Reconstructing the Kinematics of Deep Inelastic Scattering with Deep Learning

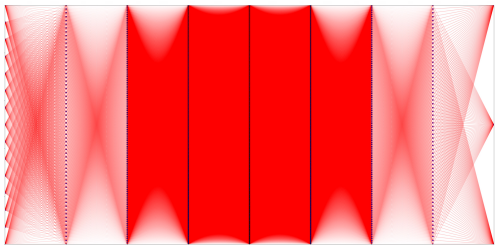
Miguel Arratia,[a,b] **Daniel Britzger,**[c] **Owen Long,**[*a] **and Benjamin Nachman**[d,e]

[a] *Department of Physics and Astronomy, University of California, Riverside, CA 92521, USA*
[b] *Thomas Jefferson National Accelerator Facility, Newport News, VA 23606, USA*
[c] *Max-Planck-Institut für Physik, Föhringer Ring 6, 80805 München, Germany*
[d] *Physics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*
[e] *Berkeley Institute for Data Science, University of California, Berkeley, CA 94720, USA*

NIMA 1025 (2022) 166164

https://arxiv.org/abs/2203.16722

(Accepted for publication in JINST)



UNIVERSITY OF CALIFORNIA UCRIVERSIDE

Overview

- Many methods for reconstructing neutral current DIS variables (Q^2, y, x) using scattered electron and Hadronic Final State (HFS) are complementary. Which method is best depends on Q^2, y, x and the strength of QED radiation.

- QED radiation can produce tails in Q^2, y, x distributions if it is not identified and handled in the event.

- Machine learning can address both of these issues to provide a single reconstruction method that optimally uses all available information in the event.

- This approach also leads to smaller uncertainties and correlations in unfolding as well as other advantages.

QED Radiation

We use the following *practical definitions* of QED radiation

Initial State Radiation (ISR): the radiated photon is closer to the electron beam direction.

Final State Radiation (FSR): the radiated photon is closer to the scattered electron direction.

We use the RAPGAP MC generator for our studies.

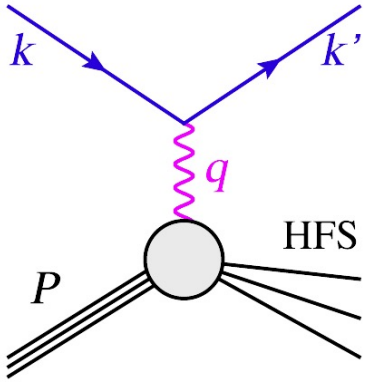
RAPGAP implements HERACLES for higher-order QED radiation.

$$p_z^{\mathrm{bal}} = 1 - \frac{\Sigma_e + \Sigma}{2\ E_0}$$

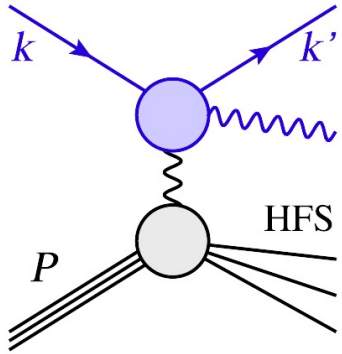
Zero if no ISR, >0 with ISR

$$p_T^{\mathrm{bal}} = 1 - \frac{p_{T,e}}{T}$$

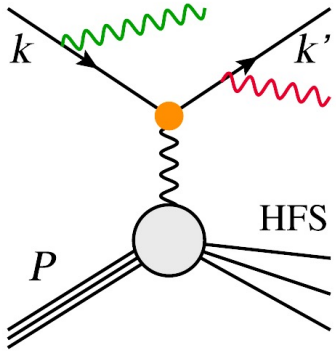
Zero if no FSR, >0 with FSR



Born-level Deep Inelastic Scattering



Radiative leptonic tensor including higher-order QED corrections



Practical implementation in MC event generators.

ISR FSR

Effective coupling

QED Radiation

We use the following *practical definitions* of QED radiation

Initial State Radiation (ISR): the radiated photon is closer to the electron beam direction.

Final State Radiation (FSR): the radiated photon is closer to the scattered electron direction.

We use the RAPGAP MC generator for our studies.

RAPGAP implements HERACLES for higher-order QED radiation.

$$p_z^{\text{bal}} = 1 - \frac{\Sigma_e + \Sigma}{2\ E_0}$$

Zero if no ISR, >0 with ISR

$$p_T^{\text{bal}} = 1 - \frac{p_{T,e}}{T}$$

Zero if no FSR, >0 with FSR

Definition of DNN regression learning targets or the "true" Q^2, y, x.

Standard definitions for events with no QED radiation in terms of beam electron ℓ , scattered electron ℓ', and beam proton p 4-vectors.

For events with QED radiation

Use post-ISR beam electron 4-vector

Use pre-FSR scattered electron 4-vector

$$q \equiv \ell - \ell'$$

$$Q^2 = -q \cdot q$$

$$y = \frac{p \cdot q}{p \cdot \ell}$$

$$x = -\frac{q \cdot q}{2\ p \cdot q}$$

Machine Learning strategy

Provide complete reconstruction information on the scattered electron and the Hadronic Final State as inputs.

Provide additional event observables that can indicate the presence of a QED radiation photon (ISR or FSR) and help quantify the radiation effects.

Feed all useful information into a regression **Deep Neural Network** (DNN) that predicts Q^2, y, and x.

DNN inputs

QED radiation:

ISR radiation (4):

E, η, and $\Delta\phi$ of photon in event closest to electron beam direction, where $\Delta\phi$ is w.r.t. the scattered electron.

$$p_z^{\mathrm{bal}} = 1 - \frac{\Sigma_e + \Sigma}{2\ E_0}$$ from event reconstruction

FSR radiation (3):

ECAL energy within $\Delta R<0.4$ around electron divided by the electron track momentum.

Number of ECAL clusters within $\Delta R<0.4$ around electron.

$$p_T^{\mathrm{bal}} = 1 - \frac{p_{T,e}}{T}$$ from event reconstruction

DIS reconstruction (Q^2, y, x):

Scattered electron (3): E, p_T, and p_z.

HFS (3): E, p_T (T), and p_z.

$\Delta\phi$ between the scattered electron and HFS.

The difference $\Sigma_e - \Sigma$.

Training samples: RAPGAP generator and DELPHES fast simulation of ATHENA.
NC DIS with $Q^2_{gen} > 200\ GeV^2$.
32 GeV < event ($E-p_z$) < 40 GeV , (±4 GeV around $2E_e$)

Machine Learning : Deep Neural Network

Sequential network with **8 layers**.

15 inputs from previous slide, transformed to zero mean, unit RMS..

Nodes per layer: N_{in}, 64, 128, 512, 1024, 512, 128, 64, N_{out}.

Activation: relu for first layer, selu for middle.

Adam optimizer, learning rate 10^{-5}.

Huber loss function with $\delta = 0.01$.

Batch size 1024.

Samples: 28 million events, evenly split for training and validation. $Q^2_{gen} > 200\ GeV^2$.

Training converges after ~40 to ~100 **epochs** in around 30 to 60 minutes on GPU machine.

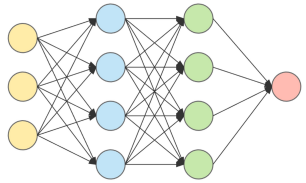
Applied this basic DNN structure to the following tasks:

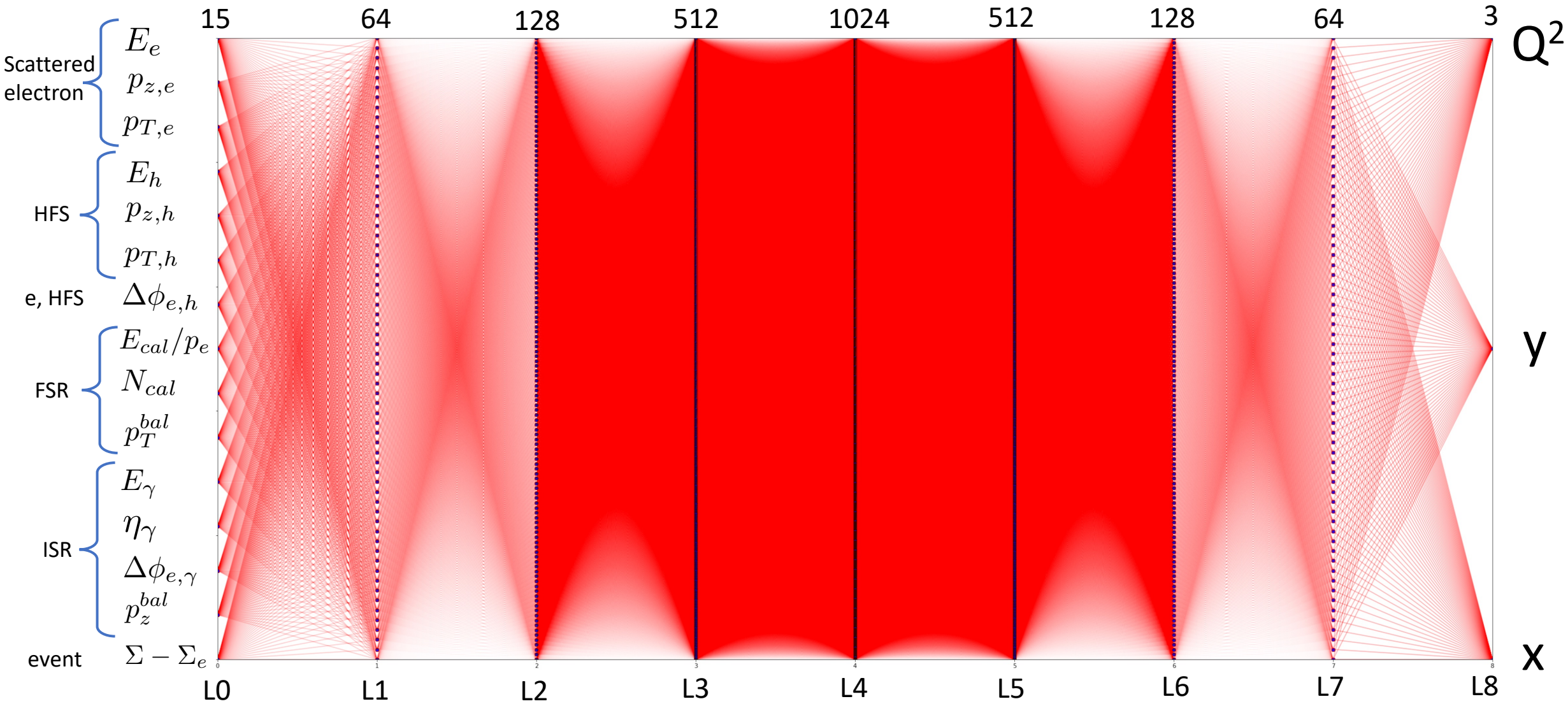
QED radiation classification : ISR, FSR, NoR

QED radiation quantification : predict p_z^{bal} and p_T^{bal}.

DIS reconstruction : Q^2, y, x.

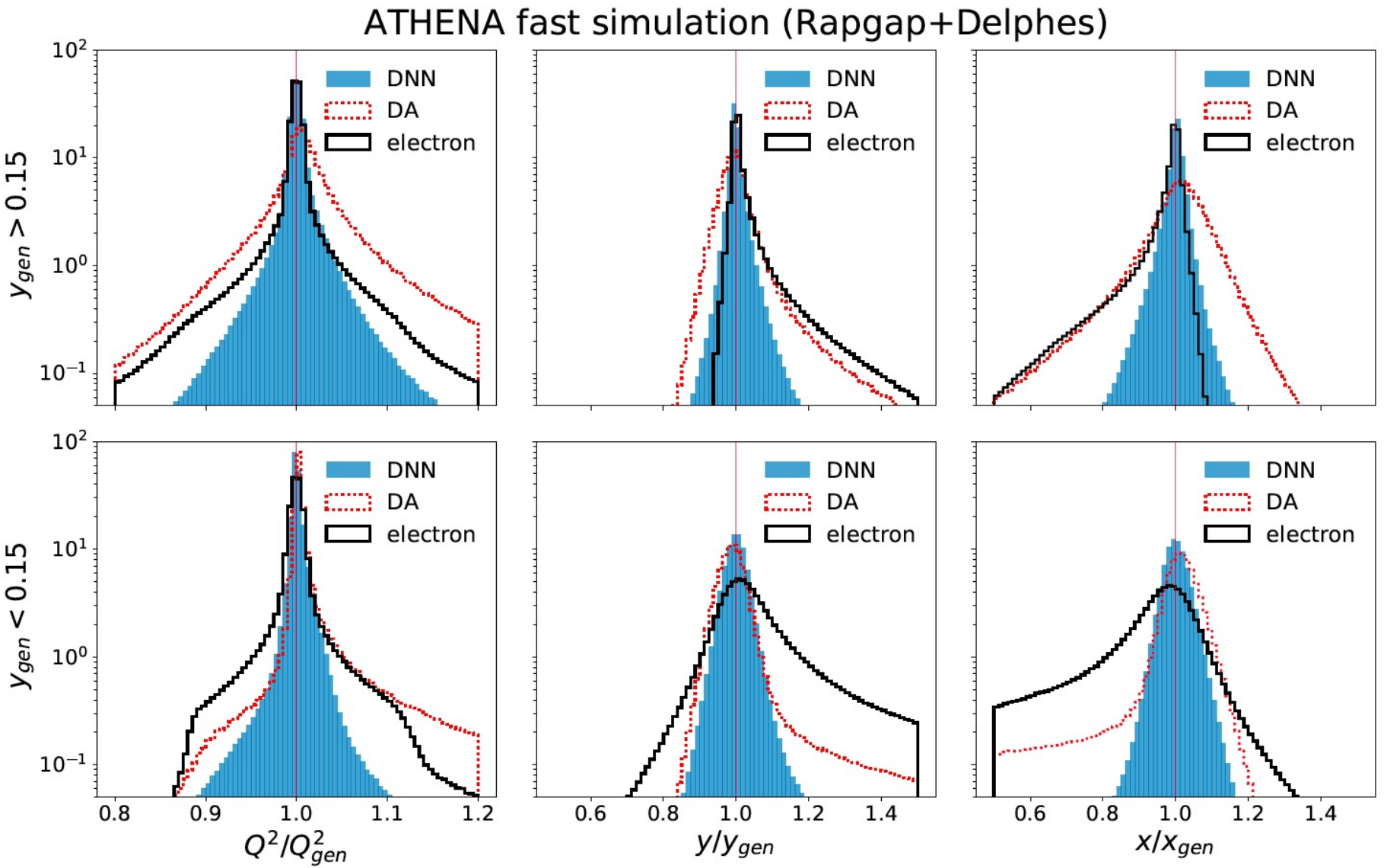
Network diagram





There are 1,197,184 connections in the network.

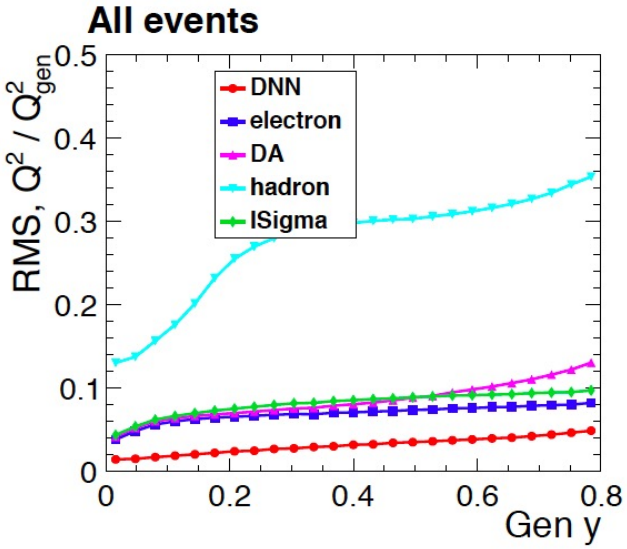
DIS reconstruction : regression DNN for Q^2, y, x

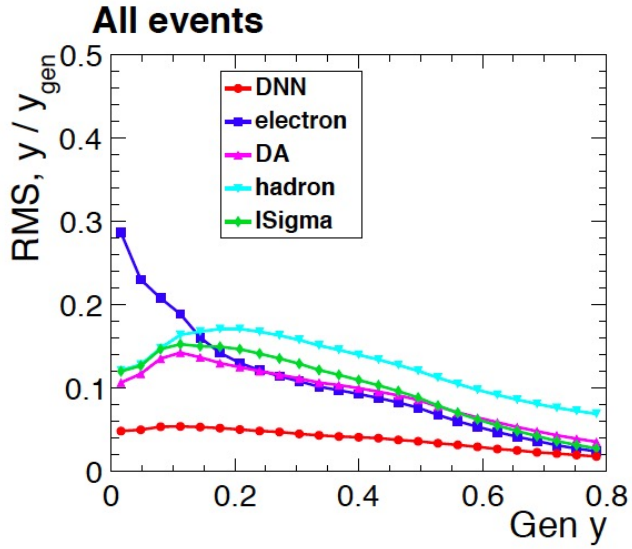


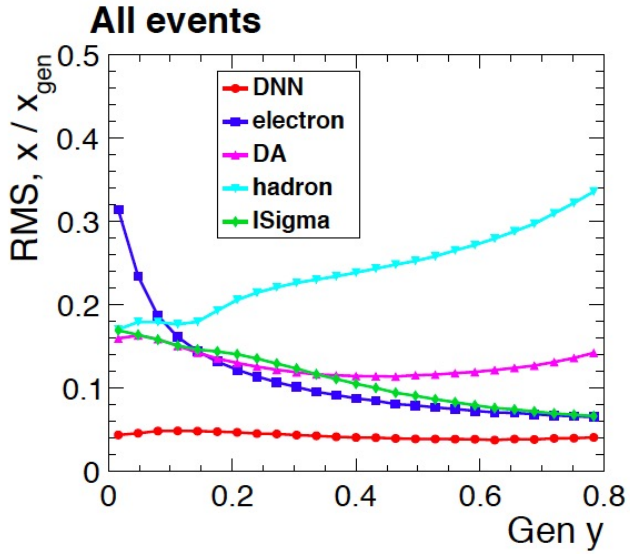
DNN has similar core resolution to best conventional method (electron at high y, DA at low y).

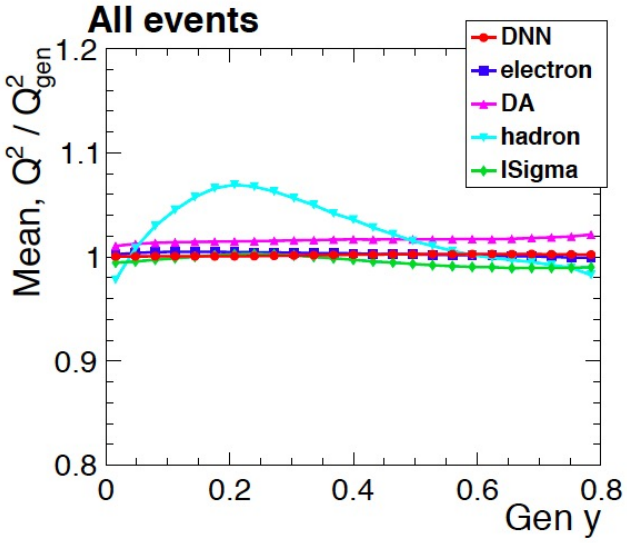
Large tails from QED radiation in conventional reconstruction methods absent in DNN.

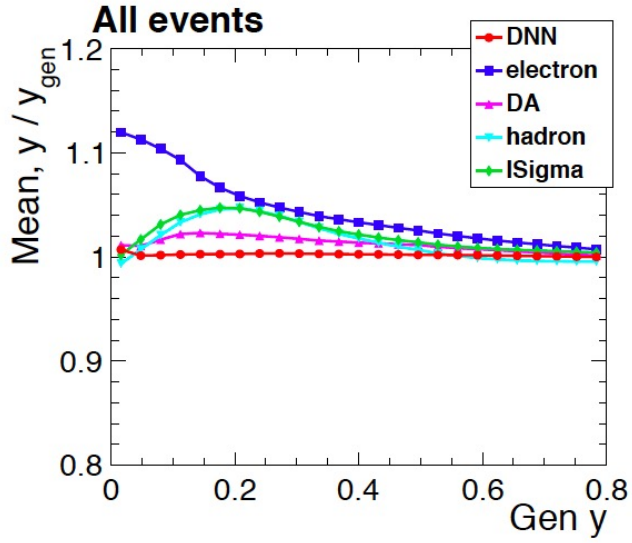
ATHENA fast simulation (Rapgap+Delphes)

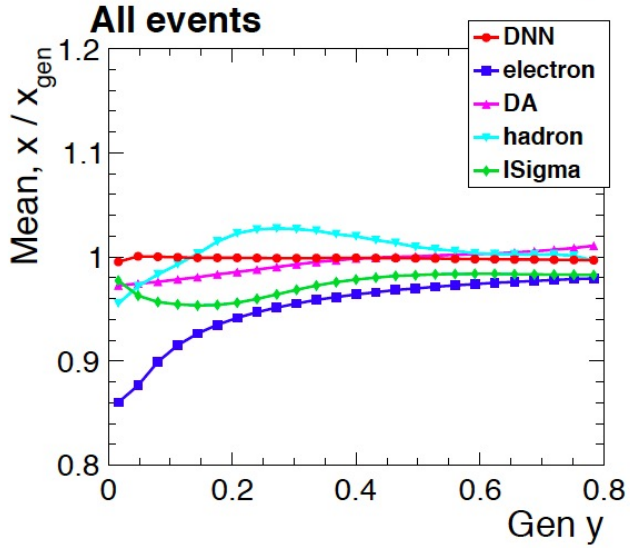












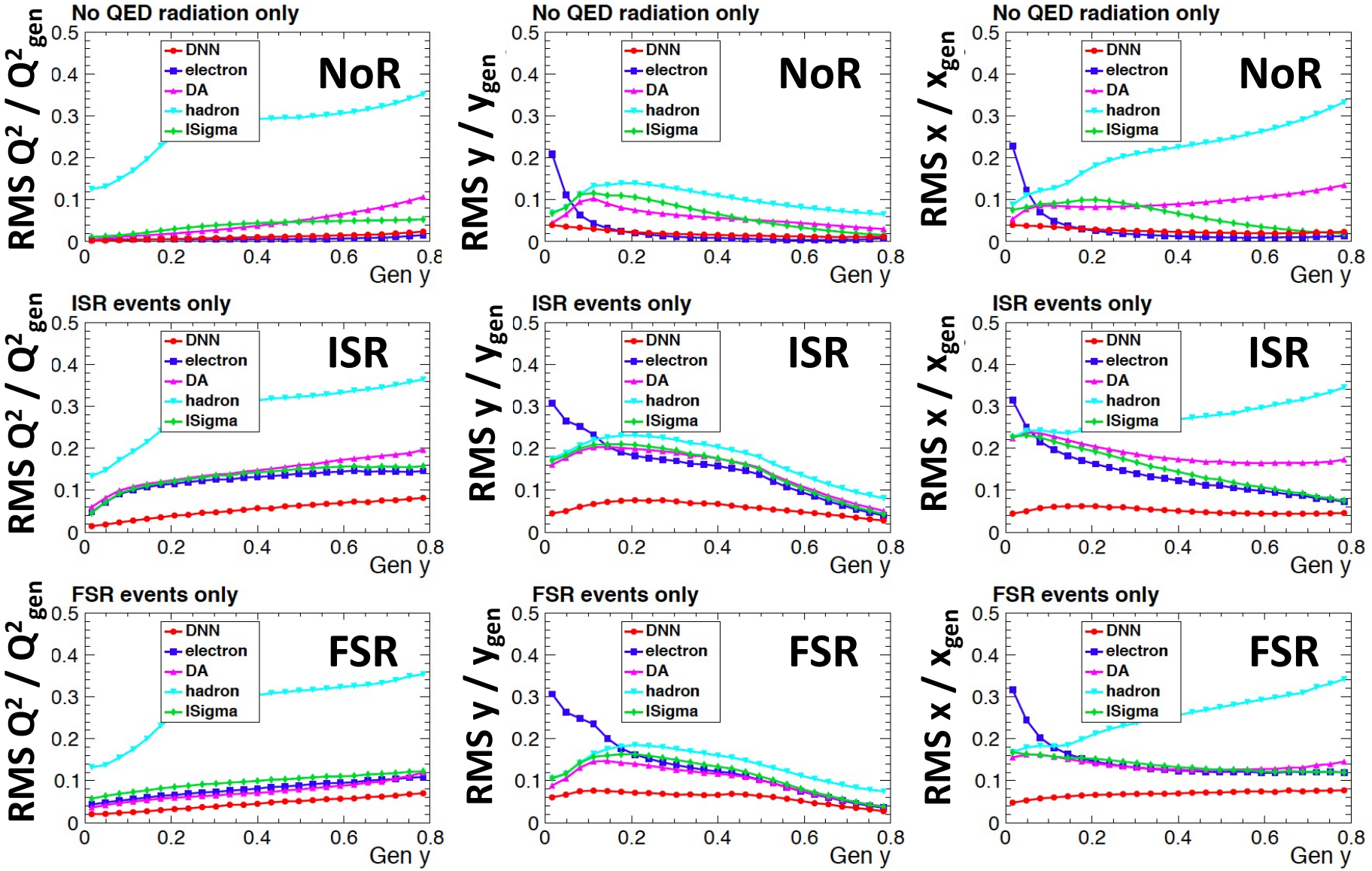
DNN outperforms all conventional reconstruction methods.

DNN has smallest RMS and essentially no bias.

RMS and mean calculated for events with measured / gen ratio between 0 and 2.

Resolution (RMS) vs y_{gen}

ATHENA fast simulation (Rapgap+Delphes)

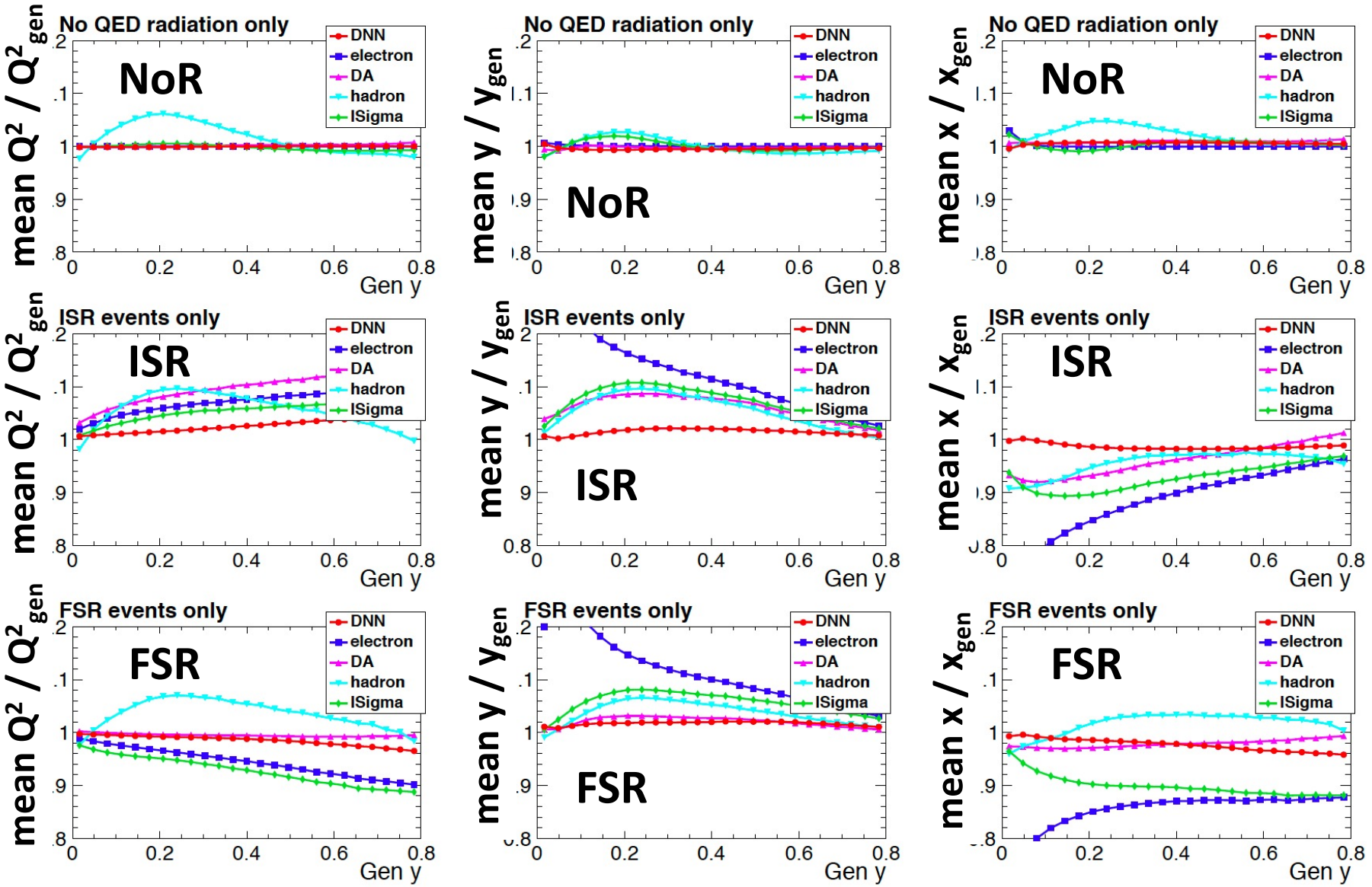


Electron method has better core resolution than DNN for y>0.15 in NoR events (no tails).

DNN resolution much less affected by QED radiation

RMS and mean calculated for events with measured / gen ratio between 0 and 2.

Bias (mean) vs y_{gen}



All methods (except hadron) are unbiased in events with no QED radiation

DNN remains unbiased in events with QED radiation, while other methods have large bias.

DNN has successfully learned how to mitigate QED radiation effects.

RMS and mean calculated for events with measured / gen ratio between 0 and 2.

Demonstration of DNN with H1 full simulation

HERA beam energies: E_e = 27.6 GeV, E_p = 920 GeV.

NC DIS with $Q^2_{gen} > 200\ GeV^2$.

RAPGAP 3.1 generator (includes HERACLES).

GEANT 3 detector simulation.

Employ standard reconstruction methods for electron and HFS.

Includes real calorimeter noise.

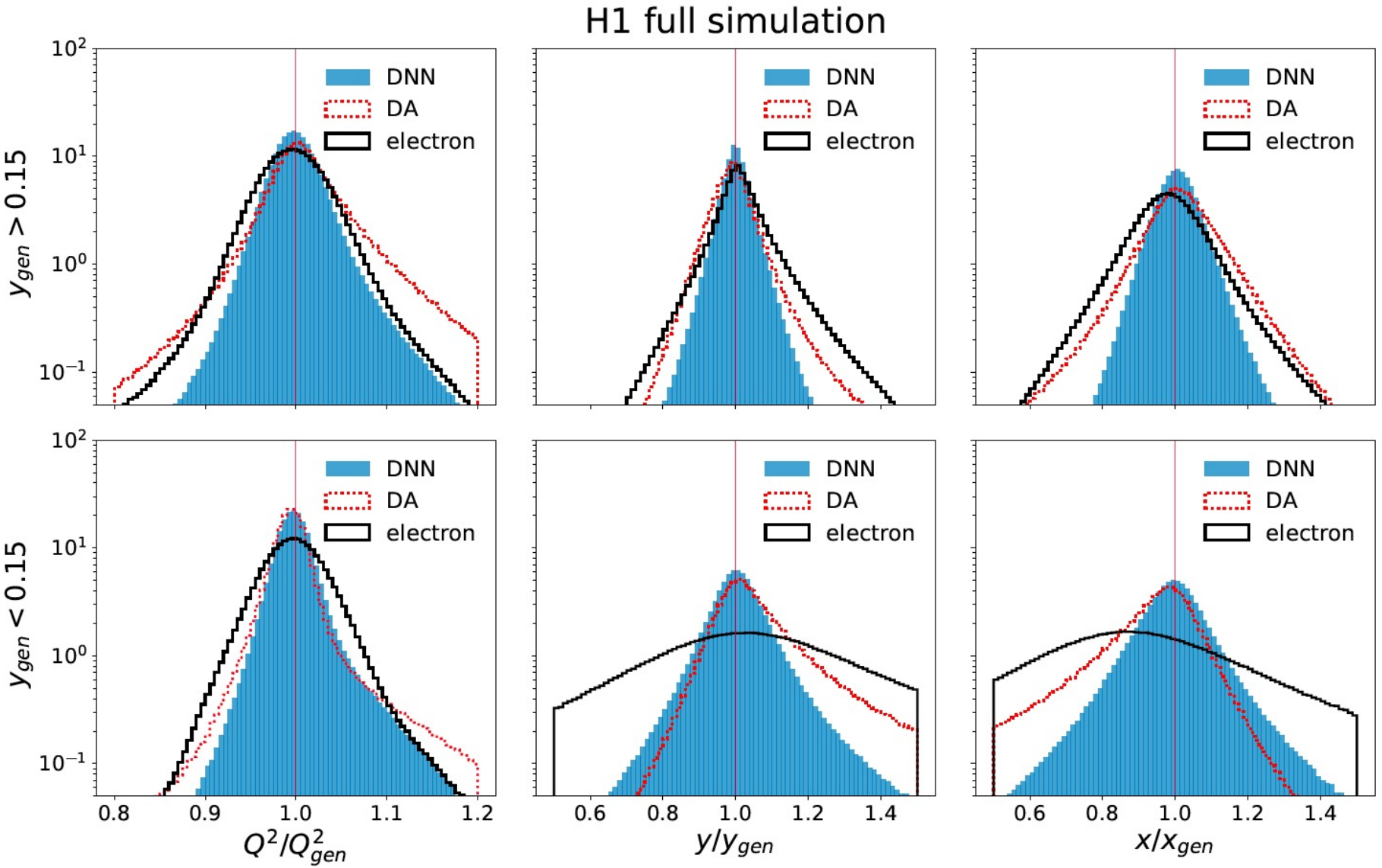
Includes run-specific conditions.

Event selection:

45 GeV < event $(E-p_z)$ < 65 GeV (±10 GeV around 2 E_e)

Around 11 million events, evenly split for training and validation.

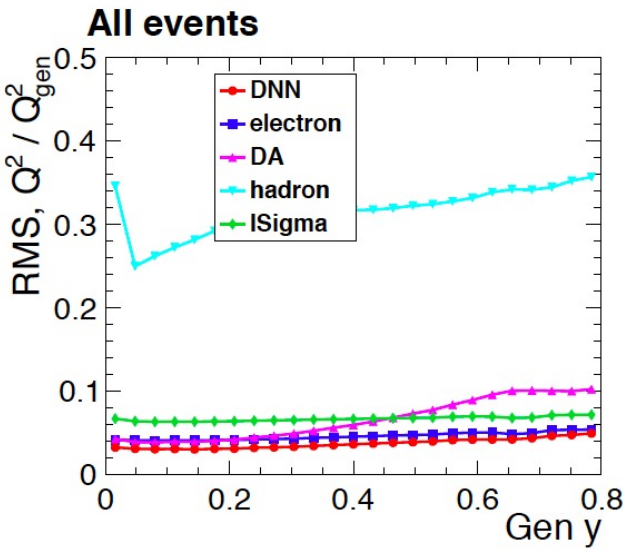
Demonstration of DNN technique with highly tuned, well understood, data-validated full simulation from the H1 experiment.

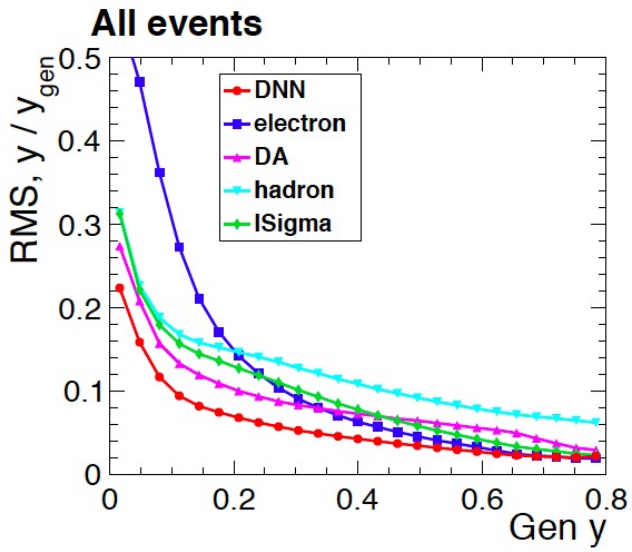


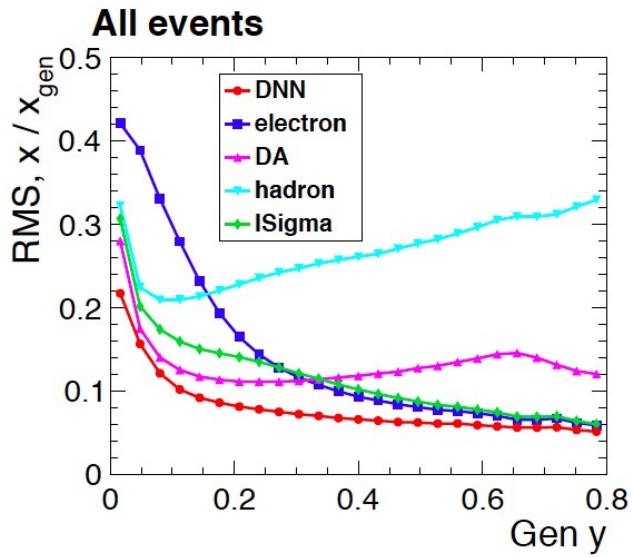
*DNN has **better** core resolution than best conventional method (electron at high y, DA at low y).*

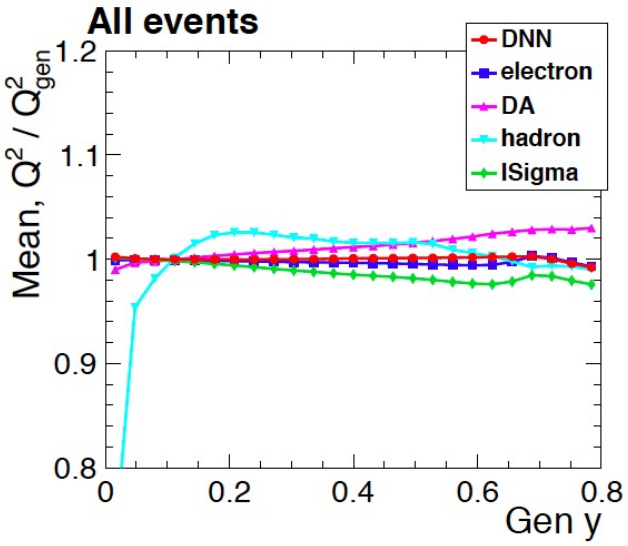
DNN distributions much more symmetric, free of large QED radiation tails.

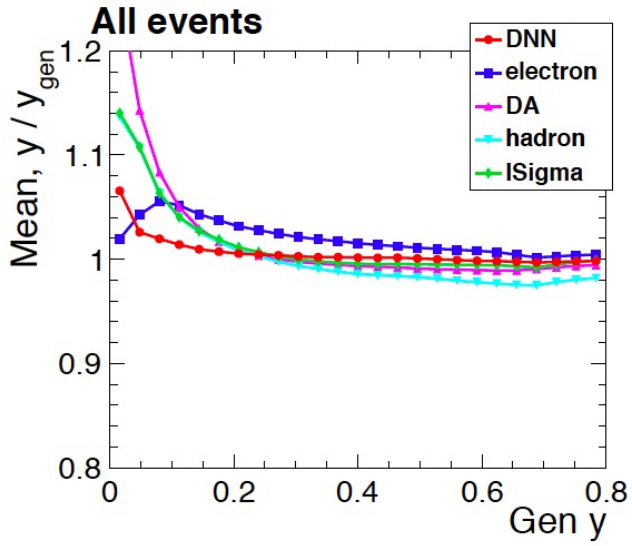
H1 full simulation

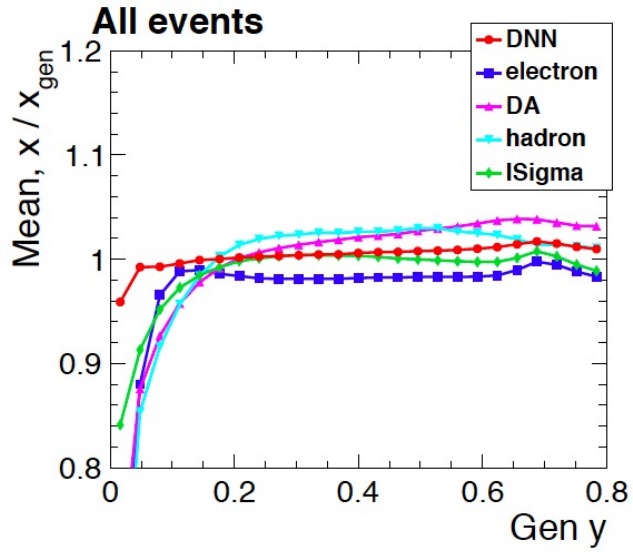






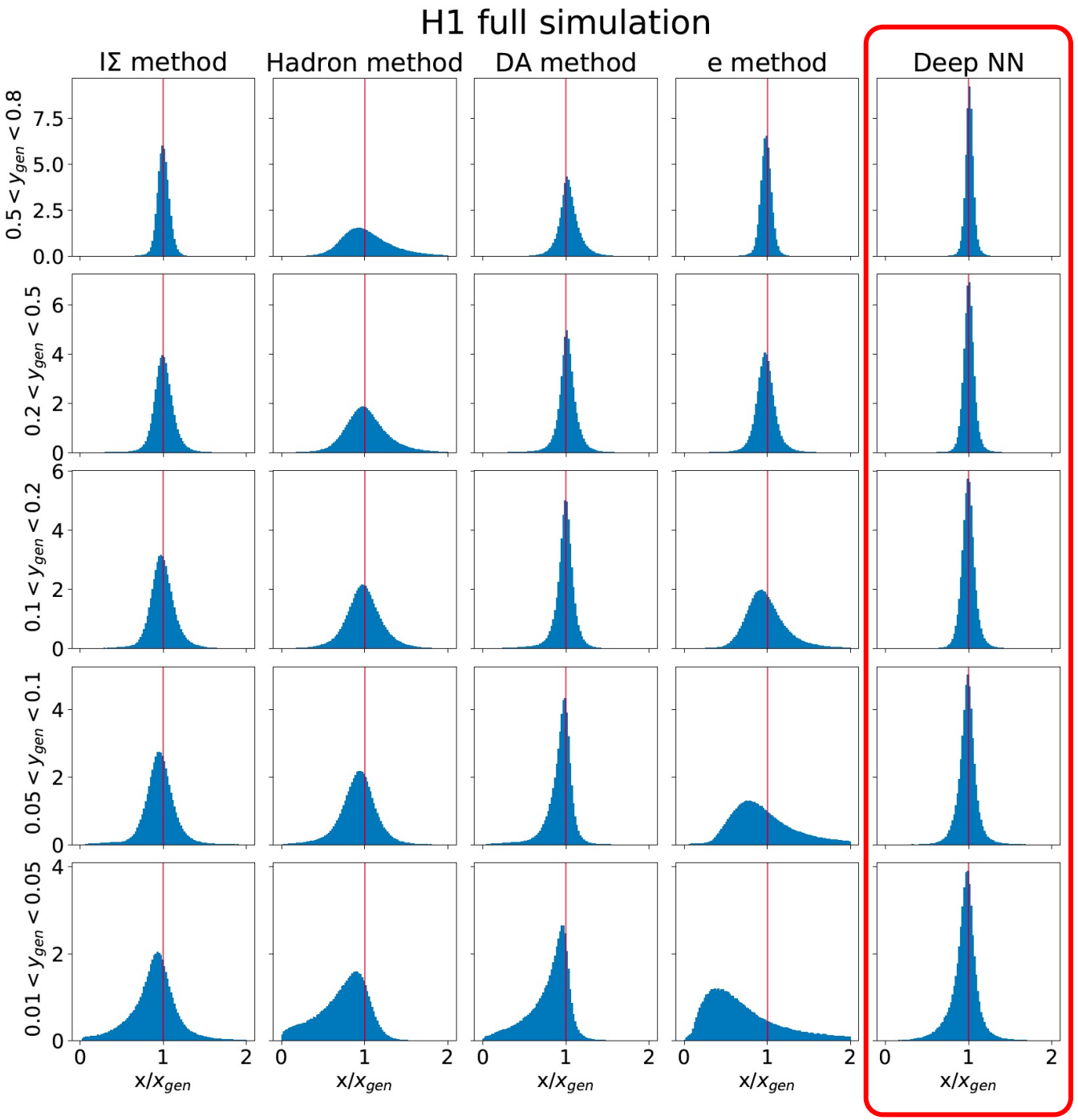






DNN resolution outperforms all conventional methods.

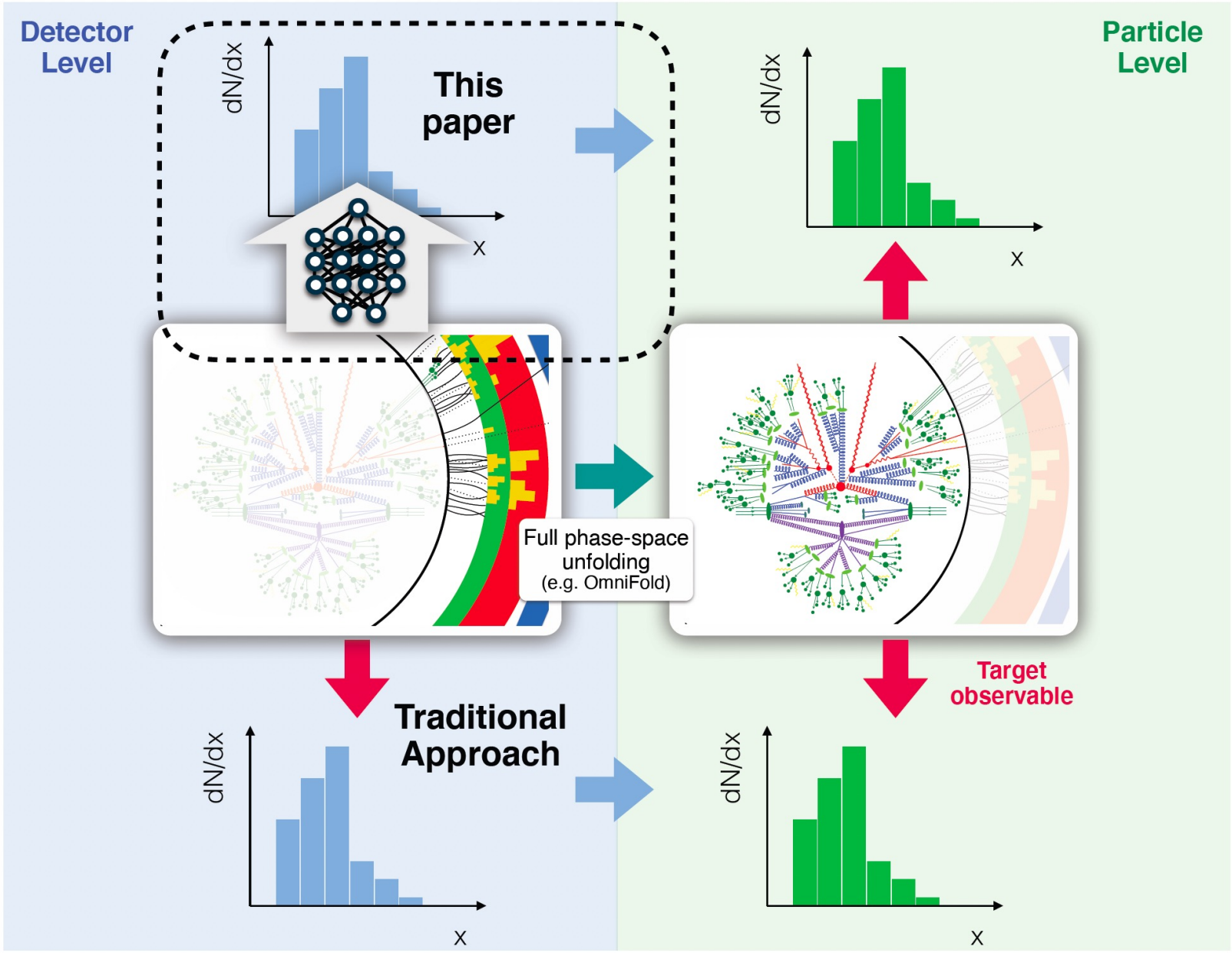
DNN resolution degrades for y<0.15, not seen in ATHENA fastsim study (but understood).



Resolution on x in bins of y_{gen}

*DNN has **better** core resolution than best conventional method (electron at high y, DA at low y).*

DNN distributions much more symmetric, free of large QED radiation tails.



Benefit of ML Approach in Unfolding

We use all useful event information in the **DNN inputs**.

We train the DNN regression *directly* for what we want (the "true" Q^2, y, x), not the particle-level version of the reconstruction variables.

Standard unfolding is still necessary, but it has less to do.

Result: unfolding response matrix is much more diagonal.

- Smaller statistical uncertainties.
- Smaller bin-to-bin correlations.

https://arxiv.org/abs/2203.16722 *(Accepted for publication in JINST)*

Advantages of Machine Learning Reconstruction in Unfolding

H1 full simulation

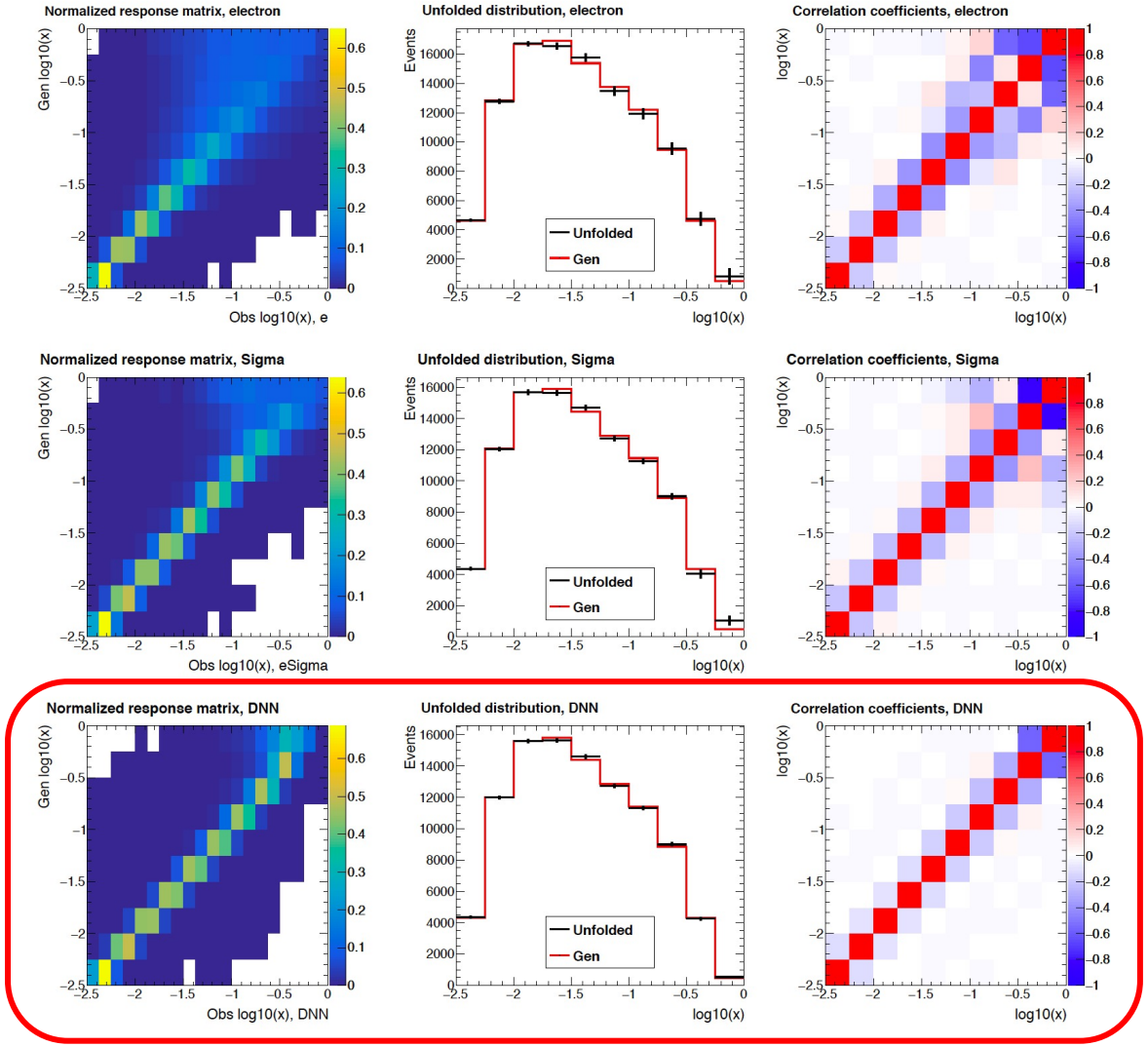
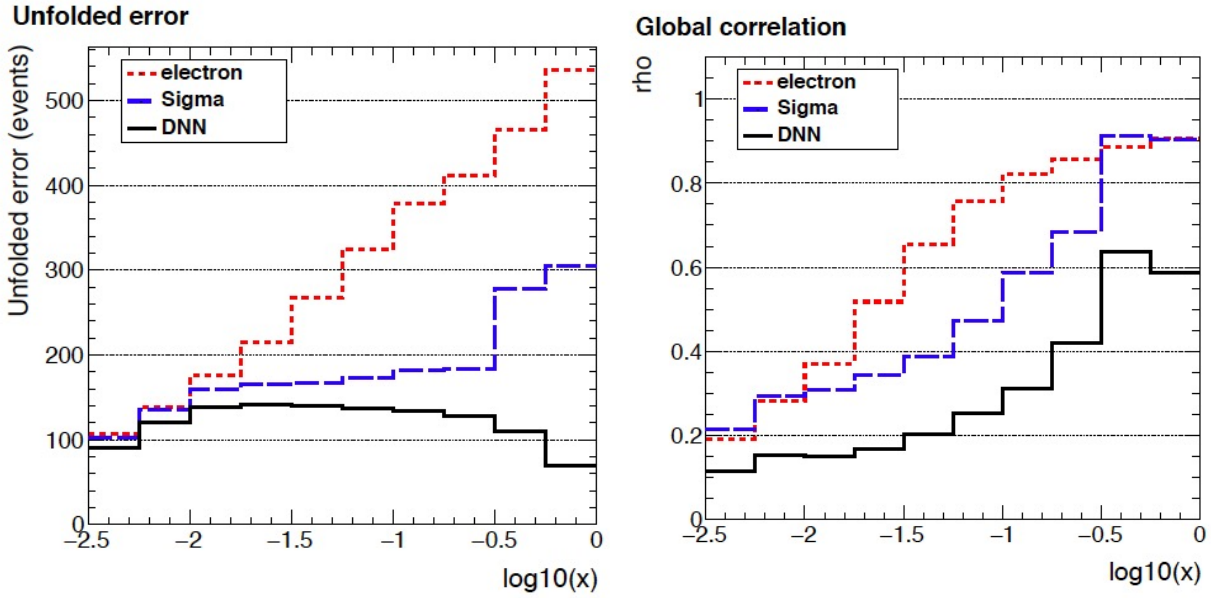


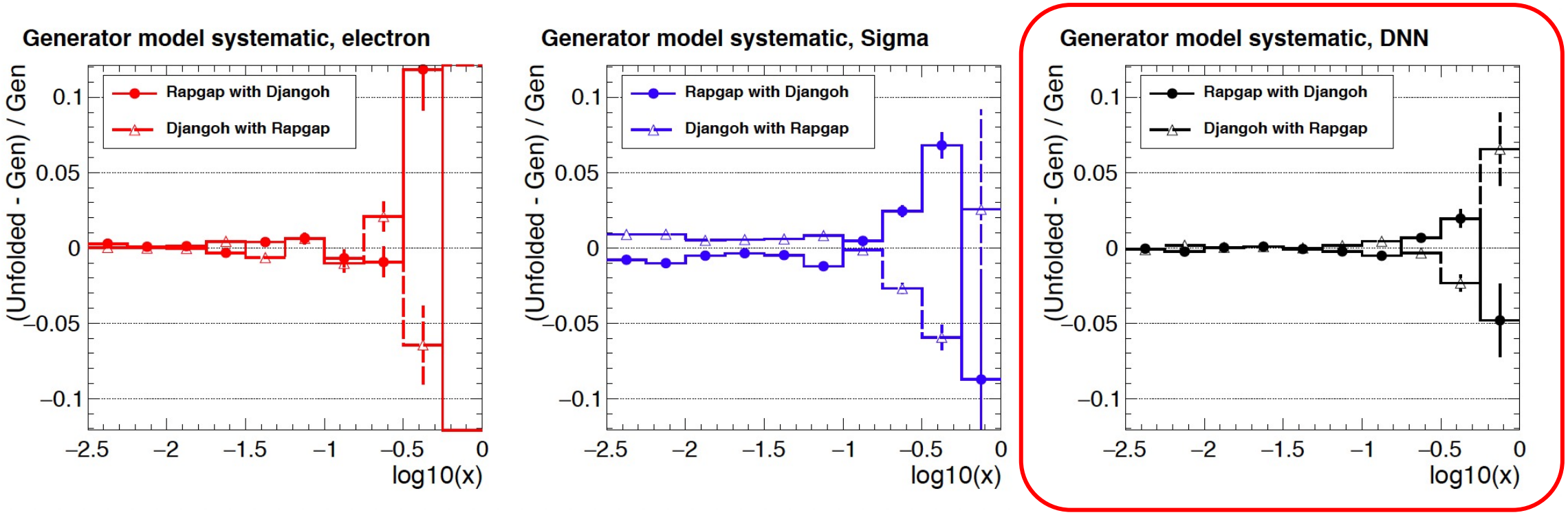
Figure 1. Examples of unfolding $\log_{10}(x)$ in DIS in one dimension for samples of 10^5 events. The response matrix (left), unfolded and gen distributions (middle), and unfolding correlation matrix (right) are shown for the electron (top), Sigma (middle), and DNN (bottom) methods.



The response matrix for the DNN is much more diagonal at high x.

This translates into smaller statistical errors and smaller correlations in the unfolding results at high x.

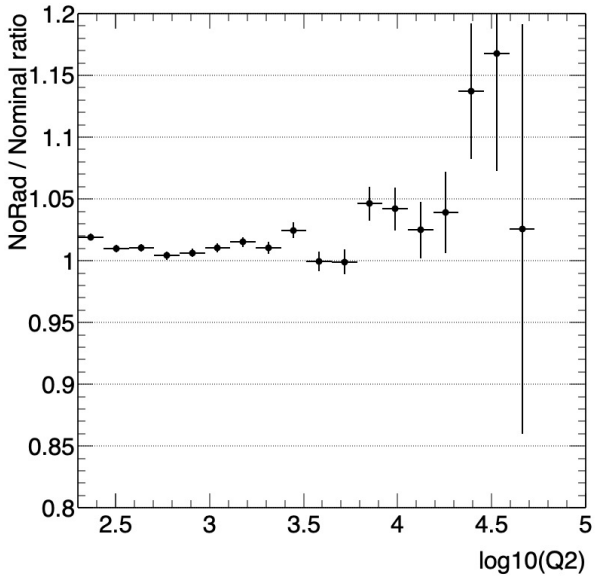
Generator Model Systematic Uncertainty in the Unfolding



Resuts of unfolding pseudodata from one generator using the response matrix from another generator.

Model systematic uncertainty is **substantially reduced in the DNN approach**.

Djangoh, NoRad / Nominal ratios (QED correction)



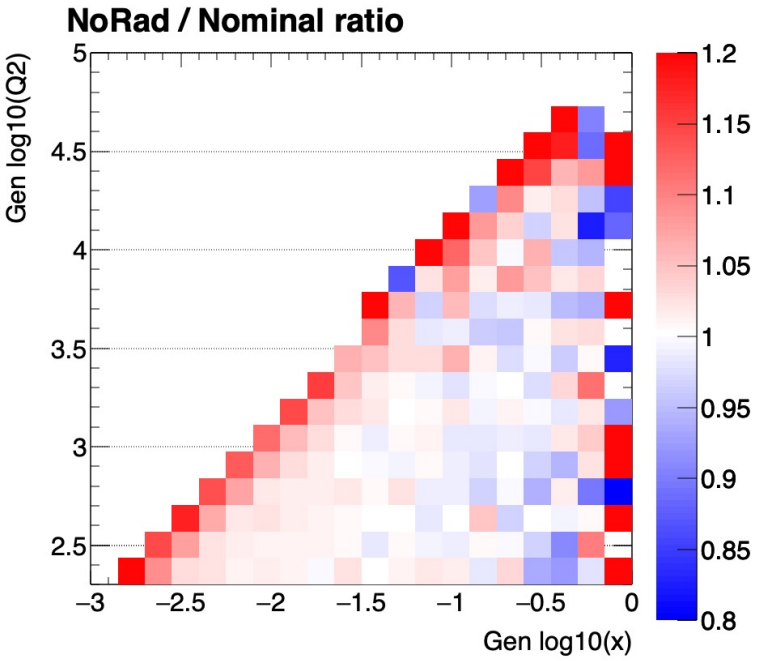
Definition of DNN regression learning targets or the "true" Q^2, y, x.

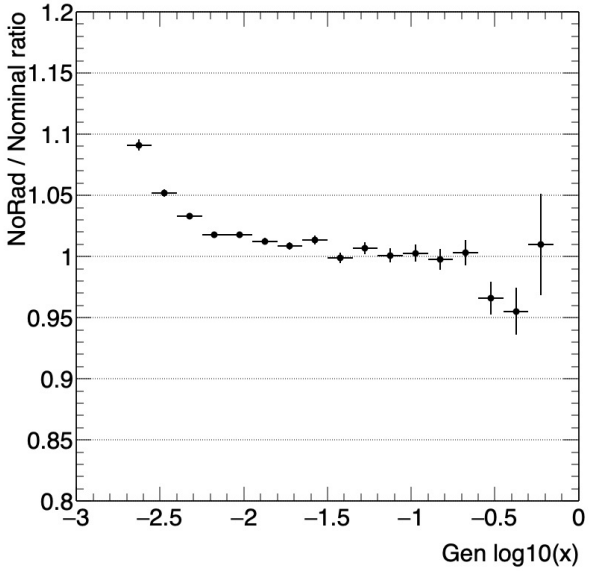
Standard definitions for events with no QED radiation in terms of beam electron ℓ , scattered electron ℓ', and beam proton p 4-vectors.

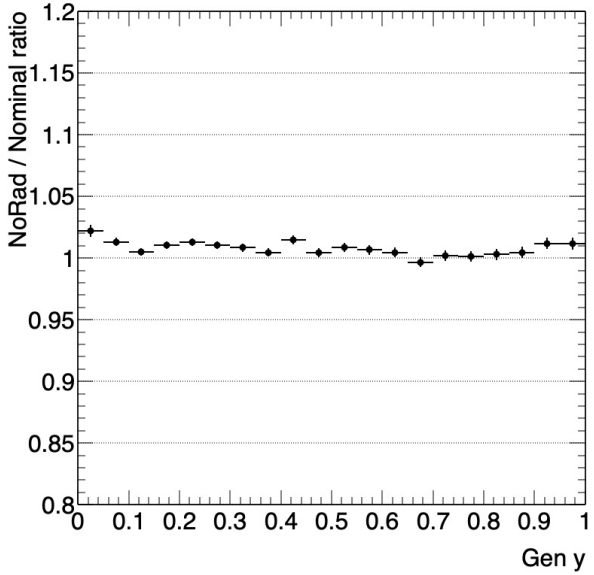
For events with QED radiation

Use post-ISR beam electron 4-vector

Use pre-FSR scattered electron 4-vector







We train the DNN directly for what we want (the true Q^2, y, and x in no-Rad, post-ISR, or pre-FSR).

The unfolding of detector effects is much improved.

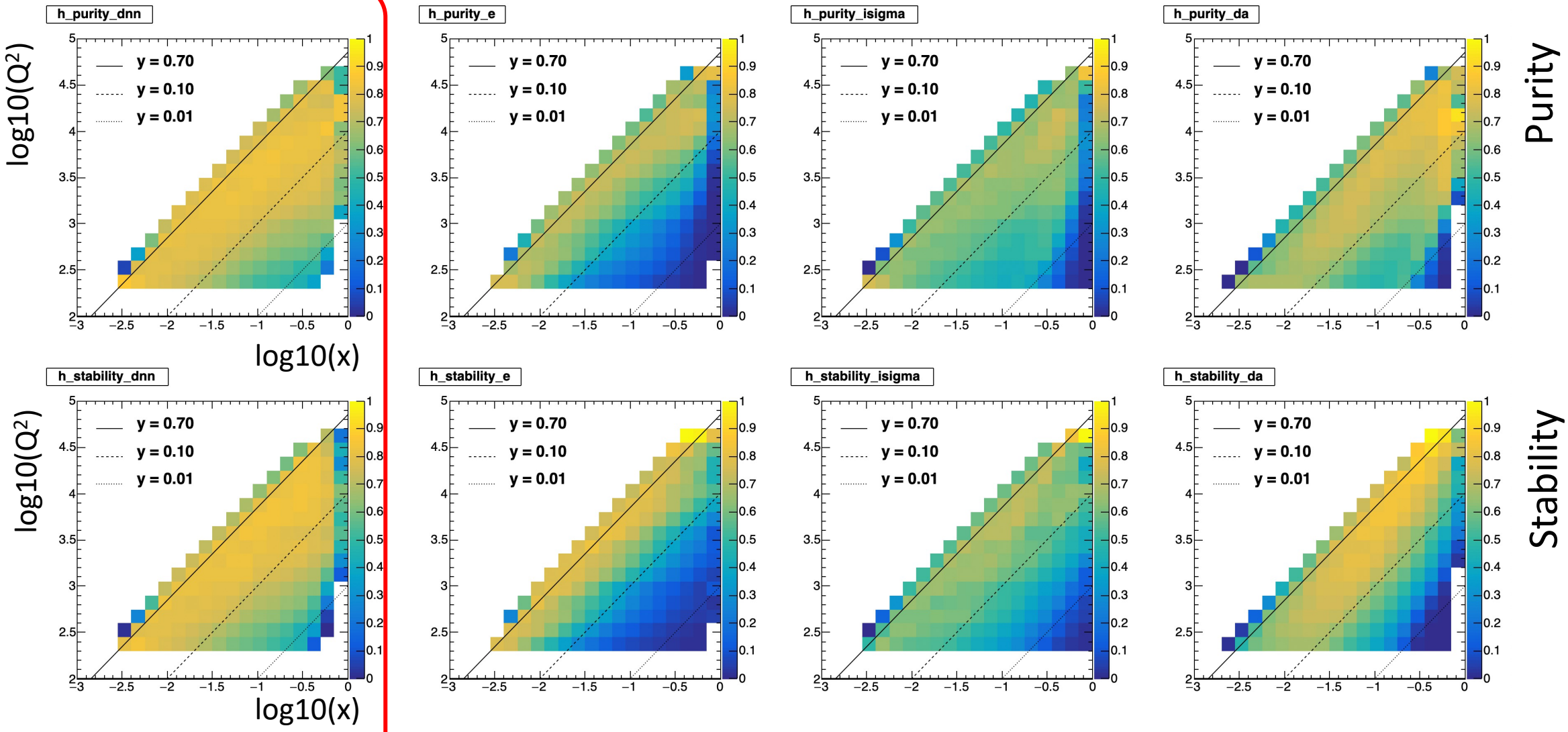
This approach also leads to a **much smaller QED correction**.

Only event selection applied is Gen Q2 > 200

H1 full simulation

Advantages of Machine Learning in Purity and Stability

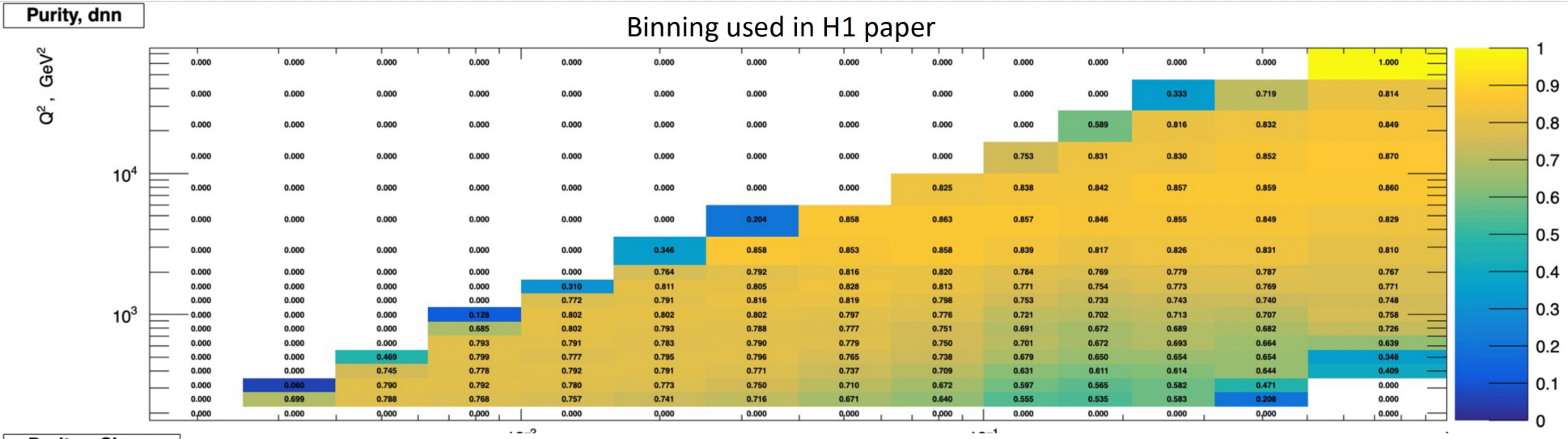
H1 full simulation



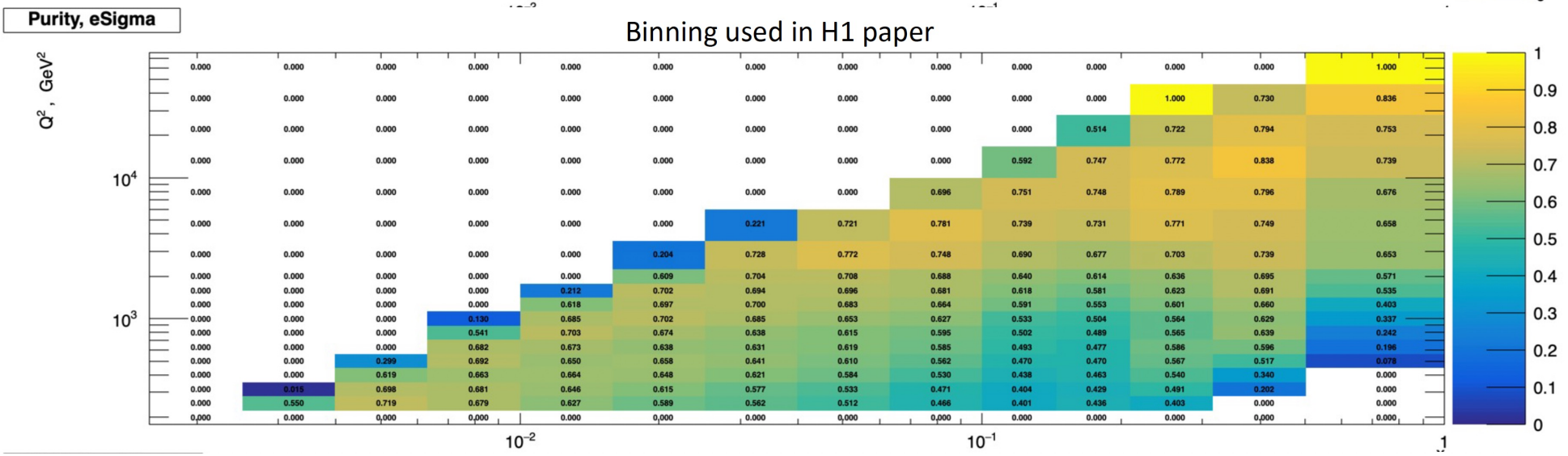
Tighter resolution leads to higher Purity and Stability. Extends the reach of measurements.

Advantages of Machine Learning in Purity and Stability

H1 full simulation



Purity for DNN



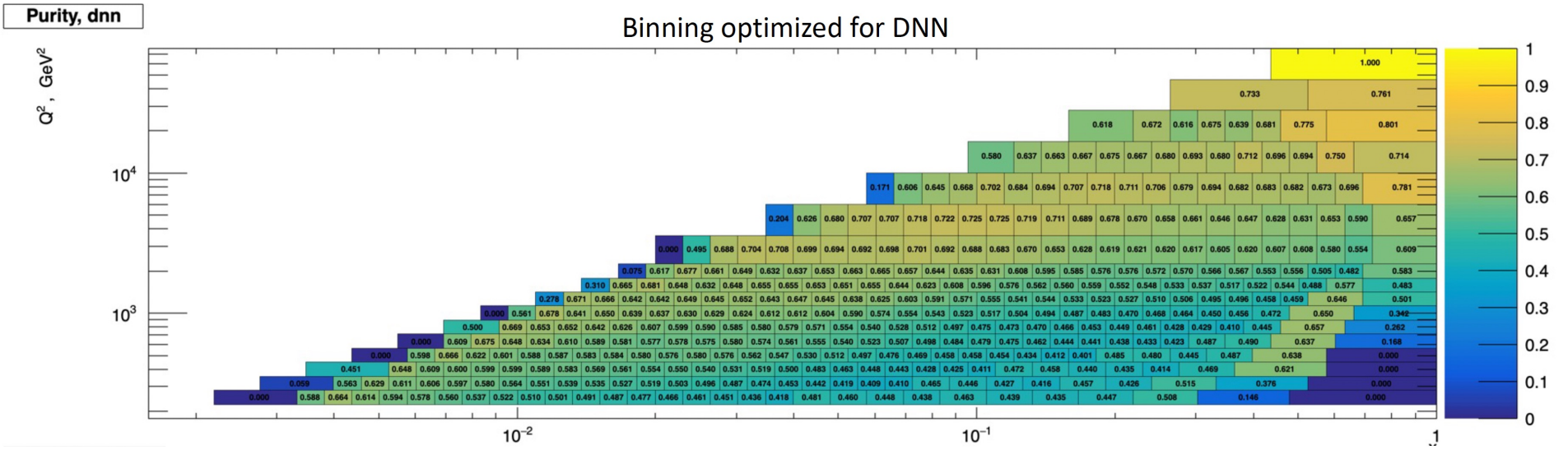
Purity for eSigma

X

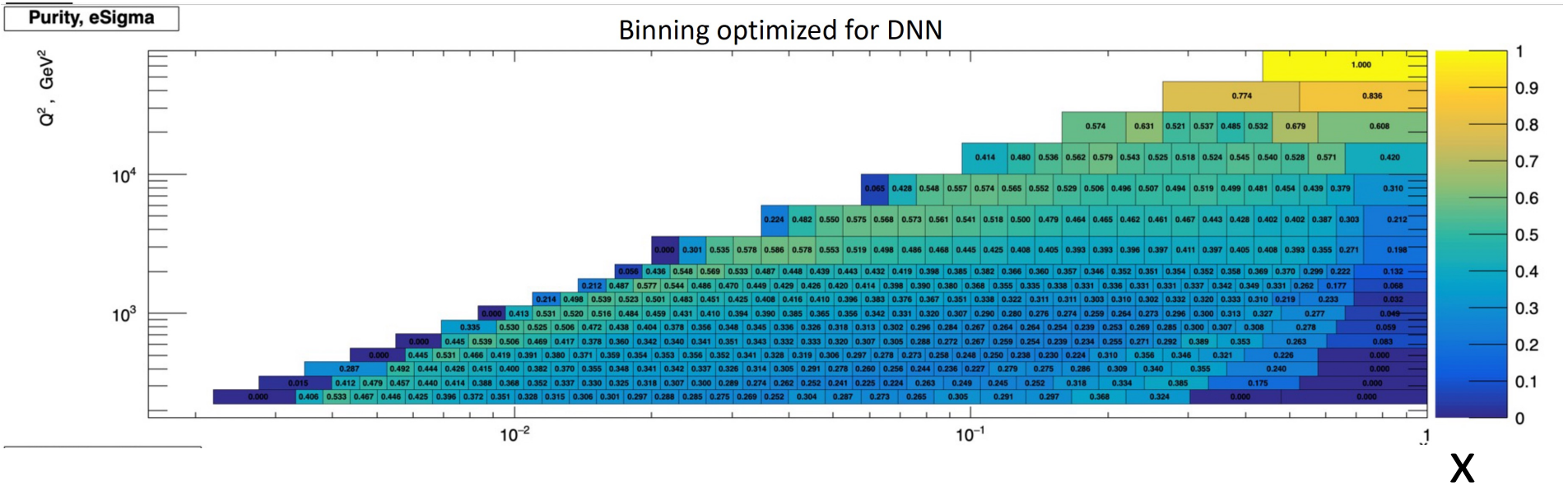
Approximate binning from H1 publication.

Advantages of Machine Learning in Purity and Stability

H1 full simulation



Purity for DNN



Purity for eSigma

Possible finer binning.

DNN extends the reach at high x.

Conclusions

We have applied modern machine learning techniques to reconstructing the kinematics of Deep Inelastic Scattering.

Our method includes observables from the event that allow QED radiation effects to be significantly reduced in the reconstruction.

The DNN approach outperforms all conventional reconstruction methods in the full range of y for $Q^2 > 200\ GeV^2$.

We have demonstrated our method in the full simulation of H1.

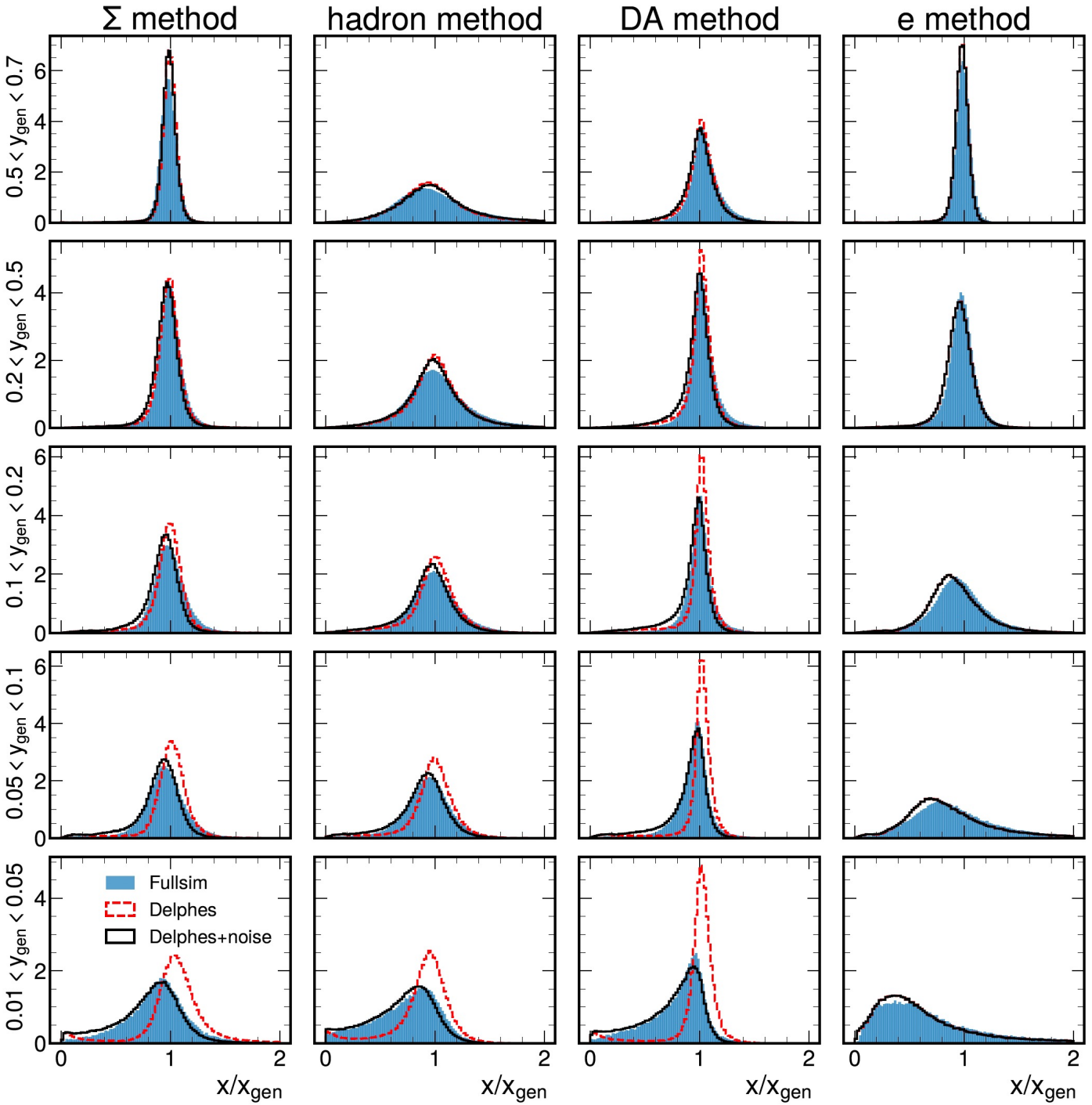
By training directly for what you want, you get better unfolding results and other advantages.

Improved resolution can extend the reach of existing data at high x.

NIMA 1025 (2022) 166164

https://arxiv.org/abs/2203.16722 *(Accepted for publication in JINST)*

Extra Slides

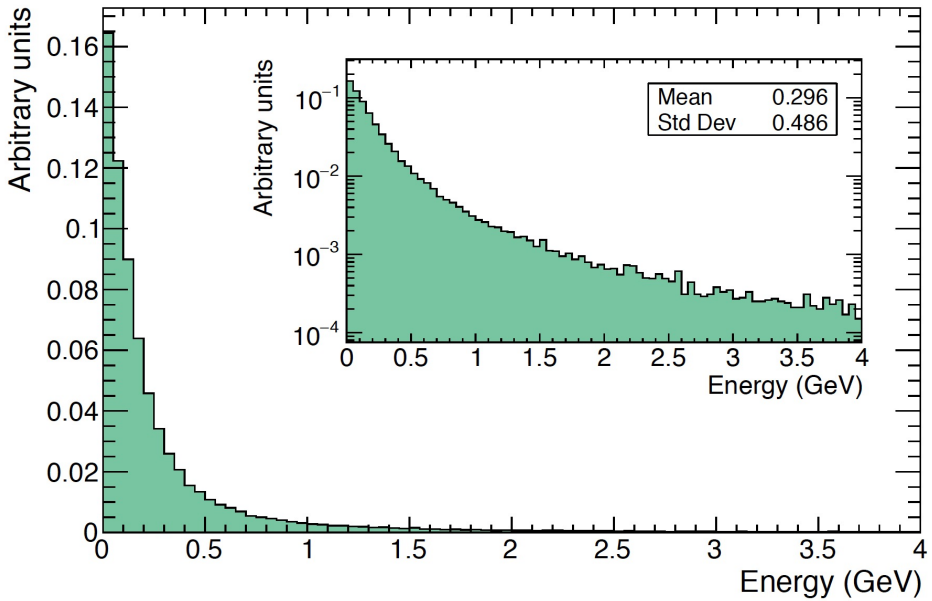


H1 fastsim vs fullsim

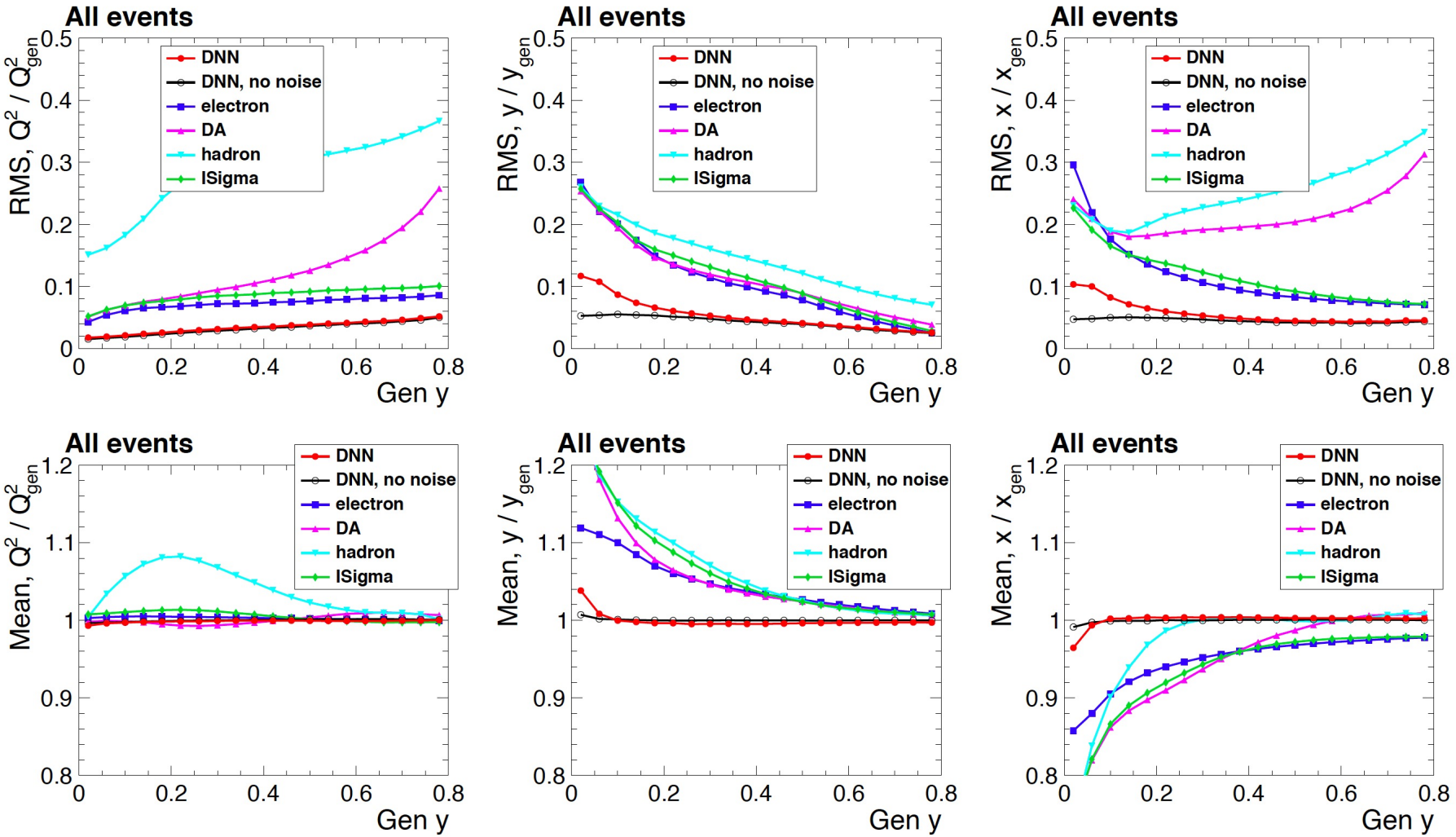
DELPHES fastsim does not include **noise hits in the calorimeter.**

Adding an additional noise-resolution-acceptance component (simple ad-hoc model) to the fastsim gives much better agreement with fullsim.

Trandom::Landau , mu=0, sigma = 0.05 GeV



ATHENA fast simulation with additive resolution effect (Rapgap+Delphes)

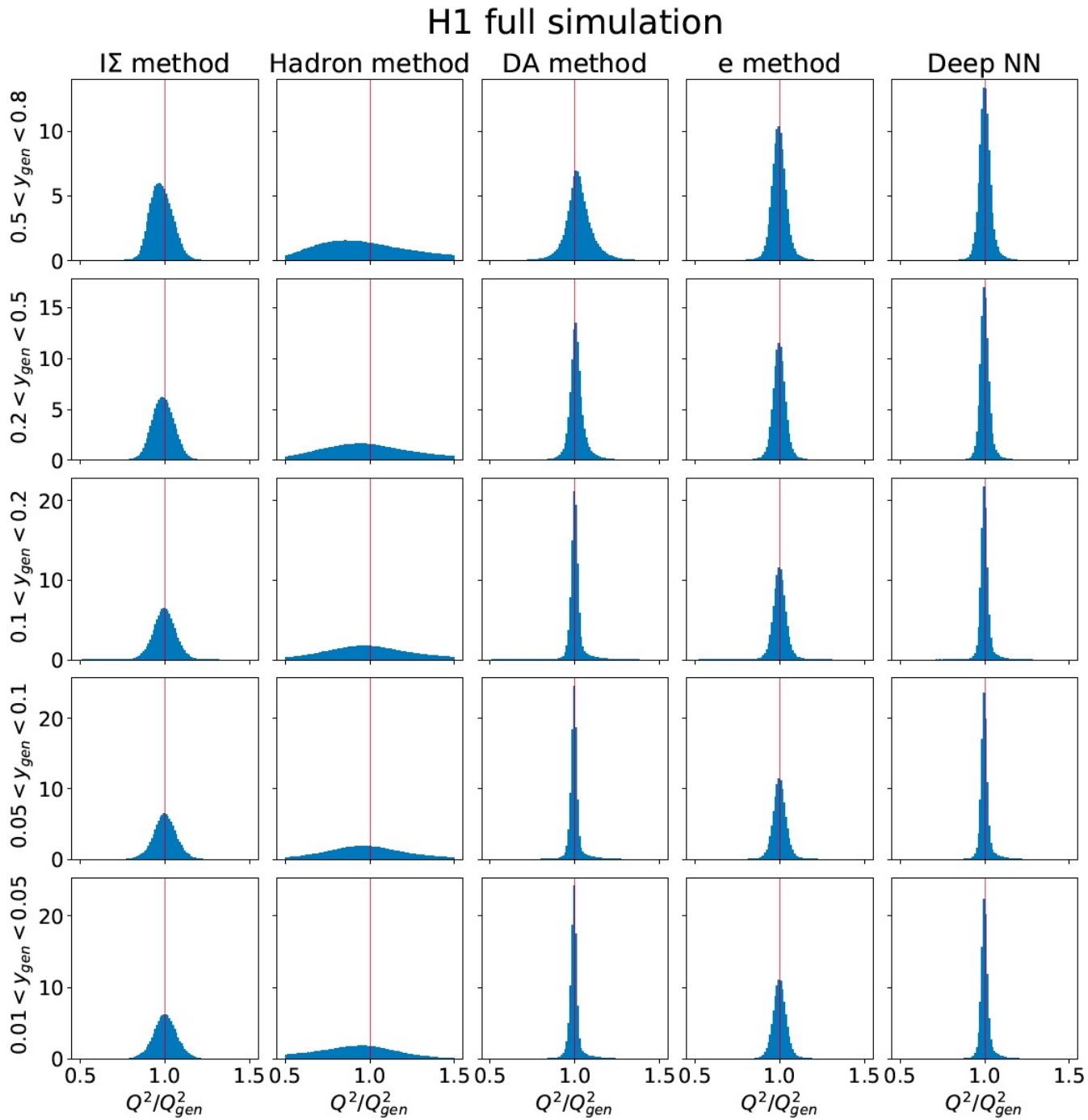


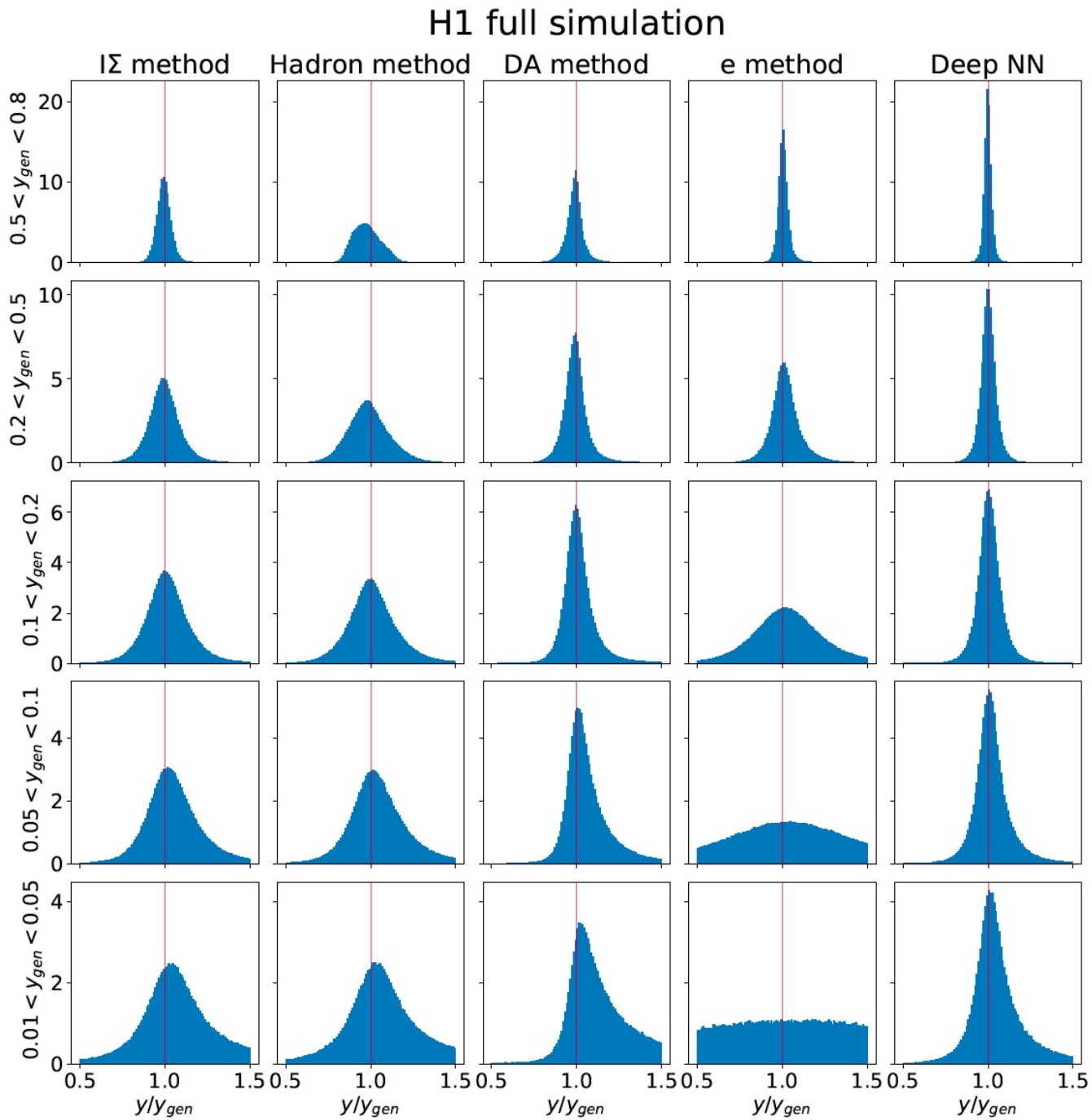
Test of adding same ad-hoc noise model to ATHENA fastsim.

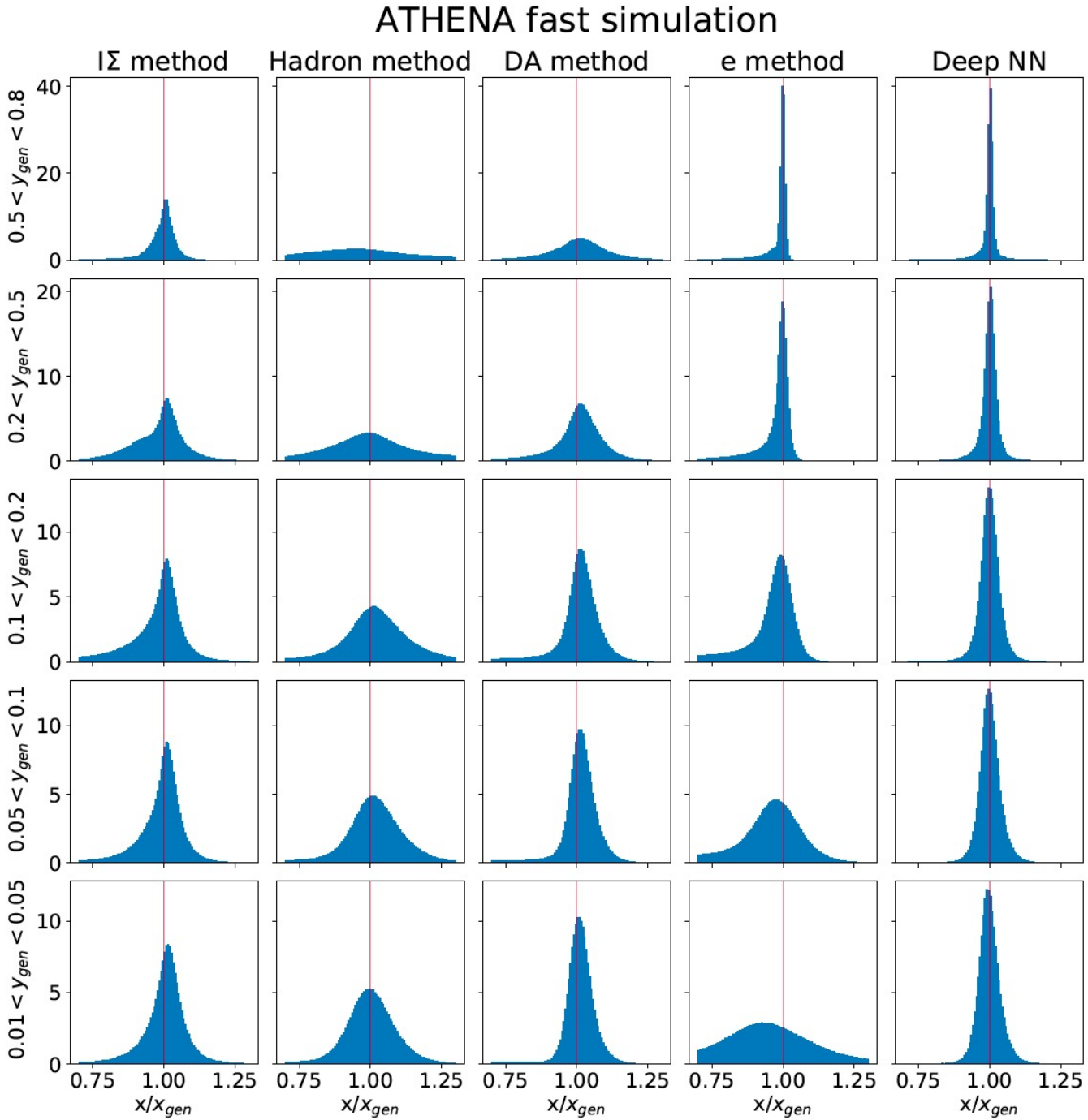
Reran entire procedure, including DNN training, with noise-added fastsim sample.

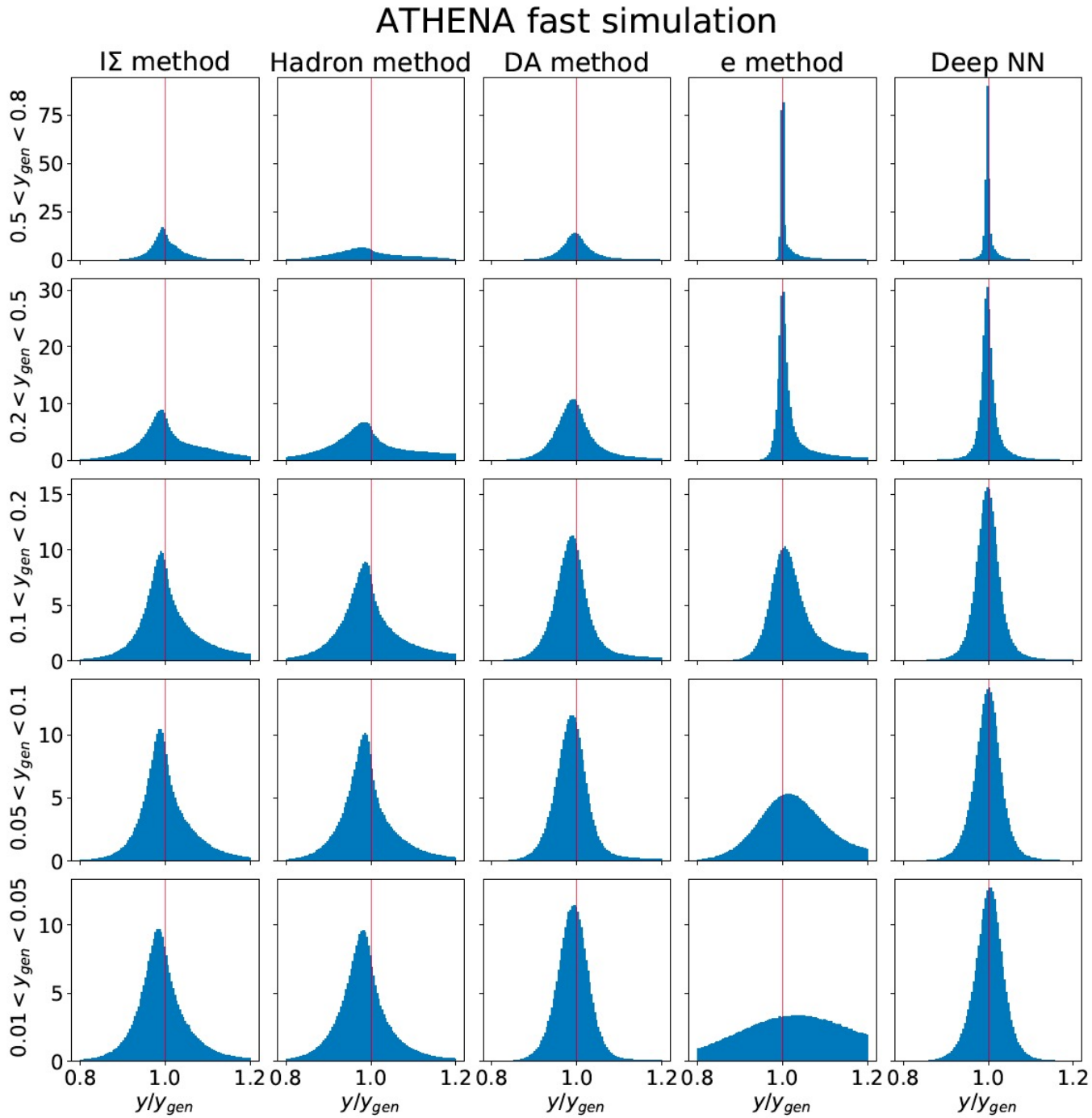
DNN resolution does degrade for y<0.15.

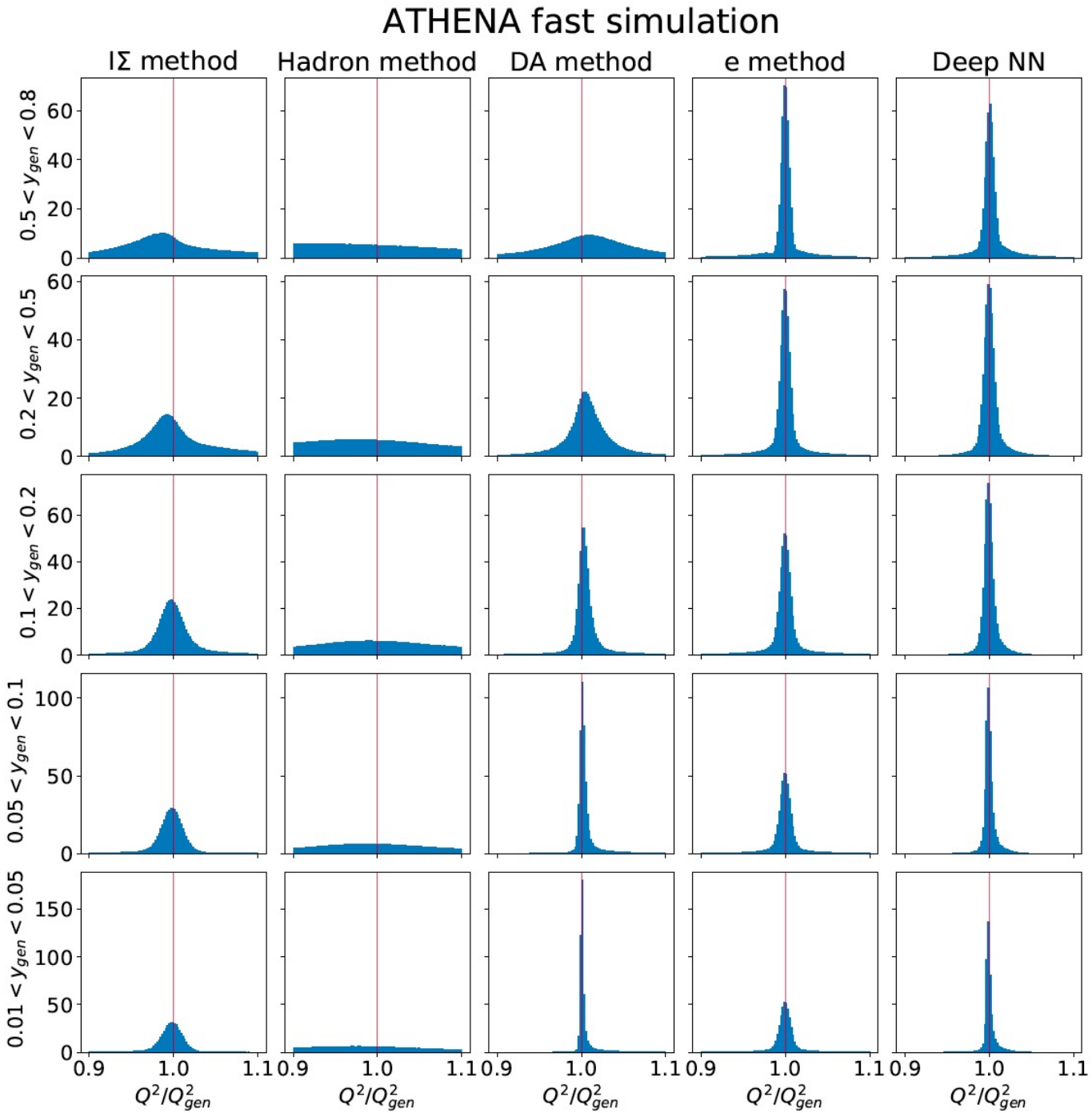
DNN still significantly outperforms conventional methods.


H1 full simulation
IΣ method
Hadron method
DA method
e method
Deep NN
0.5 < y_{gen} < 0.8
0.2 < y_{gen} < 0.5
0.1 < y_{gen} < 0.2
0.05 < y_{gen} < 0.1
0.01 < y_{gen} < 0.05
Q^2/Q^2_{gen}

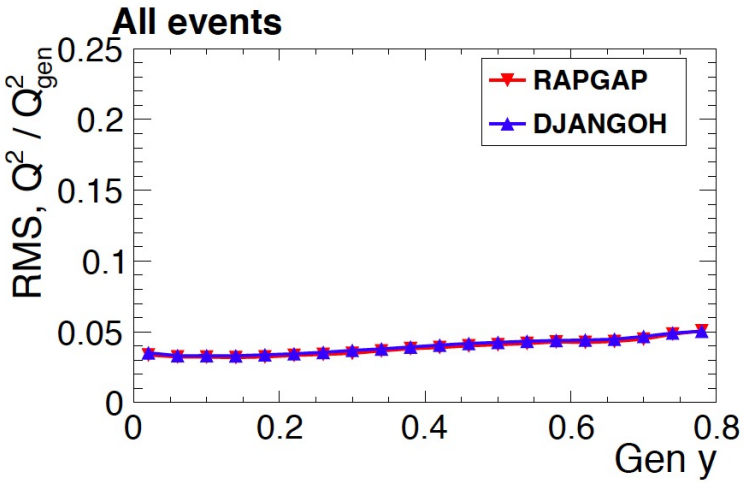

H1 full simulation
IΣ method
Hadron method
DA method
e method
Deep NN
$0.5 < y_{gen} < 0.8$
$0.2 < y_{gen} < 0.5$
$0.1 < y_{gen} < 0.2$
$0.05 < y_{gen} < 0.1$
$0.01 < y_{gen} < 0.05$
y/y_{gen}

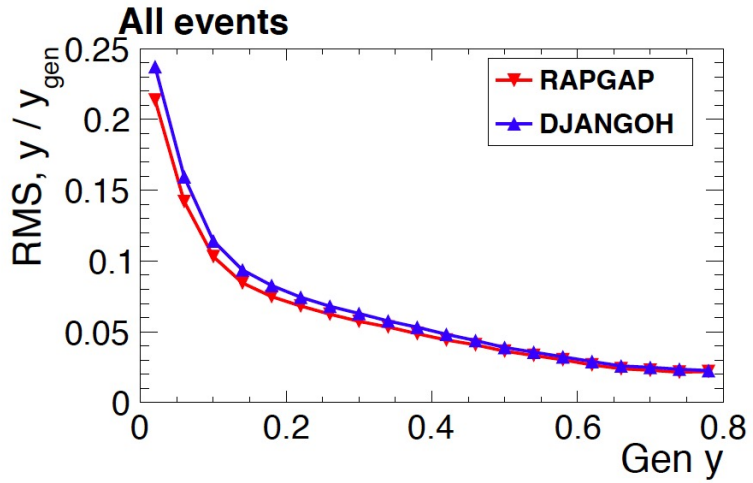

ATHENA fast simulation
IΣ method
Hadron method
DA method
e method
Deep NN
0.5 < y_{gen} < 0.8
0.2 < y_{gen} < 0.5
0.1 < y_{gen} < 0.2
0.05 < y_{gen} < 0.1
0.01 < y_{gen} < 0.05
0.75
1.00
1.25
x/x_{gen}

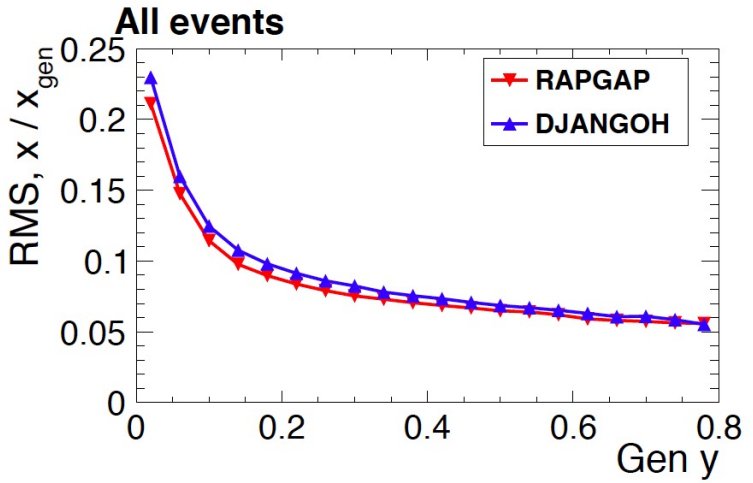

ATHENA fast simulation
IΣ method
Hadron method
DA method
e method
Deep NN
0.5 < y_{gen} < 0.8
0.2 < y_{gen} < 0.5
0.1 < y_{gen} < 0.2
0.05 < y_{gen} < 0.1
0.01 < y_{gen} < 0.05
y/y_{gen}

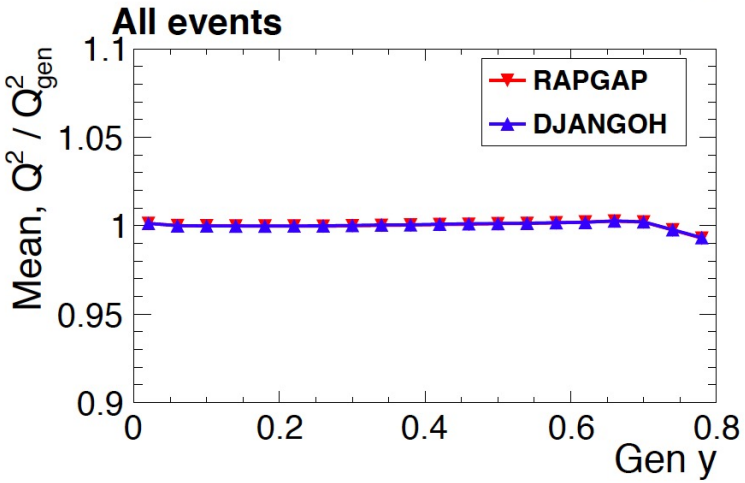

ATHENA fast simulation
IΣ method
Hadron method
DA method
e method
Deep NN
0.5 < y_{gen} < 0.8
0.2 < y_{gen} < 0.5
0.1 < y_{gen} < 0.2
0.05 < y_{gen} < 0.1
0.01 < y_{gen} < 0.05
Q^2/Q^2_{gen}

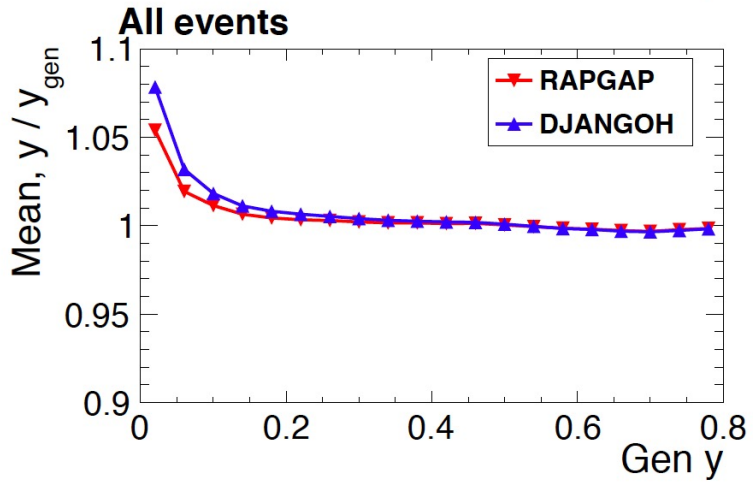
H1 full simulation

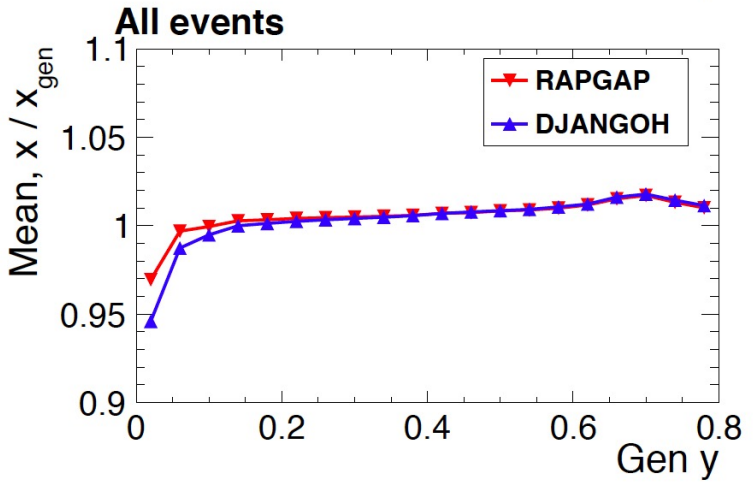






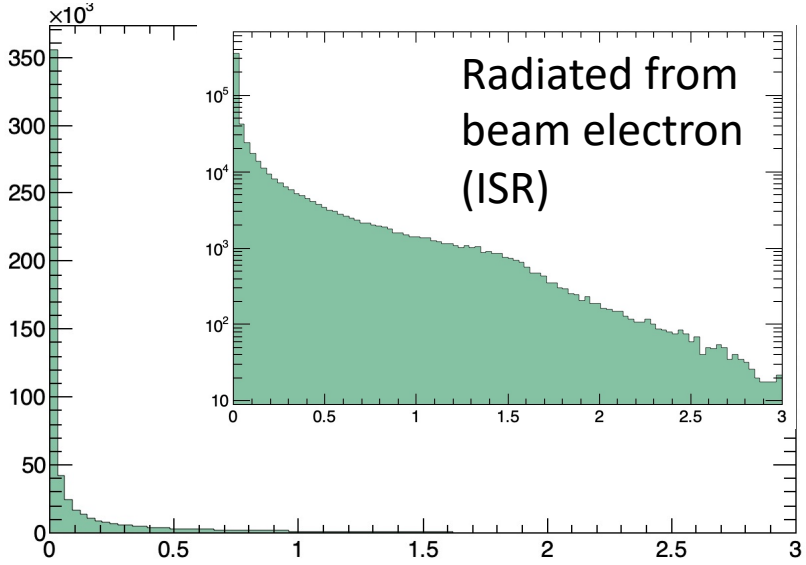




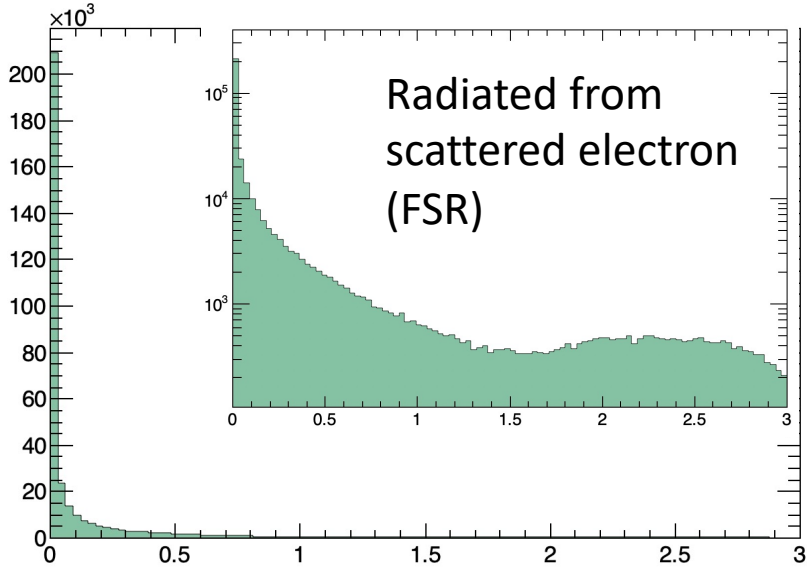


Test of using DNN trained in RAPGAP sample to make predictions in DHANGOH sample.

QED Radiation



Angle between electron and photon (radians)



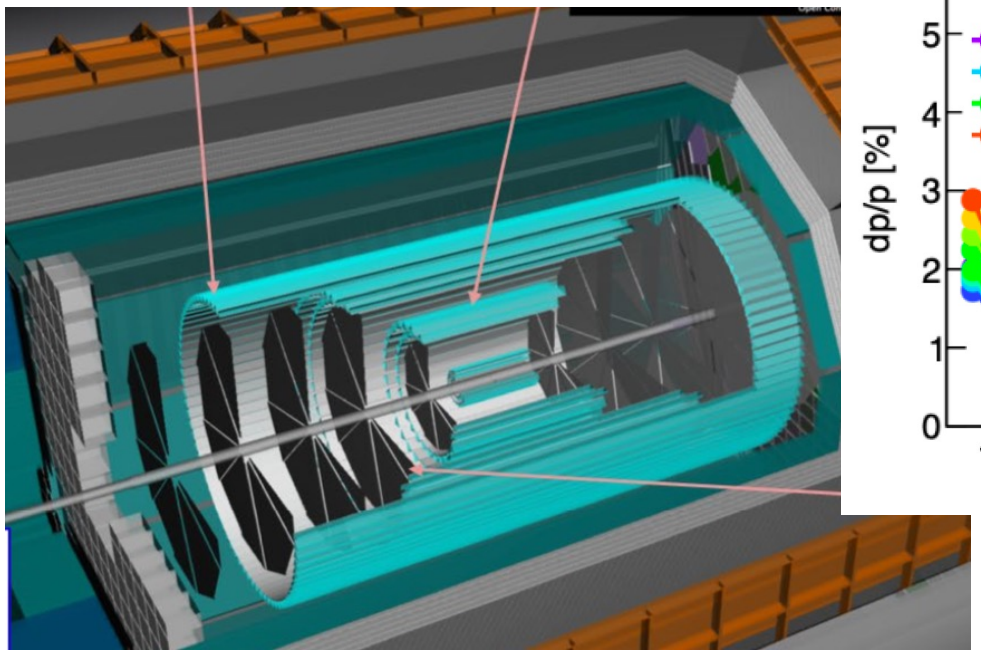
Angle between electron and photon (radians)

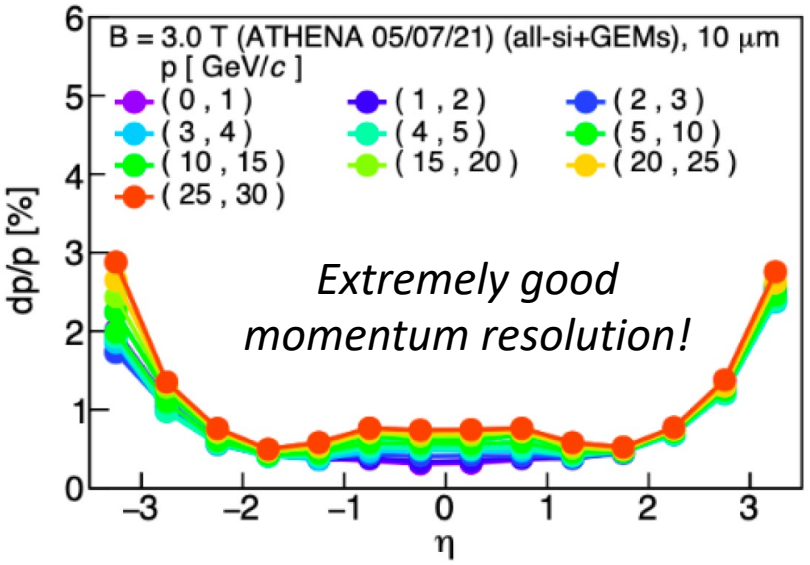
The ATHENA experiment and DELPHES fast simulation

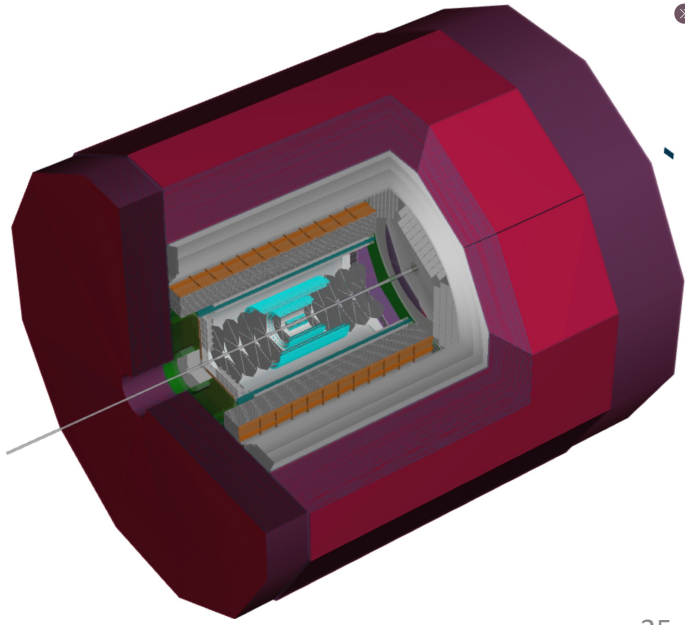
Electron Ion Collider (EIC)
beams: 275 GeV (p), 18 GeV (e)

ATHENA experiment
3 T solenoid
All silicon tracker
Very good particle ID
Large acceptance (-4 < eta < 4)

DELPHES fast simulation
Detailed momentum smearing of gen particles







Event selection for this presentation

Generated $Q^2 > 200\ GeV^2$

32 GeV < event (E-pz) < 40 GeV , (±4 GeV around $2E_e$) reduces QED radiation

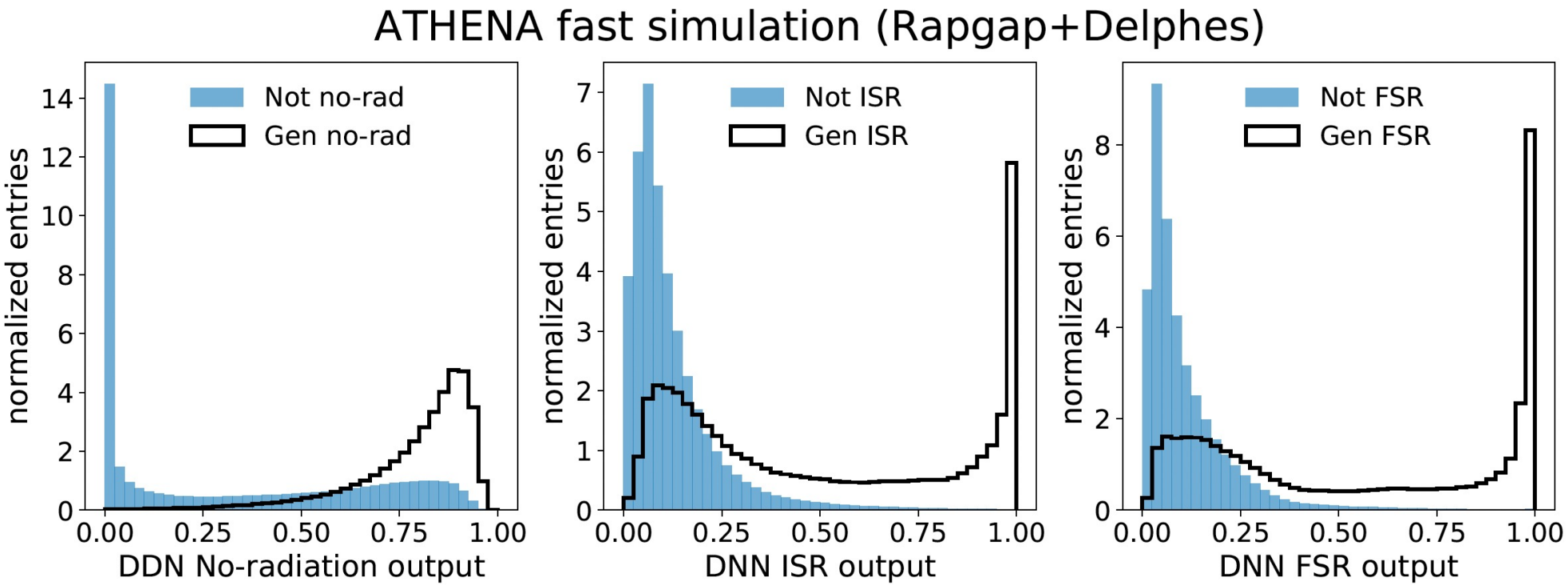
See references in extra slides for more info.

QED radiation : classification DNN

Learning targets: three binary state variables for ISR, FSR, NoR.

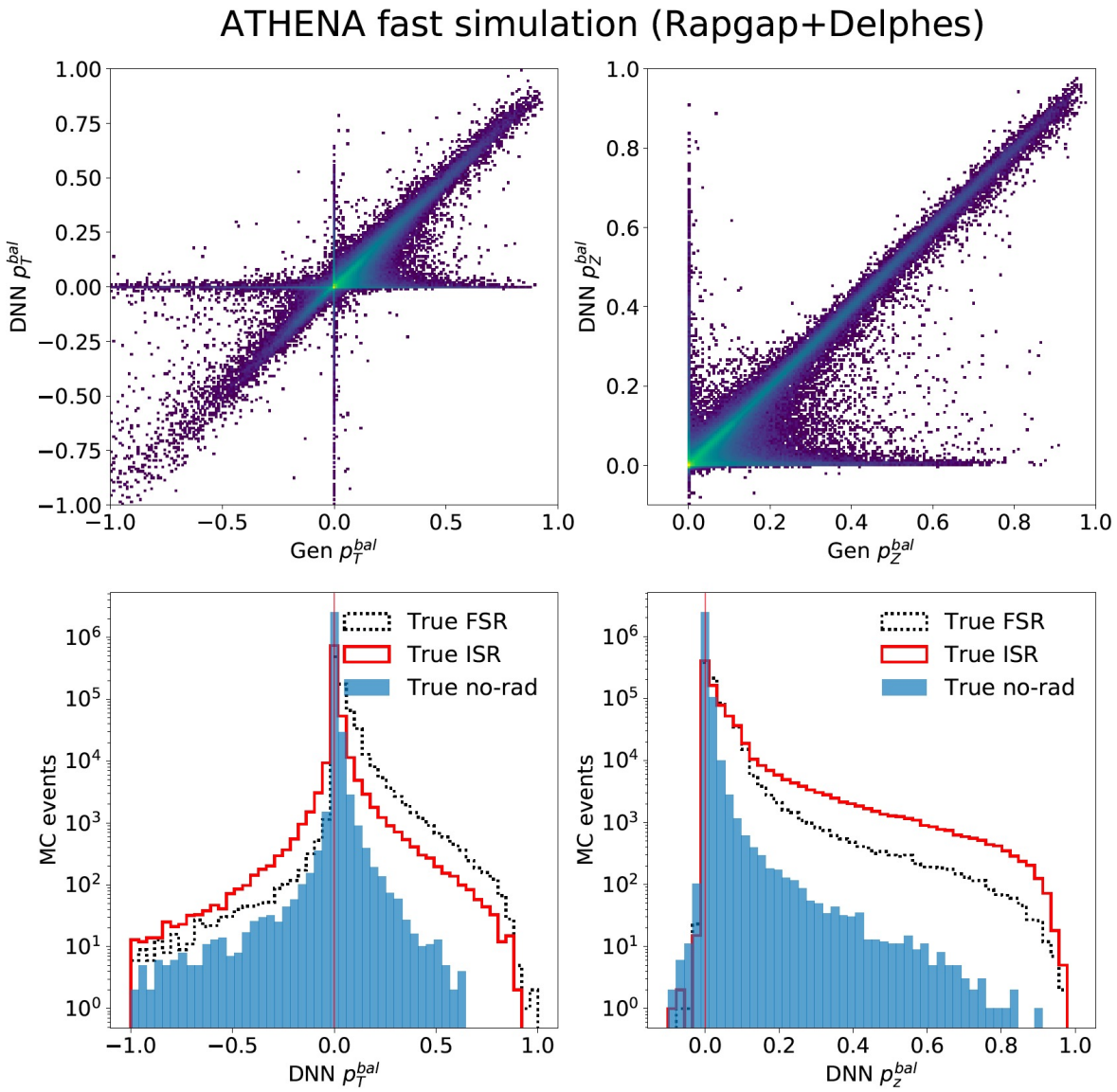
Output layer is 3 nodes for ISR, FSR, NoR with softmax activation (each is 0 to 1, with sum = 1).

Loss function is categorical cross entropy.



Some QED radiation events are strongly identified.

QED radiation : regression DNN for p_z^{bal} and p_T^{bal}



Learning targets are gen values of p_z^{bal} and p_T^{bal}.

Output layer is two nodes for p_z^{bal} and p_T^{bal} with linear activation.

Loss function is Huber.

DNN accurately estimates p_z^{bal} and p_T^{bal} in many events.

DIS reconstruction : regression DNN for Q^2, y, x

Learning targets are log of gen values of Q^2, y, and x.

Three output nodes for log of Q^2, y, and x with linear activation.

Loss function is Huber.

We tried three approaches:

1. Add 3 QED classification outputs (ISR, FSR, NoR) to 15 other inputs.
2. Add 2 QED regression outputs (p_z^{bal} and p_T^{bal}) to 15 other inputs.
3. Use the same 15 inputs as in QED DNNs.

DIS reconstruction : regression DNN for Q^2, y, x

Learning targets are log of gen values of Q^2, y, and x.

Three output nodes for log of Q^2, y, and x with linear activation.

Loss function is Huber.

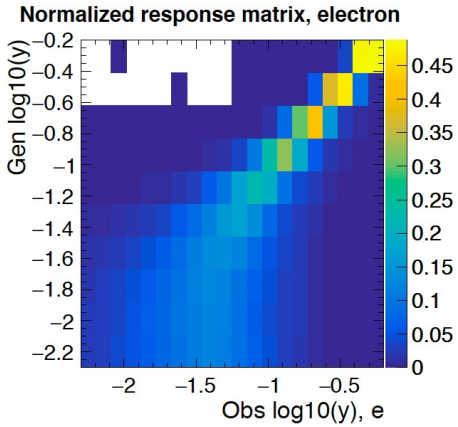
We tried three approaches:

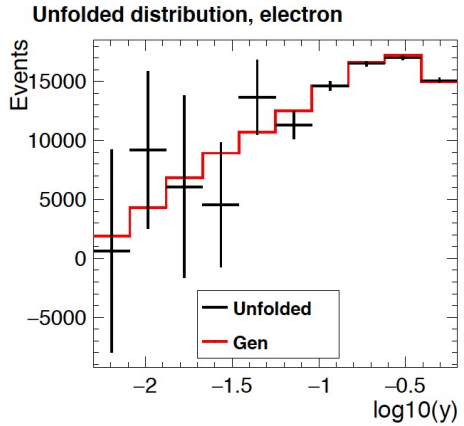
1. Add 3 QED classification outputs (ISR, FSR, NoR) to 15 other inputs.
2. Add 2 QED regression outputs (p_z^{bal} and p_T^{bal}) to 15 other inputs.
3. Use the same 15 inputs as in QED DNNs.

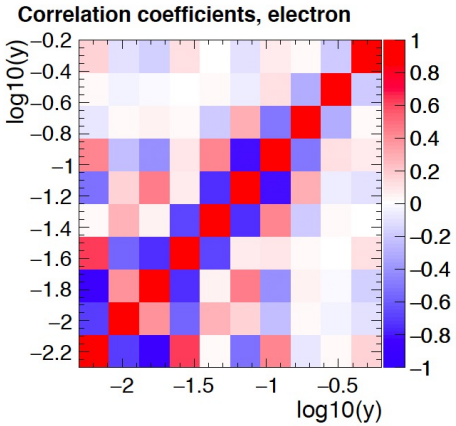
All 3 give essentially identical results!

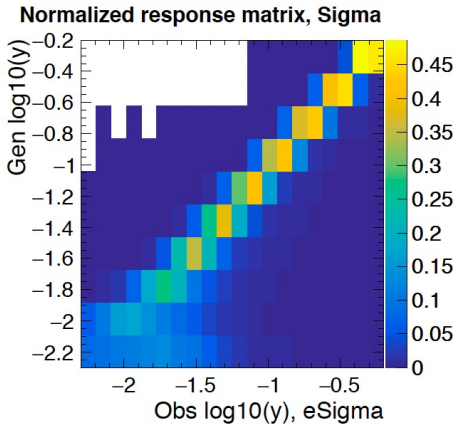
We choose the simplest option (3).

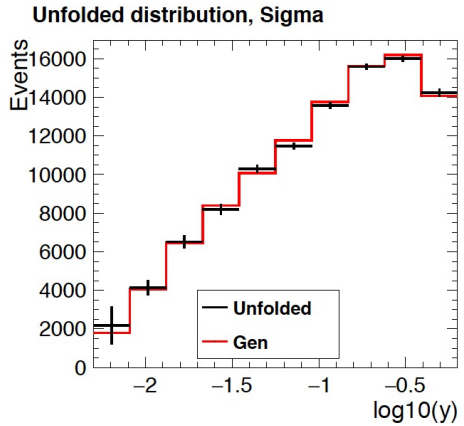
H1 full simulation

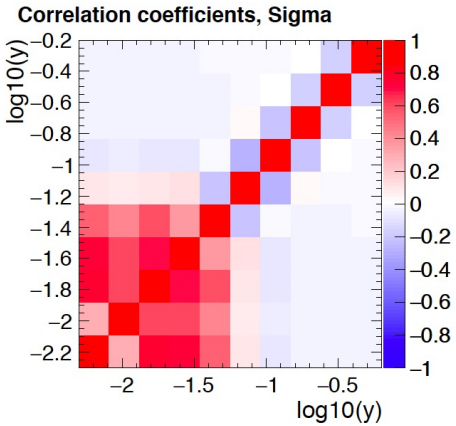


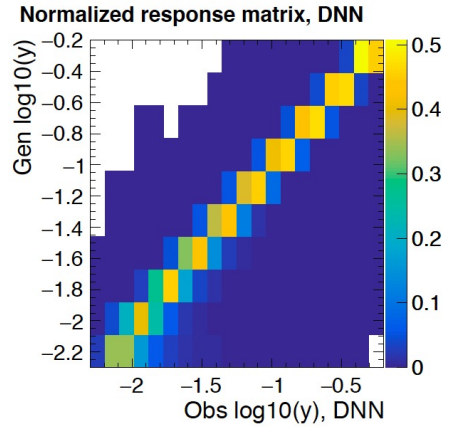


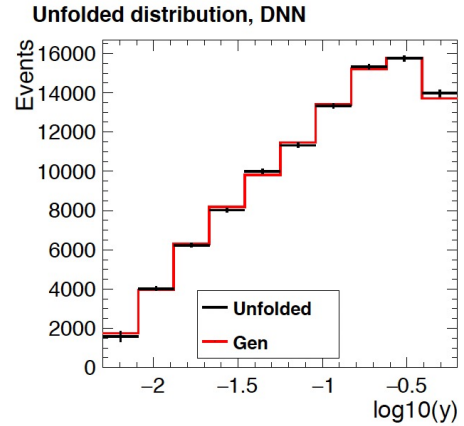


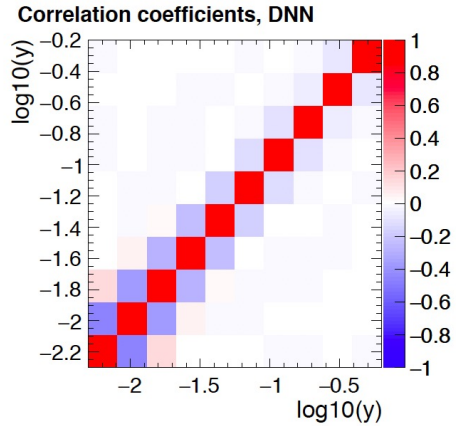




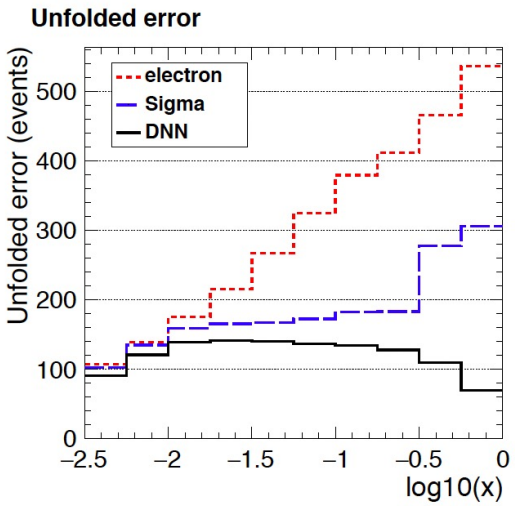


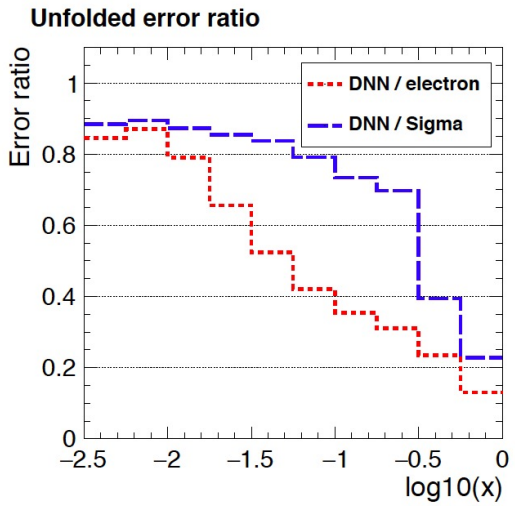


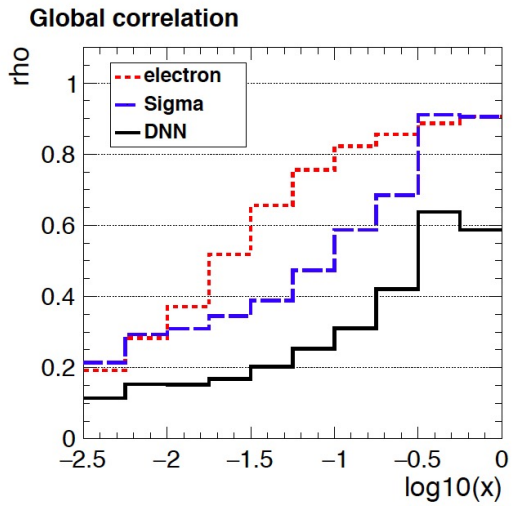


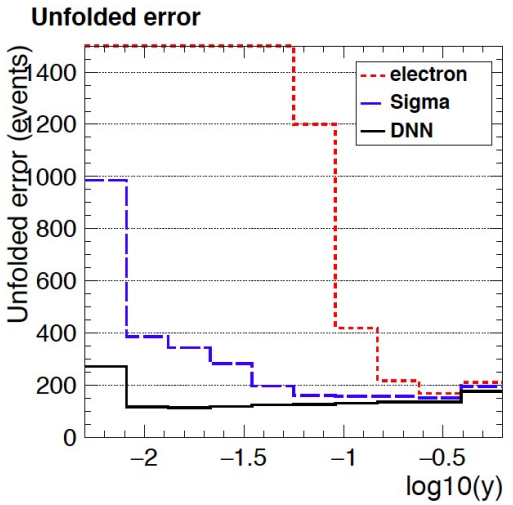


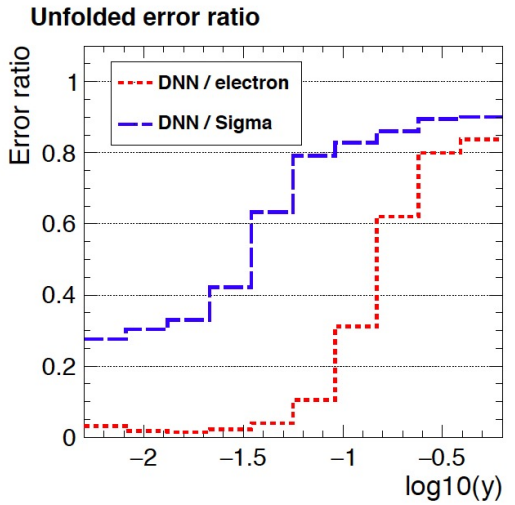
H1 full simulation

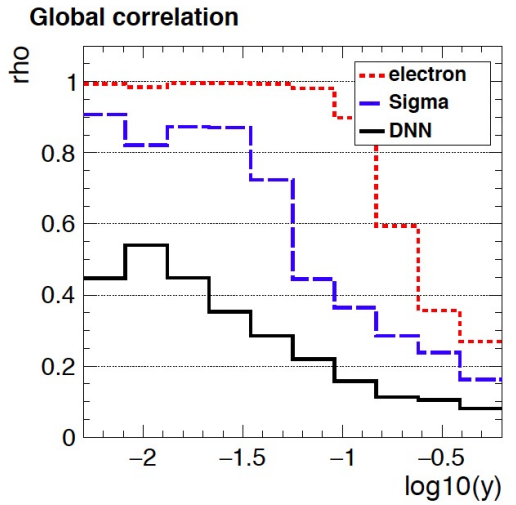




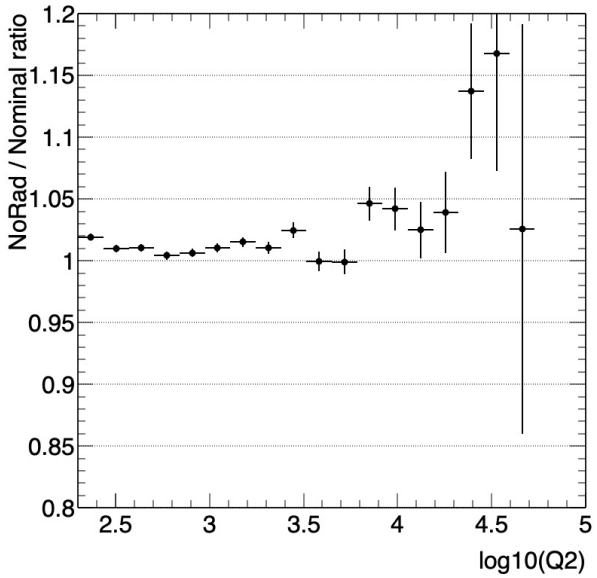








Djangoh, NoRad / Nominal ratios (QED correction)



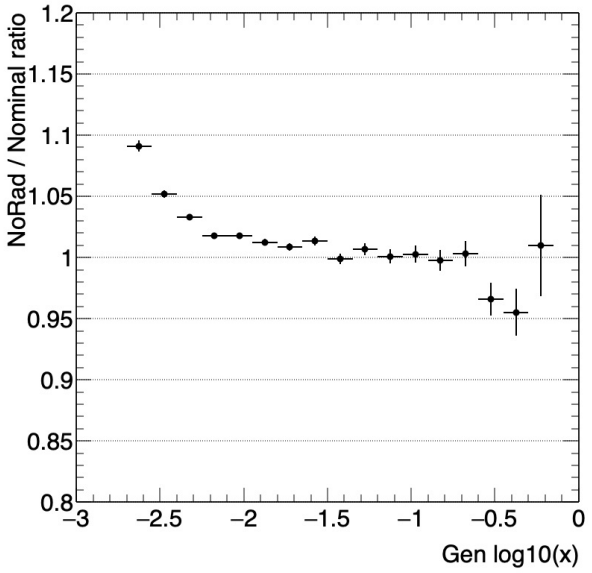
Definition of DNN regression learning targets or the "true" Q^2, y, x.

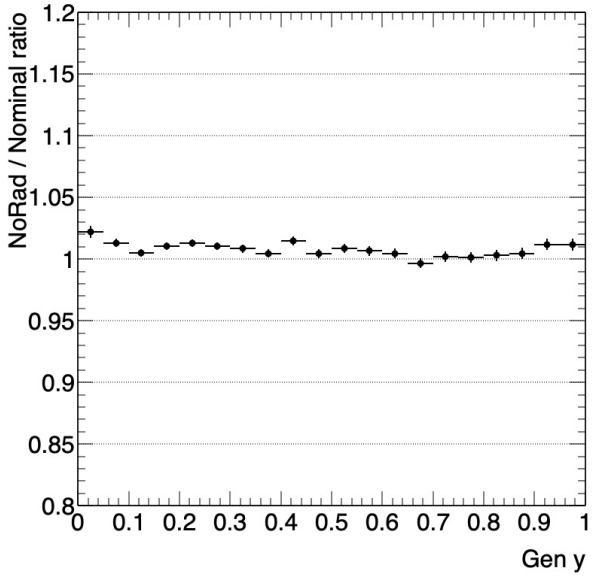
Standard definitions for events with no QED radiation in terms of beam electron ℓ , scattered electron ℓ', and beam proton p 4-vectors.

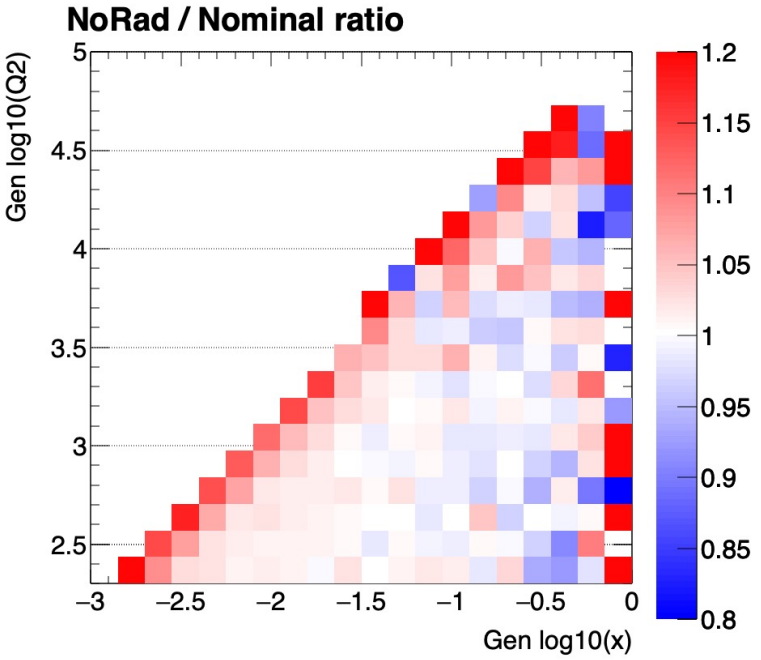
For events with QED radiation

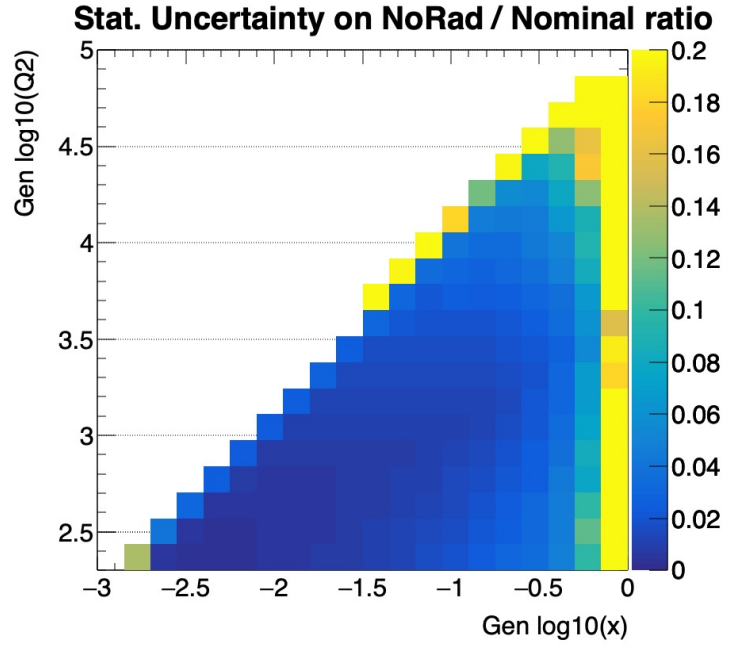
Use post-ISR beam electron 4-vector

Use pre-FSR scattered electron 4-vector









Only event selection applied is Gen Q2 > 200

H1 full simulation

arXiv:1206.7007 : H1 high Q2 Cross Section paper

5.5 Cross Section Measurement

For both the NC and CC analyses the selected event samples are corrected for detector acceptance, efficiencies and migrations using the simulation and converted to QED corrected cross sections. The quality of the simulation, in which all selection efficiency effects are included, is shown in figures 1-7 and gives a reliable determination of detector acceptance. The accessible kinematic ranges of the measurements depend on the resolution of the reconstructed kinematics and are determined by requiring the purity and stability of any measurement bin to be larger than 30% as determined from signal MC. The purity is defined as the fraction of events generated and reconstructed in a measurement bin from the total number of events reconstructed in the bin. The stability is the ratio of the number of events generated and reconstructed in a bin to the number of events generated in that bin. The detector acceptance, $\mathcal{A}$, is obtained from the ratio of stability divided by purity and corrects the measured signal event yields for detector effects including resolution smearing and selection efficiency.

Purity : denominator is histogram binned in observed x and Q2.

Stability : denominator is histogram binned in gen x and Q2.

Numerator for both Purity and Stability : histogram where both obs and gen are in same bin.

Table 1
Summary of basic reconstruction methods that employ only three out of five quantities: E_0 (electron-beam energy), E and θ (scattered electron energy and polar angle), Σ and γ (longitudinal energy–momentum balance, $\Sigma = \sum_{\mathrm{HFS}}(E_i - p_{z,i})$, and the inclusive angle of the HFS). Alternatively, the A4 method makes use of the HFS total energy E_h. Shorthand notations are used for the longitudinal energy–momentum balance of the electron, Σ_e, and for the transverse momentum of the HFS, T. The $E\Sigma T$ and $E_0 ET$ methods are under-constrained and have two solutions, referring to two possible electron polar angles, and several more are existent when using E_h (see Ref. [7] for two examples). The two bottom rows provide the equations of the Σ and $e\Sigma$-methods, which combine quantities of different basic reconstruction methods, while further methods (like the PT (rDΣ), DΣ, r$e\Sigma$ or mixed method) are found, e.g., in Ref. [9,11].

Method name	Observables	y	Q^2	$x \cdot E_p$
Electron (e)	$[E_0,\ E,\ \theta]$	$1 - \frac{\Sigma_e}{2E_0}$	$\frac{E^2 \sin^2\theta}{1-y}$	$\frac{E(1+\cos\theta)}{2y}$
Double angle (DA) [6,7]	$[E_0,\ \theta,\ \gamma]$	$\frac{\tan\frac{\gamma}{2}}{\tan\frac{\gamma}{2}+\tan\frac{\theta}{2}}$	$4E_0^2 \cot^2\frac{\theta}{2}(1-y)$	$\frac{Q^2}{4E_0 y}$
Hadron (h, JB) [4]	$[E_0,\ \Sigma,\ \gamma]$	$\frac{\Sigma}{2E_0}$	$\frac{T^2}{1-y}$	$\frac{Q^2}{2\Sigma}$
ISigma (IΣ) [9]	$[E,\ \theta,\ \Sigma]$	$\frac{\Sigma}{\Sigma+\Sigma_e}$	$\frac{E^2 \sin^2\theta}{1-y}$	$\frac{E(1+\cos\theta)}{2y}$
IDA [7]	$[E,\ \theta,\ \gamma]$	y_{DA}	$\frac{E^2 \sin^2\theta}{1-y}$	$\frac{E(1+\cos\theta)}{2y}$
$E_0 E\Sigma$	$[E_0,\ E,\ \Sigma]$	y_h	$4E_0 E - 4E_0^2(1-y)$	$\frac{Q^2}{2\Sigma}$
$E_0\theta\Sigma$	$[E_0,\ \theta,\ \Sigma]$	y_h	$4E_0^2 \cot^2\frac{\theta}{2}(1-y)$	$\frac{Q^2}{2\Sigma}$
$\theta\Sigma\gamma$ [8]	$[\theta,\ \Sigma,\ \gamma]$	y_{DA}	$\frac{T^2}{1-y}$	$\frac{Q^2}{2\Sigma}$
Double energy (A4) [7]	$[E_0,\ E,\ E_h]$	$\frac{E-E_0}{(xE_p)-E_0}$	$4E_0 y(xE_p)$	$E + E_h - E_0$
$E\Sigma T$	$[E,\ \Sigma,\ T]$	$\frac{\Sigma}{\Sigma+E\pm\sqrt{E^2+T^2}}$	$\frac{T^2}{1-y}$	$\frac{Q^2}{2\Sigma}$
$E_0 ET$	$[E_0,\ E,\ T]$	$\frac{2E_0-E\mp\sqrt{E^2-T^2}}{2E_0}$	$\frac{T^2}{1-y}$	$\frac{Q^2}{4E_0 y}$
Sigma (Σ) [9]	$[E_0,\ E,\ \Sigma,\ \theta]$	$y_{\mathrm{I}\Sigma}$	$Q^2_{\mathrm{I}\Sigma}$	$\frac{Q^2}{4E_0 y}$
eSigma ($e\Sigma$) [9]	$[E_0,\ E,\ \Sigma,\ \theta]$	$\frac{2E_0\Sigma}{(\Sigma+\Sigma_e)^2}$	$2E_0 E(1+\cos\theta)$	$\frac{E(1+\cos\theta)(\Sigma+\Sigma_e)}{2\Sigma}$

$$\Sigma = \sum_h (E_h - p_{z,h})$$

$$\tan\frac{\gamma}{2} = \frac{\Sigma}{T}$$

$$T = \sqrt{(\sum_h p_{x,h})^2 + (\sum_h p_{y,h})^2}$$