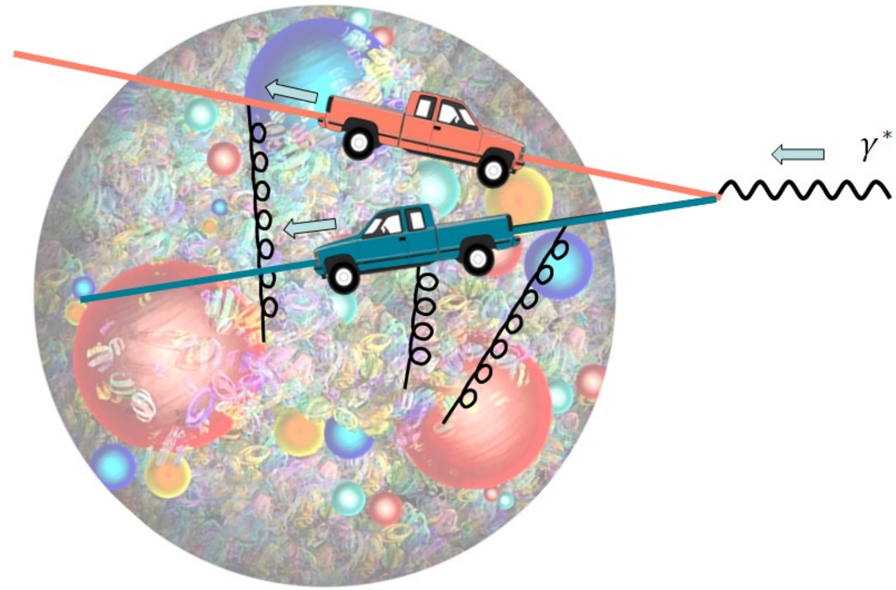
Proton's intrinsic charm remains concealed

Tim Hobbs
Argonne National Laboratory
Pavel Nadolsky
Southern Methodist University
and Fermilab

With A. Courtoy, M. Guzzi, J. Huston,
K. Xie, M. Yan, C.-P. Yuan

and members of the
CTEQ-TEA (Tung Et. Al.) working group



arXiv:2211.01387, 1707.00657
and 2205.10444

Contents

1. Terminology:
   – Nonperturbative, intrinsic (IC), fitted (FC) charm
   – What is “nonperturbative” in “nonperturbative charm”?
   – Process and scheme (in)dependence of the IC
   – SACOT-MPS factorization scheme at NNLO

   As a community, we must develop precise terminology for our quantitative (and the public’s) understanding
2. The IC models of proton structure and CT18 FC NNLO global analysis
3. Probability:
   – Estimation of PDF uncertainties
   – Comparison of CT18 FC and NNPDF4.0 IC studies

References

CTEQ-TEA analyses of fitted charm

1. T.-J. Hou et al., JHEP 02 (2018) 059; 57 pages, 19 figures: QCD factorization with the NP charm and CT14 IC NNLO pheno analysis
2. M. Guzzi, T. J. Hobbs, K. Xie, et al., arXiv:2211.01387; 10 pages: **new** CT18 FC analysis with the LHC Run-1 and 2 data

IC from nonperturbative methods and models:

1. BHPS: Brodsky, Hoyer, Peterson, Sakai, PLB 93 (1980) 451
2. BHPS3: Bluemlein, PLB 753 (2016) 619
3. Meson-Baryon models (MBM): Hobbs, Londergan, Melnitchouk, PRD 89 (2014) 074008
4. Light-front WF models: Hobbs, Alberg, Miller, PRD 96 (2017) 7, 074023
5. Dyson-Schwinger equations, lattice QCD, …

CT18 NNLO analysis and methodology: T.-J. Hou, J. Gao, T. J. Hobbs, K. Xie, et al., Phys.Rev.D 103 (2021) 1, 014013

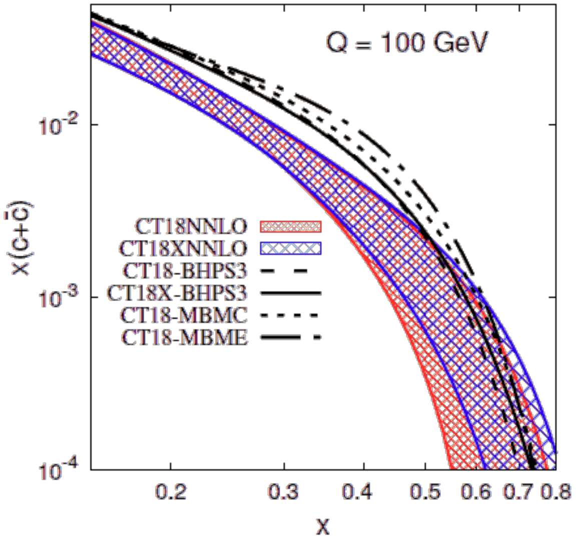
Strong goodness-of-fit criteria for PDF fits: K. Kovařík, P. Nadolsky, D. Soper, Rev.Mod.Phys. 92 (2020) 4, 045003

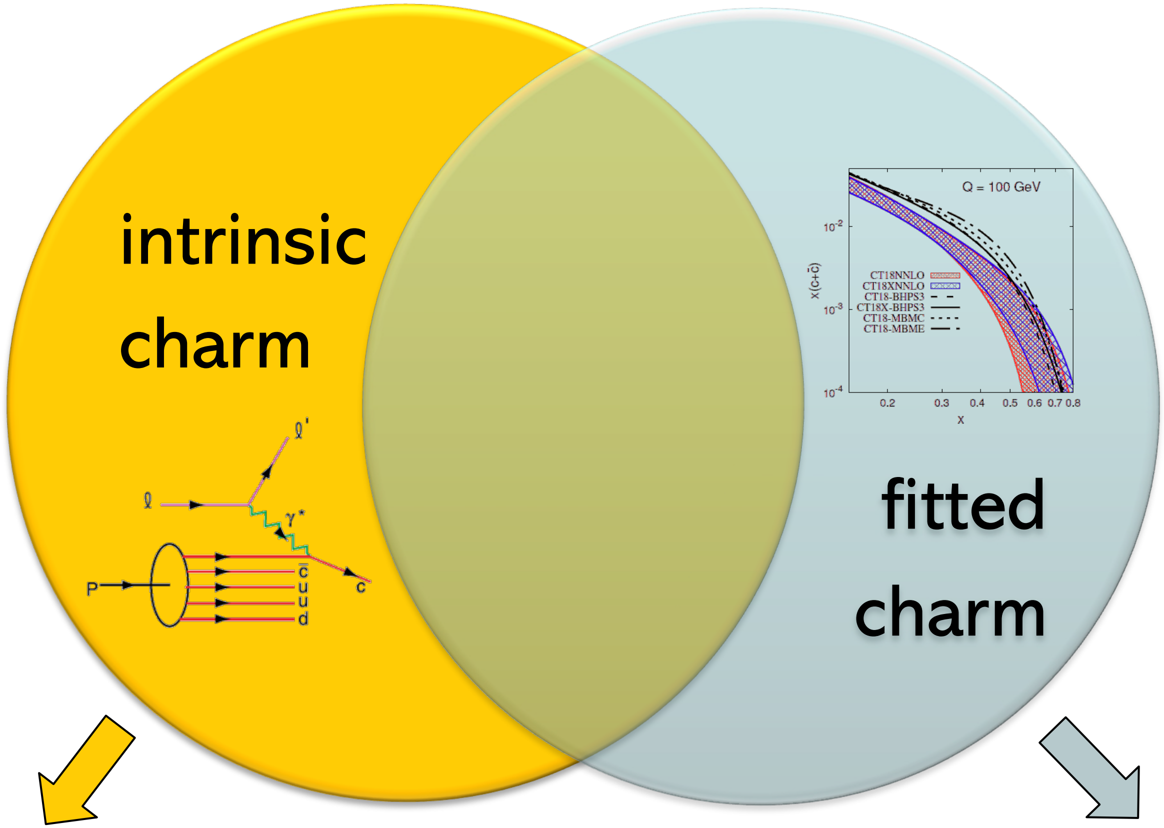
Part 1. Terminology

Do global PDF fits constrain intrinsic charm?

"Fitted charm" is a more direct term to describe the charm PDF found in the global QCD fit

Analog: the fitted charm mass





- The concept of nonperturbative methods
- Can refer to a component of the hadronic Fock state or the type of the hard process
- Predicts a typical enhancement of the charm PDF at $x \gtrsim 0.2$

Connection?

- A charm PDF parametrization at scale $Q_0 \approx 1$ GeV found by global fits [CT, NNPDF, …]
- Arises in perturbative QCD expansions over α_s and operator products
- May absorb process-dependent or unrelated radiative contributions

PDF fits may include a ``fitted charm'' PDF

``Fitted charm'' = ``higher-twist charm''
+ other (possibly not universal)
higher O(α_s) / higher power terms

QCD factorization theorem for DIS structure function $F(x, Q)$ [Collins, 1998]:

All α_s orders:

$$F(x,Q) = \sum_{a=0}^{N_f} \int_x^1 \frac{d\xi}{\xi} \mathcal{C}_a \left(\frac{x}{\xi}, \frac{Q}{\mu}, \frac{m_c}{\mu}; \alpha(\mu) \right) f_{a/p}(\xi, \mu) + \mathcal{O}(\Lambda^2/m_c^2, \Lambda^2/Q^2).$$

The PDF fits implement this formula up to (N)NLO ($N_{ord} = 1$ or 2):

PDF fits:

$$F(x,Q) = \sum_{a=0}^{N_f} \int_x^1 \frac{d\xi}{\xi} \mathcal{C}_a^{(N_{ord})} \left(\frac{x}{\xi}, \frac{Q}{\mu}, \frac{m_c}{\mu}; \alpha(\mu) \right) f_{a/p}^{(N_{ord})}(\xi, \mu).$$

The leading-power charm PDF component cancels at $Q \approx m_c$ up to a higher order

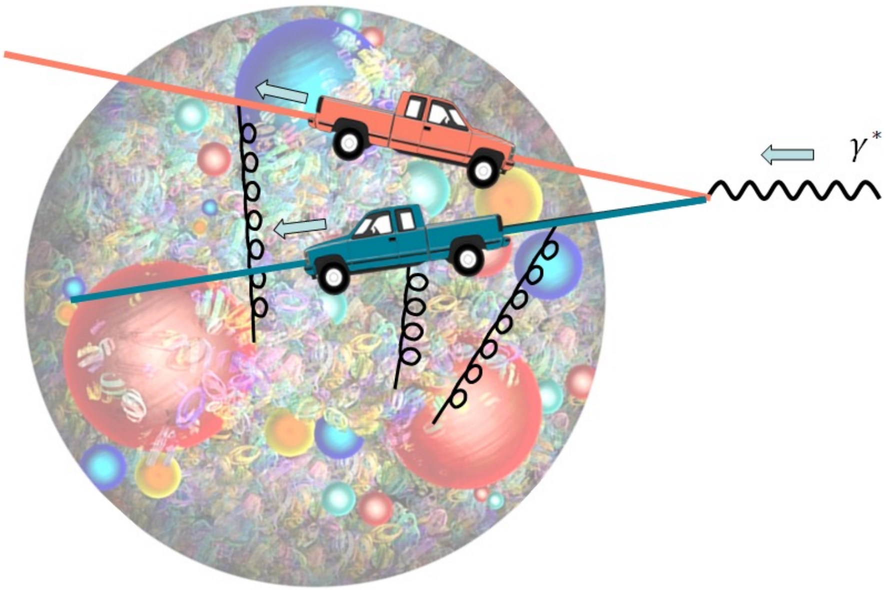
The 'fitted charm component' may approximate for missing terms of orders α_s^p with $p > N_{ord}$, or Λ^2/m_c^2, or Λ^2/Q^2

What is nonperturbative in the "nonperturbative charm"?

In perturbative QFT, extrinsic and intrinsic production mechanisms differ in topology of diagrams, not the strength of the coupling

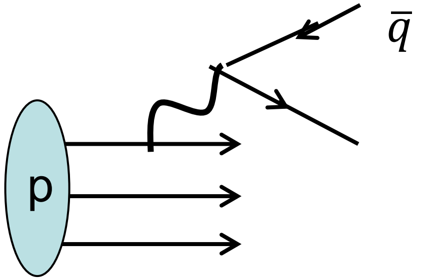
Intrinsic contributions exist in QED, weakly coupled theories

In proton scattering, the intrinsic $c\bar{c}$ contribution emerges through power-suppressed diagrams with more than 1 gluon connection to the asymptotic Fock state

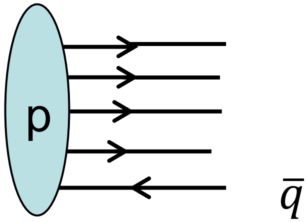


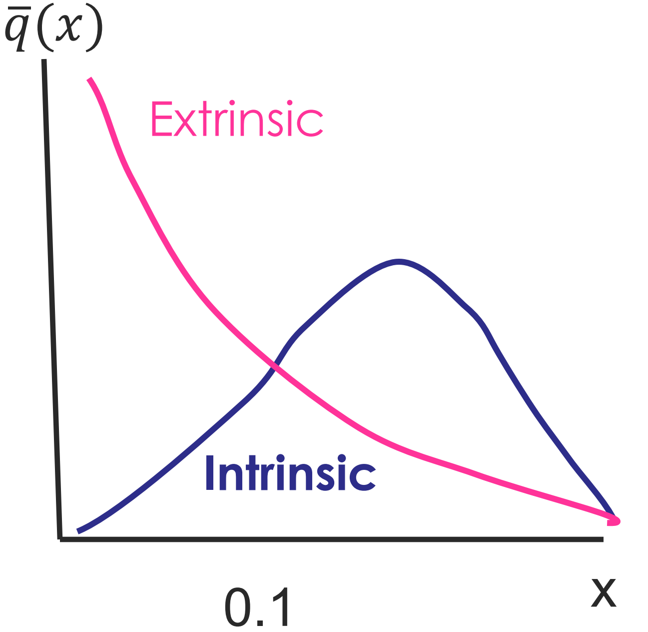
In nonperturbative models:

“Extrinsic” sea
[maps onto leading-power sea production from light flavors]



“Intrinsic” sea (excited Fock nonpert. states; beyond the leading-power production)





In perturbative QFT:

Intrinsic Chevrolets at the SSC

Stanley J. Brodsky (SLAC), John C. Collins (IIT, Chicago and Argonne), Stephen D. Ellis (Washington U., Seattle), John F. Gunion (UC, Davis), Alfred H. Mueller (Columbia U.)

Aug, 1984

10 pages

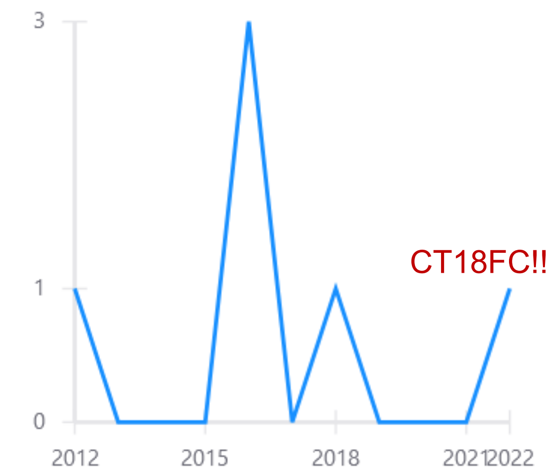
Part of DESIGN AND UTILIZATION OF THE SUPERCONDUCTING SUPER COLLIDER. PROCEEDINGS, 1984 SUMMER STUDY, SNOWMASS, USA, JUNE 23 - JULY 13, 1984

Proceedings of: 1984 DPF Summer Study on the Design and Utilization of the Superconducting Super Collider (SSC) (Snowmass 84), 227

Report number: DOE/ER/40048-21 P4, SLAC-PUB-15471

pdf links cite claim reference search 6 citations

Citations per year



An intrinsic Chevrolet is attached to the target hadron by ≥ 2 gluon propagators

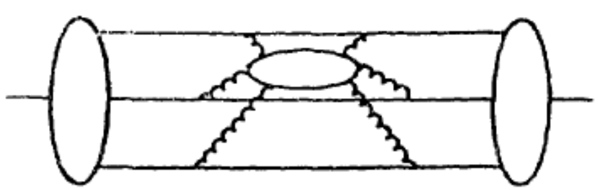


Fig. 1. Example of intrinsic heavy quark contribution to the proton wave function in QCD.

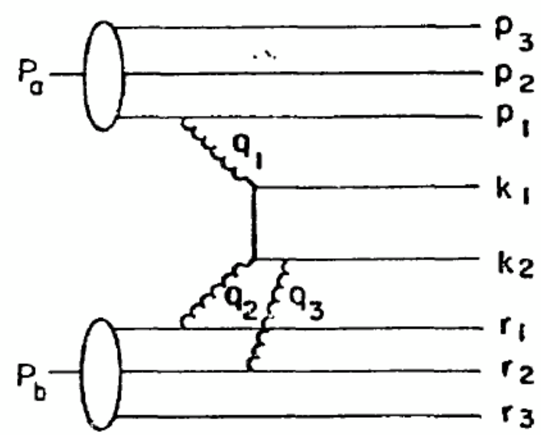


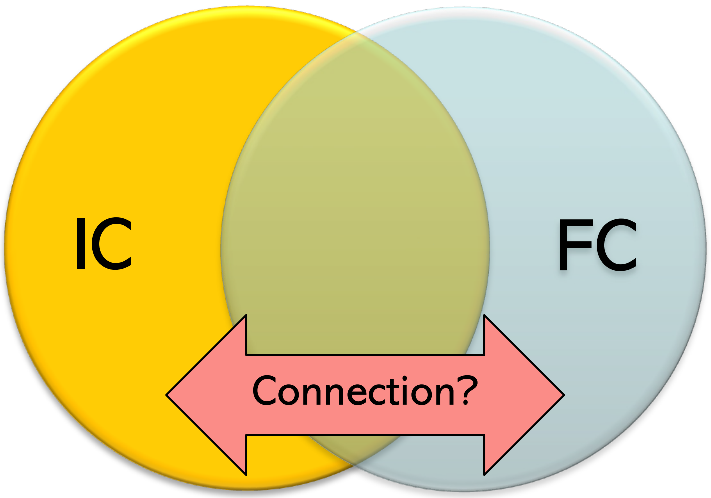
Fig. 3. Typical diagram for heavy quark production with extra gluon corresponding to intrinsic production.

IC is either process-dependent or scheme-dependent

A persistent terminological and conceptual ambiguity:
Is IC a type of a QCD observable or a nonperturbative QCD function?

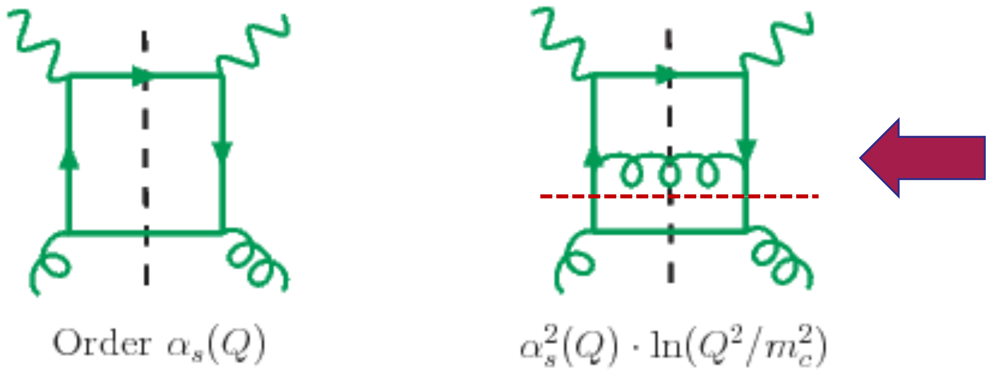
If an observable, it receives process-dependent radiative contributions.
⇒ **Process dependence**

If a nonperturbative function, it can be defined in many ways.
⇒ **Scheme dependence**



A twist-4 contribution in HERA DIS charm production ($\subset$ “intrinsic charm”)

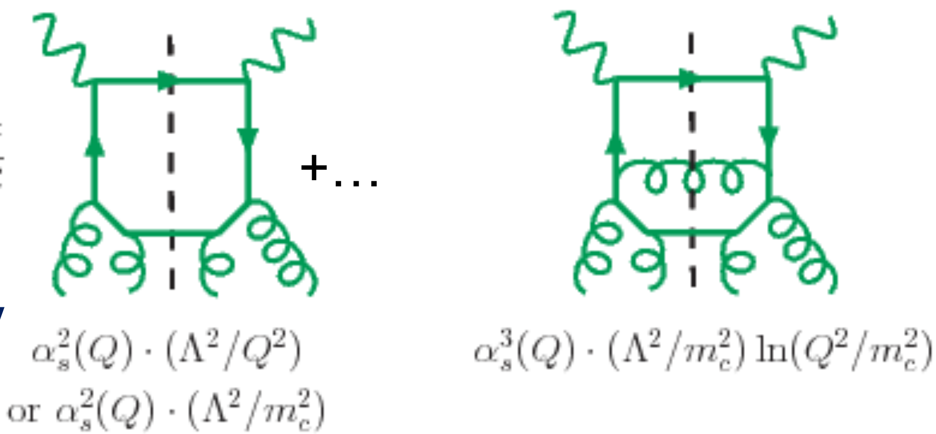
Twist-2
$\gamma^* g \to c\bar{c}$



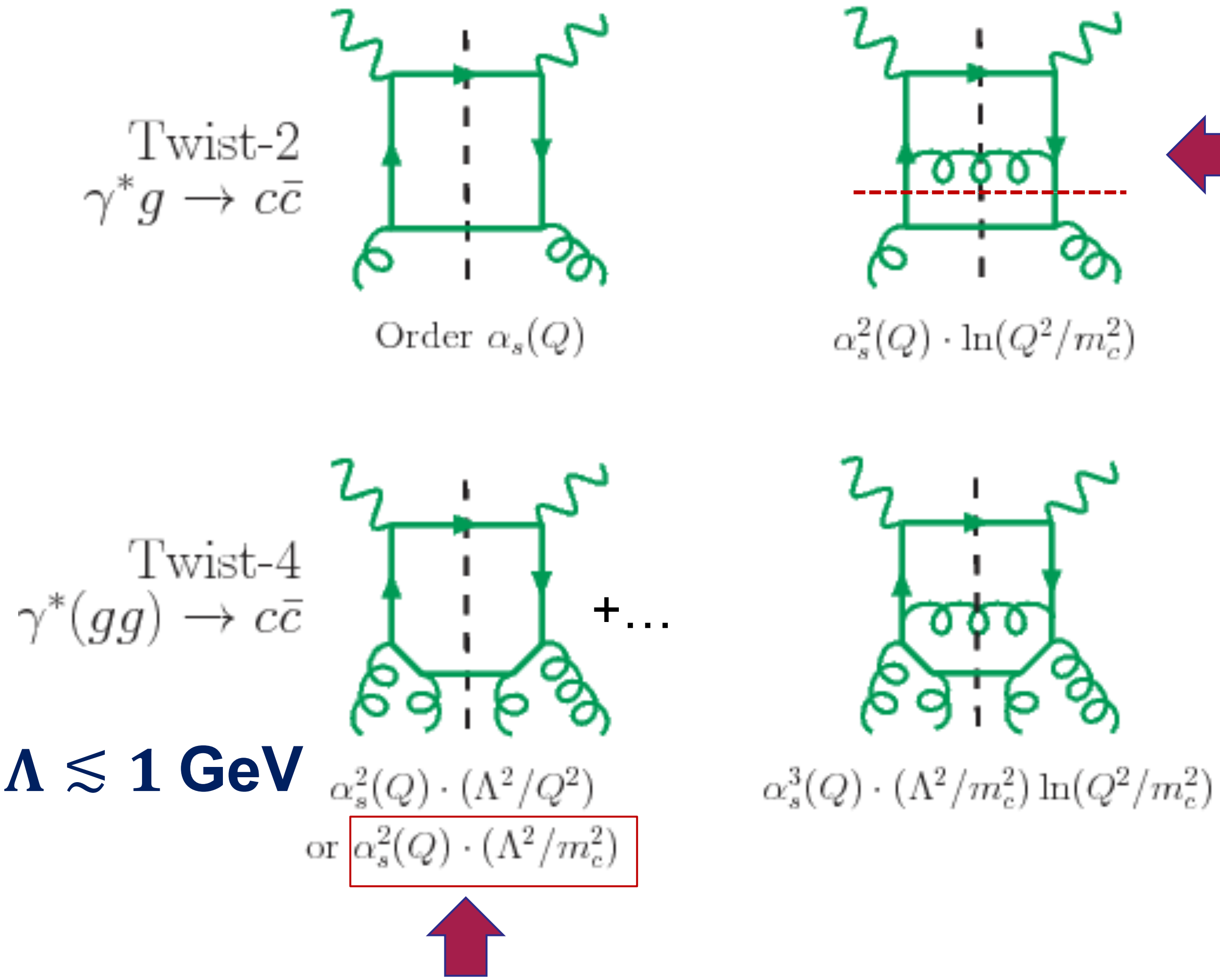
A ladder; must be resummed in $c(x, Q)$ in the $N_f = 4$ scheme at $Q^2 \gg m_c^2$; e.g., in the ACOT scheme

Twist-4
$\gamma^*(gg) \to c\bar{c}$

$\Lambda \lesssim 1$ GeV



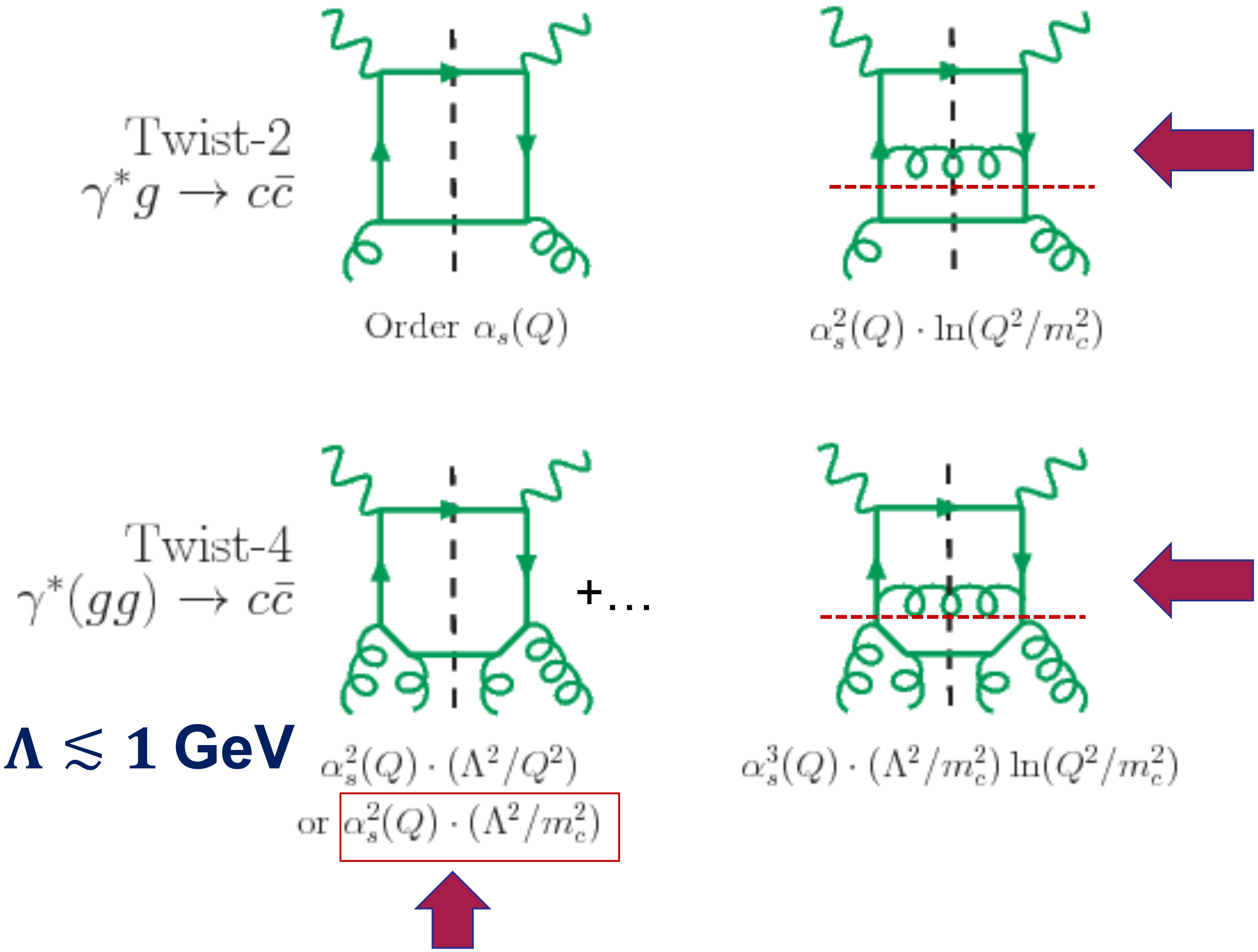
A twist-4 contribution in HERA DIS charm production ($\subset$ "intrinsic charm")



A ladder; must be resummed in $c(x, Q)$ in the $N_f = 4$ scheme at $Q^2 \gg m_c^2$; e.g., in the ACOT scheme

Can be of order 10% of the twist-2 α_s^2 term

A twist-4 contribution in HERA DIS charm production (⊂ "intrinsic charm")



A ladder; must be resummed in $c(x,Q)$ in the $N_f = 4$ scheme at $Q^2 \gg m_c^2$; e.g., in the ACOT scheme

The ladder subgraphs can be resummed as a part of $c(x,Q)$ in the $N_f = 4$ scheme at $Q^2 \gg m_c^2 > \Lambda^2$;

contributes to the boundary condition for $c(x,Q_0)$ at $Q_0 \approx m_c$;

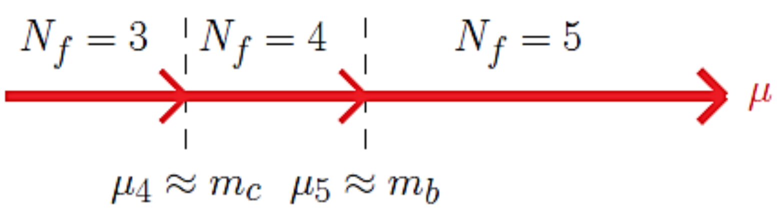
obeys twist-2 DGLAP equations.

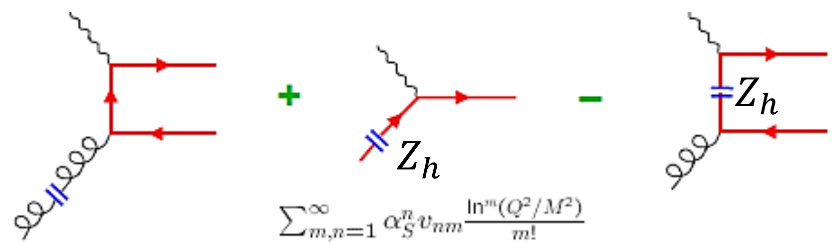
Can be of order ~10% of the twist-2 α_s^2 term

Factorization scheme dependence of the FC

In global PQCD fits, factorization is performed in a **general-mass variable flavor number scheme (GM VFNS)**.

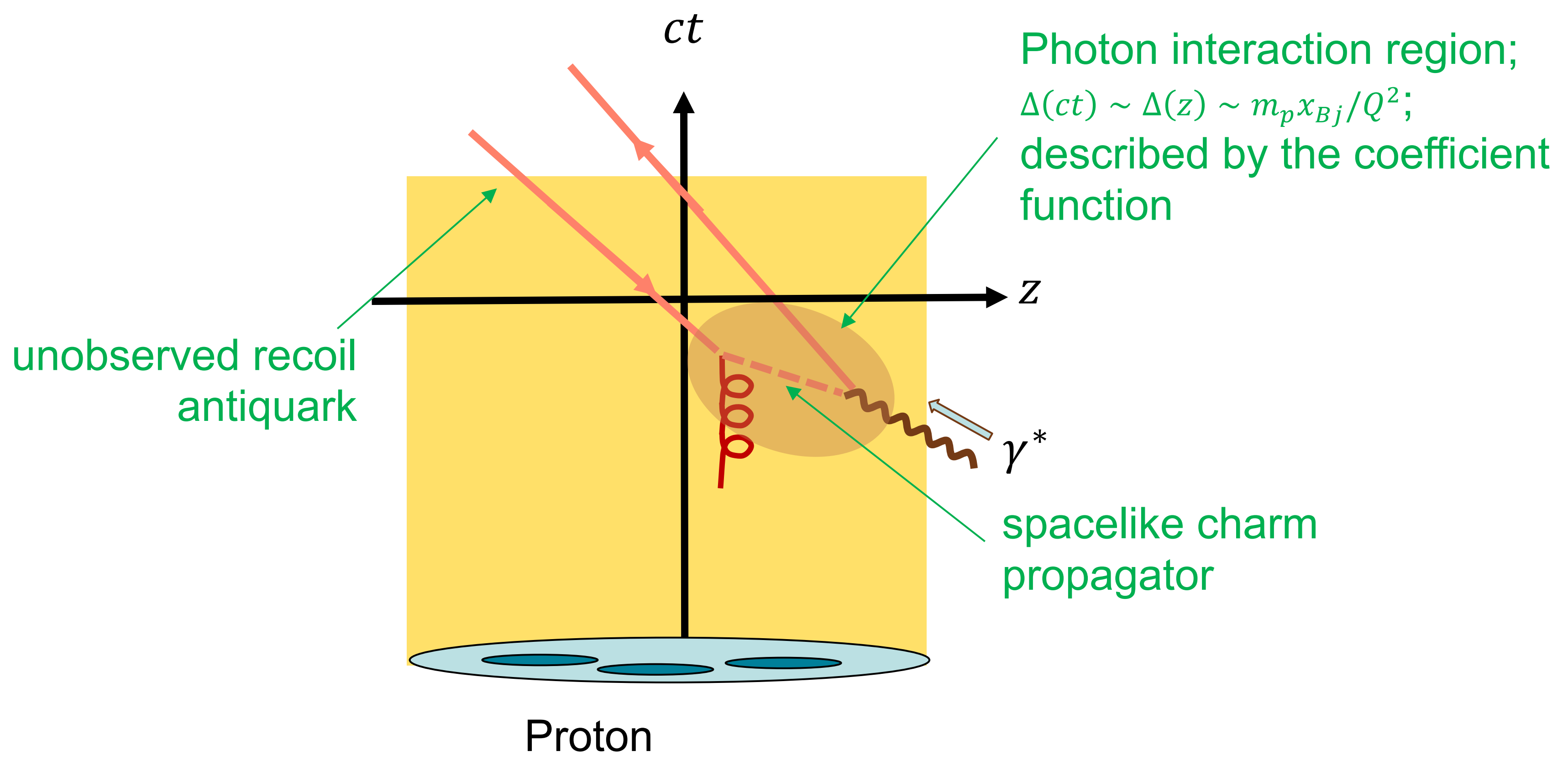
- 1 PI Green functions with charm lines are renormalized in the $\overline{\mathrm{MS}}$ scheme for $\mu > \mu_4$ and in the zero-momentum subtraction scheme for $\mu < \mu_4$.
- μ_4 is often set equal to m_c -- either the pole mass or $\overline{\mathrm{MS}}$ mass.
- Process-specific coefficient functions are factorized from universal PDFs. This factorization is under control only for the leading-power FC.
- **In all factorization schemes, the flavor-excitation coefficient functions involve a kinematic approximation [realized by operator Z_h] that puts the external heavy quark leg on the mass shell and optionally omits irrelevant m_c^2/Q^2 terms.**





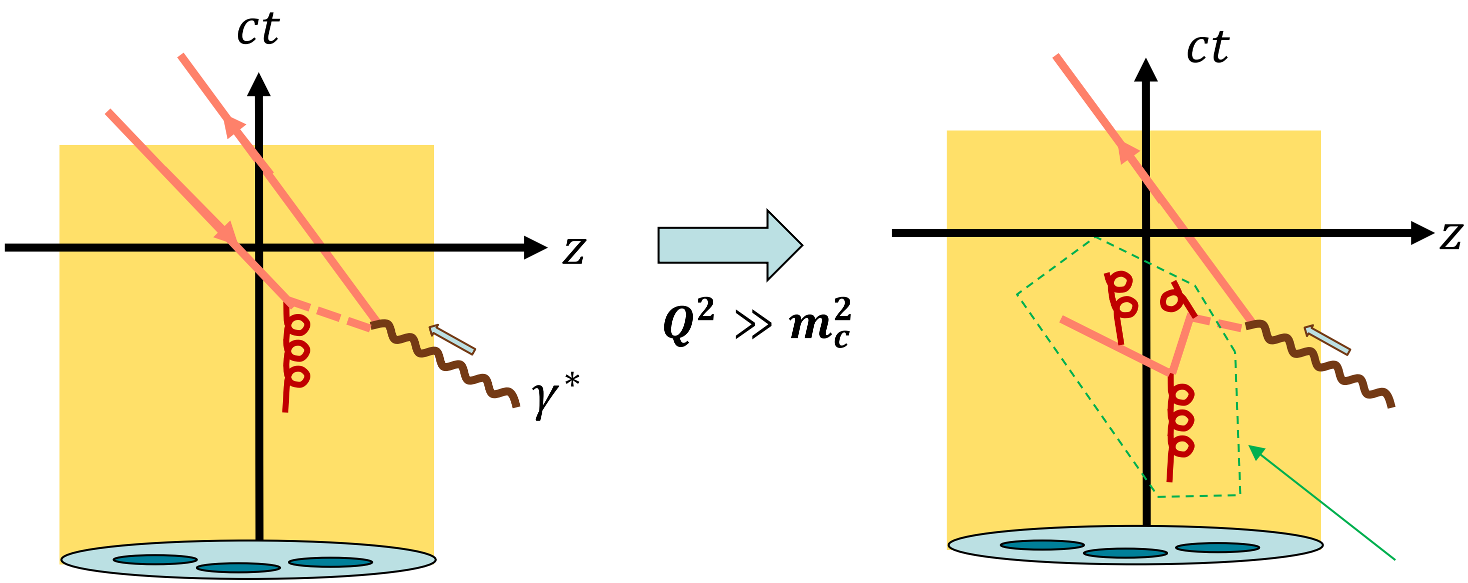
DIS in the rest frame of the proton, space-time diagram

Extrinsic production



DIS in the rest frame of the proton, leading kinematic configurations

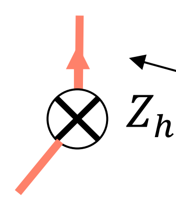
Extrinsic production



+ small terms

Proton

collinear splittings; approximated and resummed as

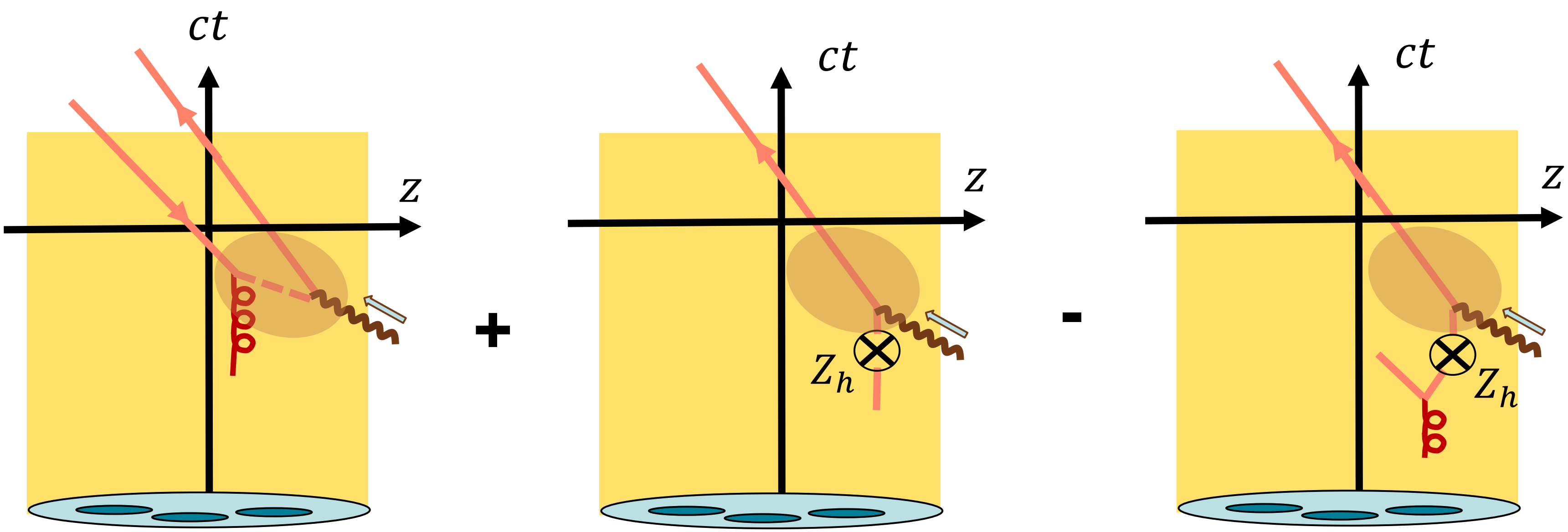


on shell

Operator Z_h applies the kinematic approximation to the coefficient function

DIS in the rest frame of the proton, leading kinematic configurations

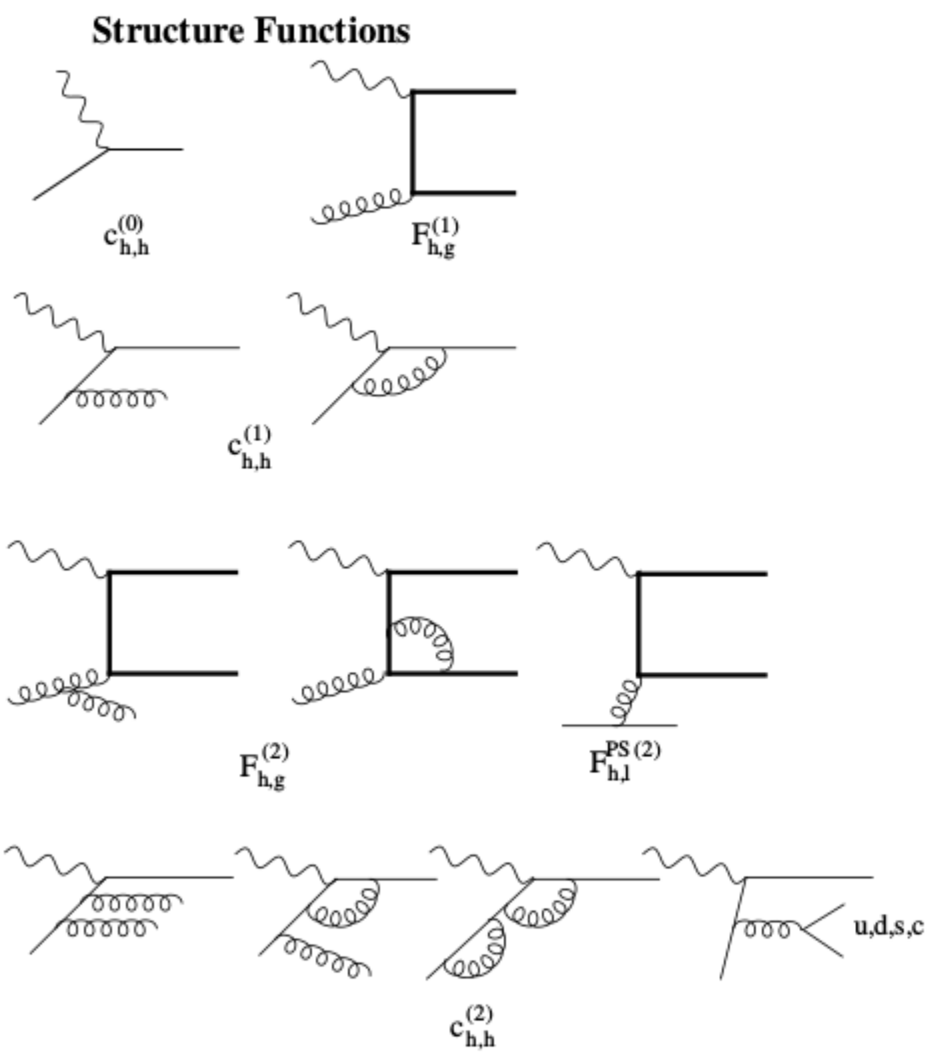
Extrinsic production, full leading order

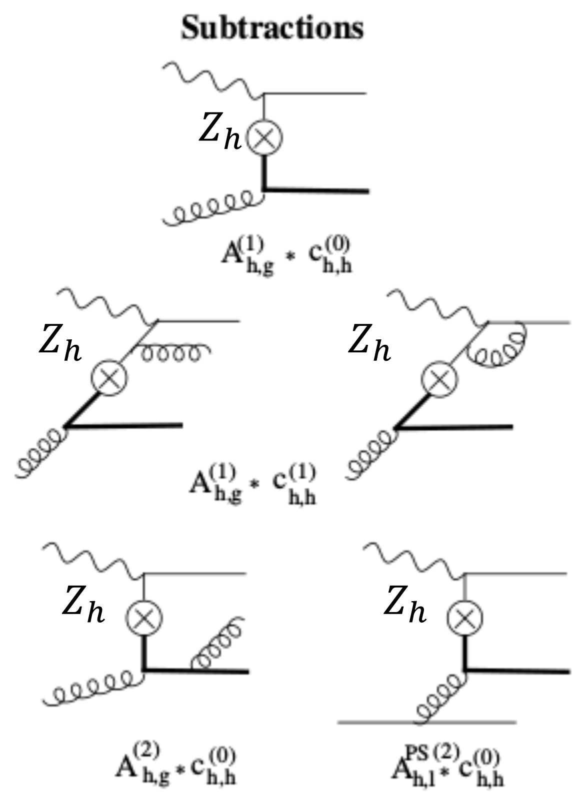


Operator Z_h applies the kinematic approximation to the coefficient function

Z_h dependence cancels to a higher order. This algorithm is continued to higher orders.

Twist-2: factorization for DIS in SACOT-MPS scheme up to NNLO

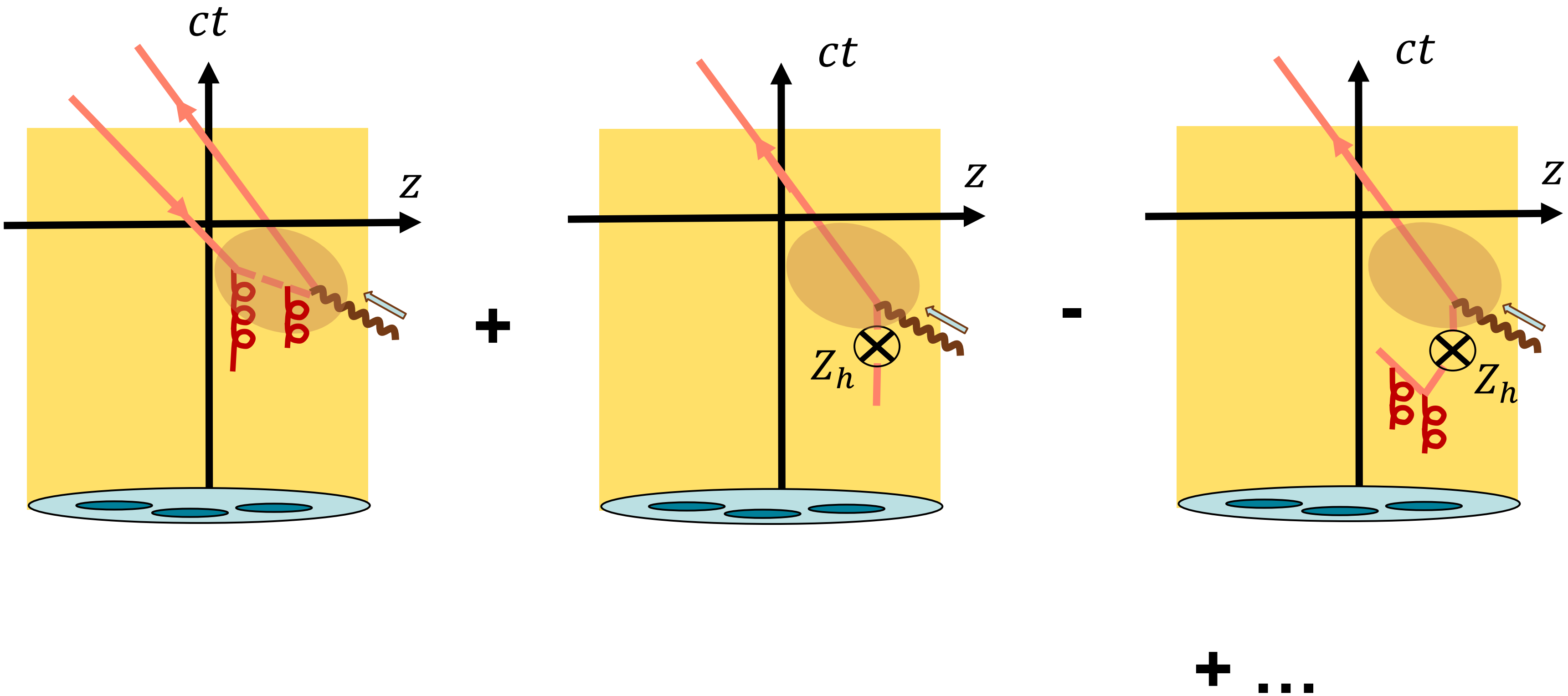




Leading-power radiative contributions to neutral-current DIS charm production in the CTEQ-TEA NNLO analysis

DIS in the rest frame of the proton, leading kinematic configurations

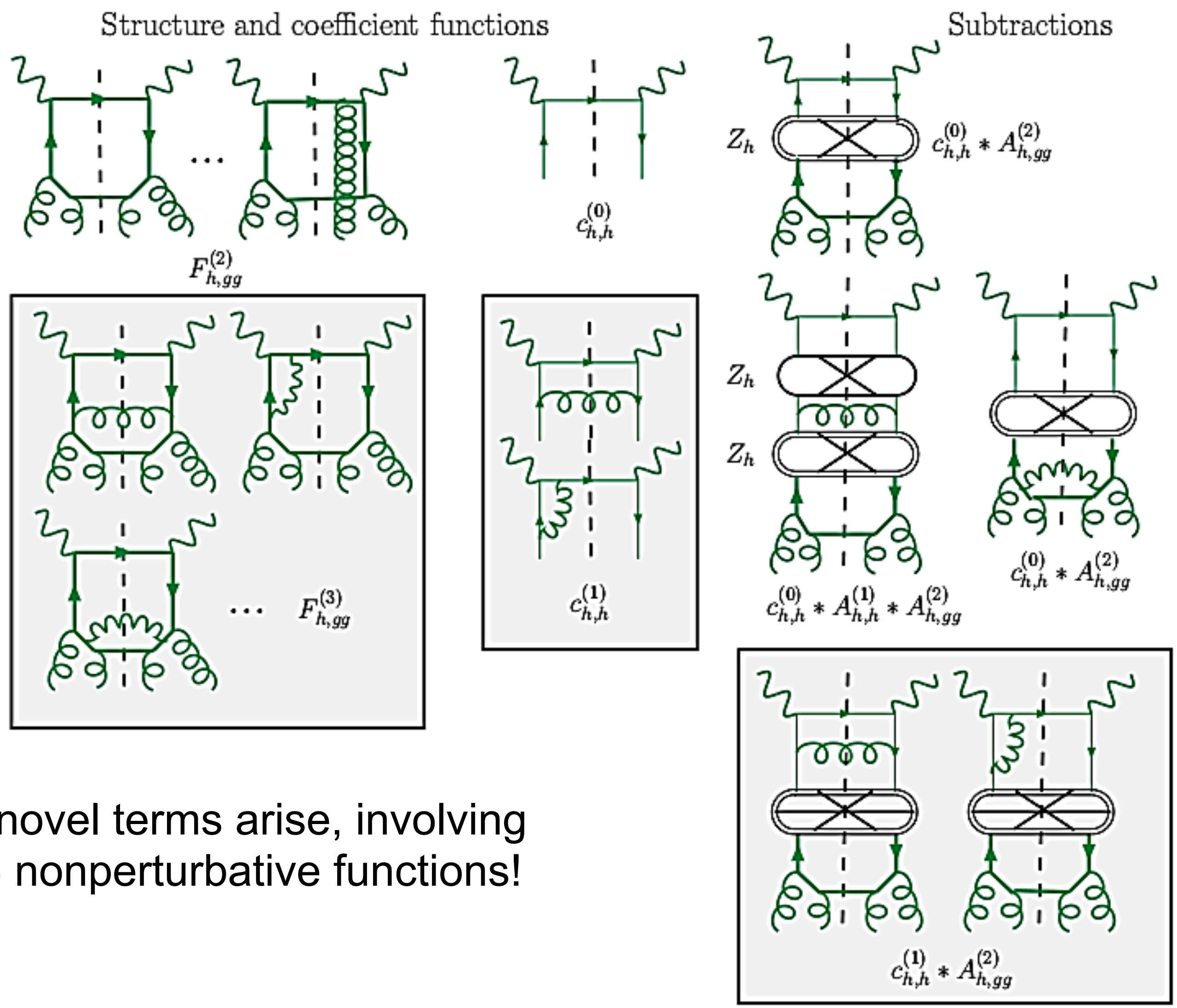
Intrinsic production: attach more gluons between the proton and charm



ACOT-like factorization for twist-4 charm contributions (an example)

α_s^2 :

α_s^3 :

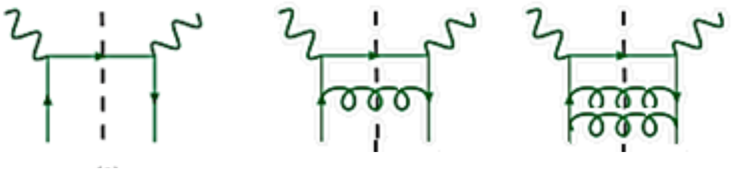


Many novel terms arise, involving twist-4 nonperturbative functions!

Fitted charm contributions, practical implementation in CT14 and CT18

Keep only $c_{h,h} \otimes f_h$:

Discard $C^{(k)}_{h,gg} \otimes f_{gg}$, etc.



In the absence of full computation, we (and other groups) make the simplest approximation:

$$F_{FC}(x, Q_0) = [c_{h,h} \otimes f^{FC}_{c/p}](x, Q_0)$$

$c_{h,h}$ **is the twist-2 charm DIS coefficient function** introduced to factorize the twist-4 ladder terms; defined according to the SACOT-MPS scheme

Start with $N_f = 3$ at $\mu_0 = m_c - \epsilon$, evolve to $\mu > \mu_0$ by incrementing N_f to 4 and 5

FC is compatible with any version of the ACOT scheme (cf. arXiv:1707.00657).

Flavor-excitation coefficient functions of these schemes differ by terms of $O(m_c^2/Q^2)$. Their overall differences are of $O(\Lambda^2/Q^2)$, i.e., within the accuracy of the factorization theorem.

SACOT
= Simplified Aivazis-Collins-Olness-Tung scheme

ACOT, PRD 50 3102 (1994); Collins, PRD 58 (1998) 094002;
Kramer, Olness, Soper, PRD (2000) 096007; Tung, Kretzer, Schmidt, J.Phys. G28 (2002) 983

The default heavy-quark scheme of CTEQ-TEA PDFs
Implementation is based upon, and closely follows, the proof of QCD factorization for DIS with massive quarks *(Collins, 1998)*

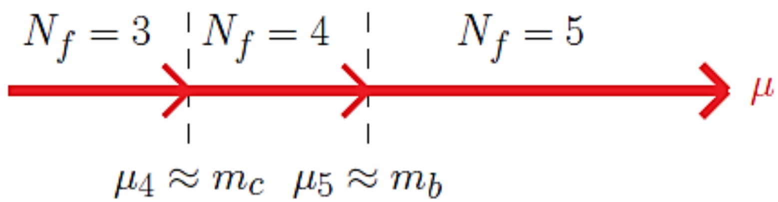
MPS/χ prescription ≡ Kinematic matching based on **m**assive **p**hase **s**pace to improve perturbative convergence near HF production threshold

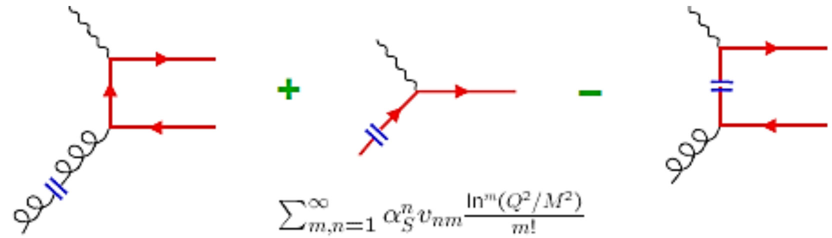
Applied
- to NNLO in NC DIS (Guzzi et al.; arXiv:1108.5112)
- to NLO in heavy-flavor hadroproduction using MCFM

(Xie, Campbell, Nadolsky, 2019-2020)

In CC dimuon DIS, the NNLO correction is relatively small (Berger, Gao, arXiv:1710.04258)

SACOT-MPS scheme: advantages

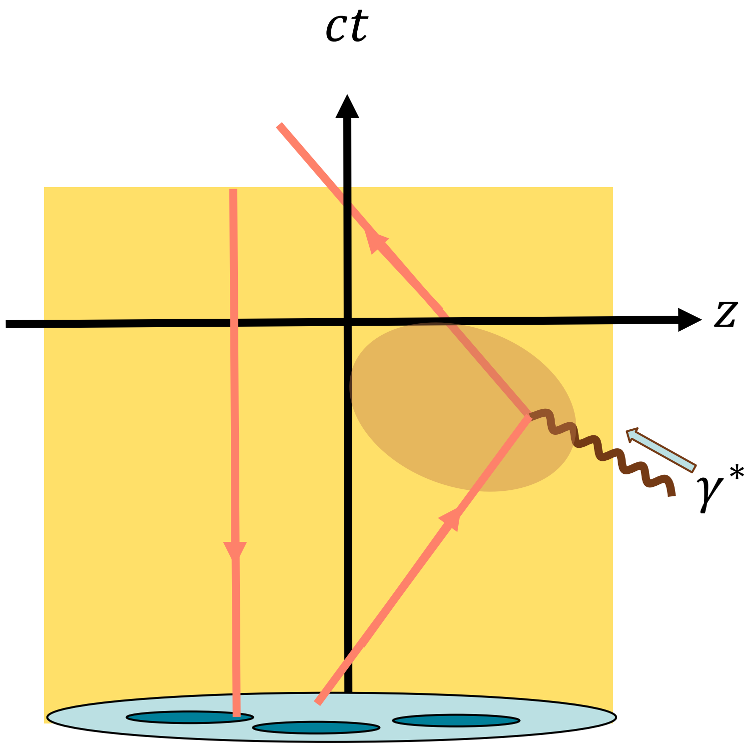




- A general-mass variable-flavor number (GM-VFN) scheme
  - Perturbatively convergent at all factorization scales $\mu \gtrsim m_Q$
  - reduces to the zero-mass $\overline{MS}$ scheme at $\mu^2 \gg m_Q^2$, without additional renormalization
  - reduces to the fixed-flavor number scheme at $\mu^2 \sim m_Q^2$
- Relatively simple
  - One value of N_f (and one PDF set) in each μ range
  - Sets $m_Q = 0$ in $|M|^2$ with incoming c or b
  - Straightforward matching based on kinematical rescaling

Other common heavy-quark schemes: FONLL, TR', SACOT-m_T,…

What about the pentaquark state?



A long-lived $|uudc\bar{c}\rangle$ state could strongly mix with the ground $|uud\rangle$ state. [No evidence exists for such mixed state.]

For this contribution, neither available factorization scheme is kinematically exact:

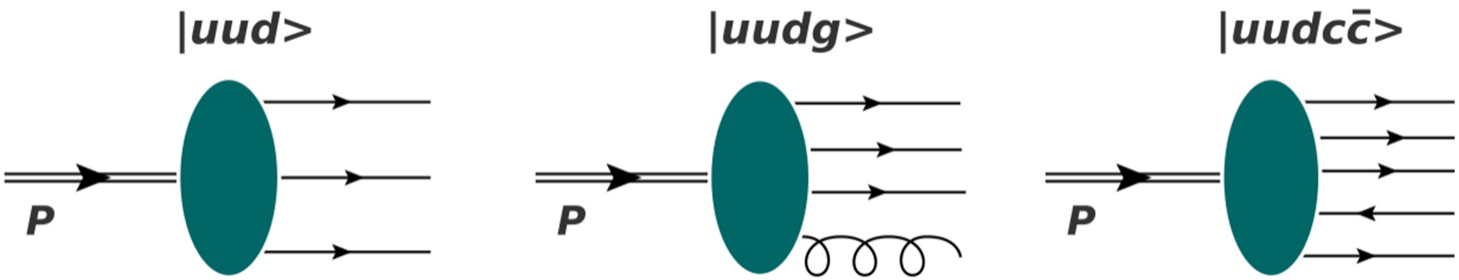
- There is a target-mass correction of order of mass of $|uudc\bar{c}\rangle$.
- There is the 2^{nd} massive antiquark in the proton remnant. [The SACOT-MPS scheme corrects for it.]

Part 2. IC models and CT18 FC global analysis

nonperturbative QCD can generate a low-scale charm PDF

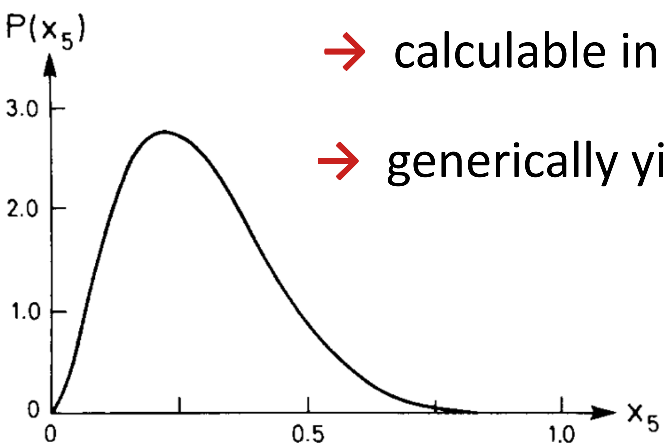
Fock expansion

Brodsky, Hoyer, Peterson, Sakai (**BHPS**); Phys. Lett. **B93** (1980) 451.



IC PDF: transition matrix element, $|\mathrm{proton}\rangle \to |uudc\bar{c}\rangle$

$$P(p \to uudc\bar{c}) \sim \left[M^2 - \sum_{i=1}^{5} \frac{k_{\perp i}^2 + m_i^2}{x_i} \right]^{-2}$$



→ calculable in old-fashioned perturbation theory; **scalar** field theory

→ generically yields valence-like shape; governed by charm masses

$$m_c = m_{\bar{c}} \implies c^{\mathrm{BHPS}}(x) = \bar{c}^{\mathrm{BHPS}}(x)$$

alternative but similar representations exist

Blumlein; Phys. Lett. **B753** (2016) 619.

meson-baryon models (MBMs): 5-quark states from hadronic interactions

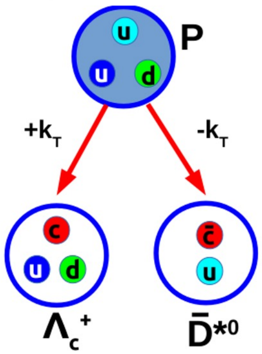
- we implement a framework which *conserves* **spin/parity**

- **nonperturbative** mechanisms are needed to break
  $c(x, Q^2 \leq m_c^2) = \bar{c}(x, Q^2 \leq m_c^2) = 0!$

We build an **EFT** which connects IC to properties of the hadronic spectrum: [TJH, J. T. Londergan and W. Melnitchouk, Phys. Rev. D89, 074008 (2014).]

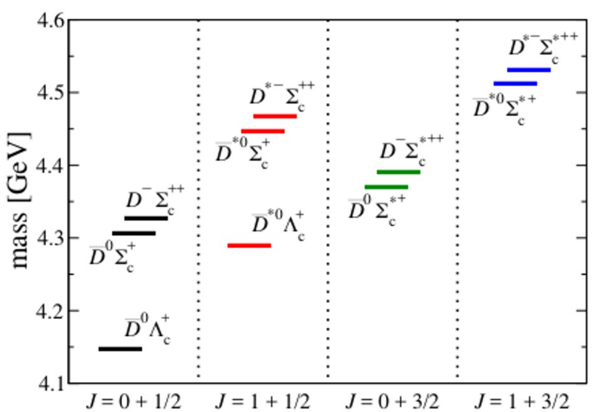
- $|N\rangle \;=\; \sqrt{Z_2}\,|N\rangle_0 \;+\; \sum_{M,B}\int dy\, f_{MB}(y)\,|M(y); B(1-y)\rangle$

$y = k^+/P^+$: k meson, P nucleon



$$c(x) = \sum_{B,M}\left[\int_x^1 \frac{d\bar{y}}{\bar{y}}\, f_{BM}(\bar{y})\, c_B\left(\frac{x}{\bar{y}}\right)\right]$$

- a similar *convolution* procedure may be used for $\bar{c}(x)$ . . .

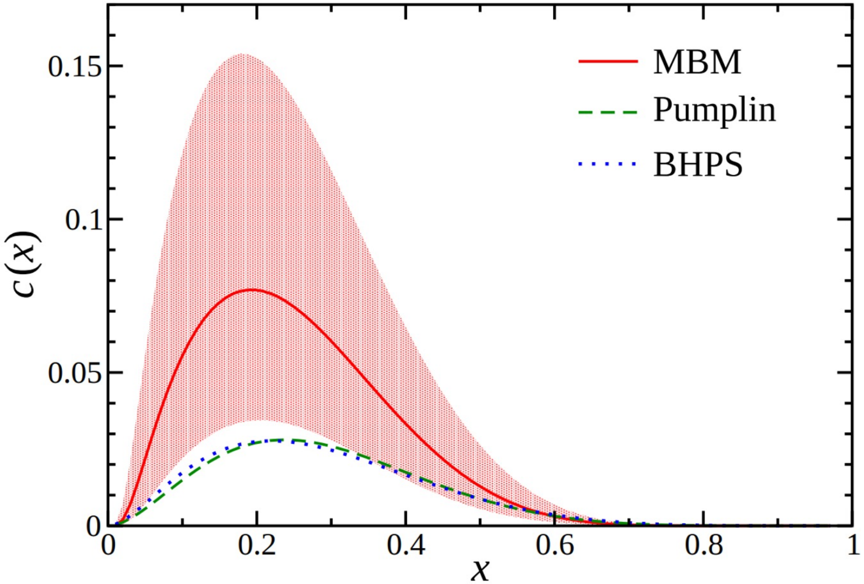


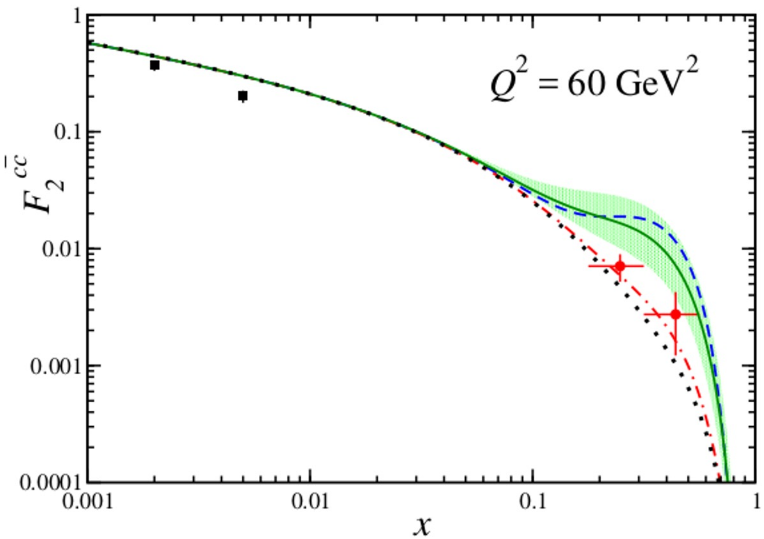
IC (MBM) depends on UV scale parameter, Λ; predicts high-*x* excess

- tune **universal** cutoff $\Lambda = \hat{\Lambda}$ to fit **ISR** $pp \to \Lambda_c X$ collider data

multiplicities, **momentum sum**:

$$\langle n \rangle^{(\mathrm{charm})}_{MB} = 2.40\% \,{}^{+2.47}_{-1.36}; \qquad P_c := \langle x \rangle_{\mathrm{IC}} = 1.34\% \,{}^{+1.35}_{-0.75}$$





$$F_2^{c\bar{c}}(x, Q^2) = \frac{4x}{9}\left[c(x, Q^2) + \bar{c}(x, Q^2)\right]$$

→ evolve to EMC scale, $Q^2 = 60\ \mathrm{GeV}^2$

low-x H1/ZEUS data check *massless* **DGLAP** evolution

IC models and **formal QCD**

models simulate nucleon wave function; aim to *mimic* nonpert QCD

- → bound-state structure driven by constituent-quark masses
- → connect to SU(4) flavor-symm breaking (in meson-baryon models [MBMs])

also, light-front models may relate PDFs, form factors, …

ex.: Hobbs, Alberg, Miller, PRD**96** (2017) 7, 074023; Sufian et al, PLB808 (2020) 135633.

$$|\Psi_P^\lambda(P^+, \boldsymbol{P}_\perp)\rangle = \frac{1}{16\pi^3} \sum_{q=c,\bar{c}} \int \frac{dx\, d^2\boldsymbol{k}_\perp}{\sqrt{x(1-x)}}\, \psi^\lambda_{q\lambda_q}(x, \boldsymbol{k}_\perp)\, |q; xP^+, x\boldsymbol{P}_\perp + \boldsymbol{k}_\perp\rangle$$

BUT: unclear mapping, IC models → systematically-improvable QCD calc.

- → based on *truncated* Fock-state or similar wave function expansions
- → no obvious mapping onto factorization theorems
- → ambiguity regarding fact. scale, μ, in IC models

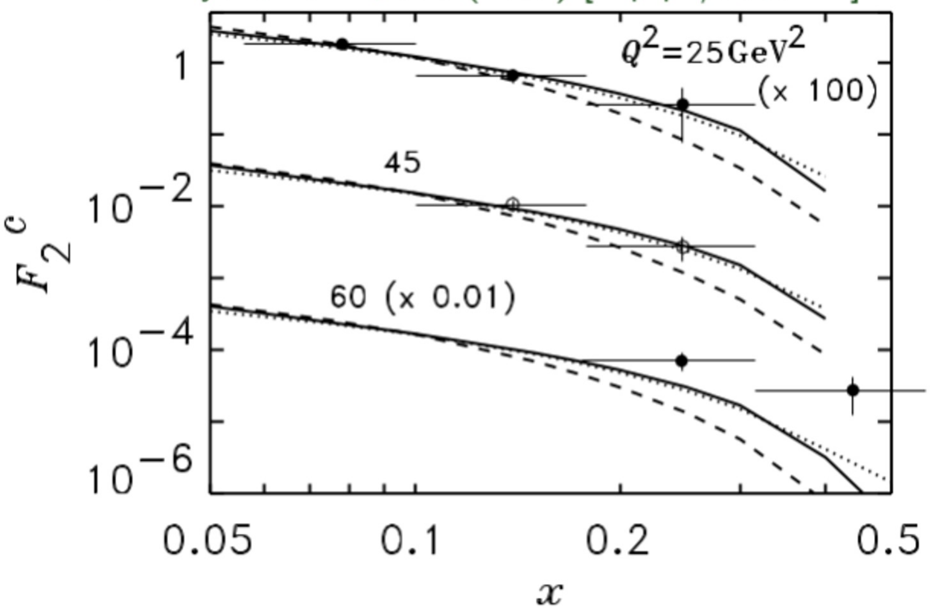
again, intrinsic charm (IC) ≠ fitted charm (FC) extracted in PDF analyses

few expts with ‘smoking gun’ sensitivity to FC; but **EMC data** (?)

historically, charm structure function data, $F_2^{c\bar{c}}$, from EMC were suggestive

J. J. Aubert *et al*. (EMC), NPB**213** (1983) 31–64.

F. M. Steffens, W. Melnitchouk and A. W. Thomas, Eur. Phys. J. C **11**, 673 (1999) [hep-ph/9903441].



- → hint of high-x excess in select Q^2 bins
- → data were analyzed only at LO
- → show anomalous Q^2 dependence
- → EMC data fit poorly in CT14 IC study

we do not include EMC in CT18 FC

CT14 IC, arXiv: 1707.00657.

Candidate NNLO PDF fits	$\chi^2/N_{\rm pts}$			
	All Experiments	HERA inc. DIS	HERA $c\bar{c}$ SIDIS	EMC $c\bar{c}$ SIDIS
CT14 + EMC (weight=0), no IC	1.10	1.02	1.26	3.48
CT14 + EMC (weight=10), no IC	1.14	1.06	1.18	2.32
CT14 + EMC in BHPS model	1.11	1.02	1.25	2.94
CT14 + EMC in SEA model	1.12	1.02	1.28	3.46

FC at LHC: *Z+c* suggested as sensitive probe

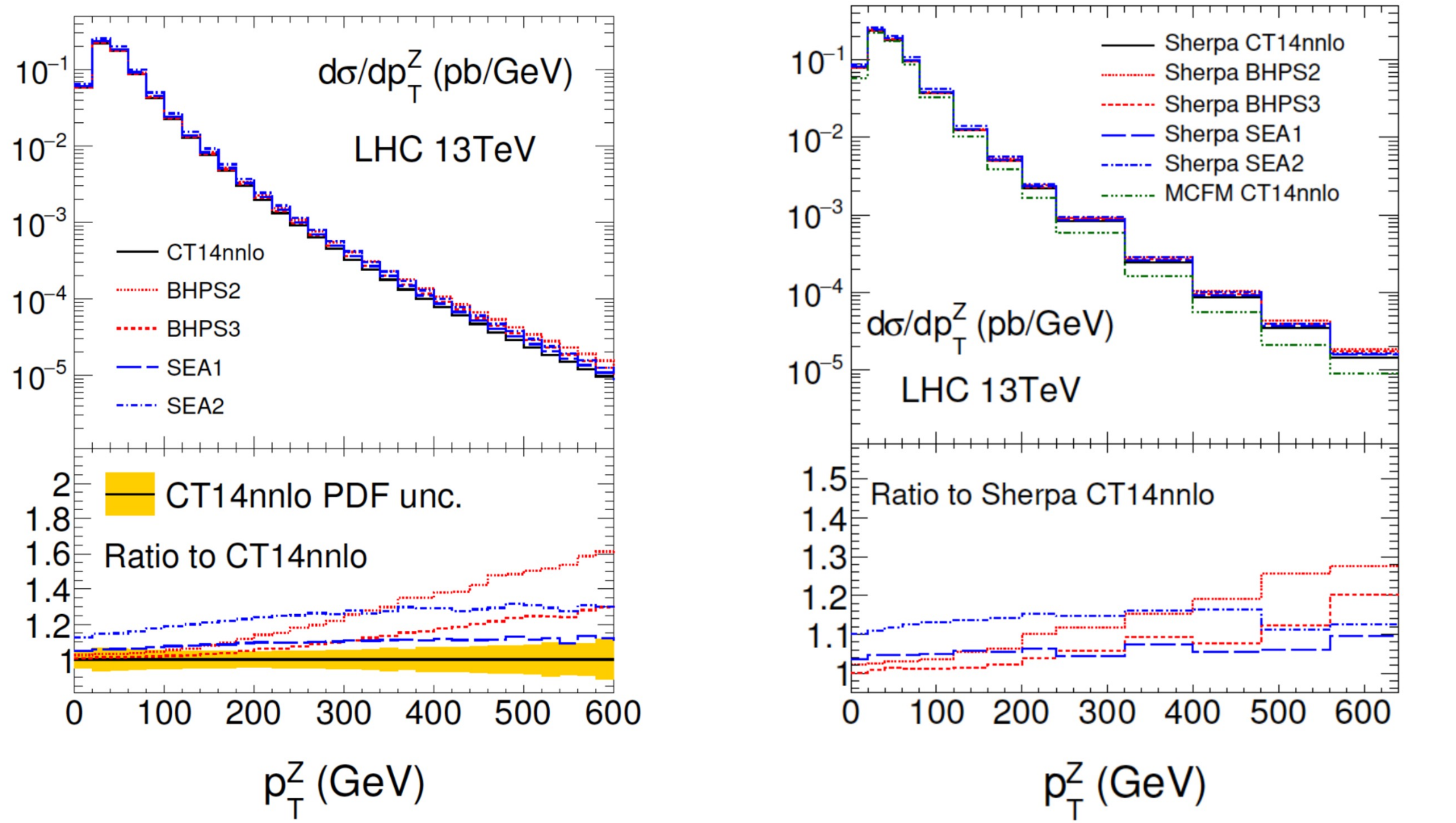
T. Boettcher, P. Ilten, M. Williams, 1512.06666; Bailas, Goncalves, 1512.06007

p_T spectra, rapidity dists nominally sensitive to high-*x* charm PDF

→ parton-shower effects can dampen high-p_T tails

Z+c NLO LHC 13 TeV

[Hou et al., arXiv:1707.00657]

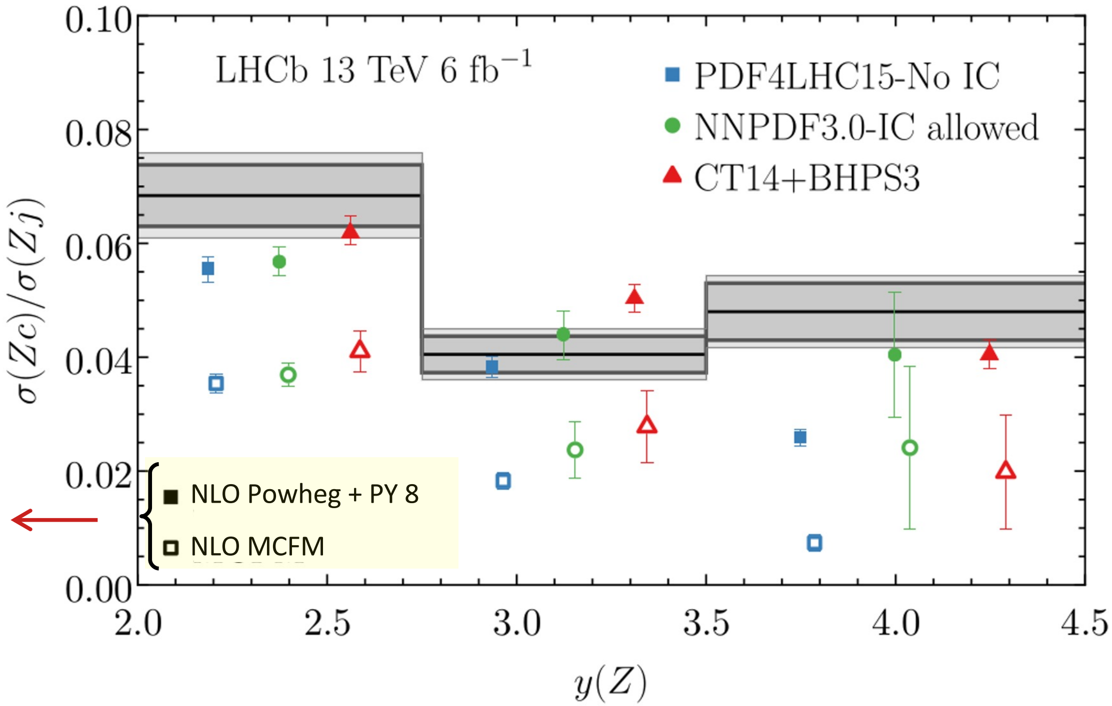


Z+c theory predictions carry sizable uncertainties

2022 LHCb 13 TeV data: (*Z+c*) / (*Z+jet*) ratios; 3 rapidity bins

→ calculated **NLO** cross-section ratio similarly depends on showering, hadronization

R. Aaij, *et al*. (LHCb); arXiv: 2109.08084.



differ due to large FSR correction ←

$$x \sim \frac{Q}{\sqrt{s}} \exp y$$

NNLO calculations recently available, but not implemented in PDF fits

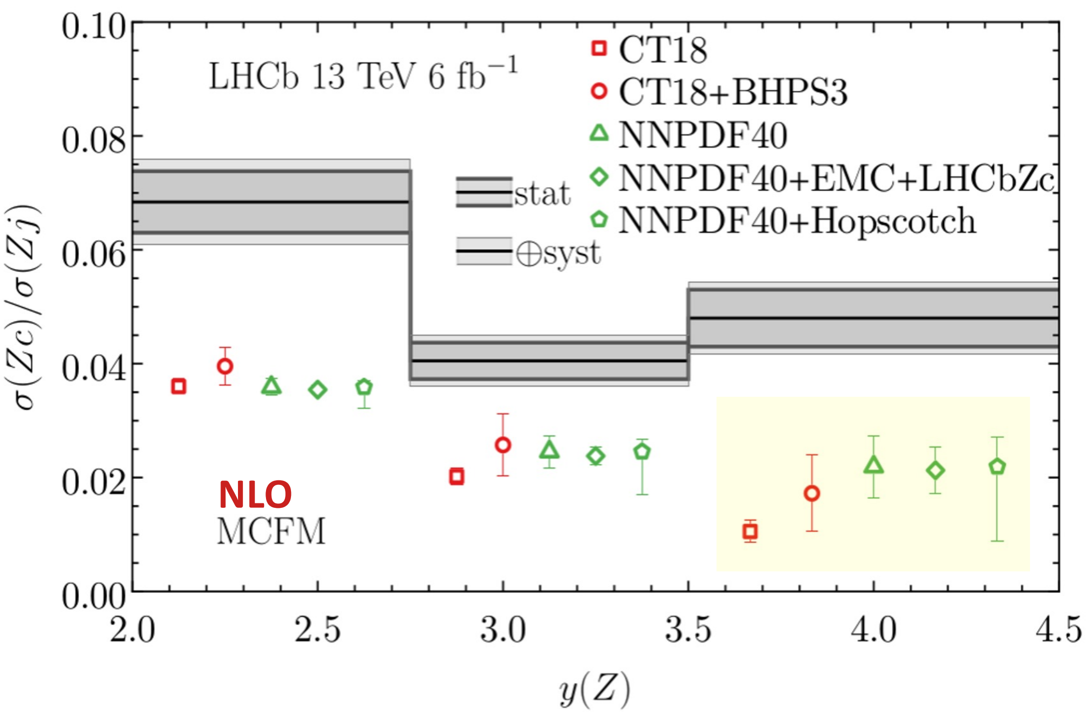
R. Gauld, *et al*.; arXiv: 2005.03016.
M. Czakon, *et al*.; arXiv: 2011.01011.

theory uncertainties currently larger than PDF variations

assuming MCFM at NLO, can vary underlying PDFs, test inclusion of FC

→ FC slightly enhances ratio; not enough to improve agreement with data

R. Aaij, *et al*. (LHCb); arXiv: 2109.08084.



$$x \sim \frac{Q}{\sqrt{s}} \exp y$$

theory accuracy not yet sufficient to leverage expt. precision for PDFs

→ need NNLO theory interface; control over showering, final-state effects

might other HEP experiments be sensitive to FC?

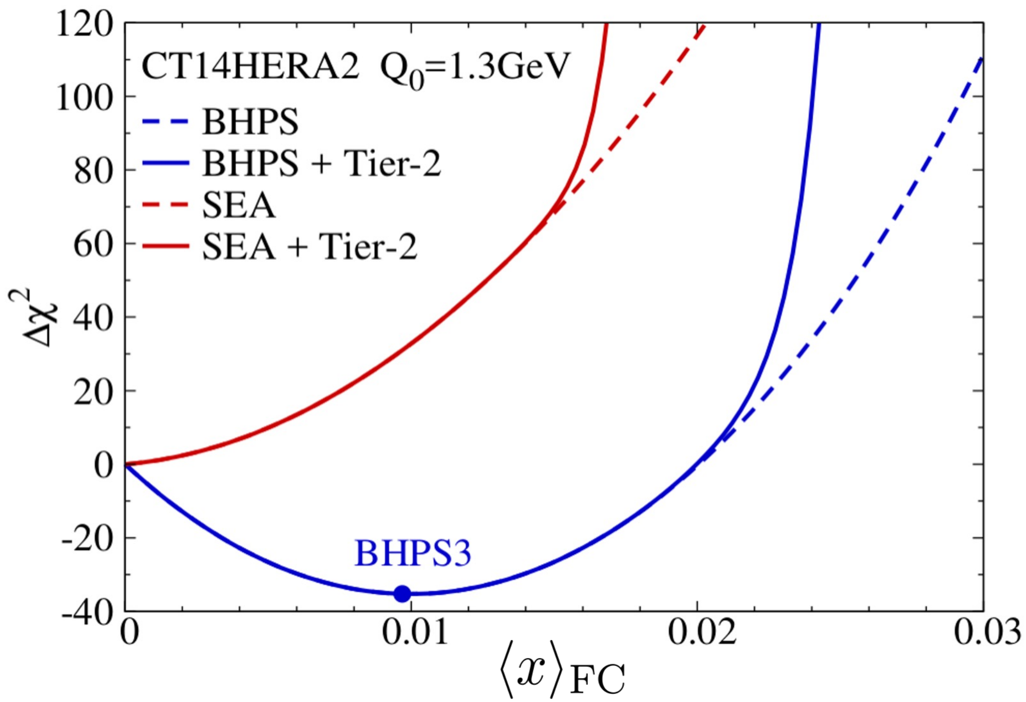
must be assessed using comprehensive global QCD analysis of PDFs

CT performed such an analysis, CT14 IC in arXiv: 1707.00657

→ found $\langle x \rangle_{\mathrm{FC}} < 2\%$, but with large uncertainty consistent with zero FC

$$\langle x \rangle_{\mathrm{FC}} = \int_0^1 dx\, x[c(x, Q_0) + \bar{c}(x, Q_0)]$$

included many details on theory and analysis of IC



since CT14 IC, many LHC measurements have been released; natural to ask if these possess *collective* sensitivity to FC

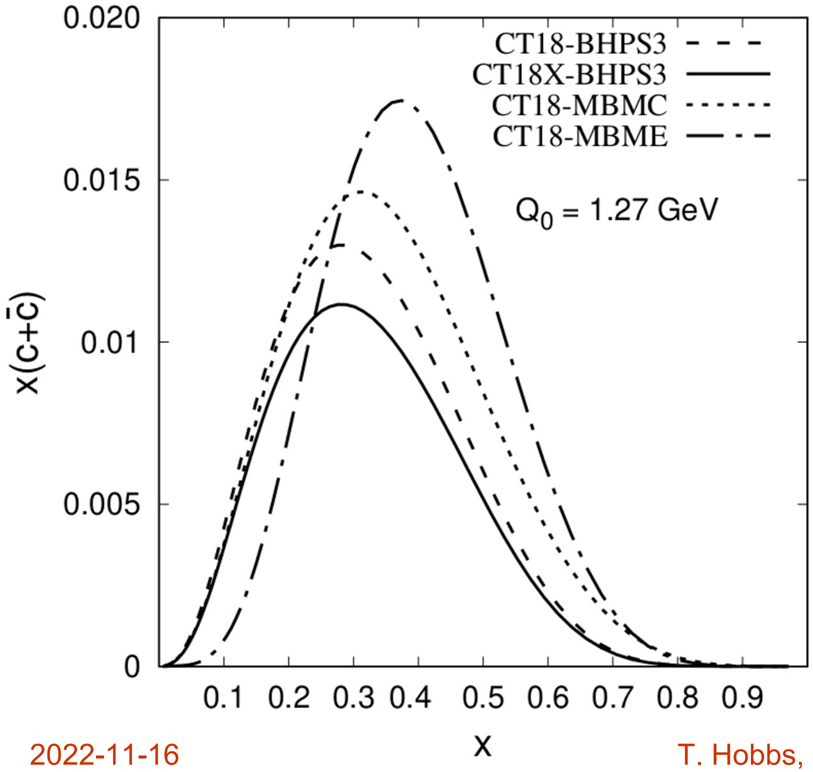
CT18 FC total charm PDFs

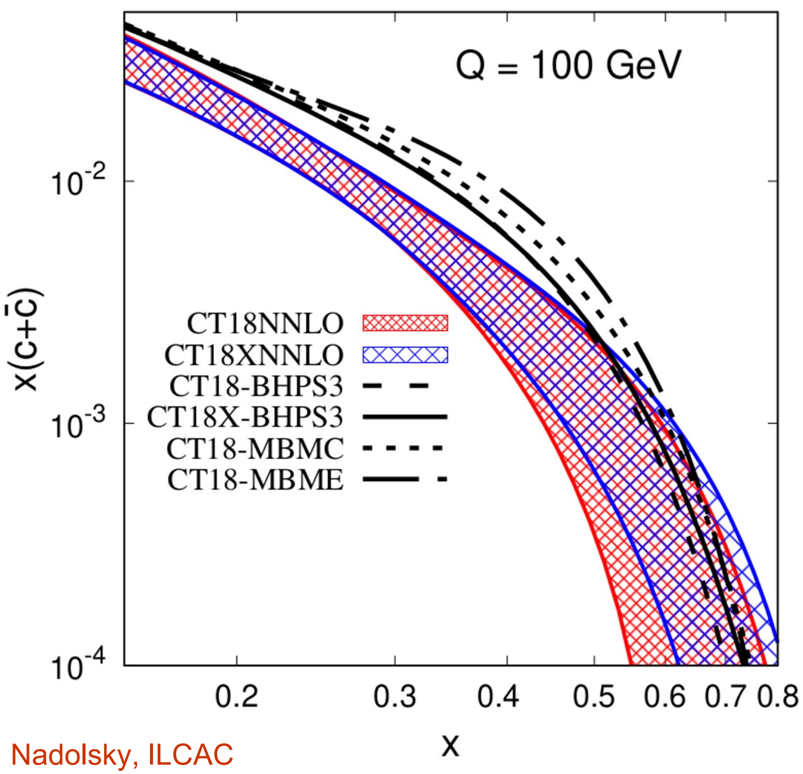
FC scenarios traverse range of high-*x* behaviors from IC models

- → fit implementation of BHPS from CT14IC (BHPS3) on CT18 or CT18X (NNLO)
- → fit two MBMs: MBMC (confining), MBME (effective mass) on CT18

investigate constraints from newer LHC data in CT18

central fits

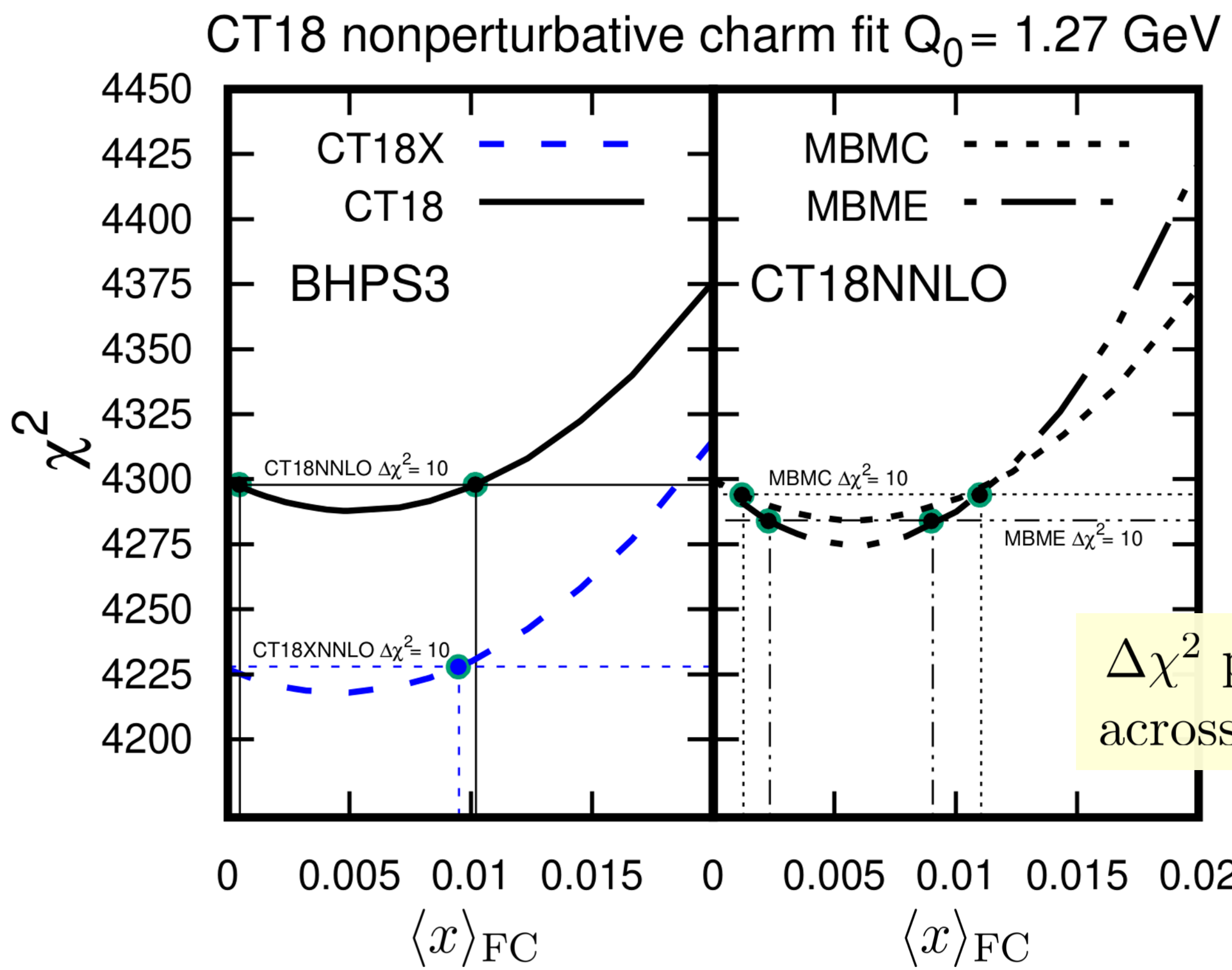




signal for FC in CT18 study, but with shallower $\Delta\chi^2$ than CT14 IC

FC uncertainty quantified by normalization via $\langle x \rangle_{\rm FC}$ for each input IC model

→ $\langle x \rangle_{\rm FC} \approx 0.5\%\ (\Delta\chi^2 \gtrsim -25)$ vs. $\langle x \rangle_{\rm FC} \approx 0.8{-}1\%\ (\Delta\chi^2 \gtrsim -40)$ CT14 IC



$\Delta\chi^2$ profiles similar across FC scenarios

FC PDF moments as F.o.M.

moments of the FC PDFs often used to characterize magnitude, asymmetry

$$\langle x^n \rangle_{c^\pm} = \int_0^1 dx\, x^n (c \pm \bar{c})[x, Q]$$

$$\langle x \rangle_{\rm FC} \equiv \langle x \rangle_{\rm c+} [Q_0 = 1.27\,{\rm GeV}]$$ ...at NNLO.

$$= 0.0048\,^{+0.0063}_{-0.0043}\,\left(^{+0.0090}_{-0.0048}\right),\ \text{CT18 (BHPS3)}$$

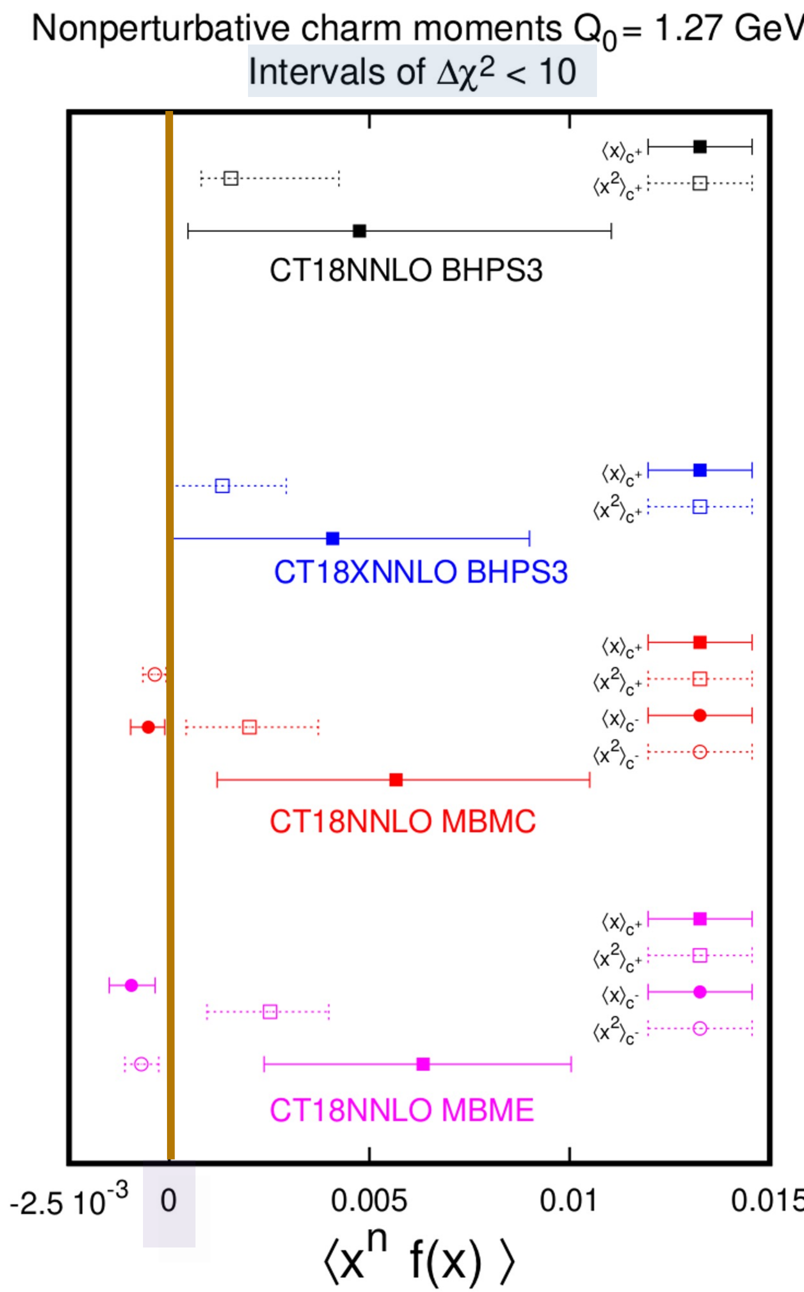
$$= 0.0041\,^{+0.0049}_{-0.0041}\,\left(^{+0.0091}_{-0.0041}\right),\ \text{CT18X (BHPS3)}$$

$$= 0.0057\,^{+0.0048}_{-0.0045}\,\left(^{+0.0084}_{-0.0057}\right),\ \text{CT18 (MBMC)}$$

$$= 0.0061\,^{+0.0030}_{-0.0038}\,\left(^{+0.0064}_{-0.0061}\right),\ \text{CT18 (MBME)}$$

$\Delta\chi^2 \leq 10$ (restrictive tolerance)

$\Delta\chi^2 \leq 30$ (~CT standard tolerance)



FC PDF moments as F.o.M.

even restrictive uncertainties give moments consistent with zero

→ broaden further for default CT tol.

→ lattice may give $\langle x \rangle_{c+}$, $\langle x^2 \rangle_{c-}$

$$\langle x \rangle_{\rm FC} \equiv \langle x \rangle_{\rm c+}[Q_0 = 1.27\,{\rm GeV}]$$

$$= 0.0048\,{}^{+0.0063}_{-0.0043}\ ({}^{+0.0090}_{-0.0048}),\ \text{CT18 (BHPS3)}$$

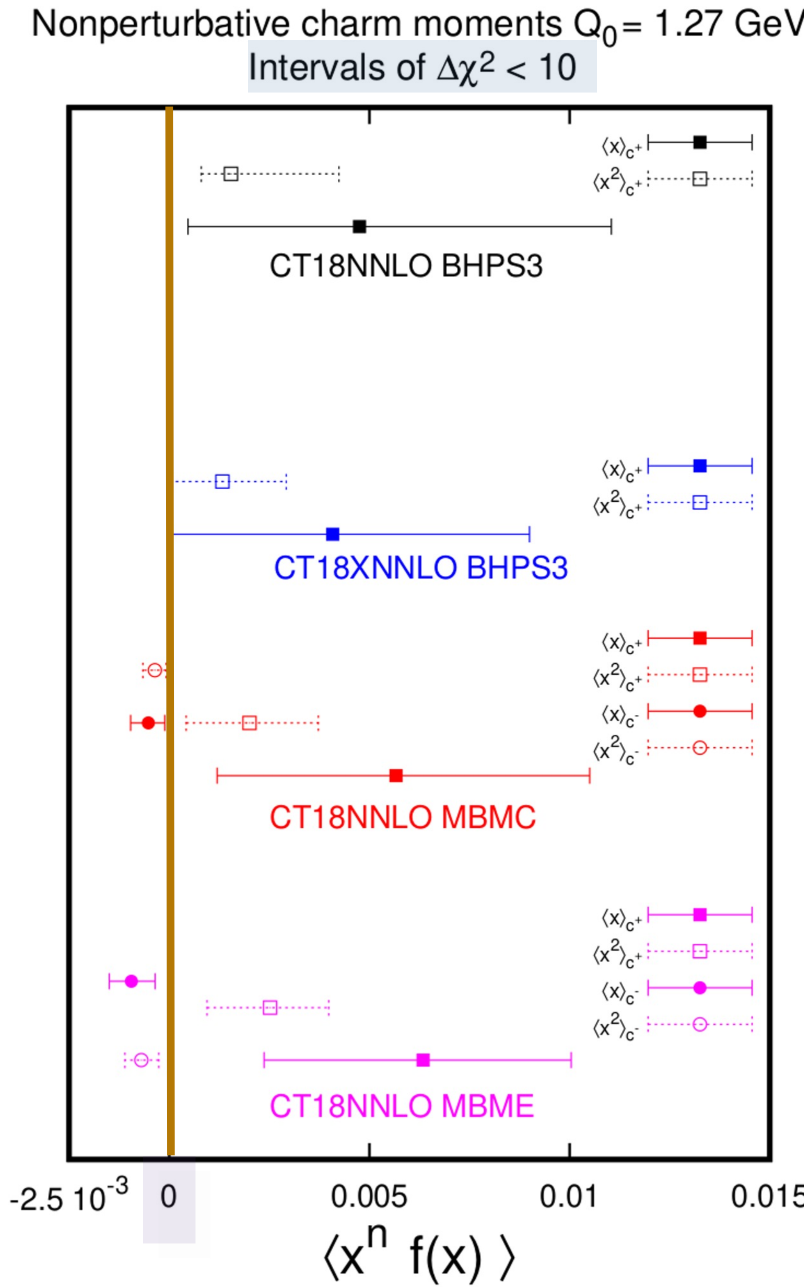
$$= 0.0041\,{}^{+0.0049}_{-0.0041}\ ({}^{+0.0091}_{-0.0041}),\ \text{CT18X (BHPS3)}$$

$$= 0.0057\,{}^{+0.0048}_{-0.0045}\ ({}^{+0.0084}_{-0.0057}),\ \text{CT18 (MBMC)}$$

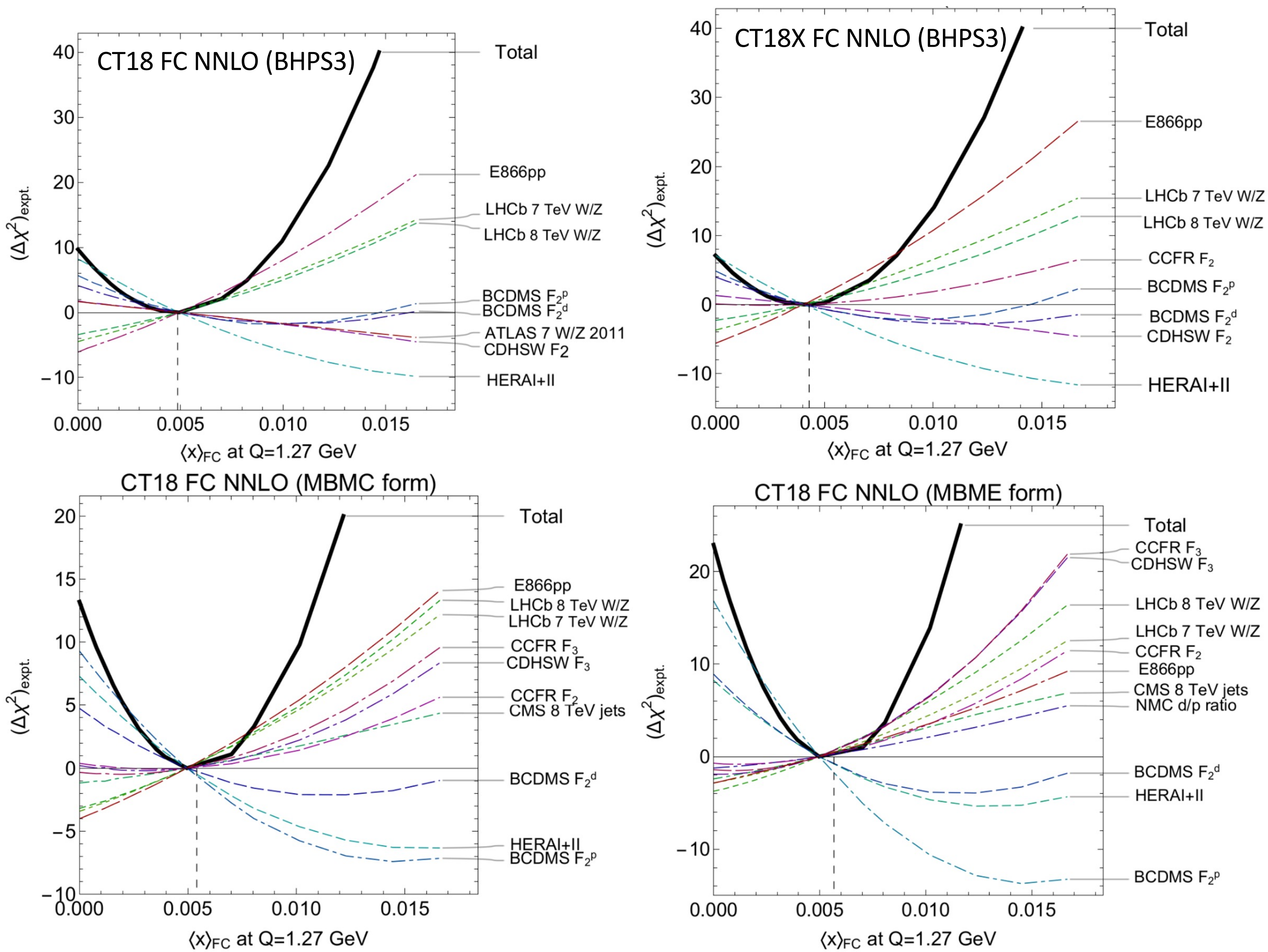
$$= 0.0061\,{}^{+0.0030}_{-0.0038}\ ({}^{+0.0064}_{-0.0061}),\ \text{CT18 (MBME)}$$

$\Delta\chi^2 \leq 10$ (restrictive tolerance)

$\Delta\chi^2 \leq 30$ (~CT standard tolerance)



data pull opposingly on $\langle x \rangle_{\rm FC}$; depend on FC scenario, enhancing error

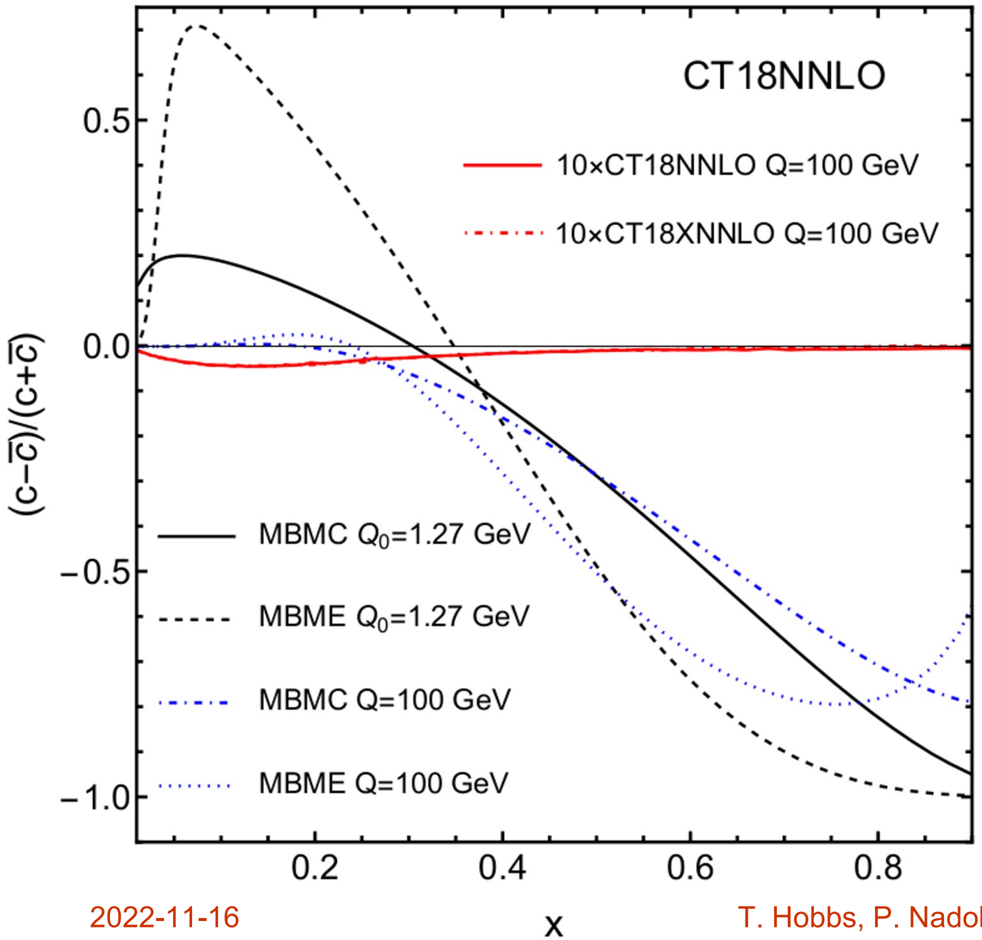


possible charm-anticharm asymmetries

pQCD only very weakly breaks $c = \bar{c}$ through HO corrections

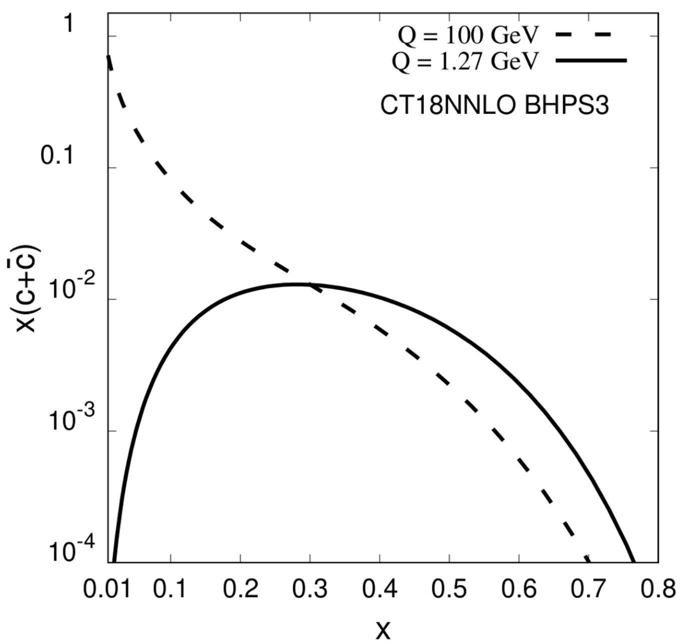
→ large(r) charm asymmetry would signal nonpert dynamics, IC

→ MBM breaks $c = \bar{c}$ through hadronic interactions



consider two MBM models as ***examples*** (not predictions)

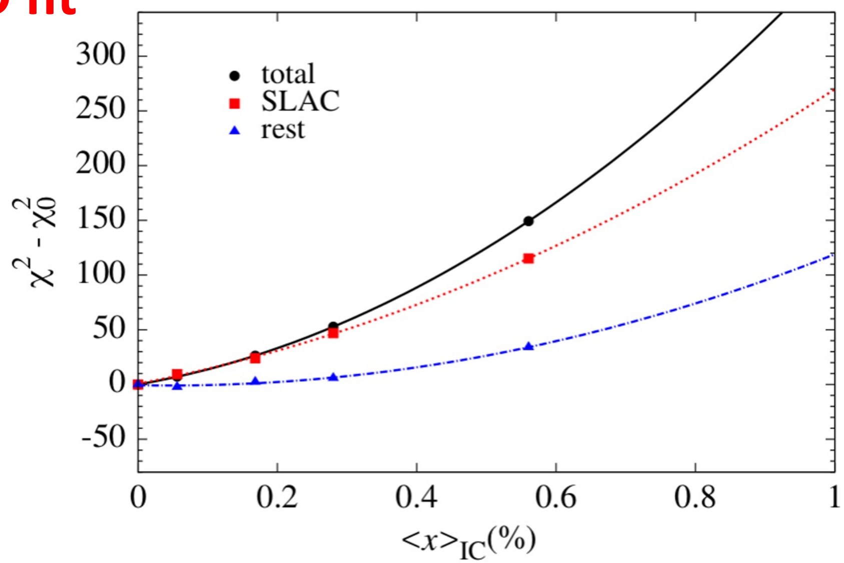
→ asym. small but ratio (left) can be bigger; will be hard to extract from data

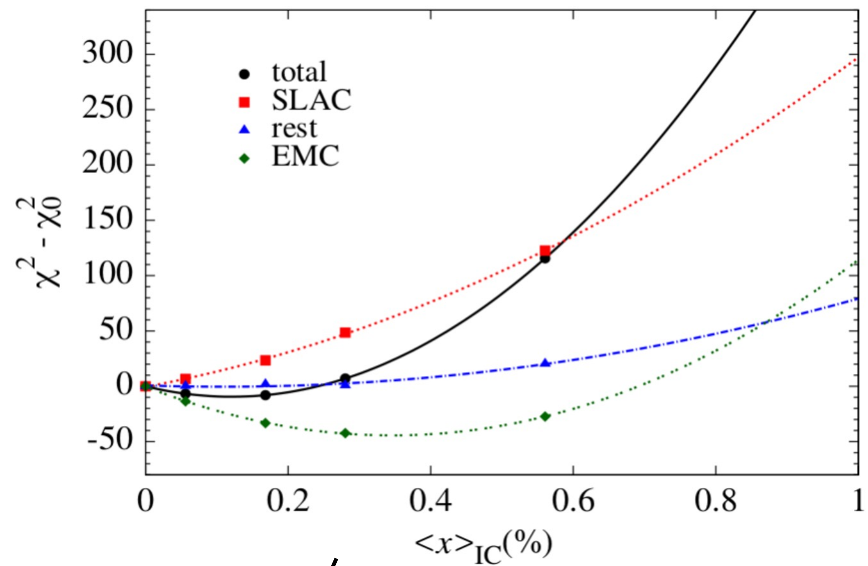


re: other studies → first full fit with EMC data; no signal for ‘IC’

P. Jimenez-Delgado, TJH, J. T. Londergan and W. Melnitchouk; PRL 114, no. 8, 082002 (2015).

NLO fit





‘SLAC + REST’ $\Longrightarrow$ $\langle x \rangle_{\rm IC} < 0.1\%$; at 5σ !

‘REST’ only $\Longrightarrow$ $\langle x \rangle_{\rm IC} < 0.1\%$; at 1σ

EMC alone: $\langle x \rangle_{\rm IC} = 0.3 - 0.4\%$

\+ SLAC/‘REST’: $\langle x \rangle_{\rm IC} = \mathbf{0.13 \pm 0.04}\%$

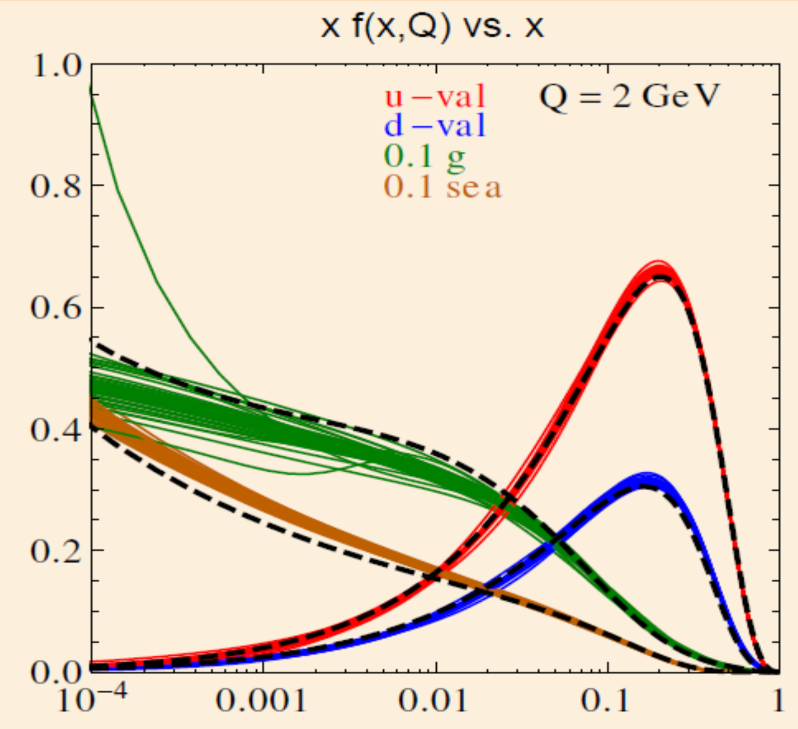
...but $F_2^{c\bar{c}}$ poorly fit — $\chi^2 \sim 4.3$ per datum!

included SLAC DIS; used more restrictive tolerance criterion, $\Delta\chi^2 = 1$

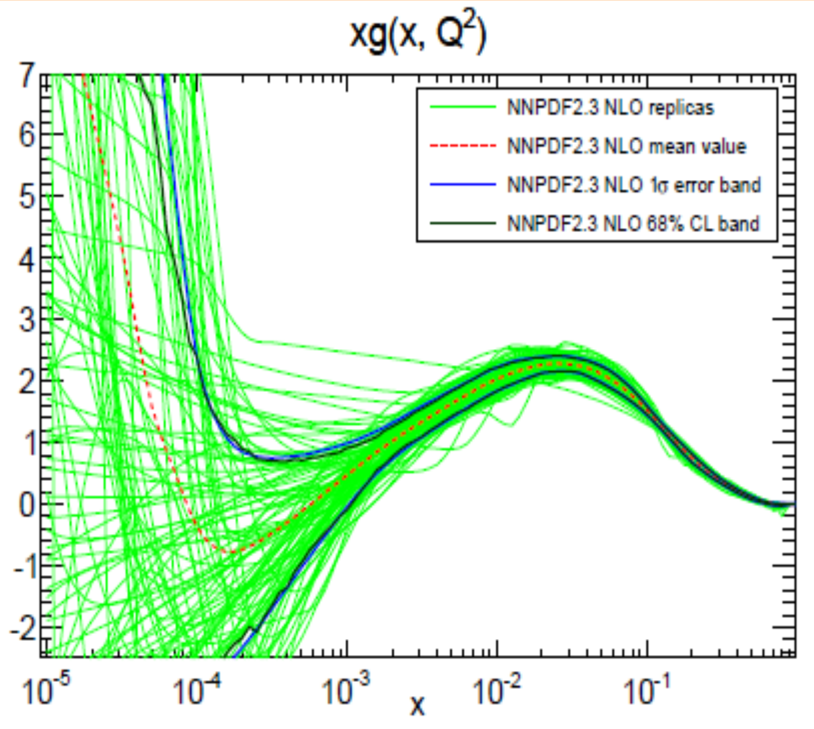
Part 3. Probability

Two types of modern error PDFs

Analytic parametrizations +
Hessian PDF eigenvector sets
(ABM, CTEQ, HERA, MMHT,…)



Neural network parameterizations
+ Monte Carlo PDF replicas
(NNPDF)



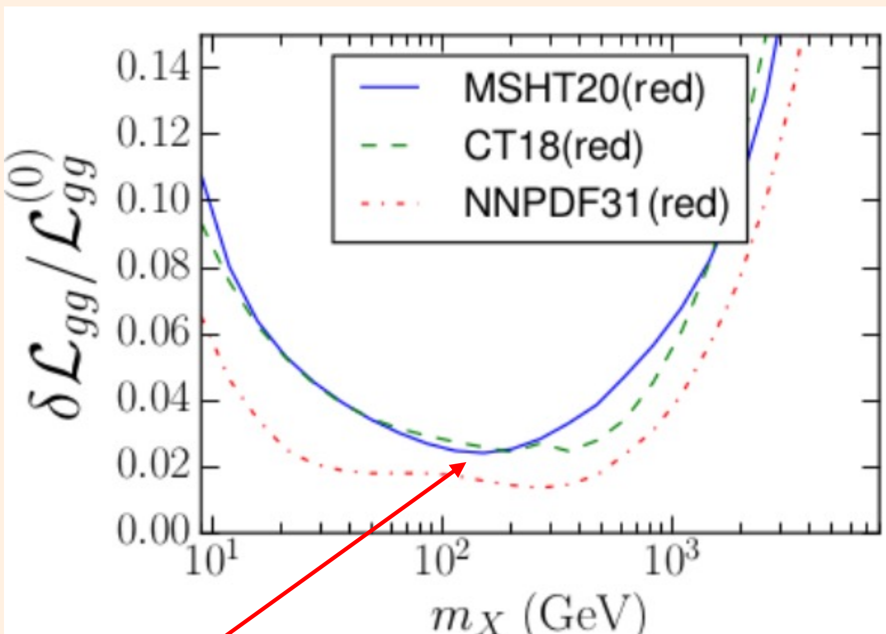
Obtained by analytic minimization of the log-likelihood χ^2

Obtained as NN replicas of an optimized architecture trained using an objective function with the log-likelihood and prior constraints

The tolerance puzzle

Relative PDF uncertainties on the gg luminosity at 14 TeV in three PDF4LHC21 fits to the **identical** reduced global data set

arXiv:2203.05506



×1.5 – 2 difference

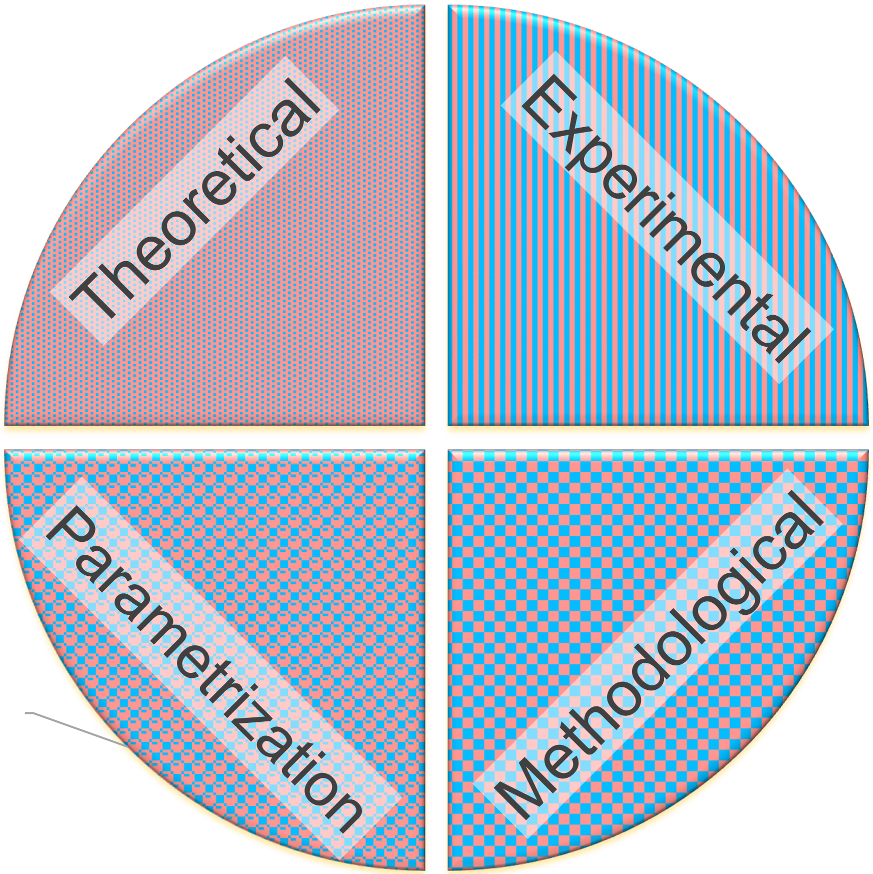
"Tolerance" determines the size of the PDF uncertainty.

While the fitted data sets are identical or similar in several such analyses, the resulting PDF sets may differ because of methodological choices adopted by the PDF fitting groups.

NNPDF3.1' and especially 4.0 (based on the NN's+ MC technique) tend to give smaller uncertainties in data-constrained regions

Is any error underestimated or overestimated?

Components of PDF uncertainty



In each category, one must maximize

- **PDF fitting accuracy** (accuracy of experimental, theoretical and other inputs)
- **PDF sampling accuracy** (adequacy of sampling of space of possible solutions)

NEW AND IMPORTANT

Fitting/sampling classification is borrowed from the statistics of large-scale surveys
[Xiao-Li Meng, *The Annals of Applied Statistics*, Vol. 12 (2018), p. 685]

Kovarik et al., arXiv: 1905.06957

Article

Unrepresentative big surveys significantly overestimated US vaccine uptake

Nature **v. 600** (2021) 695

https://doi.org/10.1038/s41586-021-04198-4

Received: 18 June 2021

Valerie C. Bradley[1,6], Shiro Kuriwaki[2,6], Michael Isakov[3], Dino Sejdinovic[1], Xiao-Li Meng[4] & Seth Flaxman[5]

SCIENCE ADVANCES | RESEARCH ARTICLE

MATHEMATICS

Models with higher effective dimensions tend to produce more uncertain estimates

Arnald Puy[1,2,3]*, Pierfrancesco Beneventano[4], Simon A. Levin[2], Samuele Lo Piano[5], Tommaso Portaluri[6], Andrea Saltelli[3,7]

The Big Data Paradox in Clinical Practice

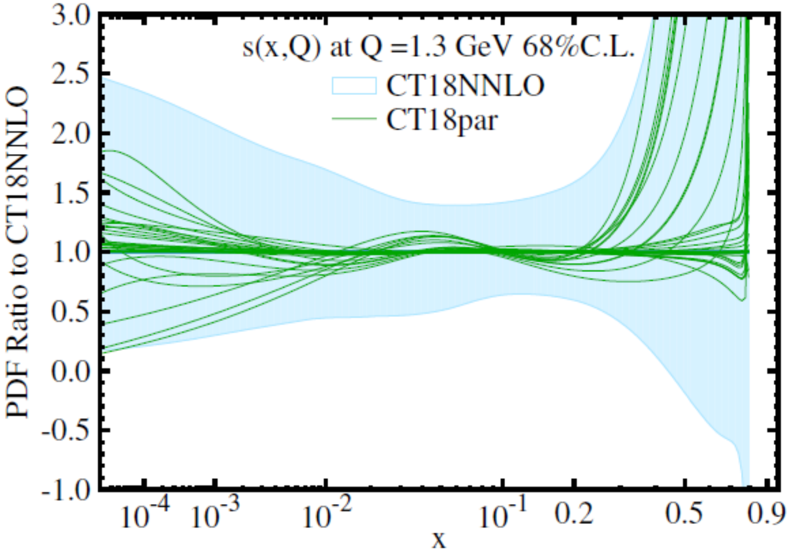
Pavlos Msaouel

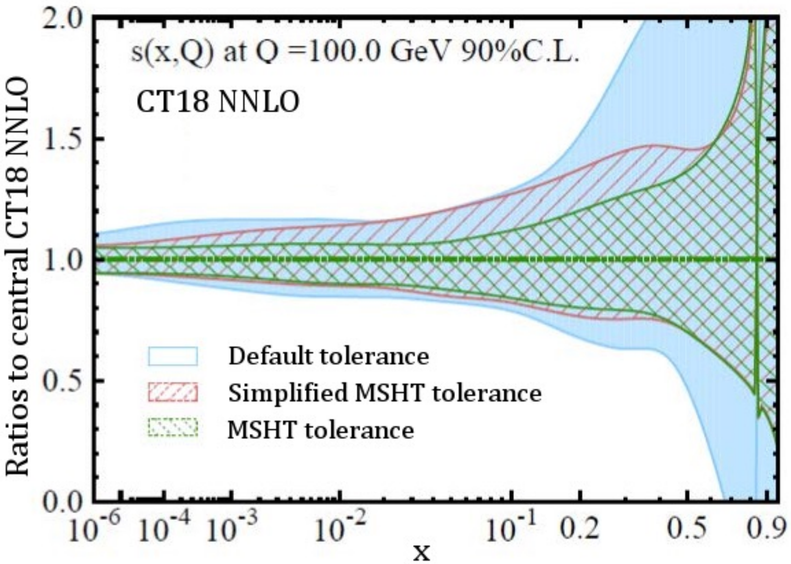
To cite this article: Pavlos Msaouel (2022) The Big Data Paradox in Clinical Practice, Cancer Investigation, 40:7, 567-576, DOI: 10.1080/07357907.2022.2084621

A flurry of recent articles point out statistical difficulties with estimating uncertainties in multiparametric models working with large data sets.
Here, subtle **sampling biases can negate** reduction of the uncertainty due to **the law of large numbers,** leading to the situation dubbed "the **big-data paradox"**:

> Unrepresentative sampling in a large number of dimensions commonly leads to wrong conclusions with high confidence.

CT and MSHT tolerances: average $\langle\Delta\chi^2\rangle \approx$10 and 30 at 68% C.L.





Tolerance can be understood in the language of sampling over acceptable PDF solutions.

Upper figure: A large part of the CT18 PDF uncertainty accounts for the sampling over 250-350 parametrization forms, possible choices of fitted experiments and fitting parameters, definitions of χ^2

Lower figure: this approach sometimes enlarges the uncertainties compared to the other groups. For example, the MSHT tolerance is based on the agreement with each included experiment for a fixed parametrization form.

More restrictive tolerance criteria may introduce a hidden uncertainty due to unrepresentative sampling.

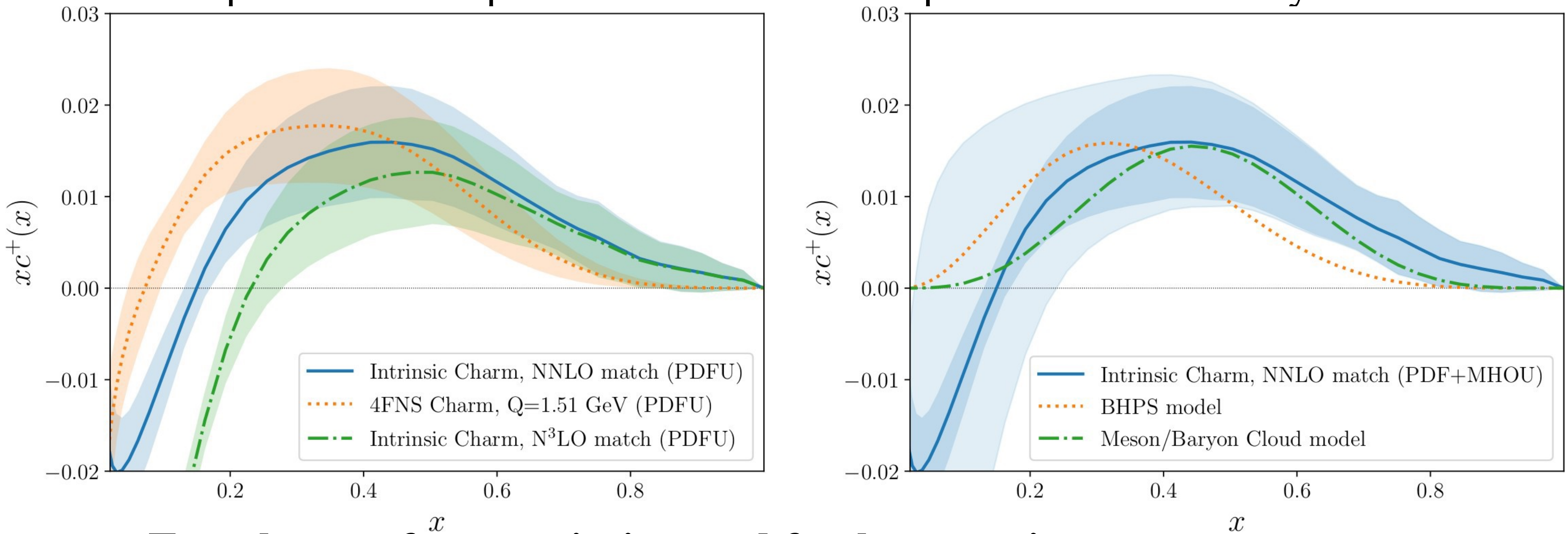
recent NNPDF IC analysis

NNPDF, Nature 608 (2022) 7923, 483.

- NNPDF have recently claimed 3σ evidence for 'IC'

  → based on local (x-dependent) deviation of FC PDF from the 'no-FC' scenario

  → implies crucial dependence on size and shape of PDF uncertainty



→ **Two classes of uncertainties need further scrutiny:**

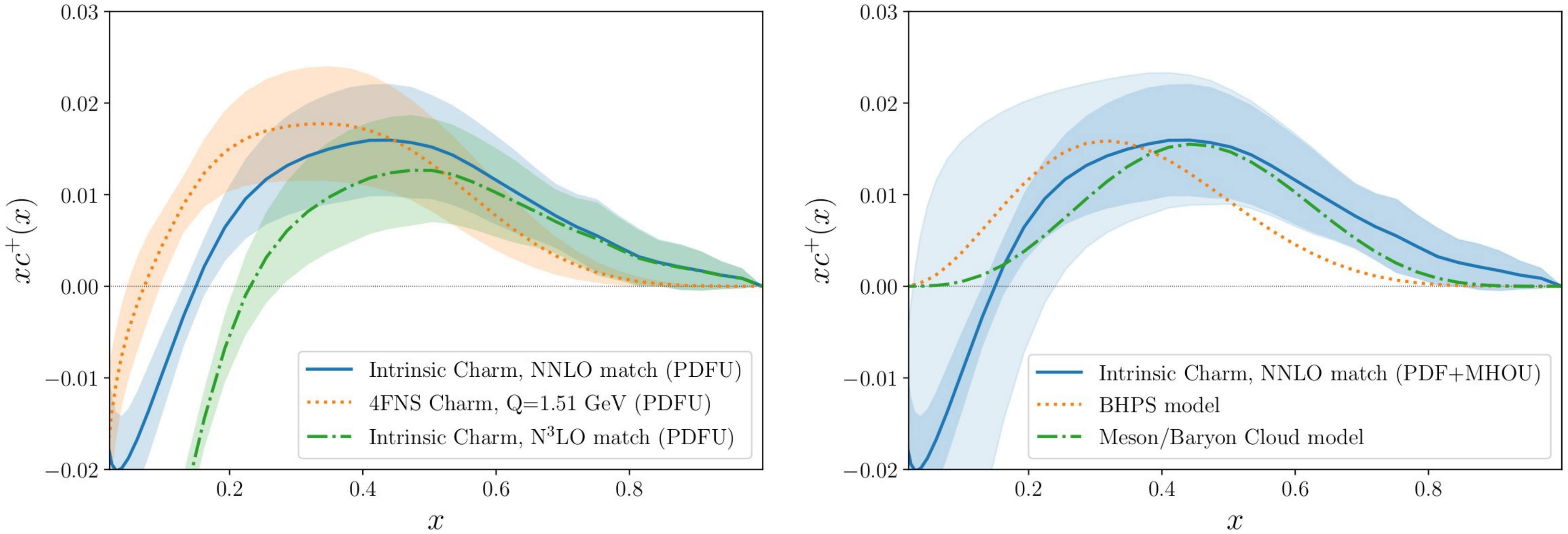
1. Missing higher-order uncertainties (MHOU): N3LO in DIS, etc.; N2LO in Z+c production
2. Parametrization sampling uncertainty (PSU): underestimation of PDFU

NNPDF IC PDFs and moments

NNPDF, Nature 608 (2022) 7923, 483.

- large perturbative instability from MHOU in DGLAP affects low-x behavior

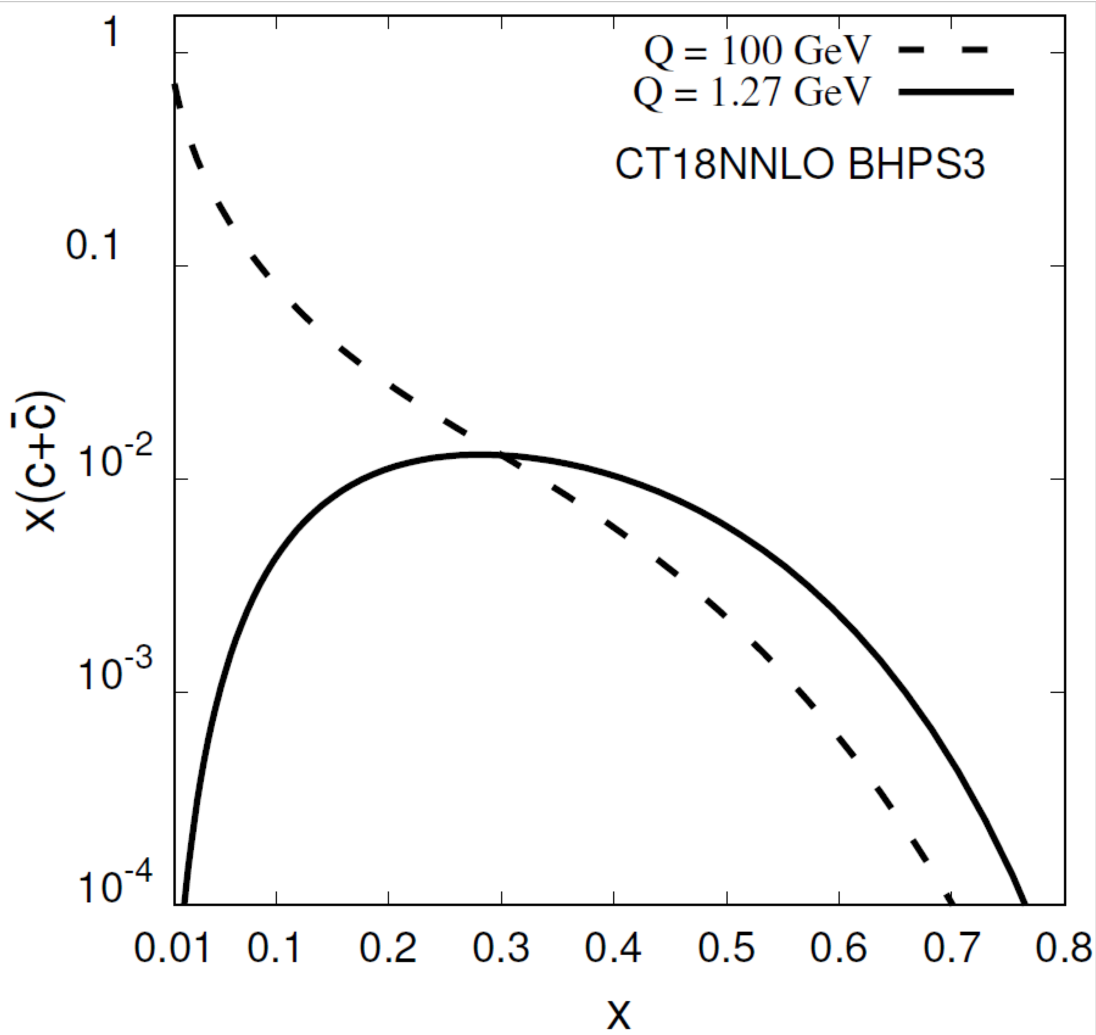
  → matching at fixed NNLO gives negative FC, unlike IC models



→ MHOU excluded to obtain a nominal charm fraction, $\langle x \rangle_{\rm FC} = 0.62 \pm 0.28\%$

→ if MHOU is included, consistency with zero: $\langle x \rangle_{\rm FC} = 0.62 \pm 0.61\%$

Backward DGLAP evolution is approximate



Data constrain the PDFs at $Q > 2$ GeV.

When PDFs are evolved at N2LO down to $Q \approx 1.3$ GeV, the charm PDF is increased at $x \gtrsim 0.3$ and decreased at $x \lesssim 0.3$.

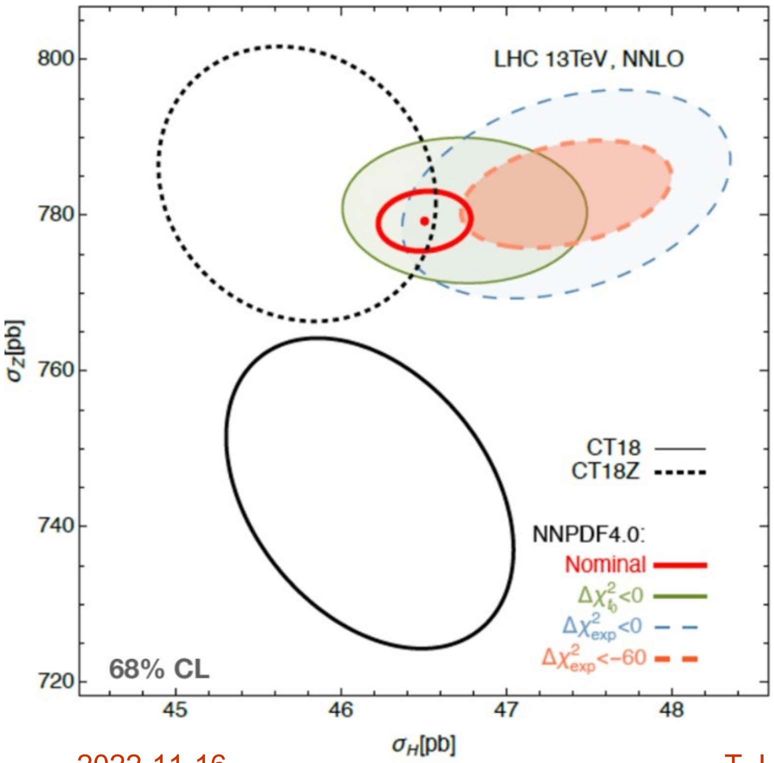
MHOU in DGLAP evolution can produce the bump-like shape.

more representative sampling can enlarge MC uncertainties

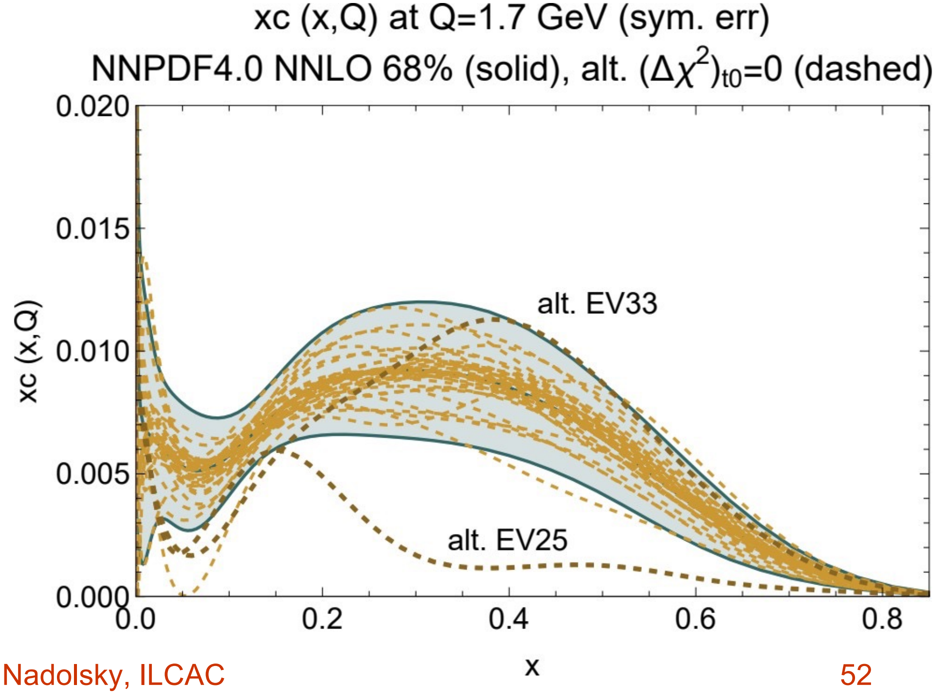
Courtoy *et al*., arXiv: 2205.10444.

- default replica-training in MC studies may omit otherwise acceptable solutions

- more comprehensive sampling with the public NNPDF4.0 code impacts PDF errors of cross sections



- substantially broadens high-x FC error

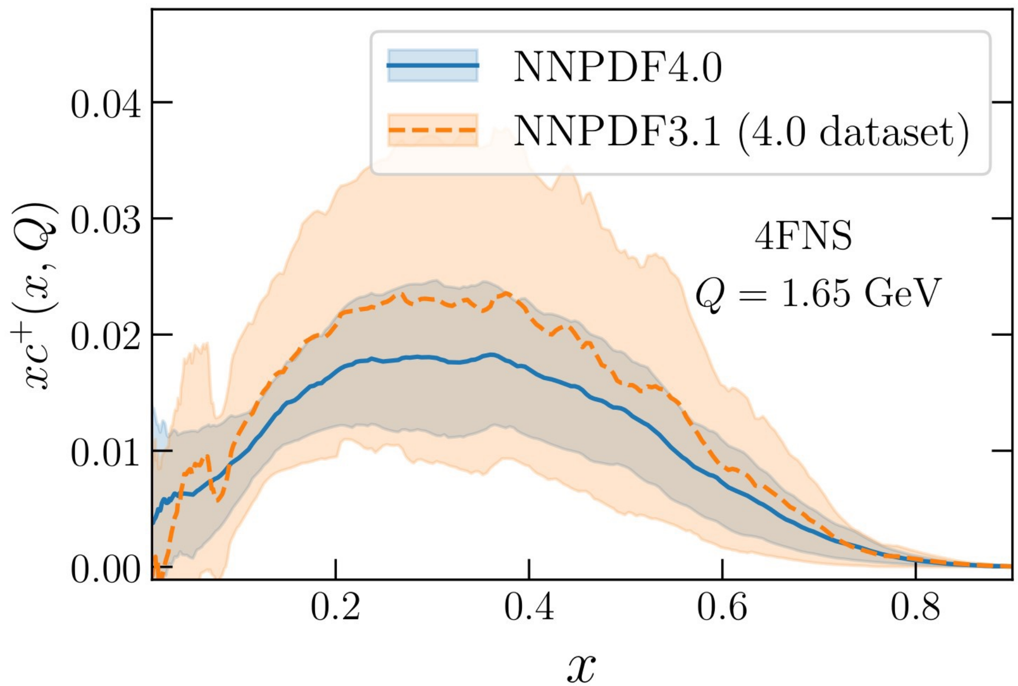


more representative sampling can enlarge MC uncertainties

Courtoy *et al*., arXiv: 2205.10444.

- default replica-training in MC studies may omit otherwise acceptable solutions

more comprehensive sampling impacts PDF errors of cross sections
→ alternate fitting methodologies (NNPDF3.1 vs. 4.0) produce significant differences in the PDF uncertainty

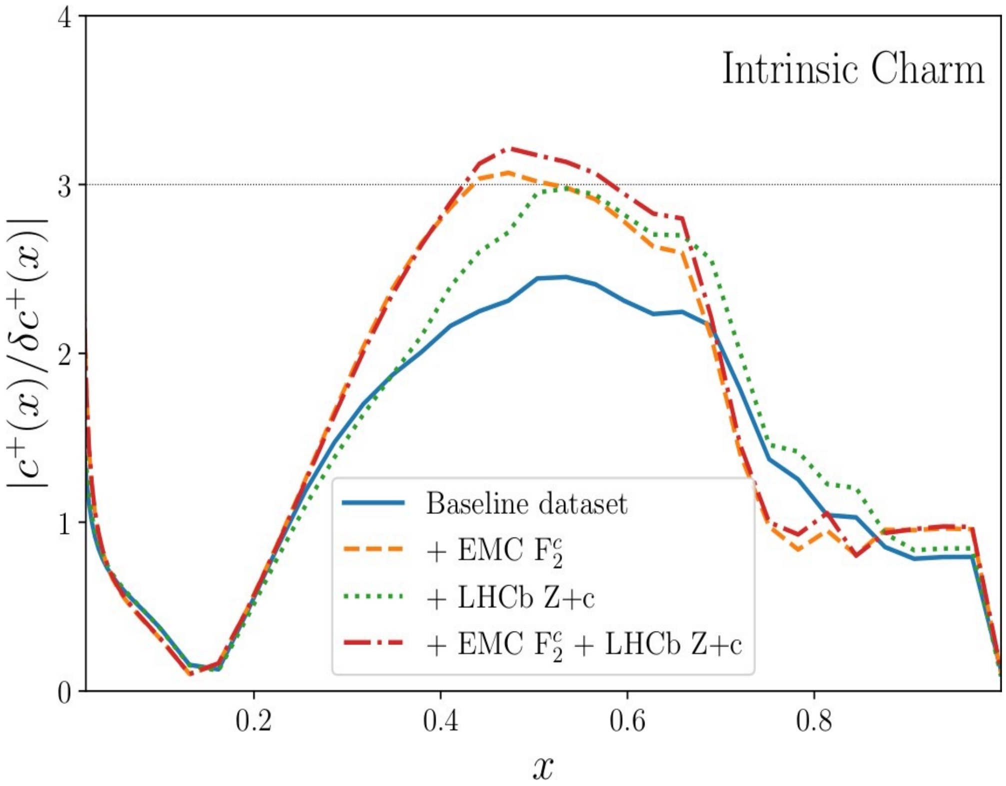


both curves based on same underlying data

Revisiting the significance in NNPDF IC

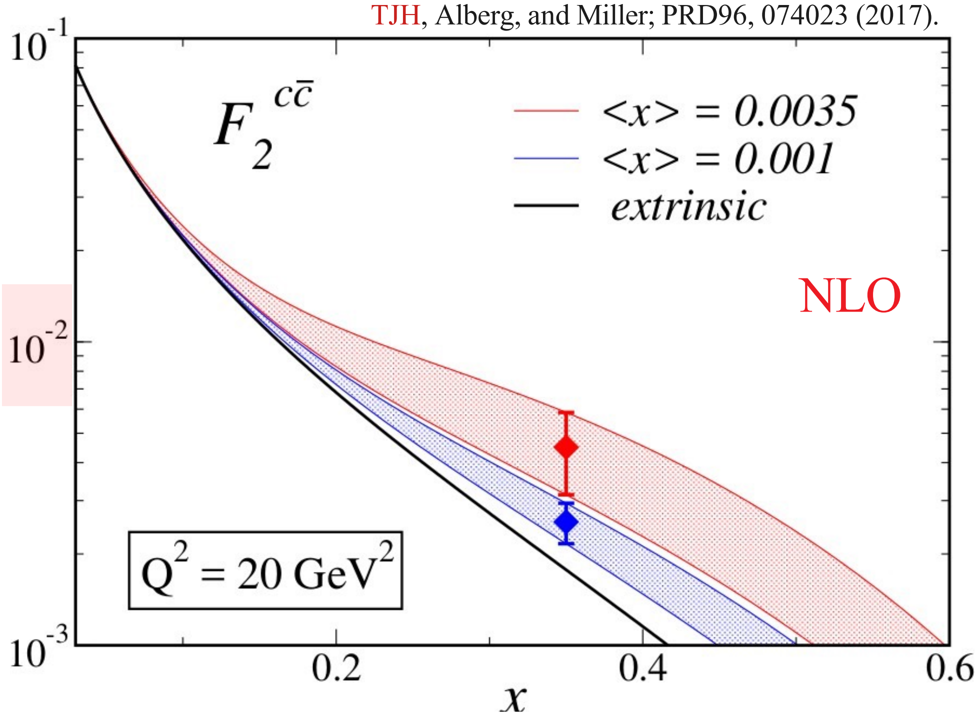
Important additional uncertainties:

- In the baseline fit due to sampling of MC replicas
- In the NLO LHCb $Z + c$ analysis due to MHOU and final-state showering
- In the EMC F_2^c due to insufficient control of syst. uncertainties and LO analysis



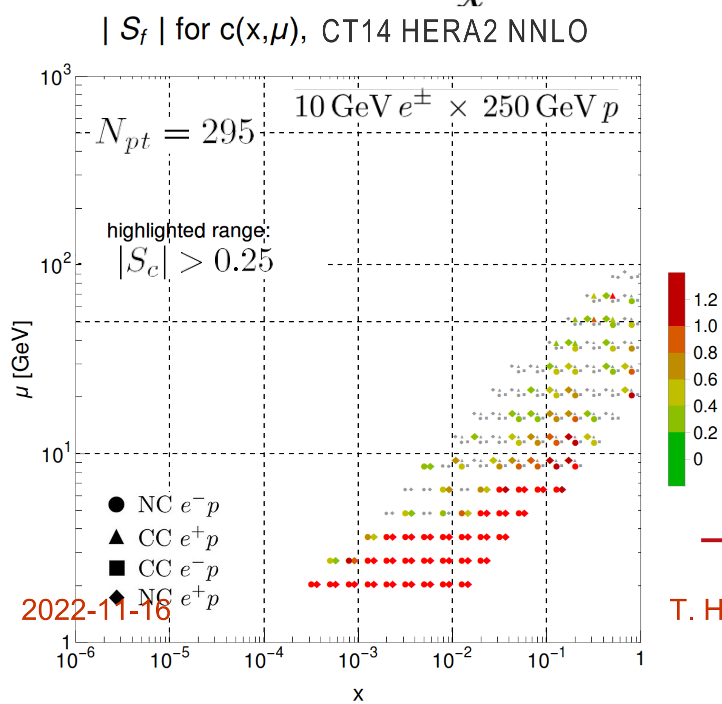
∴ No significant evidence for NNPDF4.0 IC, in compliance with CT18 FC observations

future data will inform FC



EIC + lattice QCD will constrain FC scenarios

enhanced FC momentum implied by EMC data → small high-x effects in structure function; need high precision



essential complementary input from LHC; CERN FPF

EIC will measure precisely in the few-GeV, high-x region where FC signals are to be expected

conclusions

- size, shape of nonpert charm remains indeterminate

  → theoretical ambiguities in relation between FC/IC unresolved

  → need more sensitive data; FC currently consistent with zero

concordance with enlarged error estimates: $\langle x \rangle_{\rm FC} \sim 0.5\%$, well below evidence-level

- need more NNLO and better showering calculations (*e.g.*, for *Z*+*c*)
- further progress in quantifying and estimating PDF uncertainties

opportunities to improve knowledge of FC:

- → promising experiments at LHC; EIC; CERN FPF

  → lattice data on key charm PDF moments; quasi-PDFs

  → direct benchmarking of FC among PDF fitting groups

Backup