ML to access multi-d cross sections in hadron physics

M.Battaglieri (JLab)

A(i)DAPT Working Group

LDRD program + ODU ML
Y. Li (co-PI) (ODU, CS prof.)
Y. Alanazi (ODU, CS grad.)
M. P. Kuchera (Davidson College, CS prof.)
R. Ramanujan (Davidson College, CS prof.)
L. Velasco (U. Dallas, Physics undergrad)
P. Ambrozewicz (JLab, post-doc)
T. Liu (JLab, Theory post-doc)
W. Melnitchouk (PI) (JLab, Theory)
N. Sato (co-PI) (JLab, Theory)

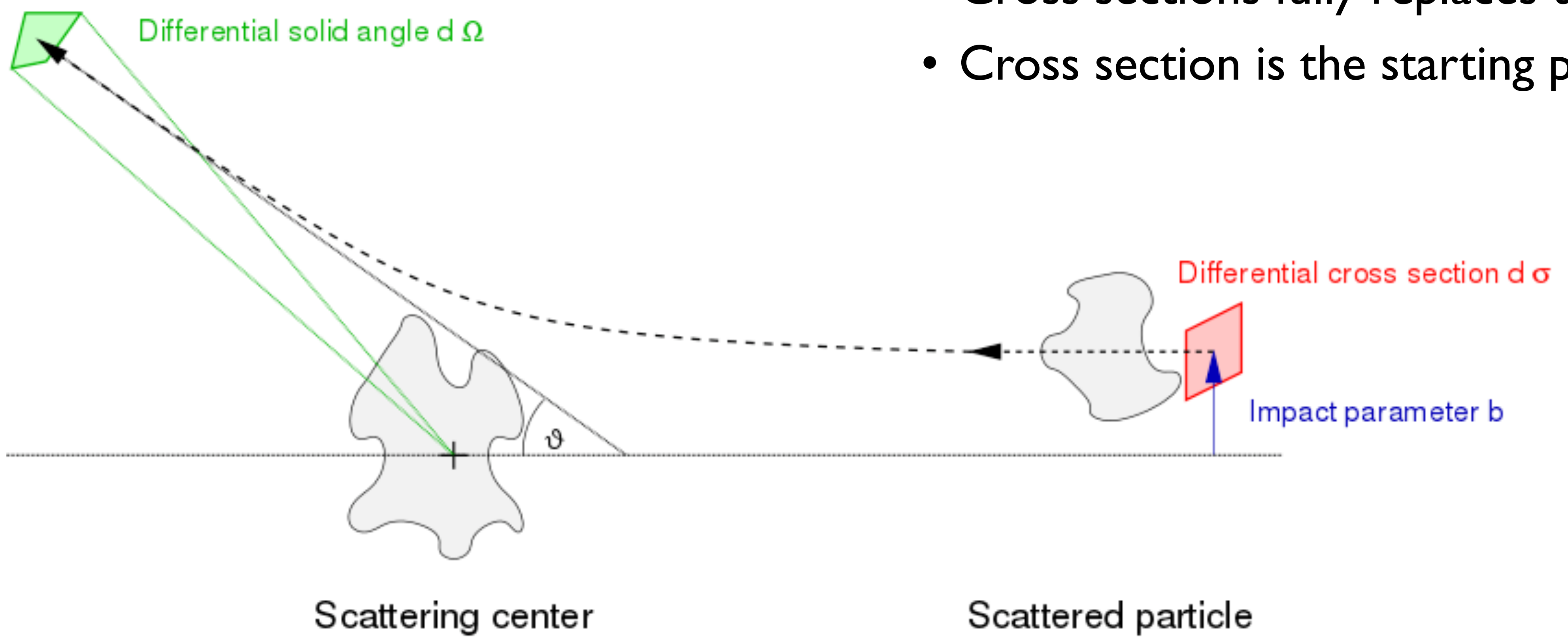
Other collaborators
A. Blin (JLab)
A. Szczepaniac (IU/JLab)
A. Pilloni (ECT*)
C. Fanelli (MIT)
G. Costantini (UniBs)
L. Biondo (UniME)
L. Marsicano (INFN-GE)
T. Roark (ODU)
T. Viducic (ODU)
N. Tyler (SCU)
M. Arratia (UCR)
S. Paul (UCR)

Hall-B Data Preservation Task Force
H. Avakian (PI)
G. Gavalian (core)
M. Ungaro (core)
V. Mokeev (core)
N. Baltzell (external)
A. Vossen (external)

The cross section in particle physics

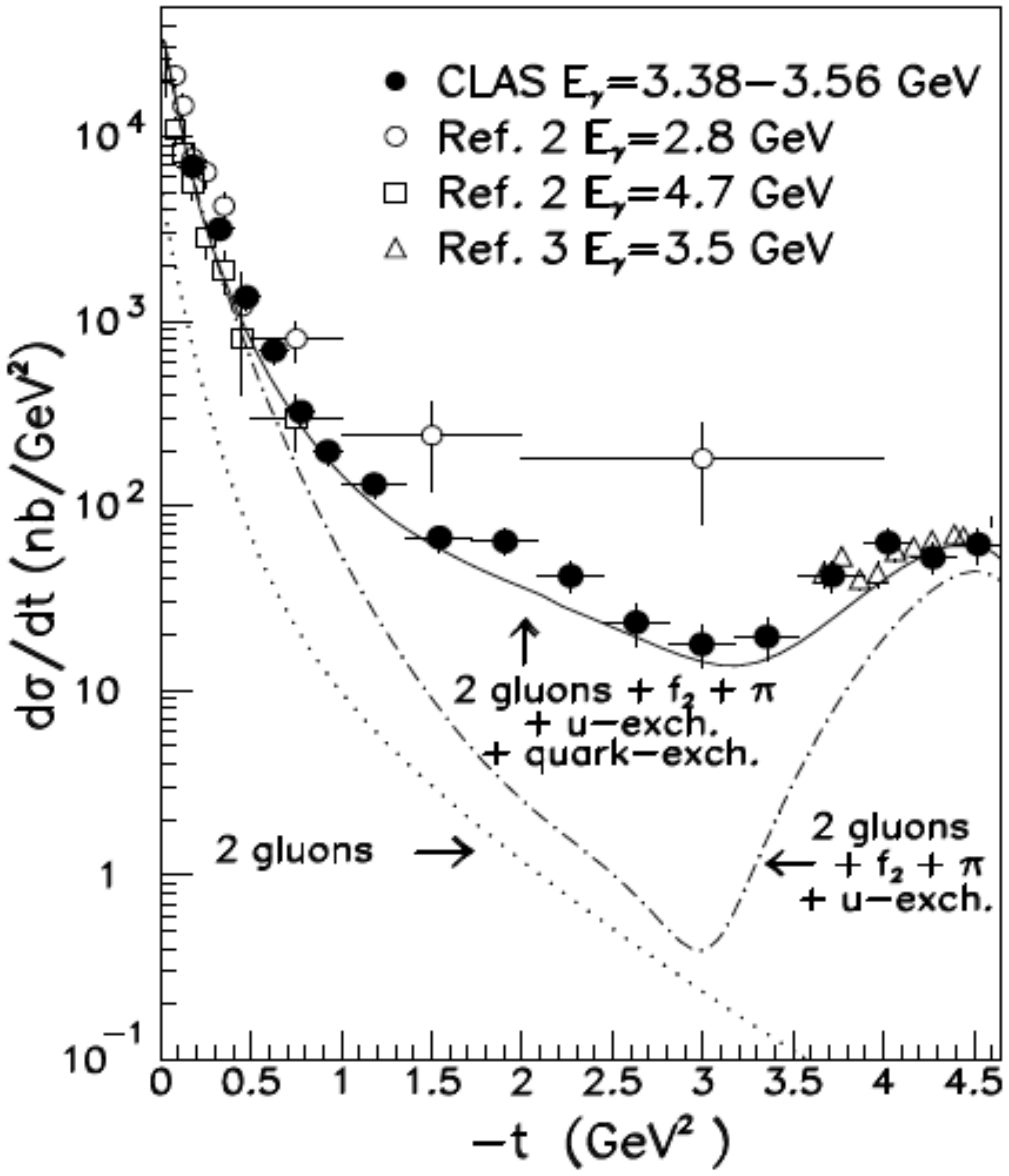
- The *cross section* is related to the transition probability between an initial to a final state
- In case of scattering, cross sections provides information about the elementary interaction
- Cross section is expressed as squared sum of scattering amplitudes (complex functions) depending on the kinematic 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

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



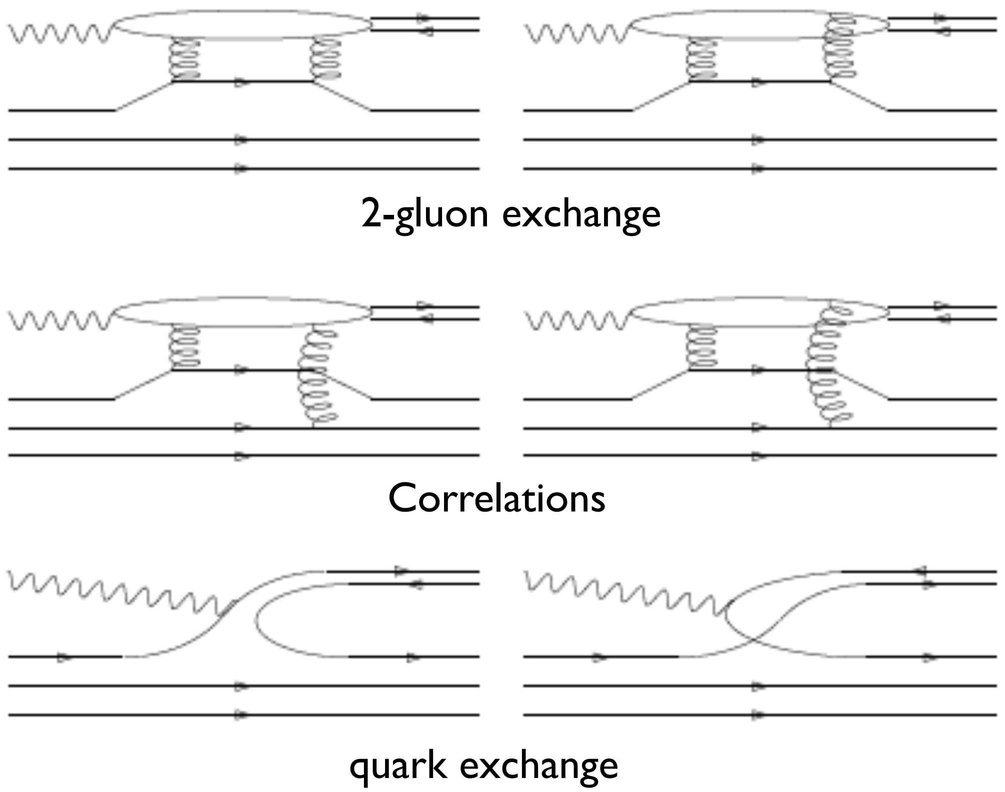
- Traditional approach: particles (4-momenta) measured into the detector, extract the relevant observables, extract physics mechanisms
- Cross section preserves this information as replacement for the original particle-by-particle scattering information

Exclusive reactions: 2 → 2



CLAS g6 ω photo production
at large momentum transfer

$\gamma\, p \to p\, \omega$



2 → 2 scattering (no polarisation)

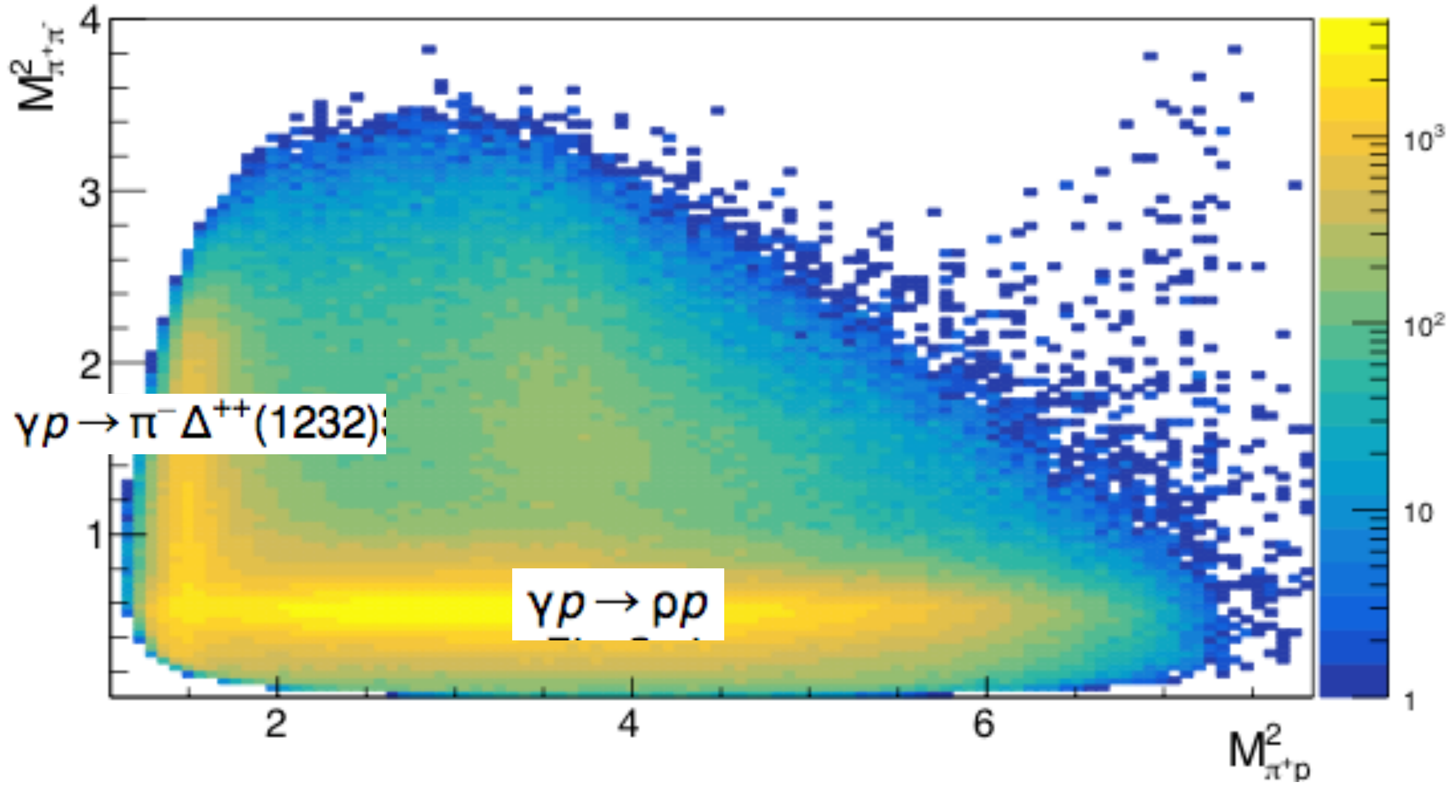
- Initial state: known
- Final state: 2 x 3
- Parameters: (2 x 3) - 4 = 2
- Possible choice: *-t* and φ
- *the* physics depends only on one variable (-t)

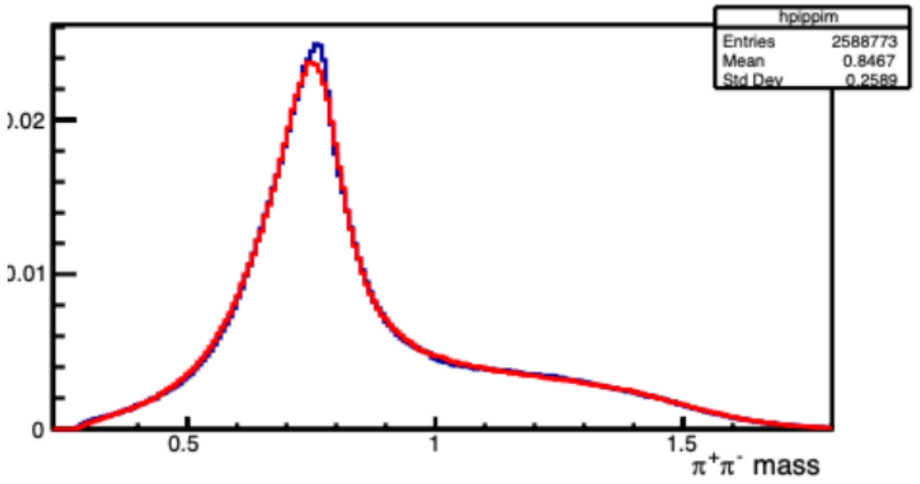
- It worked (and still works!) well if limited to channels with a single variable
- Xsec, Polarization observables, angular distribution, decay matrix, …

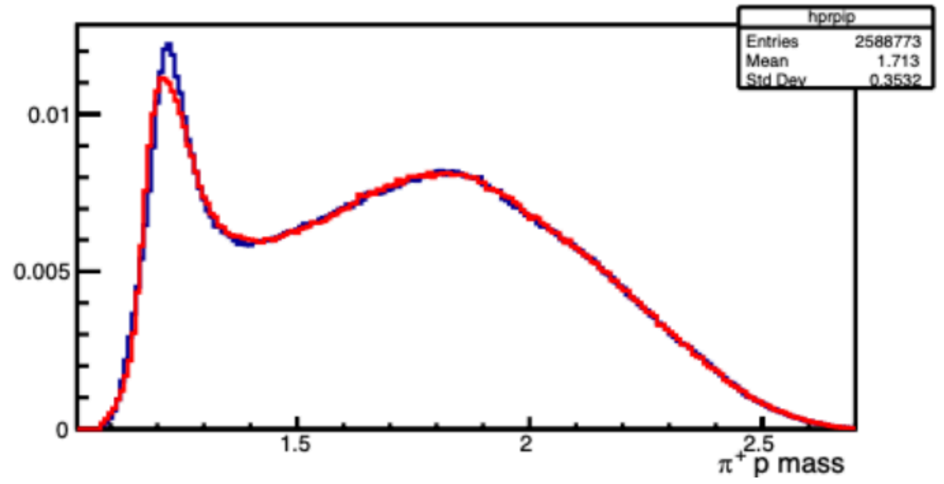
Exclusive reactions: 2 → 3

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

CLAS g11 2π photo production







2 → 3 scattering (no polarisation)

- Initial state: known
- Final state: 3 x 3
- Parameters: (3 x 3) - 4 = 5
- Possible choice: $M^2_{\pi\pi}$, $M^2_{p\pi}$ θ_π, α, φ

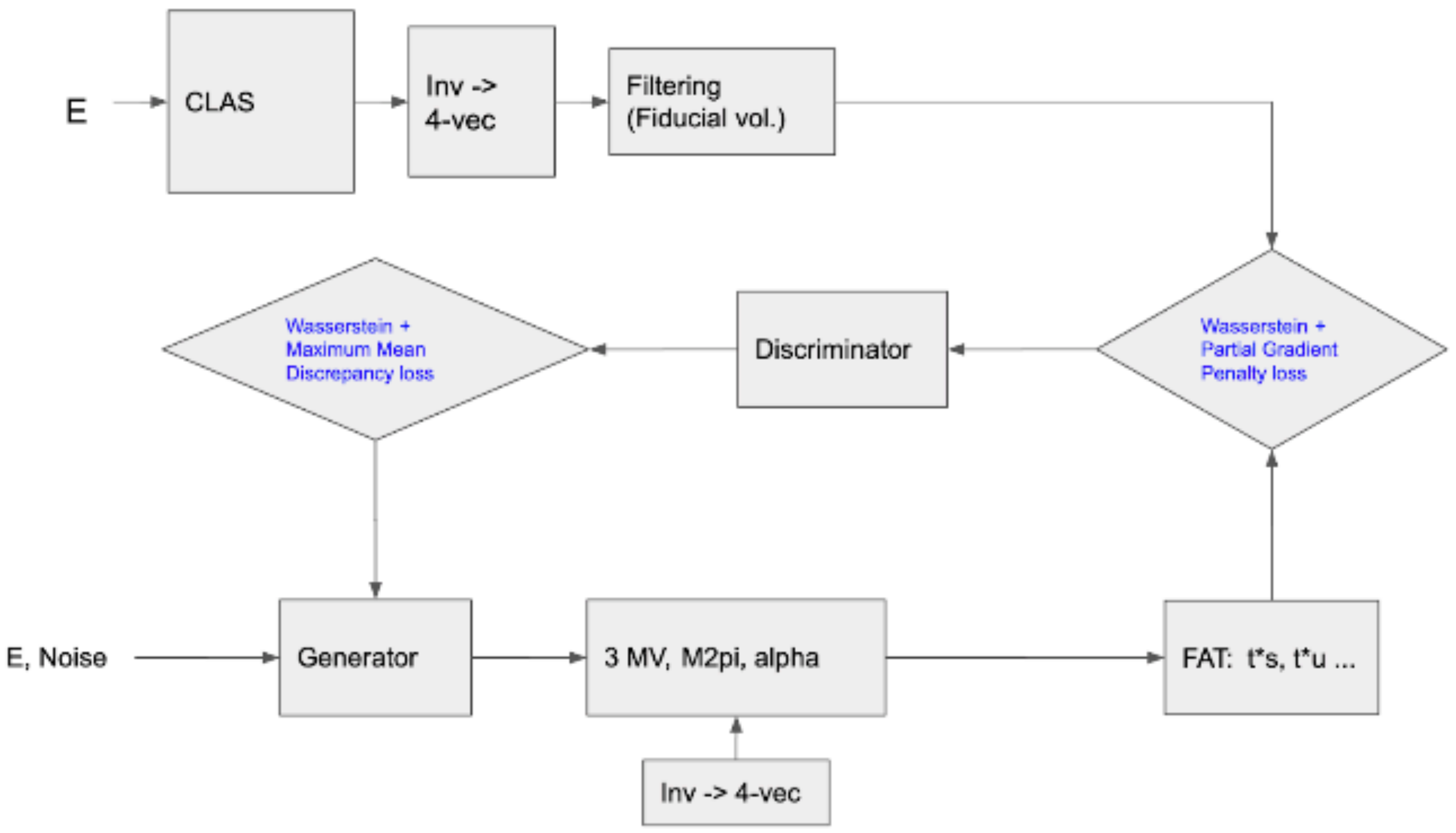
- It does not work (in practice) when you have several independent variables: multi-particle final states (spectroscopy) or multi-variable correlations (SIDIS)
- In the integration to reduce to 1-dim all correlations are lost

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

ML to access multi-d cross section

- Train a NN to generate events (synthetic) with the SAME correlations of experimental data **N.Sato talk**
- Replace the xsec with a NN (synthetic data are equivalent to data)
- Light GAN-based event generator can generate any statistics

FAT-GAN architecture **Y.Li talk**



- Tested different options: 4-moms in CM or LAB or using kinematics INVARIANTS
- Implemented a folding/unfolding procedure to take into account detector acceptance/resolution **Y.Li and L.Marsicano talks**
- Training performed on 10% of the data set to use the others for systematic studies. When optimised the training will be done on the whole data set
- Results are validated by comparing data/synt 1-dim projected and selected bins in 5-d space in LAB, CM and INVARIANTS space **E.Isupov talk**
- Error quantified to include statistical (via bootstrap) and systematic (different data samples, different reference systems) **P.Ambrozewicz talk**
- The final check is performed comparing YLM moments extracted from data and sent-data (in the assumption that any further physics observables will be extracted from YLM) **A. Hiller Blin talk**

ML to access multi-d cross section

A(i)DAPT Working Group

Goals

- Cross section: embed multi-d cross section information (correlation) in a data-trained event generator
- Preserve data in an alternative compact and efficient form (to be applied to current JLab physics program)
- Statistics: use the NN to determine the necessary statistics for a given analysis
- Statistics: overcome statistics limitation exploiting ML super-resolution
- Detector Efficiency: folding/unfolding detector effects to extract physics at vertex level (via sim or data)
- Physics analysis: incorporate Universality (of scattering amplitudes) training the NN with different kinematics of the same final state or different final states (coupled channels)
- Physics analysis: extract from the NN features related to the underlying physics
- Physics analysis: structure the NN to reflect amplitudes properties (poles, cuts, dynamics, …)

- Collaborative effort
  - ML
  - Data manipulation
  - Validation
  - Unfolding
  - Theory
- Regular weekly meeting

- develop a procedure to best fit data (ML Group)
- develop a procedure to compare synt-data to data (Validation Group)
- develop a procedure to quantify the error associated to sent-data (Data manipulation Group)
- develop a procedure to take into account the detector effect (Folfing/Unfolding Group)
- extract physics form data and sent-data and compare (Theory Group)

- Wiki page: https://clasweb.jlab.org/wiki/index.php/A%28I%29DAPT_-_AI_for_Data_Analysis_and_PresevaTion