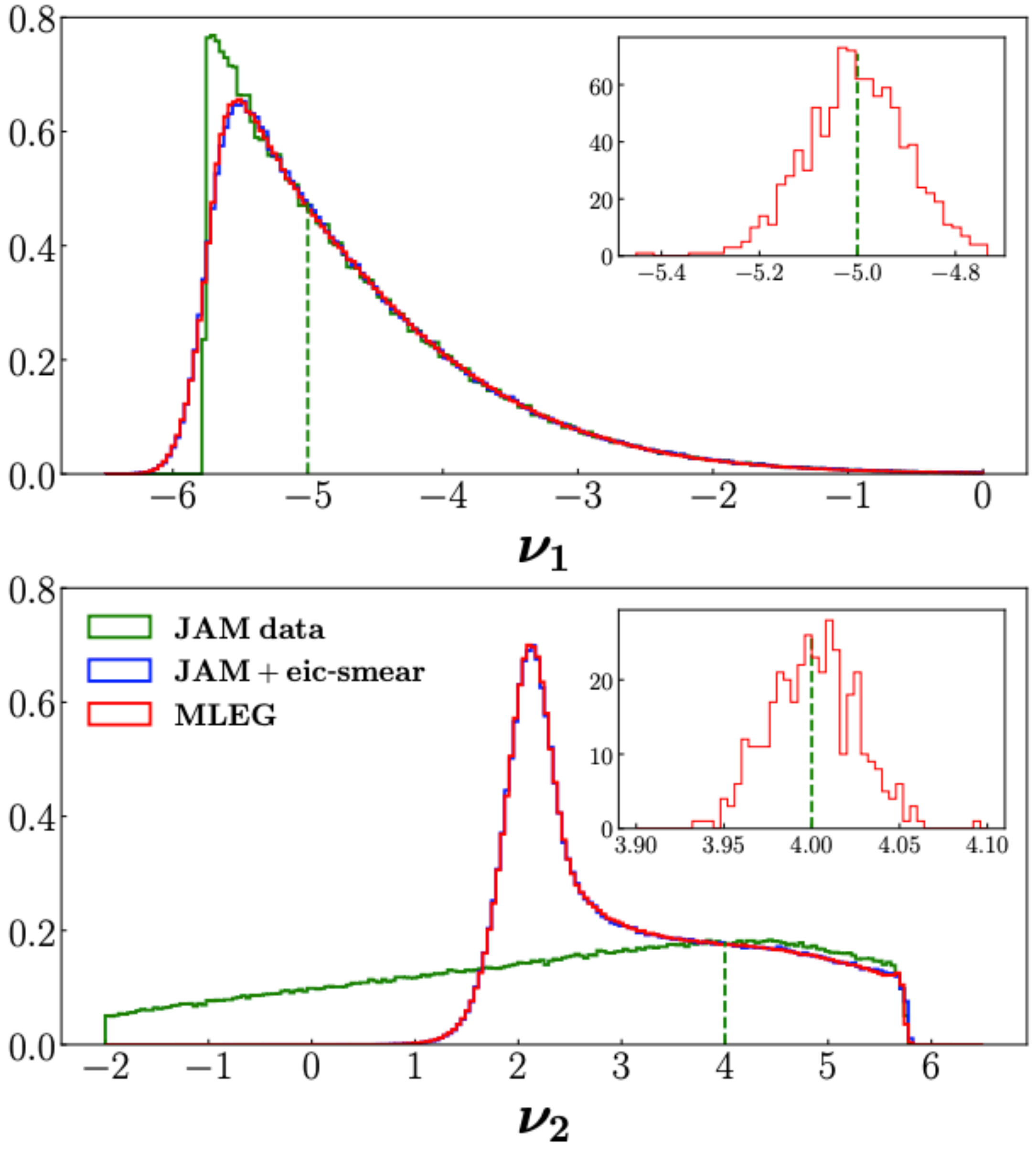


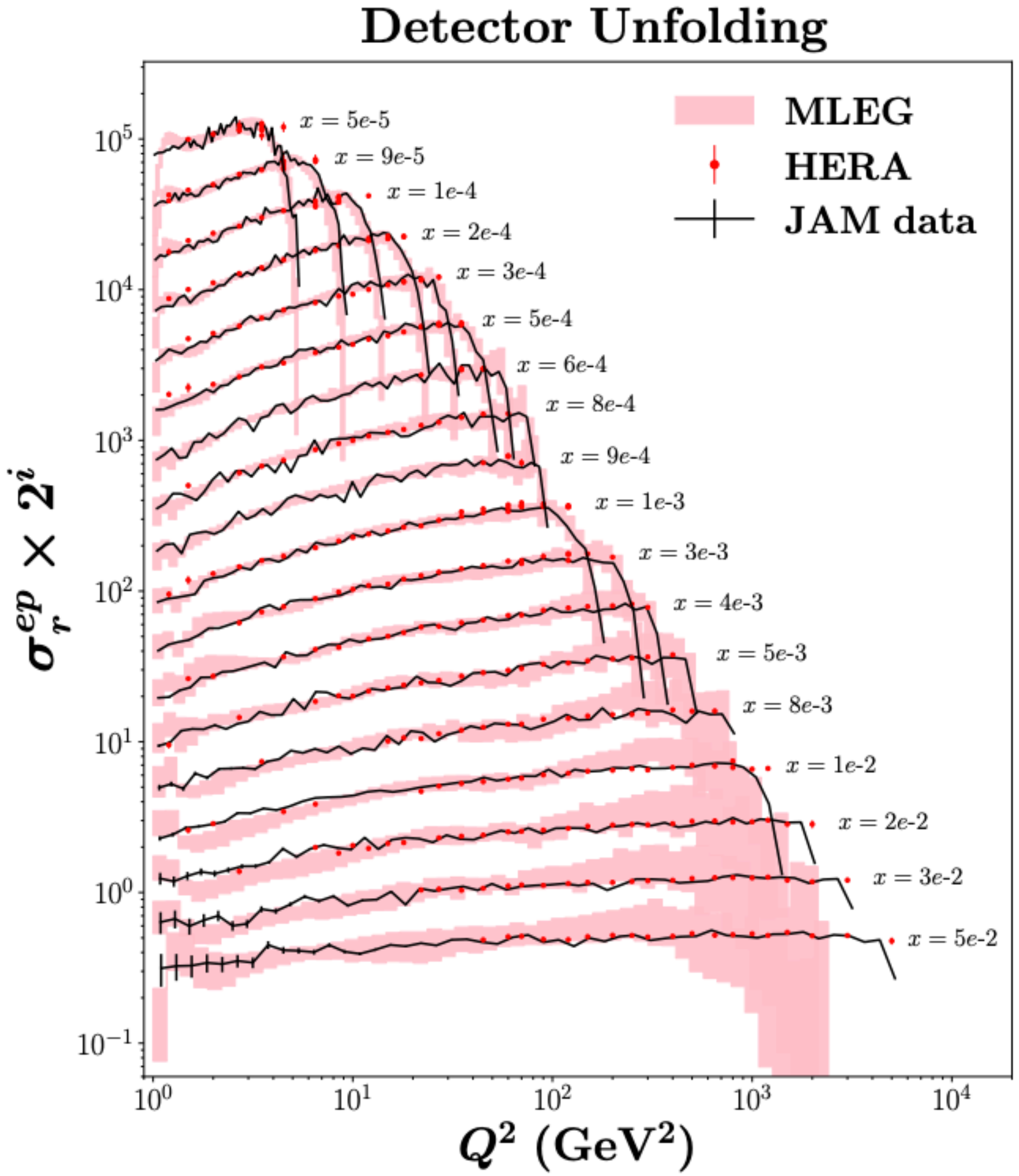
A simple study case: DIS scattering

Y. Alanazi, P. Ambrozewicz, M. Battaglieri, A.N. Hiller Blin, M.P. Kuchera, Y. Li, T. Liu, R.E. McClellan, W. Melnitchouk, E. Pritchard, M. Robertson, N. Sato, R. Strauss, and L. Velasco Phys. Rev. D **106**, 096002

II) GAN training WITH detector effects

- eic-smear introduces significant distortions to the detector level sample in particular on ν_2



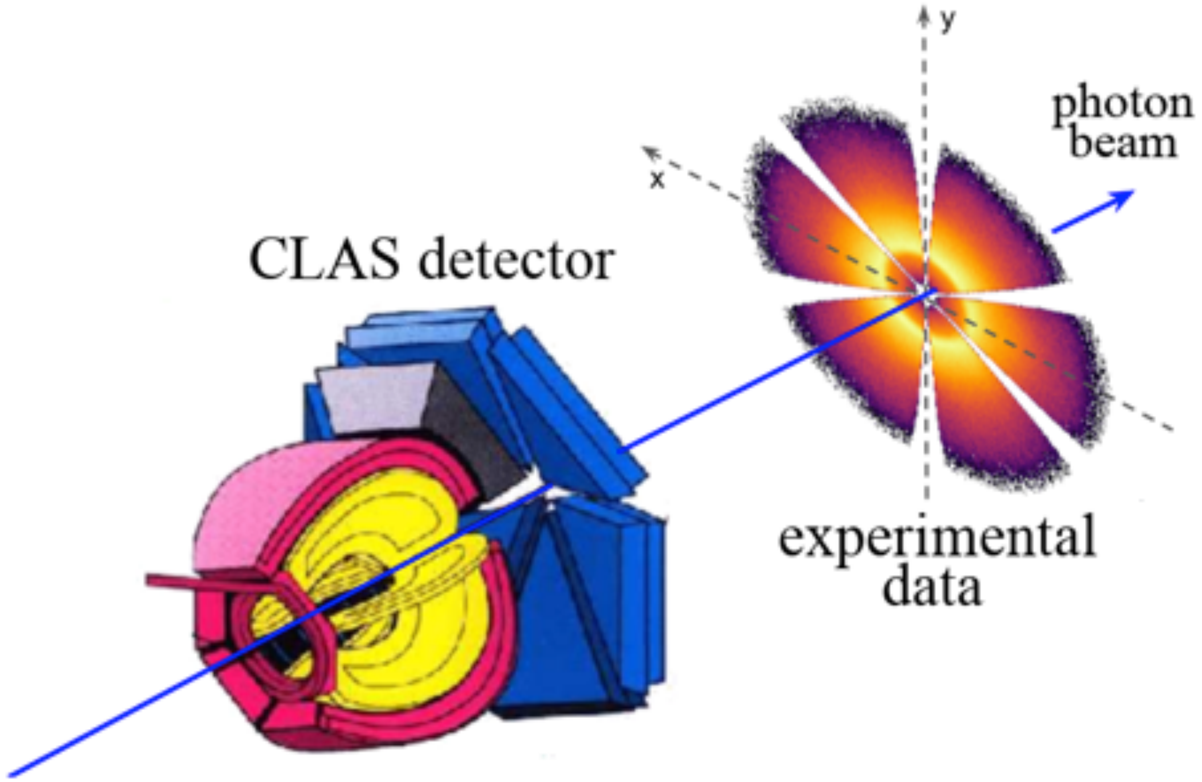


Conclusions

- The 'closure test' was successful
- 'generated' events (trained on generated)
- 'reconstructed' events (trained on reconstructed)
- 'generated' (trained on reconstructed) with larger error bars, in particular on the edge of the phase space
- The PDFs are correctly recovered

Multi-d xsec: 2π photo production

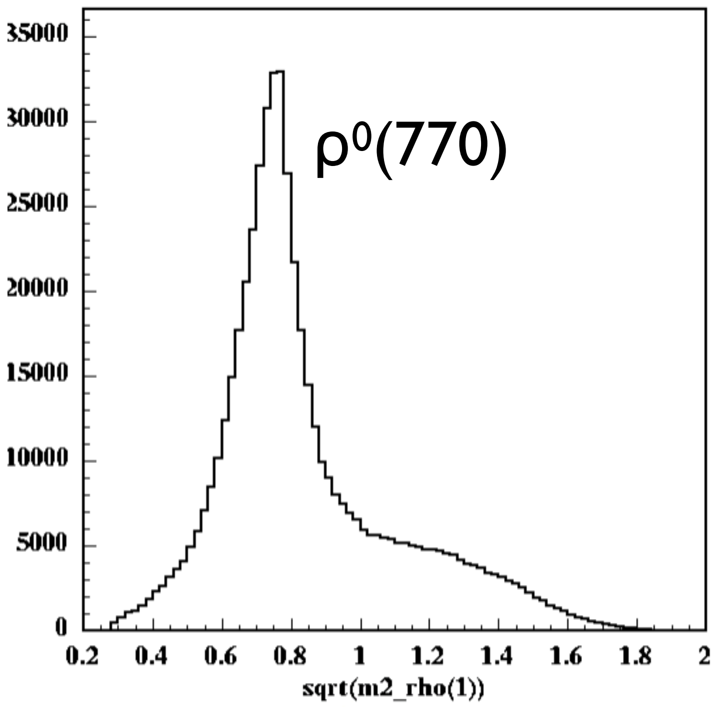
M. Battaglieri *et al.* (CLAS Collaboration) Phys. Rev. Lett. 102, 102001
M. Battaglieri *et al.* (CLAS Collaboration) Phys. Rev. D 80, 072005



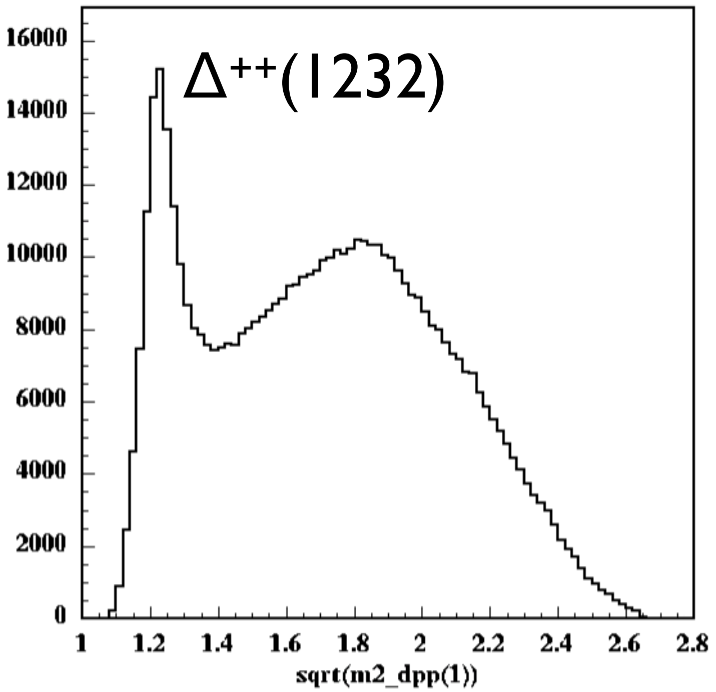
CLAS g11 kinematics

- Data set used by CLAS Collaboration for many publications
- Fiducial cuts (p, Θ, ϕ) as used in published analysis
- All four topologies are available but only focused on $\gamma\, p \rightarrow p\, \pi^+ (\pi^-)$
- Final exclusive 2π state identified by missing mass technique (variables reconstructed by energy/momentum conservation)
- Multipion background comes from $\gamma\, p \rightarrow p\, \omega^0 \rightarrow p\, \pi^+ \pi^- \pi^0$
- At E_γ=3-4 GeV reaction dynamics dominated by ρ^0 photo production ($\gamma\, p \rightarrow p\, \rho^0$) and Δ^{++} resonance excitation ($\gamma\, p \rightarrow \Delta^{++} \pi^-$)

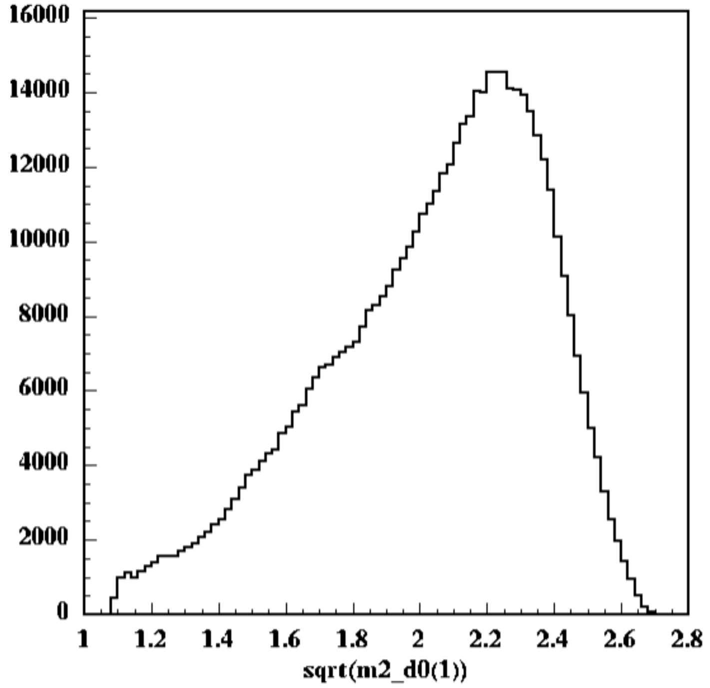
M(ππ)



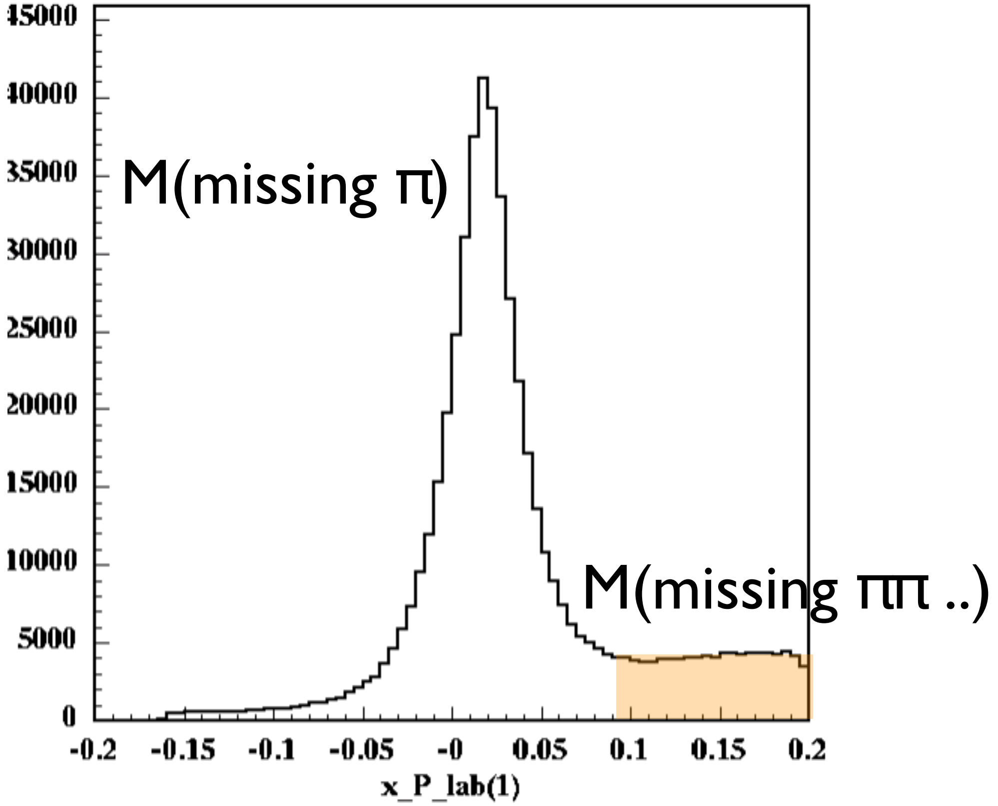
M(pπ+)



M(pπ-)



M(γ p → p π+ X)



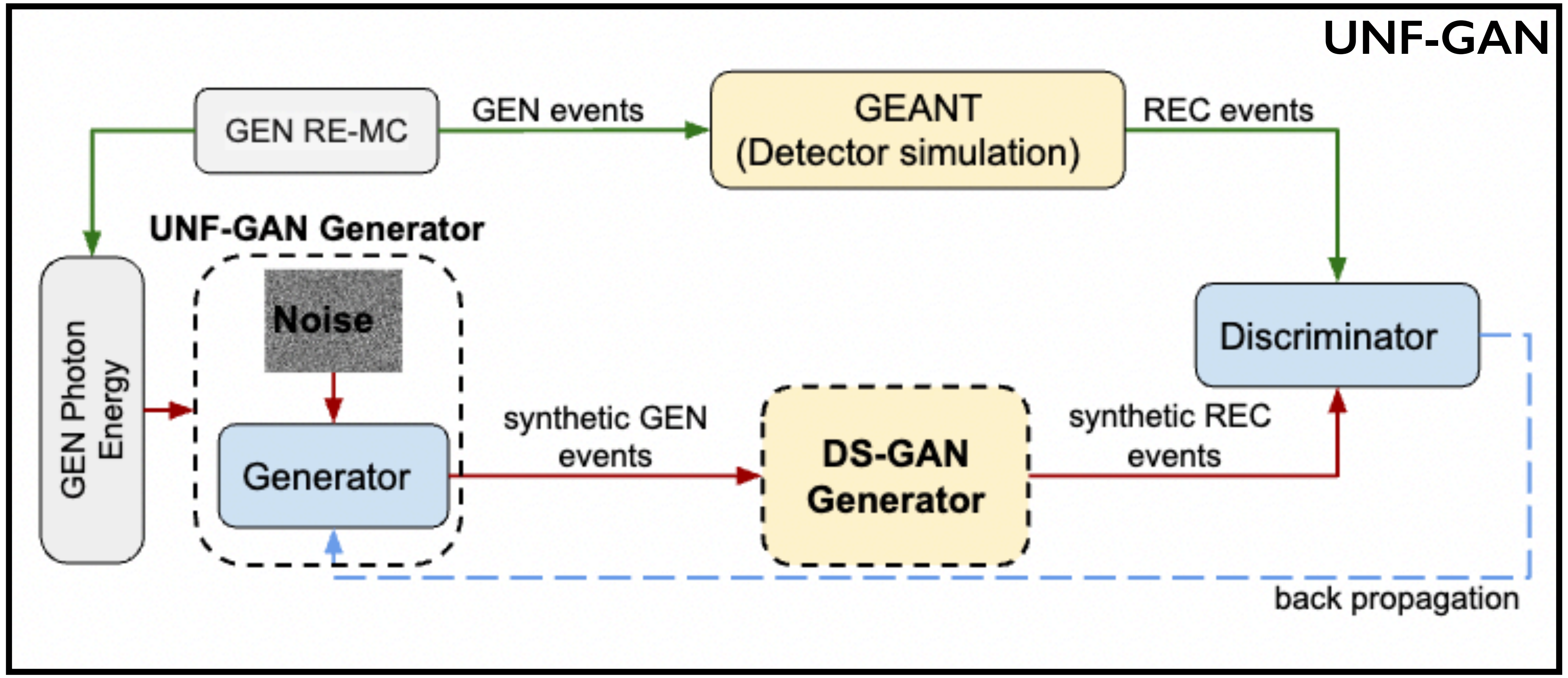
2π photo production closure test

T.Alghamdi, M.Battaglieri, A.Golda, A. Hiller Blin, L.Marsicano, W.Melnitchouk, G.Montaña, E.Isupov, Y.Li, V.Mokeev, A.Pilloni, N.Sato, A.Szczepaniak, T.Vittorini, Y.Alanazi *to appear in ArXiv*

CLOSURE TEST:
Demonstrate GANs reproduce 'true' multi-dim correlations, unfolding CLAS detector effects, comparing vertex-level (GEN) events with GANs GEN SYNT events, trained at detector-level and unfolded with a (GANs-based) detector proxy

1. Generate events with a (realistic) Monte Carlo 2π photo production model (RE-MC GEN pseudodata)
2. Apply detector effects (acceptance and resolution) via GSIM-GEANT (RE-MC REC pseudodata)
3. Deploy a secondary GAN (DS-GAN) to learn detector effects using an independent MC event generator (PS-MC) + GSIM-GEANT (GEN and REC pseudodata)
4. Deploy the unfolding GAN (UNF-GAN) that includes the DS-GAN, and train it with RE-MC REC pseudodata
5. Compare UNF-GAN GEN SYNT data to RE-MC GEN pseudodata

[if but works, replace RE-MC REC pseudo data with CLAS data in the training to unfold the vertex-level experimental distributions]

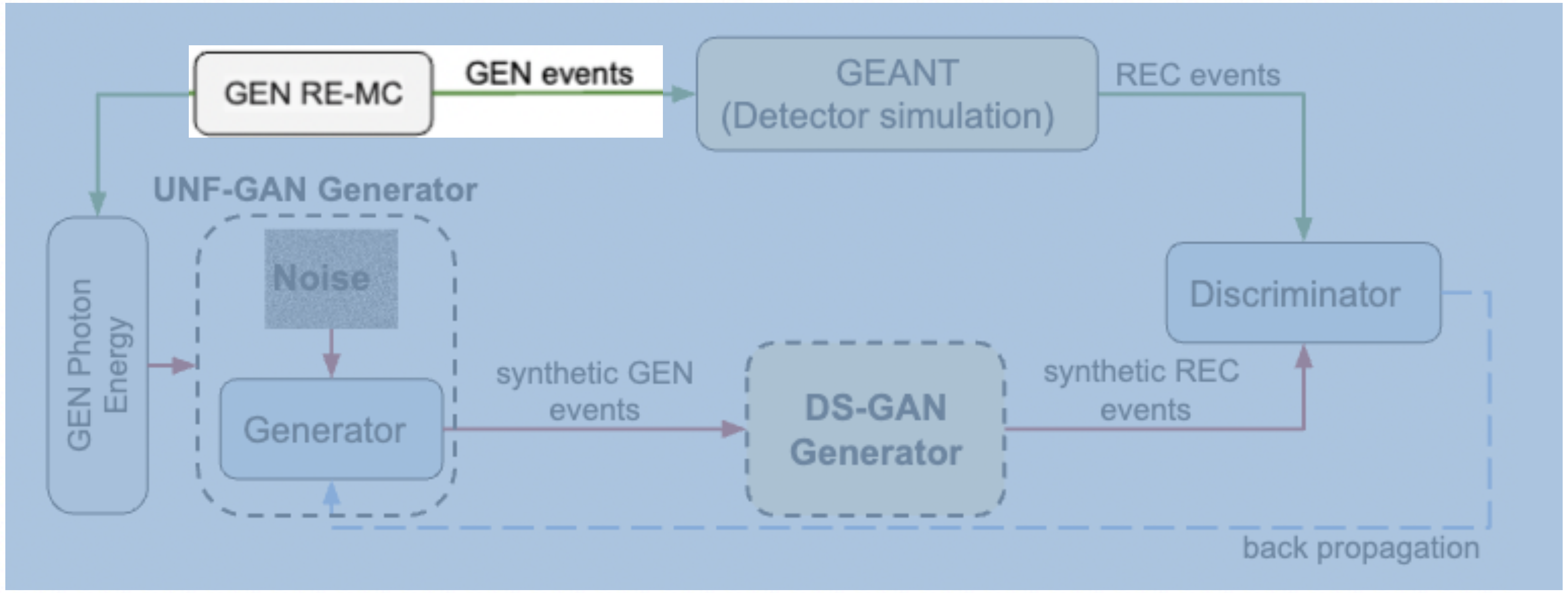


2π photo production closure test

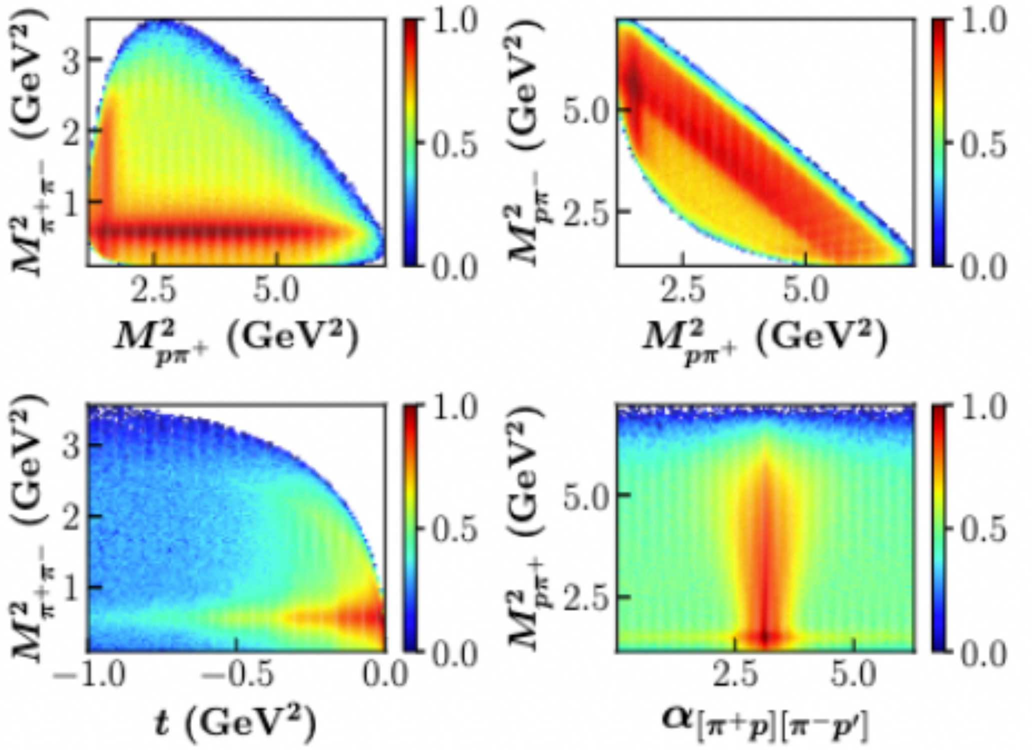
T.Alghamdi, M.Battaglieri, A.Golda, A. Hiller Blin, L.Marsicano, W.Melnitchouk, G.Montaña, E.Isupov, Y.Li, V.Mokeev, A.Pilloni, N.Sato, A.Szczepaniak, T.Vittorini, Y.Alanazi *to appear in ArXiv*

1. Generate events with a (realistic) Monte Carlo 2π photo production model (RE-MC GEN pseudodata)

- RE-MC: realistic Monte Carlo event generator to mimic real data. Includes measured xsec, angular distributions, and decay of dominant mechanisms (ρ^0, Δ^{++}, Δ^0 + a contact term)



GEN events

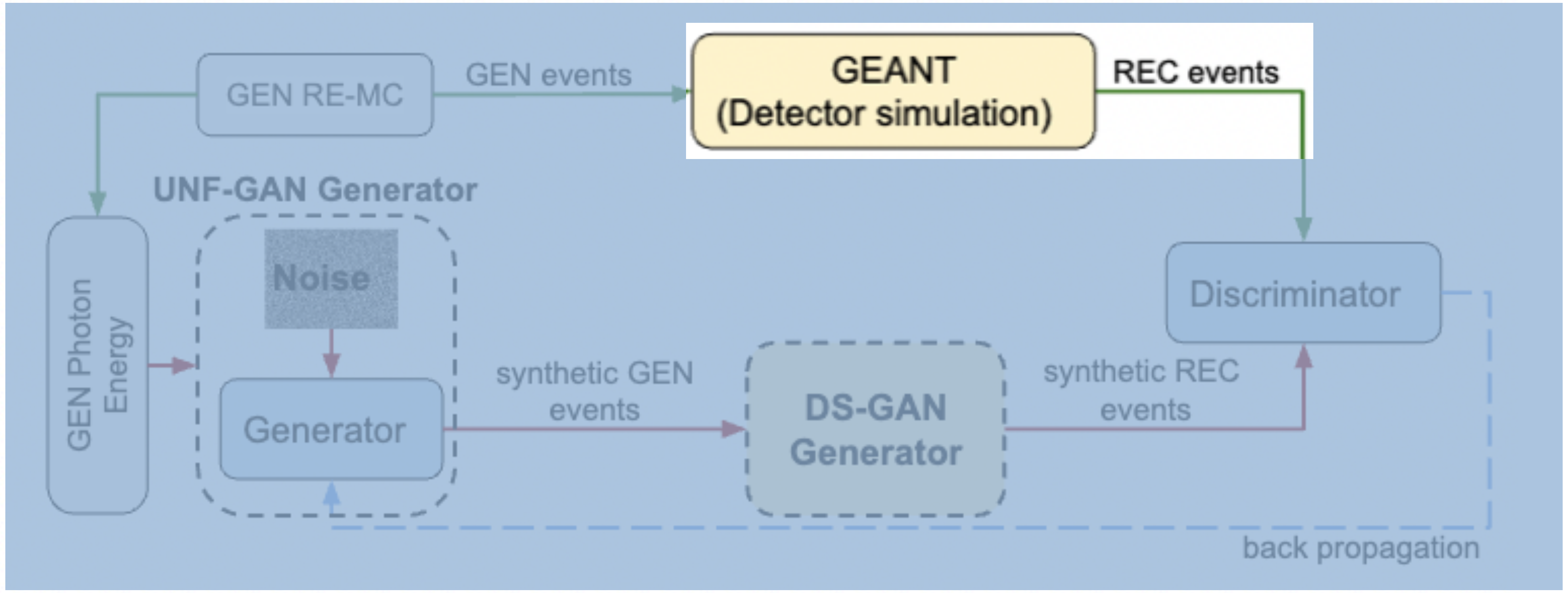


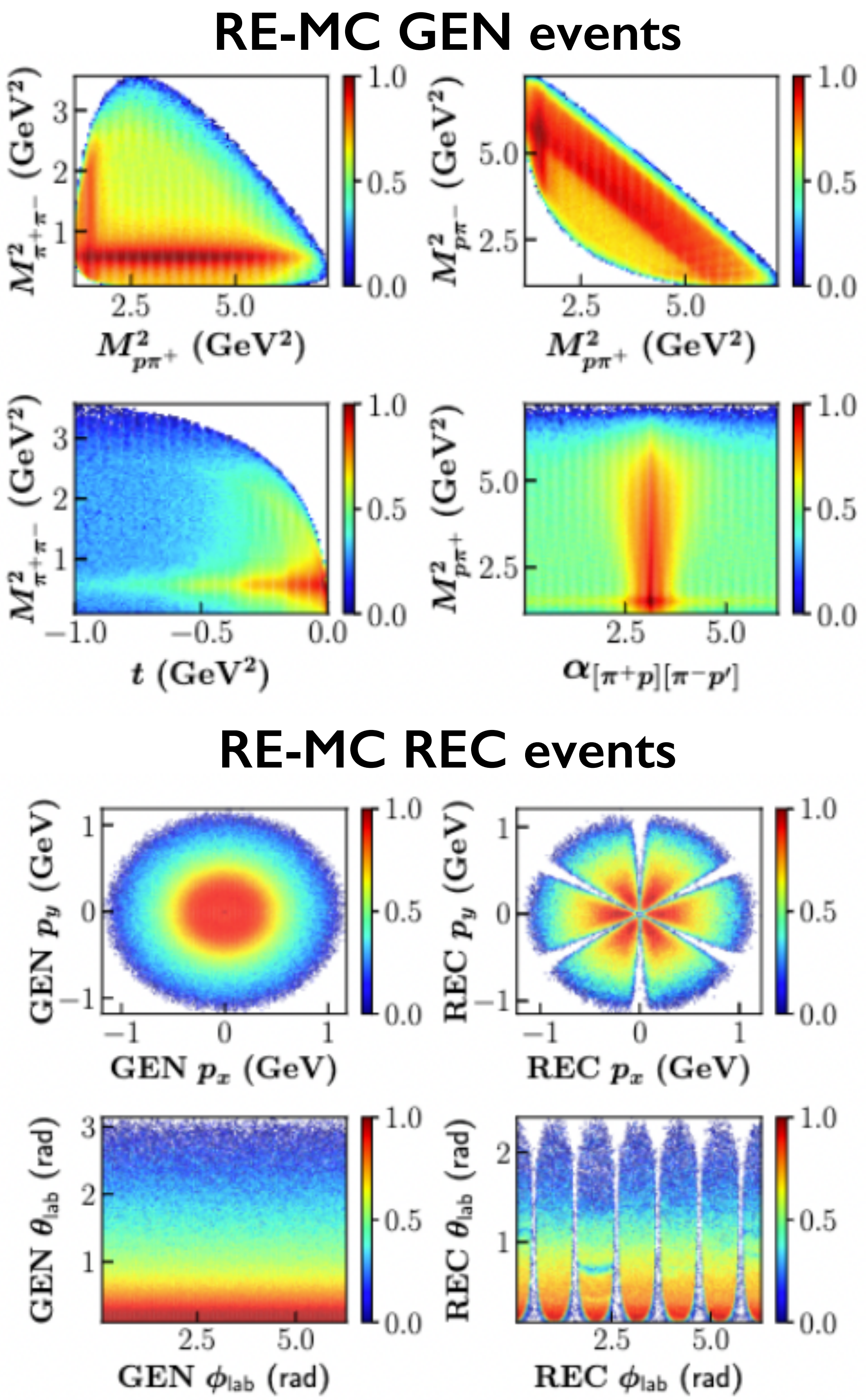
2π photo production closure test

T.Alghamdi, M.Battaglieri, A.Golda, A. Hiller Blin, L.Marsicano, W.Melnitchouk, G.Montaña, E.Isupov, Y.Li, V.Mokeev, A.Pilloni, N.Sato, A.Szczepaniak, T.Vittorini, Y.Alanazi *to appear in ArXiv*

2. Apply detector effects (acceptance and resolution) via GSIM-GEANT (RE-MC REC pseudodata)

- GSIM: detector simulation package to simulate CLAS detector effects based on GEANT3

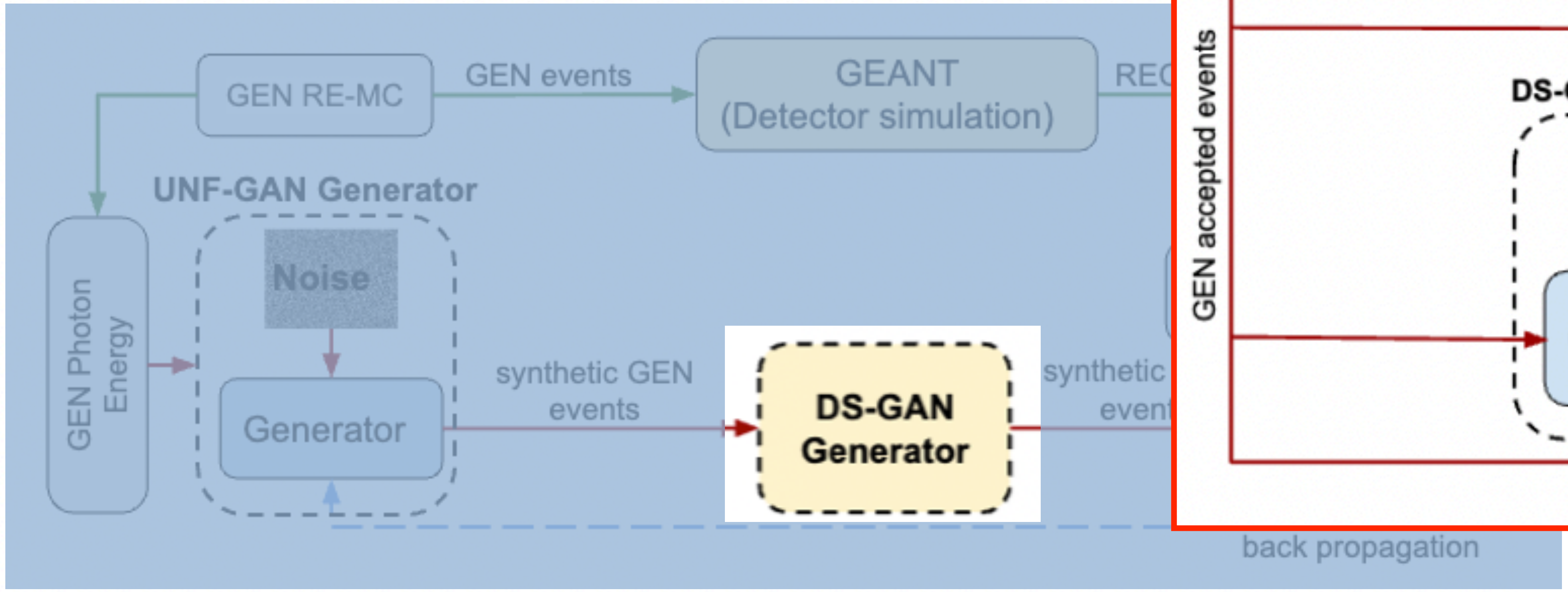


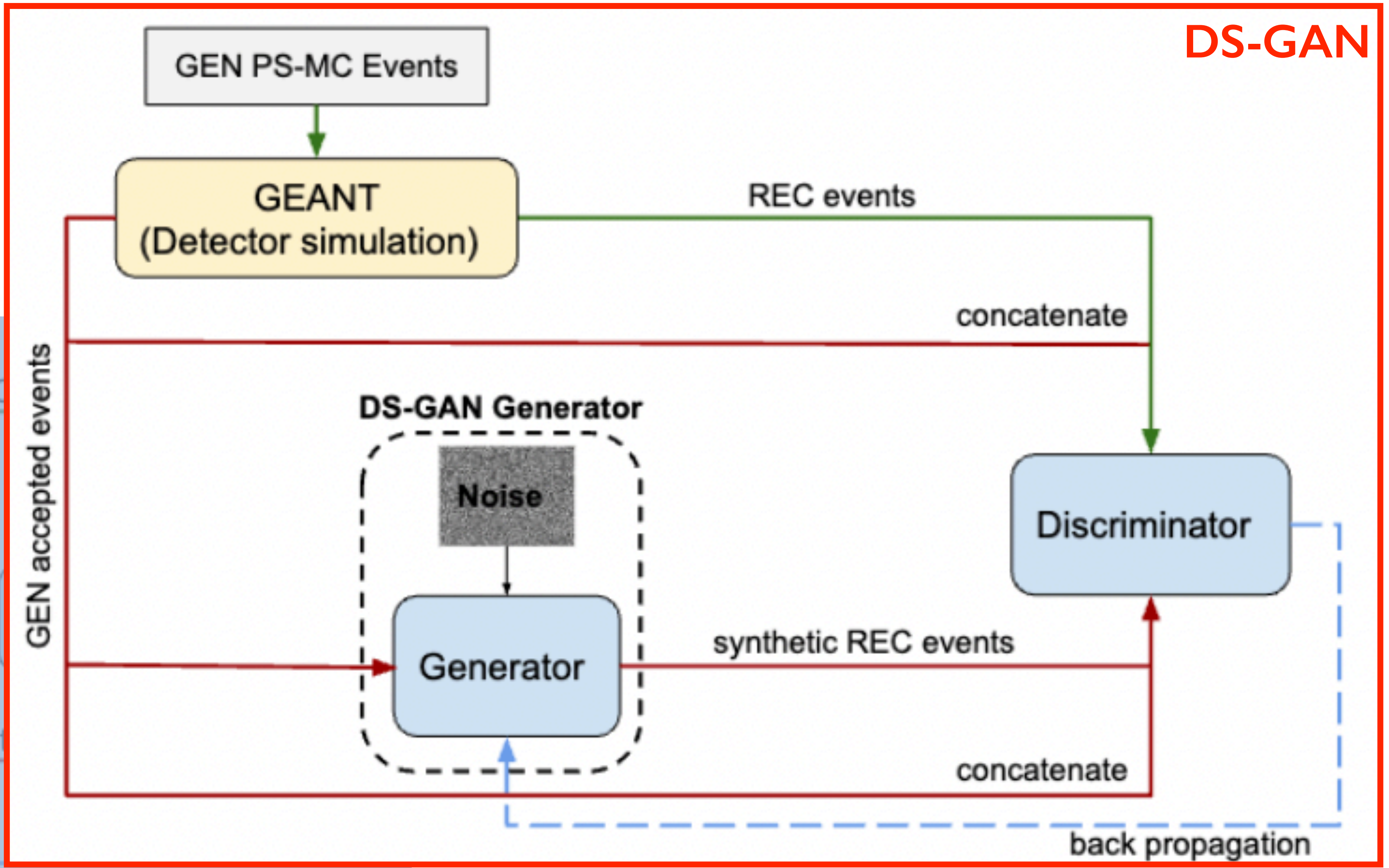


2π photo production closure test

T.Alghamdi, M.Battaglieri, A.Golda, A. Hiller Blin, L.Marsicano, W.Melnitchouk, G.Montaña, E.Isupov, Y.Li, V.Mokeev, A.Pilloni, N.Sato, A.Szczepaniak, T.Vittorini, Y.Alanazi *to appear in ArXiv*

3. Deploy a secondary GAN (DS-GAN) to learn detector effects using an independent MC event generator (PS-MC) + GSIM-GEANT (GEN and REC pseudodata)



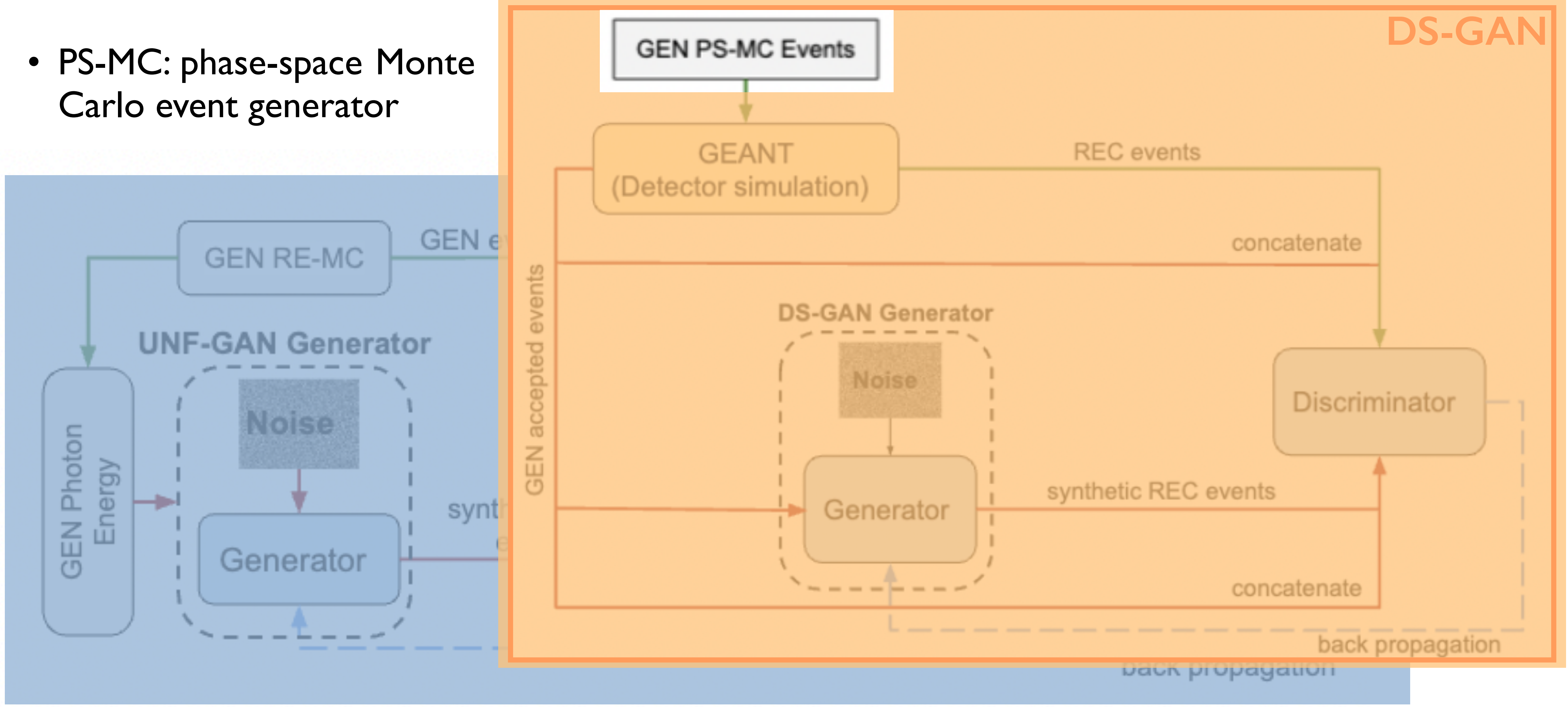


2π photo production closure test

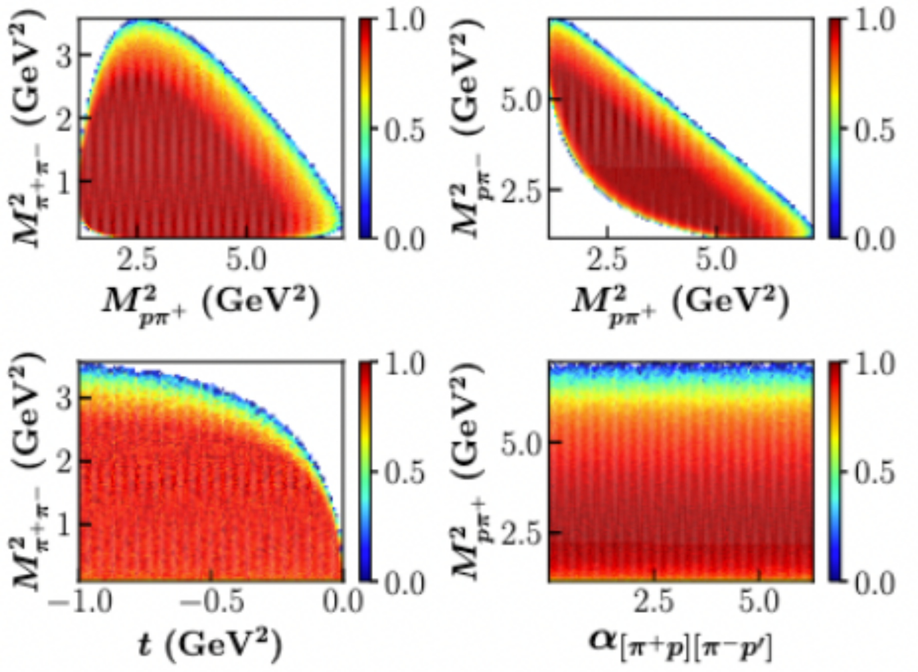
T.Alghamdi, M.Battaglieri, A.Golda, A. Hiller Blin, L.Marsicano, W.Melnitchouk, G.Montaña, E.Isupov, Y.Li, V.Mokeev, A.Pilloni, N.Sato, A.Szczepaniak, T.Vittorini, Y.Alanazi *to appear in ArXiv*

3. Deploy a secondary GAN (DS-GAN) to learn detector effects using an independent MC event generator (PS-MC) + GSIM-GEANT (GEN and REC pseudodata)

- PS-MC: phase-space Monte Carlo event generator



PS-MC GEN events

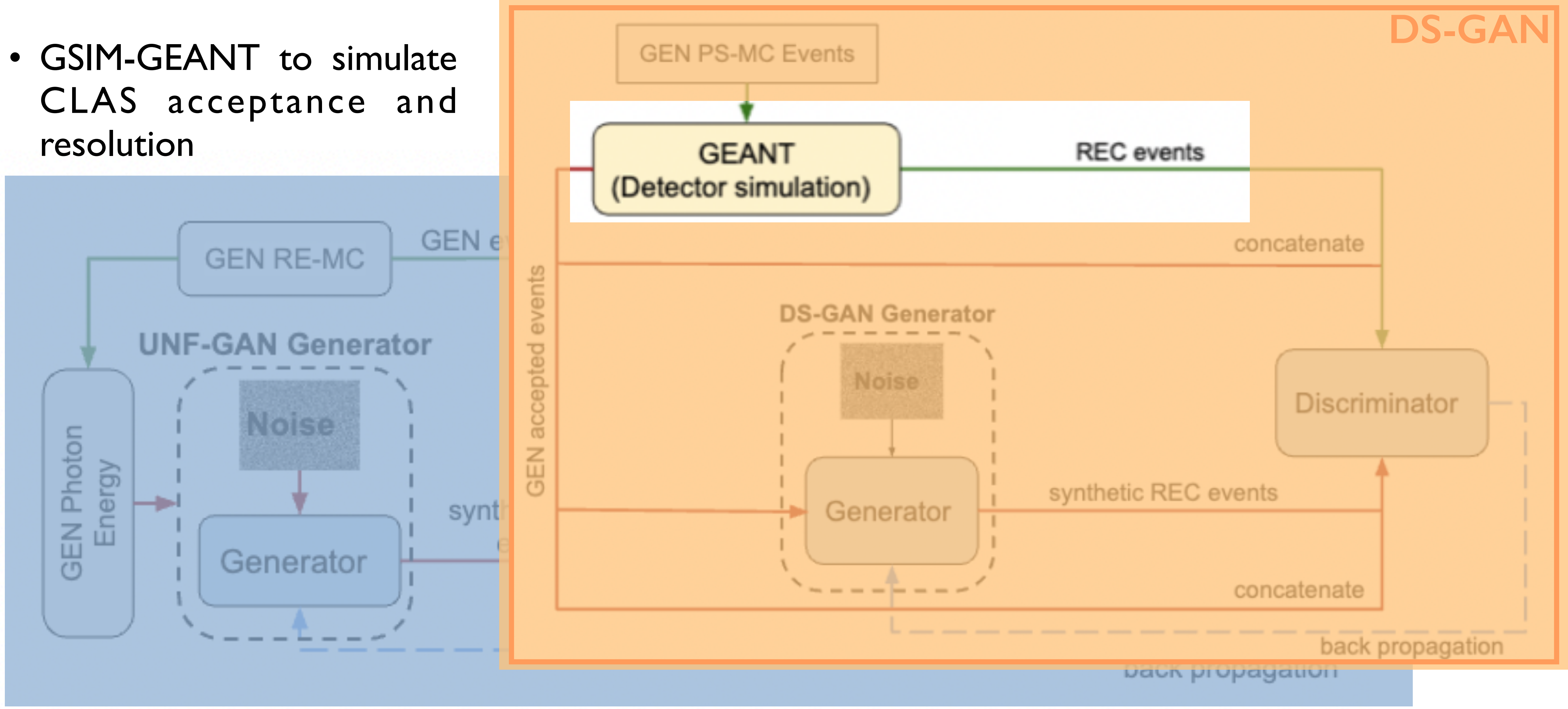


2π photo production closure test

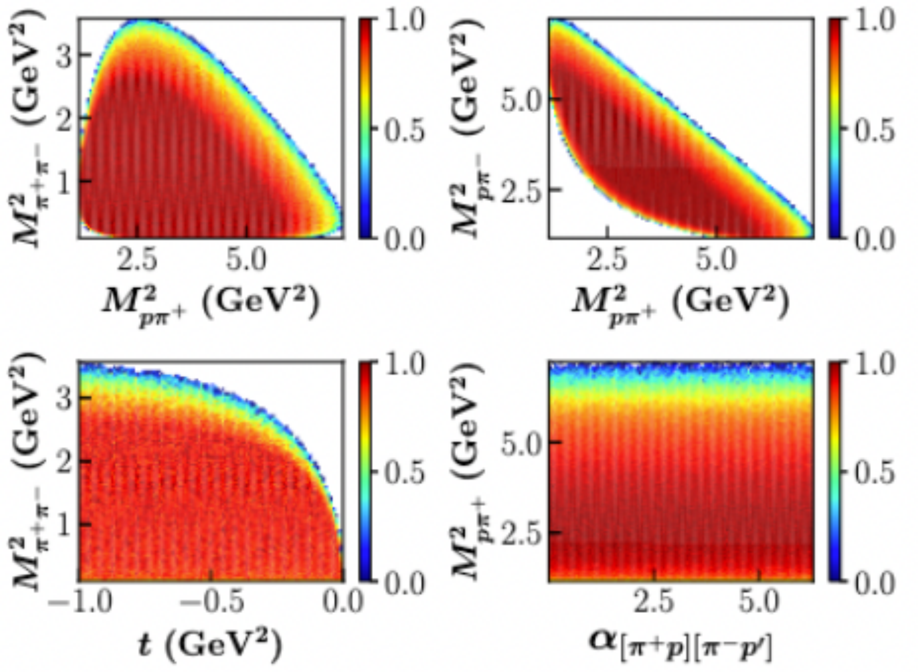
T.Alghamdi, M.Battaglieri, A.Golda, A. Hiller Blin, L.Marsicano, W.Melnitchouk, G.Montaña, E.Isupov, Y.Li, V.Mokeev, A.Pilloni, N.Sato, A.Szczepaniak, T.Vittorini, Y.Alanazi *to appear in ArXiv*

3. Deploy a secondary GAN (DS-GAN) to learn detector effects using an independent MC event generator (PS-MC) + GSIM-GEANT (GEN and REC pseudodata)

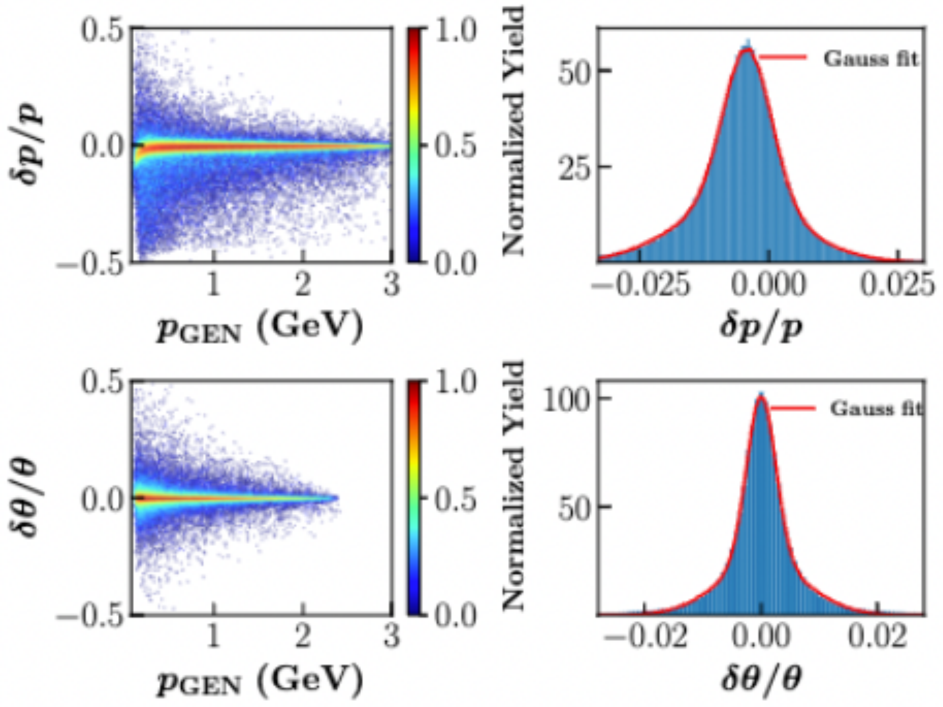
- GSIM-GEANT to simulate CLAS acceptance and resolution



PS-MC GEN events



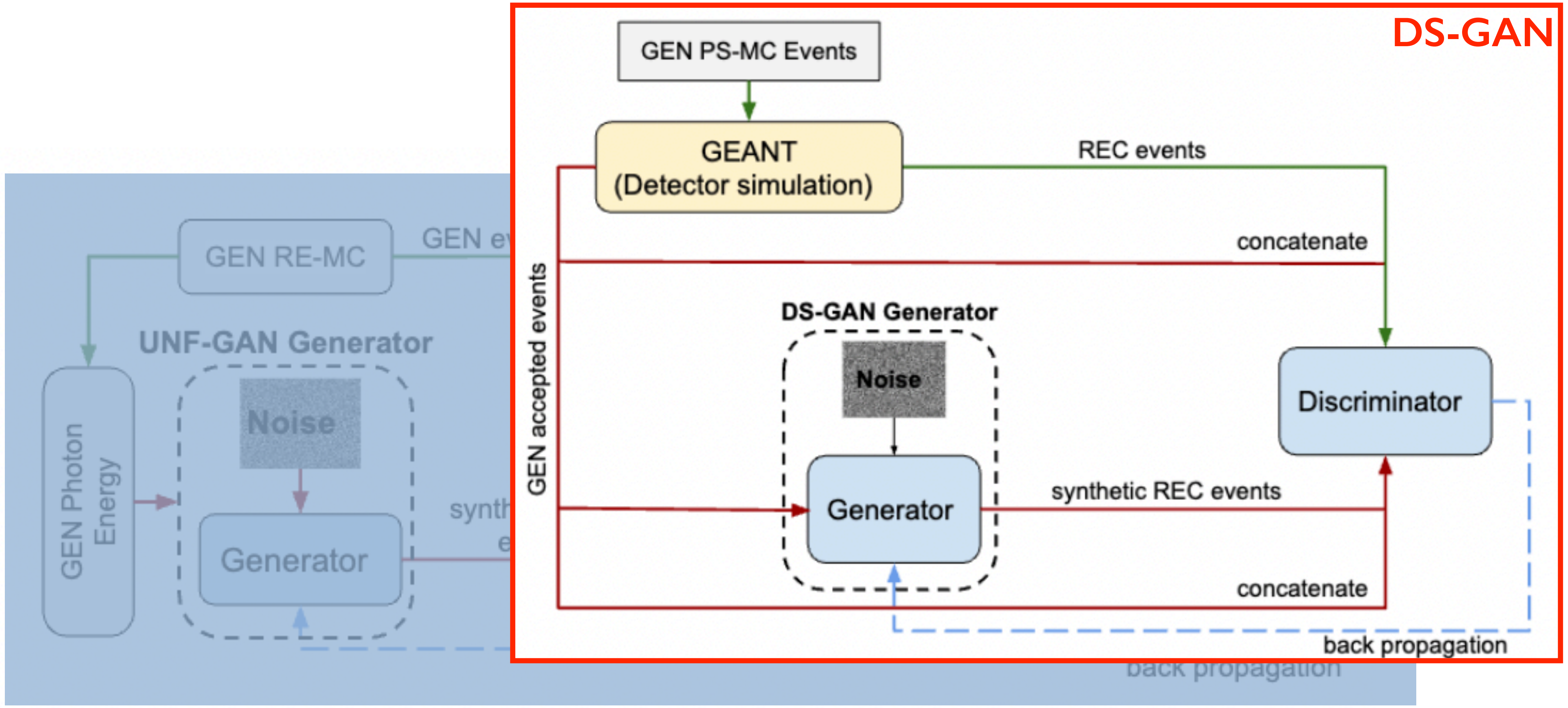
CLAS resolution



2π photo production closure test

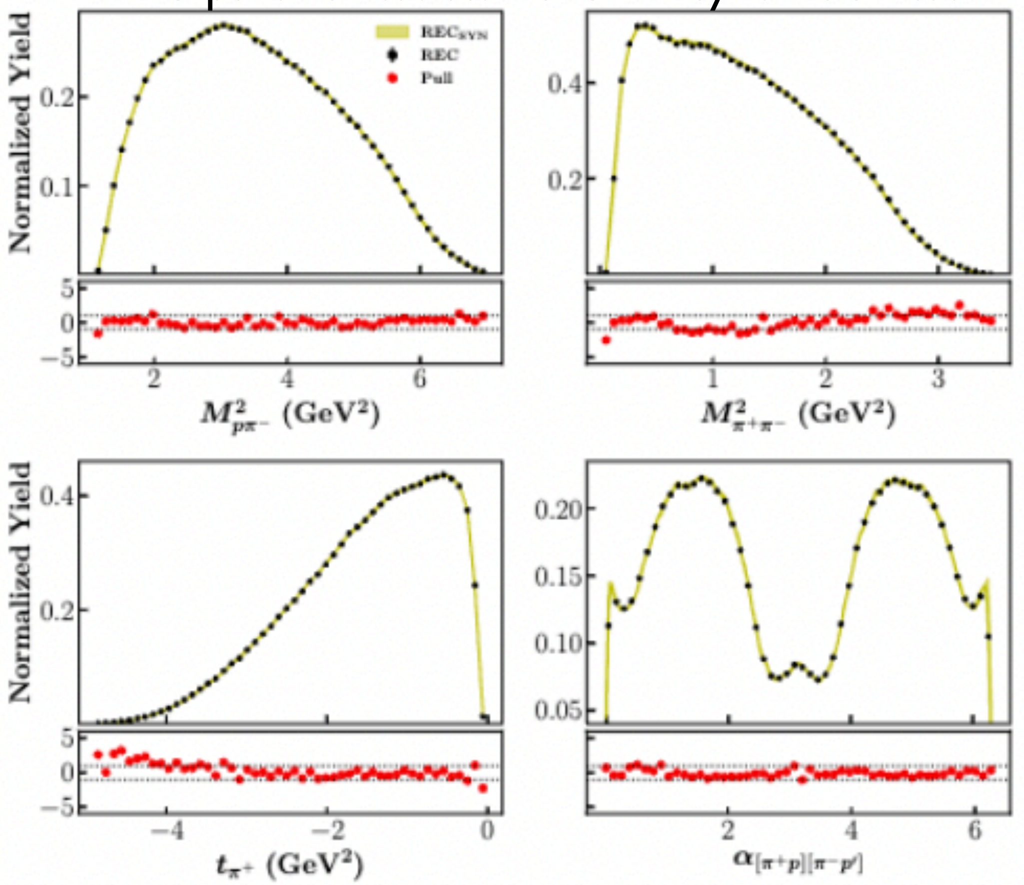
T.Alghamdi, M.Battaglieri, A.Golda, A. Hiller Blin, L.Marsicano, W.Melnitchouk, G.Montaña, E.Isupov, Y.Li, V.Mokeev, A.Pilloni, N.Sato, A.Szczepaniak, T.Vittorini, Y.Alanazi *to appear in ArXiv*

3. Deploy a secondary GAN (DS-GAN) to learn detector effects using an independent MC event generator (PS-MC) + GSIM-GEANT (GEN and REC pseudodata)

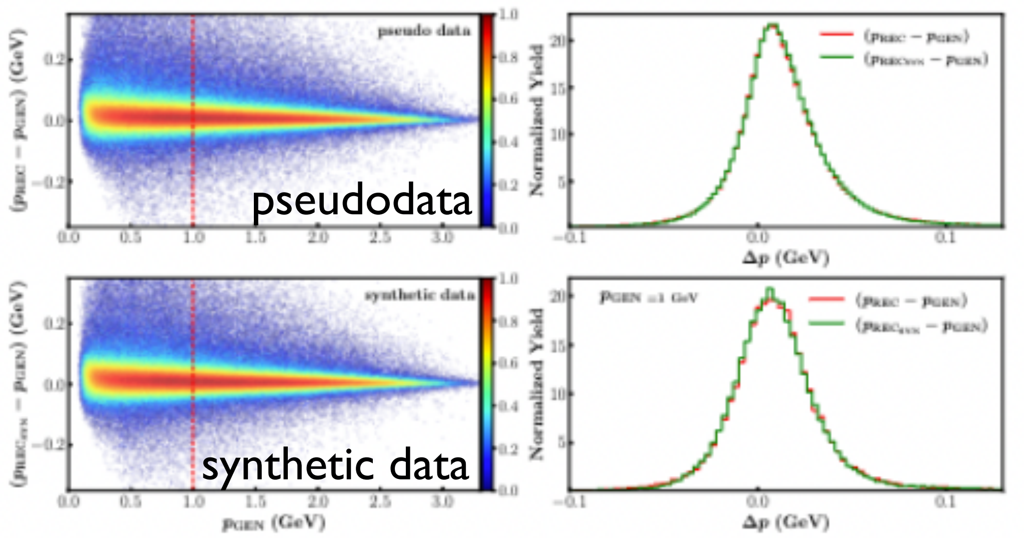


PS-MC REC events

MC pseudodata vs. DS-GAN synthetic data



CLAS resolution



Uncertainty Quantification via *pull* calculation

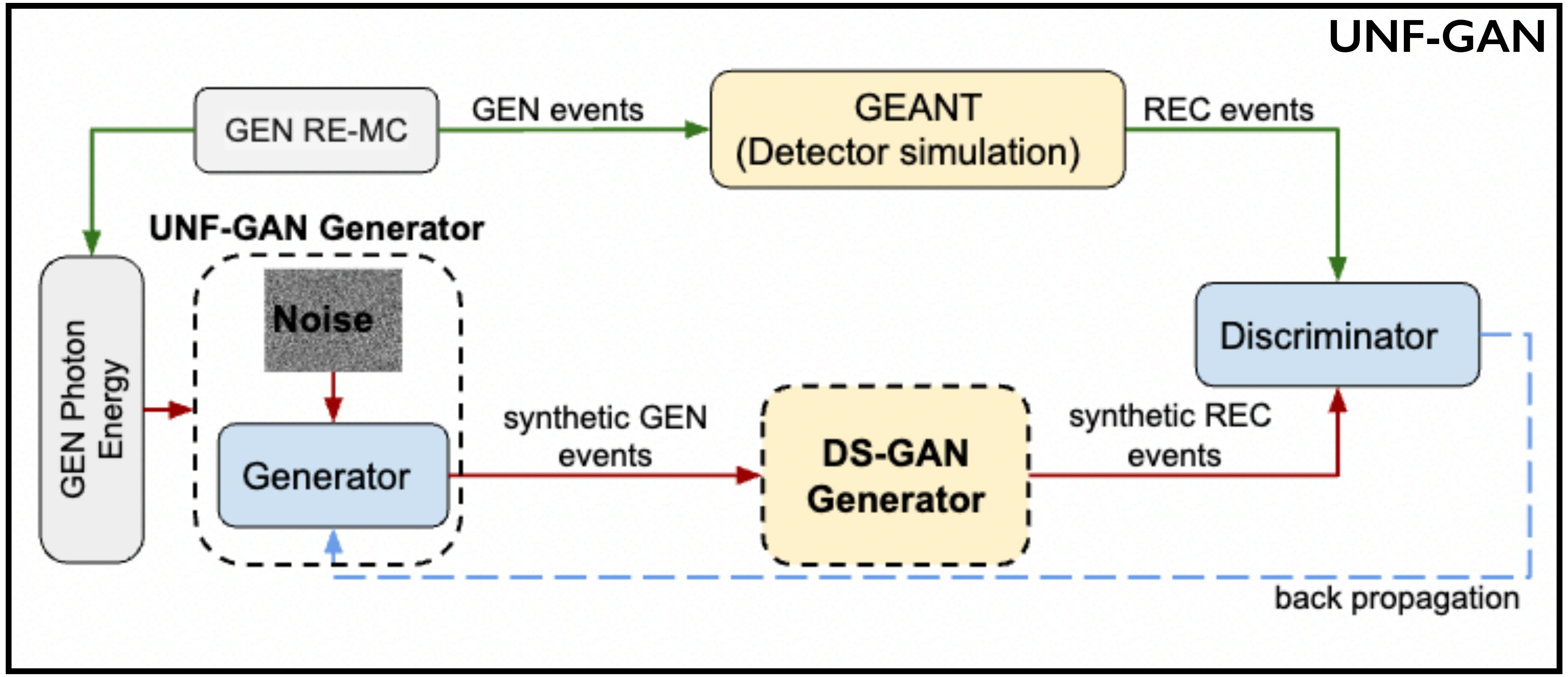
- *Bootstrap with* 20 independently trained GAN

DS-GAN learned the CLAS detector effects!

2π photo production closure test

T.Alghamdi, M.Battaglieri, A.Golda, A. Hiller Blin, L.Marsicano, W.Melnitchouk, G.Montaña, E.Isupov, Y.Li, V.Mokeev, A.Pilloni, N.Sato, A.Szczepaniak, T.Vittorini, Y.Alanazi *to appear in ArXiv*

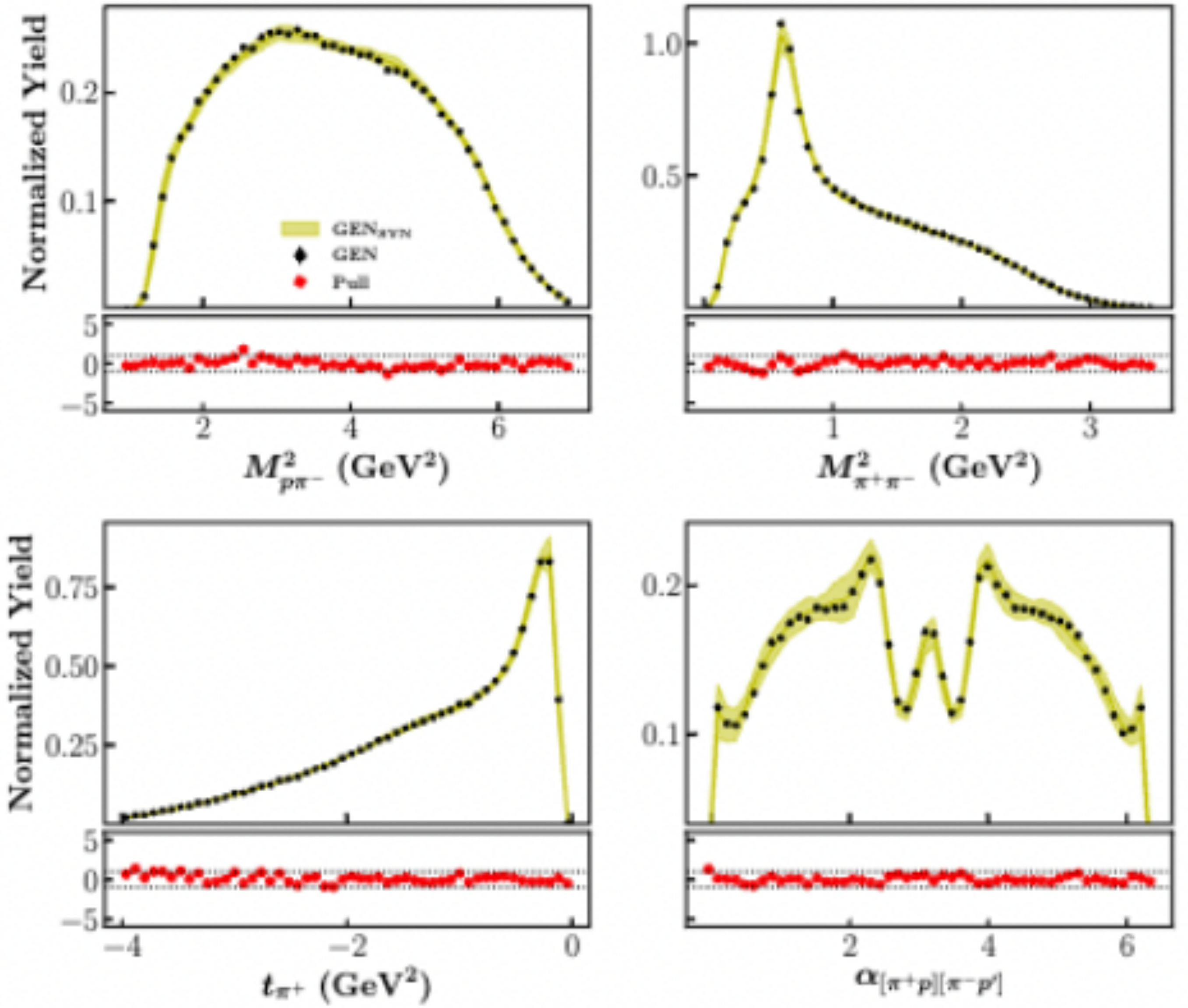
- UNF-GAN trained with RE-MC REC pseudodata (exp data proxy)
- DS-GAN used to unfold CLAS detector effects (within acceptance)



5. Compare UNF-GAN GEN SYNT data to RE-MC GEN pseudodata

Good agreement (± 1σ) at vertex-level for training variables

RE-MC GEN pseudodata vs. UNF-GAN SYN data



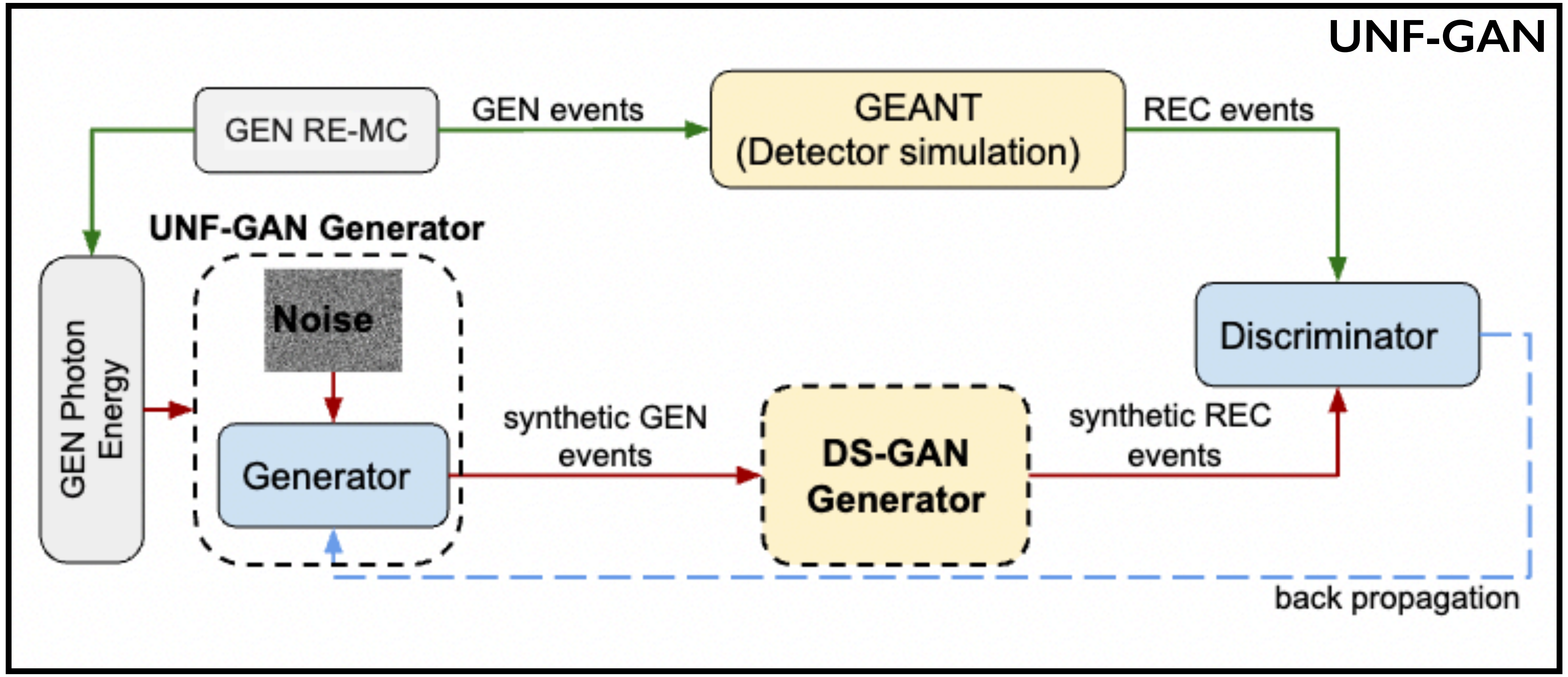
- Systematic of the full procedure (two GANs) estimated by bootstrap *with* 20+20 independently trained GANs

2π photo production closure test

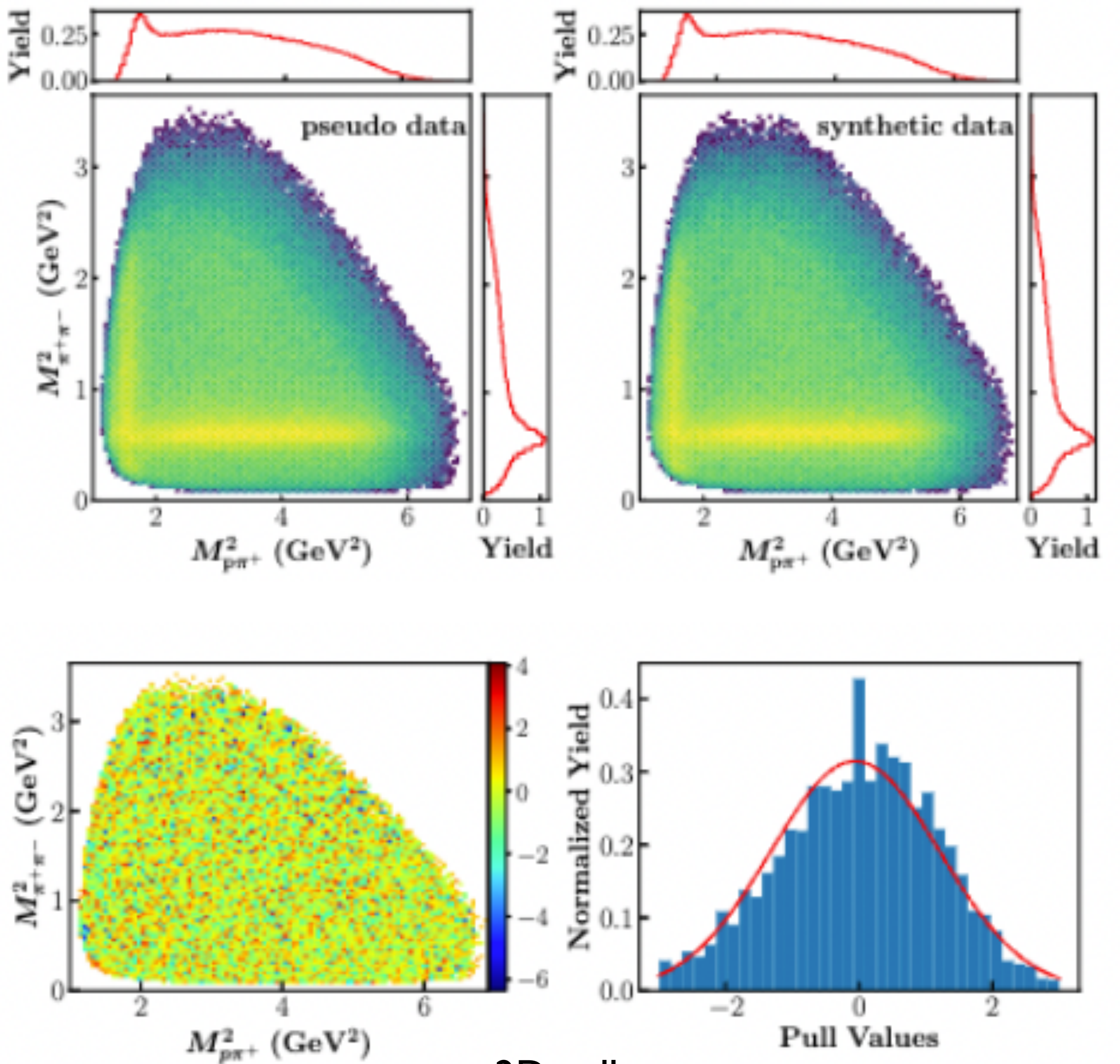
T.Alghamdi, M.Battaglieri, A.Golda, A. Hiller Blin, L.Marsicano, W.Melnitchouk, G.Montaña, E.Isupov, Y.Li, V.Mokeev, A.Pilloni, N.Sato, A.Szczepaniak, T.Vittorini, Y.Alanazi *to appear in ArXiv*

4.Deploy the unfolding GAN (UNF-GAN) that includes the DS-GAN, and train it with RE-MC REC pseudodata

- UNF-GAN trained with RE-MC REC pseudodata (exp data proxy)
- DS-GAN used to unfold CLAS detector effects (within acceptance)



RE-MC GEN pseudodata vs. UNF-GAN SYN data



2D pulls

5.Compare UNF-GAN GEN SYNT data to RE-MC GEN pseudodata

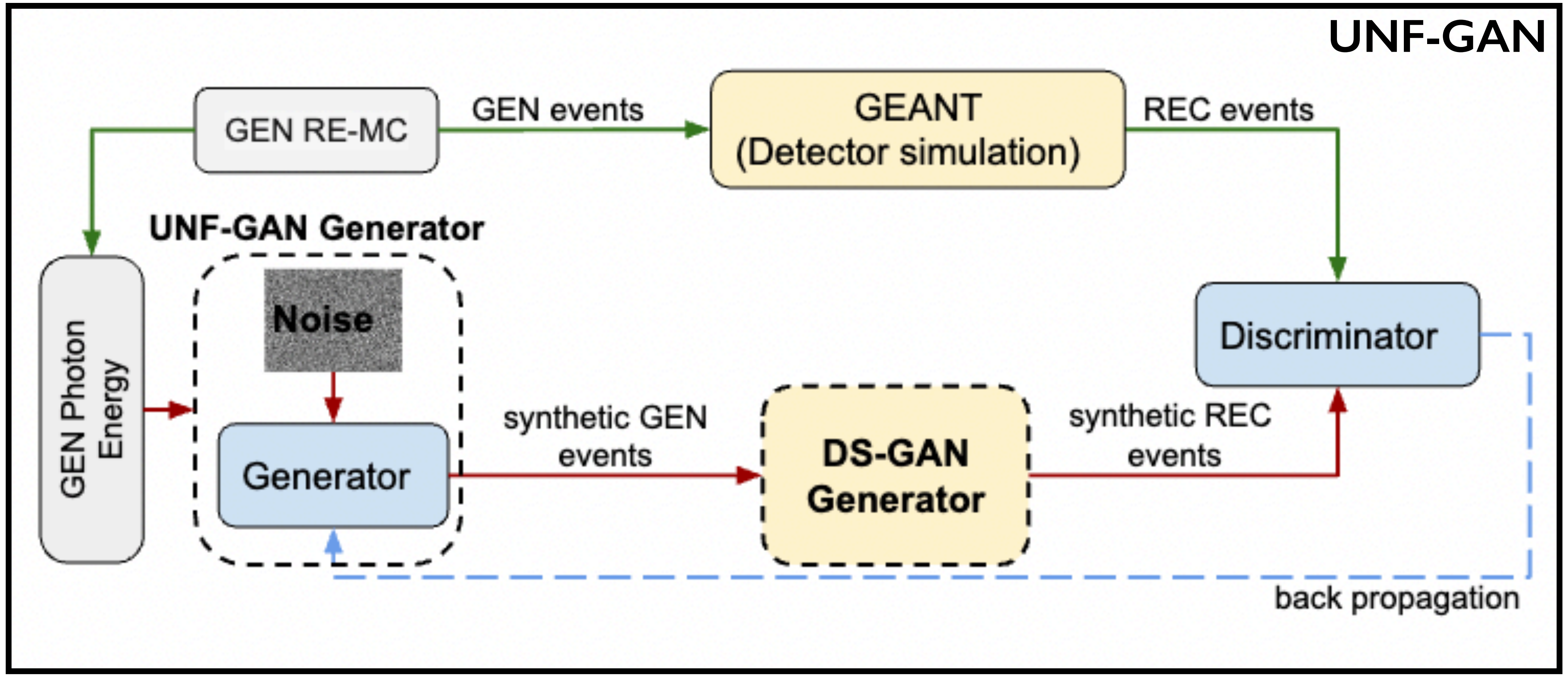
Good agreement (± 1σ) for 2D distributions (correlations)

2π photo production closure test

T.Alghamdi, M.Battaglieri, A.Golda, A. Hiller Blin, L.Marsicano, W.Melnitchouk, G.Montaña, E.Isupov, Y.Li, V.Mokeev, A.Pilloni, N.Sato, A.Szczepaniak, T.Vittorini, Y.Alanazi *to appear in ArXiv*

4.Deploy the unfolding GAN (UNF-GAN) that includes the DS-GAN, and train it with RE-MC REC pseudodata

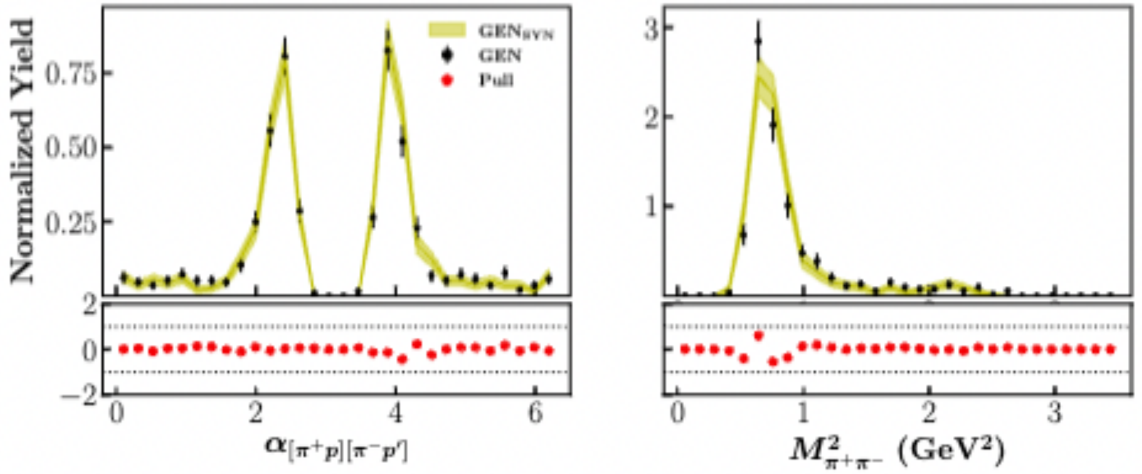
- UNF-GAN trained with RE-MC REC pseudodata (exp data proxy)
- DS-GAN used to unfold CLAS detector effects (within acceptance)



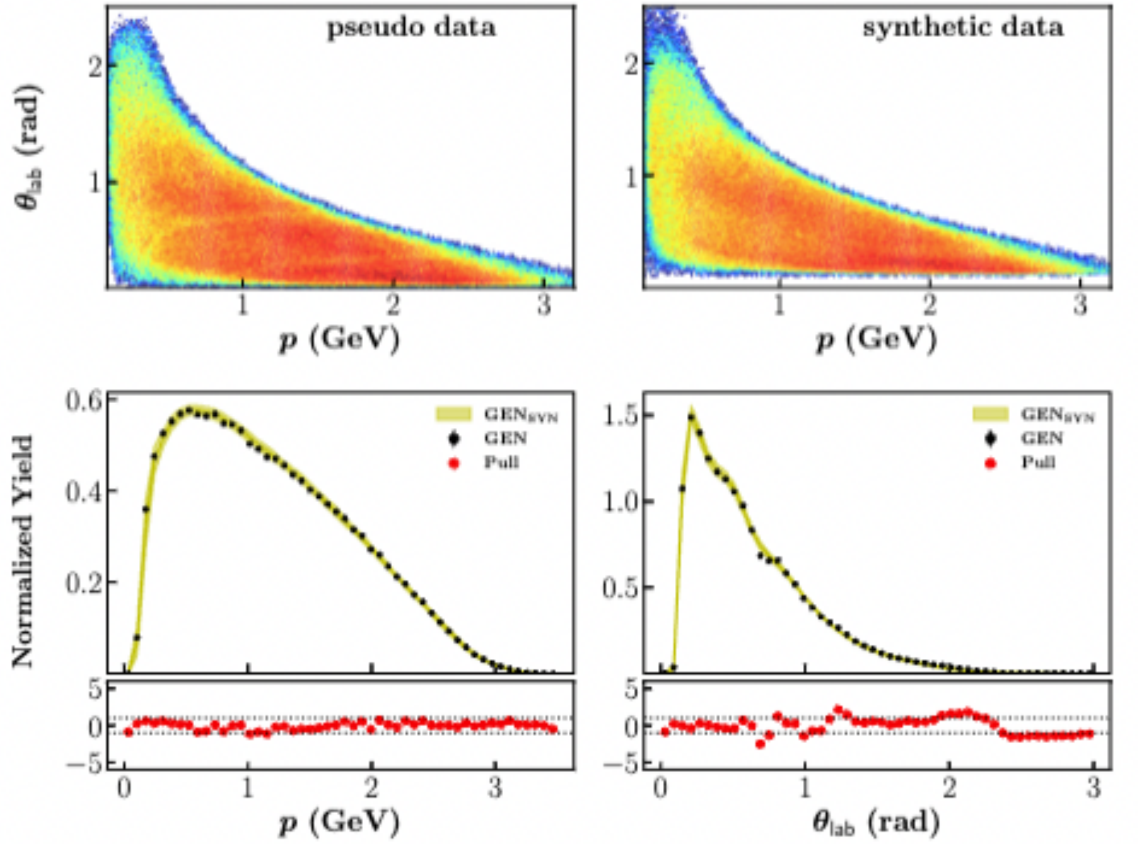
5.Compare UNF-GAN GEN SYNT data to RE-MC GEN pseudodata

Good agreement (± 1σ) for lab variables and in 4D bins

Distribution in 4D bins



RE-MC GEN pseudodata vs. UNF-GAN SYN data



This talk

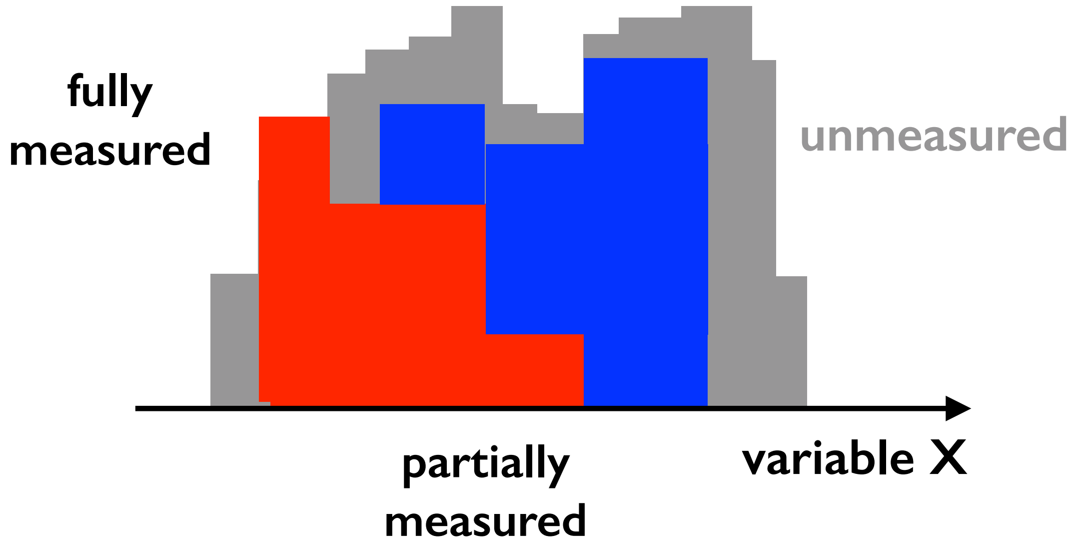
- Deploy an AI Generative Model to reproduce NP/HEP data ✓
- Detector effects unfolding: smearing ✓
- Detector effects unfolding: acceptance **in progress**
- Extract few dimensions cross-section (PDF) (e.g. inclusive electron scattering - MC) ✓
- Extend the closure test to cross-sections in a mutiD phase-space (e.g. 2-pion photoproduction - MC) ✓
- Validate the analysis procedure extracting cross-section from data (e.g. high energy CLAS-g11 2-pion data)
- Combine data of the same final state taken in different kinematics (e.g. low energy CLAS-g11 2-pion data)
- Combine data from different final states (e.g. CLAS-g11 3-pion/ω data)
- Extract cross-section and amplitudes in a 2-body reaction (e.g. $\pi\pi$ scattering - MC) **in progress**
- Extract amplitudes from a multi-particle exclusive channel (e.g. CLAS-g11 2-pion data)
- Extract amplitudes in multi- coupled-channel analysis (e.g. CLAS-g11 2-pion + 3-pion/ω data)
- Connect NN features to different physics processes (e.g. baryon and meson resonances in CLAS-g11 2-pion data)
- …
- Extract physics out of data

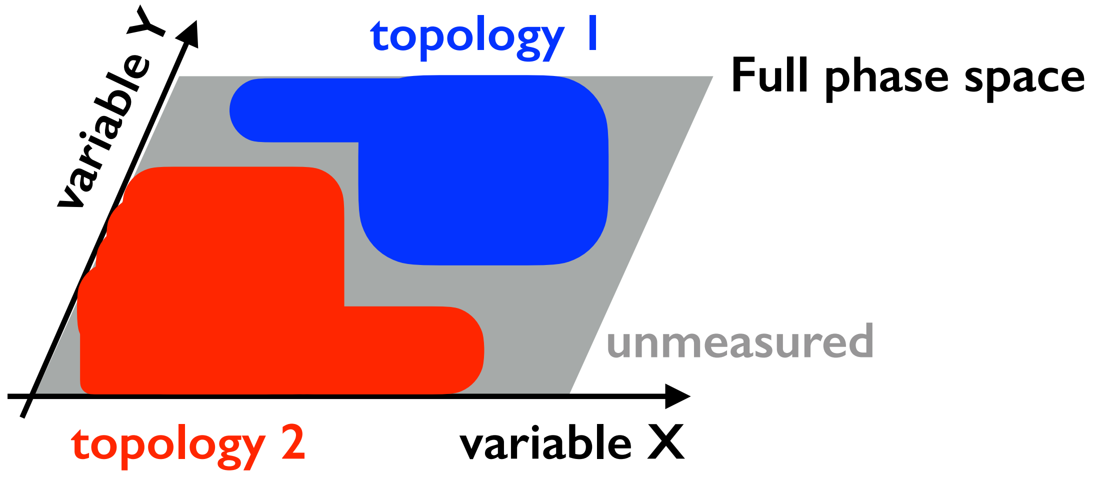
Unfolding detector's acceptance

- The reaction lives into the allowed phase space
- Detector's acceptance defines measured/unmeasured areas
- Extrapolation in the unmeasured phase space can be:

  I) model dependent
  - requires a model that matches the measured areas and extends outside
  - the extrapolation even if 'reasonable' is arbitrary

  II) model-independent
  - incorporate 'physics' in the amplitudes
  - exploits the general properties of the amplitudes (parity invariance, analyticity, unitarity, …)
  - incorporates known properties of the amplitudes
- Different combinations of particles in the final states (topologies) are mutually exclusive (not if projected on 1D)
- PDF extraction based on multiple topologies minimises the extrapolation





A(i)DAPT is working to provide:

Work in progress

- multi-topology analysis in a single framework
  - minimizes the extrapolation
- clear model-dependent extrapolation in unmeasured regions
  - quantify the model-dependence comparing different models
  - limit the model dependence to the unmeasured areas ONLY
- exploring if a model-independent extrapolation is possible

Credit: T.Vittorini, Y.Alanazi, T.Alghamdi, Y. Li

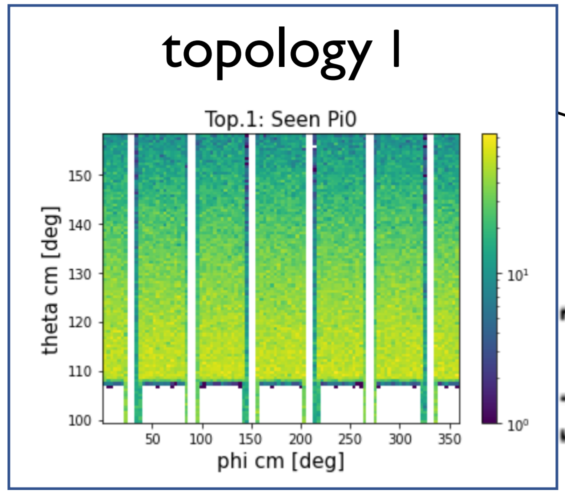
Multi-topology analysis

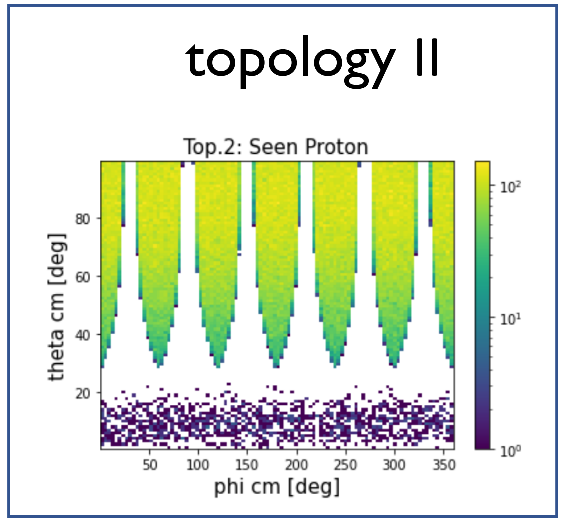
- Simple 2-body process: $\gamma p \to \Delta^+(1232) \to \pi^0 p$
- Two independent variables (at fixed energy): θ_{cm} and ϕ_{cm}
- Monte Carlo event generator
- Simple model: Breit-Wigner with two parameters: m_Δ and Γ_Δ

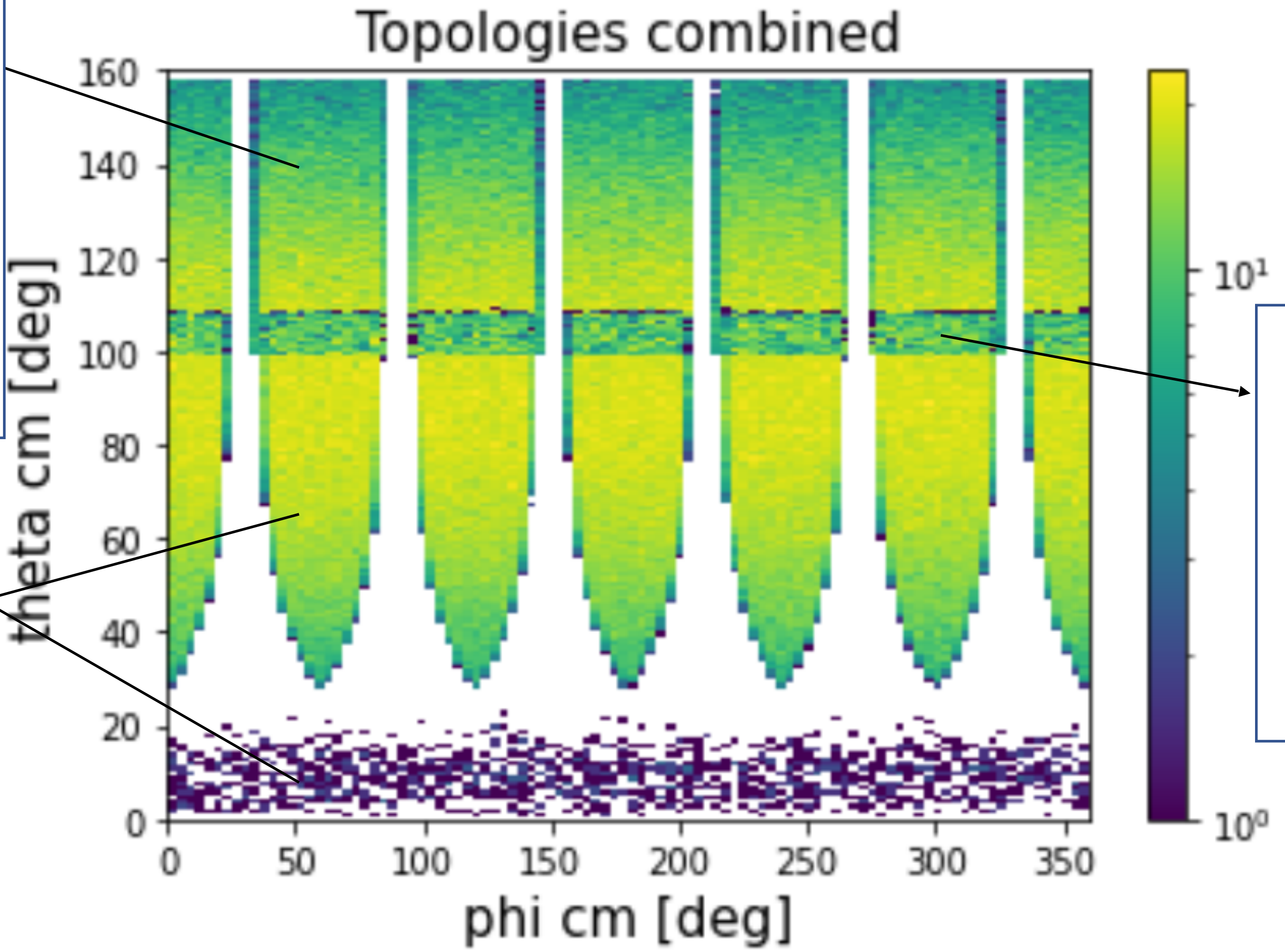
$$\frac{d\sigma}{d\Omega} \propto \frac{p_f}{p_i\, s} \sum_{\lambda_\gamma \lambda_p \lambda'_p} \left| (-)^{\lambda_\gamma} H_{|\lambda_\gamma - \lambda_p|} \frac{d^{3/2}_{\lambda_\gamma - \lambda_p, -\lambda'_p}(\theta)}{m_\Delta^2 - s - i\Gamma_\Delta m_\Delta} \right|^2$$

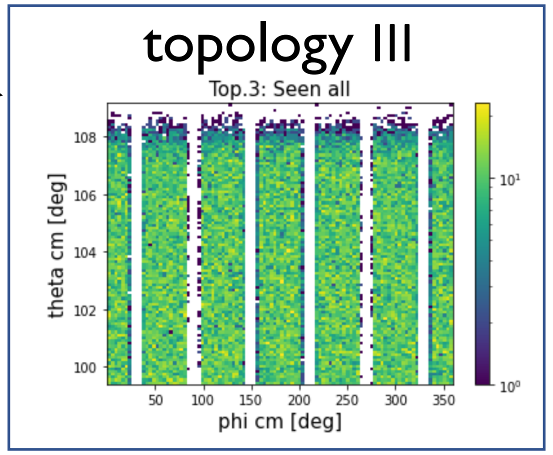
$$\propto \frac{p_f}{p_i\, s} \frac{3\left|H_{3/2}\right|^2 + 5\left|H_{1/2}\right|^2 - 3\cos 2\theta \left(\left|H_{3/2}\right|^2 - \left|H_{1/2}\right|^2\right)}{(m_\Delta^2 - s)^2 + \Gamma_\Delta^2 m_\Delta^2}$$

- Detector acceptance (CLAS) implemented via fiducial cuts (coils, minimum proton momentum and angle in the lab frame)
  - topology I: γ p → (p) π⁰ (proton missing)
  - topology II: γ p → p (π⁰) (π⁰ missing)
  - topology III: γ p → p π⁰ (all detected)
  - [topology 0: unmeasured]





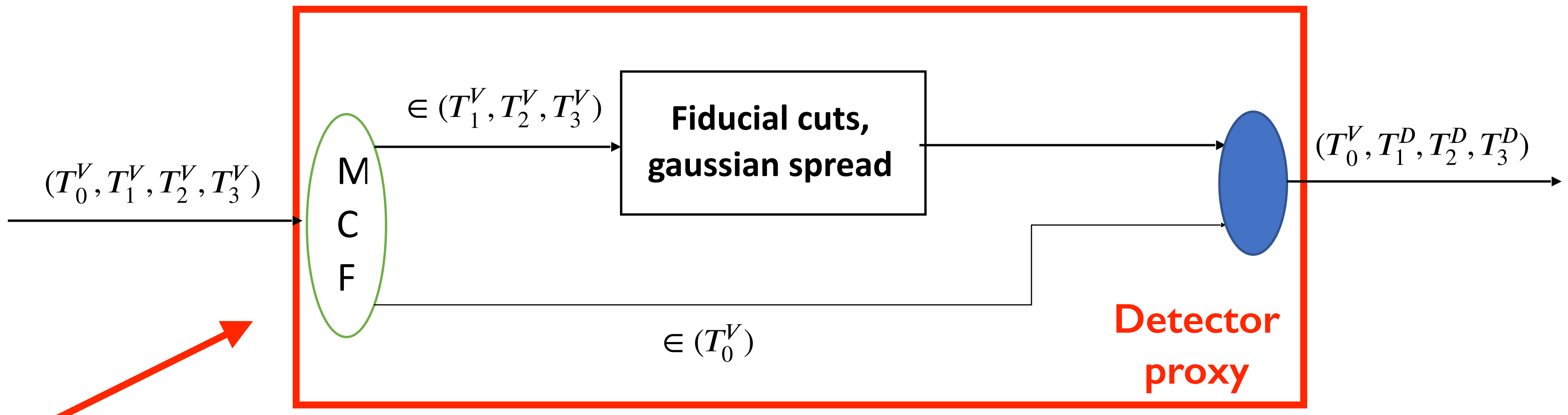


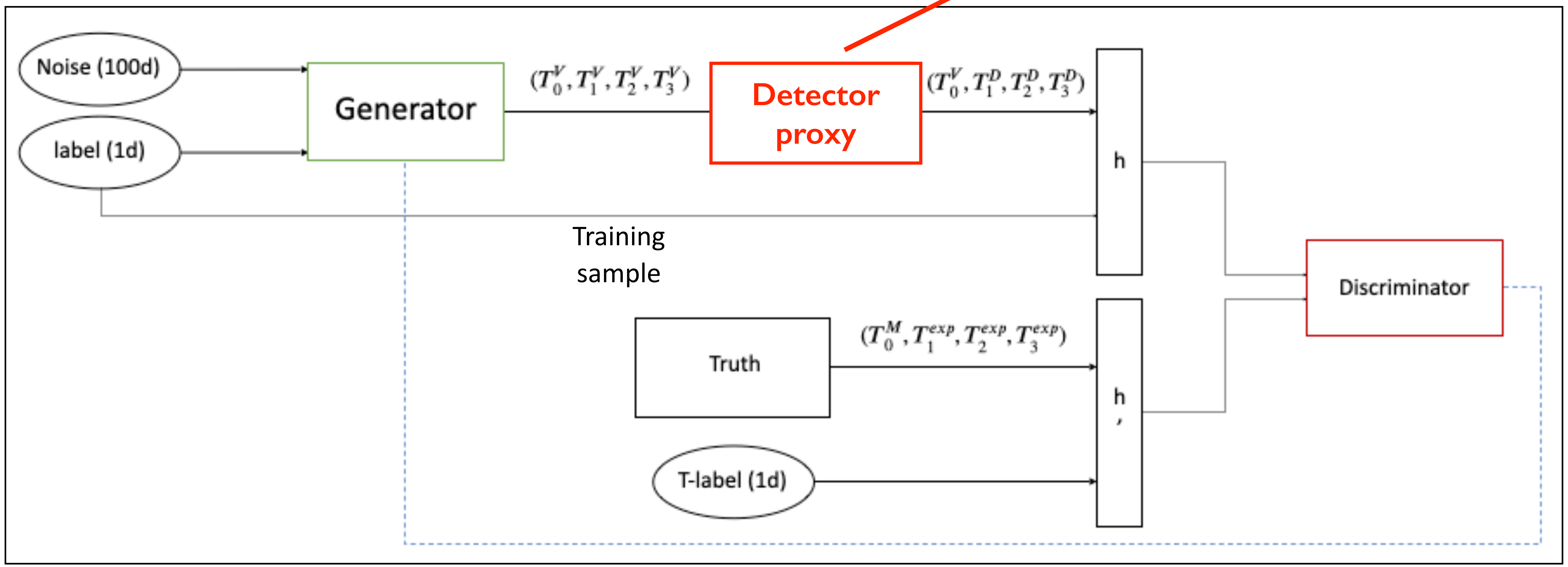


Credit: T.Vittorini ,Y.Alanazi, T.Alghamdi, Y. Li

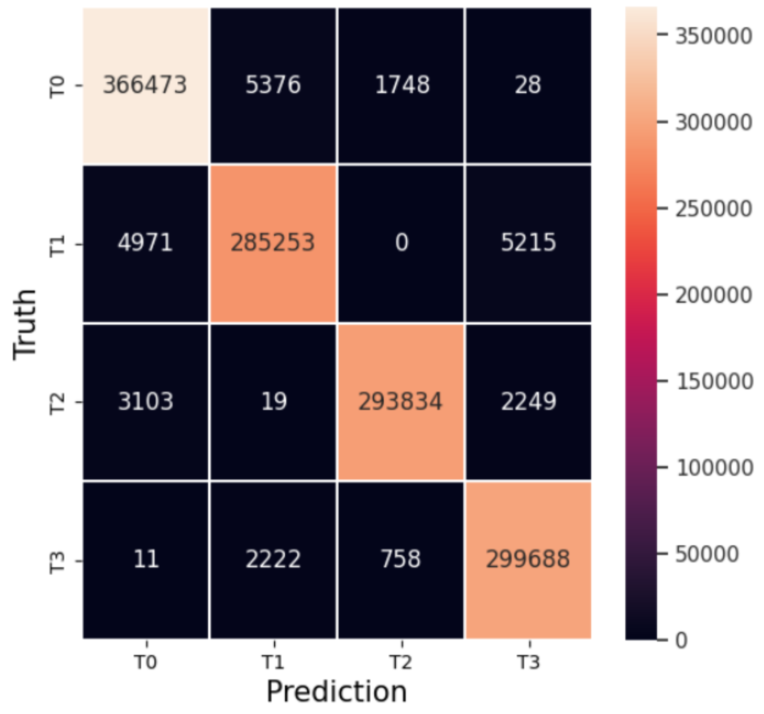
Multi-topology analysis

- GANs: generator + discriminator + multi-classifier (MCF)
- All four topologies implemented
- Training performed on REC (accepted/unmeasured) pseudodata
- Unmeasured events normalisation: $T_0^V = T_0^M(\frac{T_{tot}^{exp}}{T_{tot}^M})$





MCF accuracy ~97%



- GANs training in progress

Credit: T.Vittorini , Y.Alanazi, T.Alghamdi, Y. Li

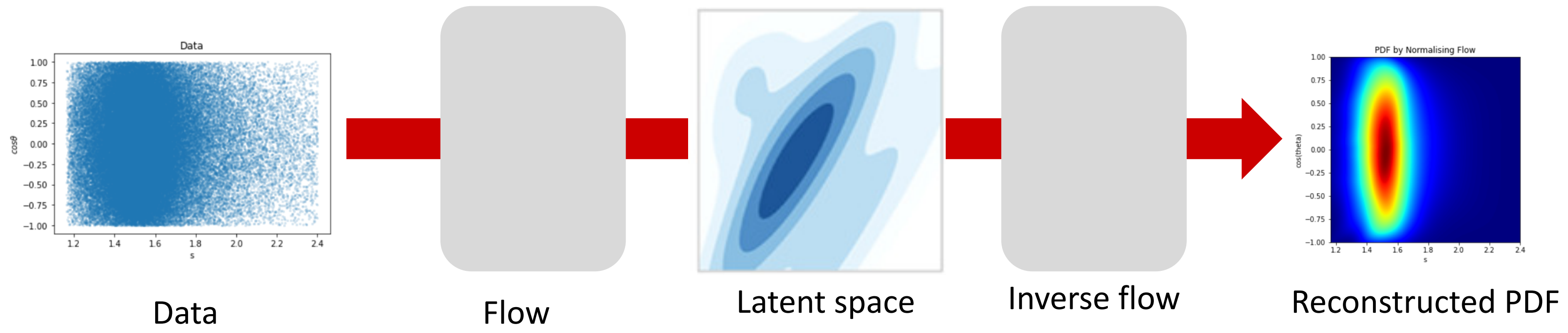
This talk

- Deploy an AI Generative Model to reproduce NP/HEP data ✓
- Detector effects unfolding: smearing ✓
- Detector effects unfolding: acceptance in progress
- Extract few dimensions cross-section (PDF) (e.g. inclusive electron scattering - MC) ✓
- Extend the closure test to cross-sections in a mutiD phase-space (e.g. 2-pion photoproduction - MC) ✓
- Validate the analysis procedure extracting cross-section from data (e.g. high energy CLAS-g11 2-pion data)
- Combine data of the same final state taken in different kinematics (e.g. low energy CLAS-g11 2-pion data)
- Combine data from different final states (e.g. CLAS-g11 3-pion/ω data)
- Extract cross-section and amplitudes in a 2-body reaction (e.g. π π scattering - MC) in progress
- Extract amplitudes from a multi-particle exclusive channel (e.g. CLAS-g11 2-pion data)
- Extract amplitudes in multi- coupled-channel analysis (e.g. CLAS-g11 2-pion + 3-pion/ω data)
- Connect NN features to different physics processes (e.g. baryon and meson resonances in CLAS-g11 2-pion data)
- …
- Extract physics out of data

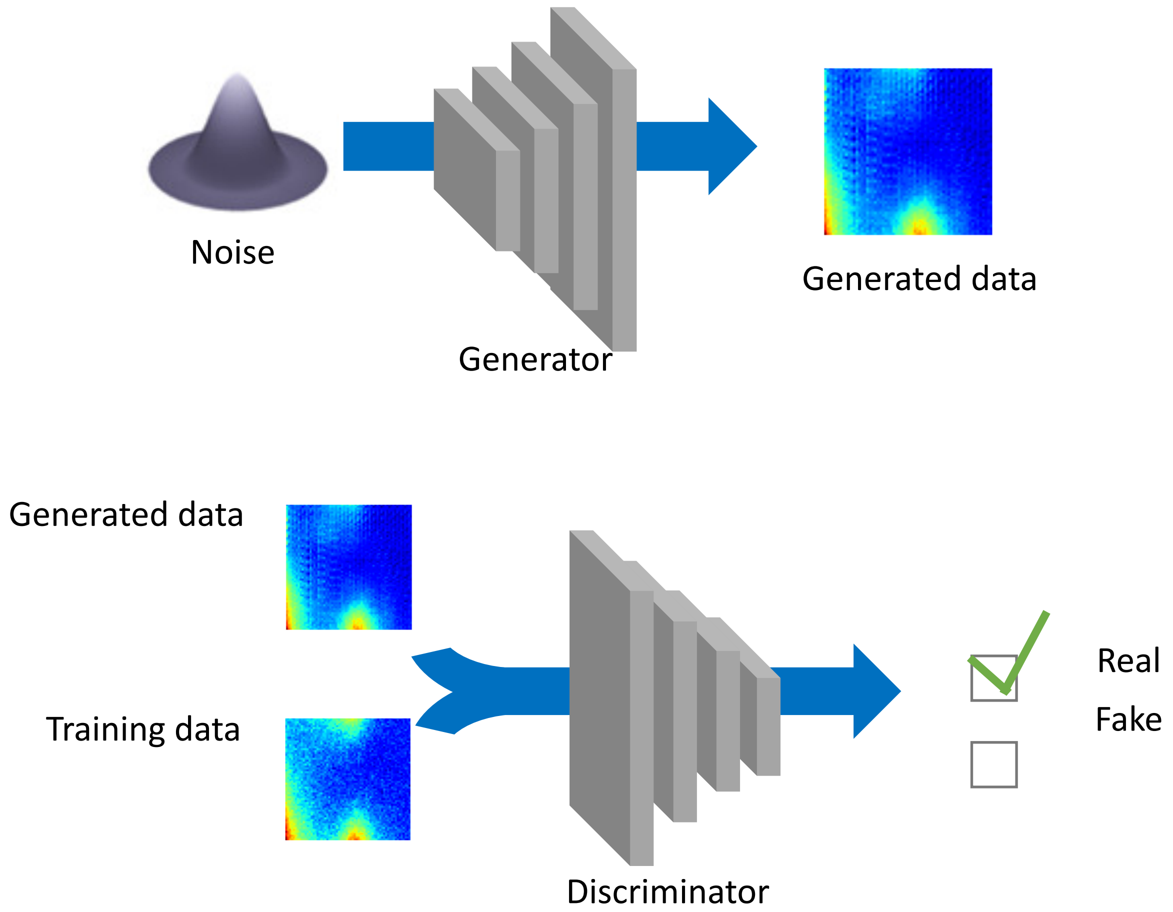
Amplitudes extraction

Goal: Train an AI model to extract amplitudes (complex numbers satisfying some physics constraints, e.g. unitarity) from events generated with Monte Carlo simulations according to a theoretical model (and eventually from experimental data)

A. Normalizing Flows:
extract differential cross section (Probability Density) from events distribution



B. Generative Adversarial Networks (GANs):
extract amplitude from differential cross sections, using unitarity constraint



Credit: G.Montaña, A.Pillioni, N.Sato

Amplitudes extraction

B. Generative Adversarial Networks (GANs):
extract amplitude from differential cross sections, using unitarity constraint

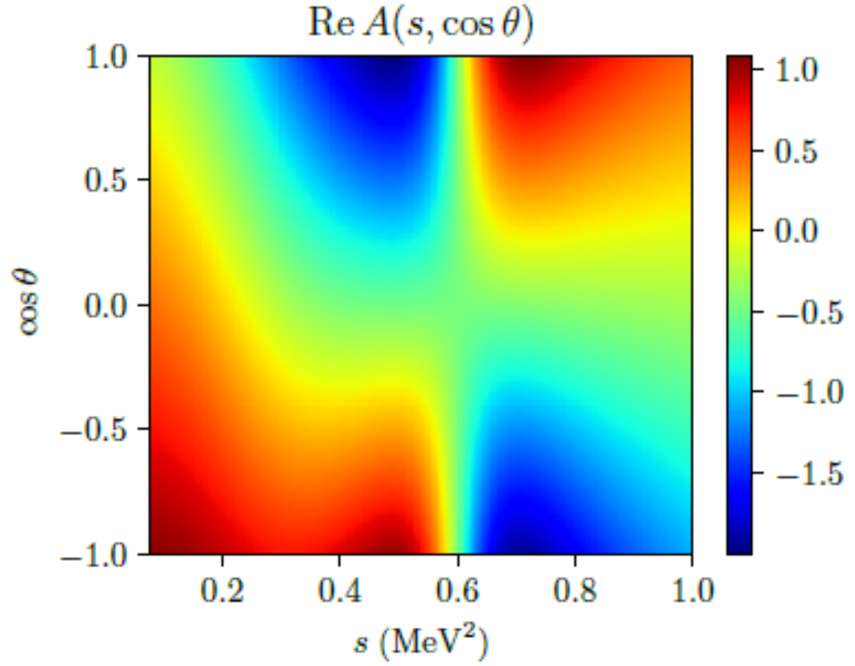
Physics model: elastic scattering $\pi^+\pi^- \to \pi^+\pi^-$

$$A(s,\cos\theta) = \sum_{\ell=0}^{n}(2\ell+1)f_\ell(s)P_\ell(\cos\theta)$$

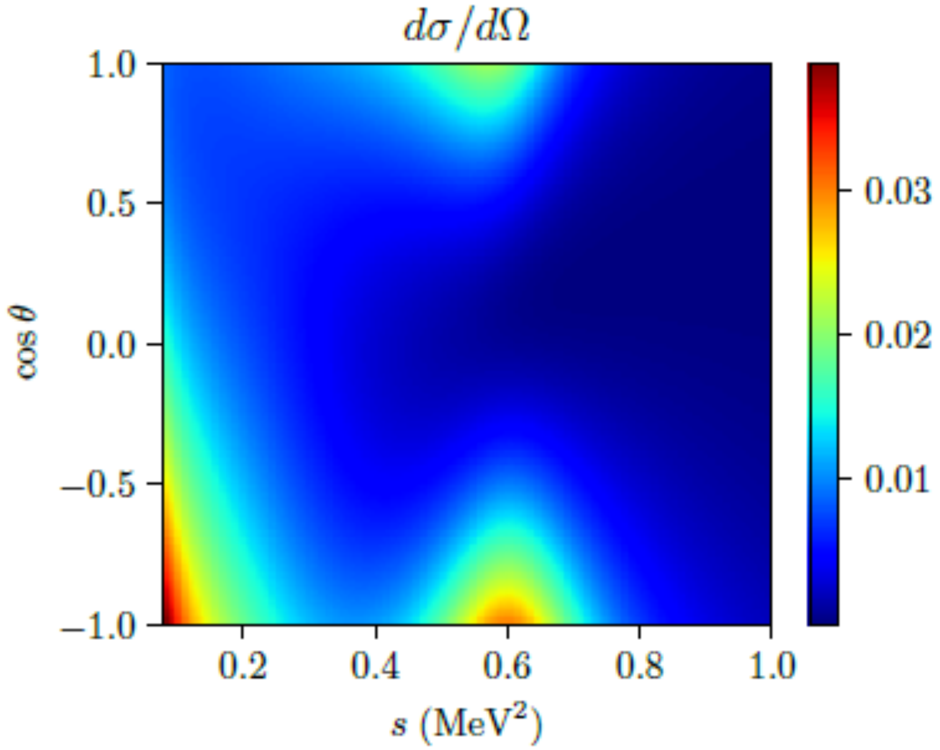
$$A(s,\cos\theta) = f_0(s) + 3f_1(s)$$

$$f_0(s) = \frac{m_\sigma\Gamma_\sigma}{m_\sigma^2 - s - i\Gamma_\sigma m_\sigma} \qquad m_\sigma = (0.4 - 0.55)\ \mathrm{GeV}\ ,\ \Gamma_\sigma = (0.4 - 0.7)\ \mathrm{GeV}$$

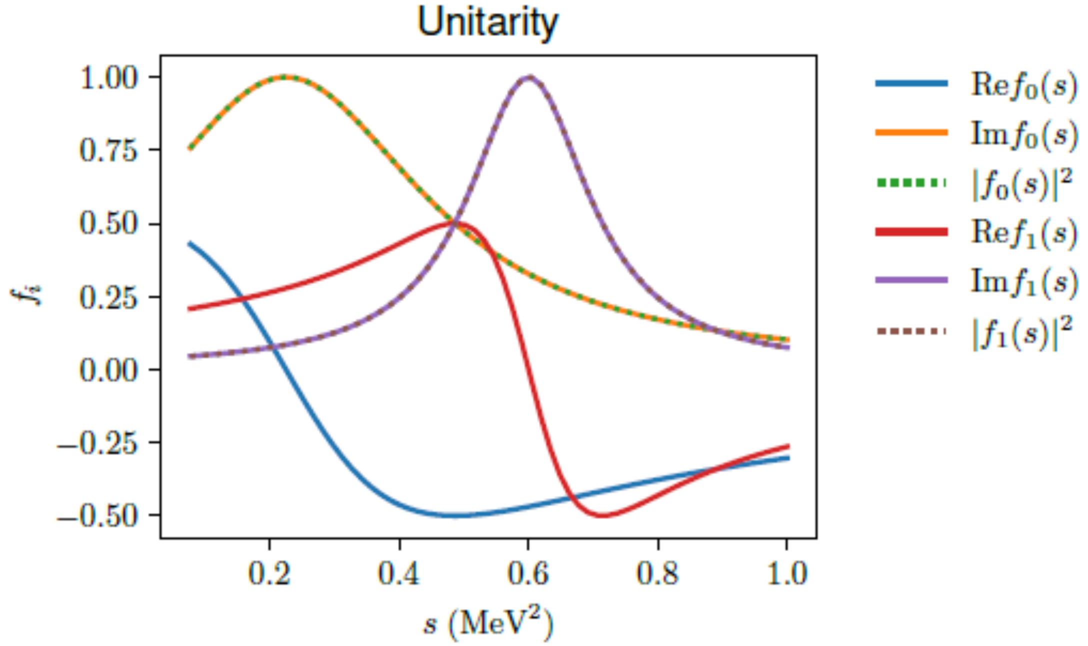
$$f_1(s) = \frac{m_\rho\Gamma_\rho}{m_\rho^2 - s - i\Gamma_\rho m_\rho} \qquad m_\rho = (0.775)\ \mathrm{GeV}\ ,\ \Gamma_\rho = (0.147)\ \mathrm{GeV}$$



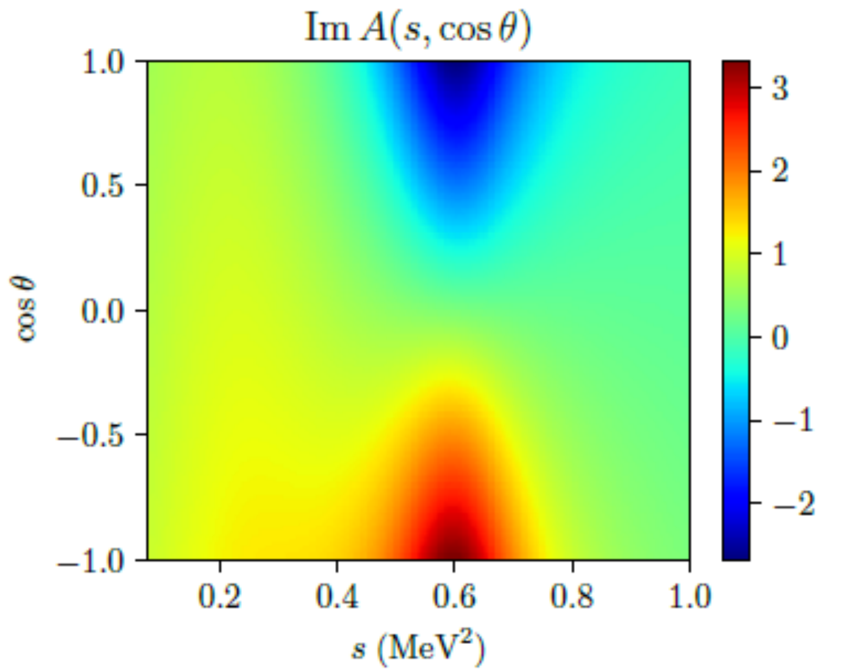
$$\frac{d\sigma}{d\Omega} = \frac{1}{64\pi^2}\frac{1}{s}|A(s,\theta)|^2$$



Partial waves satisfy the **unitarity condition**:



$$\mathrm{Im} f_l(s) = |f_l(s)|^2$$

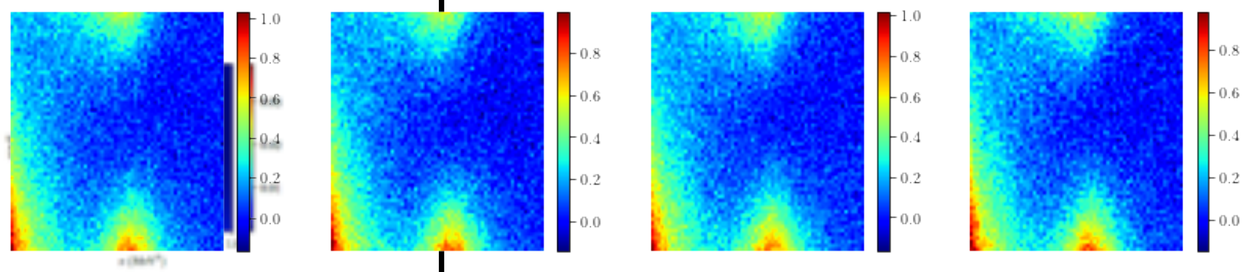


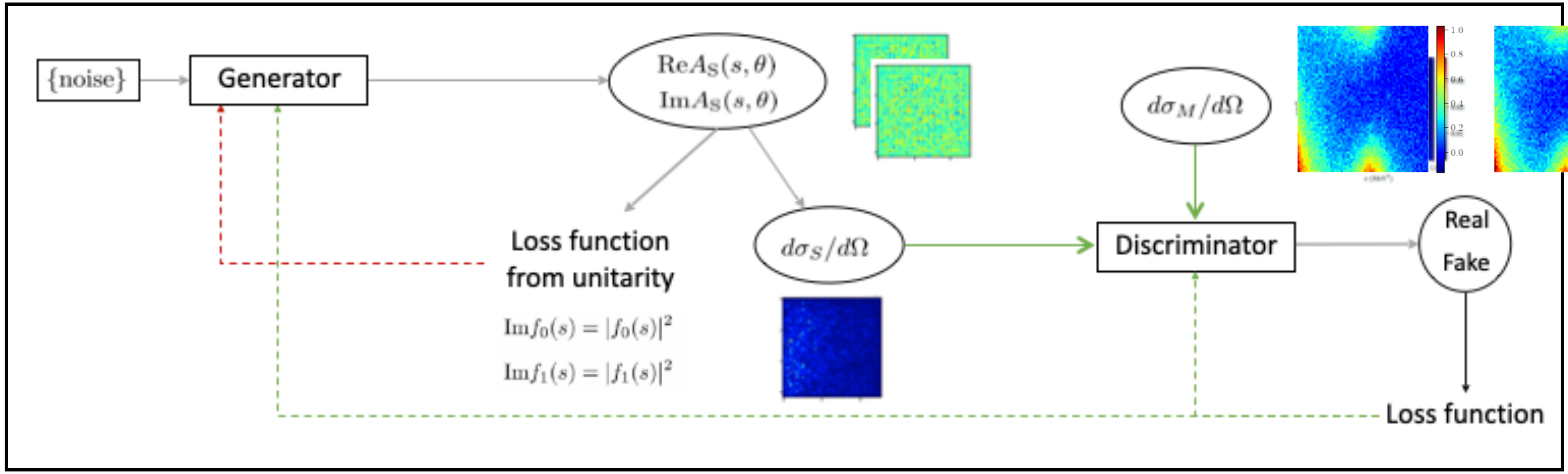
Credit: G.Montaña, A.Pillioni, N.Sato

Amplitudes extraction

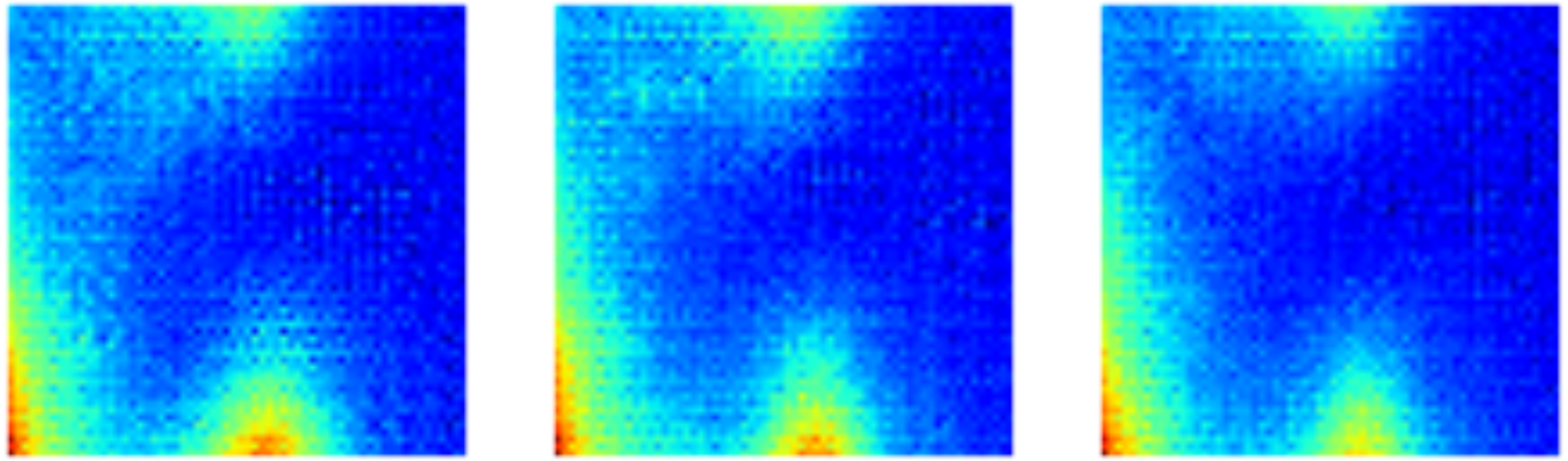
B. Generative Adversarial Networks (GANs):
extract amplitude from differential cross sections, using unitarity constraint

Training xsections generated form the model





Generated samples at the end of the training



- GANs training in progress
- from preliminary results, GANs are converging

Credit: G.Montaña, A.Pillioni, N.Sato

Summary

AI can be used to:

- Unfold detector's effects to extract physics observables at vertex-level
- Embed (multiD) xsec information (correlations) in a data-trained event gen.
- Preserve data in an alternative compact and efficient form
- Provide an alternative way to extract PDFs and amplitudes
- Incorporate Universality (of scattering amplitudes) training a NN with different kinematics of the same final state or different final states (coupled channels)
- Extract NN features related to the underlying physics

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

- Multi-step program
- We performed a positive closure test on inclusive electron scattering and multiD reactions (2pion photo production)
- We demonstrate that GANs are a viable tool to unfold detector effects (smearing) to generate a synthetic copy of data
- We demonstrate that original correlations are preserved
- We are currently working on quantifying the systematic error introduced by the detector acceptance
- The first attempt to use a model-independent procedure supervising at level of amplitudes is encouraging

Still a long way to use AI to extract physics from data in an easier and more efficient way, but, step by step, we are demonstrating this intuition is correct!