

Recent Developments in PDFs: From Optimal Observables to Collider Neutrinos



Juan Rojo, VU Amsterdam & Nikhef

Recent developments in PDFs

- **CT**
 - CT18 (NNLO)
 - * release on LHAPDF August 2019; preprint December 2019
 - * paper published January 2021
 - CT25 (NNLO): partially available on CT website since few months
- **MSHT**
 - MSHT20 (NNLO)
 - * release on LHAPDF & preprint December 2020
 - * paper published April 2021
 - N³LO
 - * release on LHAPDF & preprint July 2022
 - * paper published March 2023
- **NNPDF**
 - NNPDF4.1 (NNLO)
 - * release on LHAPDF September 2021 & preprint September 2021
 - * paper published May 2022
 - N³LO
 - * release on LHAPDF & preprint July 2024
 - * paper published July 2024
 - NNPDF4.1 (N³LO): in preparation, due 2027

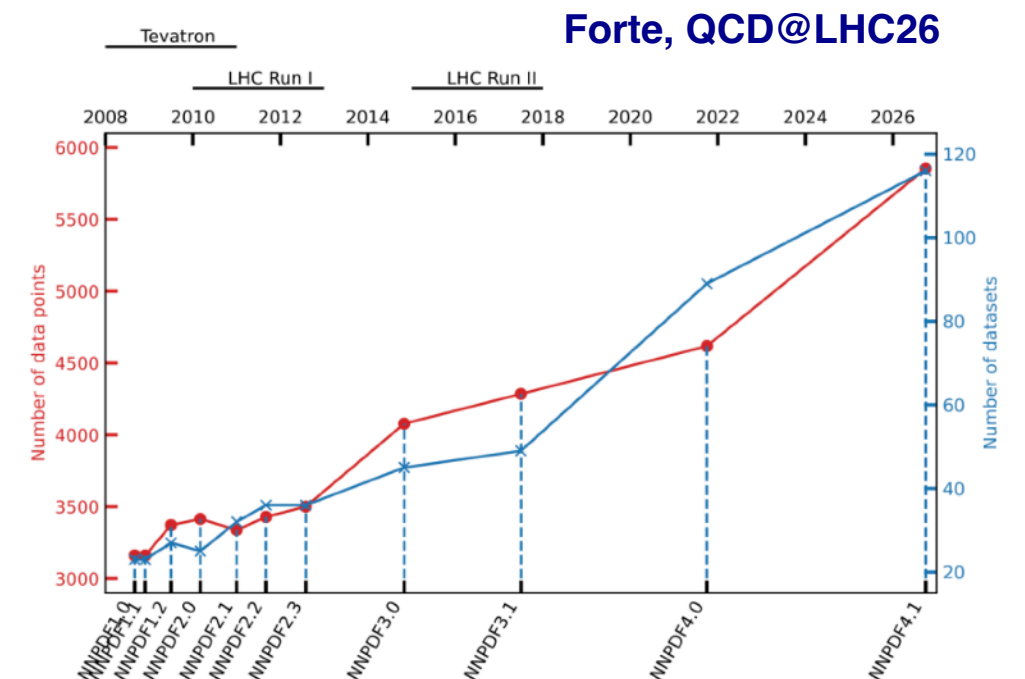
SOME MORE SPECIALIZED PDF SETS

- **ABMP**: simultaneous fit of QCD & EW parameters
- **ATLASPDF**: cross-correlation between ATLAS datasets
- **JAM**: large x , low scale JLAB data, higher twists

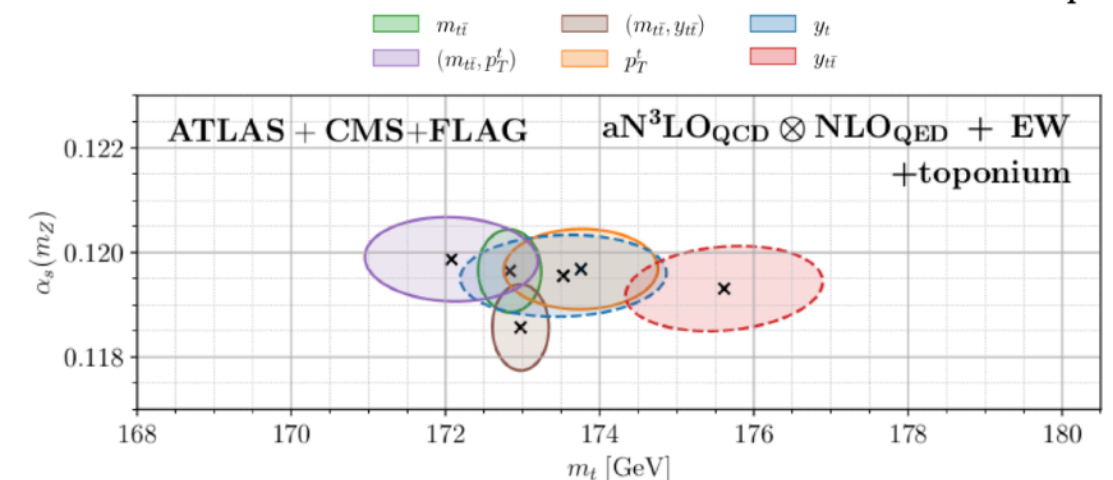
🔊 **Data**: many new LHC datasets available, most with full Run II luminosity, also SeaQuest, ...

🔊 **Theory**: exact NNLO grids, (p)N³LO calculations, MHOU in theory covariance matrix, electroweak/QED corrections, small- p_T resummation, higher twists / jet power corrections, ...

🔊 **Methodology**: joint fits of PDFs with (B)SM parameters, closure tests, novel approaches to PDF fitting (Bayesian inference, physics-informed AI, neural tangent kernel), novel diagnostics metrics ,...

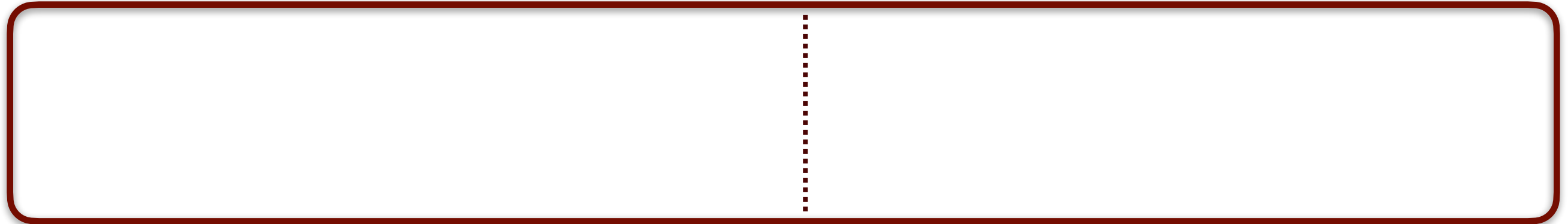


SIMULTANEOUS DETERMINATION OF PDFs, α_s , m_{top}

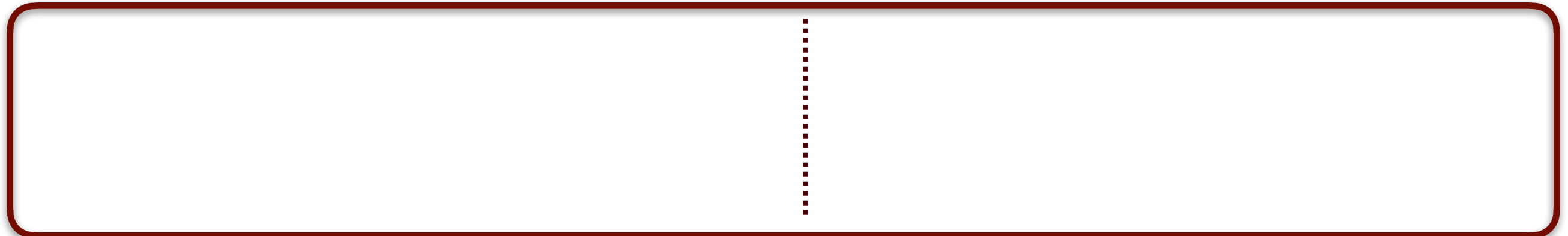


Recent developments in PDFs

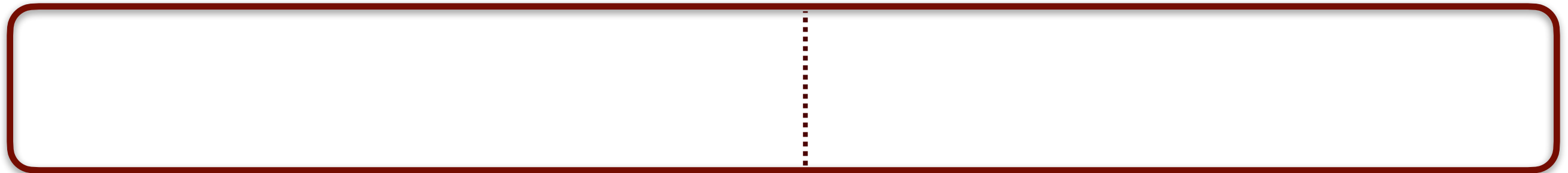
Problem

A large, empty rectangular box with a dark red border and rounded corners, intended for the user to describe the problem.

Question

A large, empty rectangular box with a dark red border and rounded corners, intended for the user to describe the question.

Solution

A large, empty rectangular box with a dark red border and rounded corners, intended for the user to describe the solution.

Recent developments in PDFs

Problem

More data does not necessarily result in **better PDFs** (tensions, inconsistencies, theory and experimental systematics, ...)

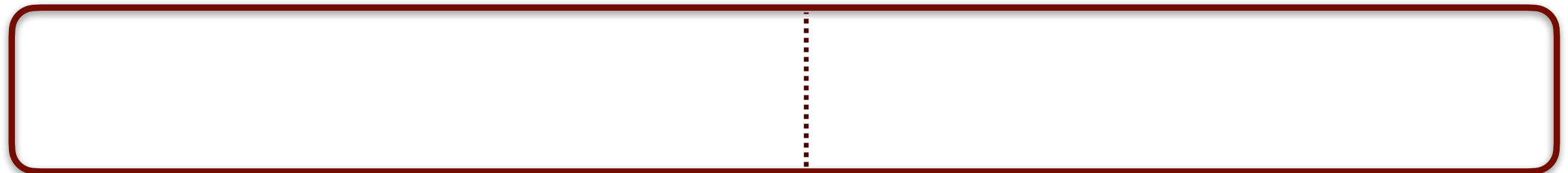


Question

How to bypass restrictions of **traditional, low-dimensional, binned observables** for global PDF determinations?



Solution



Recent developments in PDFs

Problem

More data does not necessarily result in **better PDFs** (tensions, inconsistencies, theory and experimental systematics, ...)



Question

How to bypass restrictions of **traditional, low-dimensional, binned observables** for global PDF determinations?



Solution

Integrate **Neural Simulation Based Inference** in global PDF fitting pipeline

Recent developments in PDFs

Problem

More data does not necessarily result in **better PDFs** (tensions, inconsistencies, theory and experimental systematics, ...)

PDF constraints from **pp collisions** are less direct than from DIS and risk **fitting away BSM signals** in high- p_T tails

Question

How to bypass restrictions of **traditional, low-dimensional, binned observables** for global PDF determinations?

Can we constrain proton and nuclear structure **at the LHC using elementary particle beams**?

Solution

Integrate **Neural Simulation Based Inference** in global PDF fitting pipeline

Recent developments in PDFs

Problem

More data does not necessarily result in **better PDFs** (tensions, inconsistencies, theory and experimental systematics, ...)

PDF constraints from **pp collisions** are less direct than from DIS and risk **fitting away BSM signals** in high- p_T tails

Question

How to bypass restrictions of **traditional, low-dimensional, binned observables** for global PDF determinations?

Can we constrain proton and nuclear structure **at the LHC using elementary particle beams**?

Solution

Integrate **Neural Simulation Based Inference** in global PDF fitting pipeline

Use **TeV-energy neutrinos and muons at the LHC** as precision probes of QCD

Recent developments in PDFs

Problem

More data does not necessarily result in **better PDFs** (tensions, inconsistencies, theory and experimental systematics, ...)

PDF constraints from **pp collisions** are less direct than from DIS and risk **fitting away BSM signals** in high- p_T tails

Question

How to bypass restrictions of **traditional, low-dimensional, binned observables** for global PDF determinations?

Can we constrain proton and nuclear structure **at the LHC using elementary particle beams**?

Solution

Integrate **Neural Simulation Based Inference** in global PDF fitting pipeline

Use **TeV-energy neutrinos and muons at the LHC** as precision probes of QCD

Both topics heavily rely on progress in **high-precision QCD calculations!**

Optimal Observables for PDF fits

*Barrué, Benato, Güven, Hammou, ter Hoeve, Krause, Li, Mantani, Rojo, Sánchez
Cruz, Schöfbeck, Ubiali, Wang, arXiv:2604.13157*

Experimental inputs for global PDF fits

- 📌 Global PDF fits consider exclusively **binned, parton-level, low-multiplicity observables**
- 📌 Released as **binned multi-Gaussian** likelihoods (via HepData)
- 📌 Can be evaluated with **high precision** in QCD (NNLO or N³LO) and EW corrections

$$E^{(k)} = \frac{1}{n_{\text{dat}}} \sum_{i,j=1}^{n_{\text{dat}}} \left(D_i^{(k)} - T_i^{(k)} \right) (\text{cov})_{ij}^{-1} \left(D_j^{(k)} - T_j^{(k)} \right)$$

Experimental inputs for global PDF fits

- 📌 Global PDF fits consider exclusively **binned, parton-level, low-multiplicity observables**
- 📌 Released as **binned multi-Gaussian** likelihoods (via HepData)
- 📌 Can be evaluated with **high precision** in QCD (NNLO or N³LO) and EW corrections

$$E^{(k)} = \frac{1}{n_{\text{dat}}} \sum_{i,j=1}^{n_{\text{dat}}} \left(D_i^{(k)} - T_i^{(k)} \right) (\text{cov})_{ij}^{-1} \left(D_j^{(k)} - T_j^{(k)} \right)$$

binning, unfolding, and final-state **multiplicity reduction** lead to **information loss** and introduce possible (theory) biases.

can we do better for PDF fits, as we move towards the HL-LHC era

Experimental inputs for global PDF fits

- 📌 Global PDF fits consider exclusively **binned, parton-level, low-multiplicity observables**
- 📌 Released as **binned multi-Gaussian** likelihoods (via HepData)
- 📌 Can be evaluated with **high precision** in QCD (NNLO or N³LO) and EW corrections

$$E^{(k)} = \frac{1}{n_{\text{dat}}} \sum_{i,j=1}^{n_{\text{dat}}} \left(D_i^{(k)} - T_i^{(k)} \right) (\text{cov})_{ij}^{-1} \left(D_j^{(k)} - T_j^{(k)} \right)$$

binning, unfolding, and final-state multiplicity reduction lead to **information loss** and introduce possible (theory) biases.

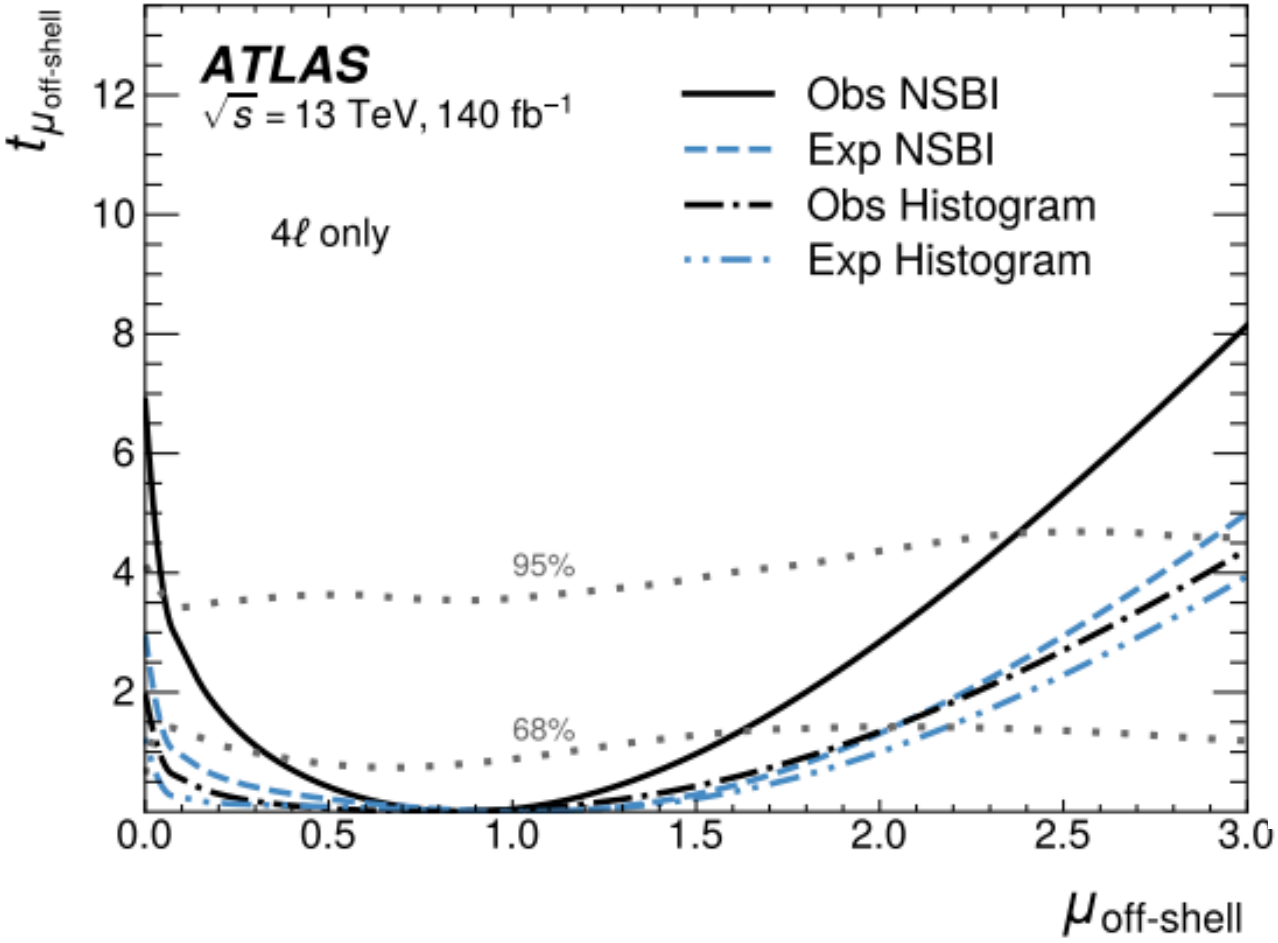
can we do better for PDF fits, as we move towards the HL-LHC era

- 📌 We know how to carry out **unbinned** (event-by-event) measurements composed by a **large number of final-state features**
- 📌 We know how to do parameter inference in large dimensionality spaces with AI techniques
- 📌 Inference technology for unbinned data available, both from pheno groups and within ATLAS/CMS

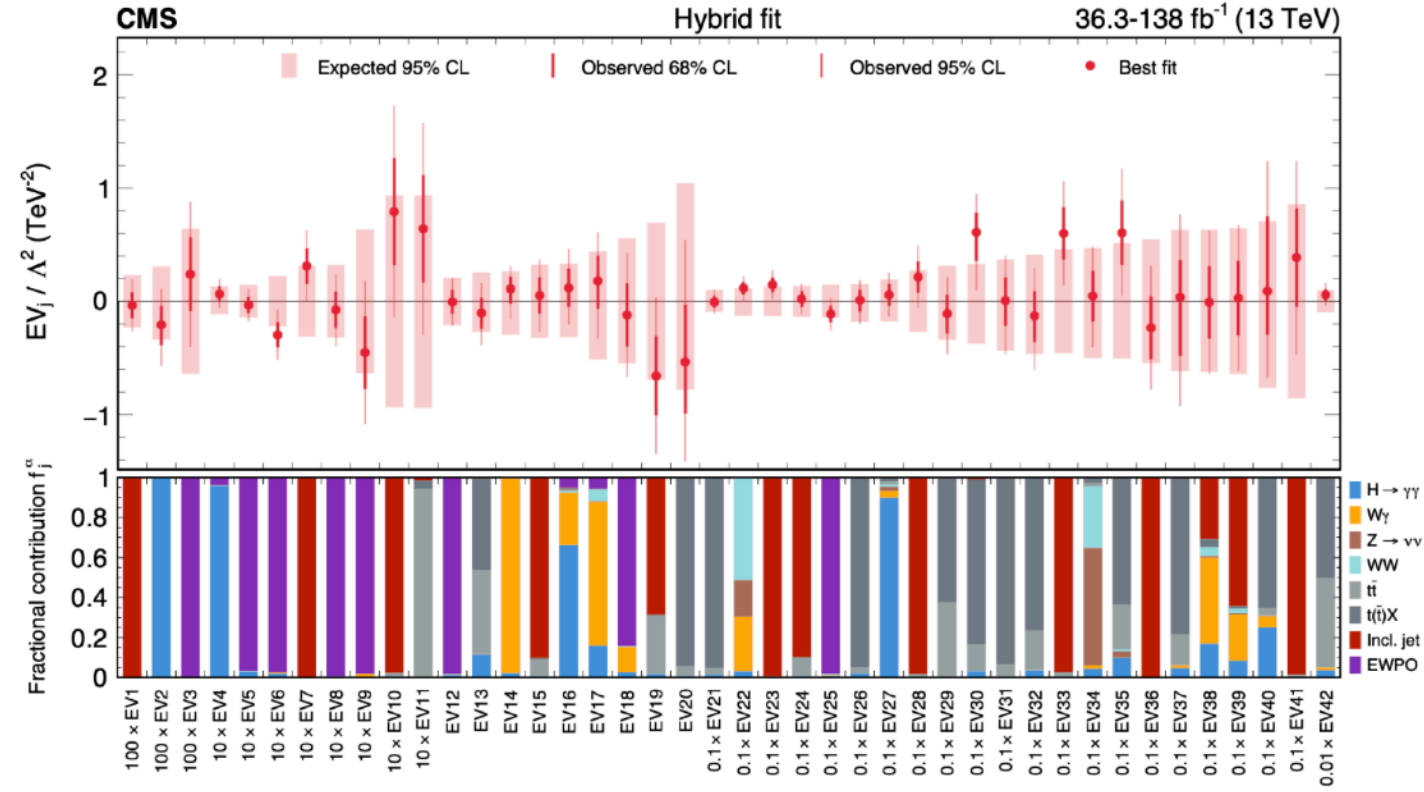
Can we coalesce these ideas to use unbinned data in PDF fits?

Unbinned analyses at the LHC

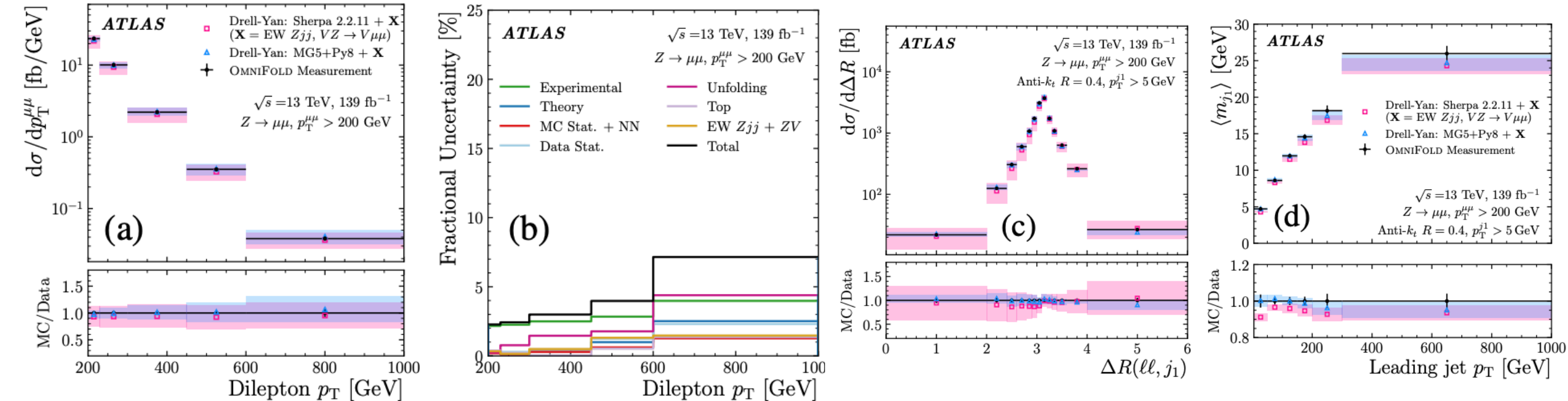
NSBI for Higgs width



Detector-level EFT interpretation



Unbinned Z+jets (24 final-state features)

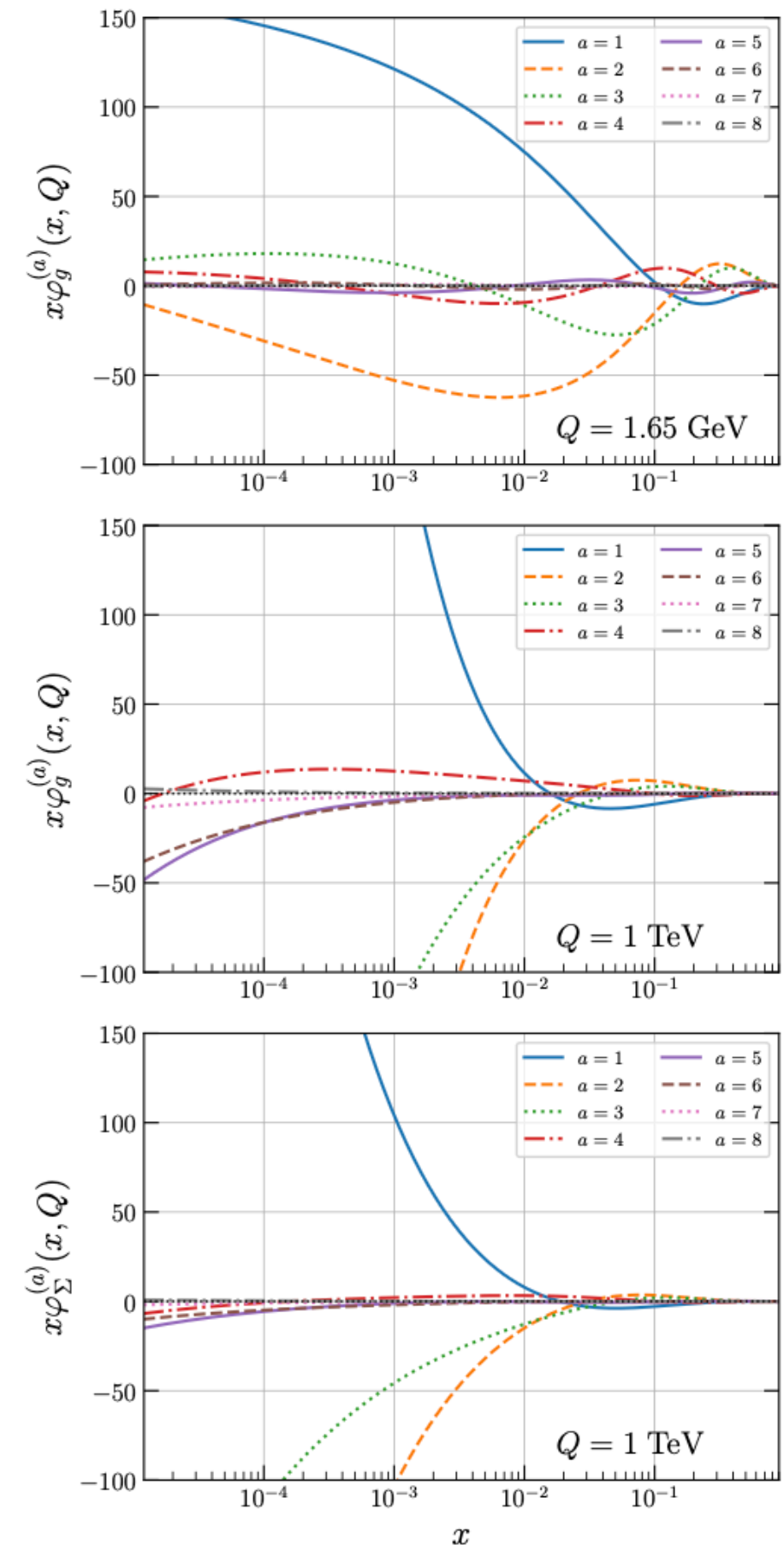
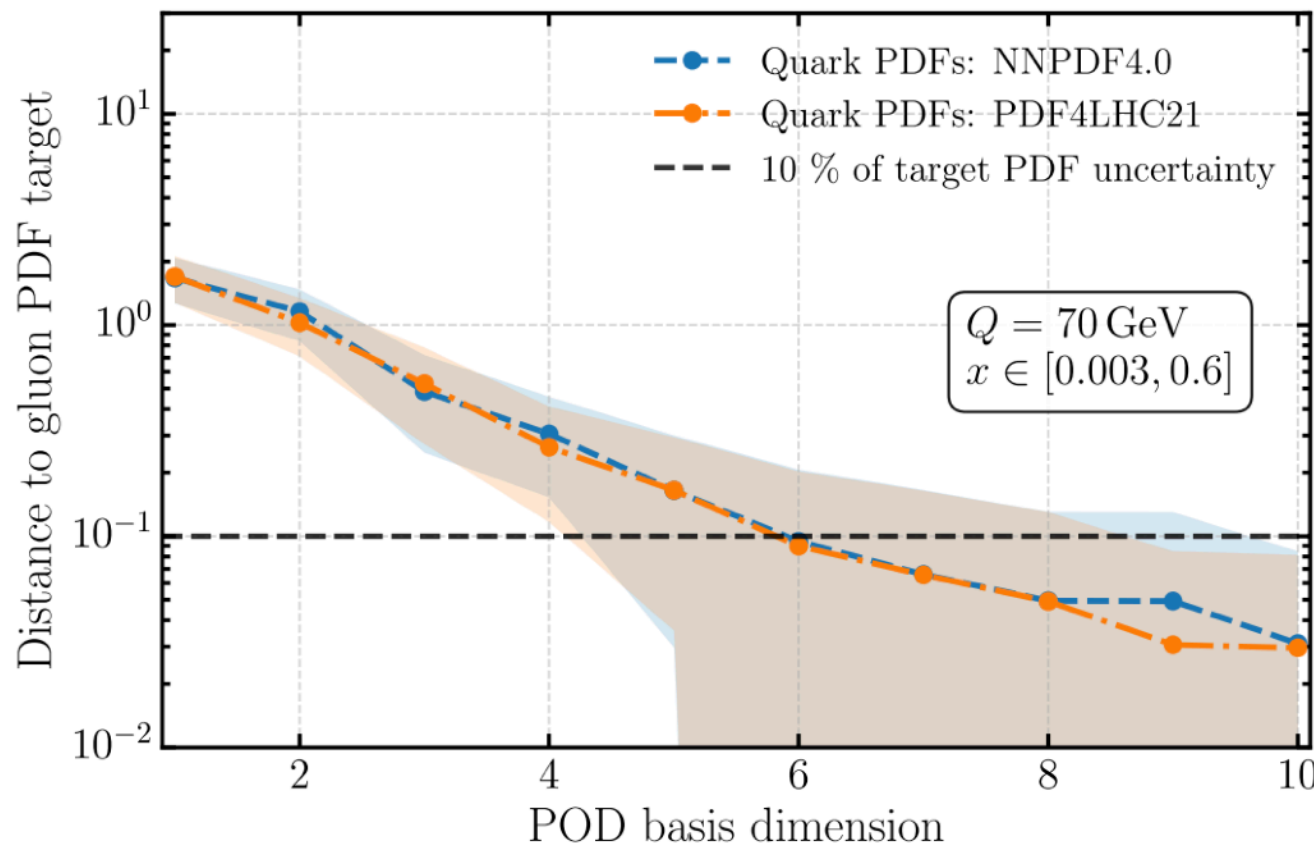


A linear model for the gluon PDF

- Sample the space of **physically admissible gluons**
- Quark PDFs fixed to a reference set (PDF4LHC21)
- Construct Proper Orthogonal Decomposition (SVD on x grid):
linear model for gluon PDF:

$$f_g(x, Q_0, \mathbf{c}) = \underbrace{\varphi_g^{(0)}(x, Q_0)}_{\text{mean gluon}} + \sum_{a=1}^N \underbrace{c_a}_{\text{Pols to be fitted}} \underbrace{\varphi_g^{(a)}(x, Q_0)}_{\text{Basis functions}}$$

- Theory constraints (DGLAP, sum rules) satisfied



NSBI for PDF determinations

🔧 **Extended likelihood** for an unbinned dataset of N_{obs} events with detector-level features $\mathbf{x}$, Poles $\mathbf{c}$ (parametrising PDF dependence) and nuisance parameters $\boldsymbol{\nu}$:

$$L(\mathcal{D}|\mathbf{c}, \boldsymbol{\nu}) = \underset{\substack{\nearrow \\ \text{Total yields}}}{\text{Pois}(N_{\text{obs}} \mid \mathcal{L}(\boldsymbol{\nu}) \sigma(\mathbf{c}, \boldsymbol{\nu}))} \underset{\substack{\downarrow \\ \text{Nuisances}}}{\mathcal{N}(\boldsymbol{\nu}|\mathbf{0}, \mathbf{1})} \prod_{j=1}^{N_{\text{obs}}} \underset{\substack{\nearrow \\ \text{Event density}}}{p(\mathbf{x}_j|\mathbf{c}, \boldsymbol{\nu})}$$

🔧 **Likelihood ratio**: factorises PDF parameters from exp. & th. nuisances

$$\frac{d\sigma(\mathbf{x}|\mathbf{c}, \boldsymbol{\nu})}{d\sigma(\mathbf{x}|\mathbf{0}, \mathbf{0})} \simeq S(\mathbf{x}, \boldsymbol{\nu}) R(\mathbf{x}, \mathbf{c})$$

🔧 **ML surrogates** with exact analytical structure

$$\hat{R}(\mathbf{x}, \mathbf{c}) = 1 + \sum_a c_a \hat{R}_a(\mathbf{x}) + \sum_{a,b \leq a} c_a c_b \hat{R}_{ab}(\mathbf{x}), \quad \hat{S}(\mathbf{x}, \boldsymbol{\nu}) = \exp\left(\nu_A \hat{\Delta}_A(\mathbf{x})\right)$$

Boosted Information Trees

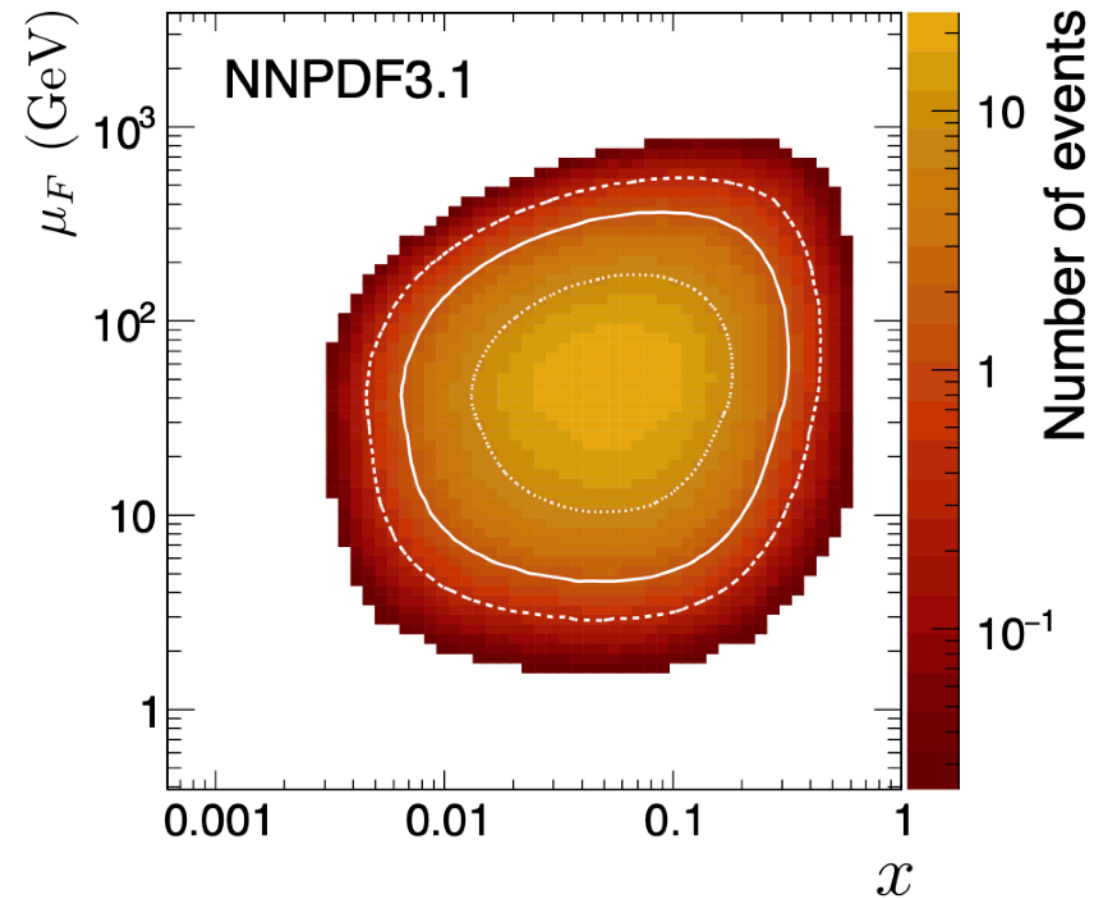
parametric neural network

🔧 **PDF inference** from log-likelihood ratio on Asimov

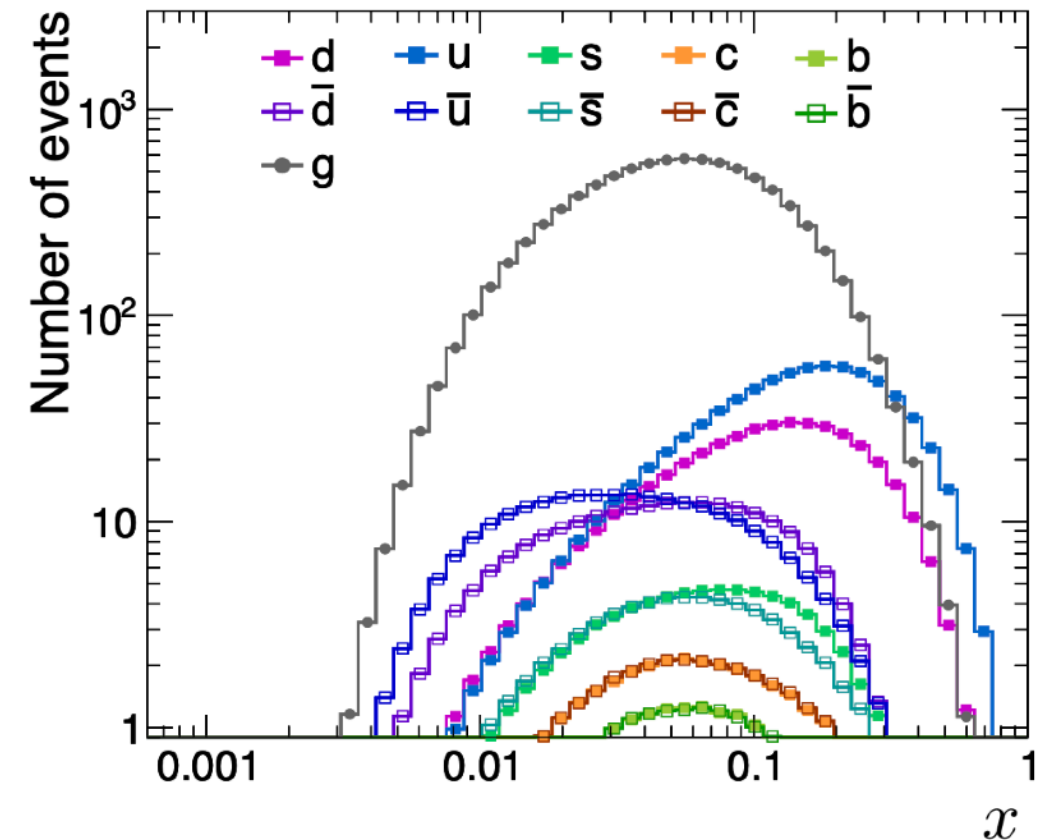
$$q_c(\mathcal{D}) = -2 \left[\max_{\boldsymbol{\nu}} \log \frac{L(\mathcal{D}|\mathbf{c}, \boldsymbol{\nu})}{L(\mathcal{D}|\mathbf{0}, \mathbf{0})} - \max_{\mathbf{c}', \boldsymbol{\nu}'} \log \frac{L(\mathcal{D}|\mathbf{c}', \boldsymbol{\nu}')}{L(\mathcal{D}|\mathbf{0}, \mathbf{0})} \right]$$

Unbinned top-quark pair production at 13 TeV

- Simulation:** POWHEG-v2 NLO + PYTHIA 8 (CP5) + Delphes CMS model, normalised to NNLO+NNLL
- Dilepton channel** with $L = 137 \text{ fb}^{-1}$ (full Run 2): 2 isolated leptons, Z-mass veto, > 2 jets, ≥ 2 b-tags, top kinematics
- Partonic coverage:** $10^{-2} \leq x \leq 0.5$, gg dominant
- 16 event-level features**, both top-level and lepton-level: no double counting since event probabilities are evaluated from simulation (impossible with binned data!)
- Reference **binned measurement:** 16 bins in $(m_{t\bar{t}}, y_{t\bar{t}})$

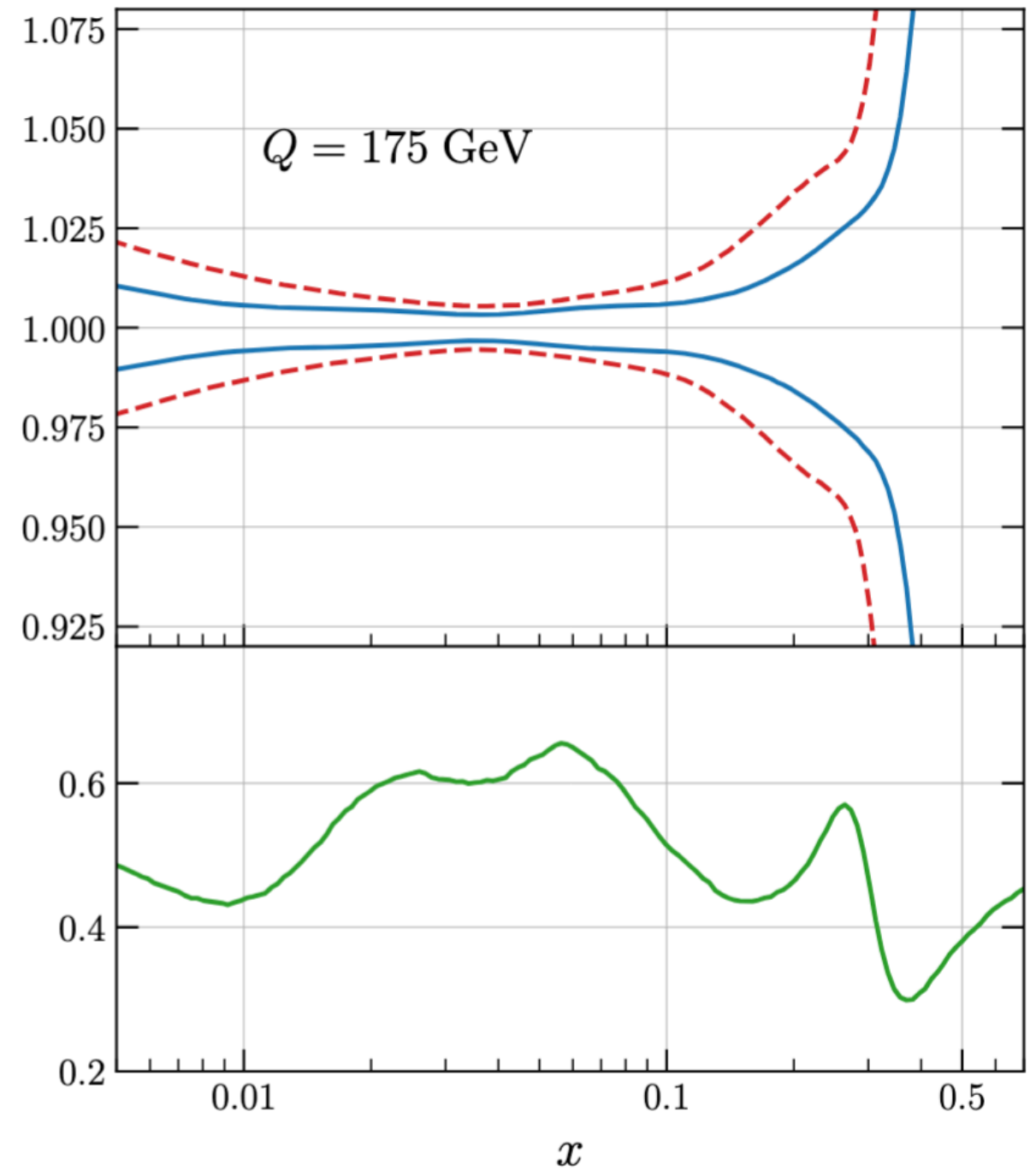
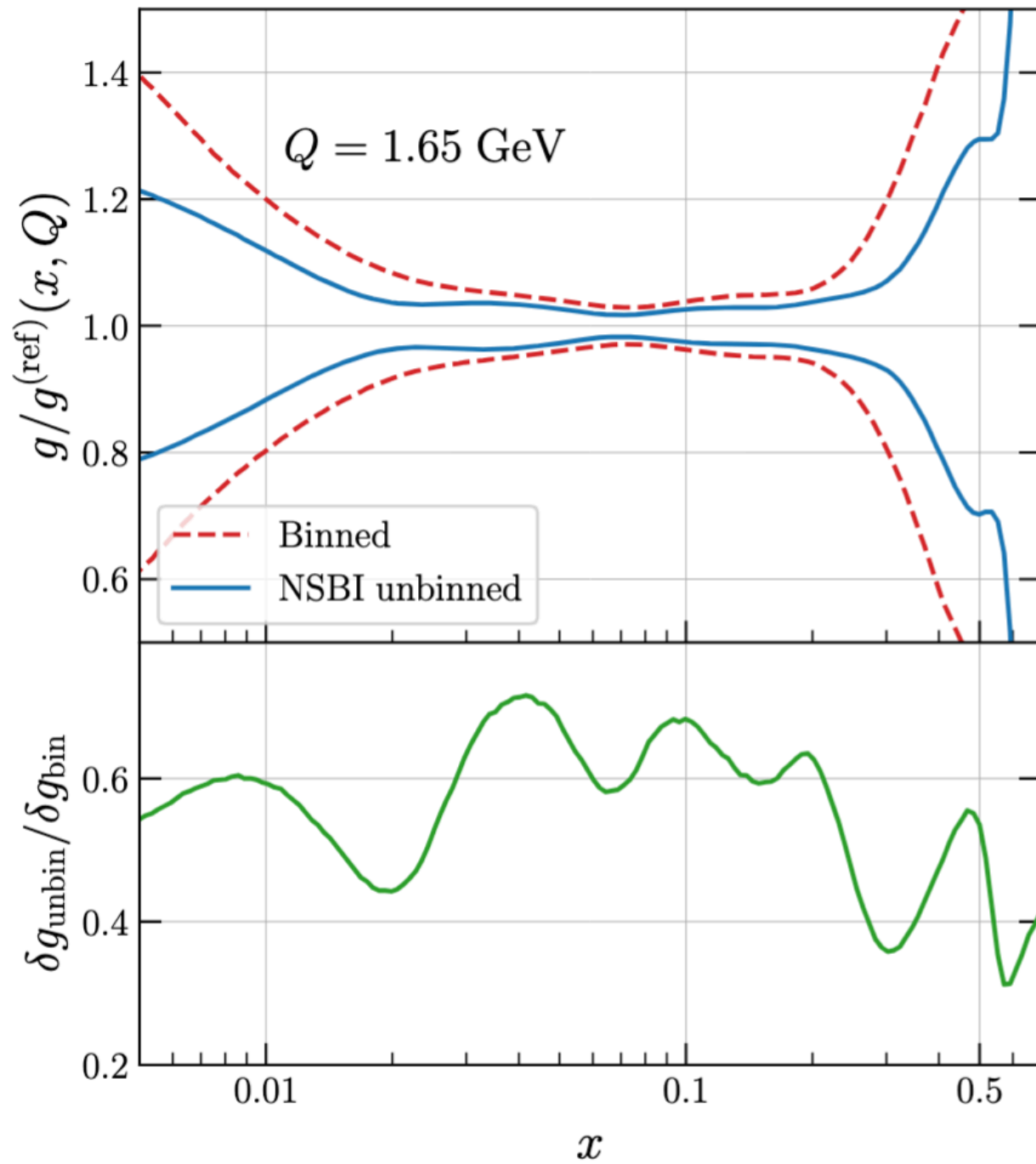


Observable	Description	Observable	Description
$m(t\bar{t})$	invariant mass of the top quark pair	$p_T(t\bar{t})$	p_T of the top quark pair
$y(t\bar{t})$	rapidity of the top quark pair	$\Delta\eta(t\bar{t})$	difference of η of the $t\bar{t}$ system
$\Delta \eta (t\bar{t})$	absolute rapidity difference of the $t\bar{t}$ system	$p_T(t)$	p_T of the top quark
$p_T(\bar{t})$	p_T of the anti-top quark	$y(t)$	rapidity of the top quark
$y(\bar{t})$	rapidity of the anti-top quark	$p_T(\ell_0)$	p_T of the leading lepton
$p_T(\ell_1)$	p_T of the subleading lepton	$p_T(\ell\bar{\ell})$	p_T of the $\ell\bar{\ell}$ system
$m(\ell\bar{\ell})$	invariant mass of the dilepton system	$\eta(\ell\bar{\ell})$	pseudorapidity of the $\ell\bar{\ell}$ system
$\Delta\eta(\ell\bar{\ell})$	difference of η of the $\ell\bar{\ell}$ system	$\Delta \eta (\ell\bar{\ell})$	difference of $ \eta $ of the $\ell\bar{\ell}$ system



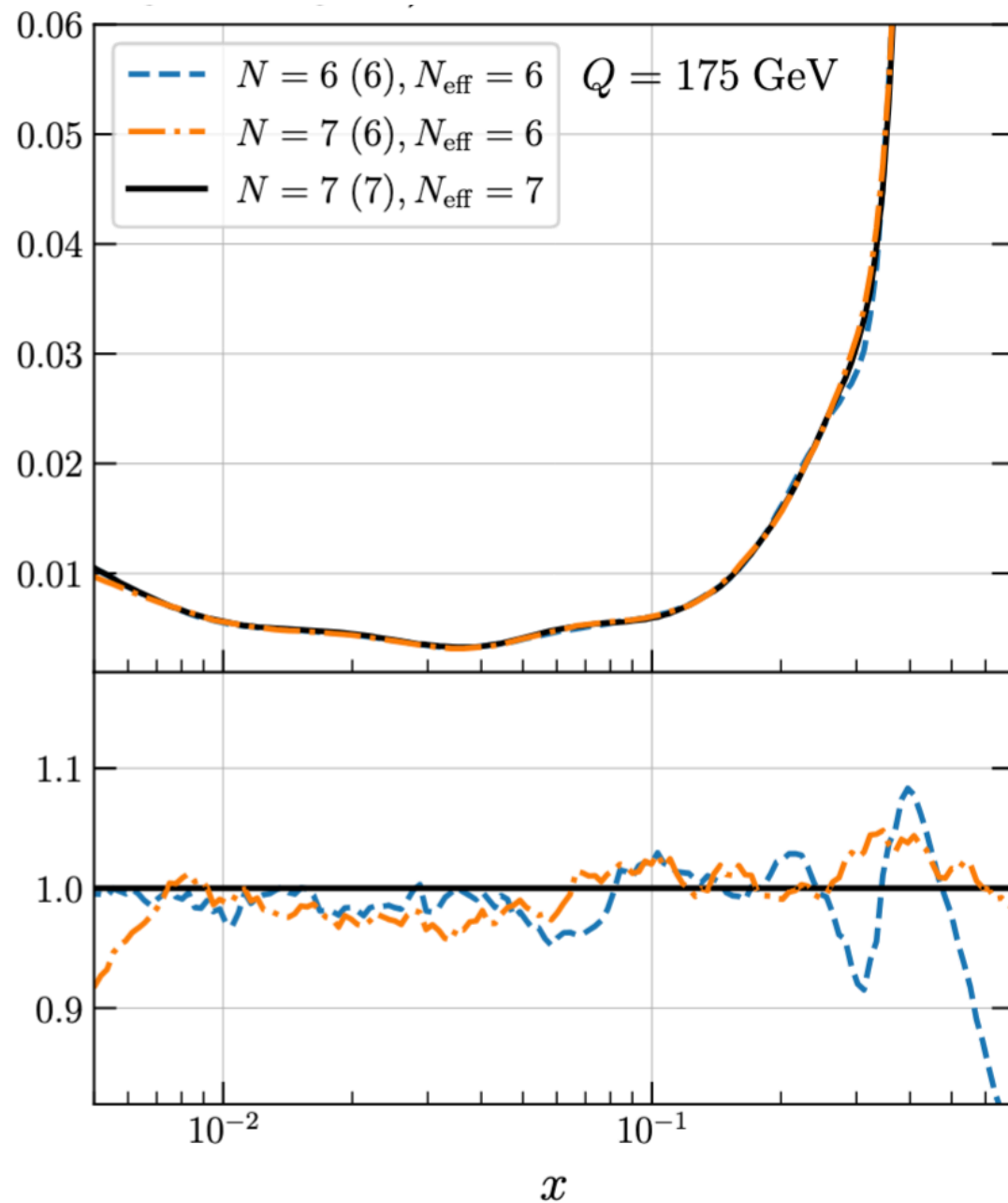
Results

Gluon PDF from $t\bar{t}$ ($\sqrt{s} = 13$ TeV, $\mathcal{L} = 137$ fb $^{-1}$, $N = 7$)

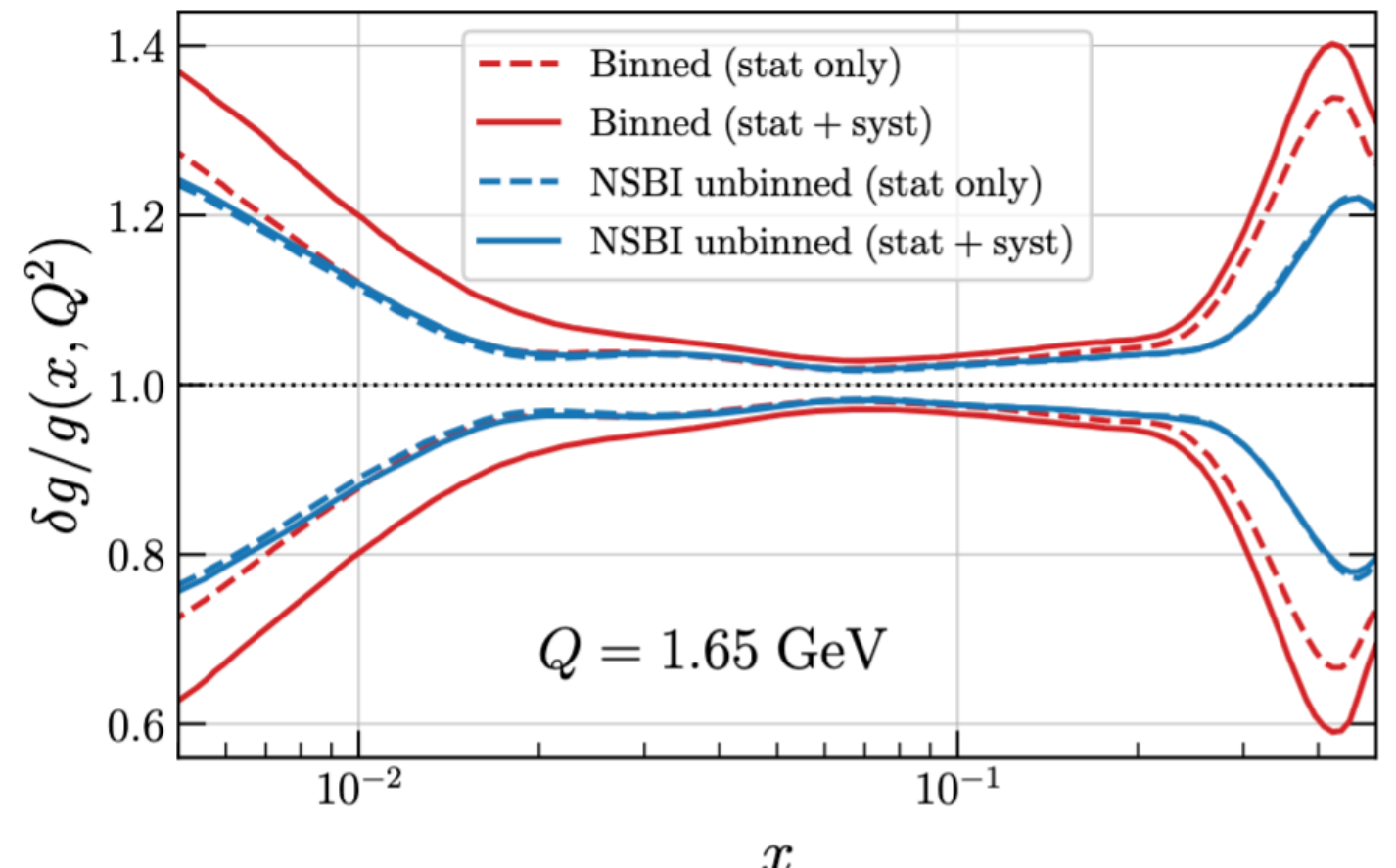


Factor 2 improvement in precision thanks to NSBI

Results

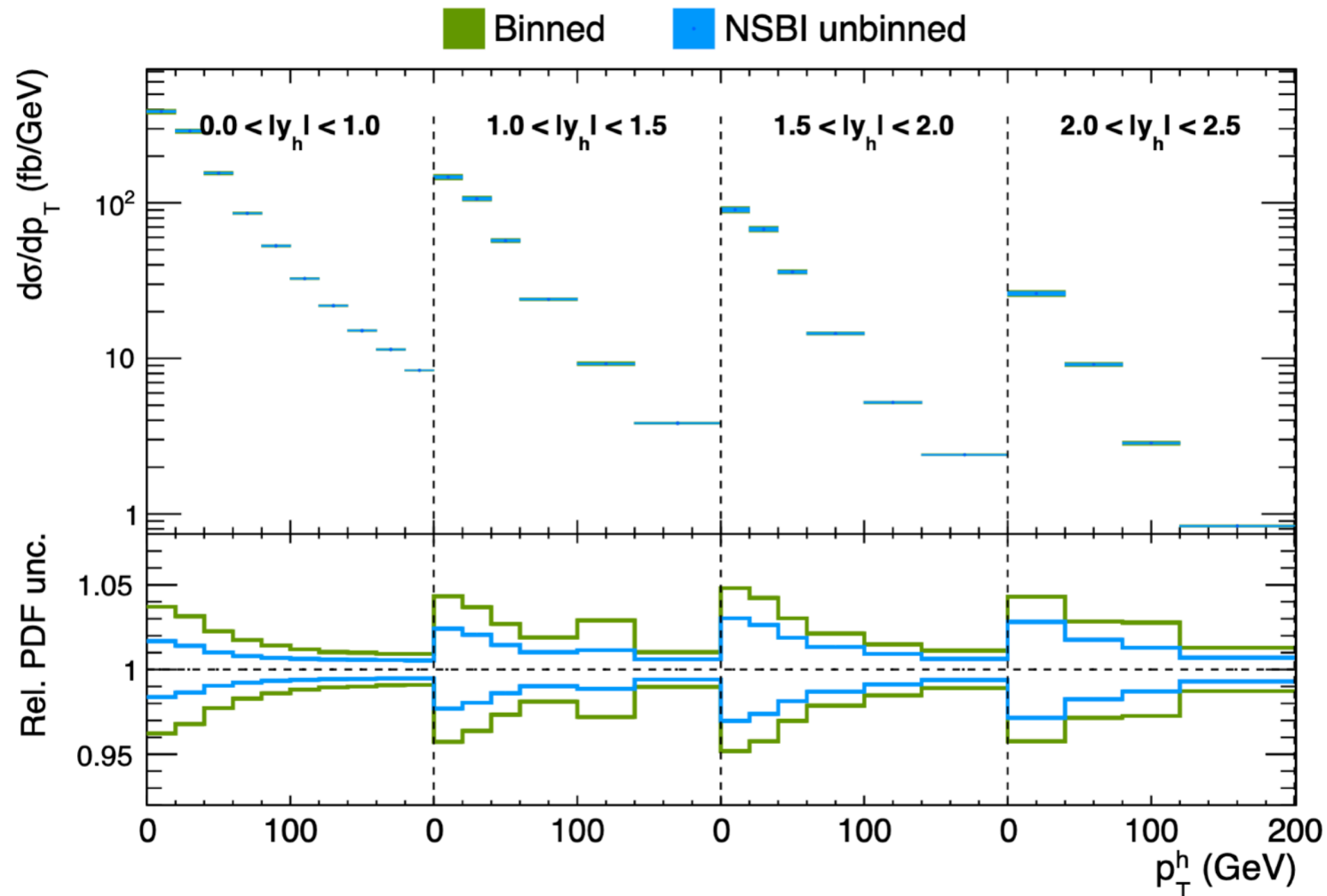


- 📌 **Stability** upon dimensionality of linear model (determined by the data via Fisher information)
- 📌 Reduced impact of **systematic uncertainty modelling** (correlations naturally built in)
- 📌 NSBI unbinned **computationally much more intensive** due to training of ML surrogates



Implications for Higgs Production

- 📌 Propagate the NSBI gluon (binned and unbinned) to **differential Higgs production in ggF**
- 📌 LHC experiments can carry out an **internal calibration of proton structure** from their own unbinned data, reducing reliance on external data sets or theory assumptions
- 📌 Systematic uncertainties **consistently propagated** from calibration measurement to Higgs predictions



NSBI for PDFs: towards a new paradigm

Done

First NSBI determination of proton structure from high-dimensional unbinned data, demonstrated on Asimov $t\bar{t}$ events

Experimental and theoretical **systematics learned event by event** and profiled in the likelihood

Major sensitivity gains over the binned analysis Enables an **internal calibration of proton structure** by the LHC experiments

NSBI for PDFs: towards a new paradigm

Done

First NSBI determination of proton structure from high-dimensional unbinned data, demonstrated on Asimov $t\bar{t}$ events

Experimental and theoretical **systematics learned event by event** and profiled in the likelihood

Major sensitivity gains over the binned analysis Enables an **internal calibration of proton structure** by the LHC experiments

Next

Release the likelihood (surrogates, nuisance model) so that PDF fitters outside the experiments can use it

Integrate **unbinned alongside binned** measurements in the global fits: likelihood interface, fast-grid surrogates

Extend to **quark PDFs and other processes**: Drell-Yan, jets, W +charm; NNLO+PS to reduce MHOU

NSBI for PDFs: towards a new paradigm

Done

First NSBI determination of proton structure from high-dimensional unbinned data, demonstrated on Asimov $t\bar{t}$ events

Experimental and theoretical **systematics learned event by event** and profiled in the likelihood

Major sensitivity gains over the binned analysis Enables an **internal calibration of proton structure** by the LHC experiments

Next

Release the likelihood (surrogates, nuisance model) so that PDF fitters outside the experiments can use it

Integrate **unbinned alongside binned** measurements in the global fits: likelihood interface, fast-grid surrogates

Extend to **quark PDFs and other processes**: Drell-Yan, jets, W +charm; NNLO+PS to reduce MHOU

NNLO event generation & high-precision shower algorithms are essential to realise this new paradigm of unbinned, detector-level, AI-assisted measurements for PDF determinations

Proton Structure from LHC Neutrinos & Muons

[1] J. M. Cruz-Martinez, JR. et al The LHC as a Neutrino-Ion Collider, *Eur. Phys. J. C* 84 (2024) 369 [arXiv:2309.09581]

[2] R. Francener, V. P. Goncalves, F. Kling, P. Krack, JR, Deep-inelastic scattering at TeV energies with LHC muons, *Eur. Phys. J. C* 85 (2025) 1098 [arXiv:2506.13889]

[3] J. Adhikary, JR, et al., Scientific program for the Forward Physics Facility, *Eur. Phys. J. C* 85 (2025) 430 [arXiv:2411.04175]

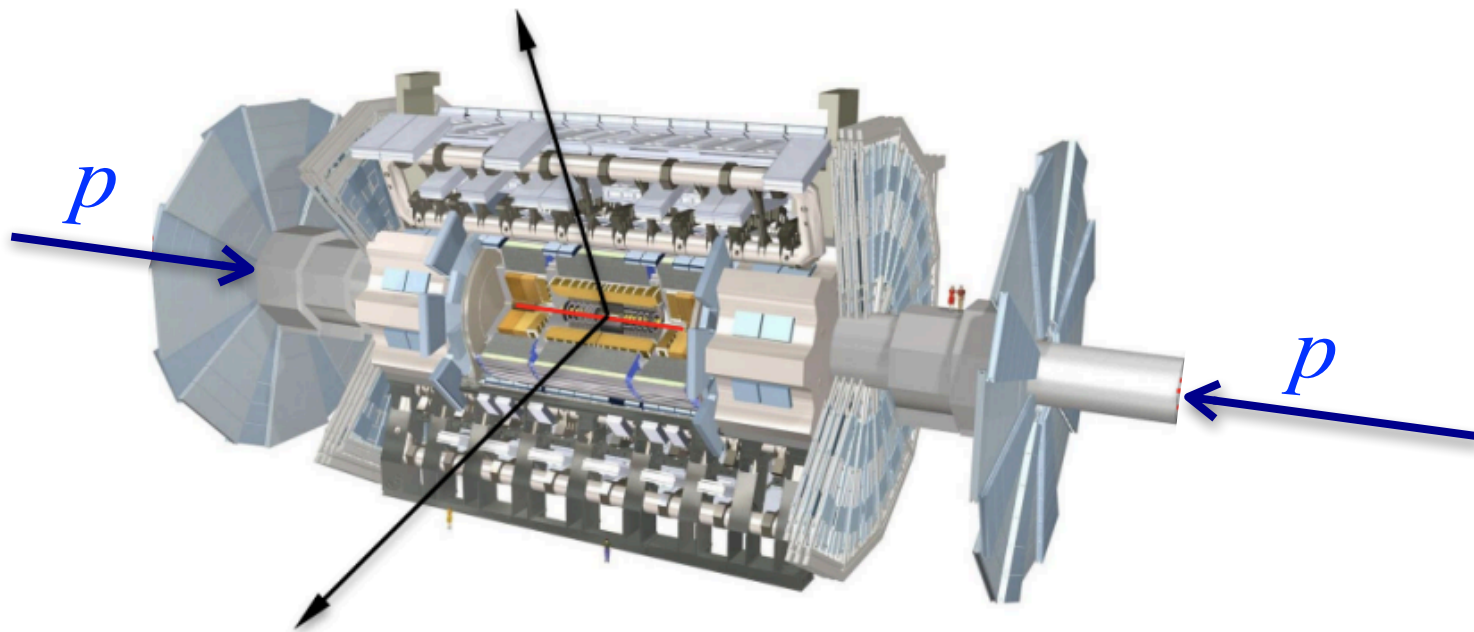
[4] J. L. Feng, JR et al., The Forward Physics Facility at the High-Luminosity LHC, *J. Phys. G* 50 (2023) 030501 [arXiv:2203.05090]

[5] P. Hernandez-Sainz, F. Kling, J. Koorn, JR, Event Generation for Neutrino and Muon DIS at the LHC, in preparation

Proton Imaging with Neutrinos

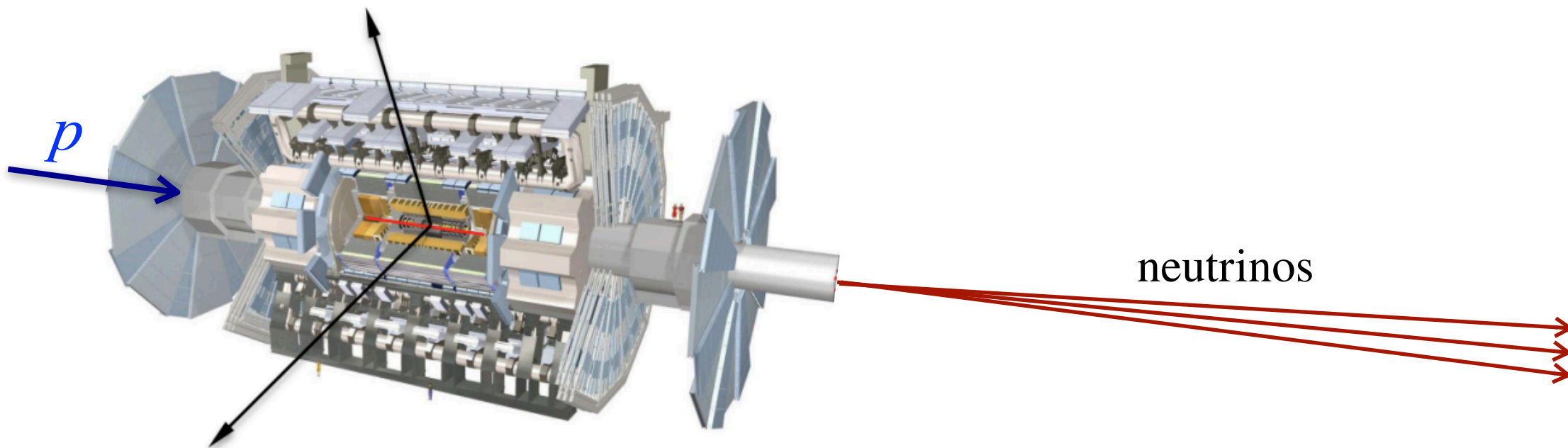
ATLAS @ LHC

Higgs boson ($m_h \sim 125 m_p$)



Proton Imaging with Neutrinos

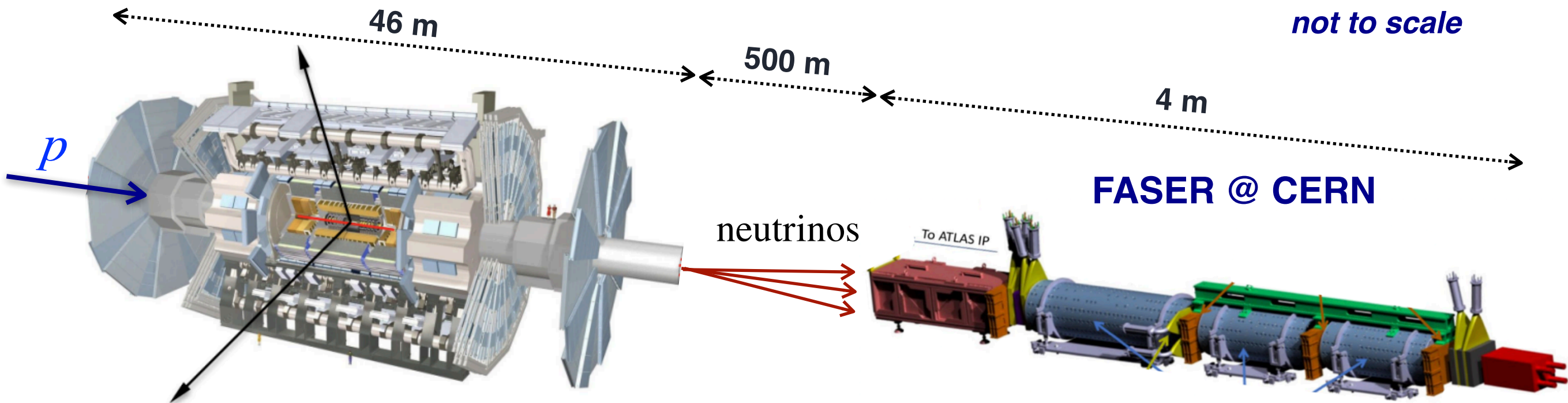
ATLAS @ LHC



- 📌 **Large fluxes of most energetic neutrinos** produced in lab ($\sigma_\nu \propto E_\nu$)
- 📌 For 15 years: forward region **not instrumented** & neutrinos **lost for science**

Proton Imaging with Neutrinos

ATLAS @ LHC



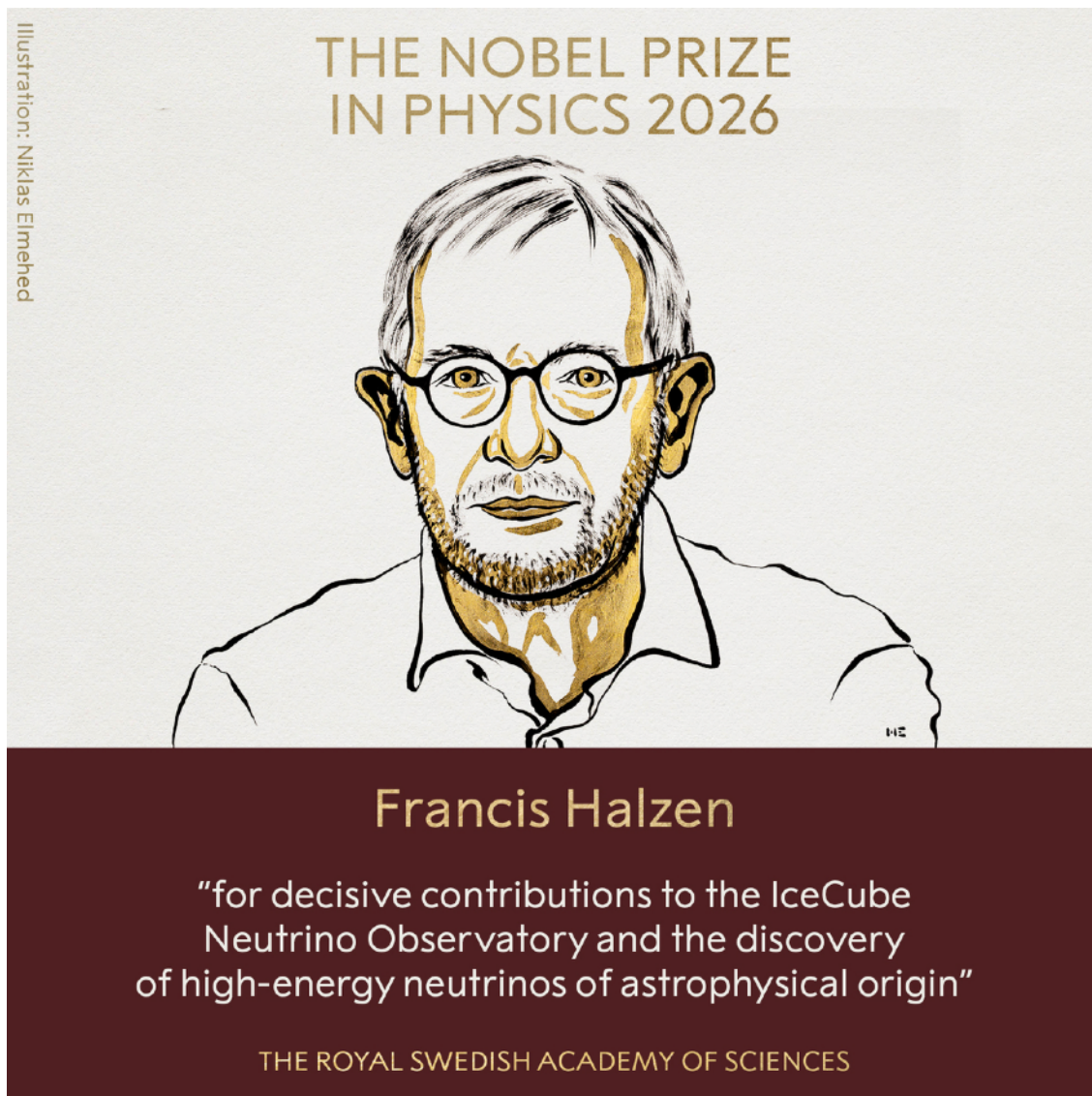
FASER @ CERN

- 📌 **Large fluxes of most energetic neutrinos** produced in lab ($\sigma_\nu \propto E_\nu$)
- 📌 For 15 years: forward region **not instrumented** & neutrinos **lost for science**

FASER & SND@LHC (2023):

first observation of LHC neutrinos!

Proton Imaging with Neutrinos



in order for IceCube to:

- make precise measurements of the flux of cosmic neutrinos,
- measure neutrino oscillations at PeV energy over cosmic distances,

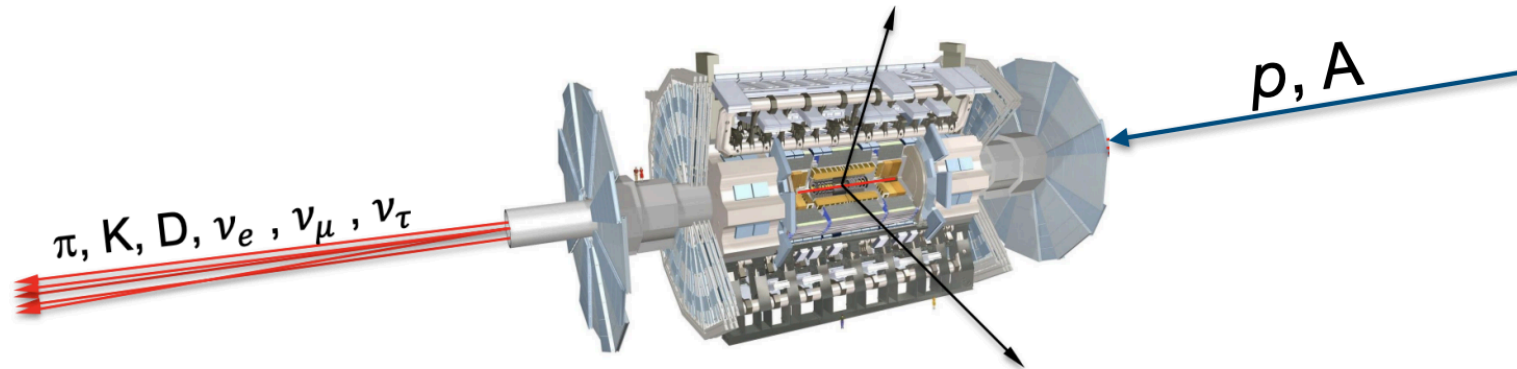
we need accelerator measurements of the high energy and large rapidity charm production cross sections,

*F. Halzen, keynote talk at 1st **Forward Physics Facility** workshop*

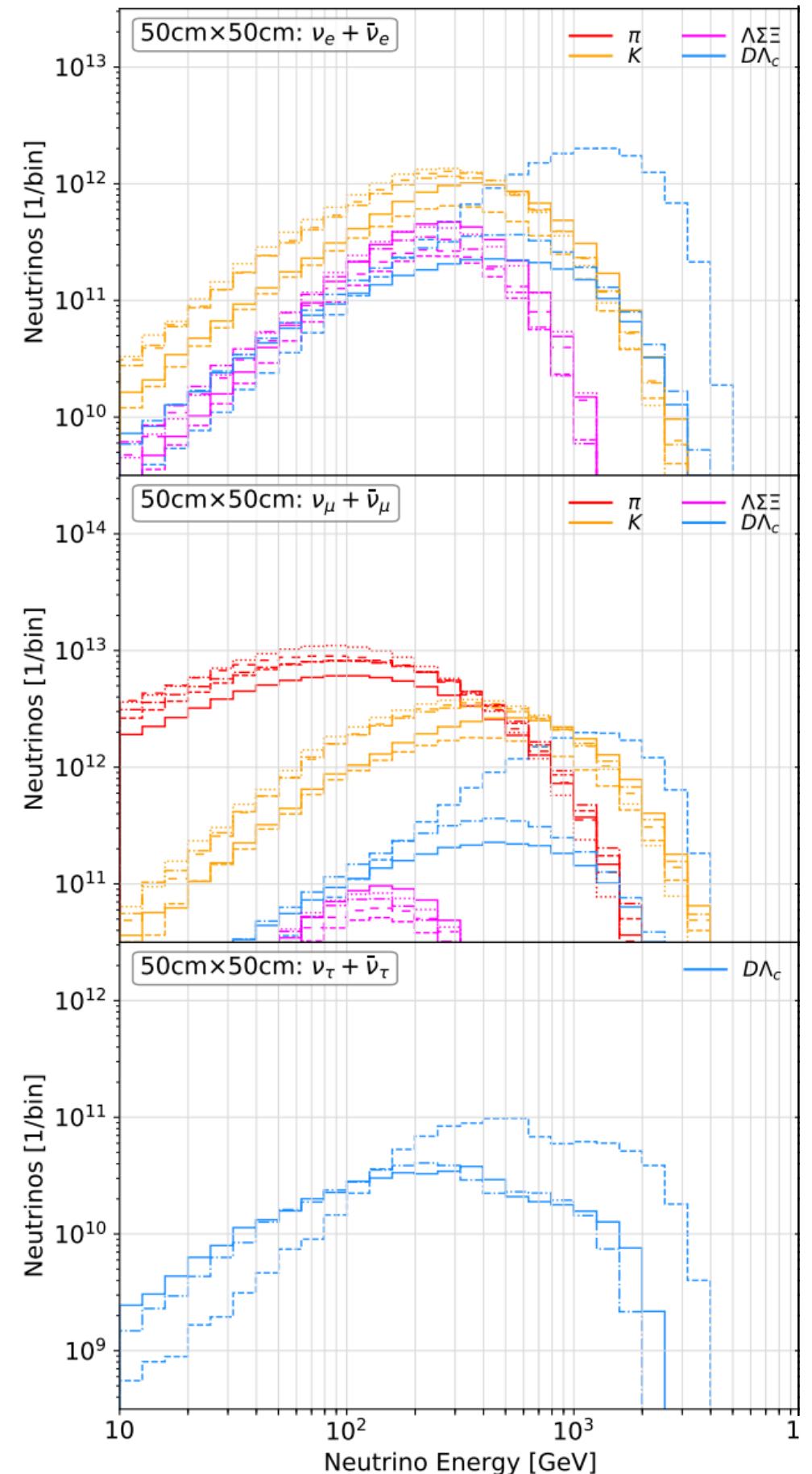
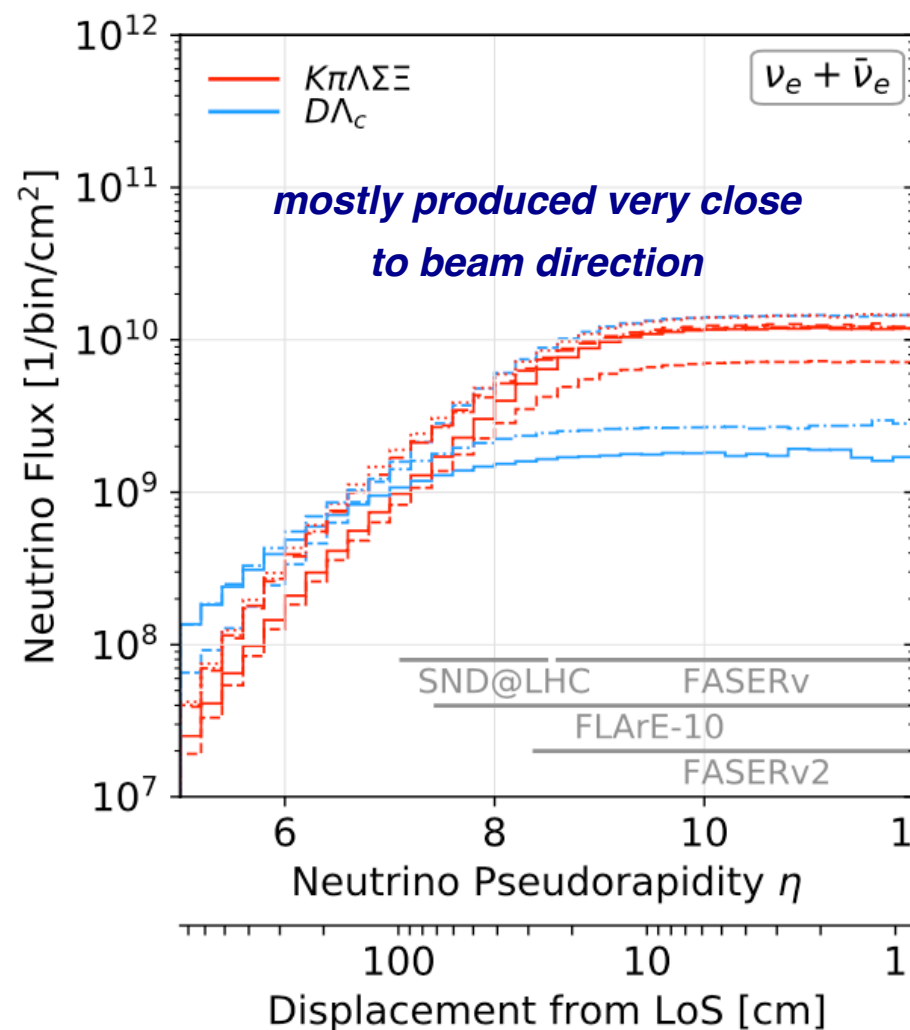
cosmic ray + air $\rightarrow$ charm particle $\rightarrow$ neutrino + muon + ...

"at colliders all cosmic ray physics is in the beam pipe" Tom Gaisser

Neutrinos at the LHC

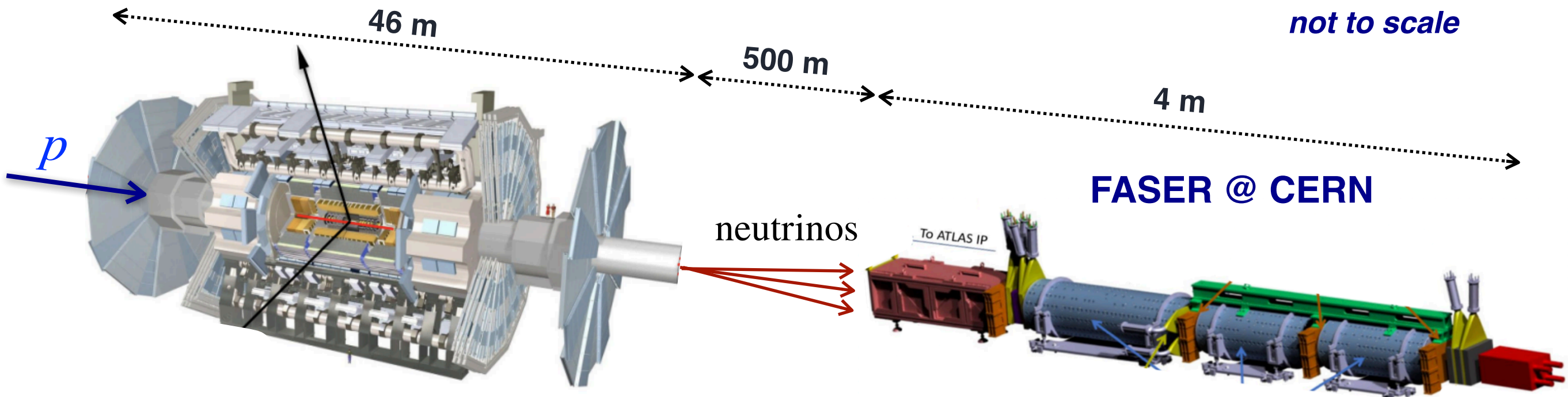


- **electron neutrinos** mostly from D -meson decays above 500 GeV, below it mostly from kaon decays
- **muon neutrino** flux dominated by pion & kaon decays
- **tau neutrinos** entirely from D -meson decays

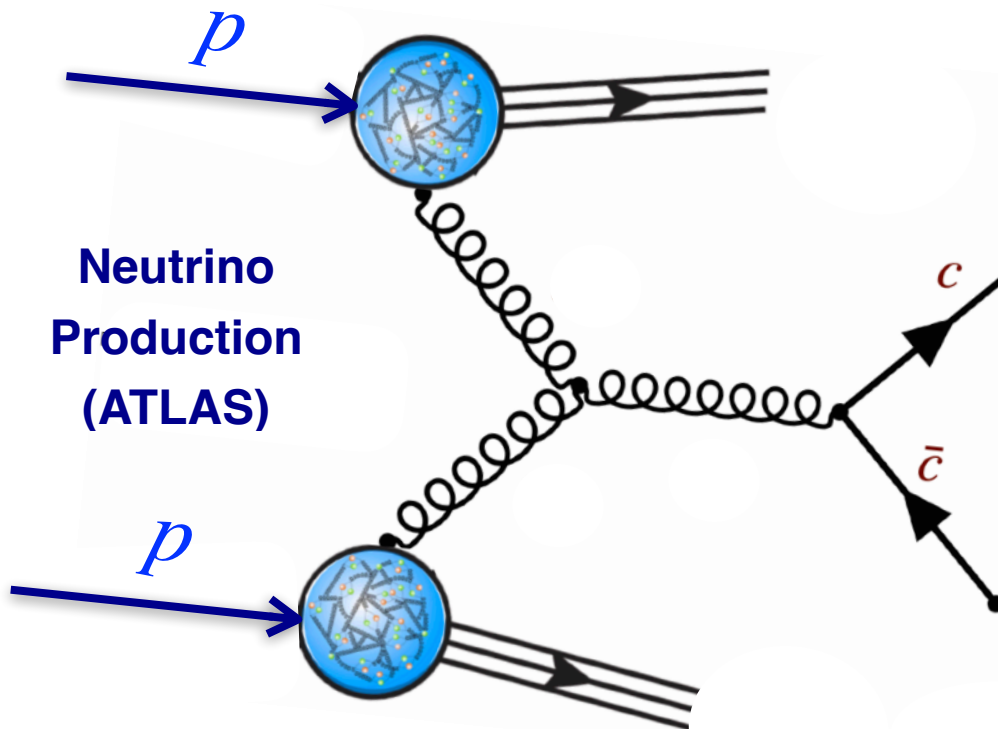


Proton Imaging with Neutrinos

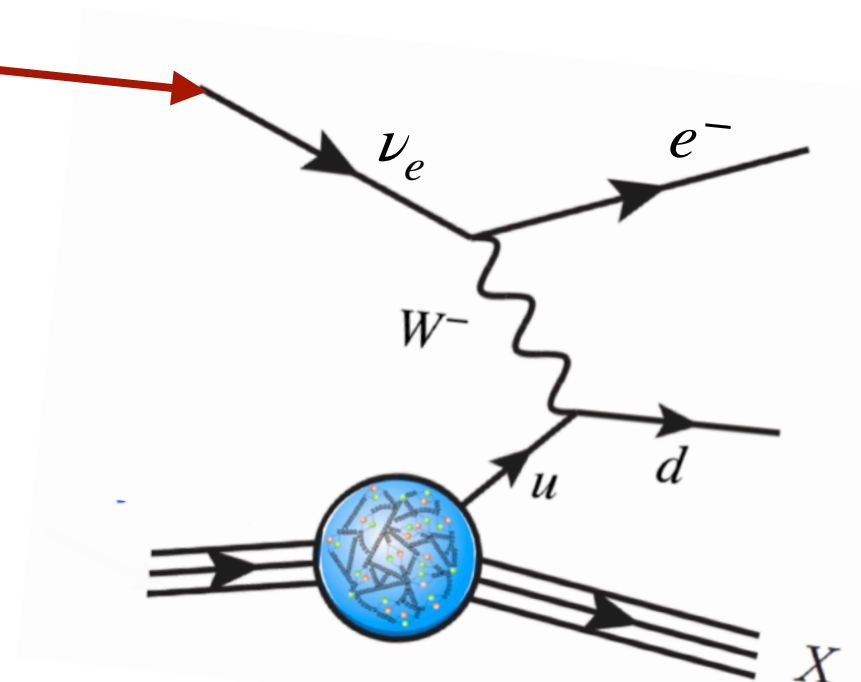
ATLAS @ LHC



Neutrino
Production
(ATLAS)



Neutrino Scattering (FASER)

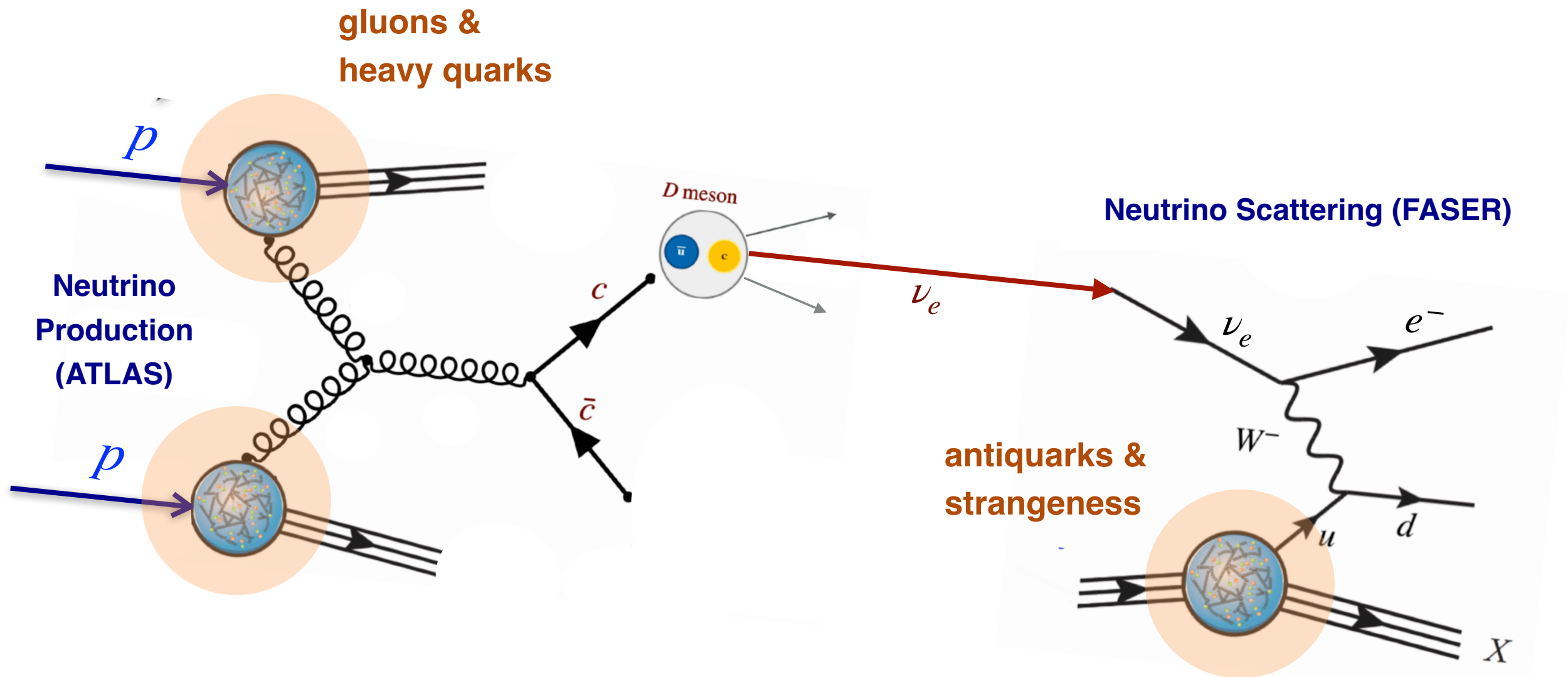


QCD & Proton Structure at FASER

Our goal: develop a **comprehensive theoretical infrastructure**
modelling neutrino production & scattering @ LHC

→ **transform FASER from a search experiment into
a precision microscope for proton structure**

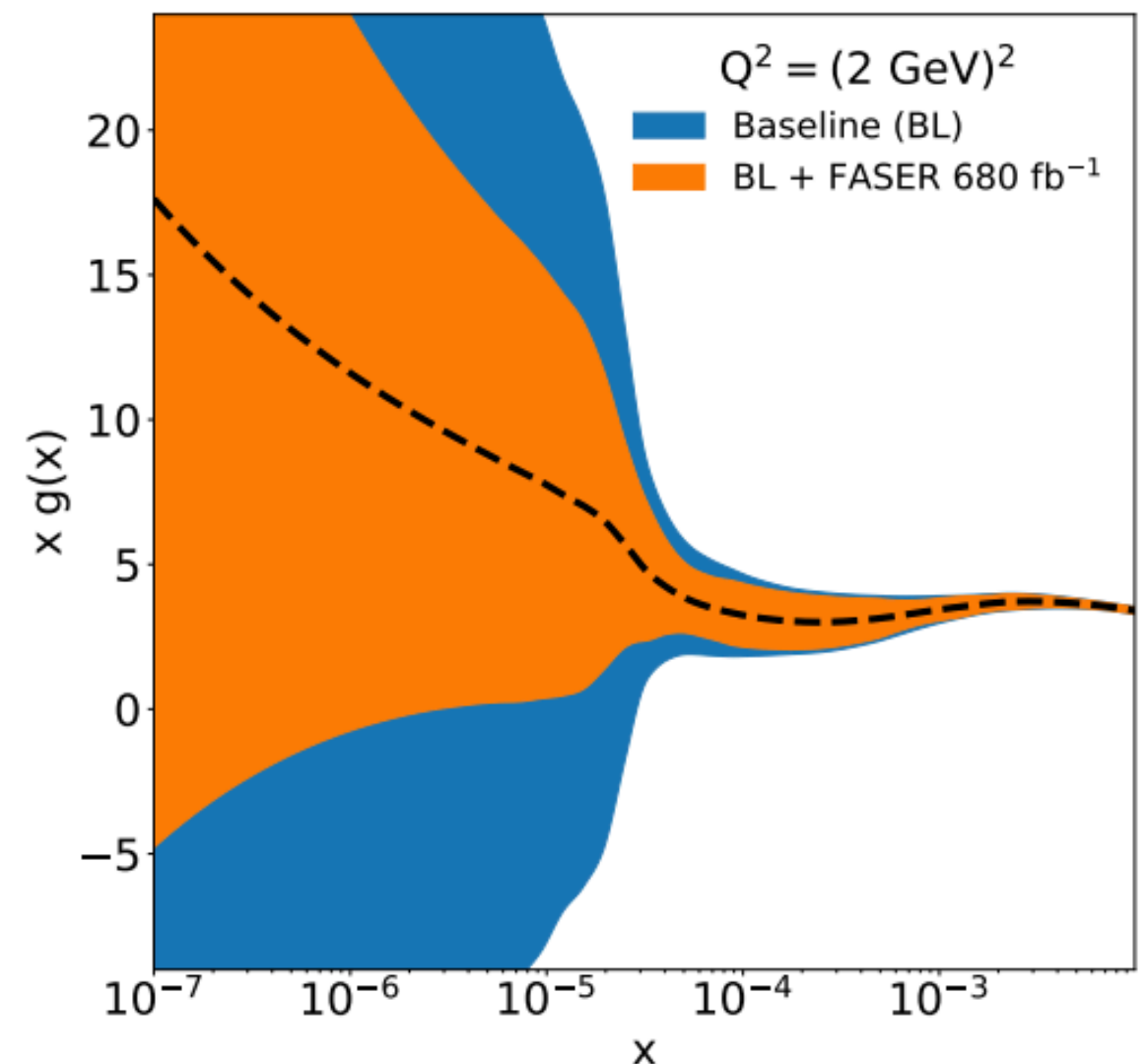
enabling the next **LHC discoveries**



Small-x QCD at FASER

- 📍 Determine **gluon content of proton** from forward charm production
- 📍 Search for **non-linear QCD / CGC** effects, study charm fragmentation
- 📍 Signal and background predictions for **neutrino telescopes**: IceCube, KM3NET

gluon imaging at FASER: precise predictions for **prompt neutrino fluxes** up to $E_\nu \sim 100$ PeV

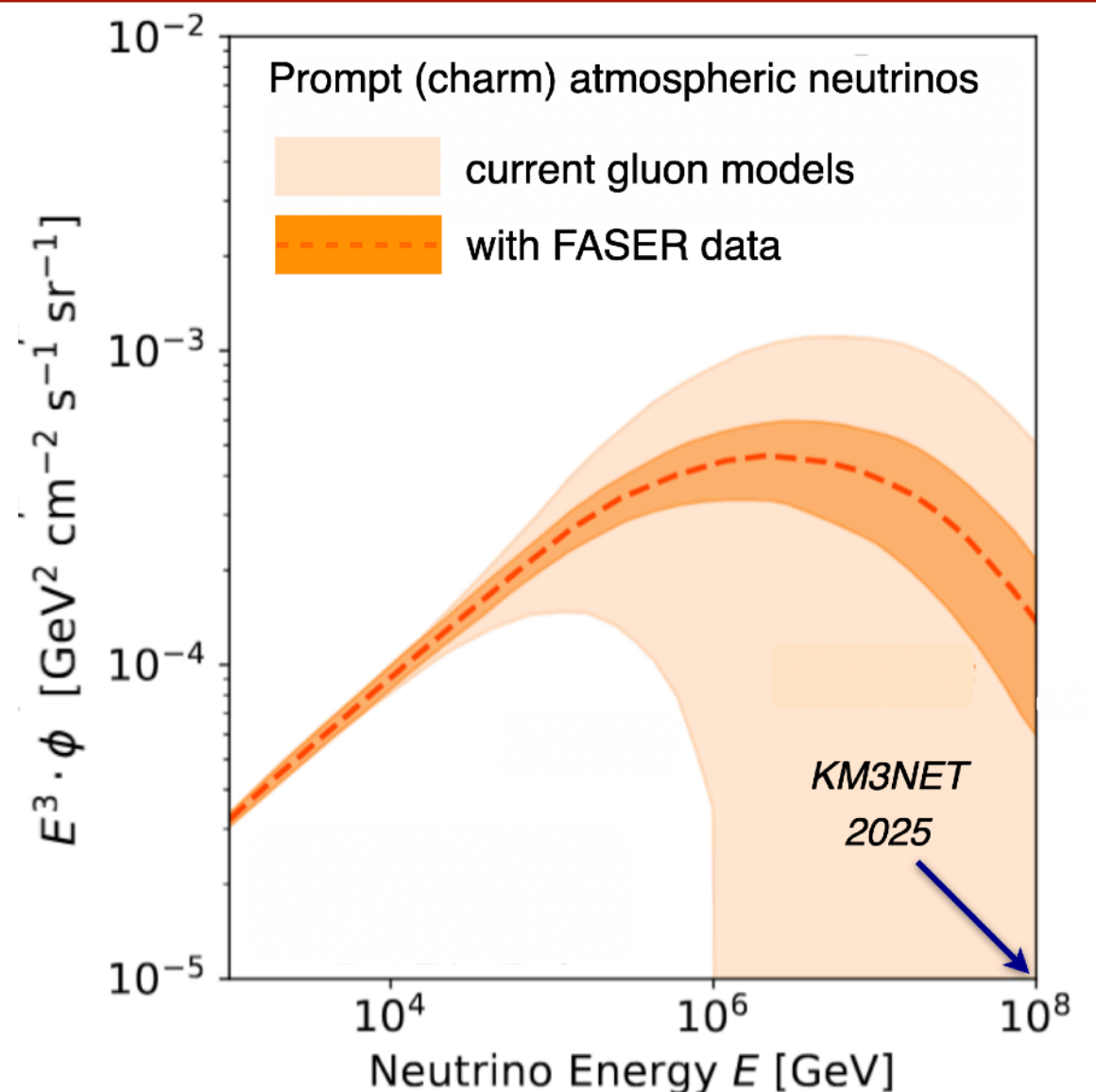


Small-x QCD at FASER

- 📍 Determine **gluon content of proton** from forward charm production
- 📍 Search for **non-linear QCD / CGC** effects, study charm fragmentation
- 📍 Signal and background predictions for **neutrino telescopes**: IceCube, KM3NET

gluon imaging at FASER: precise predictions for **prompt neutrino fluxes** up to $E_\nu \sim 100$ PeV

KM3NET25: most energetic elementary particle ever detected!

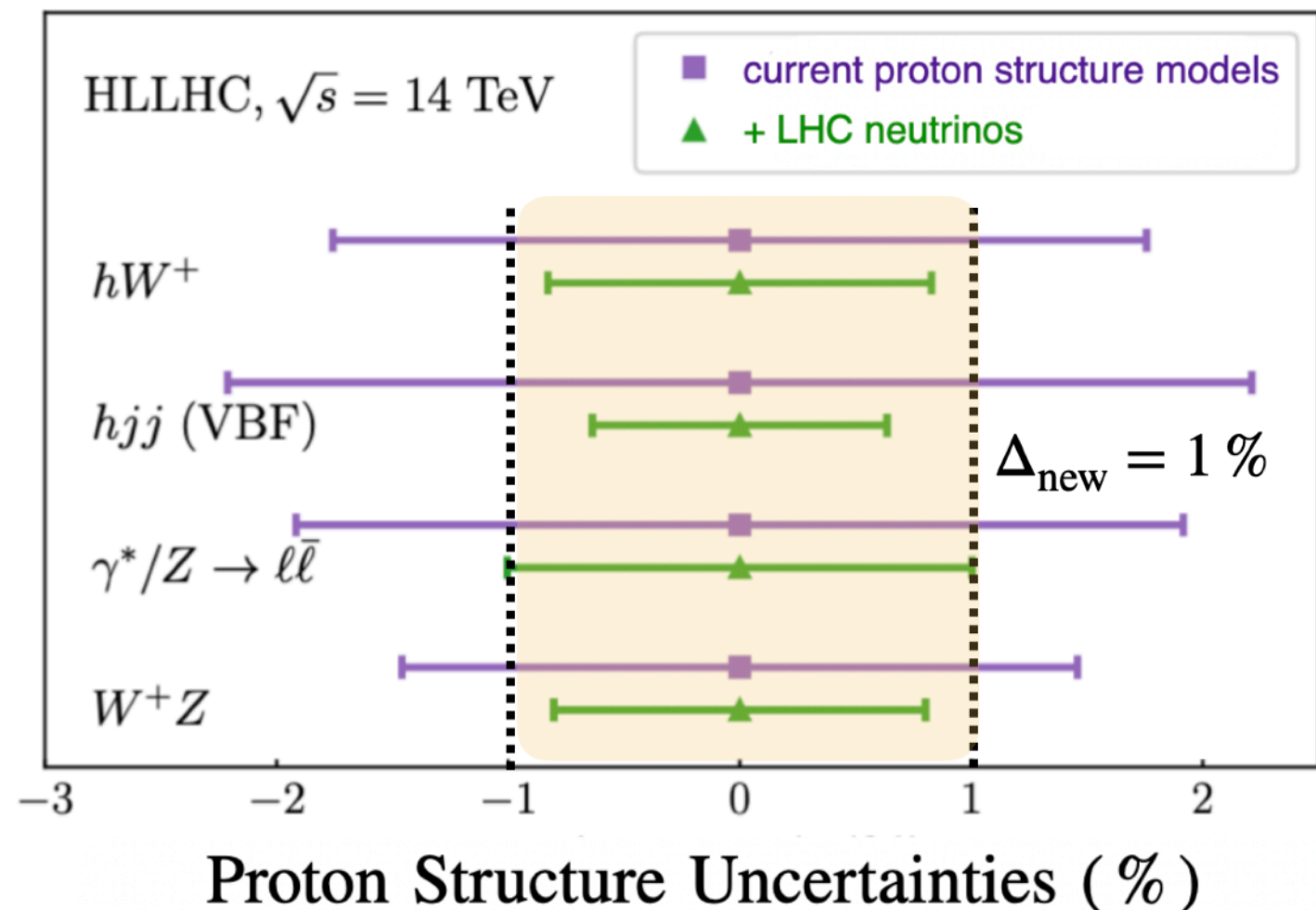


Neutrino DIS at TeV energies

- 📌 Constrain quark/antiquark flavour separation from neutrino and muon DIS at the TeV
- 📌 Tailored predictions for **Higgs** and **new heavy particles** at LHC, with large- x PDFs fixed from robust “low-energy” observables (prevent fitting away BSM)
- 📌 Anticipate **Electron-Ion Collider** flagship measurements e.g. charm production in muon DIS

up to factor 3 of sensitivity
improvement thanks for **LHC
neutrino constraints** as compared
to PDF4LHC21 reference

*Subtle BSM signals at the LHC
demand this precision*

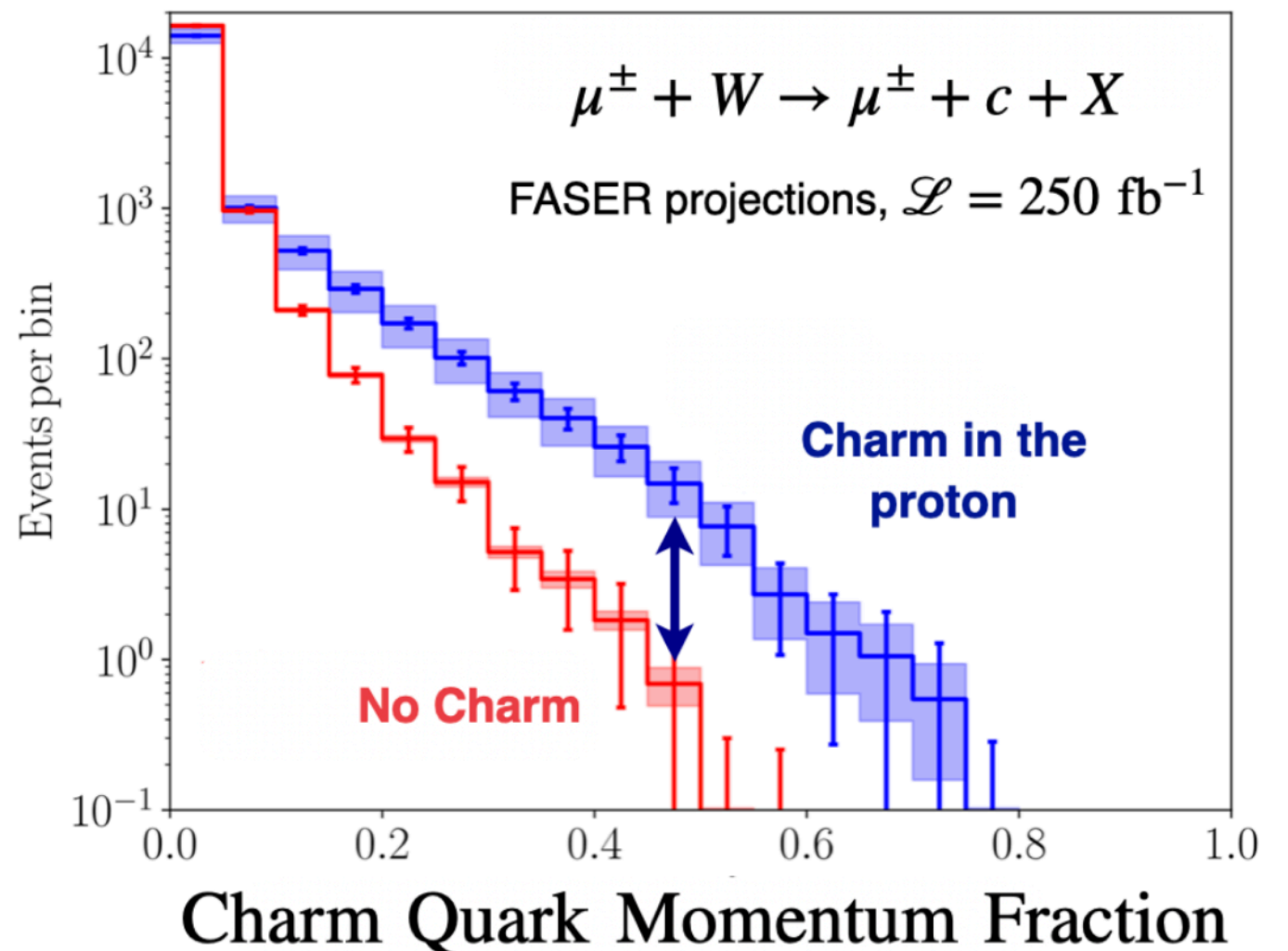


Muon DIS at TeV energies

- 📍 Constrain quark/antiquark flavour separation from neutrino and muon DIS at the TeV
- 📍 Tailored predictions for **Higgs** and **new heavy particles** at LHC, with large- x PDFs fixed from robust “low-energy” observables (prevent fitting away BSM)
- 📍 Anticipate **Electron-Ion Collider** flagship measurements e.g. charm production in muon DIS

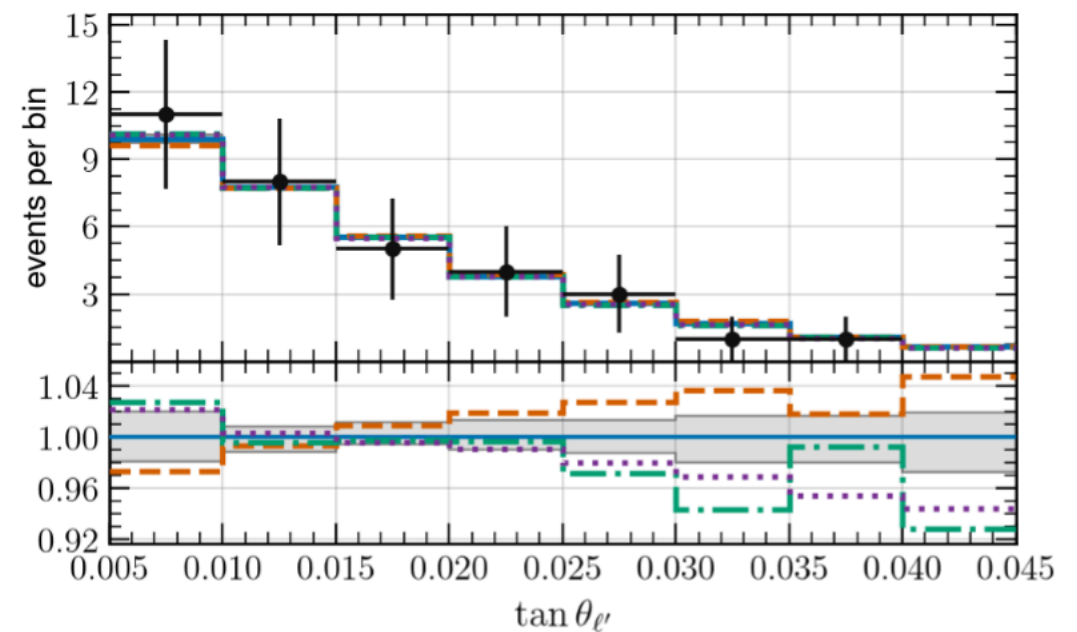
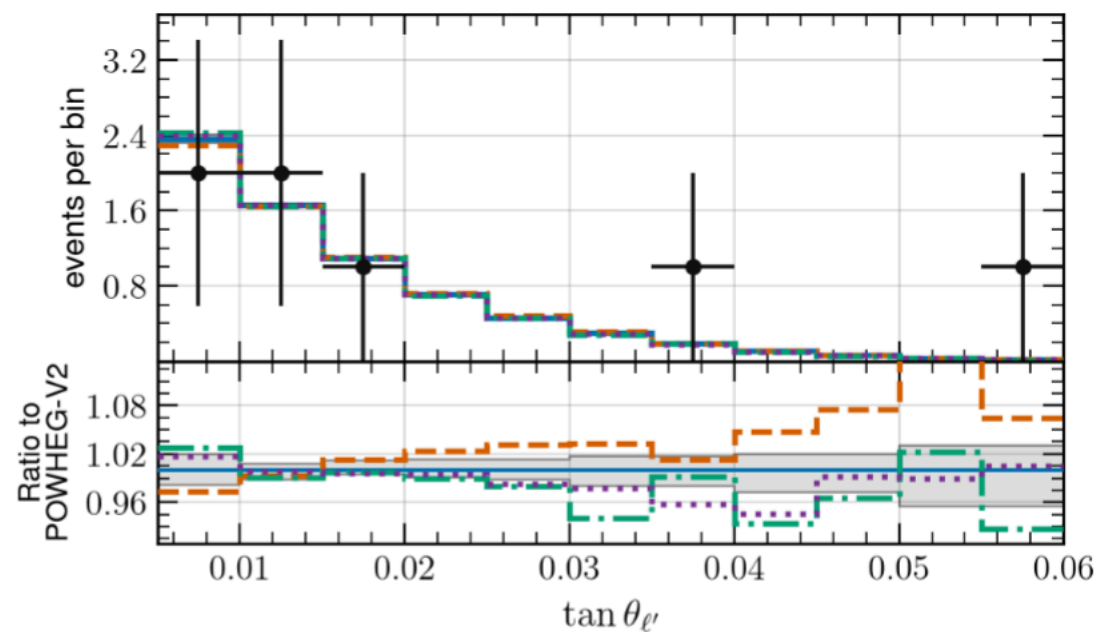
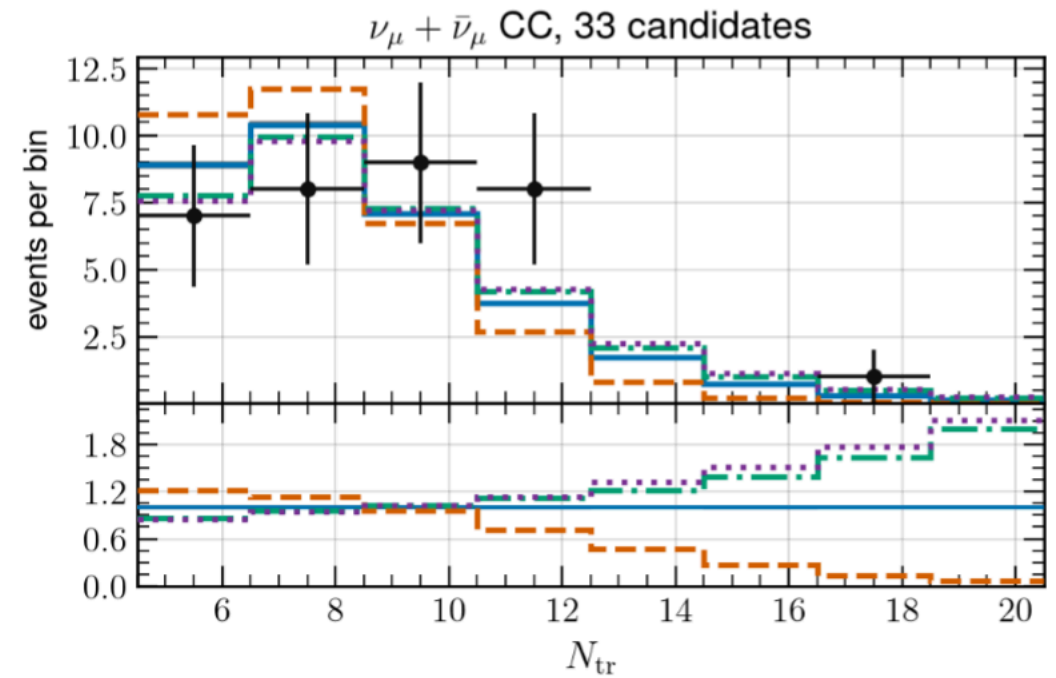
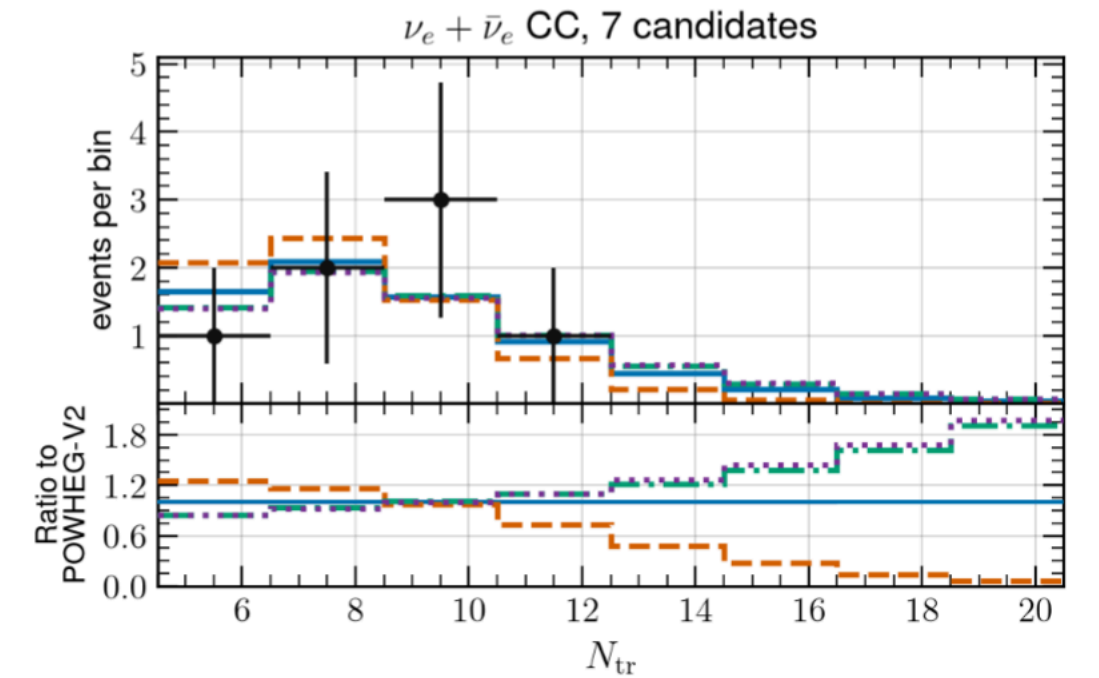
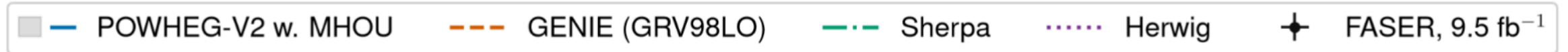
already with data on tape:

Definitive evidence for
**charm quarks in the
proton** within reach



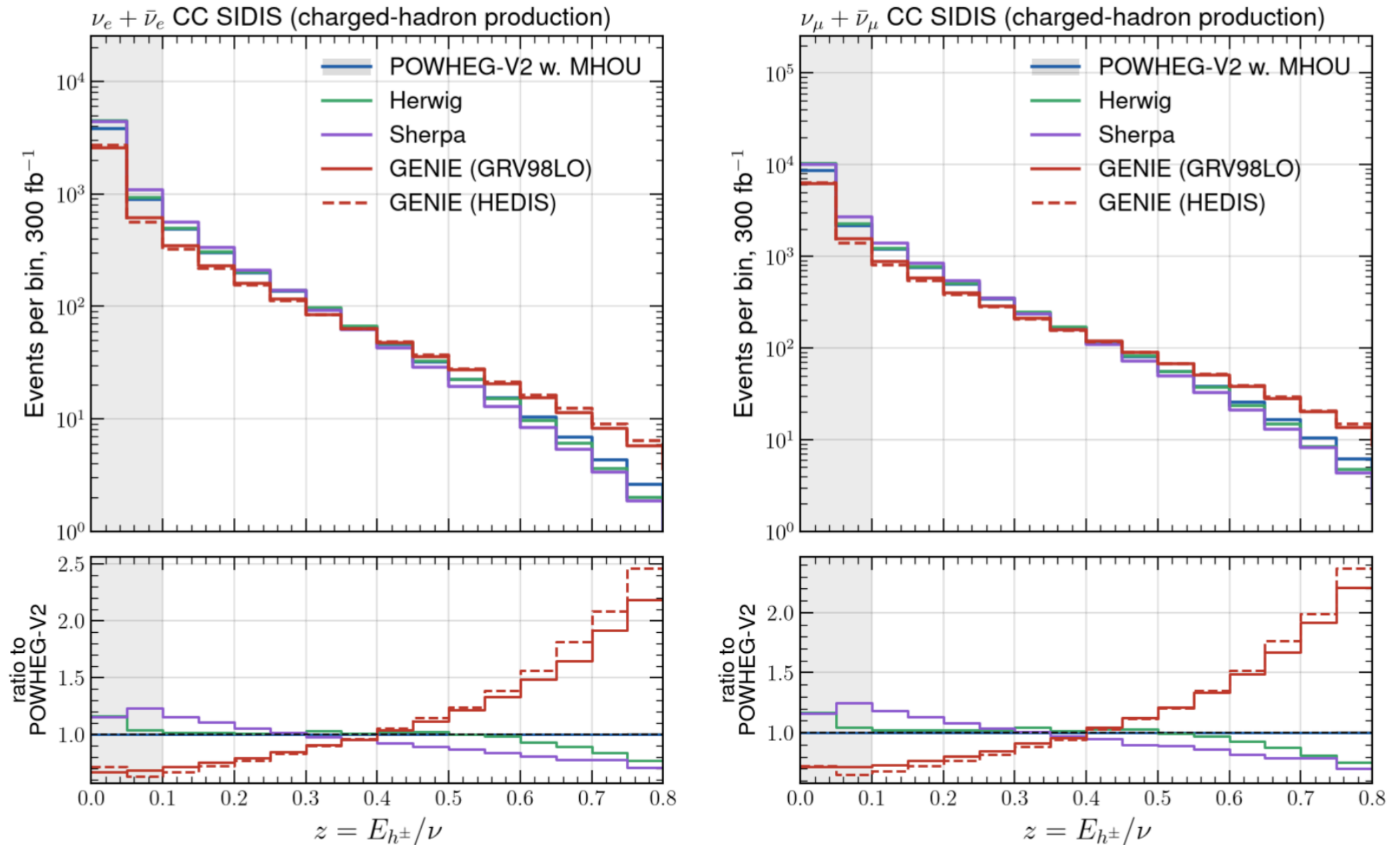
Event generation for LHC neutrinos

Comparison with FASER ν data (normalised to observed yields)



Precision DIS event generation instrumental for LHC neutrino/muon physics

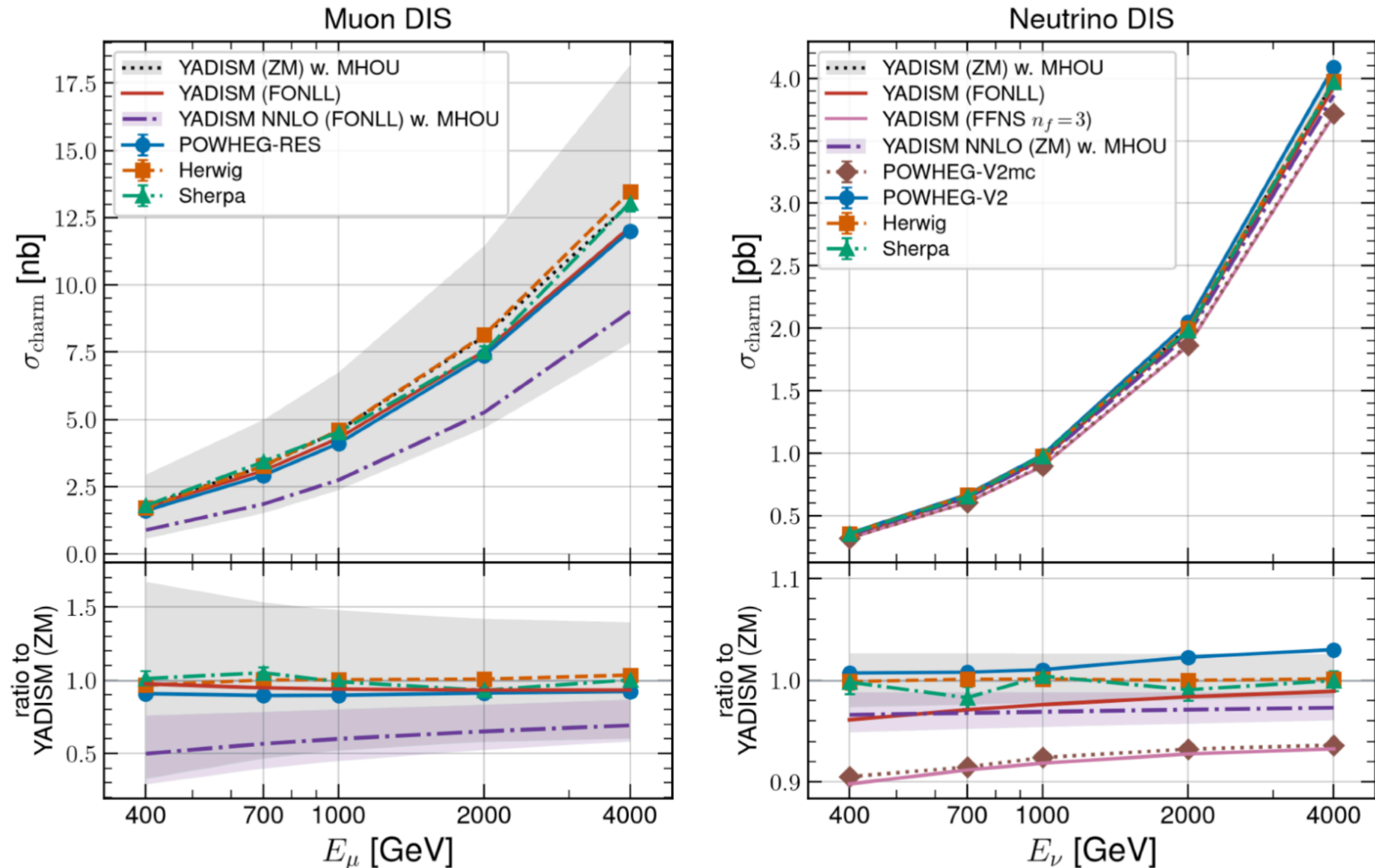
Neutrino Semi-inclusive DIS



The first ever measurement of **neutrino SIDIS at TeV energies** could be done with Run 3 data: unique tests of QCD and **fragmentation functions**

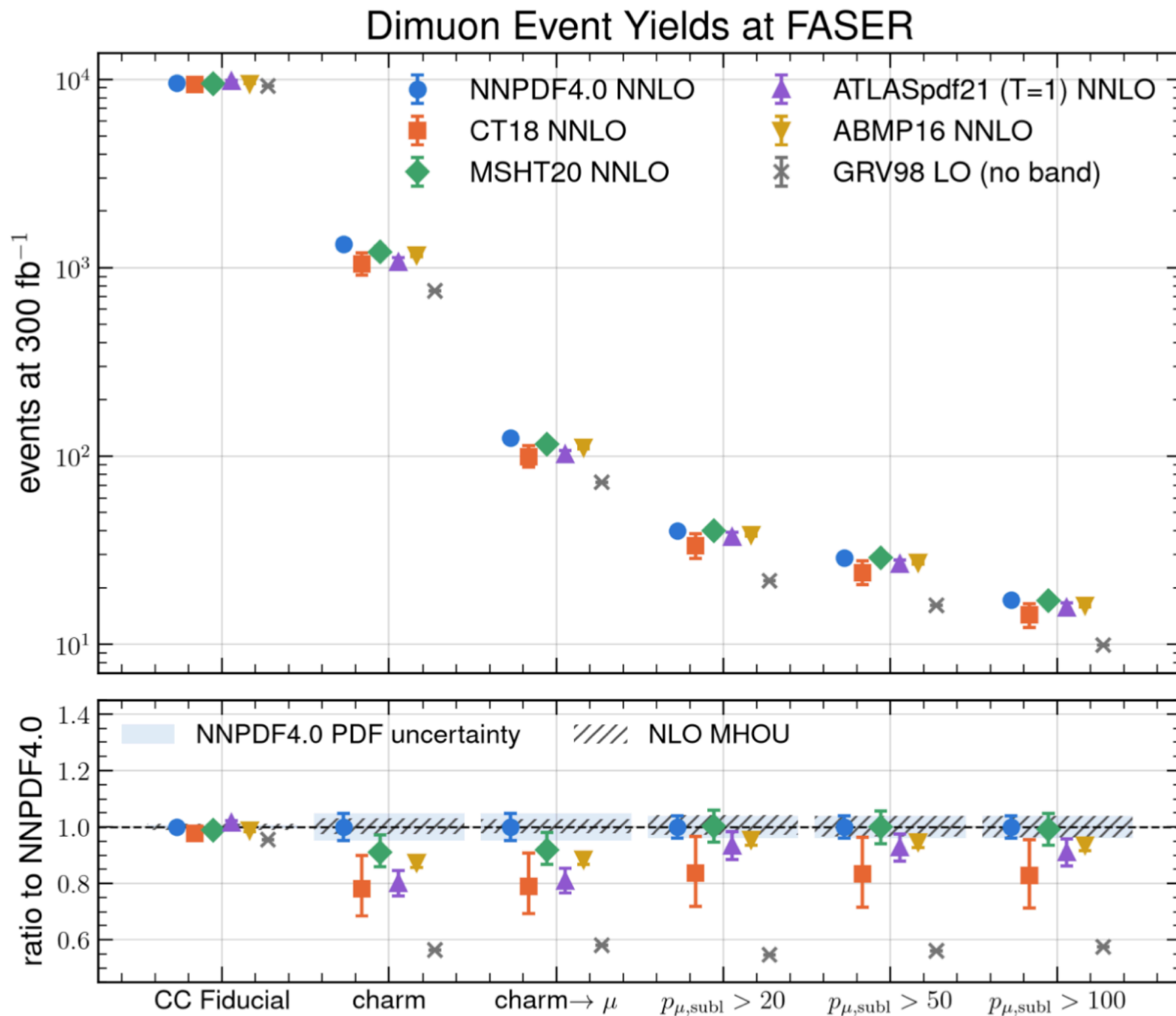
Charm production for TeV DIS

Fiducial NLO charm production cross-section vs beam energy



- 🔍 Excellent agreement between NLO generators and analytical calculations for common mass scheme
- 🔍 Muon DIS: large higher order corrections, due to Q^{-4} enhancement of the cross-section

Charm production for TeV DIS



Significant PDF dependence: **sensitivity to strange PDF**, requires careful modelling of higher order QCD corrections (both yields and acceptances)

FASER as a QCD experiment

📌 Short term (with **FASER Run 3 data**):

- 📌 Constraints on the **small- x gluon PDF** from charm production down to $x = 10^{-7}$
- 📌 Search for **deviations from DGLAP** in forward hadron production
- 📌 First measurement of muon & neutrino **DIS structure functions at TeV** energies
- 📌 Constraints on **intrinsic charm** from muon DIS
- 📌 Constraints on **strangeness** from charm production (& dimuon) in neutrino DIS
- 📌 Fragmentation functions from neutrino SIDIS

📌 Medium term (with **FASER Run 3+4 data**)

- 📌 **Muon and neutrino DIS structure functions** integrated in global PDF fits
- 📌 **Unbinned (NSBI) observables** for optimal sensitivity to proton structure
- 📌 **NNLO (+NNLL?) DIS event generators** tuned to FASER muon and neutrino data
- 📌 Constraints on **cold nuclear matter** (tungsten target)

NNLO/N³LO calculations and event generation & high-precision shower algorithms are essential to realise the potential of FASER as a precision QCD experiment