Nuclear PDFs: Status, challenges, prospects

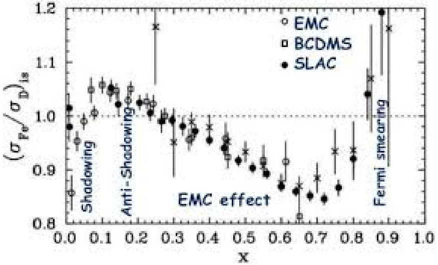
DIS2025

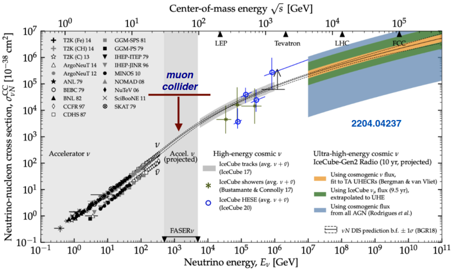
Emanuele R. Nocera

Università degli Studi di Torino and INFN — Torino

Cape Town, South Africa - 24 March 2025

Why nuclear parton distribution functions?





nuclei do not behave as a simple incoherent superposition of protons and neutrons

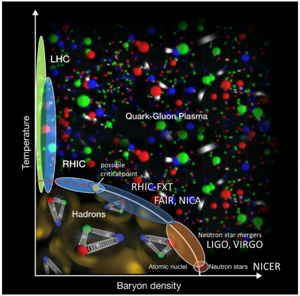
nPDFs enter theoretical predictions of signal and background events at high-energy neutrino observatories such as KM3NET and IceCube

search for exotic forms of QCD matter, such as the gluon-dominated Color Glass Condensate

interplay with proton PDFs, given the data used

[See Ann. Rev. Nucl. Part. Science 74 (2024) 49]

[See M. Klasen, WG1 Wed. afternoon]



Determining nuclear PDFs from experimental data

Assumption:
fundamental interactions are the same in the vacuum and in the medium,
but PDFs are different, *i.e.* nuclear effects are reabsorbed into nPDFs

Keep standard QCD framework and assume factorisation also for nuclei

Use the same DGLAP equations and coefficient functions as for proton PDFs

Require the same theoretical constraints (*e.g.* sum rules, positivity, ...)

$$\sigma_{\rm DIS}^{\ell+A} = \sum_i \underbrace{f_i^{(A,Z)}(x,\mu^2)}_{\text{nuclear PDFs, obey usual DGLAP}} \otimes \underbrace{\hat{\sigma}_{\rm DIS}^{\ell+i}(x,\mu^2)}_{\text{usual QCD coefficient functions}} + \text{p.s. corrections}$$

Nuclear PDFs are linear combinations of bound proton and neutron ($f_i^{(p,n),A}$) PDFs

$$f_i^{(A,Z)}(x,\mu^2) = \frac{Z}{A} f_i^{p,A}(x,\mu^2) + \frac{Z-A}{A} f_i^{n,A}(x,\mu^2)$$

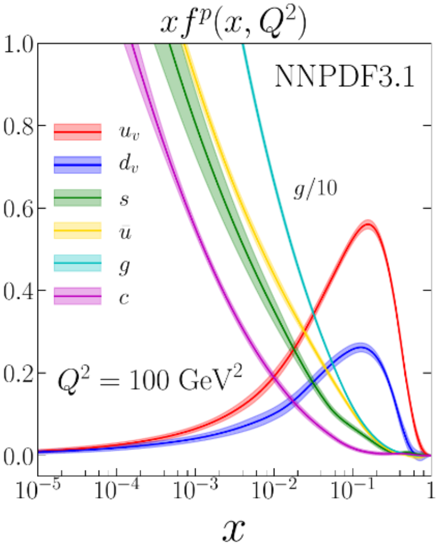
A relationship between the bound proton PDFs and the nucleon PDFs is assumed

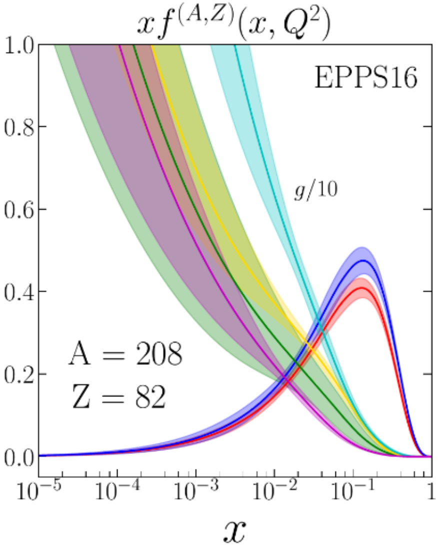
$$f_i^{p,A}(x,\mu^2) = R_i^A(x,\mu^2) f_i^p(x,\mu^2)$$

Restrictive framework, but it allows for making testable predictions for many hard probes
In principle interactions could be different in the medium or factorisation could not hold
Additional complication: nuclear modifications of final-state hadrons

1. Status of nuclear PDF fits

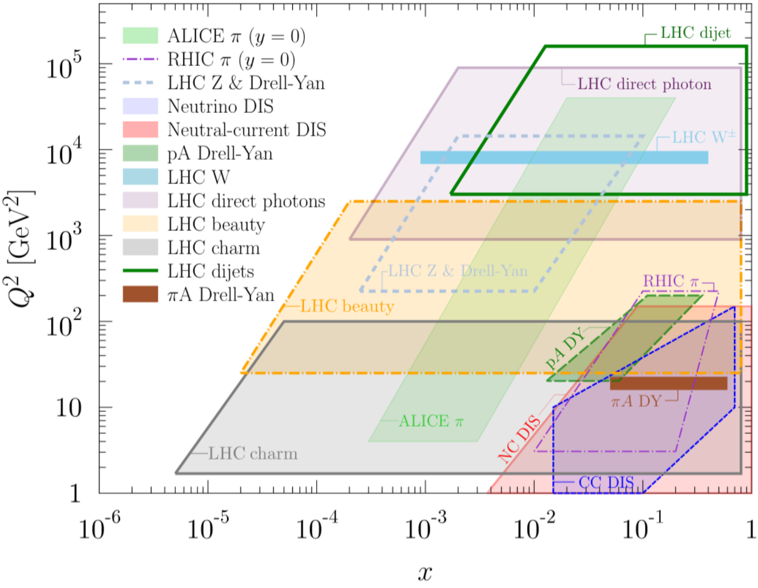
Proton and nuclear PDFs *vis-à-vis*





[Figure adapted from Ann.Rev.Nucl.Part.Sci. 70 (2020) 43]

Overview of experimental data



[Figure taken from Ann. Rev. Nucl. Part. Science 74 (2024) 49]

Overview of nuclear PDF determinations

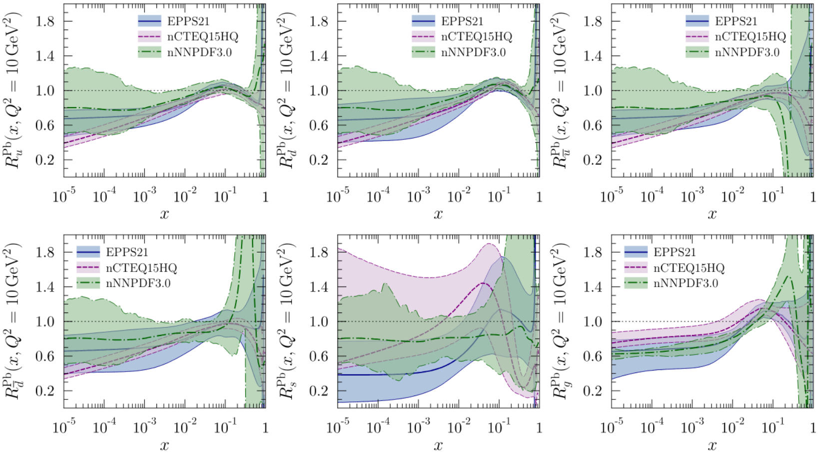
	KSASG20	TUJU21	EPPS21	nNNPDF3.0	nCTEQ15HQ
Order in α_s	NLO & NNLO	NLO & NNLO	NLO	NLO	NLO
lA NC DIS	✓	✓	✓	✓	✓
νA CC DIS	✓	✓	✓	✓	
pA DY	✓		✓	✓	✓
πA DY			✓		
RHIC dAu π^0,$\pi^\pm$			✓		✓
LHC pPb π^0,$\pi^\pm$,$K^\pm$					✓
LHC pPb dijets			✓	✓	
LHC pPb HQ			✓	✓ reweight	✓ ME fitting
LHC pPb W,Z		✓	✓	✓	✓
LHC pPb γ				✓	
Q, W cut in DIS	1.3, 0.0 GeV	1.87, 3.5 GeV	1.3, 1.8 GeV	1.87, 3.5 GeV	2.0, 3.5 GeV
$p_{\rm T}$ cut in HQ,inc.-h	N/A	N/A	3.0 GeV	0.0 GeV	3.0 GeV
Data points	4353	2410	2077	2188	1496
Free parameters	9	16	24	256	19
Error analysis	Hessian	Hessian	Hessian	Monte Carlo	Hessian
Free-proton PDFs	CT18	own fit	CT18A	~NNPDF4.0	~CTEQ6M
Free-proton corr.	no	no	yes	yes	no
HQ treatment	FONLL	FONLL	S-ACOT	FONLL	S-ACOT
Indep. flavours	3	4	6	6	5

KSAG20 [PRD 104 (2021) 034010] TUJU21 [PRD 105 (2022) 094031] EPPS21 [EPJ C82 (2022) 413]
nNNPDF3.0 [EPJ,C82 (2022) 507] nCTEQ15HQ [PRD 105 (2022) 114043]

Current PDF sets differ in three respects:
the data set, the perturbative accuracy, and the fitting methodology

[Slide adapted from P. Paakkinen]

Comparison of nuclear PDF sets



Evidence of shadowing and anti-shadowing in all partons

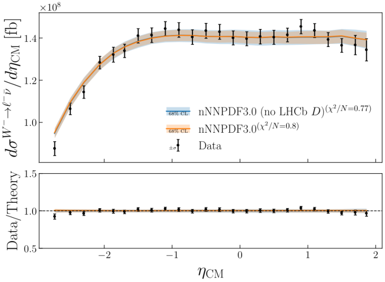
Evidence of EMC effect in EPPS21 for u and d

Overall qualitative agreement across PDF sets, except for nCTEQ strange

Difference in uncertainties depends on differences in methodologies and in data sets

Impact of data

W bosons in pPb at 8.16 TeV
CMS, Run II [PLB 800 (2020) 135048]

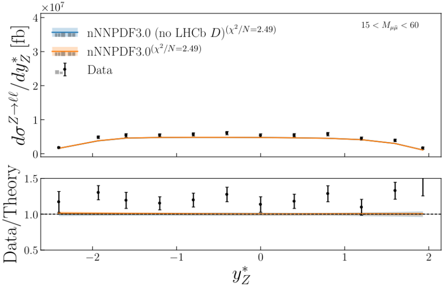


sensitivity to quark flavour separation and to the gluon through quark-gluon correlations at small x and high Q

nCTEQ15WZSIH, TUJU21 and nNNPDF3.0 fit absolute cross section, EPPS21 fit nuclear modification ratios

All find a good description of the data set

Z bosons in pPb at 8.16 TeV
CMS, Run II [JHEP 05 (2021) 182]



nNNPDF3.0 include both low-mass and on-peak data pp/pPb studied in EPPS21 but not included

Both EPPS21 and nNNPDF3.0 observe data/theory tensions: shift over rapidity

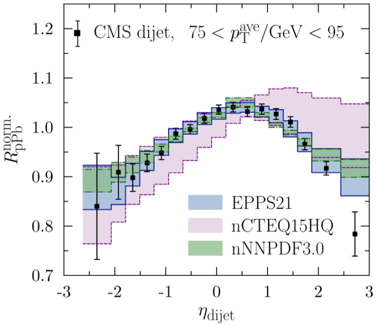
NNLO to cure for the low-mass data?

[See also the impact of neutrino dimuon data, JHEP 09 (2024) 043; S. Yrjänheikki, WG1 Tue. morning]

Impact of data

Dijets in pPb/pp at 5.02 TeV

CMS [PRL 121 (2018) 062002]

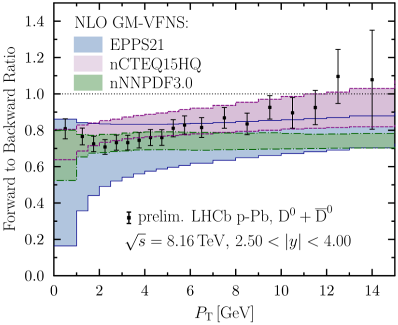


Drastic reduction of nPDF uncertainties
Important constraint for the gluon nPDF

Both EPPS21 and nNNPDF3.0
do not fit the most forward data points
missing data correlations?
NNLO? Nonperturbative effects?

D^0 mesons in pPb/pp at 5.02 TeV

LHCb [JHEP 10 (2017) 090]



Drastic reduction of nPDF uncertainties
Important constraint for the gluon nPDF

nNNPDF3.0: POWHEG+PYTHIA
large scale uncertainty, only forward data
EPPS21: S-ACOT-M_T GM-VFNS
large scale uncertainties not seen

[Importance of mass schemes: V. Bertone WG1 Tue. morning]

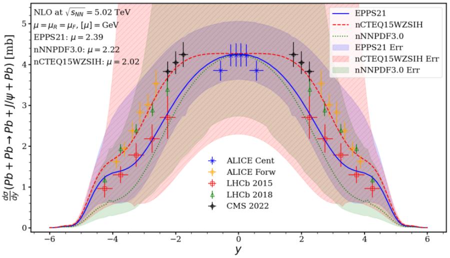
Impact of data

Quarkonia: B, J/Ψ, Υ

ALICE, LHCb, CMS

quarks enter at NLO

$$\mathcal{M}^{\gamma A\to J/\psi A} \sim f^A_{\rm gluon}(\mu)\otimes T_g(\mu) + f^A_{\rm quark}(\mu)\otimes T_q(\mu)$$



[PRC 106 035202; ibid. 107 044912]

Quadratic dependence of PDFs

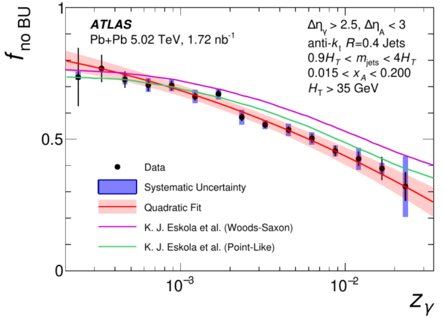
Large scale dependence at NLO

Only gluons at LO, quarks dominate at NLO

nCTEQ15WZSIH reproduces the shape thanks to its hugely enhanced strange PDF

Ultraperipheral collisions

ATLAS [PRD 111 (2025) 052006]



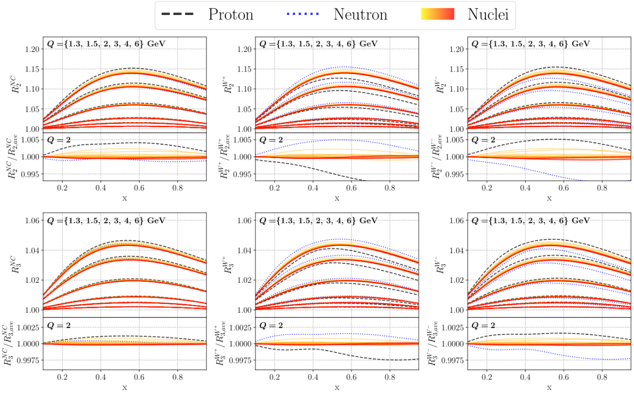
Photoproduction of dijets potentially a good constraint on nuclear PDFs

Nuclei in UPCs often approximated as point-like objects (not optimal)

Model uncertainty from neutron-class

Impact of Target Mass Corrections

$$\overbrace{F_2^{A,\mathrm{TMC}}(x_N,Q^2)}^{\textbf{TMC}} = \underbrace{\left(\frac{x_N^2}{\xi_N^2 r_N^3}\right)\overbrace{F_2^{A,(0)}(\xi_N,Q^2)}^{\substack{\textbf{No-TMC}\\ \textbf{with } \xi\to\mathbf{x}}}}_{\textbf{Leading-TMC}} + \underbrace{\left(\frac{6M_N^2 x_N^3}{Q^2 r_N^4}\right) h_2^A(\xi_N,Q^2)}_{\textbf{h-term}} + \underbrace{\left(\frac{12M_N^4 x_N^4}{Q^4 r_N^5}\right) g_2^A(\xi_N,Q^2)}_{\textbf{g-term}}$$

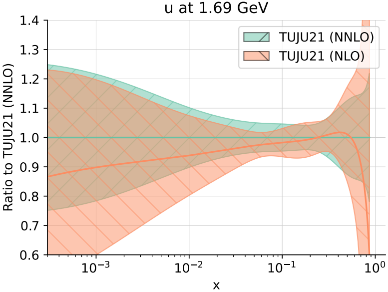


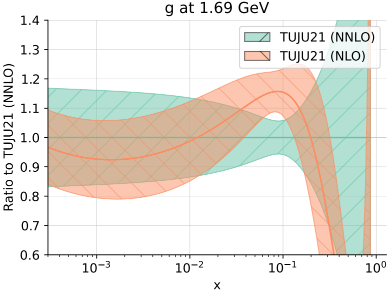
[Prog.Part.Nucl.Phys. 136 (2024) 104096]

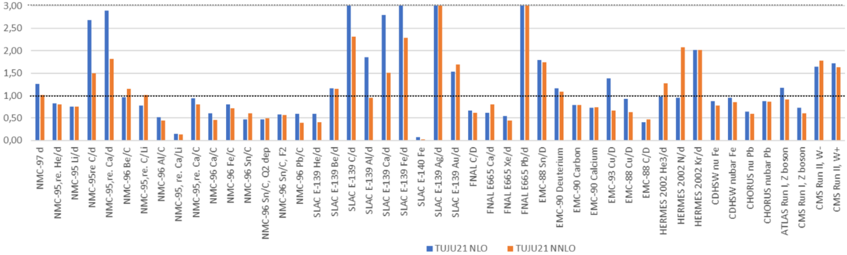
TMCs potentially large at large x, small Q, they are well caputered by leading TMCs

TMCs approximately independent from A, they can be parametrised and included in a fit

Impact of higher-order corrections





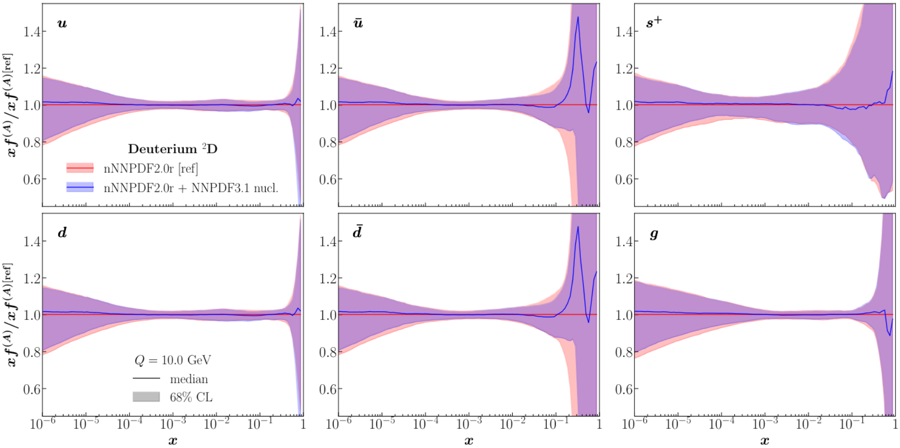


TUJU21: $\chi^2/N_{\rm dat} = 0.94$ (NLO) $\chi^2/N_{\rm dat} = 0.84$ (NNLO) $N_{\rm dat} = 2410$

[PRD 105 (2022) 094031]

Impact of proton baseline fit

What's the impact of varying the proton baseline fit in the nPDF determination?

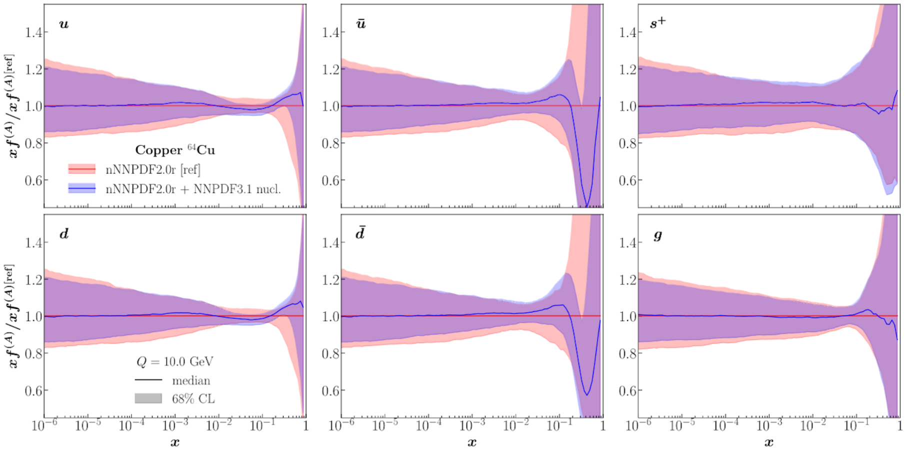


Case 1: remove the deuteron data from the proton baseline

(NMC, SLAC, BCDMS, E866, E906)

Impact of proton baseline fit

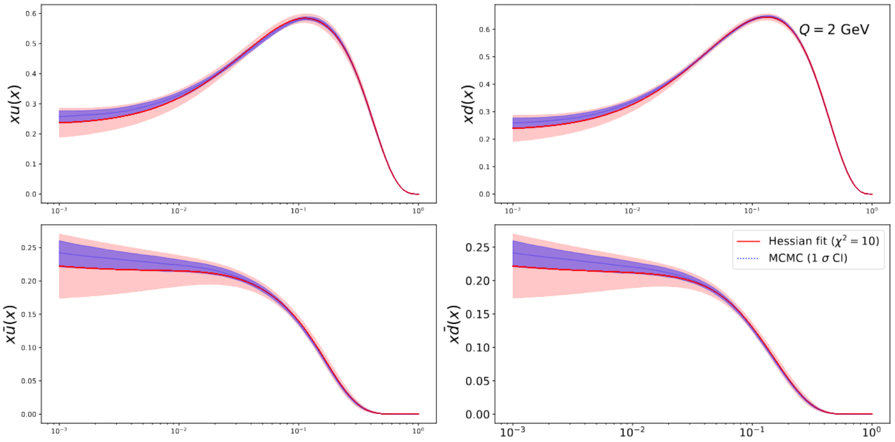
What's the impact of varying the proton baseline fit in the nPDF determination?



Case 2: remove the heavy nuclei data from the proton baseline

(CHORUS, NuTeV, E605)

Impact of the methodology: MCMC



[POSDIS2024 (2025) 058]

Markov Chain Monte Carlo to sample the probability distribution of nPDF parameters

The proposal distribution is a multivariate Gaussian with self-learned covariance

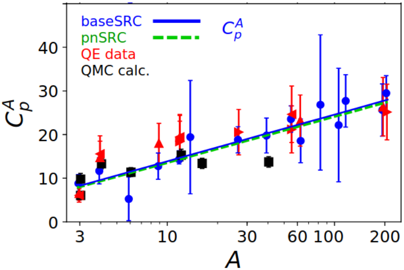
A uniform prior is used

Impact of the fitting paradigm

Split the nuclear PDF into two contributions:
one from mean-field interaction and one from short range correlated nucleon pairs

$$f_i^{(A,Z)} = \frac{Z}{A} f_i^{p,A} + \frac{N}{A} f_i^{n,A}$$

$$f_i^{(A,Z)} \frac{Z}{A} \left[(1 - C_p^A) f_i^p + C_p^A f_i^{\mathrm{SRC}p} \right] + \frac{N}{A} \left[(1 - C_n^A) f_i^n + C_n^A f_i^{\mathrm{SRC}n} \right]$$



χ^2/N_{data}	DIS	DY	W/Z	JLab	χ^2_{tot}	$\frac{\chi^2_{\mathrm{tot}}}{N_{\mathrm{DOF}}}$
traditional	0.85	0.97	0.88	0.72	1408	0.85
baseSRC	0.84	0.75	1.11	0.41	1300	0.80
pnSRC	0.85	0.84	1.14	0.49	1350	0.82
N_{data}	1136	92	120	336	1684	

baseSRC C_p^A, C_n^A independent
pnSRC $C_p^A = N/Z C_n^A$

Description of the data similar to or better than the traditional approach

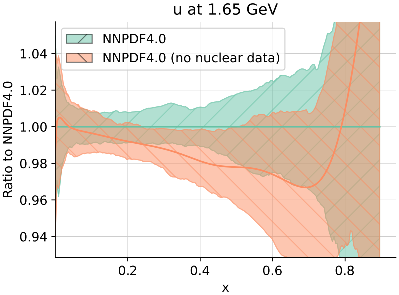
Fractions of SRC pairs agree with extractions from the low-energy quasielastic data

Potential to link high-energy partonic properties and lower-energy nuclear physics

[PRL 133 (2024) 152502; see also F. Olness, WG1 Wed. afternoon]

2. Challenges and prospects in nuclear PDF fits

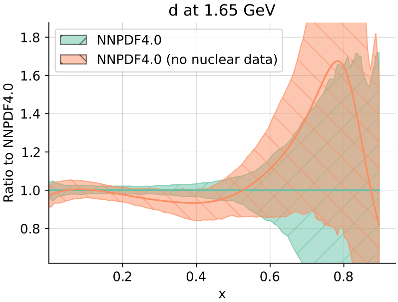
Challenge: nuclear data in proton PDF fits

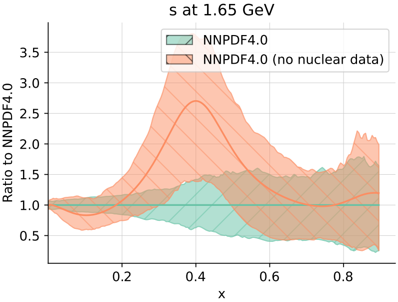


Nuclear data sets have a significant impact on proton PDFs

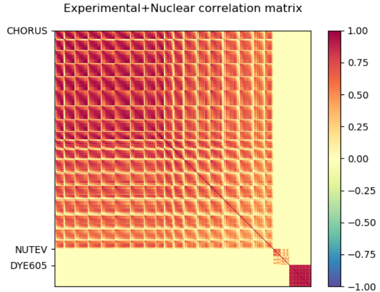
This despite nuclear data sets are a (limited) subset of the global data set (in NNDPF4.0, 1417 out of 4618 data points)

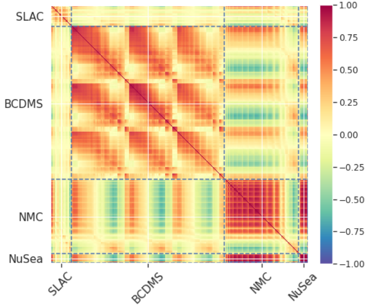
on deuteron (NC DIS): SLAC, BCDMS, NMC
on heavy nuclei (CC DIS): CHORUS, NuTeV
on both (FT DY): E866, E906, E605





Nuclear uncertainty in PDF determination

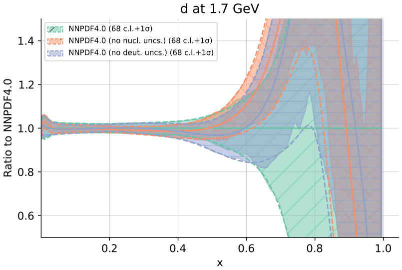




Effect of nuclear uncertainties relevant at large x
to reconcile FT DIS with LHC DY data

$\chi^2_{\rm tot} = 1.17 \to \chi^2_{\rm tot} = 1.26$ (no nucl. uncs.)
$\chi^2_{\rm LHCb} = 1.54 \to \chi^2_{\rm tot} = 1.76$ (no nucl. uncs.)

The bulk of the effect is due to nuclear uncertainties for heavy nuclei
deuteron uncertainties have a comparatively smaller effect at intermediate values of x



[EPJ C79 (2019) 282; EPJ C81 (2021) 37]

Prospect: *integrated* proton and nuclear PDF fits

Store PDF-independent predictions in interpolation grids

Idea not new (APPLgrid, fastNLO)

Interpolate PDFs with polynomials

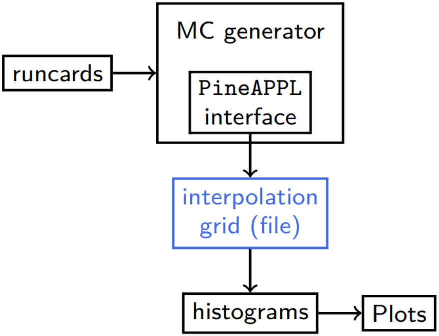
$$f = \sum_i f^i L_i$$

Convolution turns into sums

$$\frac{d\sigma}{d\mathcal{O}} = \sum_f \int dx \frac{d\hat{\sigma}}{d\mathcal{O}} = \sum_{f,i} f^i G_i$$

with interpolation grid

$$G_i = \int dx L_i \hat{\sigma}$$



PineAPPL originally developed to handle QCDxEW corrections [JHEP 12 (2020) 108]

Now interfaced to a wide range of MC generators (MadGraph, Matrix, MCFM, NNLOjet, SHERPA)

Increasingly used in a variety of HEP projects (NNPDF, xFitter, nCTEQ, ...)

More and more PineAPPL grids are being uploaded to Ploughshare

PineAPPLv1 [T. Jezo, E.R. Nocera, T. Rabemananjara, C. Schwan, T. Sharma, J. Wissmann, in preparation]

Supports an arbitrary number of convolutions

exclusive particle production in proton-proton and proton-nucleus collisions (4 convolutions)

Prospect: *integrated* proton and nuclear PDF fits

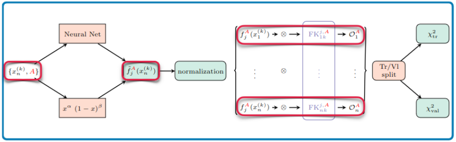
Idea: deploy the NNPDF4.0 methodology
to a simultaneous determination of proton and nuclear PDFs

Proton PDF ($A = 1$ boundary) automatically reflected in nuclear PDFs

Nuclear DIS data contribute consistently to both proton and nuclear PDFs

Take advantage of automated generation of theoretical observables [arXiv: 2302.12124]

Playground to optimise GPU parallelisation (Netherlands eScience Centre)



Methodology Improvements NNPDF4.0

- **Stochastic Gradient Descent** (SGD) for NN training using Tensorflow (easily changeable)
- **Automated Optimisation of Hyperparameters** (scanning of the parameter space)
- Methodology Validation using **Closure Tests**, **Future Tests**, and **Parametrisation Basis Independence**
- Proton, Deuteron, and heavy Nuclear PDFs are **fitted simultaneously** in a Self-Consistent Way ⇔ **Fitting Smoothly the A-dependence**

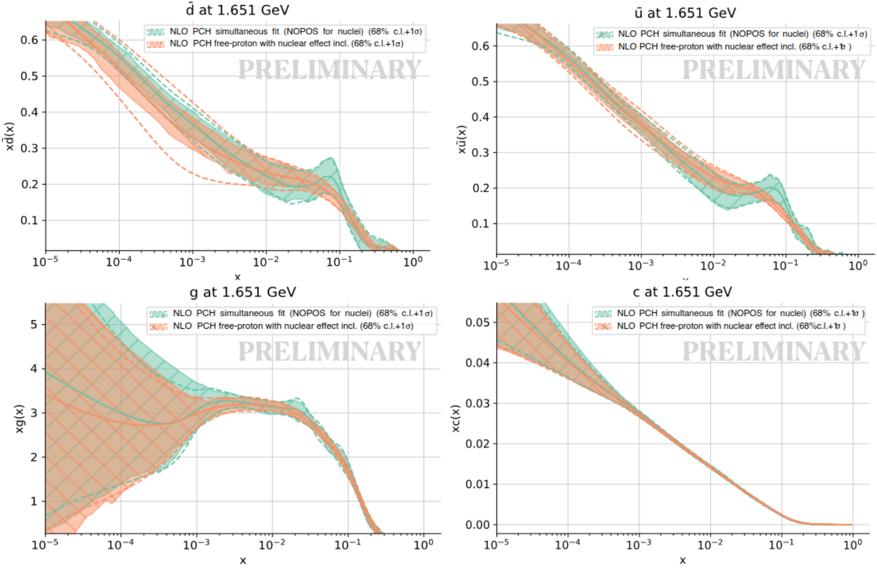
$$\chi^2_{\rm tot} = \sum_A \chi^2_{t_0}(A)$$

Dataset	NNPDF4.0	Integrated
NMC	1.63	1.62
HERACOMB	1.18	1.16
CDF	1.29	1.24
ATLAS	1.57	1.38
CMS	1.44	1.53
LHCb	1.54	1.46

Dataset	nNNPDF3.0	Integrated
NMC	0.88	1.61
SLAC	1.09	1.38
EMC	0.78	1.80
FNAL	1.01	0.50
BCDMS	0.82	1.35
LHC	1.09	1.15

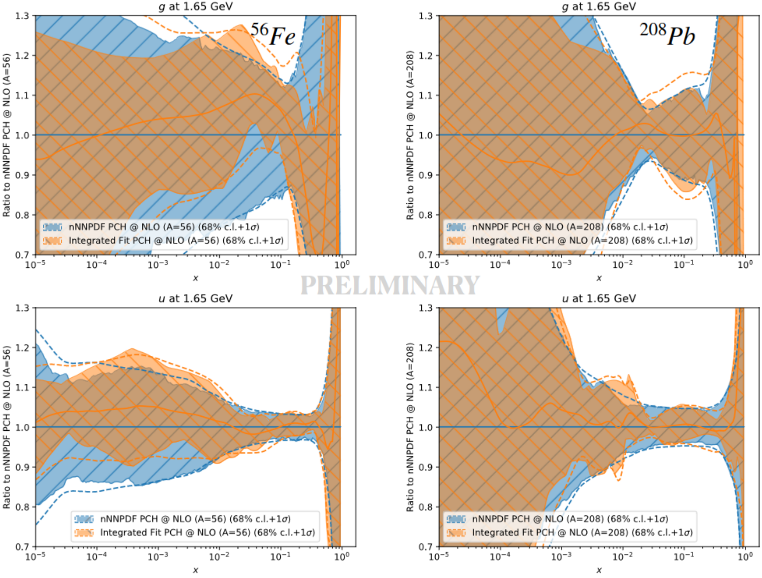
[NNPDF, in preparation; slide by courtesy of T.R. Rabemananjara]

Prospect: *integrated* proton and nuclear PDF fits



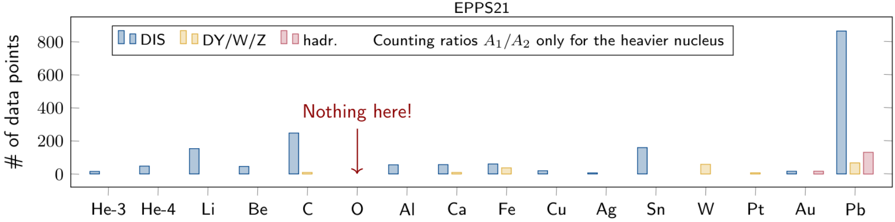
$A = 1$: generally good agreement, although further optimisation is still needed

Prospect: *integrated* proton and nuclear PDF fits



$A > 1$: generally good agreement, although better for Lead than for Iron

Challenge: nuclear data sets are somewhat limited



About 50% of the data are for Pb

Good coverage of DIS measurements for different A (but only fixed-target)
DY data on nuclear targets are more scarce, but the A coverage is fair
Hadronic observables are available only for heavy nuclei

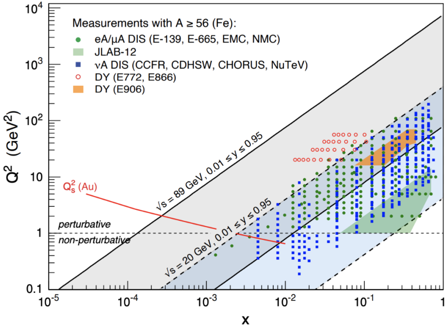
For example, nPDFs are a major source of uncertainty in testing small-system energy loss with OO [PRL 126 (2021) 192301; PRD 105 (2022) 074040]

Light-ion runs at the LHC could:
complement other light-nuclei DY data with W and Z production (strangeness)
give first direct constraint (dijets, D-mesons) on light-nuclei (small-x) distributions

Example: pO dijet production at 9.9 TeV (CMS), however, there are two problems
absolute cross sections very sensisitve to the used free-proton PDF
a pp reference is not expected at 9.9 TeV, therefore no cancellations occur

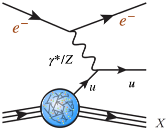
Prospect: two new experimental facilities are on sight

The Electron-Ion Collider (EIC)



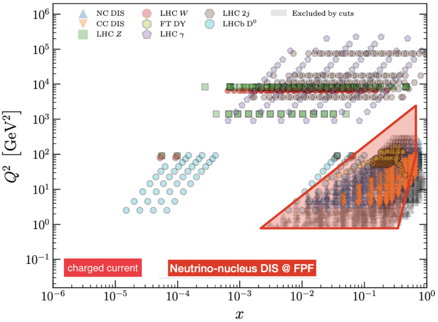
NC $e^- N$ DIS

in a wide kinematic region



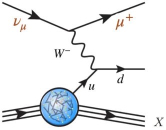
variety of nuclear targets (He, Cu, Au)

The Forward Physics Facility (FPF)



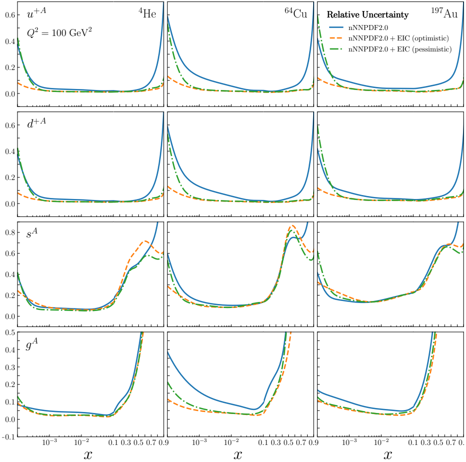
CC νN DIS

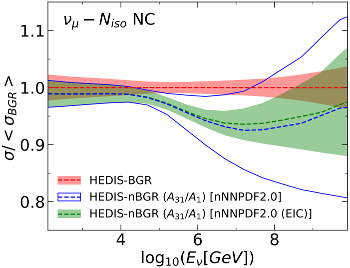
in a restricted kinematic region

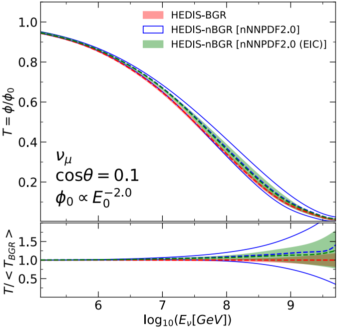


single nuclear target envisioned (W)

Nuclear PDFs at the EIC (nNNPDF2.0)







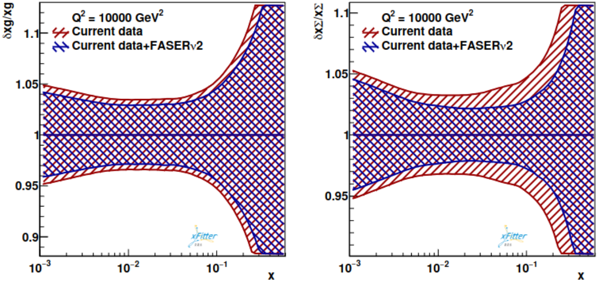
Impact from pseudodata [PRD 103 (2021) 096005]

e^-N CC DIS, N=^{4}He, ^{12}C, ^{40}Ca, ^{64}Cu, ^{197}Au

$E_\ell \times E_p$ [GeV]: 18× 100; 10× 110; 5× 41

$\mathcal{L} = 10$ fb^{-1}; $\sigma_u = 1.5/2.3\%$; $\sigma_c = 2.5/4.3\%$

Nuclear PDFs at the FPF (EPPS21)



Impact from pseudodata

$\nu/\bar{\nu}$N CC DIS, N=^{74}W

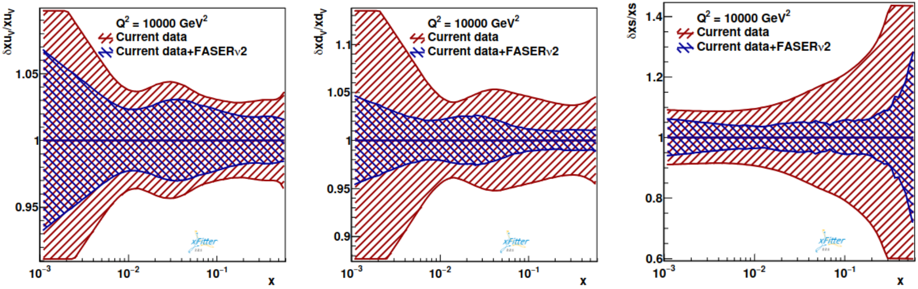
Statistical uncertainties only

FASERν setup

$E_\ell > 100$ GeV ($\delta E_\ell = 30\%$)

$\tan\theta < 0.5$ ($\delta\theta = 1$ mrad)

Hessian profiling of EPPS21



[EPJ C84 (2024) 369]

3. Conclusions

Summary

Ample progress in incorporating new data in global nPDF fits
LHCb pPb/pp D^0-meson production data puts unprecedented constraints on the gluon nPDF
which shows evidence of shadowing and antishadowing

Work in progress towards more global NNLO fits
this requires developments in Monte Carlo generators to handle asymmetric pN collisions

Ongoing work to understand correlations between proton and nuclear PDF analyses
promising integrated QCD fits are ongoing

The future will possibly bring a wealth of precise new measurements
LHC Run III, sPHENIX, SMOG@LHCb, FoCal@ALICE, EIC, FPF

Summary

Ample progress in incorporating new data in global nPDF fits
LHCb pPb/pp D^0-meson production data puts unprecedented constraints on the gluon nPDF which shows evidence of shadowing and antishadowing

Work in progress towards more global NNLO fits
this requires developments in Monte Carlo generators to handle asymmetric pN collisions

Ongoing work to understand correlations between proton and nuclear PDF analyses
promising integrated QCD fits are ongoing

The future will possibly bring a wealth of precise new measurements
LHC Run III, sPHENIX, SMOG@LHCb, FoCal@ALICE, EIC, FPF

Thank you