CHARM IN THE PROTON

STEFANO FORTE
UNIVERSITÀ DI MILANO & INFN

UNIVERSITÀ DEGLI STUDI DI MILANO
DIPARTIMENTO DI FISICA

Istituto Nazionale di Fisica Nucleare

LHCb ITALY MEETING

MILANO, FEBRUARY 1, 2023

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 740006

CHARM IN THE PROTON

THE NNPDF COLLABORATION

RICHARD D. BALL, ALESSANDRO CANDIDO, JUAN CRUZ MARTINEZ
STEFANO FORTE, TOMMASO GIANI, FELIX HEKHORN,
KIRILL KUDASHKIN, GIACOMO MAGNI AND JUAN ROJO

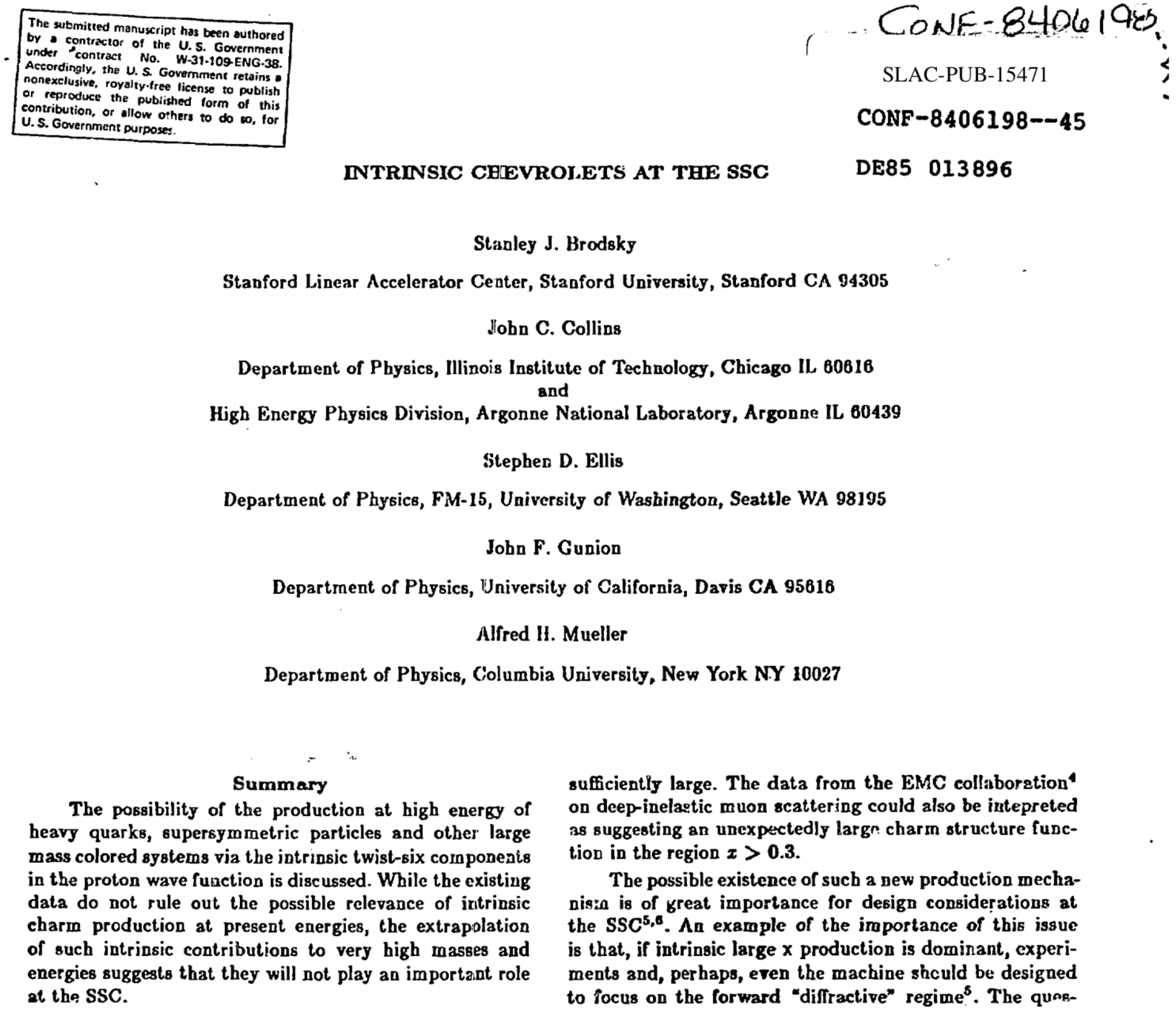
AMSTERDAM-EDINBURGH-INFN-MILAN-NIKHEF

LHCB ITALY MEETING

MILANO, FEBRURARY 1, 2023

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 740006

“INTRINSIC” CONSTITUENTS IN THE PROTON AT THE SSC (1984)



CONF-8406198S

SLAC-PUB-15471

CONF-8406198--45

DE85 013896

INTRINSIC CHEVROLETS AT THE SSC

Stanley J. Brodsky

Stanford Linear Accelerator Center, Stanford University, Stanford CA 94305

John C. Collins

Department of Physics, Illinois Institute of Technology, Chicago IL 60616
and
High Energy Physics Division, Argonne National Laboratory, Argonne IL 60439

Stephen D. Ellis

Department of Physics, FM-15, University of Washington, Seattle WA 98195

John F. Gunion

Department of Physics, University of California, Davis CA 95616

Alfred H. Mueller

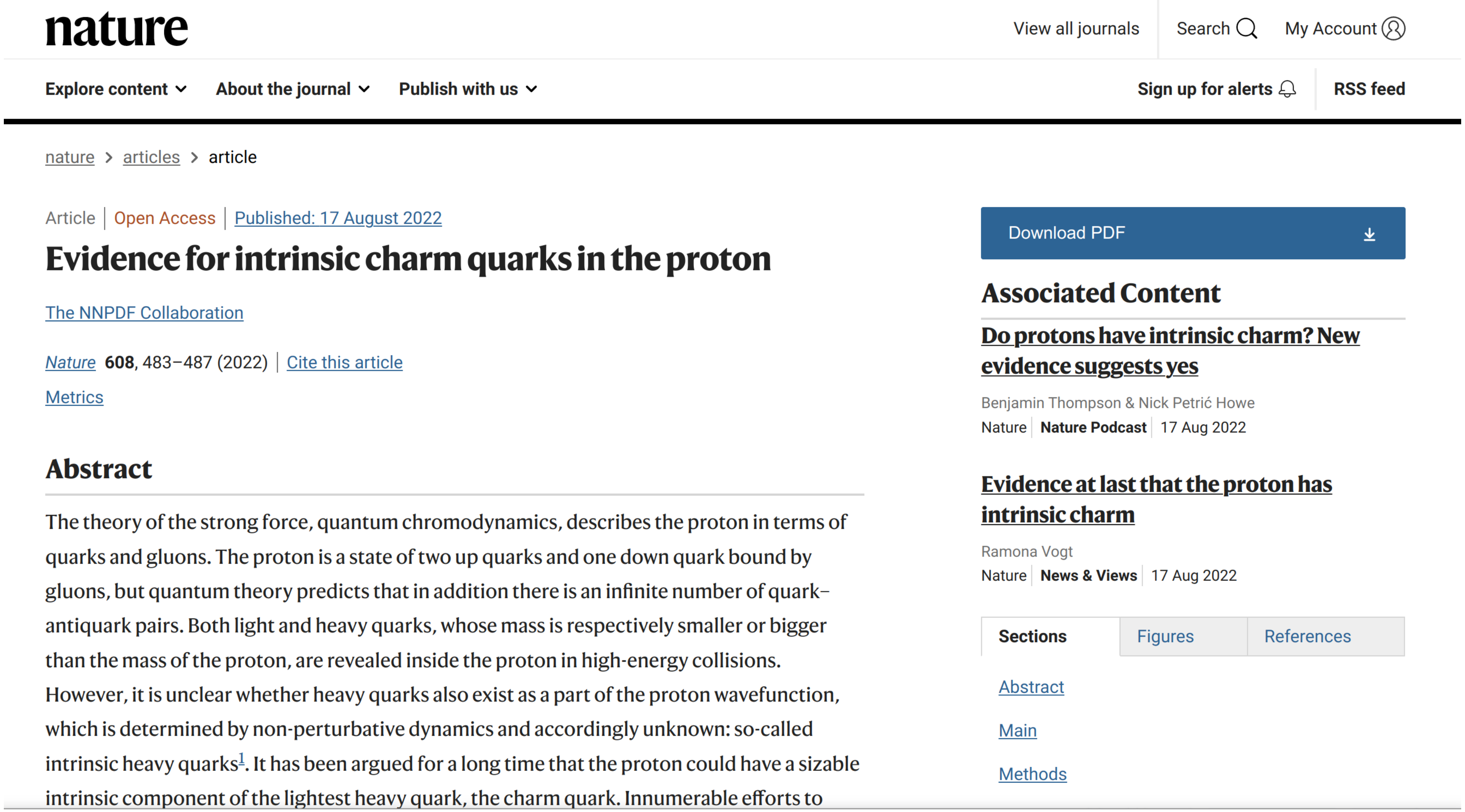
Department of Physics, Columbia University, New York NY 10027

Summary

The possibility of the production at high energy of heavy quarks, supersymmetric particles and other large mass colored systems via the intrinsic twist-six components in the proton wave function is discussed. While the existing data do not rule out the possible relevance of intrinsic charm production at present energies, the extrapolation of such intrinsic contributions to very high masses and energies suggests that they will not play an important role at the SSC.

sufficiently large. The data from the EMC collaboration[4] on deep-inelastic muon scattering could also be intepreted as suggesting an unexpectedly large charm structure function in the region $x > 0.3$.

The possible existence of such a new production mechanism is of great importance for design considerations at the SSC[5,6]. An example of the importance of this issue is that, if intrinsic large x production is dominant, experiments and, perhaps, even the machine should be designed to focus on the forward "diffractive" regime[5]. The ques-



Article | Open Access | Published: 17 August 2022

Evidence for intrinsic charm quarks in the proton

The NNPDF Collaboration

Nature **608**, 483–487 (2022) | Cite this article

Metrics

Associated Content

Do protons have intrinsic charm? New evidence suggests yes

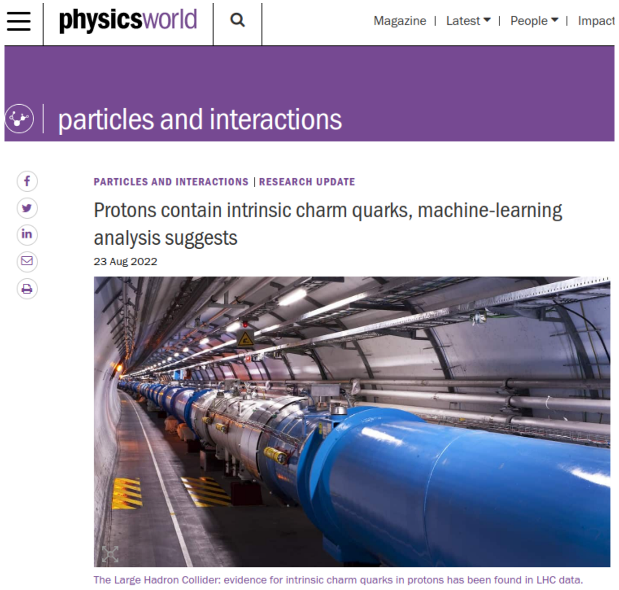
Benjamin Thompson & Nick Petrić Howe
Nature | **Nature Podcast** | 17 Aug 2022

Evidence at last that the proton has intrinsic charm

Ramona Vogt
Nature | **News & Views** | 17 Aug 2022

Abstract

The theory of the strong force, quantum chromodynamics, describes the proton in terms of quarks and gluons. The proton is a state of two up quarks and one down quark bound by gluons, but quantum theory predicts that in addition there is an infinite number of quark–antiquark pairs. Both light and heavy quarks, whose mass is respectively smaller or bigger than the mass of the proton, are revealed inside the proton in high-energy collisions. However, it is unclear whether heavy quarks also exist as a part of the proton wavefunction, which is determined by non-perturbative dynamics and accordingly unknown: so-called intrinsic heavy quarks[1]. It has been argued for a long time that the proton could have a sizable intrinsic component of the lightest heavy quark, the charm quark. Innumerable efforts to



PARTICLES AND INTERACTIONS | RESEARCH UPDATE

Protons contain intrinsic charm quarks, machine-learning analysis suggests

23 Aug 2022

The Large Hadron Collider: evidence for intrinsic charm quarks in protons has been found in LHC data.

Dienstag, 08. November 2022

Startseite / Forschung aktuell / Protonen mit Charm: Auch schwere Quarks finden sich in Kernbausteinen

Protonen mit Charm: Auch schwere Quarks finden sich in Kernbausteinen

Grotelüschen, Frank | 18. August 2022, 16:41 Uhr

NIEUWS

Proton bevat een wonderlijk extra deeltje: de 'charm quark'

Protonen, fundamentele bouwstenen van alle materie, blijken een nieuw ingrediënt te bevatten: de 'charm quark'. Natuurkundigen reageren opgetogen: 'Verbazingwekkend dat er nog iets nieuws valt te leren over een oude bekende als het proton.'

Frank Rensen 17 augustus 2022, 21:55

Tutti i prodotti sono selezionati in piena autonomia editoriale. Se acquisti uno di questi prodotti, potremmo ricevere una commissione.

VALENTINA GUGLIELMO SCIENZA 29.08.2022

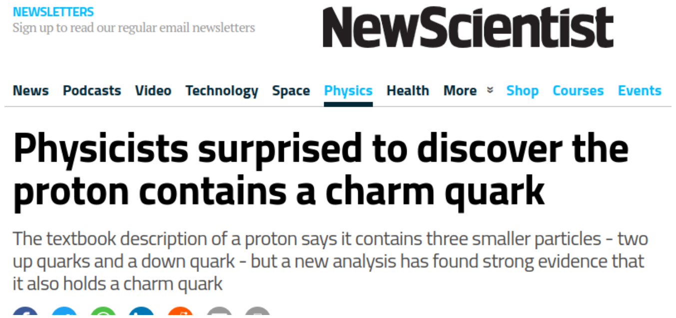
Un nuovo studio fa luce su una sorprendente caratteristica della struttura dei protoni

Pubblicato sulla rivista *Nature* spiega come anche i *quark charm*, insieme ai più noti e leggeri *quark up* e *quark down*, siano da annoverarsi tra i componenti intrinseci dei costituenti atomici, confermando un'ipotesi elaborata oltre 40 anni fa

Physique · 3 min de lecture

Une étude confirme que le proton possède un quark charm intrinsèque

Fleur Brosseau · 19 août 2022



Physicists surprised to discover the proton contains a charm quark

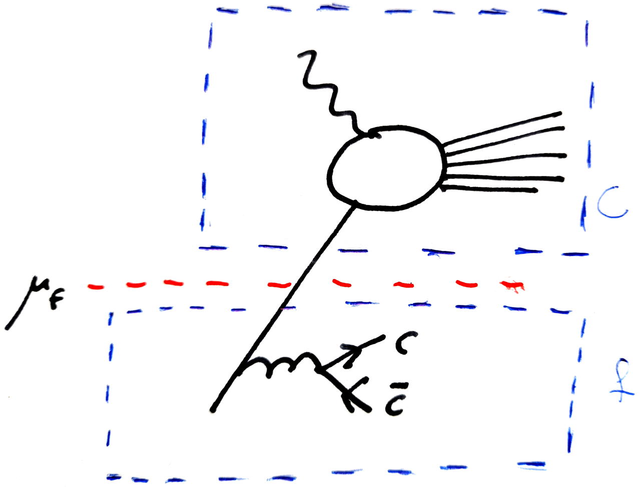
The textbook description of a proton says it contains three smaller particles - two up quarks and a down quark - but a new analysis has found strong evidence that it also holds a charm quark

WHAT IS
INTRINSIC CHARM?

WHAT IS NOT INTRINSIC CHARM?

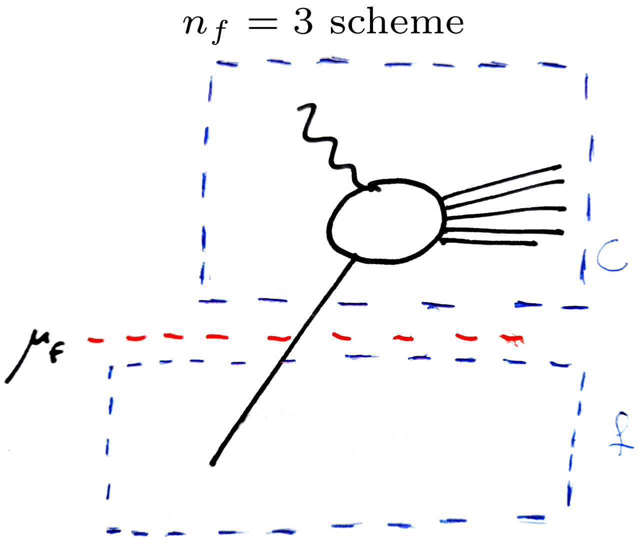
- FACTORIZED STRUCTURE FUNCTION: $F_2 = \sum_i C_i \otimes f_i$
- COLLINEAR PARTON RADIATION ⇒ MASS SINGULARITIES INCLUDED IN PDFs f_i UP TO μ_f
- CHARM PDF PERTURBATIVELY GENERATED BY QCD EVOLUTION

perturbative charm generation



- MASSIVE QUARKS ⇒ NO COLLINEAR SINGULARITY
- .
- .

WHAT IS NOT INTRINSIC CHARM?



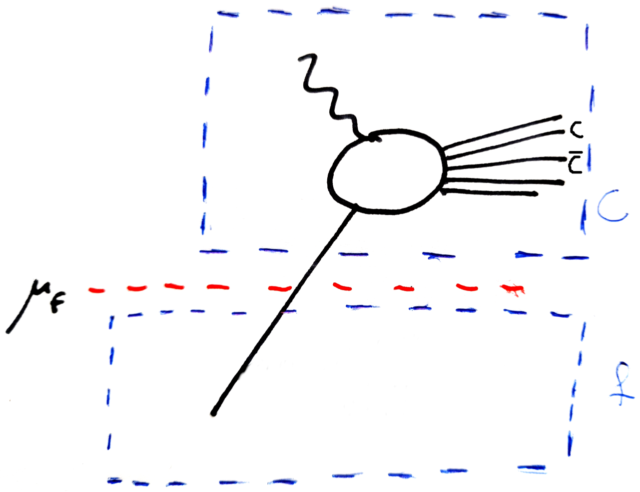
- FACTORIZED STRUCTURE FUNCTION: $F_2 = \sum_i C_i \otimes f_i$
- COLLINEAR PARTON RADIATION ⇒ MASS SINGULARITIES INCLUDED IN PDFS f_i UP TO μ_f
- CHARM PDF PERTURBATIVELY GENERATED BY QCD EVOLUTION

- MASSIVE QUARKS ⇒ NO COLLINEAR SINGULARITY
- MAY WORK IN $n_f = 3$ SCHEME ⇒ CHARM DECOUPLES FROM PDF EVOLUTION
- .

WHAT IS NOT INTRINSIC CHARM?

- FACTORIZED STRUCTURE FUNCTION: $F_2 = \sum_i C_i \otimes f_i$
- COLLINEAR PARTON RADIATION ⇒ MASS SINGULARITIES INCLUDED IN PDFS f_i UP TO μ_f
- CHARM PDF PERTURBATIVELY GENERATED BY QCD EVOLUTION

$n_f = 3$ scheme



- MASSIVE QUARKS ⇒ NO COLLINEAR SINGULARITY
- MAY WORK IN $n_f = 3$ SCHEME ⇒ CHARM DECOUPLES FROM PDF EVOLUTION
- PERTURBATIVE CHARM PRODUCTION IN COEFFICIENT FUNCTION

DECOUPLING

- DECOUPLING SCHEME ⇒ HEAVY FLAVOR GRAPHS
  SUBTRACTED AT ZERO MOMENTUM (Collins, Wilczek, Zee, 1978)
- $N_f = 3$ ACTIVE FLAVORS IN β FUNCTION & EVOLUTION EQUATIONS
- DECOUPLING VS $\overline{\text{MS}}$ ⇔ DIFFERENT RENORMALIZATION & FACTORIZATION SCHEMES

MATCHING

- PDFS, α_s IN $N_f = 3$ & $N_f = 4$
  RELATED BY MATCHING CONDITIONS
- DETERMINED BY COMPUTING

  OPERATOR MATRIX ELEMENTS

  IN EITHER SCHEME AND EQUATING:

  NNLO (Buza, et al., 1998),

  N^3LO (Ablinger, Blümlein et al, 2009-2022)

OME CONTRIBUTING
TO THE CHARM PDF
SOLID ⇒ HEAVY; DASHED ⇒ LIGHT

M. Buza et al.: Charm

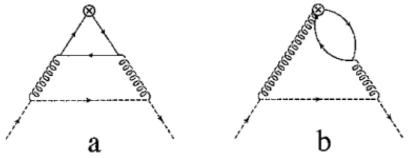


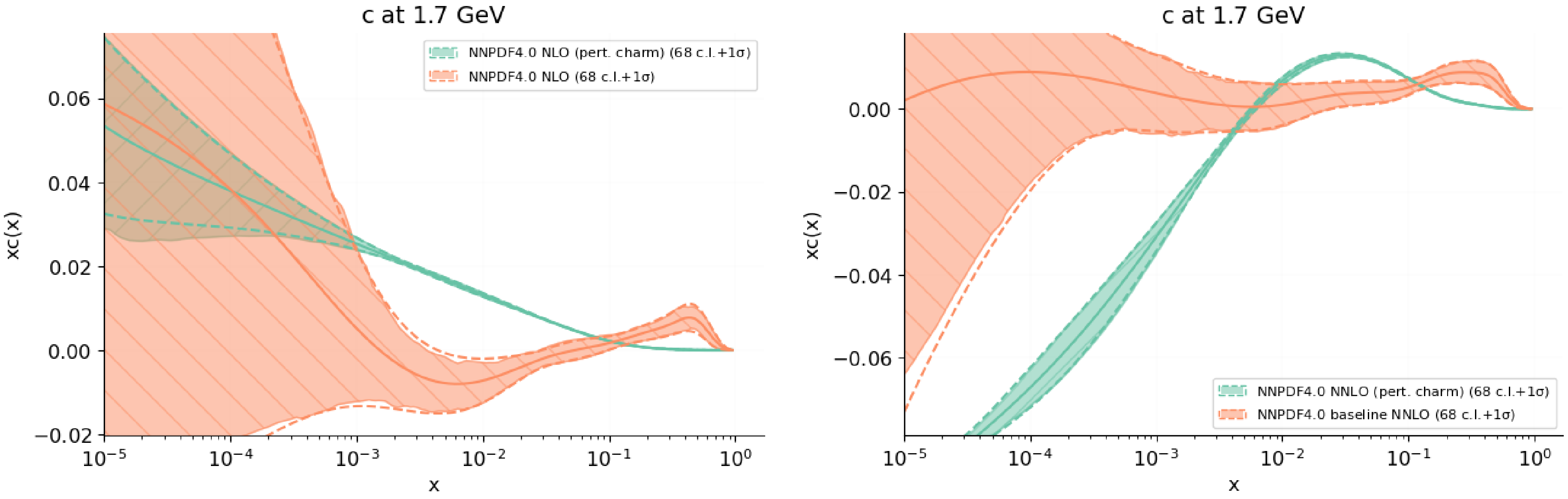
Fig. 2. $O(\alpha_s^2)$ contributions to the purely-singlet OME $A_{q'q}^{\mathrm{PS}}$. Here q and q' are represented by the *dashed* and *solid lines* respectively. In the case of $q' = H$ these graphs contribute to the heavy-quark OME A_{Hq}^{PS}

WHAT IS NOT INTRINSIC CHARM?

PERTURBATIVE CHARM

- IN $N_f = 3$ SCHEME CHARM PDF VANISHES
- IN $N_f = 4$ SCHEME, CHARM DETERMINED BY PERTURBATIVE MATCHING
- STARTING AT NNLO (TWO LOOPS) DOES NOT VANISH AT ANY SCALE
- .

PERTURBATIVE CHARM PDF, $n_f = 4$ SCHEME, Q=1.7 GEV



NLO

NNLO

WHAT IS NOT INTRINSIC CHARM?

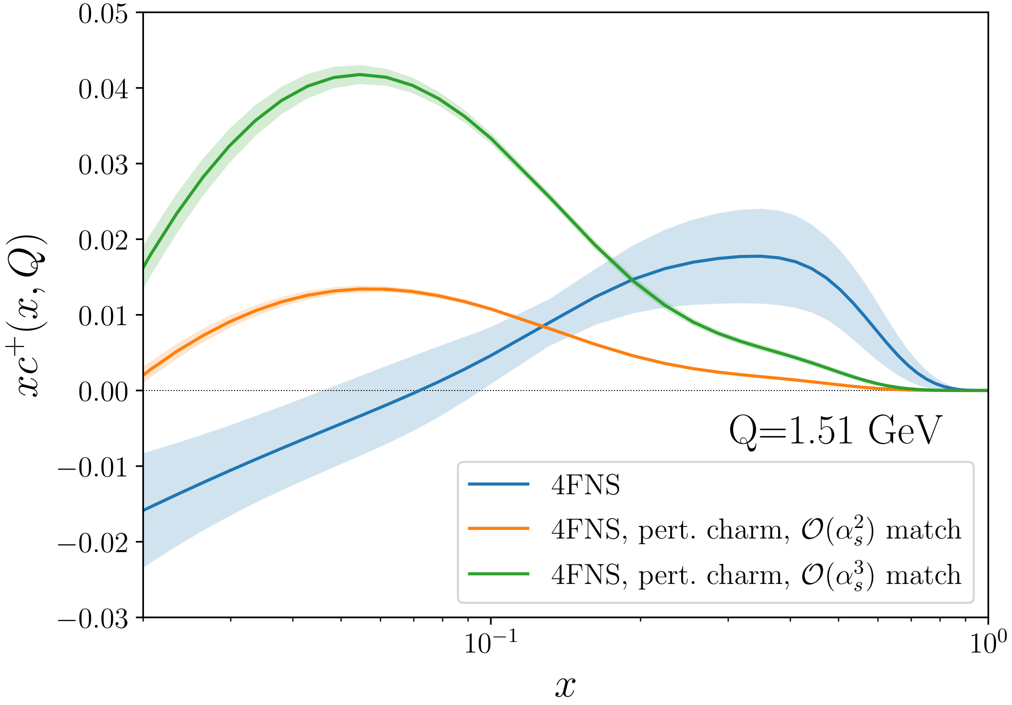
PERTURBATIVE CHARM

- IN $N_f = 3$ SCHEME CHARM PDF VANISHES
- IN $N_f = 4$ SCHEME, CHARM DETERMINED BY PERTURBATIVE MATCHING
- STARTING AT NNLO (TWO LOOPS) DOES NOT VANISH AT ANY SCALE
- SLOW PERTURBATIVE CONVERGENCE ⇒ NNLO AND N^{3}LO DIFFER SIGNIFICANTLY

PERTURBATIVE CHARM PDF, $n_f = 4$ SCHEME, Q=1.7 GEV

NNLO VS. N^{3}LO

(NLO IDENTICALLY ZERO)



WHAT IS INTRINSIC CHARM?

- DEFINE CHARM PDF AS OME:

$$\langle p|\bar{c}\gamma^{\mu_1}D^{\mu_2}\ldots D^{\mu_n}c|p\rangle = A_c^n p^{\mu_1}\ldots p^{\mu_n} - \text{traces}$$

$$A_c^n = \int_0^1 dx\, x^{n-1} c(x)$$

- DO NOT FACTOR CHARM MASS SINGULARITIES INTO OME
- ⇒ CHOOSE $n_f = 3$ SCHEME
- CHARM PDF PURELY INTRINSIC, SCALE-INDEPENDENT

INTRINSIC CHARM IS CHARM IN THE $N_F = 3$ (DECOUPLING) SCHEME

HOW CAN ONE MEASURE INTRINSIC CHARM?

(ALMOST) LIKE ANY OTHER PDF

- DETERMINE PDFS FROM DATA
- GO TO $n_f = 3$ SCHEME
- LOOK AT RESULT

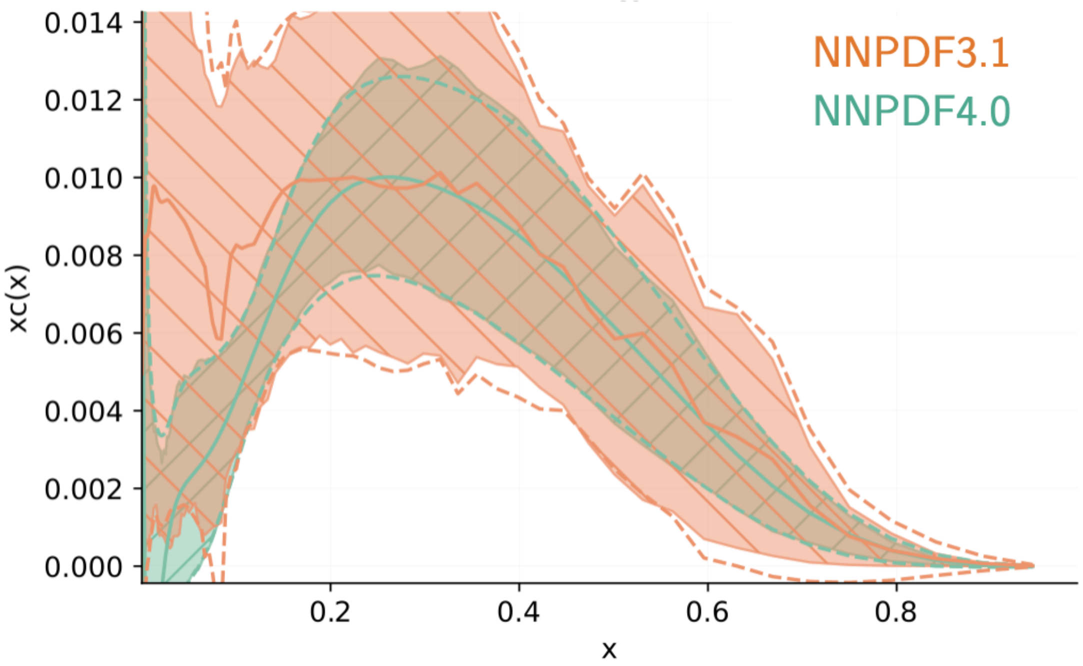
TWO POSSIBILITIES

- PARAMETRIZE PDFS IN $n_f = 3$ SCHEME AND MATCH UP FOR FITTING
- PARAMETRIZE PDFS IN $n_f = 4$ SCHEME AND MATCH DOWN FOR DETERMINING IC
- REMEMBER: MATCHING PERTURBATIVELY UNSTABLE ⇒ LARGE MHOU

DISCOVERY OF INTRINSIC CHARM

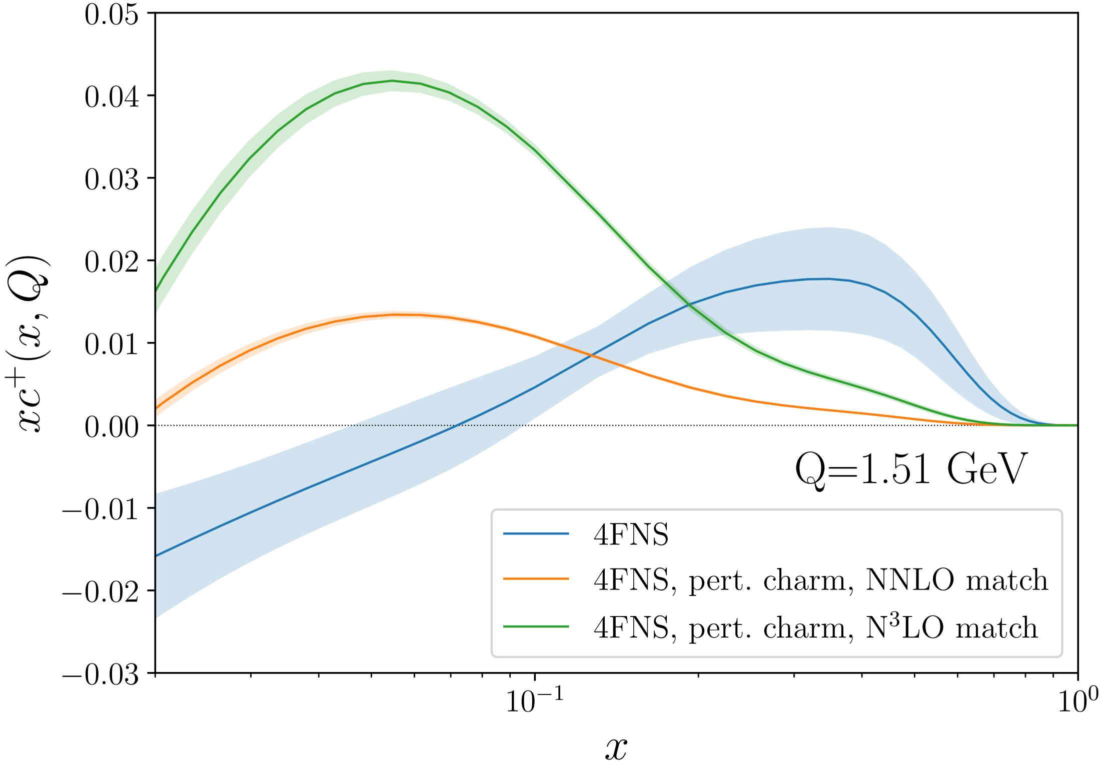
THE CHARM PDF ($n_f = 4$ SCHEME) NNPDF4.0 VS. NNPDF3.1

- NNPDF CHARM PDF $\Rightarrow$ DETERMINED FROM THE DATA ALONG WITH ALL OTHER PDFS:
  - MORE REALISTIC UNCERTAINTIES
  - INDEPENDENT OF MATCHING CONDITIONS:
    - STABLE UPON VARIATION OF m_c
    - INSENSITIVE TO MHO CORRECTIONS
- NNPDF4.0 AND NNPDF3.1 FULL AGREEMENT
- NNPDF4.0 SIGNIFICANTLY SMALLER UNCERTAINTIES THANKS TO ML METHDOLOGY



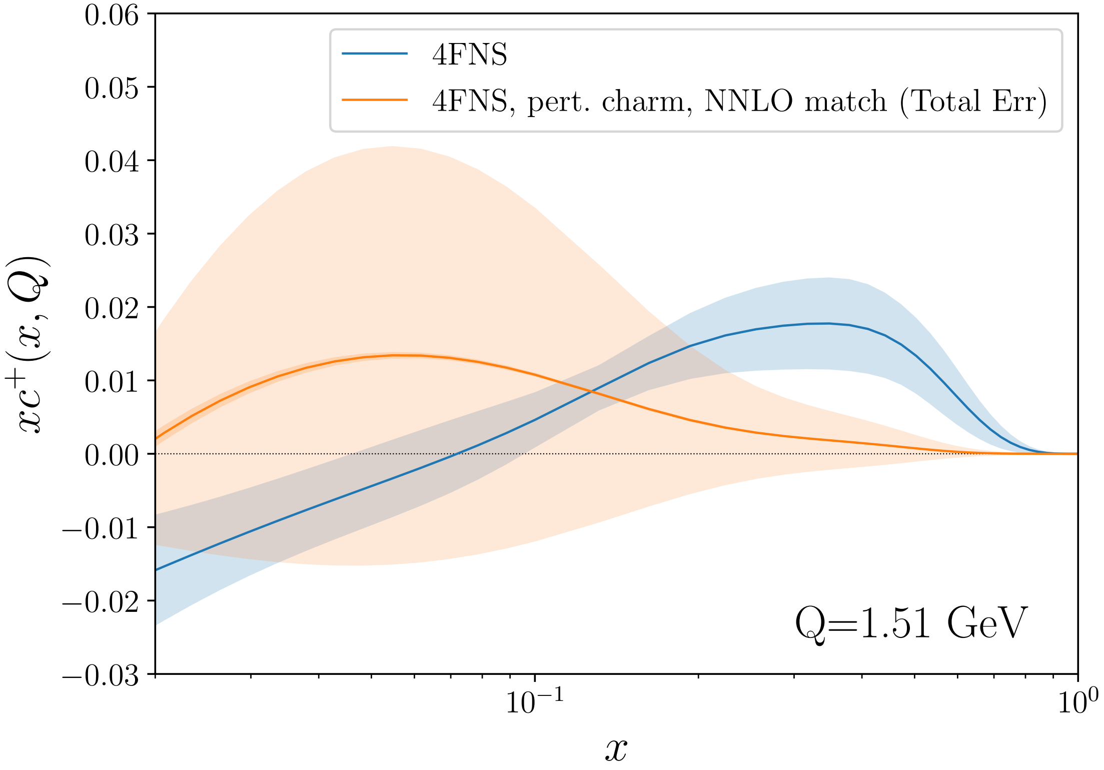
THE CHARM PDF ($n_f = 4$ SCHEME) FITTED VS PERTURBATIVE

- NNPDF CHARM PDF ⇒ DETERMINED FROM THE DATA ALONG WITH ALL OTHER PDFS:
  - MORE REALISTIC UNCERTAINTIES
  - INDEPENDENT OF MATCHING CONDITIONS:
    - STABLE UPON VARIATION OF m_c
    - INSENSITIVE TO MHO CORRECTIONS
- NNPDF4.0 RESULT DIFFERS SIGNIFICANTLY FROM PERTURBATIVE CHARM



THE CHARM PDF ($n_f = 4$ SCHEME) FITTED VS PERTURBATIVE

- NNPDF CHARM PDF ⇒ DETERMINED FROM THE DATA ALONG WITH ALL OTHER PDFS:
  - MORE REALISTIC UNCERTAINTIES
  - INDEPENDENT OF MATCHING CONDITIONS:
    - STABLE UPON VARIATION OF m_c
    - INSENSITIVE TO MHO CORRECTIONS
- RESULT DIFFERS SIGNIFICANTLY FROM PERTURBATIVE CHARM EVEN ACCOUNTING FOR MHOU
- INTRINSIC-LIKE, VALENCE-LIKE BUMP AT LARGE x



IS IT INTRINSIC CHARM?

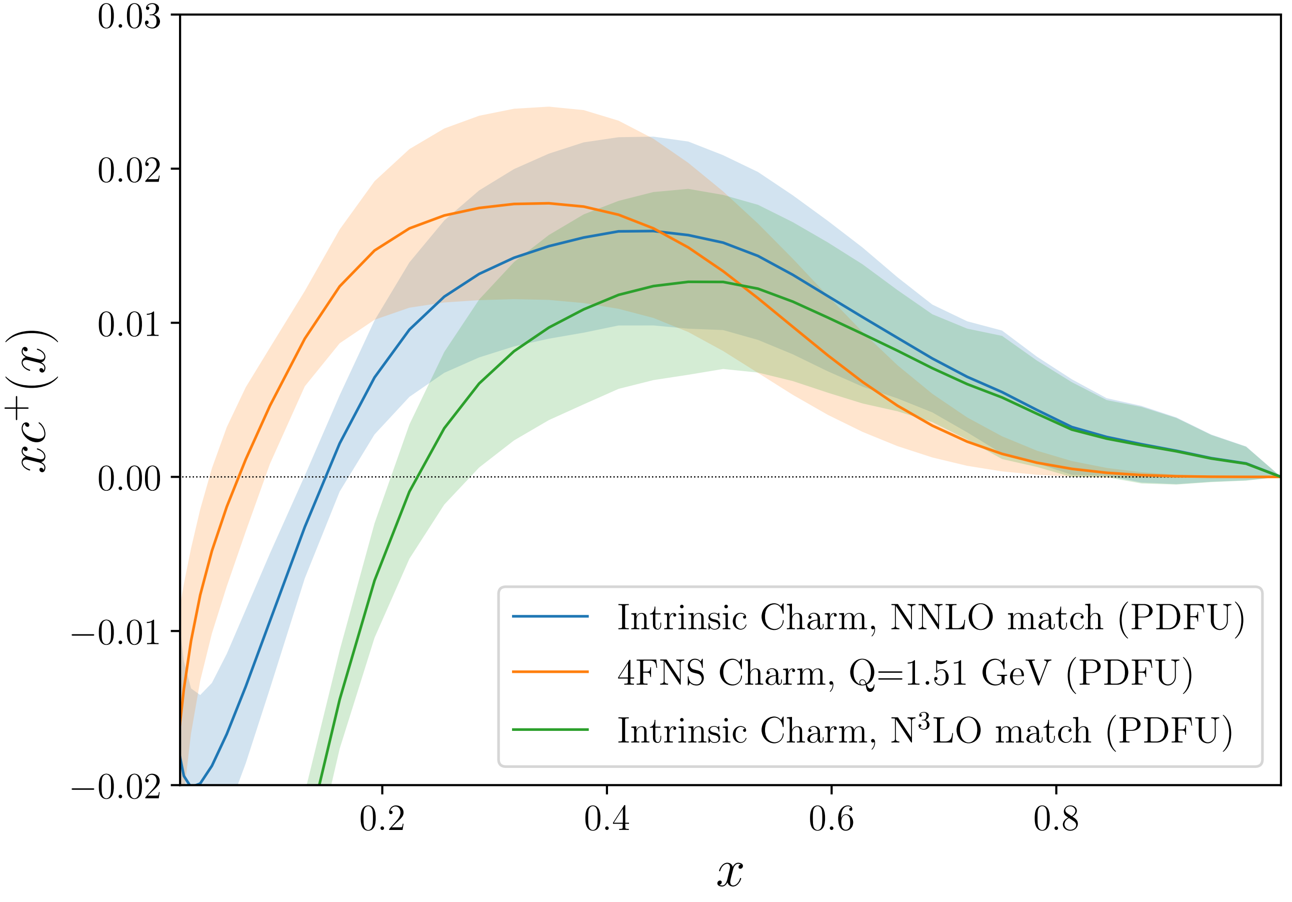
RECALL LARGE & PERTURBATIVELY UNSTABLE CONVERSION TO $N_f = 3$ SCHEME

INTRINSIC CHARM....

THE EKO CODE (Candido, Hekhorn, Magni, 2022)

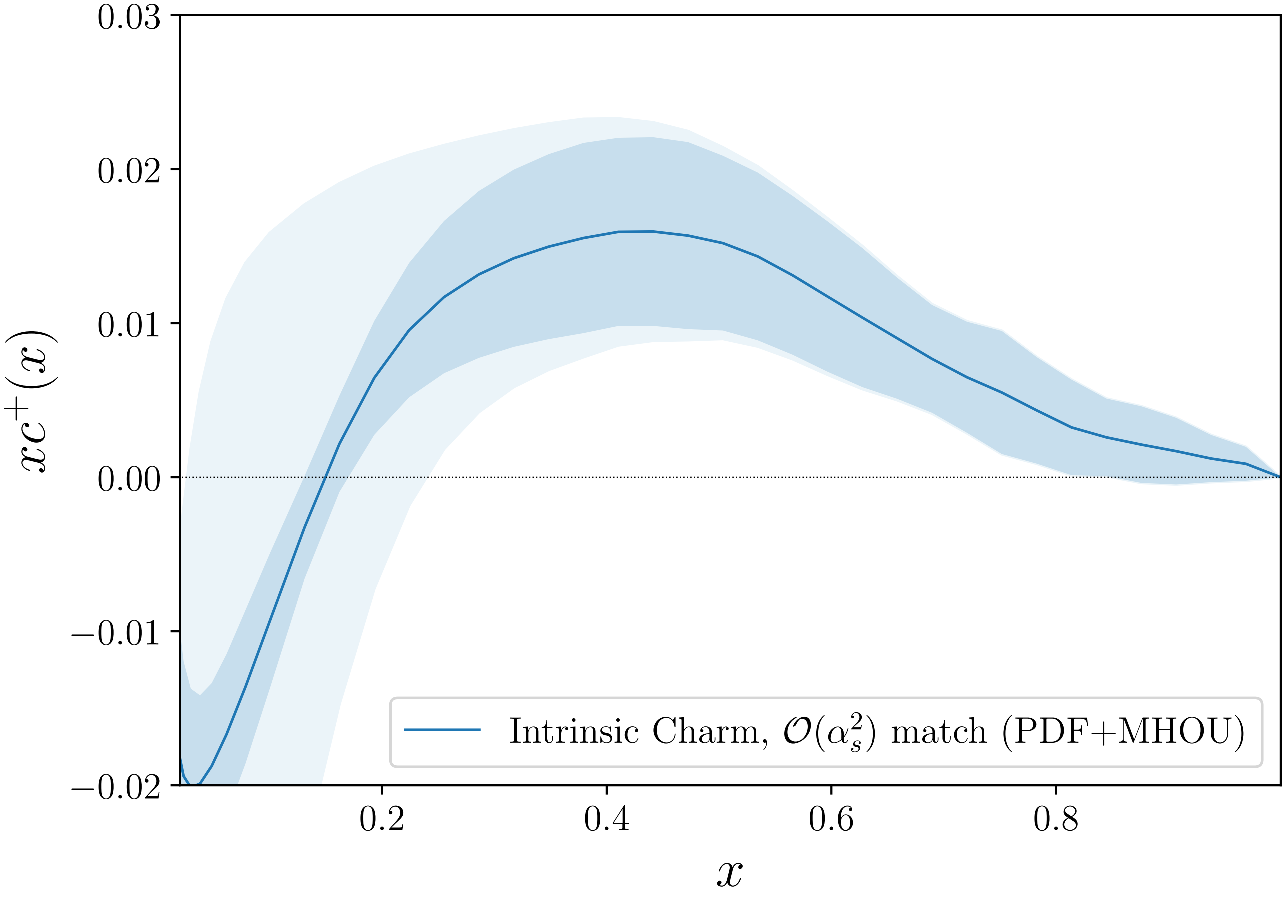
- IMPLEMENTS DIRECT & INVERSE EVOLUTION & MATCHING
- N^3LO MATCHING (Blümlein, Ablinger et al.) ALSO IMPLEMENTED

CHARM PDF: $N_f = 4$ VS $N_f = 3$ (NNLO & N^3LO CONVERSION)



INTRINSIC CHARM!

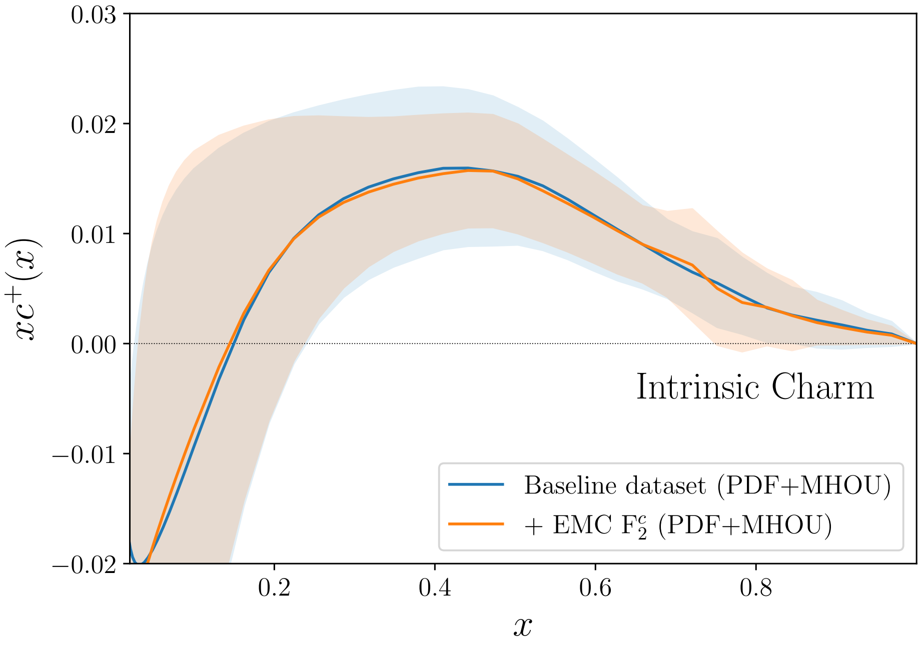
- MHOU ESTIMATED FROM N^3LO-NNLO DIFFERENCE
  - LARGE UNCERTAINTY AT SMALL x
  - NEGLIGIBLE UNCERTAINTY IN VALENCE REGION
- COMPATIBLE WITH ZERO AT SMALL x
- CLEAR EVIDENCE FOR INTRINSIC VALENCE PEAK



MORE DATA
EMC 1983

- DIRECT MEASUREMENT OF THE CHARM STRUCTURE FUNCTION F_2^c
- EVIDENCE FOR INTRINSIC CHARM CLAIMED, BUT EXPERIMENT DISPUTED
- NOT INCLUDED IN DEFAULT NNPDF4.0

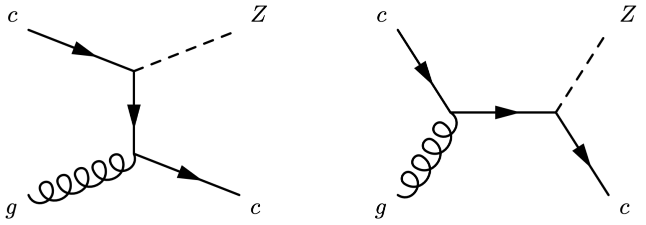
INTRINSIC CHARM WITH EMC DATA INCLUDED



COMPLETE CONSISTENCY!

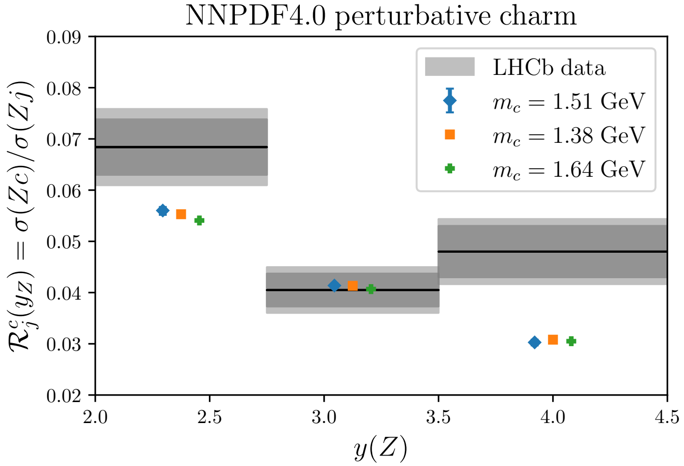
MORE DATA
LHCb 2021

MEASUREMENT OF Z+CHARM PRODUCTION

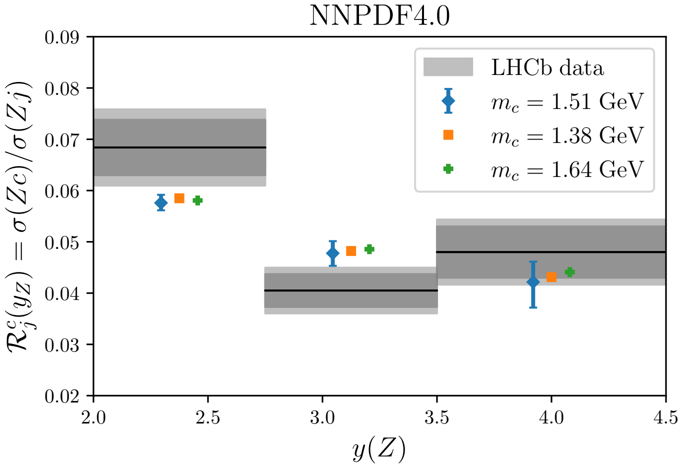


DATA VS THEORY PREDICTION

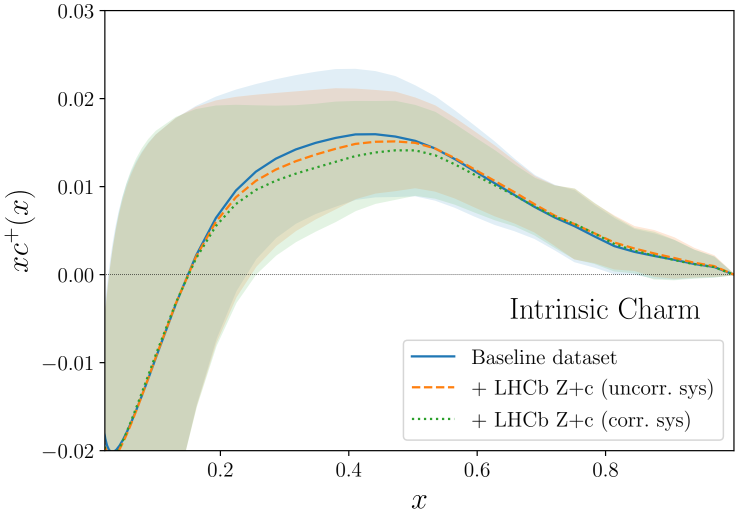
NO INTRINSIC CHARM



NNPDF4.0 INTRINSIC CHARM

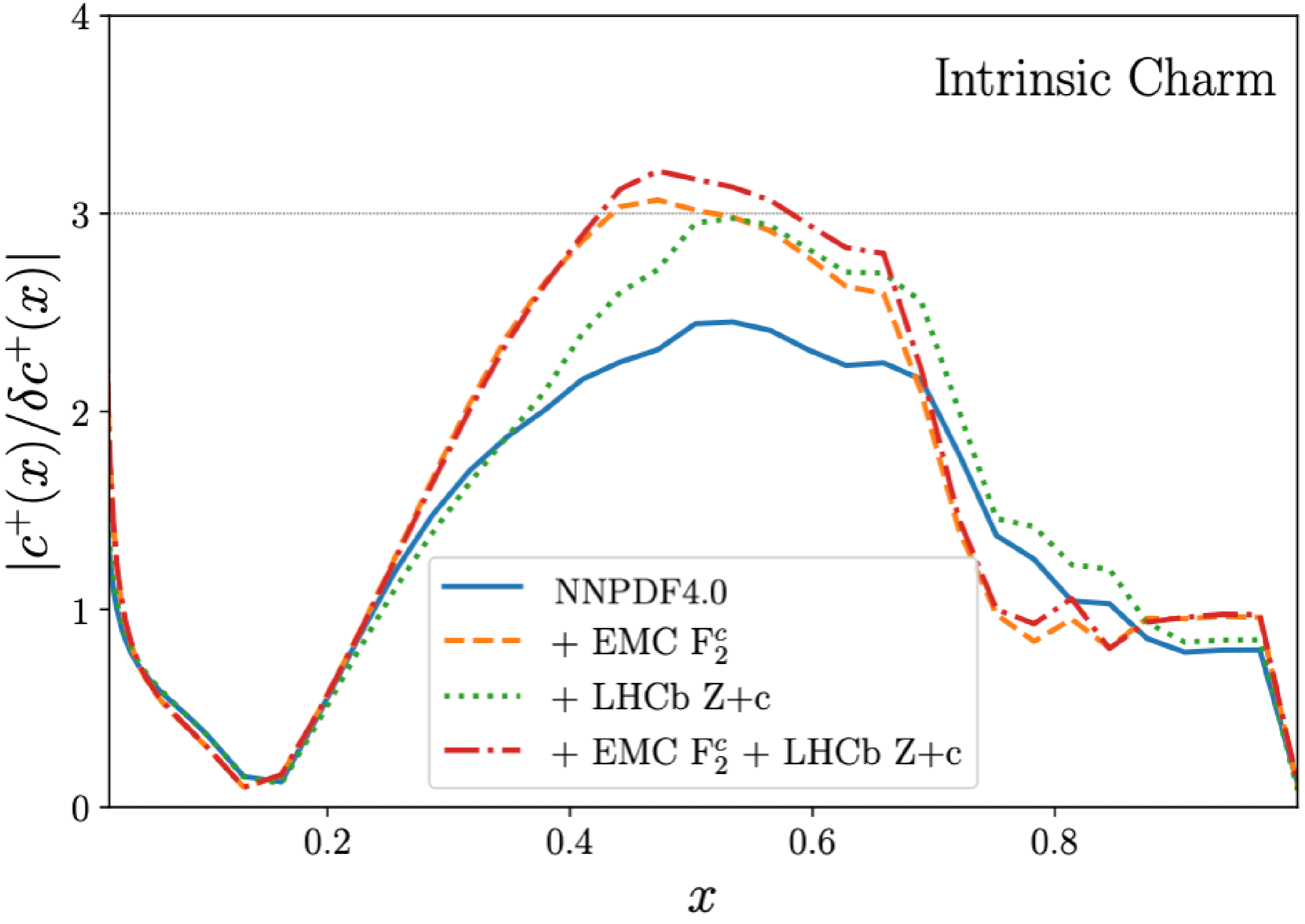


INTRINSIC CHARM WITH LHCb DATA INCLUDED: COMPLETE CONSISTENCY



INTRINSIC CHARM DISCOVERY

THE CURRENT BEST EVIDENCE



MORE THAN 3 σ EVIDENCE

THE ROAD AHEAD

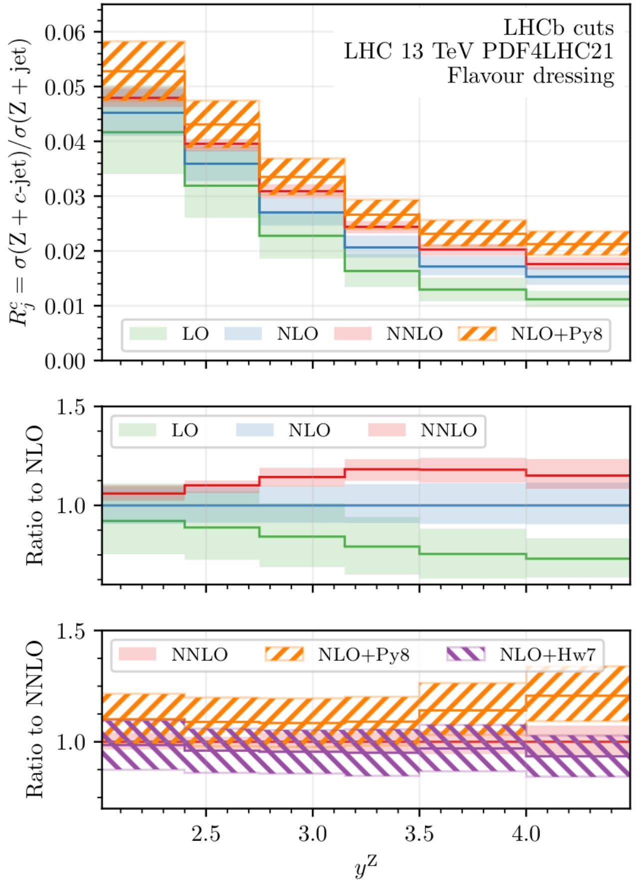
THE PROBLEM OF FLAVOR-TAGGED JETS

- ANTI-k_t WITH FLAVOR TAG IS NOT IRC SAFE
- FLAVOR k_t ALGORITHM PROPOSED LONG AGO (Banfi, Salam, Zanderighi, 2006)
  $\Rightarrow$ REQUIRES KNOWLEDGE OF FLAVOR AT ALL STAGES OF CLUSTERING
- ALTERNATIVE PROPOSALS BASED ON DIFFERENT RECOMBINATION
  $\Rightarrow$ DO NOT REDUCE TO ANTI-k_t, NOT IRC SAFE, OR BOTH

FLAVOR DRESSING

(Gauld, Huss, Stagnitto, 2022)

LHCb FLAVOR JET RATIO



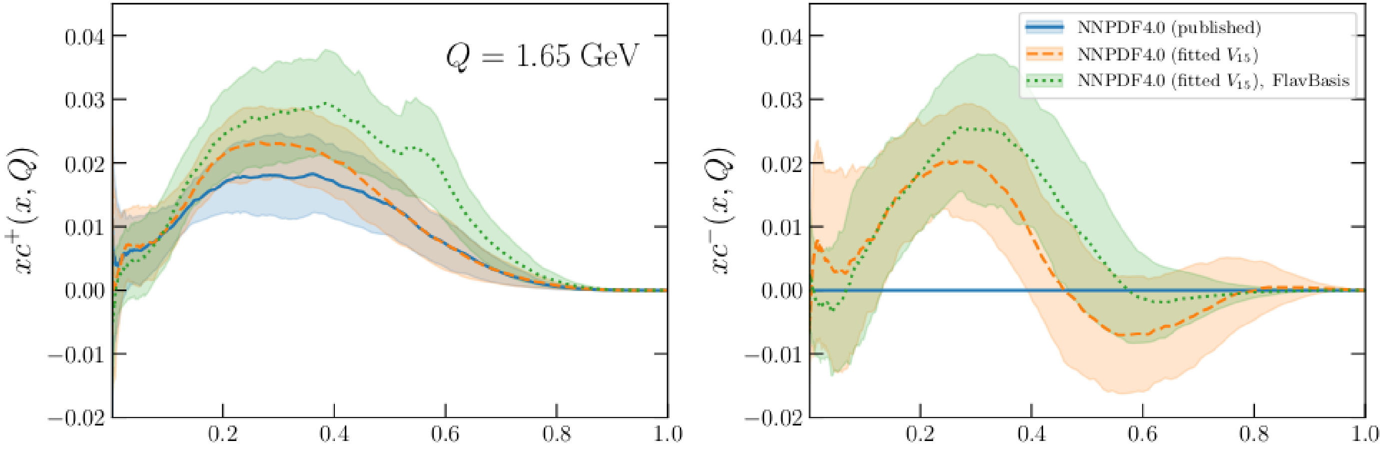
- DEFINE JETS USING ANTI-k_t (FLAVOR AGNOSTIC)
- CONSTRUCT FLAVORED CLUSTERS BASED ON IRC-SAFE RECOMBINATION
- NNLO RESULTS CLOSE TO NLO+MC
- FULL NNLO ANALYSIS NOW POSSIBLE

(Gauld et al in preparation)

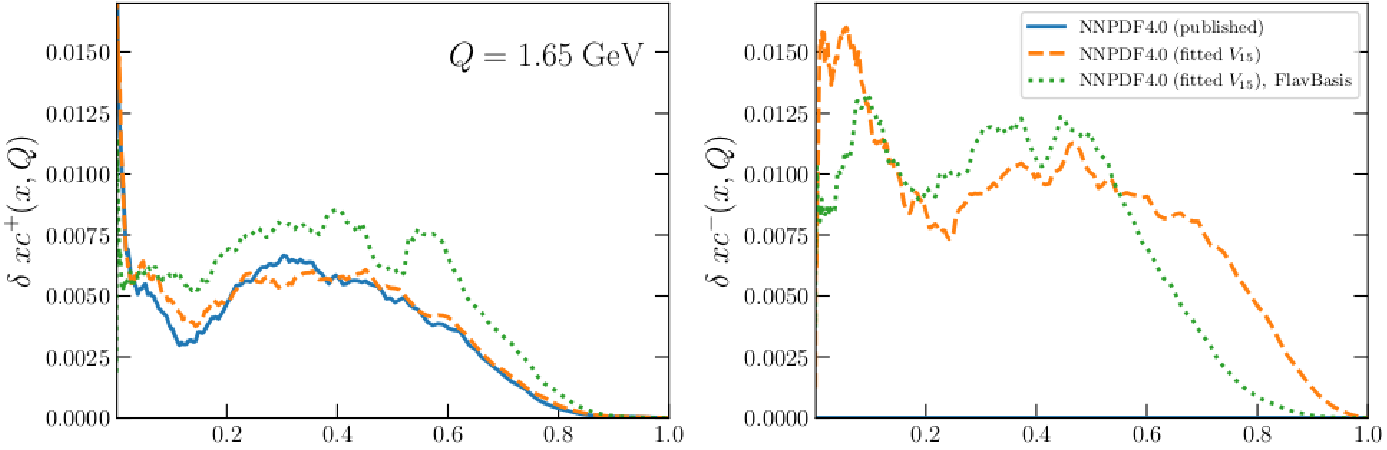
THE CHARM ASYMMETRY

- PARAMETRIZE $c^{\pm} = c \pm \bar{c}$ INDEPENDENTLY → ONLY FIRST MOMENT MUST VANISH
- $c \neq \bar{c} \Rightarrow$ SURELY INTRINSIC

PDFs



RELATIVE UNCERTAINTIES



- SMALLER UNCERTAINTY ON $c^{=} \Rightarrow$ SIGNIFICANCE INCREASED TO 3σ FOR BASELINE
- ABOUT 1.4σ EVIDENCE FOR $c^- \neq 0$
- DIRECTLY MEASURABLE?

SUMMARY

WE FITTED THE CHARM PDF IN ORDER TO GET

- REALISTIC ERROR ESTIMATE
- NO STRONG DEPENDENCE ON CHARM MASS
- NO SENSITIVITY TO MHOU IN MATCHING CONDITION

WE FOUND

- LARGE UNCERTAINTIES AND CHARM COMPATIBLE WITH ZERO AT SMALL x
- THREE-σ EVIDENCE FOR AN INTRINSIC CHARM VALENCE PEAK

TO DO

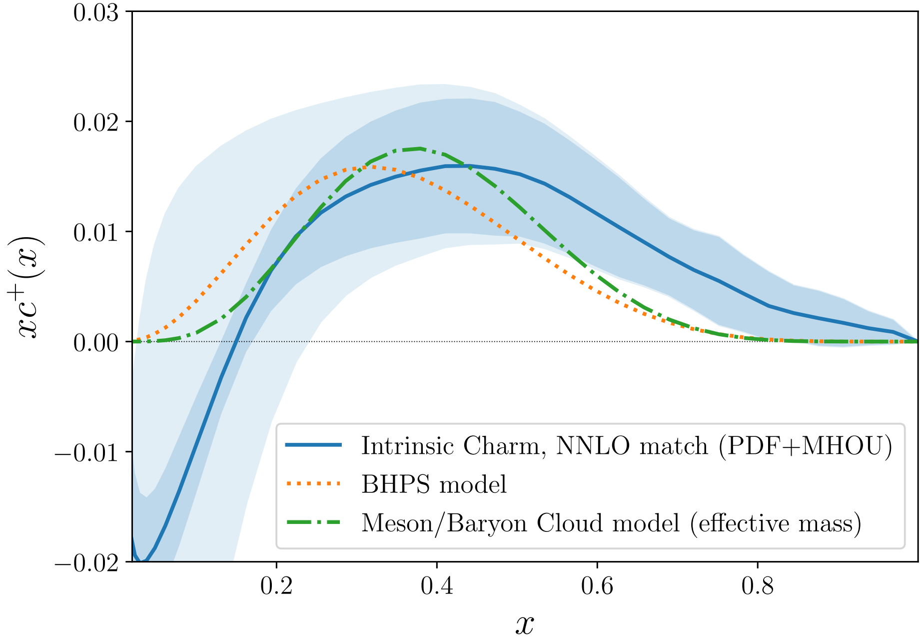
- BETTER CHARMED JET DEFINITIONS $\Rightarrow$ NNLO
- MORE DATA $\Rightarrow$ FIVE σ EVIDENCE
- $c-\bar{c}$ ASYMMETRY PHENOMENOLOGY

EXTRAS

MODELS

- SHAPE OF INTRINSIC CHARM PREDICTED BY MODELS
- FOCK-SPACE WAVE FUNCTION (Brosky, Hoyer, Peterson, Sakai, 1980)
- MESON CLOUD (Hobbs, Londergan, Melnitchouk, 2014)

NNPDF4.0 INTRINSIC CHARM VS. MODELS

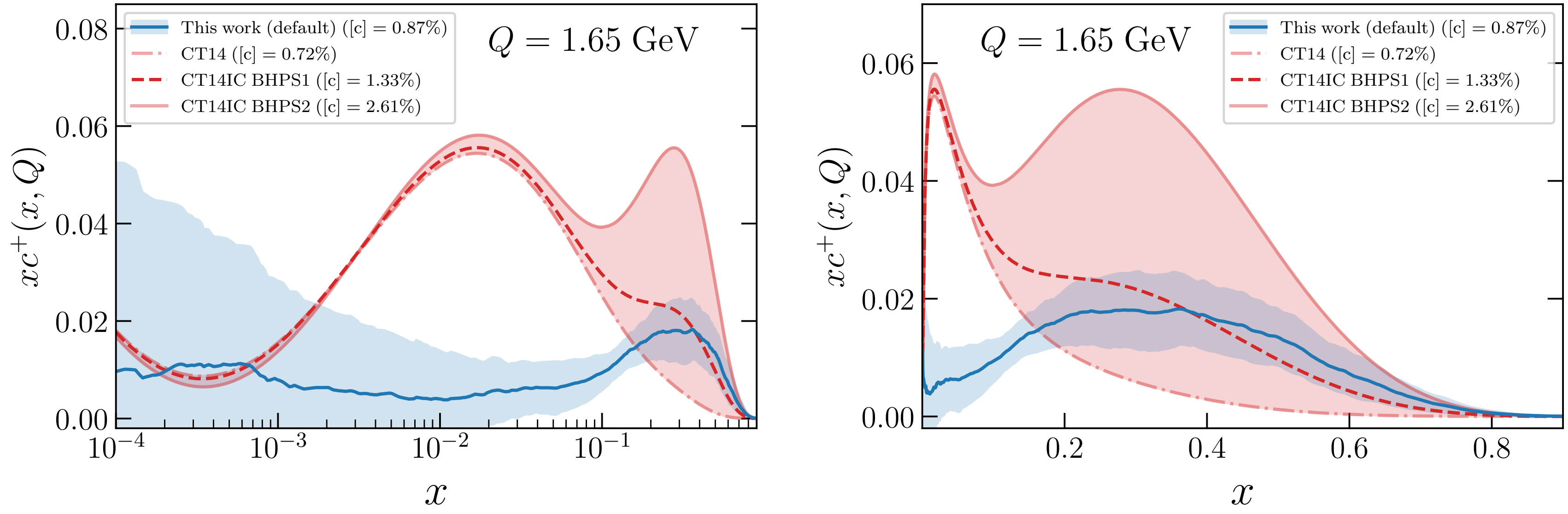


QUALITATIVE AGREEMENT

MODEL FITS

- CT FITTED BPHS NORMALIZATION USING S-ACOT (PROBLEMATIC)
- LOW $Q_0 = 1.3$ GEV
- BEST FIT $\langle c \rangle \sim 1\%$ (CT14) OR$\langle c \rangle \sim 0.5\%$ (CT18)

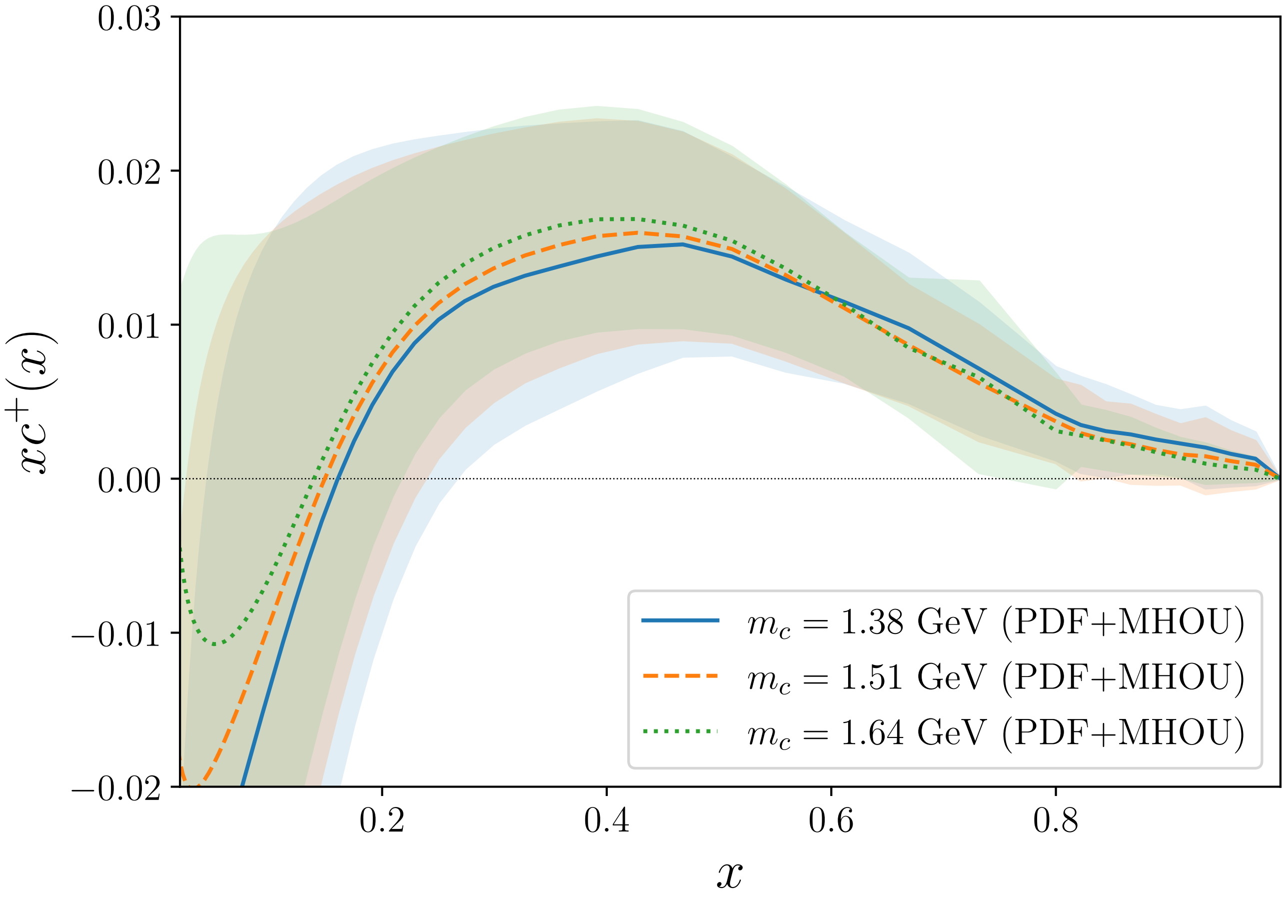
NNPDF4.0 INTRINSIC CHARM VS. CT14 FIT



- GOOD AGREEMENT W. NNPDF AT LARGE x FOR SIMILAR NORM.
- HUGE PERTURBATIVE BUMP AT SMALL x

STABILITY:

CHARM MASS

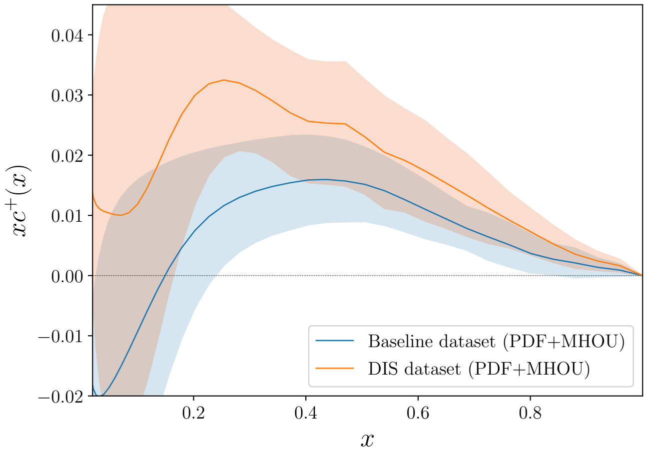


- NEGLIGIBLE DEPENDENCE ON m_c (UNLIKE PERTURBATIVE CHARM)

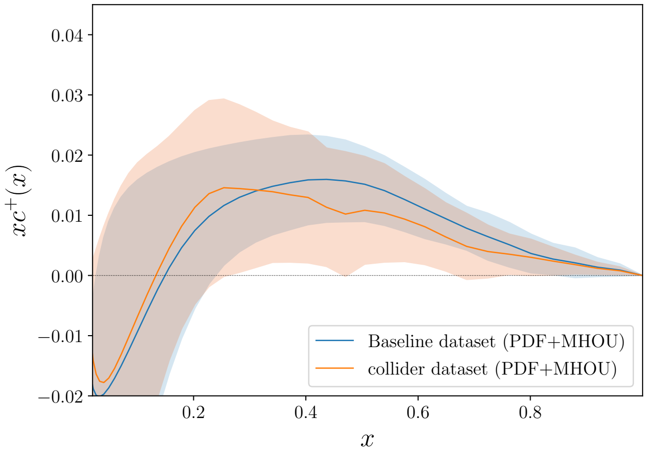
WHICH DATA DRIVE THE ANSWER?:

DATA SUBSETS $n_F = 3$

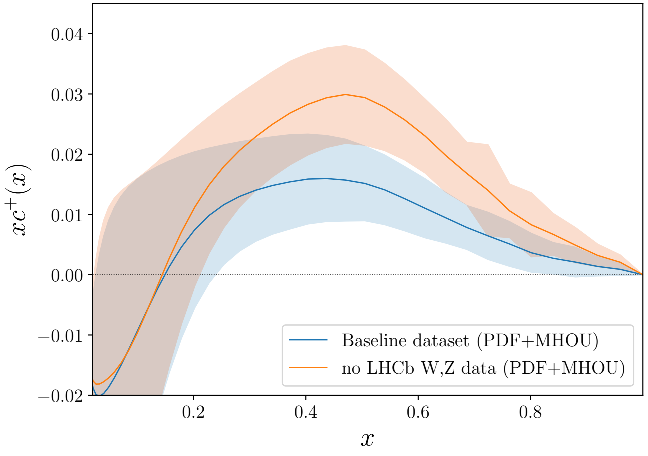
DIS ONLY



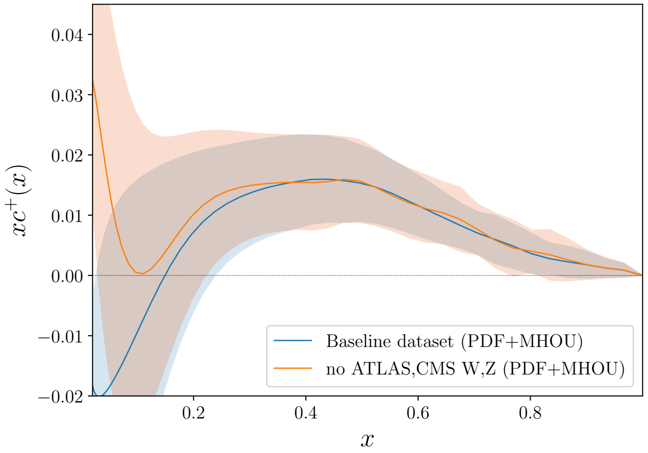
COLLIDER ONLY



NO LHCB



NO ATLAS/CMS DY



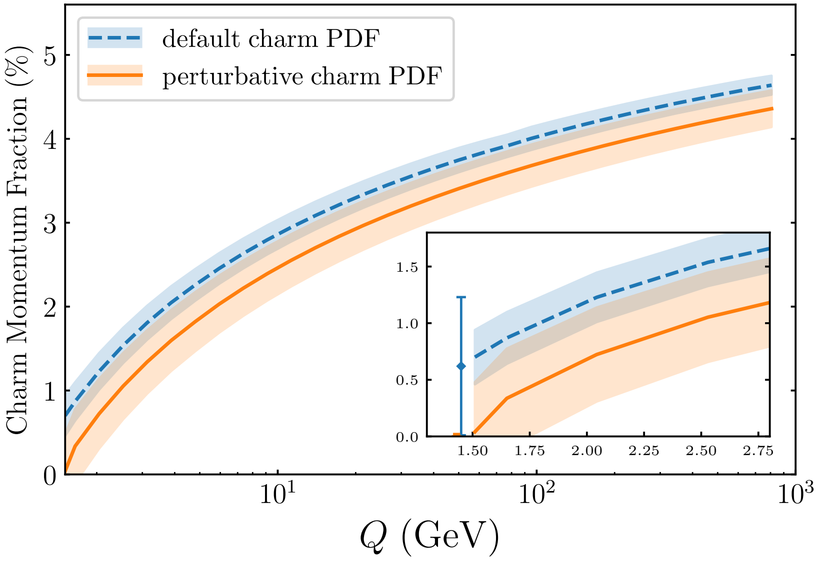
- ALL DATASETS IN AGREEMENT
- COLLIDER DATA MORE IMPORTANT THAN DIS DATA FOR PRECISION
- LHCB W, Z SIGNIFICANT IMPACT

THE CHARM MOMENTUM FRACTION

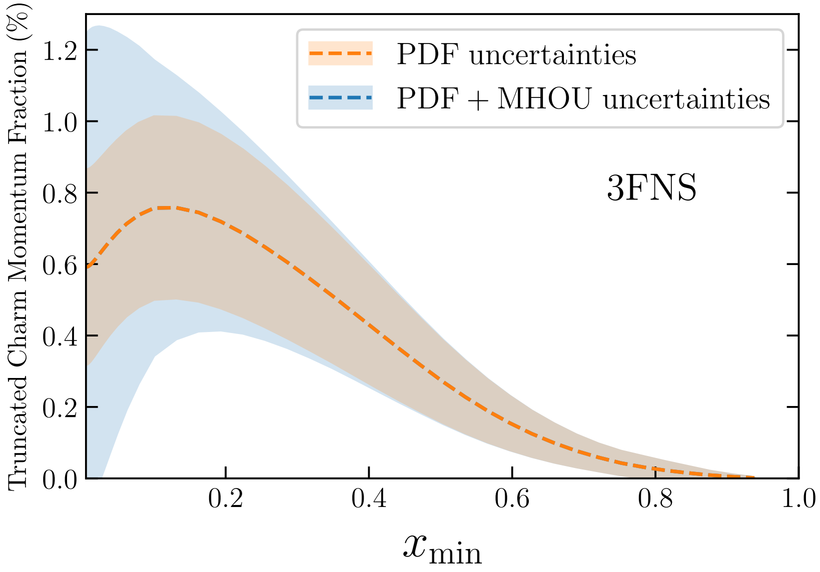
- $n_f = 4$, $Q = 1.65$ GEV: FITTED $\langle c \rangle = 0.87 \pm 0.23_{\rm pdf}\%$ VS.
  PERTURBATIVE $\langle c \rangle = 0.346 \pm 0.005_{\rm pdf} \pm 0.44_{\rm mhou}\%$
- $n_f = 3$, FITTED $\langle c \rangle = 0.62 \pm 0.28_{\rm pdf} \pm 0.54_{\rm mhou}\%\%$ VS.

  PERTURBATIVE $\langle c \rangle = 0\%$

$n_f = 4$: FITTED VS. PERTURBATIVE



$n_f = 3$: MOMENTUM INTEGRAL



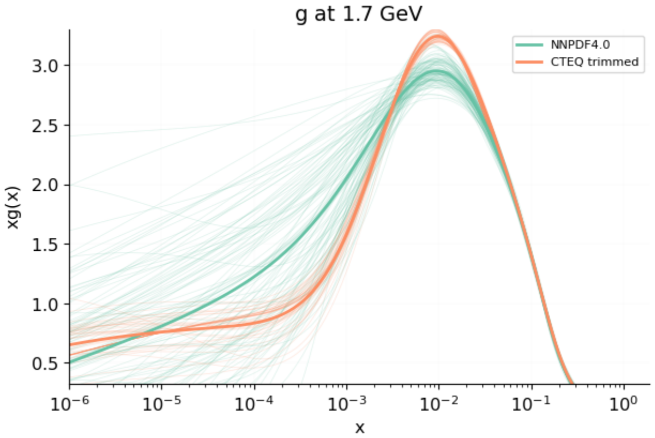
- $n_f = 4$ MOMENTUM FRACTION DETERMINED TO GOOD ACCURACY
- LARGE MHOU AT SMALL $x \Rightarrow$ TOTAL INTRINSIC MOMENTUM FRACTION

  COMPATIBLE WITH ZERO

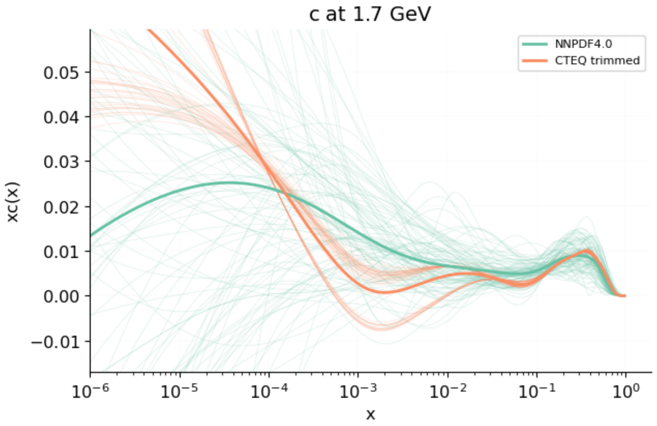
“HOPSCOTCH” OVERLEARNT PDFs

PDFs

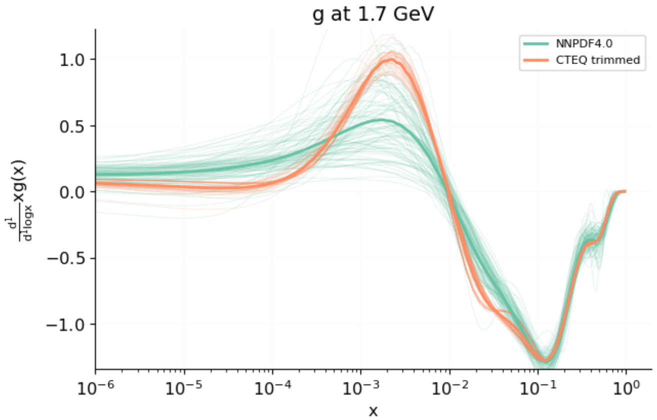
gluon

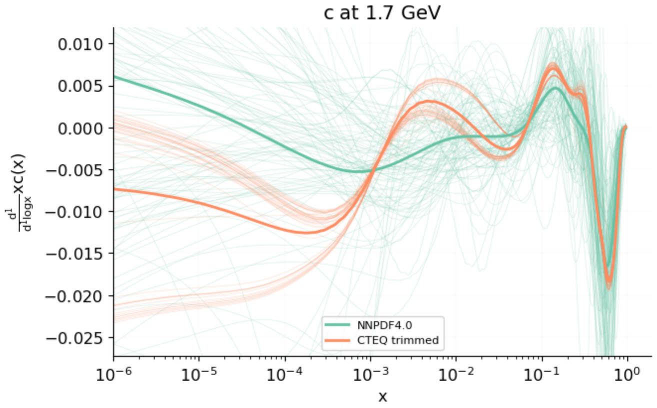


charm



KINETIC ENERGY (DIFFERENTIAL ARCLENGTH)

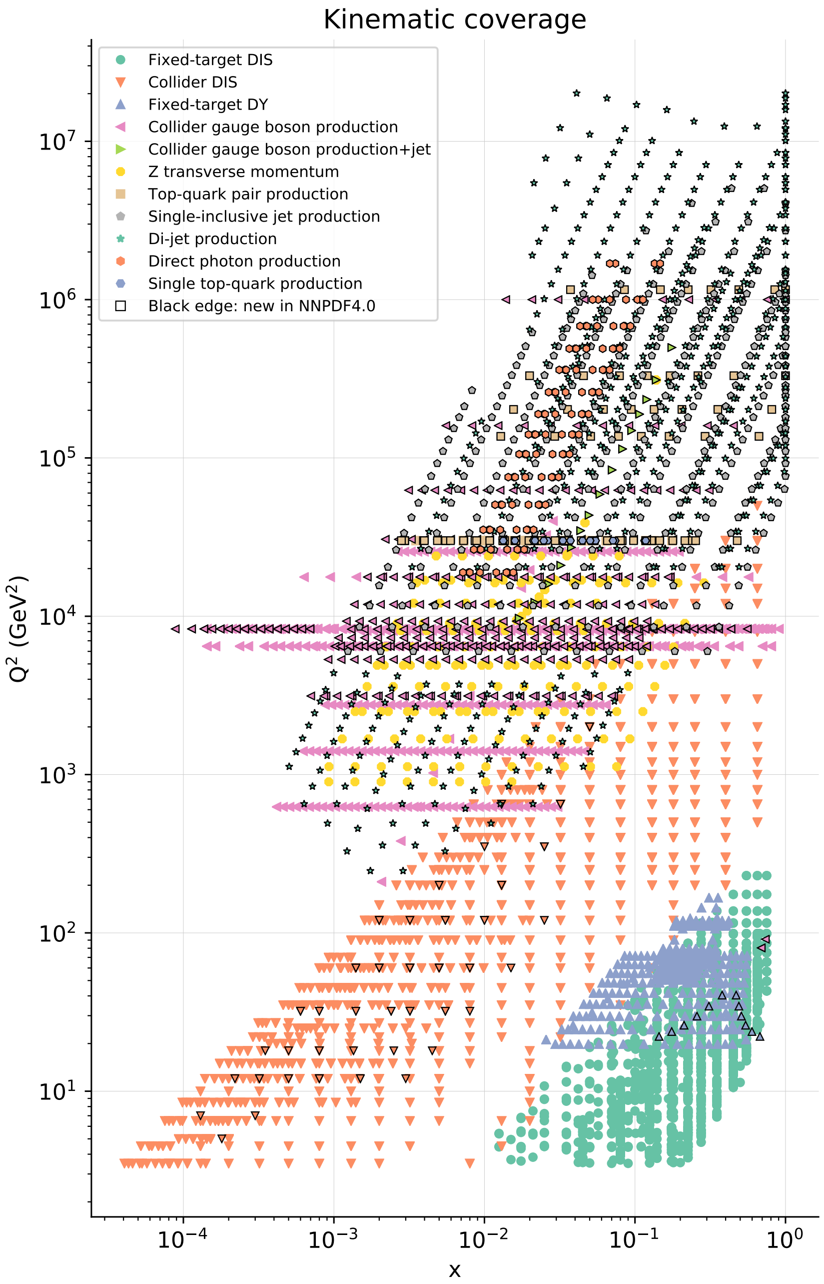




- HS OVERLEARNT BY CONSTRUCTION
- CAN BE REPRODUCED IN NNPDF4.0 BY FORCING OVERLEARNING

PDF DETERMINATION:
NNPDF4.0

DATA



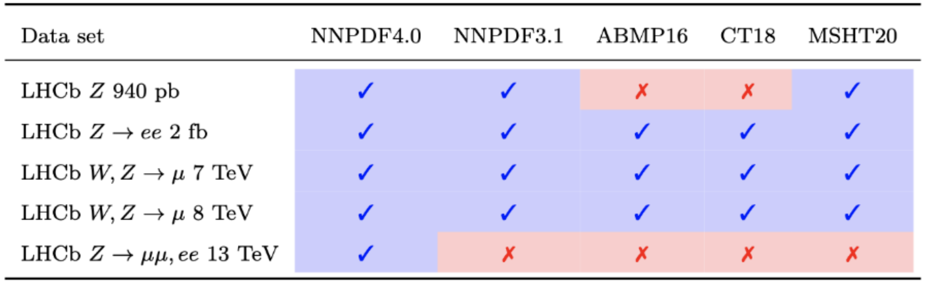
- compared to NNPDF3.1/PDF4LHC21 about 50 new datasets & 400 extra datapoints
- full DIS and FT DY dataset
  - as in NNPDF3.1: final HERA, NMC, BCDMS, CHORUS, NuTeV
  - now also NOMAD neutrino
  - SeaQuest DY
- full 7 TeV and 8 TeV dataset & extensive use of 13 TeV data:
  - W, Z production: rapidity distributions, asymmetries, Z p_T distributions
  - top pair production: all available distributions
  - single-inclusive jets
- several new processes:
  - prompt photon
  - single top
  - dijets
  - HERA jets

about 4000 datapoints

THE LARGEST DATASET

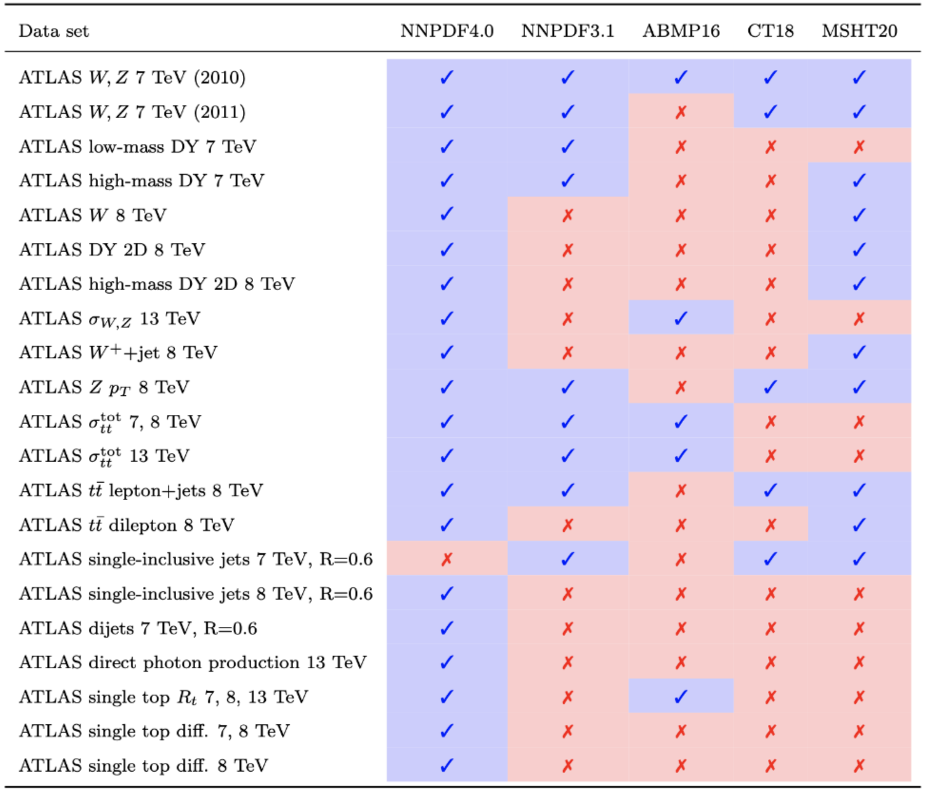
LHC DATA

LHCb



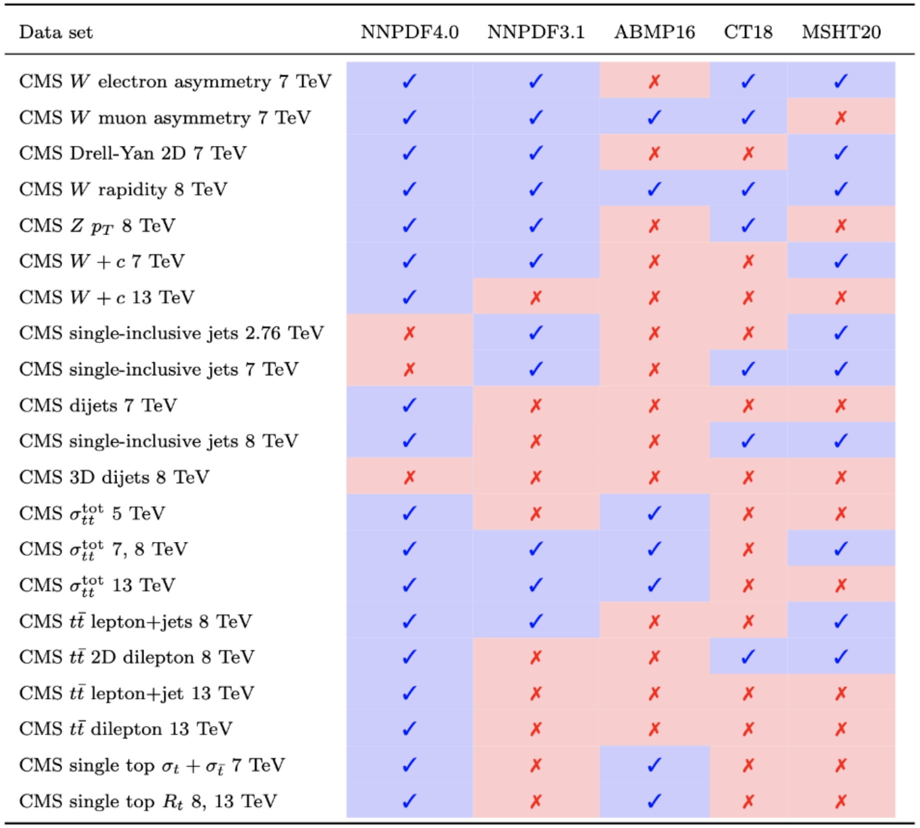
Data set	NNPDF4.0	NNPDF3.1	ABMP16	CT18	MSHT20
LHCb Z 940 pb	✓	✓	✗	✗	✓
LHCb $Z \to ee$ 2 fb	✓	✓	✓	✓	✓
LHCb $W, Z \to \mu$ 7 TeV	✓	✓	✓	✓	✓
LHCb $W, Z \to \mu$ 8 TeV	✓	✓	✓	✓	✓
LHCb $Z \to \mu\mu, ee$ 13 TeV	✓	✗	✗	✗	✗

ATLAS



Data set	NNPDF4.0	NNPDF3.1	ABMP16	CT18	MSHT20
ATLAS W, Z 7 TeV (2010)	✓	✓	✓	✓	✓
ATLAS W, Z 7 TeV (2011)	✓	✓	✗	✓	✓
ATLAS low-mass DY 7 TeV	✓	✓	✗	✗	✗
ATLAS high-mass DY 7 TeV	✓	✓	✗	✗	✓
ATLAS W 8 TeV	✓	✗	✗	✗	✓
ATLAS DY 2D 8 TeV	✓	✗	✗	✗	✓
ATLAS high-mass DY 2D 8 TeV	✓	✗	✗	✗	✓
ATLAS $\sigma_{W,Z}$ 13 TeV	✓	✗	✓	✗	✗
ATLAS W^++jet 8 TeV	✓	✗	✗	✗	✓
ATLAS Z p_T 8 TeV	✓	✓	✗	✓	✓
ATLAS $\sigma_{tt}^{\rm tot}$ 7, 8 TeV	✓	✓	✓	✗	✗
ATLAS $\sigma_{tt}^{\rm tot}$ 13 TeV	✓	✓	✓	✗	✗
ATLAS $t\bar{t}$ lepton+jets 8 TeV	✓	✓	✗	✓	✓
ATLAS $t\bar{t}$ dilepton 8 TeV	✓	✗	✗	✗	✓
ATLAS single-inclusive jets 7 TeV, R=0.6	✗	✓	✗	✓	✓
ATLAS single-inclusive jets 8 TeV, R=0.6	✓	✗	✗	✗	✗
ATLAS dijets 7 TeV, R=0.6	✓	✗	✗	✗	✗
ATLAS direct photon production 13 TeV	✓	✗	✗	✗	✗
ATLAS single top R_t 7, 8, 13 TeV	✓	✗	✓	✗	✗
ATLAS single top diff. 7, 8 TeV	✓	✗	✗	✗	✗
ATLAS single top diff. 8 TeV	✓	✗	✗	✗	✗

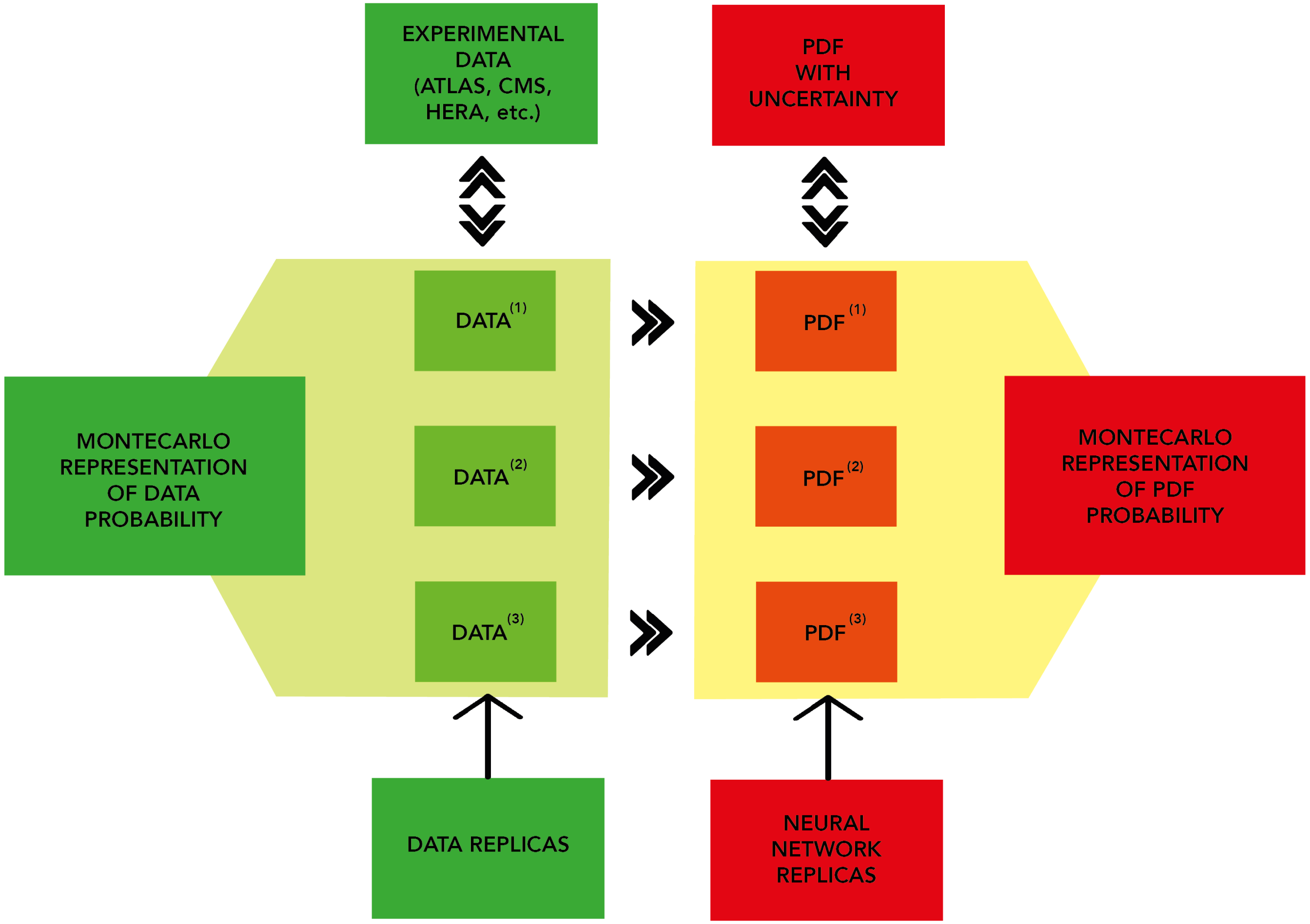
CMS



Data set	NNPDF4.0	NNPDF3.1	ABMP16	CT18	MSHT20
CMS W electron asymmetry 7 TeV	✓	✓	✗	✓	✓
CMS W muon asymmetry 7 TeV	✓	✓	✓	✓	✗
CMS Drell-Yan 2D 7 TeV	✓	✓	✗	✗	✓
CMS W rapidity 8 TeV	✓	✓	✓	✓	✓
CMS Z p_T 8 TeV	✓	✓	✗	✓	✗
CMS $W+c$ 7 TeV	✓	✓	✗	✗	✓
CMS $W+c$ 13 TeV	✓	✗	✗	✗	✗
CMS single-inclusive jets 2.76 TeV	✗	✓	✗	✗	✓
CMS single-inclusive jets 7 TeV	✗	✓	✗	✓	✓
CMS dijets 7 TeV	✓	✗	✗	✗	✗
CMS single-inclusive jets 8 TeV	✓	✗	✗	✓	✓
CMS 3D dijets 8 TeV	✗	✗	✗	✗	✗
CMS $\sigma_{tt}^{\rm tot}$ 5 TeV	✓	✗	✓	✗	✗
CMS $\sigma_{tt}^{\rm tot}$ 7, 8 TeV	✓	✓	✓	✗	✓
CMS $\sigma_{tt}^{\rm tot}$ 13 TeV	✓	✓	✓	✗	✗
CMS $t\bar{t}$ lepton+jets 8 TeV	✓	✓	✗	✗	✓
CMS $t\bar{t}$ 2D dilepton 8 TeV	✓	✗	✗	✓	✓
CMS $t\bar{t}$ lepton+jet 13 TeV	✓	✗	✗	✗	✗
CMS $t\bar{t}$ dilepton 13 TeV	✓	✗	✗	✗	✗
CMS single top $\sigma_t + \sigma_{\bar{t}}$ 7 TeV	✓	✗	✓	✗	✗
CMS single top R_t 8, 13 TeV	✓	✗	✓	✗	✗

THE NNPDF METHODOLOGY

REPLICA SAMPLE OF FUNCTIONS ⇔ PROBABILITY DENSITY IN FUNCTION SPACE
KNOWLEDGE OF LIKELIHHOD SHAPE (FUNCTIONAL FORM) NOT NECESSARY



EXPERIMENTAL DATA (ATLAS, CMS, HERA, etc.)

PDF WITH UNCERTAINTY

MONTECARLO REPRESENTATION OF DATA PROBABILITY

DATA$^{(1)}$

DATA$^{(2)}$

DATA$^{(3)}$

PDF$^{(1)}$

PDF$^{(2)}$

PDF$^{(3)}$

MONTECARLO REPRESENTATION OF PDF PROBABILITY

DATA REPLICAS

NEURAL NETWORK REPLICAS

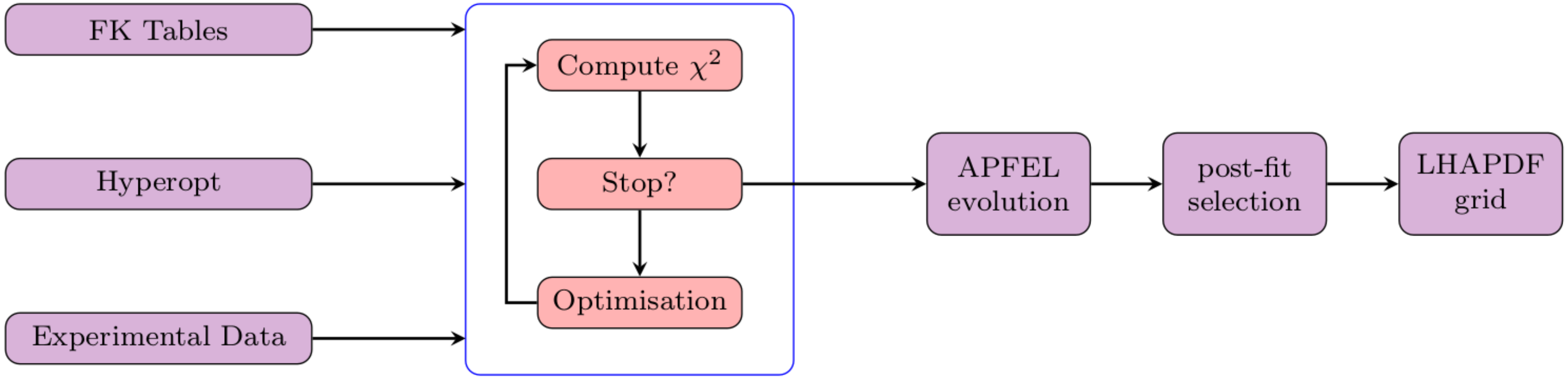
FINAL PDF SET: $f_i^{(a)}(x,\mu)$;
i =up, antiup, down, antidown, strange, antistrange, charm, gluon; $j = 1, 2, \ldots N_{\mathrm{rep}}$

THE NNPDF CODE STRUCTURE

- MODULAR PYTHON-BASED CODE
- HIGH DEGREE PARALLELIZATION & HARDWARE ACCELERATION

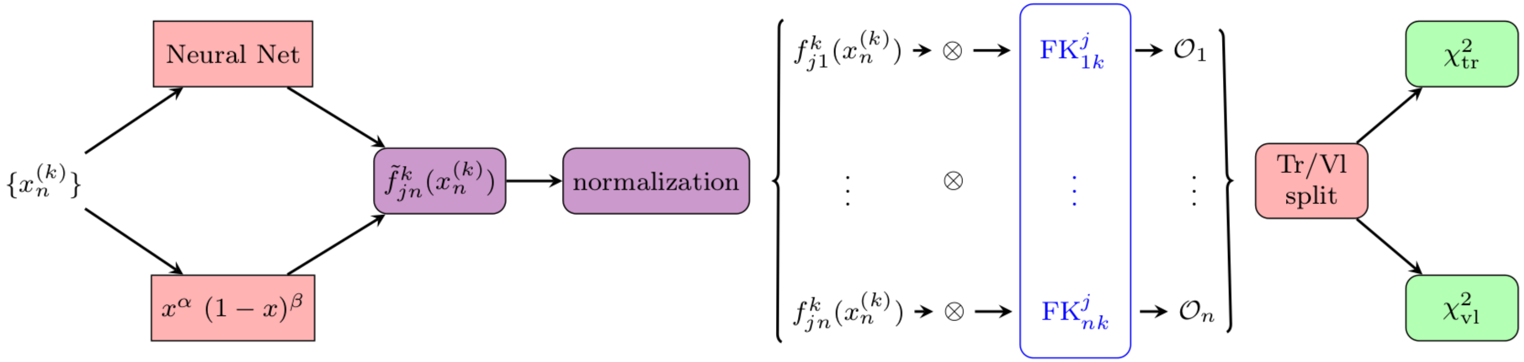
AVERAGE FITTING TIME PER REPLICA AND USE OF RESOURCES
SAME DATASET FOR OLD AND NEW METHODOLOGIES IN CPU AND GPU
CPU: INTEL(R) CORE(TM) I7-4770 AT 3.40GHZ; GPU: NVIDIA TITAN V

	NNPDF31 CODEBASE	NNPDF40 CODEBASE IN CPU	NNPDF40 CODEBASE IN GPU
TIME	15.2 H.	38 ± 5 MIN.	6.6 MIN.
RAM USE	1.5 GB	6.1 GB	NA



MINIMIZATION AND CROSS-VALIDATION

- DATA REPLICAS ⇒ PDF REPLICAS
- EACH PDF REPLICA: PREPROCESSED NEURAL NET
- NEURAL NET ⇒ OBSERVABLES
- RANDOM TRAINING-VALIDATION SPLIT, χ^2 TO TRAINING DATA REPLICAS MINIMIZED
- TRAINING STOPS IF VALIDATION χ^2 GROWS FOR A WHILE (PATIENCE)
- LOWEST VALIDATION χ^2 ⇒ OPTIMAL FIT



$\{x_n^{(k)}\}$

Neural Net

$x^\alpha\ (1-x)^\beta$

$\tilde{f}^k_{jn}(x_n^{(k)})$

normalization

$f^k_{j1}(x_n^{(k)}) \rightarrow \otimes \longrightarrow \mathrm{FK}^j_{1k} \rightarrow \mathcal{O}_1$

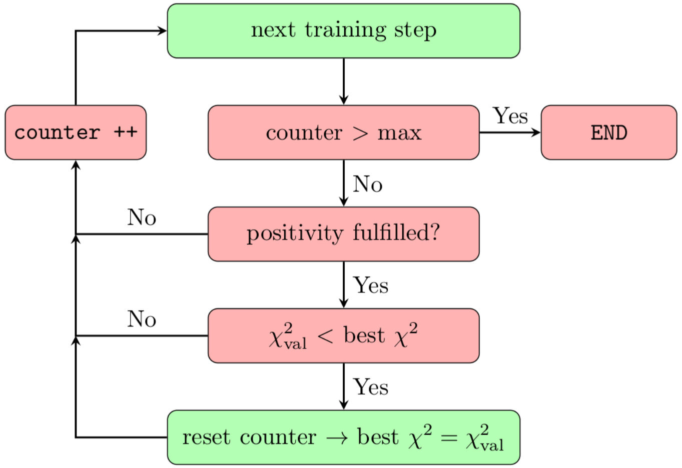
$\vdots \quad \otimes \quad \vdots \quad \vdots$

$f^k_{jn}(x_n^{(k)}) \rightarrow \otimes \longrightarrow \mathrm{FK}^j_{nk} \rightarrow \mathcal{O}_n$

Tr/Vl split

$\chi^2_{\rm tr}$

$\chi^2_{\rm vl}$



next training step

counter > max

Yes

END

No

positivity fulfilled?

No

Yes

$\chi^2_{\rm val} <$ best χ^2

No

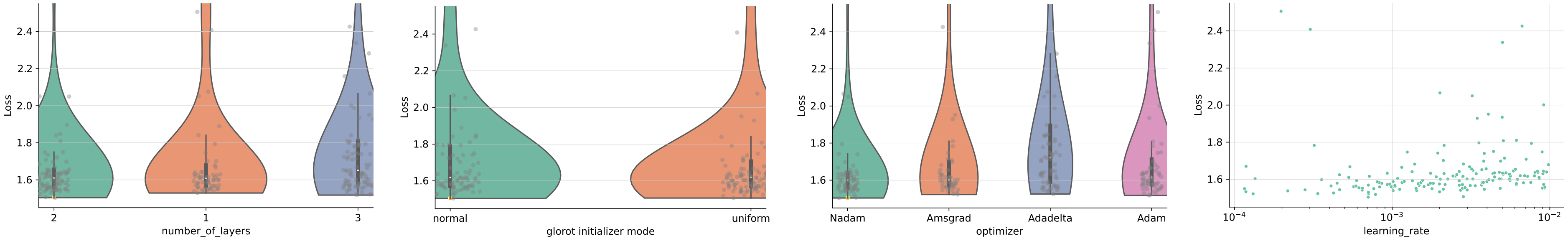
Yes

reset counter → best $\chi^2 = \chi^2_{\rm val}$

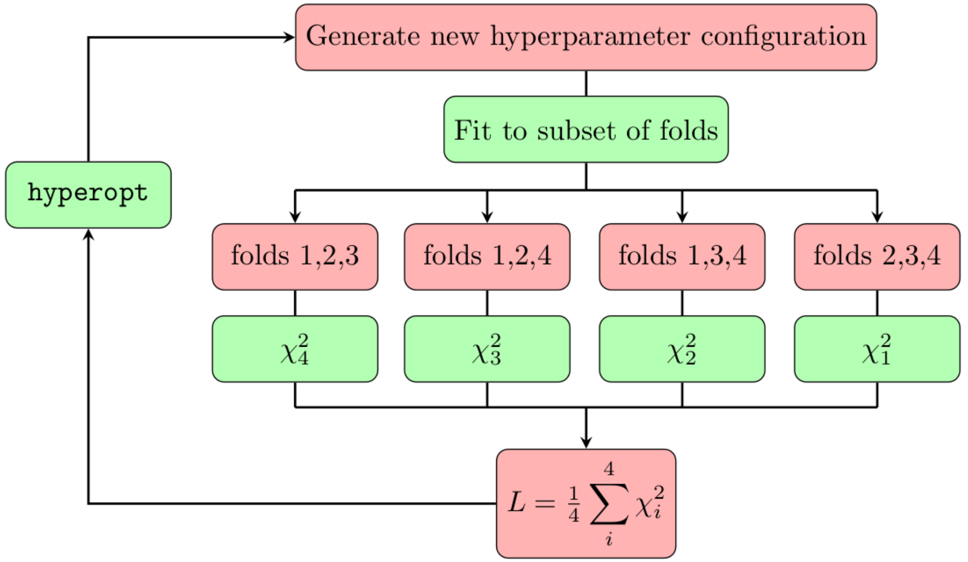
`counter ++`

HYPEROPTIMIZATION

- PARAMETRIZATION AND MINIMIZATION PARAMETERS VARIED
- SCAN OF PARAMETER SPACE
- BAYESIAN UPDATING LEADS TO BEST METHODOLOGY



K-FOLDING



$$L = \frac{1}{4}\sum_{i}^{4} \chi^2_i$$

- HYPEROPTIMIZATION ⇒ OVERFITTING (χ^2 TOO GOOD)
- CHECK GENERALIZATION POWER: K-FOLDING
  - DIVIDE DATA IN FOLDS
  - EXCLUDE ONE FOLD IN TURN FROM FIT
  - OPTIMIZE ON THE χ^2 OF THE EXCLUDED FOLDS
  - BEST AVERAGE OR BEST WORST

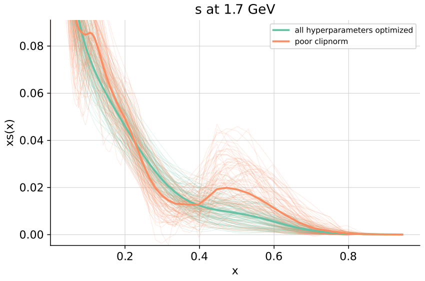
Fold 1		
CHORUS σ^{ν}_{CC}	HERA I+II inc NC e^+p 920 GeV	BCDMS p
LHCb Z 940 pb	ATLAS W, Z 7 TeV 2010	CMS Z p_T 8 TeV (p_T^{ll}, y_{ll})
DY E605 σ^{p}_{DY}	CMS Drell-Yan 2D 7 TeV 2011	CMS 3D dijets 8 TeV
ATLAS single-$\bar{t}$ y (normalised)	ATLAS single top R_t 7 TeV	CMS $t\bar{t}$ rapidity $y_{t\bar{t}}$
CMS single top R_t 8 TeV		

Fold 2		
HERA I+II inc CC e^-p	HERA I+II inc NC e^+p 460 GeV	HERA comb. $\sigma^{\mathrm{red}}_{b\bar{b}}$
NMC p	NuTeV $\sigma^{\bar{\nu}}_{c}$	LHCb $Z \to ee$ 2 fb
CMS W asymmetry 840 pb	ATLAS Z p_T 8 TeV (p_T^{ll}, M_{ll})	D0 $W \to \mu\nu$ asymmetry
DY E886 σ^{p}_{DY}	ATLAS direct photon 13 TeV	ATLAS dijets 7 TeV, R=0.6
ATLAS single antitop y (normalised)	CMS $\sigma^{\mathrm{tot}}_{tt}$	CMS single top $\sigma_t + \sigma_{\bar{t}}$ 7 TeV

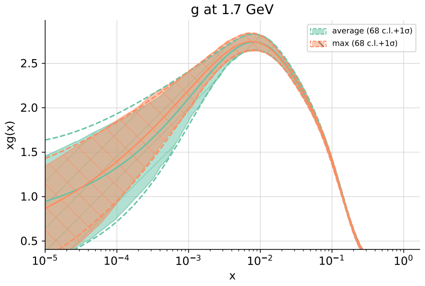
Fold 3		
HERA I+II inc CC e^+p	HERA I+II inc NC e^+p 575 GeV	NMC d/p
NuTeV σ^{ν}_{c}	LHCb $W, Z \to \mu$ 7 TeV	LHCb $Z \to ee$
ATLAS W, Z 7 TeV 2011 Central selection	ATLAS W^++jet 8 TeV	ATLAS HM DY 7 TeV
CMS W asymmetry 4.7 fb	DYE 866 $\sigma^{d}_{\mathrm{DY}}/\sigma^{p}_{\mathrm{DY}}$	CDF Z rapidity (new)
ATLAS $\sigma^{\mathrm{tot}}_{tt}$	ATLAS single top y_t (normalised)	CMS $\sigma^{\mathrm{tot}}_{tt}$ 5 TeV
CMS $t\bar{t}$ double diff. $(m_{t\bar{t}}, y_t)$		

Fold 4		
CHORUS $\sigma^{\bar{\nu}}_{CC}$	HERA I+II inc NC e^+p 820 GeV	LHCb $W, Z \to \mu$ 8 TeV
LHCb $Z \to \mu\mu$	ATLAS W, Z 7 TeV 2011 Fwd	ATLAS W^-+jet 8 TeV
ATLAS low-mass DY 2011	ATLAS Z p_T 8 TeV (p_T^{ll}, y_{ll})	CMS W rapidity 8 TeV
D0 Z rapidity	CMS dijets 7 TeV	ATLAS single top y_t (normalised)
ATLAS single top R_t 13 TeV	CMS single top R_t 13 TeV	

NO K-FOLDING



K-FOLDING VARIATION

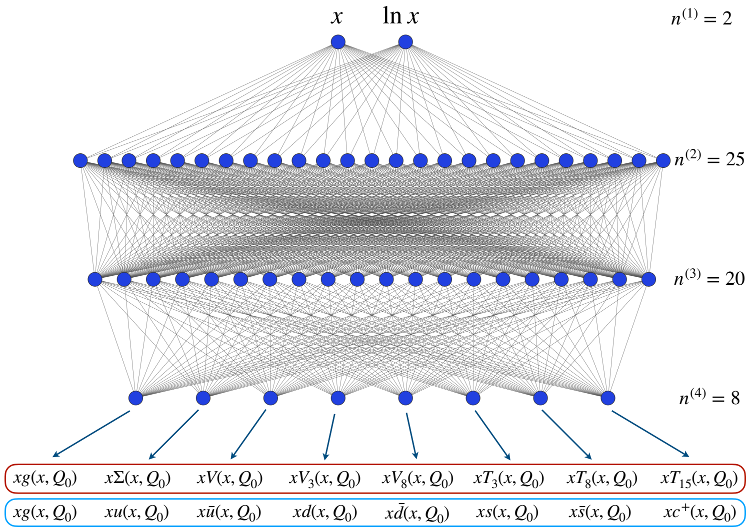


THE ML METHODOLOGY

HYPEROPTIMIZED PARAMETERS

Parameter	NNPDF4.0	L as in Eq. (3.21)	Flavour basis Eq. (3.2)
Architecture	25-20-8	70-50-8	7-26-27-8
Activation function	hyperbolic tangent	hyperbolic tangent	sigmoid
Initializer	`glorot_normal`	`glorot_uniform`	`glorot_normal`
Optimizer	`Nadam`	`Adadelta`	`Nadam`
Clipnorm	6.0×10^{-6}	5.2×10^{-2}	2.3×10^{-5}
Learning rate	2.6×10^{-3}	2.5×10^{-1}	2.6×10^{-3}
Maximum # epochs	17×10^{3}	45×10^{3}	45×10^{3}
Stopping patience	10% of max epochs	12% of max epochs	16% of max epochs
Initial positivity $\Lambda^{(\mathrm{pos})}$	185	106	2
Initial integrability $\Lambda^{(\mathrm{int})}$	10	10	10

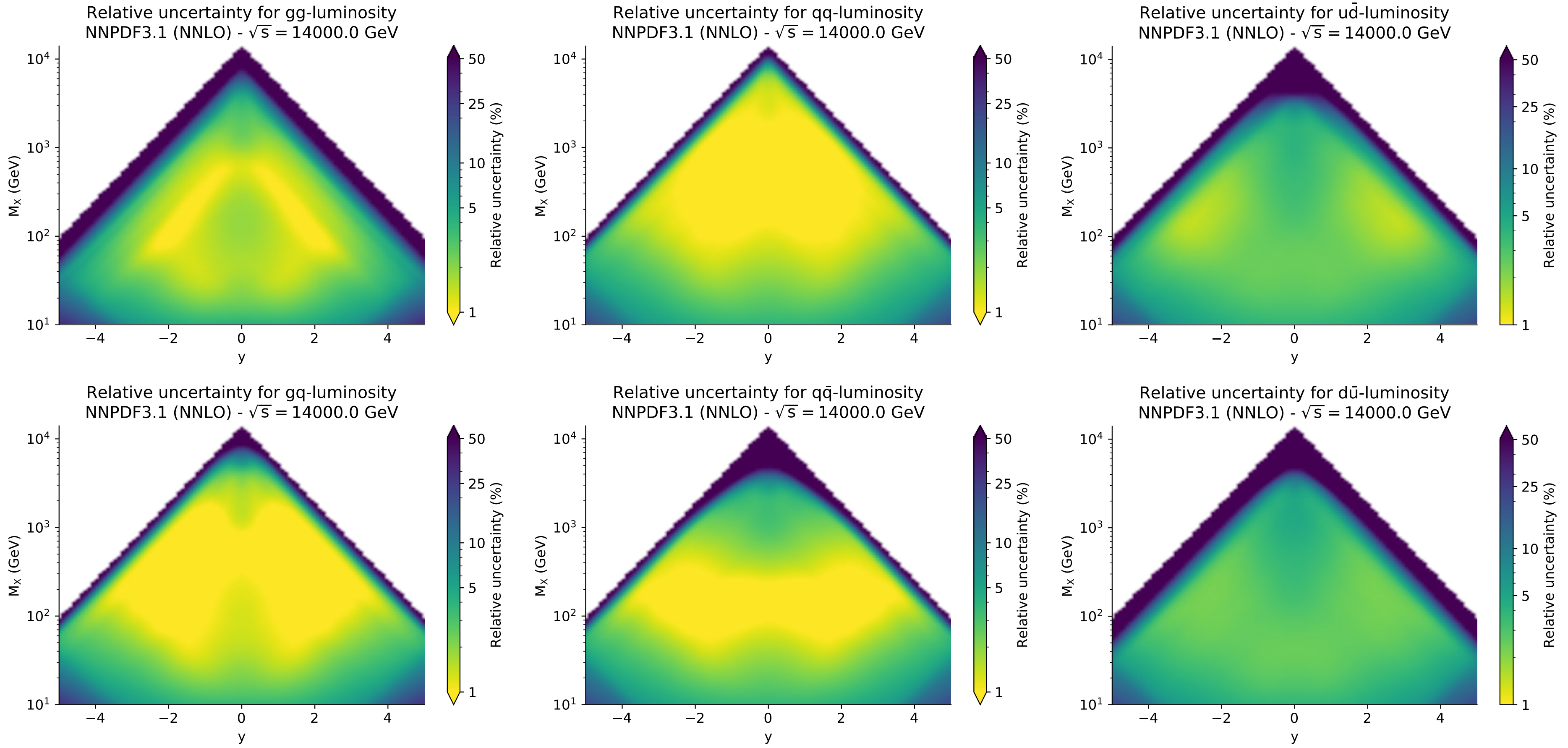
NN ARCHITECTURE



- HYPEROPT ADAPTS TO EXTERNAL CHOICES (E.G. PARAMETRIZATION BASIS)
- SIMILAR RESULTS CAN BE OBTAINED WITH RATHER DIFFERENT SETTINGS
- ~ 800 FREE PARAMETERS

UNCERTAINTIES: FROM NNPDF3.1...

GLUON SINGLET FLAVORS



Relative uncertainty for gg-luminosity
NNPDF3.1 (NNLO) - $\sqrt{s} = 14000.0$ GeV

Relative uncertainty for qq-luminosity
NNPDF3.1 (NNLO) - $\sqrt{s} = 14000.0$ GeV

Relative uncertainty for $u\bar{d}$-luminosity
NNPDF3.1 (NNLO) - $\sqrt{s} = 14000.0$ GeV

Relative uncertainty for gq-luminosity
NNPDF3.1 (NNLO) - $\sqrt{s} = 14000.0$ GeV

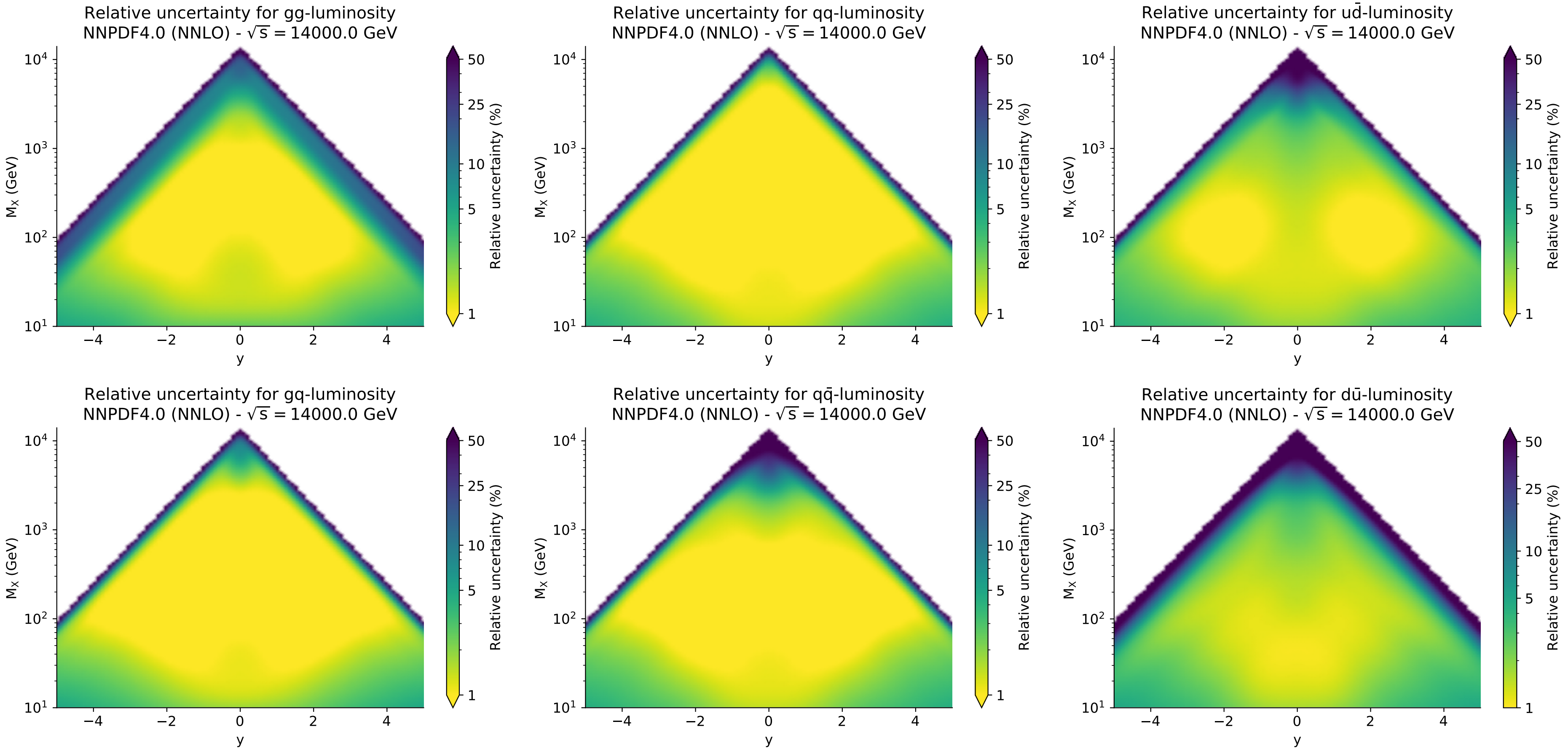
Relative uncertainty for $q\bar{q}$-luminosity
NNPDF3.1 (NNLO) - $\sqrt{s} = 14000.0$ GeV

Relative uncertainty for $d\bar{u}$-luminosity
NNPDF3.1 (NNLO) - $\sqrt{s} = 14000.0$ GeV

- TYPICAL UNCERTAINTIES IN DATA REGION: SINGLET $\sim 3\%$, NONSINGLET $\sim 5\%$
- DATA REGION: $10^2 \lesssim M_X \lesssim 10^3$ TEV, $-2 \lesssim y \lesssim 2$

UNCERTAINTIES: ...TO NNPDF4.0

GLUON SINGLET FLAVORS



- TYPICAL UNCERTAINTIES IN DATA REGION: SINGLET $\sim 1\%$, NONSINGLET $\sim 2-3\%$
- DATA REGION: $10 \lesssim M_X \lesssim 3 \cdot 10^3$ TeV, $-4 \lesssim y \lesssim 4$