Impact of the EIC on Collinear PDFs and the Strong Coupling

Electron-Ion Collider UK Gathering 2025, York

MANCHESTER 1824
The University of Manchester

Thomas Cridge

9th December 2025

THE ROYAL SOCIETY

In collaboration with MSHT colleagues - T.C., L.A. Harland-Lang, R.S. Thorne, and others - N. Armesto, F. Giuli, P. Newman, B. Schmookler, K. Wichmann

Status of Global PDF Fitting in MSHT

- Global fit of collinear unpolarised PDFs. More than 60 different datasets.
- More than 5000 datapoints included over wide range of (x, Q^2): $10^{-4} \lesssim x \lesssim 0.8$ and $2\ \text{GeV}^2 \lesssim Q^2 \lesssim 10^6\ \text{GeV}^2$.
- Robust methodology with developments on all three fronts:
  1. Theoretical - Vast majority of processes included have full NNLO QCD theory, with NLO EW where relevant. Recent extension to approximate N3LO with theoretical uncertainties for first time. MSHT: 2207.04739.
  2. Experimental - Many new datasets, more precise, more channels, more differential. MSHT: 2012.04684, 2510.03753.
  3. Methodological - Extended parameterisation, 52 PDF parameters - allow fitting to accuracy $< 1\%$. Closure tests performed to examine central value and uncertainties. MSHT: 2407.07944.

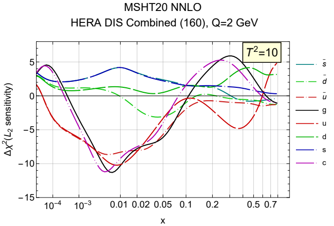
Status of Global PDF Fitting in MSHT

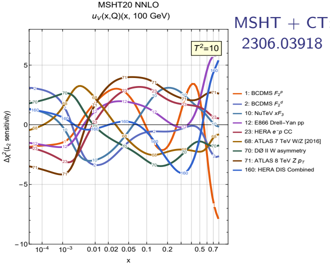
- Global fit of collinear unpolarised PDFs. More than 60 different datasets.
- More than 5000 datapoints included over wide range of (x, Q^2): $10^{-4} \lesssim x \lesssim 0.8$ and $2\ \text{GeV}^2 \lesssim Q^2 \lesssim 10^6\ \text{GeV}^2$.
- Robust methodology with developments on all three fronts:
  1. Theoretical - Vast majority of processes included have full NNLO QCD theory, with NLO EW where relevant. Recent extension to approximate N3LO with theoretical uncertainties for first time. MSHT: 2207.04739.
  2. Experimental - Many new datasets, more precise, more channels, more differential. MSHT: 2012.04684, 2510.03753.
  3. Methodological - Extended parameterisation, 52 PDF parameters - allow fitting to accuracy $< 1\%$. Closure tests performed to examine central value and uncertainties. MSHT: 2407.07944.

What can the EIC contribute to this? $\Rightarrow$ Precise, new DIS data.

Impact of Current DIS data on PDFs:

- HERA DIS data forms backbone of global PDFs still (even with new LHC data).
- Supplemented by fixed target DIS experiments to constrain high x, e.g. BCDMS.
- Tensions exist between datasets which reduce precision.
- Larger experimental plus theoretical uncertainties, e.g. higher twists at low Q^2, suppress impact of old fixed target DIS.

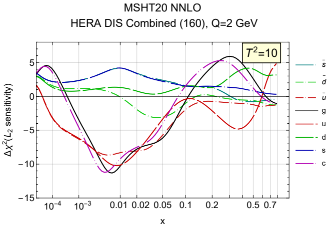


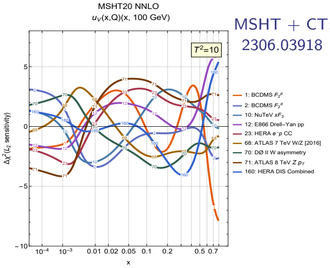


MSHT + CT: 2306.03918

Impact of Current DIS data on PDFs:

- HERA DIS data forms backbone of global PDFs still (even with new LHC data).
- Supplemented by fixed target DIS experiments to constrain high x, e.g. BCDMS.
- Tensions exist between datasets which reduce precision.
- Larger experimental plus theoretical uncertainties, e.g. higher twists at low Q^2, suppress impact of old fixed target DIS.





MSHT + CT: 2306.03918

$\Rightarrow$ EIC precise new DIS data at moderate Q^2 will help!

EIC Impact on Collinear PDFs

More information in articles:
Armesto, TC, Giuli, Harland-Lang, Newman, Schmookler, Thorne, Wichmann: 2309.11269
Phys.Rev.D 109 (2024) 5, 5.

EIC Kinematic Coverage:

- Consider NC and CC DIS at EIC.
  - Higher x coverage, still at moderate Q^2.
  - Complements HERA data.
  - EIC less sensitive to higher twists.
  - Study here - generate pseudodata for e^-p data with updated beam energies, configurations, lumis and uncertainty projections.
  - Kinematic coverage: $Q^2 > 2$ GeV^2, $0.01 < y < 0.95$, $W^2 > 15$ GeV^2.
  - Only highest $\sqrt{s}$ has CC DIS.

Armesto, TC et al 2309.11269

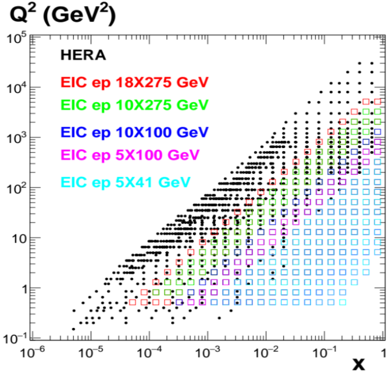


TABLE I. Beam energies, center-of-mass energies and annual integrated luminosities of the different configurations considered for the EIC.

e-beam energy (GeV)	p-beam energy (GeV)	$\sqrt{s}$ (GeV)	Integrated lumi (fb^{-1})
18	275	141	15.4
10	275	105	100
10	100	63	79.0
5	100	45	61.0
5	41	29	4.4

EIC Kinematic Coverage:

- Consider NC and CC DIS at EIC.
  - Higher x coverage, still at moderate Q^2.
  - Complements HERA data.
  - EIC less sensitive to higher twists.
  - Study here - generate pseudodata for e^-p data with updated beam energies, configurations, lumis and uncertainty projections.
  - Kinematic coverage: $Q^2 > 2\ \mathrm{GeV}^2$, $0.01 < y < 0.95$, $W^2 > 15\ \mathrm{GeV}^2$.
  - Only highest $\sqrt{s}$ has CC DIS.

Armesto, TC et al 2309.11269

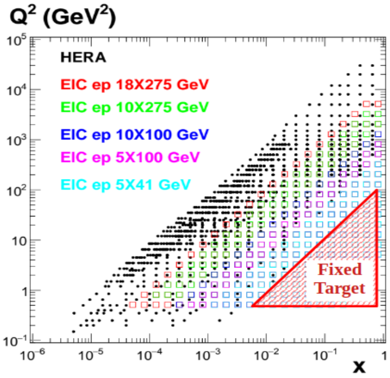
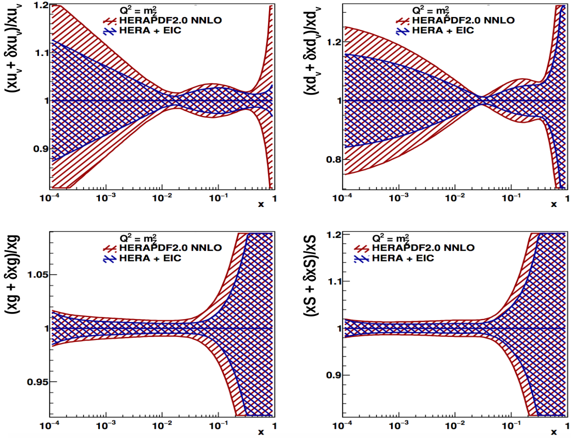


TABLE I. Beam energies, center-of-mass energies and annual integrated luminosities of the different configurations considered for the EIC.

e-beam energy (GeV)	p-beam energy (GeV)	$\sqrt{s}$ (GeV)	Integrated lumi (fb^{-1})
18	275	141	15.4
10	275	105	100
10	100	63	79.0
5	100	45	61.0
5	41	29	4.4

PDF Impact in HERAPDF:

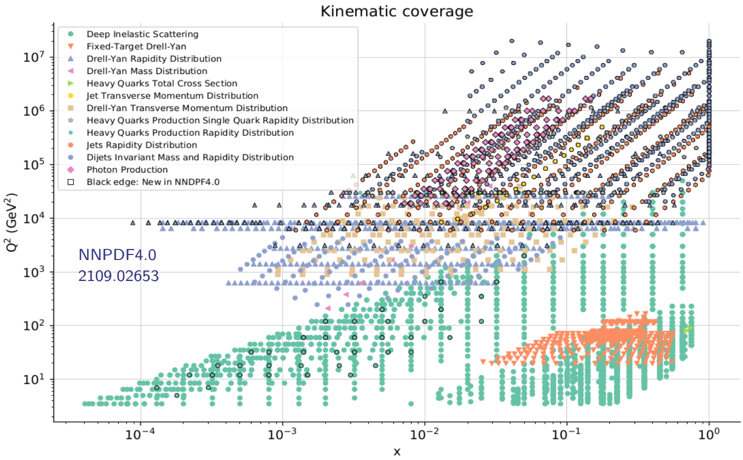
- Observed large reductions in PDF uncertainty when EIC data added on top of HERAPDF, no fixed target or LHC data.



Armesto, TC et al 2309.11269

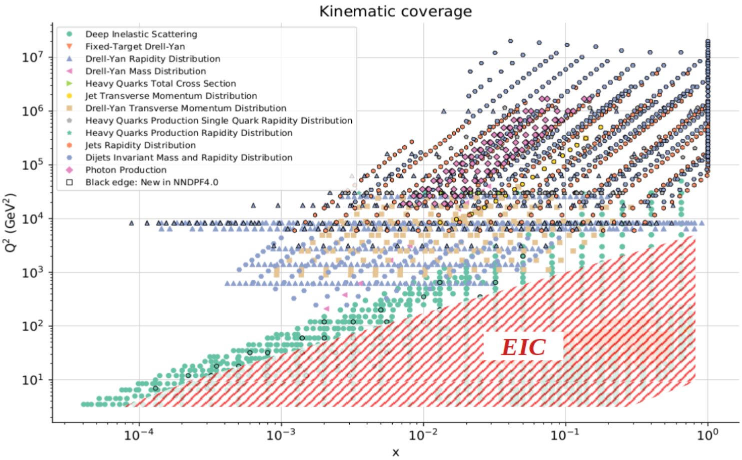
PDF Impact in MSHT:

- Does the same hold for global PDFs? Also fixed target and LHC data.



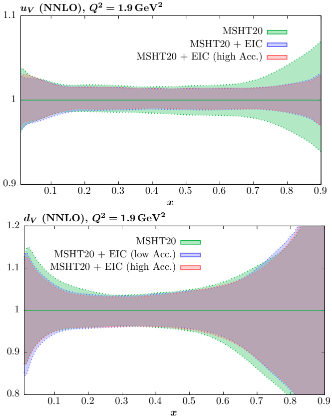
PDF Impact in MSHT:

- Does the same hold for global PDFs? Also fixed target and LHC data.



PDF Impact in MSHT:

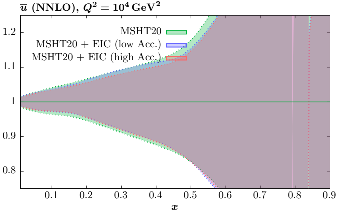
- Add the pseudodata to global MSHT PDFs at NNLO:
- Largest impact on u PDF at large x as
$$\sigma_{e^-p}^{\mathrm{NC\ DIS}} \propto \sum_i Q_i^2 f_i(x)$$
$\Rightarrow$ Uncertainty reduced by up to 50%.
- Smaller impact on d PDF.
- Impact of larger y acceptance negligible as different beam energy configurations provide constraints.
- Positron or deuteron data would increase constraints on d PDF. As would tagged DIS studies and PVDIS with polarised electrons. (see e.g. CJ/JAM in 2103.05419.)

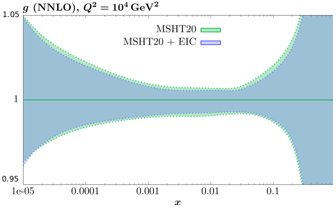


Armesto, TC et al 2309.11269

PDF Impact in MSHT:

- Add pseudodata to global MSHT PDFs at NNLO:
- Inclusive DIS has smaller impact on sea quarks, where uncertainties are larger.
- Mild reduction in gluon uncertainty across all x.
- Comes from scaling violations, $dF_2/dQ^2 \sim \alpha_S g$.
- Similar EIC constraints seen in HERAPDF but greater in magnitude there as it's not a global PDF fit.
- Also investigated sensitivity to small-x $\ln(1/x)$ resummation
  - no difference in fit quality observed.

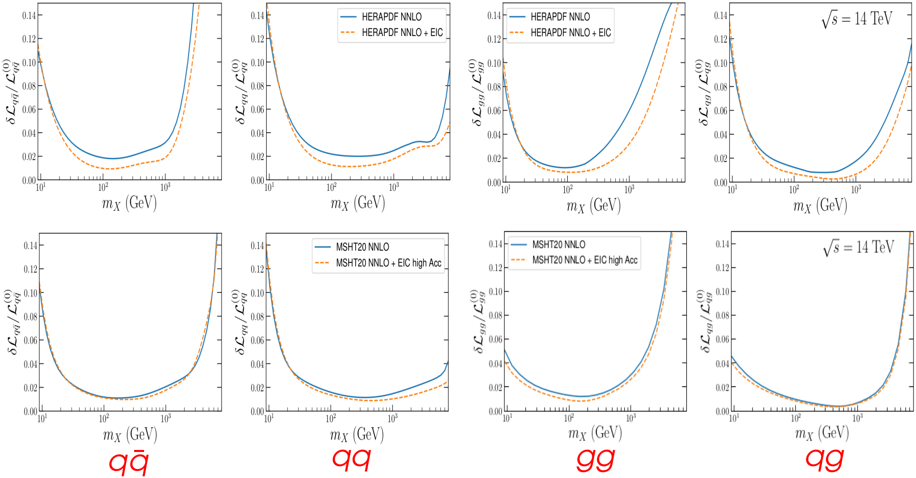




Armesto, TC et al 2309.11269

PDF Luminosity Impact in MSHT:

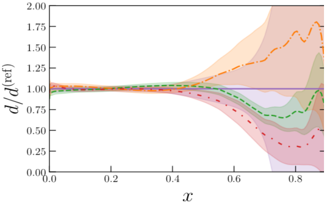
- Knock-on impact on PDF luminosity uncertainties in HERA/MSHT:

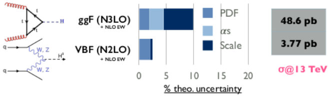


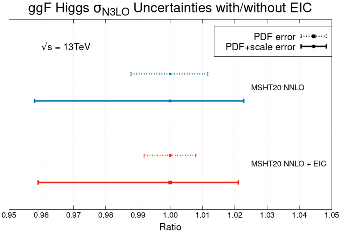
Armesto, TC et al 2309.11269

Consequences for Phenomenology:

- Why is this important?
- High x PDF (quark or gluon) uncertainties currently grow rapidly.
- Limits sensitivity to BSM physics at large invariant masses.
- Reason is lack of data and tensions observed between fixed target/LHC data $\Rightarrow$ EIC can help resolve these!
- Gluon uncertainty key for Higgs production cross-section uncertainty.
- Observe reduction in gg luminosity PDF uncertainty from 1.2% to 0.8% $\Rightarrow$ impact on $gg \to H$ cross-section.







EIC Impact on Strong Coupling Determination

More information in articles:
Harland-Lang, TC, Thorne, Newman, Wichmann 2512.06092.

Determination of the Strong Coupling Constant:

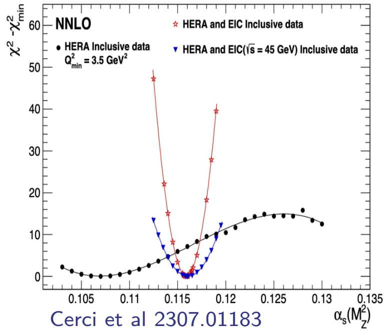
- $\alpha_S(M_Z^2)$ sensitivity in global PDF fit come from:
  - Direct $\alpha_S(M_Z^2)$ dependence in coefficient functions.

$$\mathbf{C}(\alpha_S) = \alpha_S^i[\mathbf{C}_0 + \alpha_S\mathbf{C}_1 + \alpha_S^2\mathbf{C}_2 + \alpha_S^3\mathbf{C}_3 + \ldots]$$

  - Indirect $\alpha_S(M_Z^2)$ dependence through PDF evolution.

$$\frac{df}{d\log\mu_F^2} = \begin{bmatrix} P_{qq} & n_f P_{qg} \\ P_{gq} & P_{gg} \end{bmatrix} \begin{bmatrix} \Sigma \\ g \end{bmatrix}$$

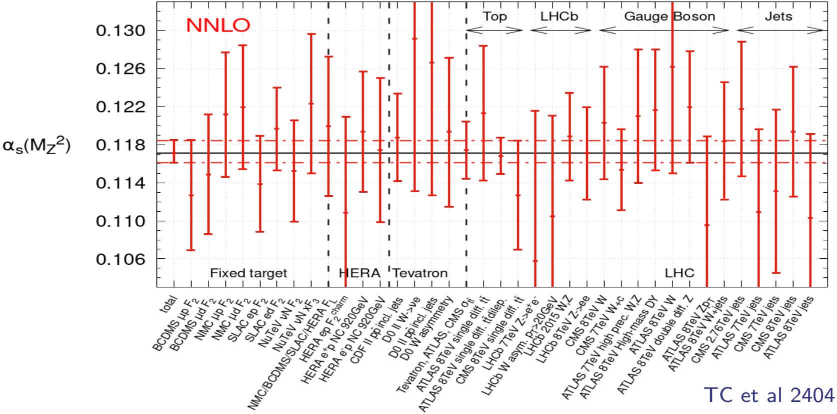
- DIS has limited sensitivity indirectly via scaling violations.
- HERA at low/intermediate x driven by gluon splitting, hard to disentangle α_S.
- EIC at higher x driven by non-singlet splitting, so α_S less correlated to g.
- Improved precision + more datapoints on structure function evolution.



Cerci et al 2307.01183

Determination of the Strong Coupling Constant:

- Can we improve our global bounds? Again have fixed target and LHC data which also bound α_S.
- First consider NNLO, MSHT recently obtained: $\alpha_{S,\mathrm{NNLO}}(M_Z^2) = 0.1171 \pm 0.0014$ (previous version in PDG).

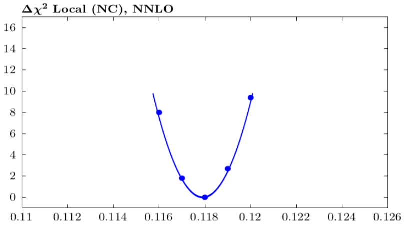


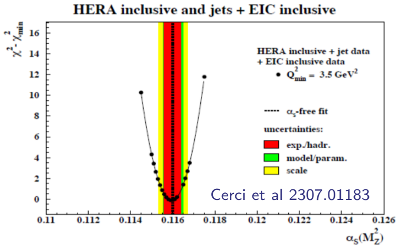
TC et al 2404.02964

Determination of the Strong Coupling:

Out today!

- Utilise EIC pseudodata generated at NNLO and with $\alpha_S(M_Z^2) = 0.118$ (larger than PDF preference of 0.117). Fit simultaneously PDF+α_S.
- Examine χ^2 profile of EIC pseudodata to determine its bounds on α_S.





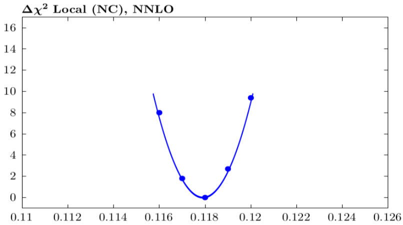
- Local (i.e. EIC data only) χ^2 profile prefers $\alpha_S(M_Z^2) = 0.118$ by construction, constraining power similar to HERAPDF+EIC α_S study.

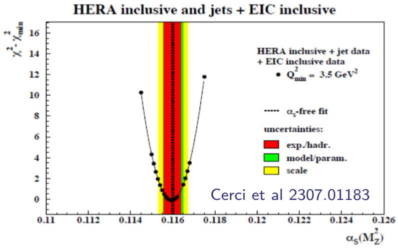
Harland-Lang, TC, Thorne, Newman, Wichmann 2512.06092

Determination of the Strong Coupling:

Out today!

- Utilise EIC pseudodata generated at NNLO and with $\alpha_S(M_Z^2) = 0.118$ (larger than PDF preference of 0.117). Fit simultaneously PDF+α_S.
- Examine χ^2 profile of EIC pseudodata to determine its bounds on α_S.



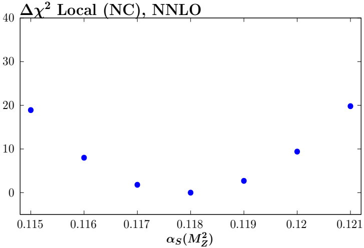


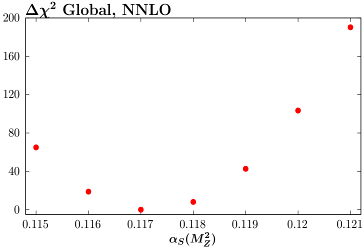
- Local (i.e. EIC data only) χ^2 profile prefers $\alpha_S(M_Z^2) = 0.118$ by construction, constraining power similar to HERAPDF+EIC α_S study.
- But what about the global α_S? $\Rightarrow$ this determines the best fit and bounds.

Harland-Lang, TC, Thorne, Newman, Wichmann 2512.06092

Determination of the Strong Coupling: Out today!

- Utilise EIC pseudodata generated at NNLO and with $\alpha_S(M_Z^2) = 0.118$ (larger than PDF preference of 0.117). Fit simultaneously PDF+α_S.
- Examine χ^2 profile of EIC pseudodata to determine its bounds on α_S.



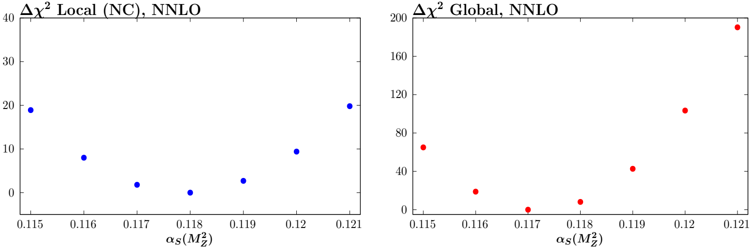


Harland-Lang, TC, Thorne, Newman, Wichmann 2512.06092

Determination of the Strong Coupling:

Out today!

- Utilise EIC pseudodata generated at NNLO and with $\alpha_S(M_Z^2) = 0.118$ (larger than PDF preference of 0.117). Fit simultaneously PDF+α_S.
- Examine χ^2 profile of EIC pseudodata to determine its bounds on α_S.



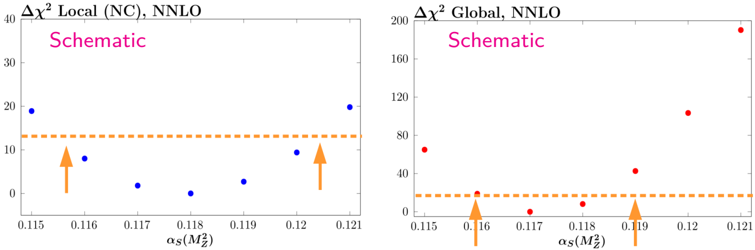
- Bounds set via dynamical tolerance $\Delta\chi^2 < (1 - \frac{\xi_{68}}{\xi_{50}}\chi_0^2) \sim 13$ for EIC NC data.

Harland-Lang, TC, Thorne, Newman, Wichmann 2512.06092

Determination of the Strong Coupling:

Out today!

- Utilise EIC pseudodata generated at NNLO and with $\alpha_S(M_Z^2) = 0.118$ (larger than PDF preference of 0.117). Fit simultaneously PDF+α_S.
- Examine χ^2 profile of EIC pseudodata to determine its bounds on α_S.

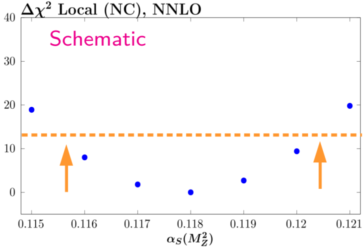


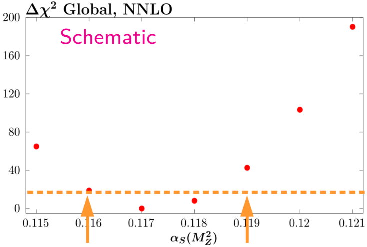
- Bounds set via dynamical tolerance $\Delta\chi^2 < (1 - \frac{\xi_{68}}{\xi_{50}}\chi_0^2) \sim 13$ for EIC NC data.

Harland-Lang, TC, Thorne, Newman, Wichmann 2512.06092

Determination of the Strong Coupling: Out today!

- Utilise EIC pseudodata generated at NNLO and with $\alpha_S(M_Z^2) = 0.118$ (larger than PDF preference of 0.117). Fit simultaneously PDF+α_S.
- Examine χ^2 profile of EIC pseudodata to determine its bounds on α_S.





- Bounds set via dynamical tolerance $\Delta\chi^2 < (1 - \frac{\xi_{68}}{\xi_{50}}\chi_0^2) \sim 13$ for EIC NC data. Upper bound on α_S found not competitive.
- Lower bound ~ -0.0015 competitive to global fit, similar to SLAC/NMC d which currently set limits.

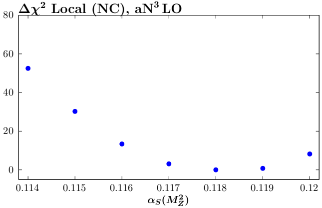
Harland-Lang, TC, Thorne, Newman, Wichmann 2512.06092

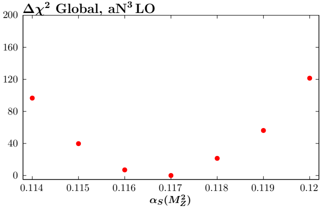
Determination of the Strong Coupling:

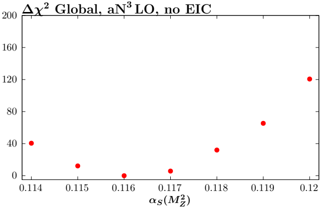
Out today!

- What about approximate N3LO? MSHT first aN3LO PDF+α_S extraction, what effect does EIC have?

TC et al 2404.02964







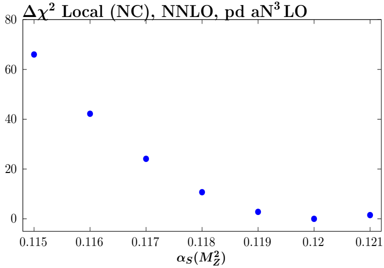
Harland-Lang, TC, Thorne, Newman, Wichmann 2512.06092

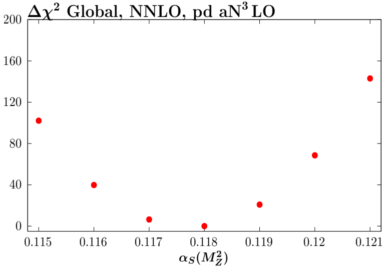
- Profiles similar to NNLO, slightly shifted down as expected.

Determination of the Strong Coupling:

Out today!

- BUT ⇒ precise bounds depend on $\alpha_S(M_Z)$ pseudodata generated with and on the precise theory used. Consistency usually assumed.
- More realistic: some inconsistency between data/theory etc.
- Model by generating pseudodata at aN3LO and fit at NNLO.

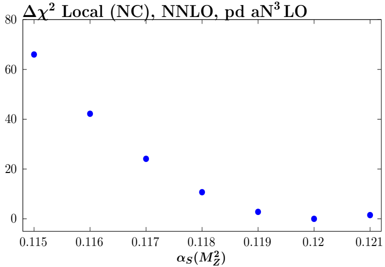




Harland-Lang, TC, Thorne, Newman, Wichmann 2512.06092

Determination of the Strong Coupling: Out today!

- BUT ⇒ precise bounds depend on $\alpha_S(M_Z)$ pseudodata generated with and on the precise theory used. Consistency usually assumed.
- More realistic: some inconsistency between data/theory etc.
- Model by generating pseudodata at aN3LO and fit at NNLO.



- Local (EIC) preference ~ 0.120 as higher α_S compensates for missing higher order corrections. ⇒ MHOU theory uncertainties important.
- Global preference therefore more shifted up to ~ 0.118.

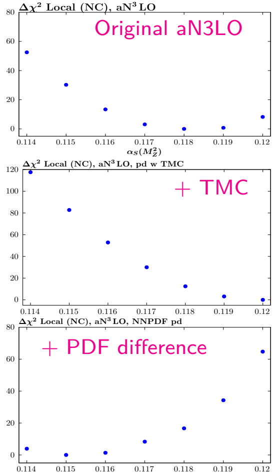
Harland-Lang, TC, Thorne, Newman, Wichmann 2512.06092

Determination of the Strong Coupling: Out today!

- What effects are important to consider?
  - Missing Higher Order Uncertainties
  - Target Mass Corrections (TMCs)
  - PDF differences.

$$F_2^{\mathrm{TMC}}(x,Q^2) \approx \frac{x^2}{\xi^2 r^3} F_2^{(0)}(\xi,Q^2)\left[1+\frac{6x\mu\xi}{r}(1-\xi)^2\right]$$

- Model with inconsistent pseudodata.
- All increase theory uncertainty $\Rightarrow$ relevant theory ingredients that should be accounted for.

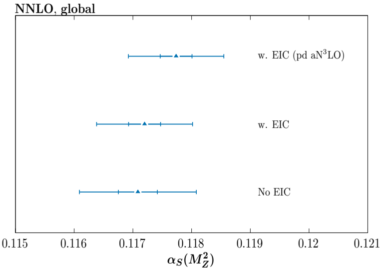


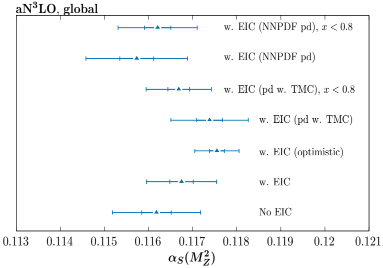
Harland-Lang, TC, Thorne, Newman, Wichmann 2512.06092

Determination of the Strong Coupling:

Out today!

- Overall results:





- EIC can help constrain $\alpha_S(M_Z)$ at both NNLO and aN3LO.
- Precise impact depends on consistency with existing global PDF inputs.
- Theoretical uncertainties are important to account for in uncertainty.
- Our aN3LO results include both MHOU and TMC uncertainties and tolerance, latter help account for dataset and PDF differences.

Harland-Lang, TC, Thorne, Newman, Wichmann 2512.06092

Conclusions:

- EIC provides important constraints on collinear PDFs in its own right.
- Constrains proton in high x low/moderate Q^2 region, complementary to HERA at lower x and current/future LHC data at higher Q^2.
- Combination with collider programs elsewhere enhances this further.
- EIC can constrain $\alpha_S(M_Z)$, competitive with current PDF fit bounds.
- Important to account for theory uncertainties and inconsistencies.
- Interplay of preferred α_S and uncertainty on bounds is often neglected.
- Improved precision on PDFs and α_S directly reduces uncertainties on many key SM processes.
- Also potential constraints from F_L, SIDIS, tagged DIS, heavy quarks.

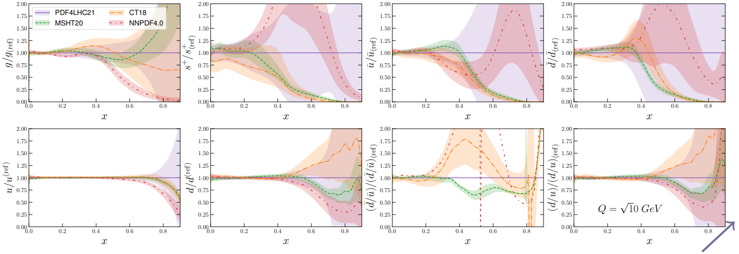
See Armesto, TC et al 2309.11269 and today's Harland-Lang, TC, Thorne et al 2512.06092.

Acknowledgements

T. Cridge acknowledges funding funding by Research Foundation-Flanders (FWO) project number: 12E1323N and by the Royal Society through Grant URF\R1\251540.

High x PDF Comparison

- High x PDFs important for BSM searches, yet quite unconstrained.
- High x PDFs constrained by fixed target, asymmetries, LHC (e.g. jets, top, Zp_T). Use of high x low Q^2 data limited by Q^2, W^2 cuts.
- PDFs at very large x and low Q are connected to collider measurements at lower x and high Q by evolution.



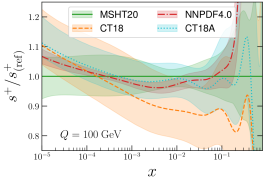
- Quite large spread of the PDFs at high x + uncertainties grow rapidly!
- Both related to fact we have *limited data in this region*:
  - Data differences/tensions can have a larger effect.
  - More sensitive to methodological differences + theoretical assumptions.

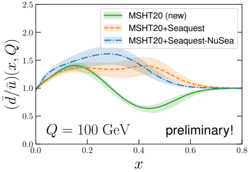
Extrapolation

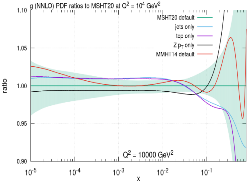
High x PDF Comparison

Data effects

- Strangeness raised by inclusion of ATLAS high precision 7, 8 TeV W, Z data - not in CT18.
- Overall strangeness is balance of this LHC precision DY data with older NuTeV dimuon data.
- $\bar{d}/\bar{u}$ raised at $x \sim 0.4$ by Seaquest data. Included only in NNPDF4.0. Seaquest tension with NuSea?
- Recent STAR data on W^+/W^- may also be relevan
- High x gluon affected by balance of LHC jet, top an Zp_T data + treatment of correlated systematics' issues.
- High x at low Q^2 connected to lower x at higher Q^2 by evolution $\Rightarrow$ data at lower x may have indirect effects. Sum rules connect different x regions.



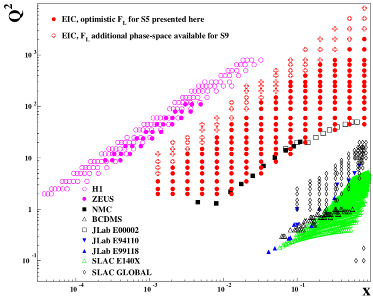


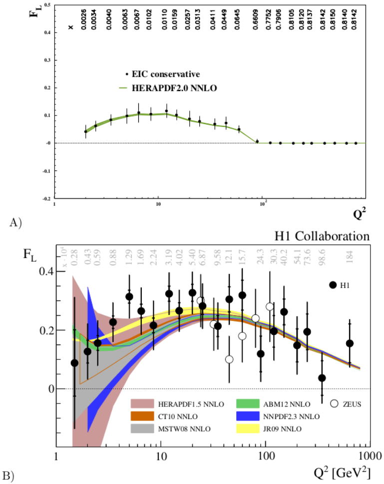


Further PDF Constraints - F_L in HERAPDF:

- Additional direct sensitivity to the gluon from F_L measurements. $F_L \sim \alpha_S g$
- Possible over larger range than HERA.
- Separate by Rosenbluth method using:

$$\sigma_{red}^{NC} \sim F_2 - \frac{y^2}{Y_+} F_L$$



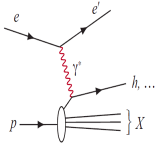


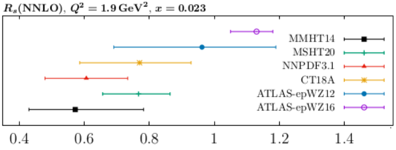
Jiménez-Lopez, Newman, Wichmann 2412.16123

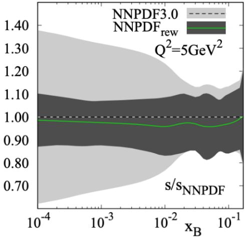
Further PDF Constraints - Strangeness:

- Limited strangeness sensitivity from inclusive DIS EIC measurements.
- Use SIDIS - parton content of outgoing hadron is connected to fragmenting parton and via CC/NC vertex to the parton in the proton.
- Pickup uncertainties from fragmentation functions.
- Similar to νDIS already used from NuTeV, which provides main constraint on $s - \bar{s}$ asymmetry, and from future FPF at CERN.
- Proton strangeness observed to be larger at LHC.
- Further s constraints come from charm jets.

(see e.g. Arratia, Hobbs et al in 2006.12520)



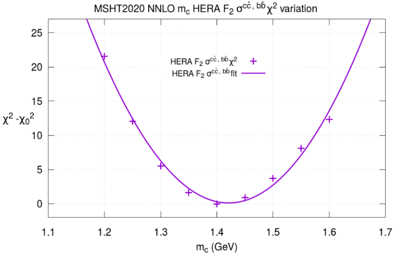




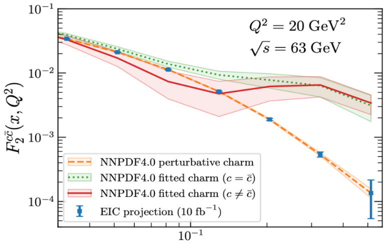
NNPDF in 2103.05419

Further PDF Constraints - Heavy Quarks:

- Measurements of charm and bottom structure functions will be extended to higher x.
- Gives sensitivity to high x heavy quark PDFs, and to heavy quark masses. E.g. used HERA data in MSHT (lower left).
- Recent suggestions of a fitted charm component of proton at high x by NNPDF, using EMC F_2^c and LHCb (Z+c) data.
- Several questions in community about this $\Rightarrow$ can be resolved by EIC.



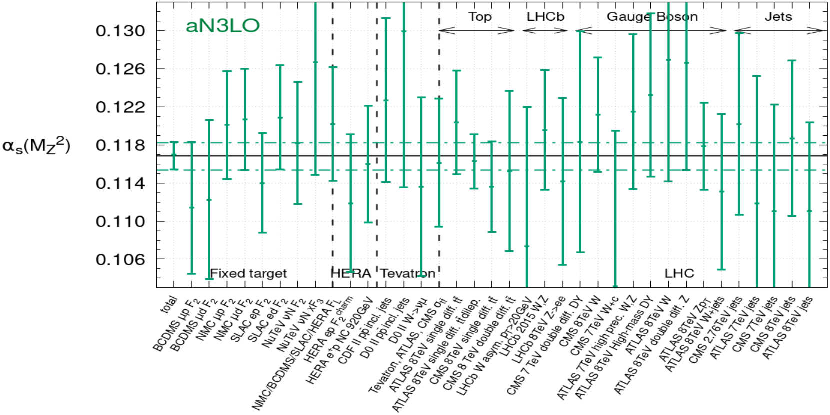
MSHT, TC et al 2106.10289



NNPDF, R.D. Ball et al 2311.00743

Determination of the Strong Coupling Constant:

- MSHT recently performed first determination in PDF fit at approximate N3LO: $\alpha_{S,\mathrm{aN3LO}}(M_Z^2) = 0.1170 \pm 0.0016$.



TC et al 2404.02964

MSHT20 Dataset Type $\alpha_S(M_Z)$ preferences

- Different data subsets prefer different $\alpha_S(M_Z)$ values.
- DY/Jets tend to prefer higher/lower $\alpha_S(M_Z)$.
- As for PDFs: reflects data tensions, data theory inconsistency, etc.
- Requires enlarged uncertainty definition $\Rightarrow$ Tolerance.
- α_S uncertainty at aN3LO corresponds to $\Delta\chi^2 \approx 16$.