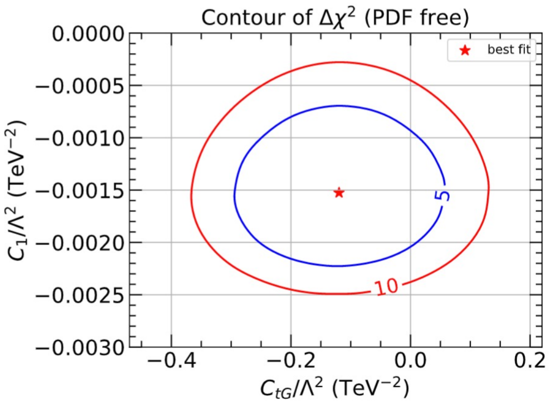
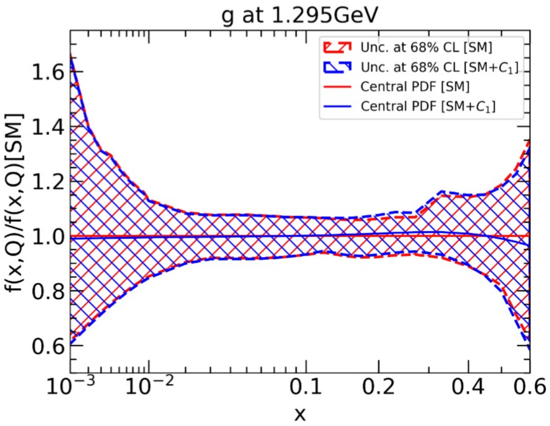
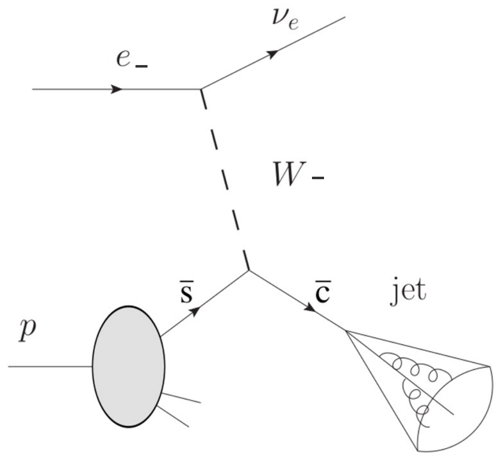
PDFs and BSM searches in the EIC era







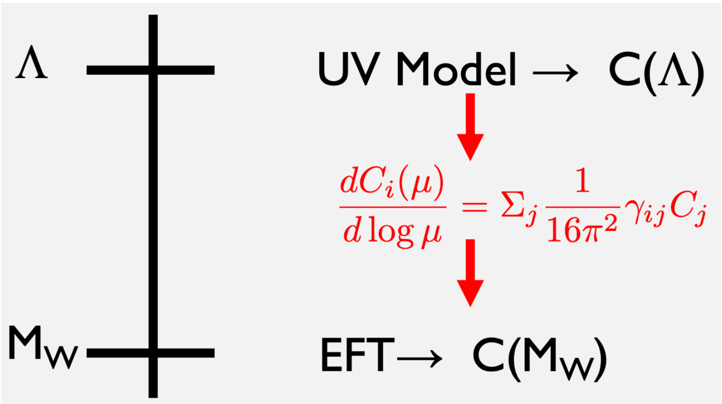
Tim Hobbs

Argonne NATIONAL LABORATORY

CTEQ

[CTEQ-TEA](#) (Tung Et. Al.) working group

S. Dawson



UV Model → C(Λ)

$$\frac{dC_i(\mu)}{d\log\mu} = \Sigma_j \frac{1}{16\pi^2}\gamma_{ij}C_j$$

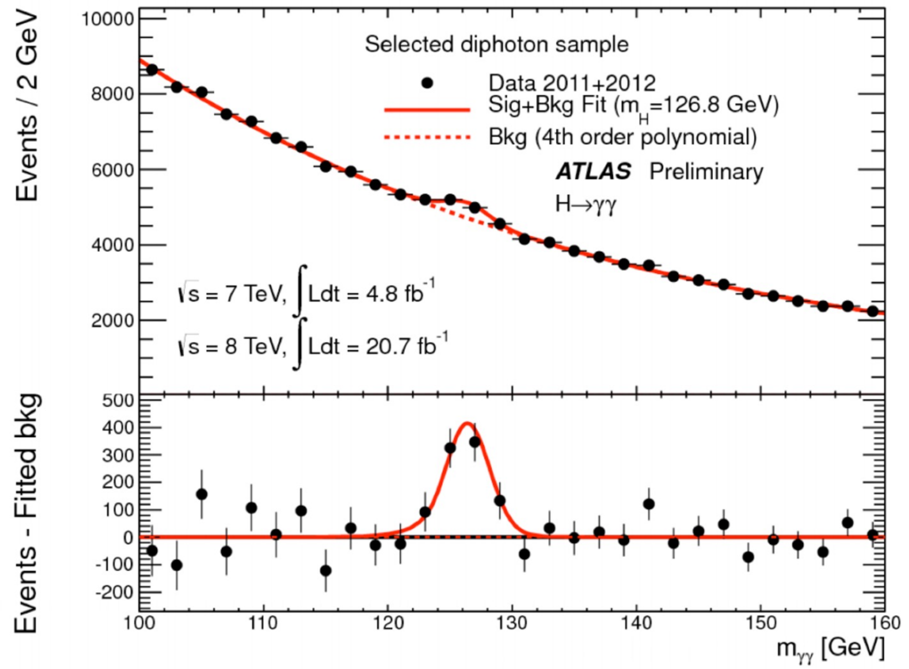
EFT→ C(M_W)

searching for physics beyond the Standard Model (BSM)

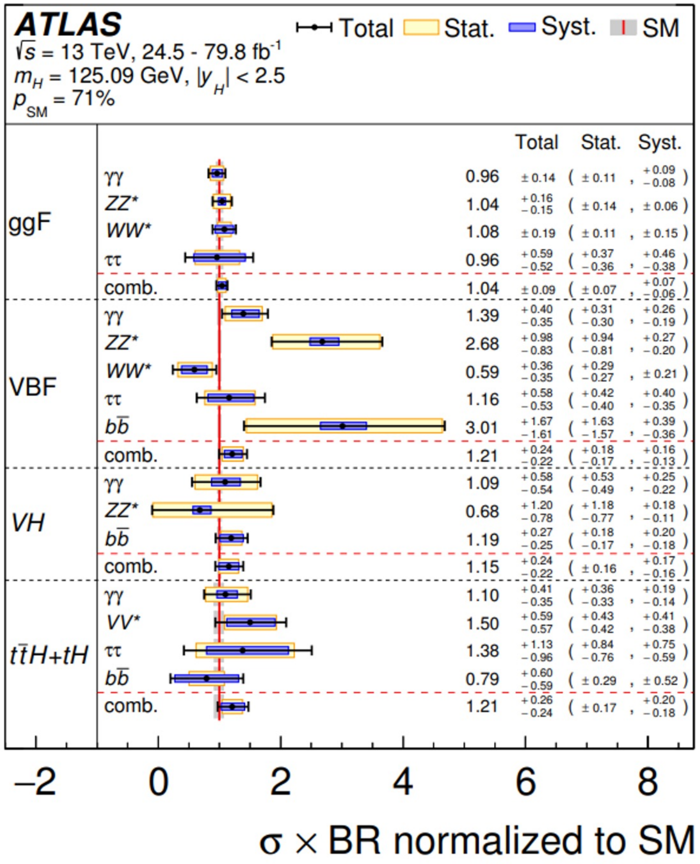
→ discovery searches

e.g., examining cross sections in previously unprobed kinematical regions

Higgs discovery, 2012



Higgs prod·decay/SM (PDG)

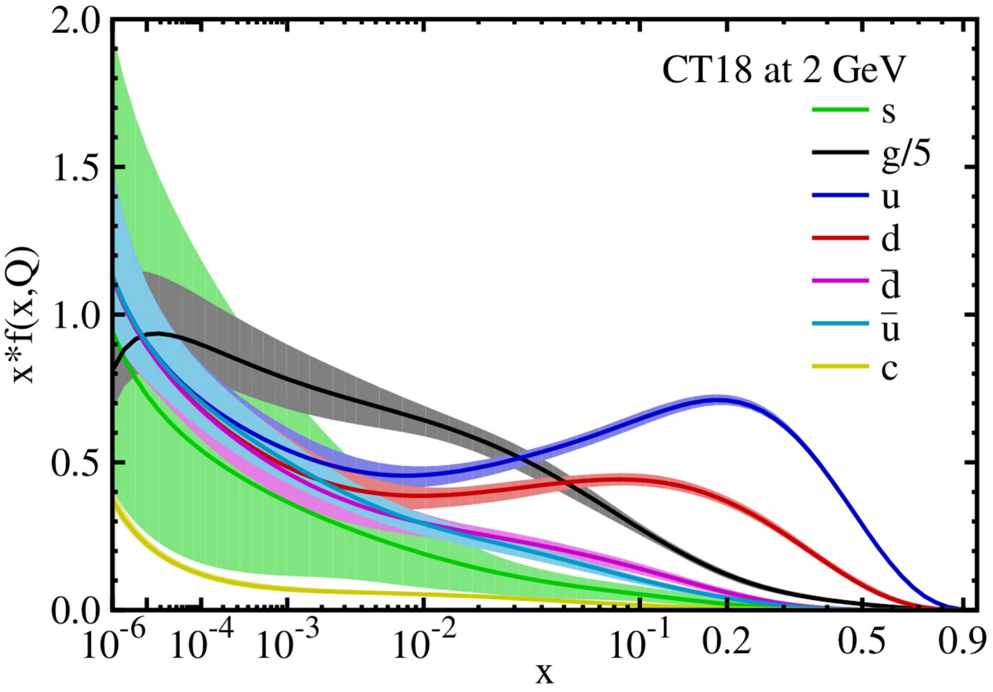


→ still no SUSY: precision searches

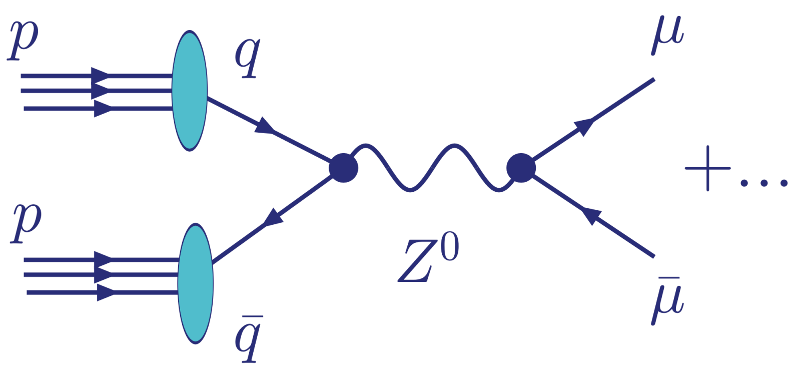
testing the Standard Model through extremely fine measurements

(deviations could reveal presence of new particles/interactions!)

PDFs are a linchpin for BSM searches



- <u>today</u>: PDFs from HEP, BSM-search perspective
  - theory, methodology for determinations from hadronic data



→ QCD <u>factorization theorem</u>; here, for Drell-Yan processes (e.g., LHC):

$$\sigma_{pp\to\ell\bar{\ell}X} = \sum_{a,b=q,\bar{q},g} \int_0^1 d\xi_1 \int_0^1 d\xi_2 \, \widehat{\sigma}_{ab\to Z\to\ell\bar{\ell}} \left(\frac{x_1}{\xi_1}, \frac{x_2}{\xi_2}; \frac{Q}{\mu}\right) f_{a/p}(\xi_1,\mu^2) f_{b/p}(\xi_2,\mu^2)$$

$$+ \, \mathcal{O}\left(\Lambda^2_{QCD}/Q^2\right)$$

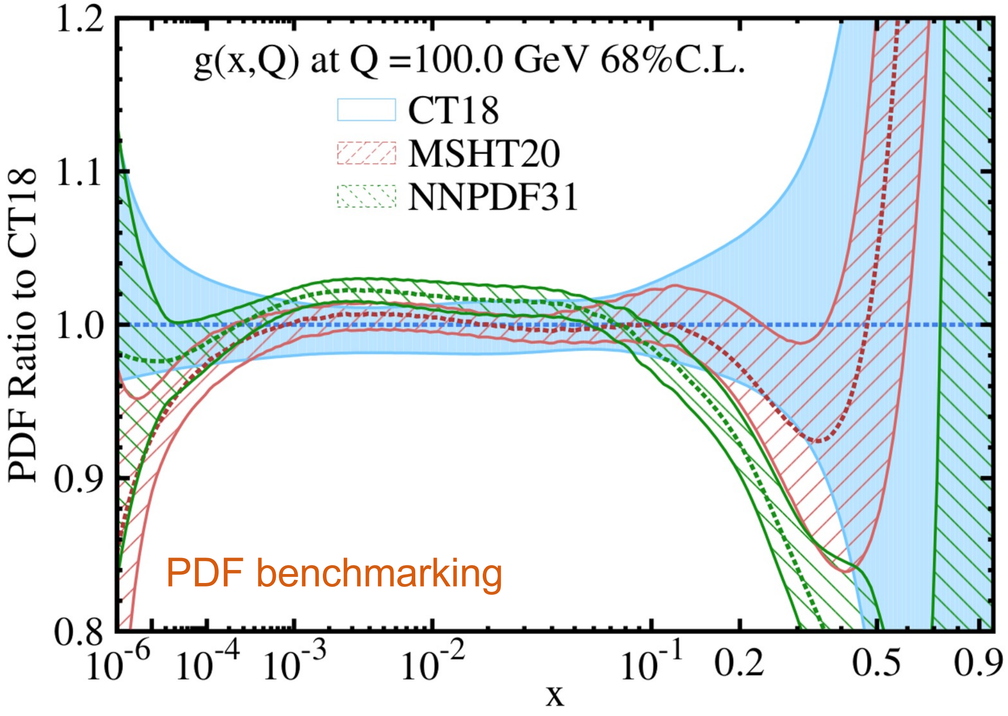
PDF global analyses have matured to high level of precision

BSM searches, upcoming programs → reductions to PDF uncertainties

→ analyses are large and challenging: (N)NNLO; NLO EW; **<u>many subtleties</u>**

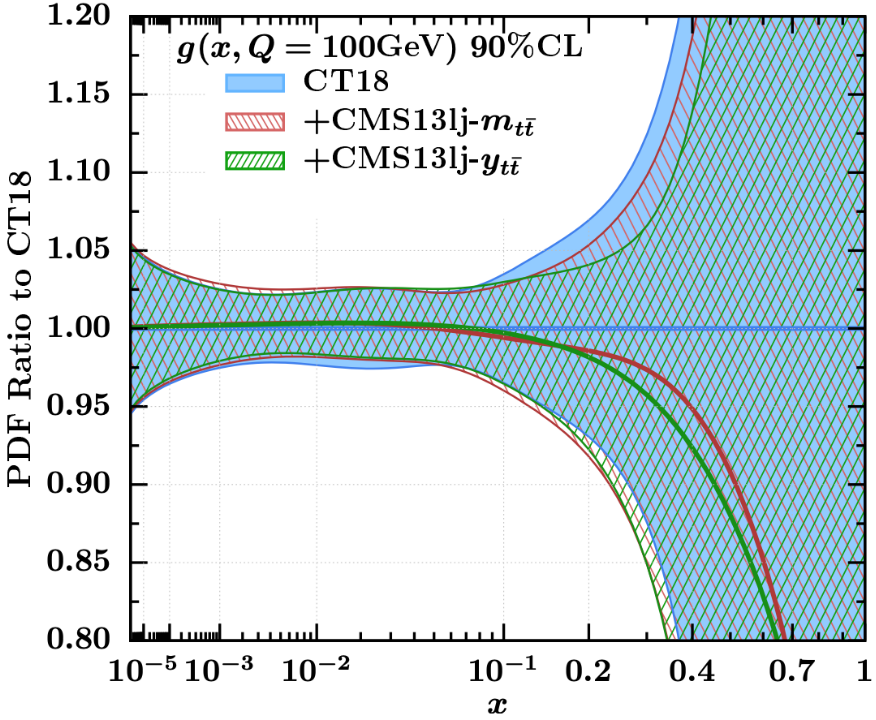
statistical, methodological assumptions; theory settings; data curation; …

<u>CT18:</u> Hou, Gao, TJH et al, PRD**103** (2021) 1, 014013



PDF benchmarking

<u>CT18 top update</u>, Ablat et al; arXiv: 2307.11153

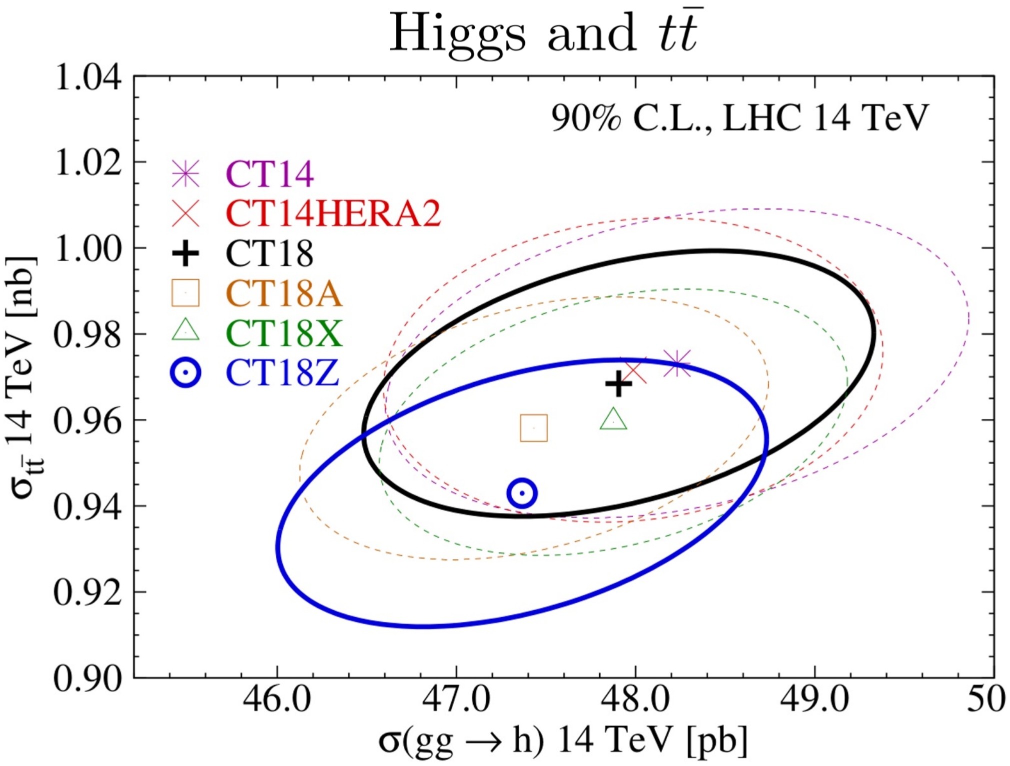


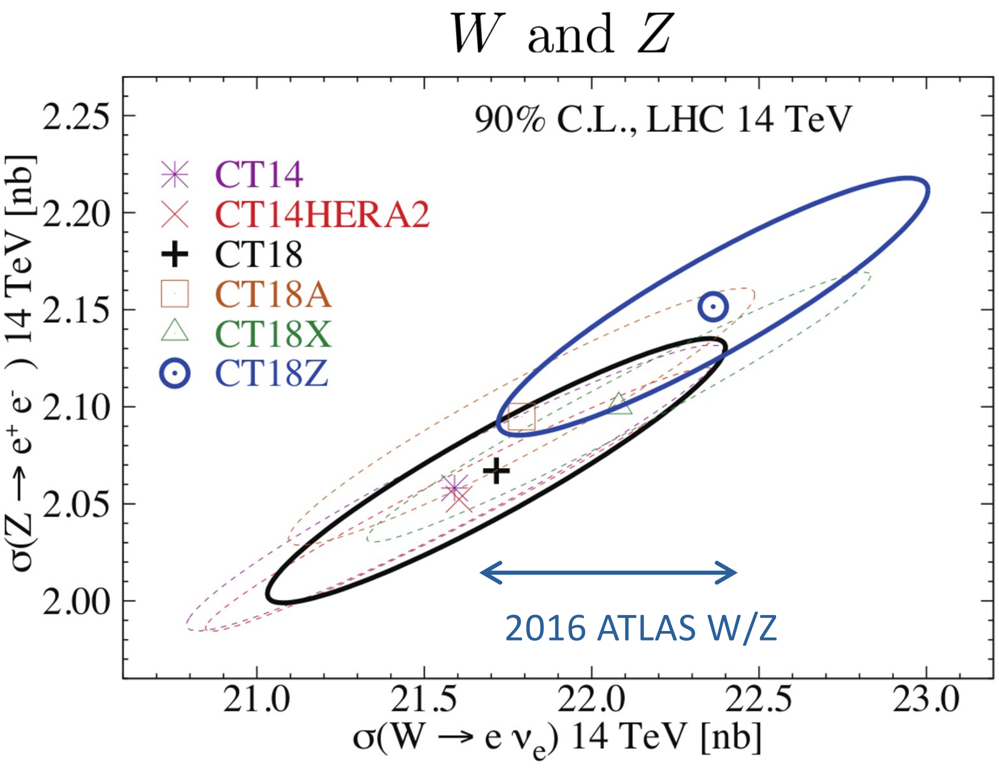
→ significant interest in jet production, top data: connections to gluon PDF

→ PDF fits, <u>SM-only</u>! …however, data are often **inputs to BSM (EFT) analyses**

SM theory predictions from global analyses

from (N)NNLO analyses, state-of-the-art predictions for fundamental LHC observables → *e.g.*, **total cross sections at 14 TeV**





$$\mu_R = \mu_F = m_t; m_{W,Z}; m_H$$

Higgs, NNLO QCD: iHixs v1.3

$t\bar{t}$, NNLO+NNLL: Top++ v2.0

NNLO QCD: Vrap v0.9

significant PDF-driven uncertainties; also, systematic effects: W cross sections sensitive to inclusion of 2016 7 TeV ATLAS inclusive W/Z data

BSM-sensitive channels limited by PDF uncertainties

→ includes many observables: $\sigma_H\,,\ \sin^2\theta_W\,,\ M_W\,,\ \cdots$

$$M_W^{\rm CDF} = 80,433.5 \pm 6.4\ ({\rm stat}) \pm 6.9\ ({\rm syst})\ {\rm MeV}$$

→ *e.g.*, recent CDF M$_W$ measurement: <u>significant</u> PDF dependence

$$M_T \equiv \sqrt{2(p_T^\ell p_T^\nu - \vec{p}_T^{\,\ell}\cdot\vec{p}_T^{\,\nu})}$$

Gao, Liu, Xie: 2205.03942 [hep-ph]

δM_W in MeV	CT10	CT18	MMHT14	NNPDF4.0	CT14	MSHT20
$W^+\ \langle M_T\rangle$ (NLO)	$0^{+12.1}_{-12.9}$	$+1.4^{+21.8}_{-20.0}$	$-10.3^{+11.6}_{-11.1}$	$-17.1^{+7.4}_{-7.4}$	$-16.2^{+23.5}_{-19.1}$	$-24.8^{+16.8}_{-11.9}$
$W^-\ \langle M_T\rangle$ (NLO)	$0^{+13.5}_{-15.2}$	$-5.7^{+14.0}_{-19.5}$	$+1.1^{+8.6}_{-10.3}$	$+7.5^{+4.9}_{-4.9}$	$-9.6^{+12.8}_{-15.3}$	$-4.5^{+8.3}_{-7.5}$
$W^\pm\ \langle M_T\rangle$ (NLO)	$0^{+9.8}_{-11.4}$	$-2.3^{+14.4}_{-16.8}$	$-4.5^{+8.2}_{-8.5}$	$-4.4^{+4.6}_{-4.6}$	$-12.8^{+16.6}_{-15.1}$	$-14.3^{+10.9}_{-8.0}$
$W^+\ \langle M_T\rangle$ (LO)	$0^{+10.8}_{-11.4}$	$-6.5^{+14.1}_{-10.0}$	$-5.7^{+8.1}_{-7.1}$	$-14.1^{+5.8}_{-5.8}$	$-4.1^{+15.0}_{-12.9}$	$-14.4^{+10.2}_{-7.3}$
$W^-\ \langle M_T\rangle$ (LO)	$0^{+8.9}_{-11.4}$	$-7.2^{+10.1}_{-12.5}$	$+3.1^{+8.3}_{-9.9}$	$+3.5^{+4.5}_{-4.5}$	$-7.0^{+6.2}_{-8.9}$	$+2.1^{+6.3}_{-4.9}$
$W^\pm\ \langle M_T\rangle$ (LO)	$0^{+5.2}_{-7.0}$	$-0.6^{+7.6}_{-7.4}$	$-1.2^{+5.3}_{-5.9}$	$-5.0^{+3.0}_{-3.0}$	$-5.6^{+8.0}_{-8.4}$	$-5.9^{+5.9}_{-4.2}$

→ extracted from transverse mass template fits

→ examine mean transverse mass as proxy

→ **large PDF variations; outdistance 1-sigma uncertainties**

theory for (precision) electroweak observables: PDF dependence

theory predictions for EW boson production sensitive
to high-x PDFs: *e.g.*, $d(x)$ at $\boldsymbol{x} \sim \mathbf{1}$, which is poorly constrained

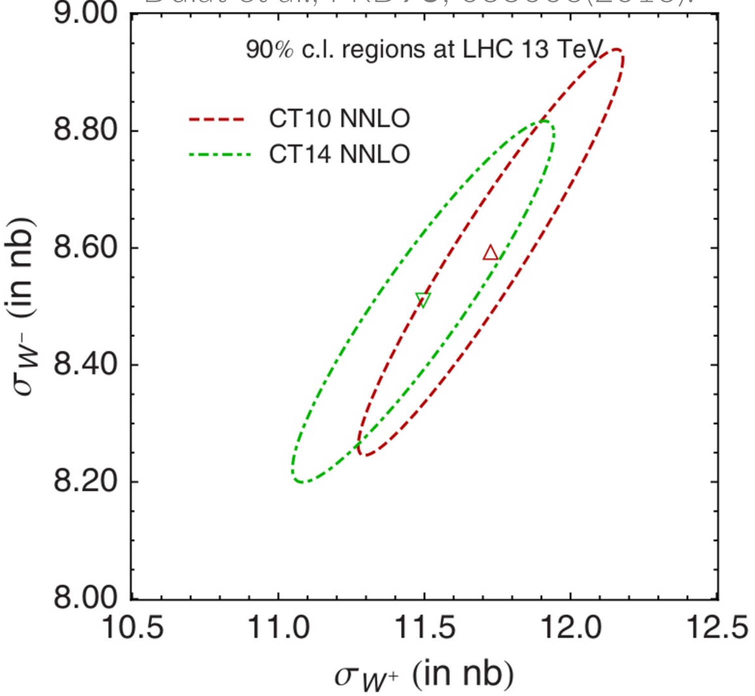
$$x_{1,2} = \frac{M}{\sqrt{s}} e^{\pm y}$$

$$\frac{d\sigma}{dy}(pp \to W^- X) = \frac{2\pi G_F}{3\sqrt{2}} x_1 x_2 \Big(\cos^2\theta_C \{ d(x_1)\bar{u}(x_2) + \bar{u}(x_1) d(x_2) \} + \sin^2\theta_C \{ s(x_1)\bar{u}(x_2) + \bar{u}(x_1) s(x_2) \} \Big)$$

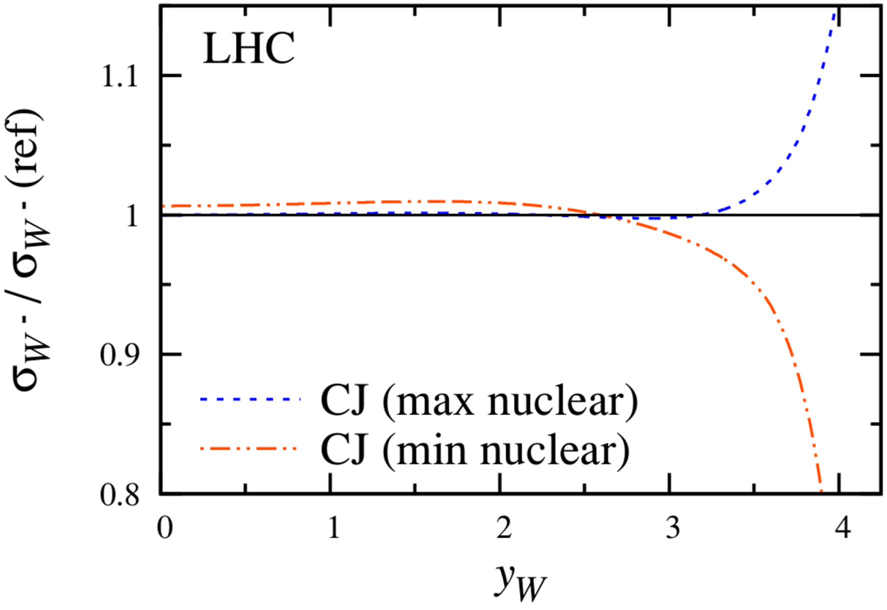
flavor separation is
crucial in these channels

see also: Accardi, TJH, Jing, Nadolsky; 2102.01107

Dulat et al., PRD**93**, 033006(2016).



Brady et al., JHEP**06** (2012) 019.

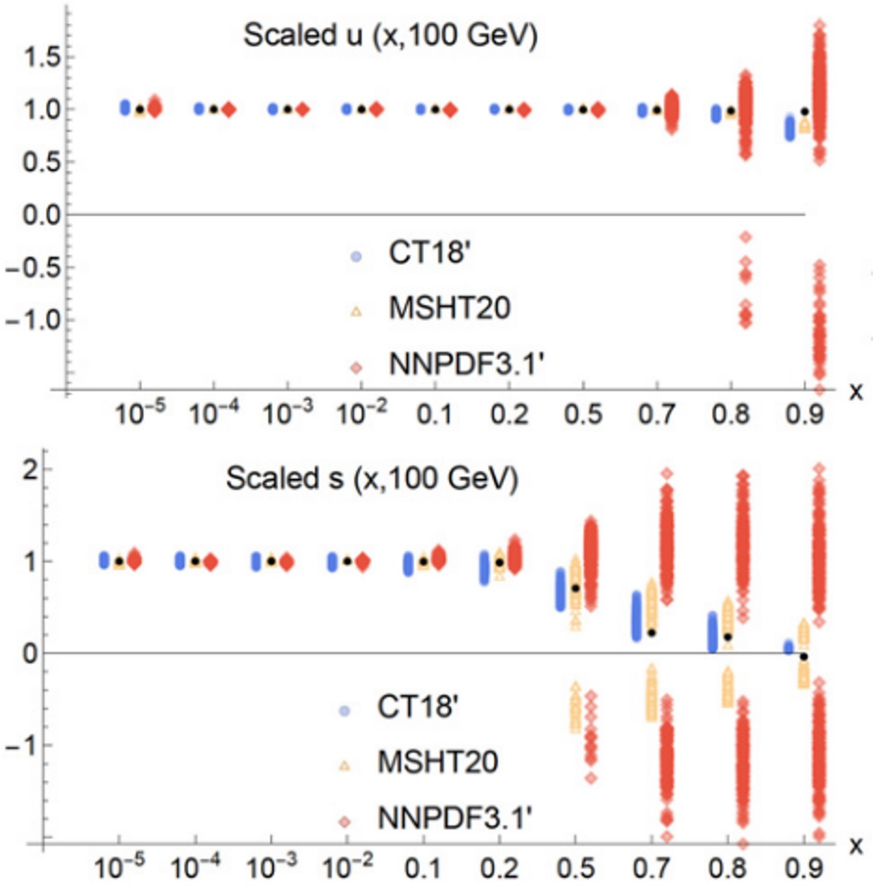


understanding PDFs and their uncertainties: high *x*

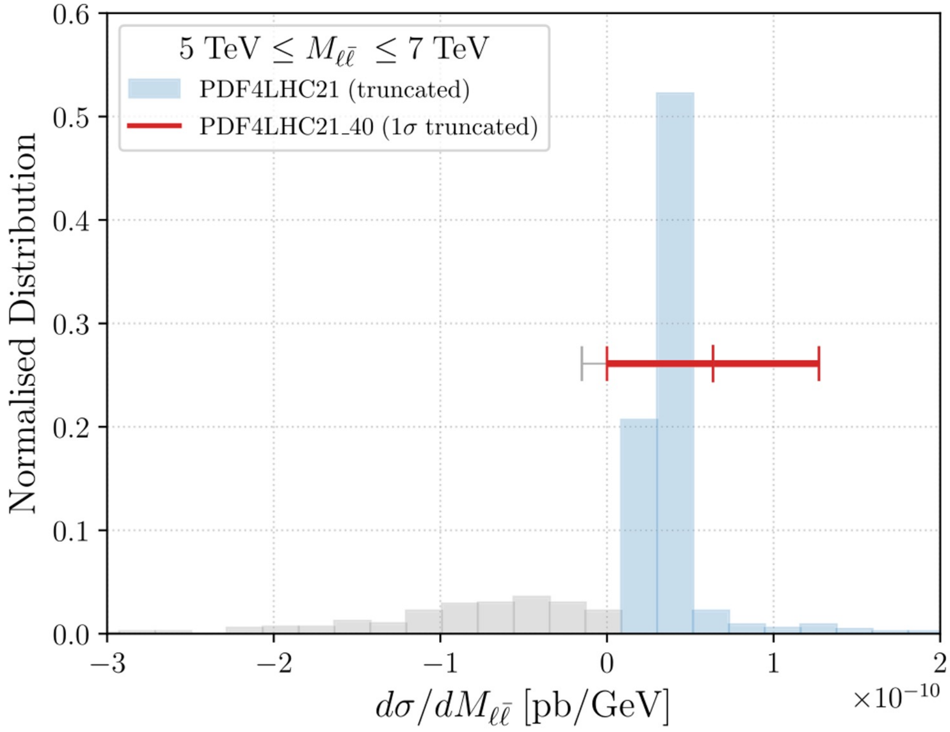
PDF4LHC21 benchmarking: J.Phys.G 49 (2022) 8, 080501.

MC sampling of high-*x* PDFs can sometimes produce irregularities

→ *e.g.*, positive-definiteness not always guaranteed for *x* → 1



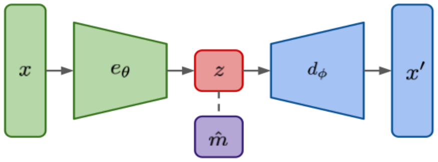
→ can produce subtle but non-negligible phenomenological consequences:



strong need for high-*x* sensitive data: (HL-)EIC

Courtoy, Huston Nadolsky, Xie, Yan, Yuan: 2205.10444

aside: PDF parametrizations, ML generative models

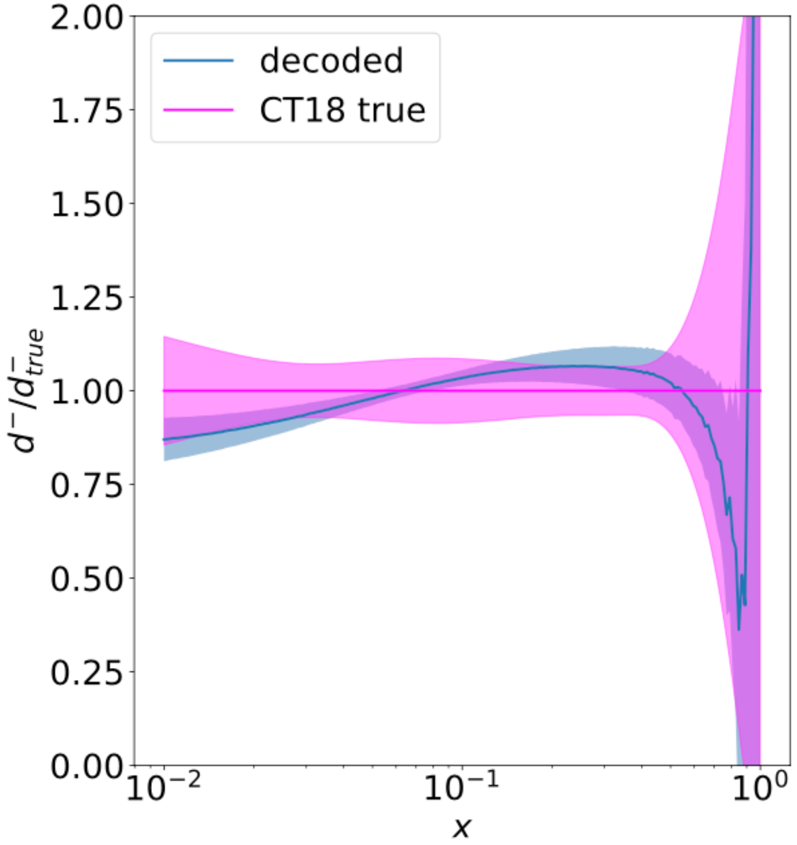
Name	Diagram	Loss	Recreates PDFs	Tractable Latent	Free Latent Dimension	Moment Constraint
AE-CL		$\mathcal{L} = \|x - d_\phi(e_\theta(x))\|_2^2 + \|z - \hat{m}\|_2^2$	✓	✓	✓	✓

→ parametrization dependence, MC uncertainty sampling are serious bottlenecks to PDF precision

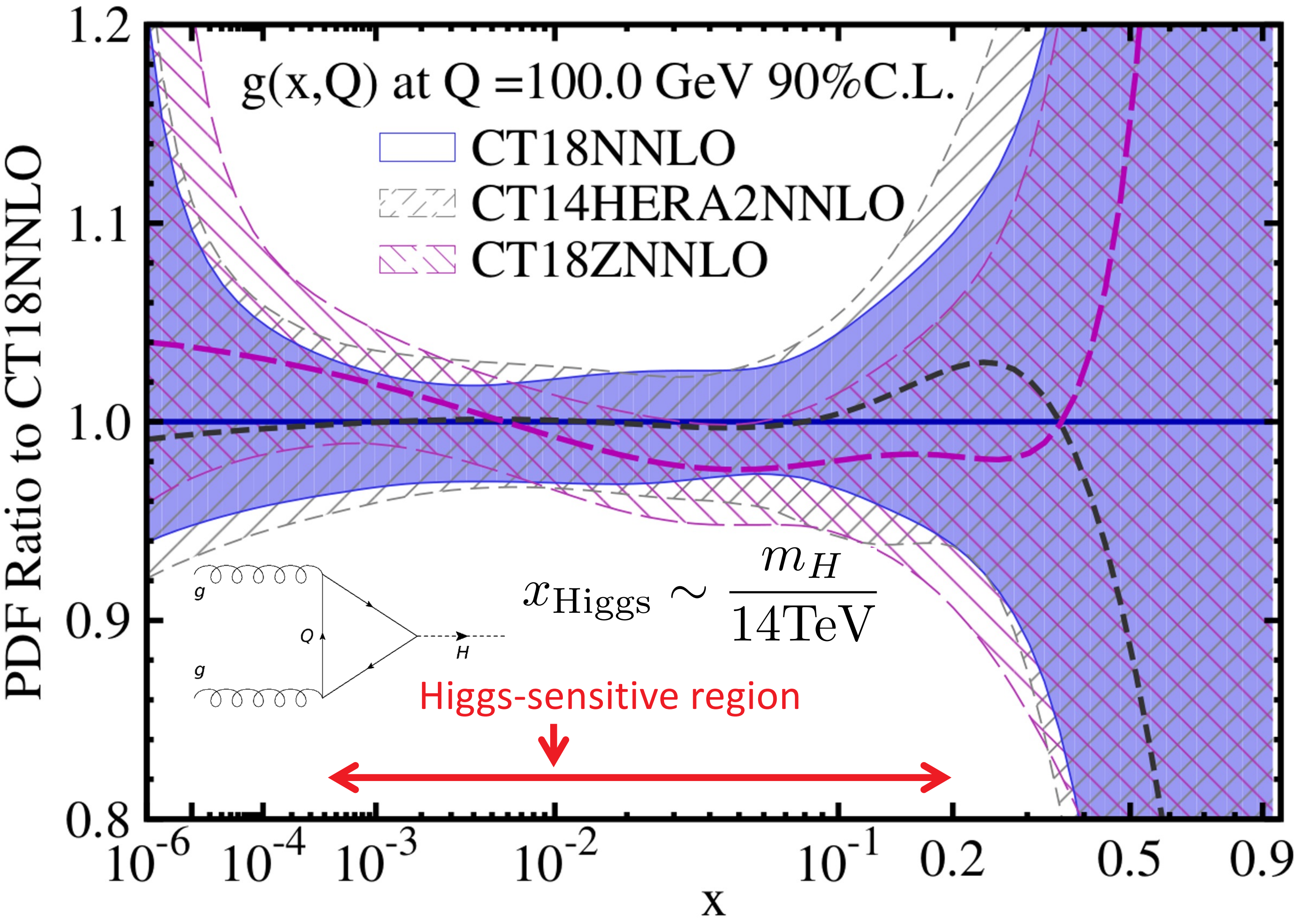
- ML models (encoder-decoder networks): testbeds to disentangle PDF parametrizations

  → can generatively predict PDFs

Kriesten, TJH; 2312.02278



outside extrapolation region, PDF errors still challenging

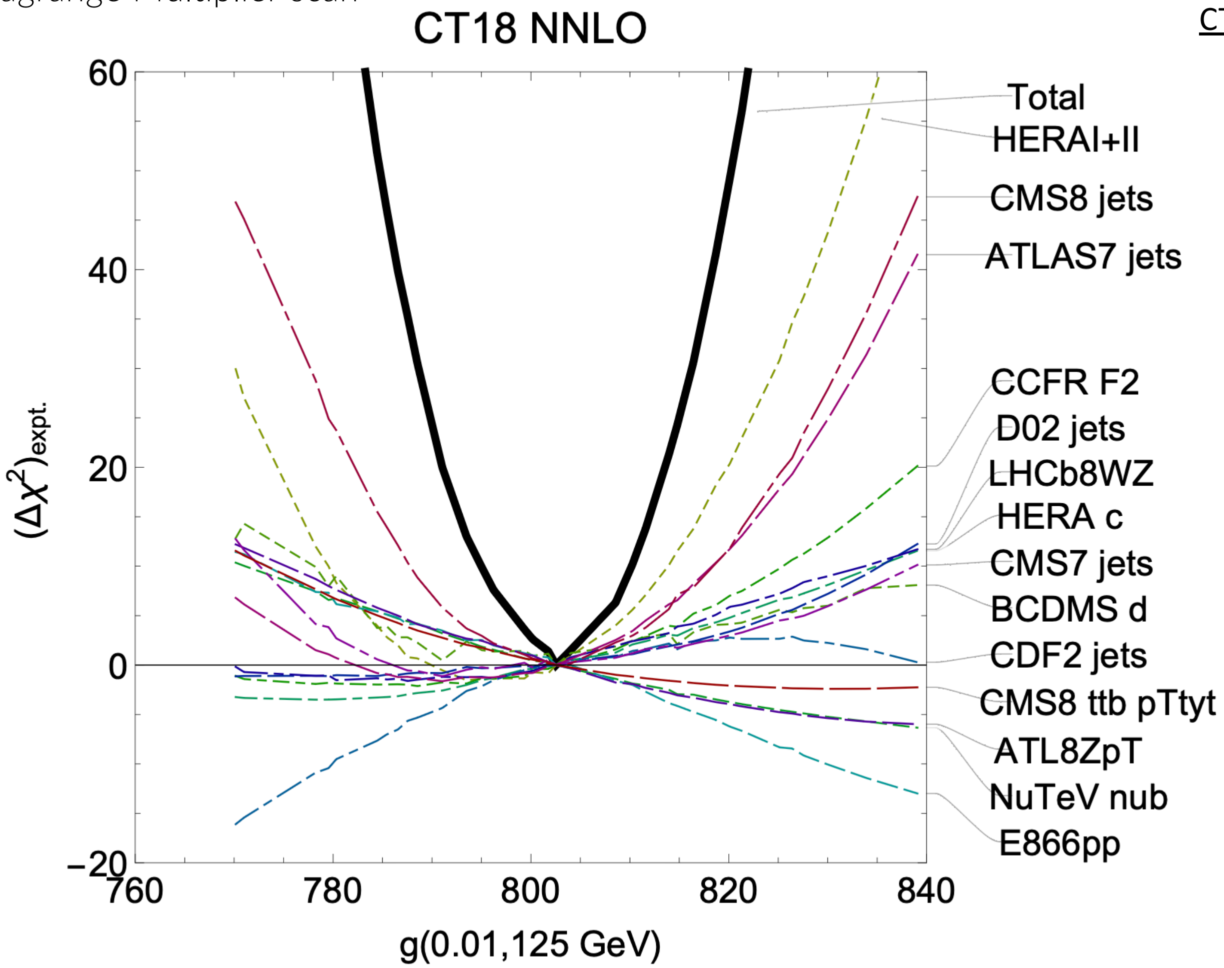


knowledge of the gluon content of the nucleon directly translates into constraints on SM Higgs production

compatibility of fitted data sets is a crucial question

tensions among individual fitted experiments drive a larger PDF uncertainty

Lagrange Multiplier scan



CT18: PRD**103** (2021) 1, 014013

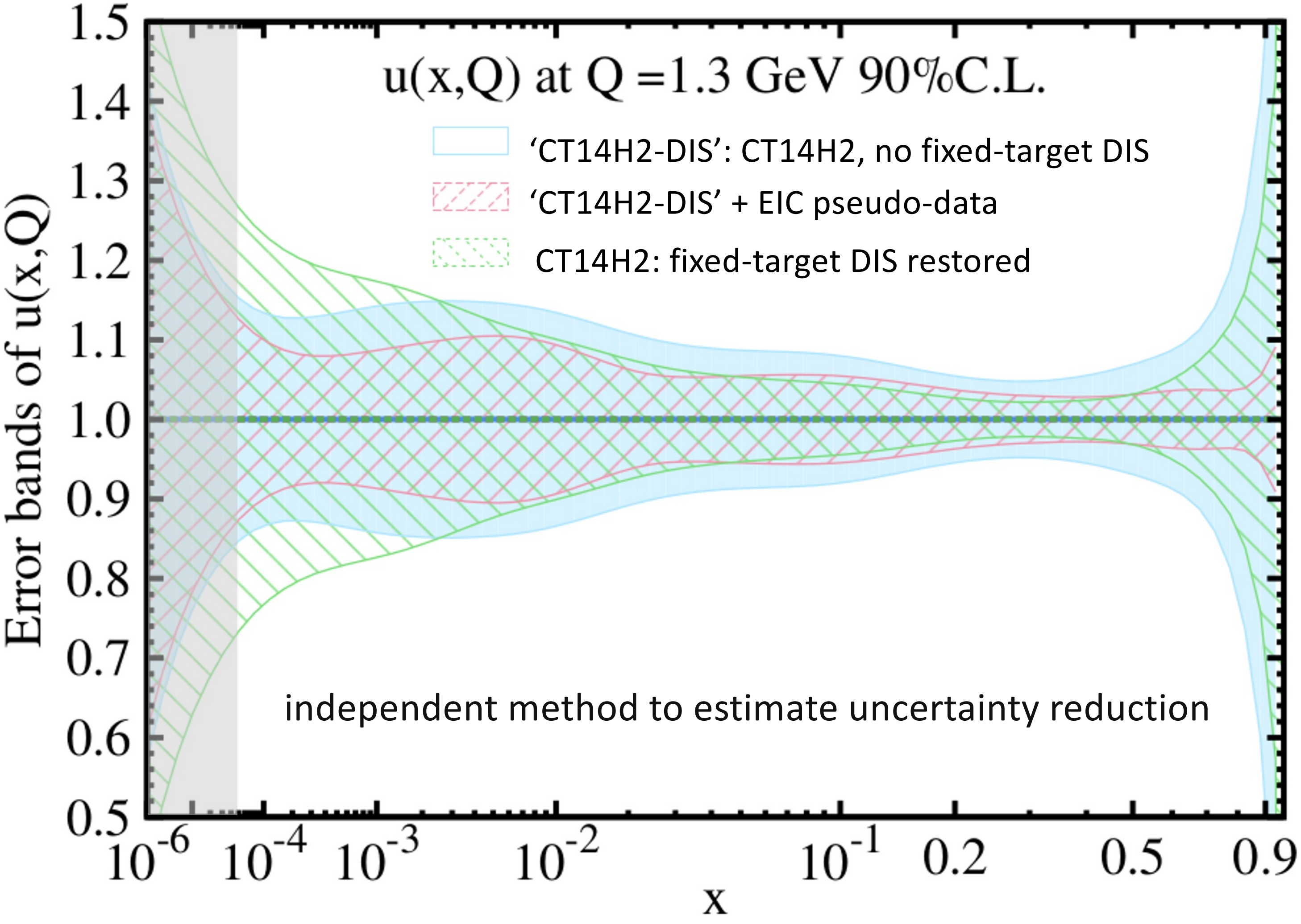
examine change in χ^2 as PDF continuously varies away from fitted central value

larger gluon… and Higgs cross section… favored by some expts [like E866pp], but not others [like 8 TeV CMS jets]

serious impediment to higher precision in PDFs and resulting theory predictions

EIC sensitive to PDFs → strong HEP implications

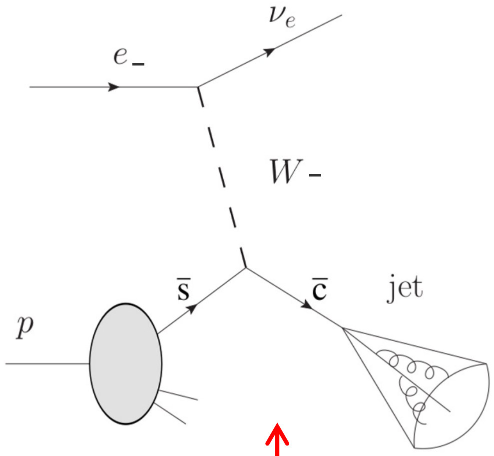
[impact studies made with error profiling methods]



independent method to estimate uncertainty reduction

DIS charm-jet prod.

Arratia, Furletova, TJH, Olness, Sekula
PRD **103** (2021) 7, 074023



1-yr inclusive EIC dataset drives steep reductions in PDF uncertainties

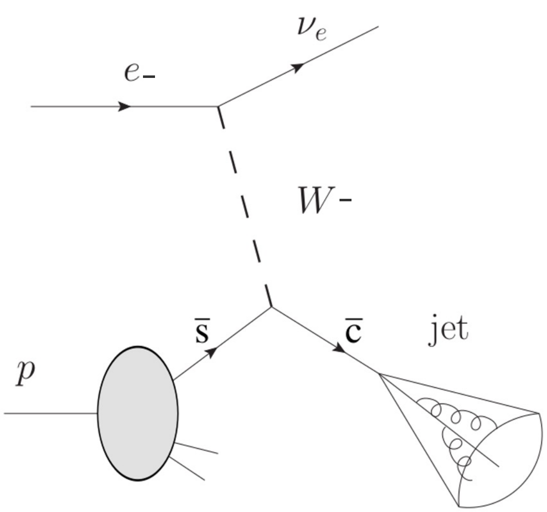
→ just inclusive DIS; many other channels with PDF sensitivity; precision QCD tests

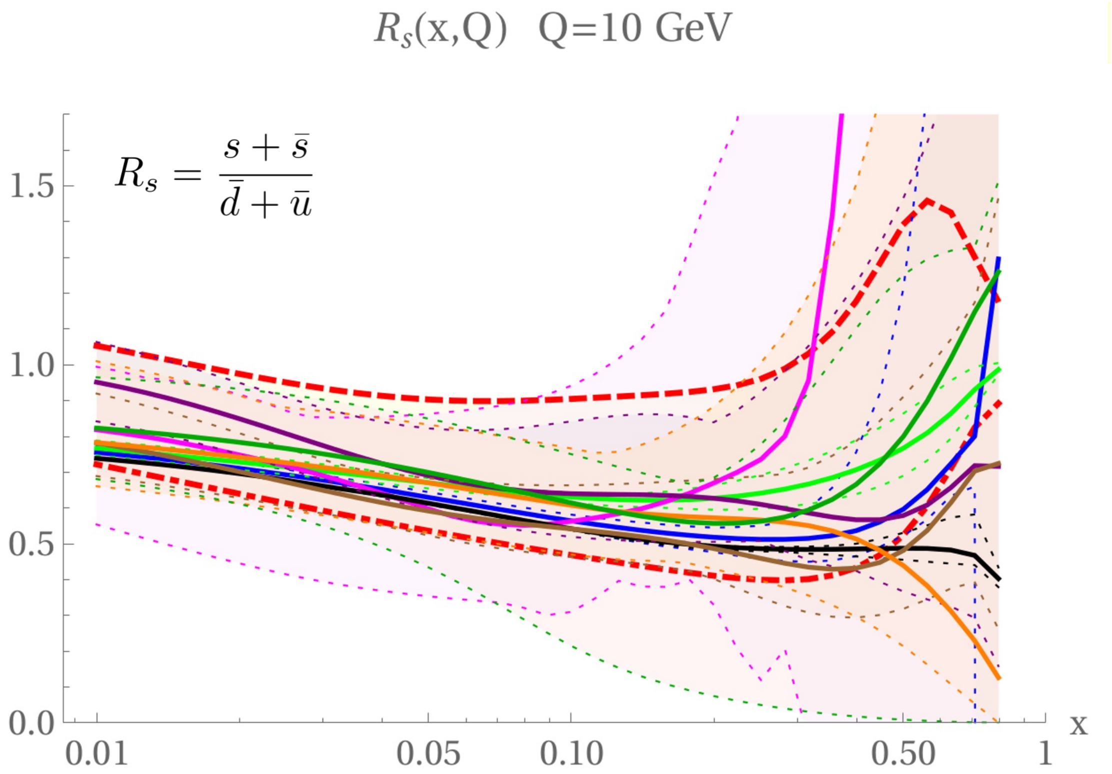
precision QCD through jet and heavy-flavor production

DIS jet production, including through charge-current interactions, provides further access to quark-level information

Arratia, Furletova, TJH, Olness, Sekula; PRD **103** (2021) 7, 074023.

100 fb^{-1} CC DIS (10M simulated events), at 10x275 GeV (e^- on p); Q^2 > 100 GeV^2





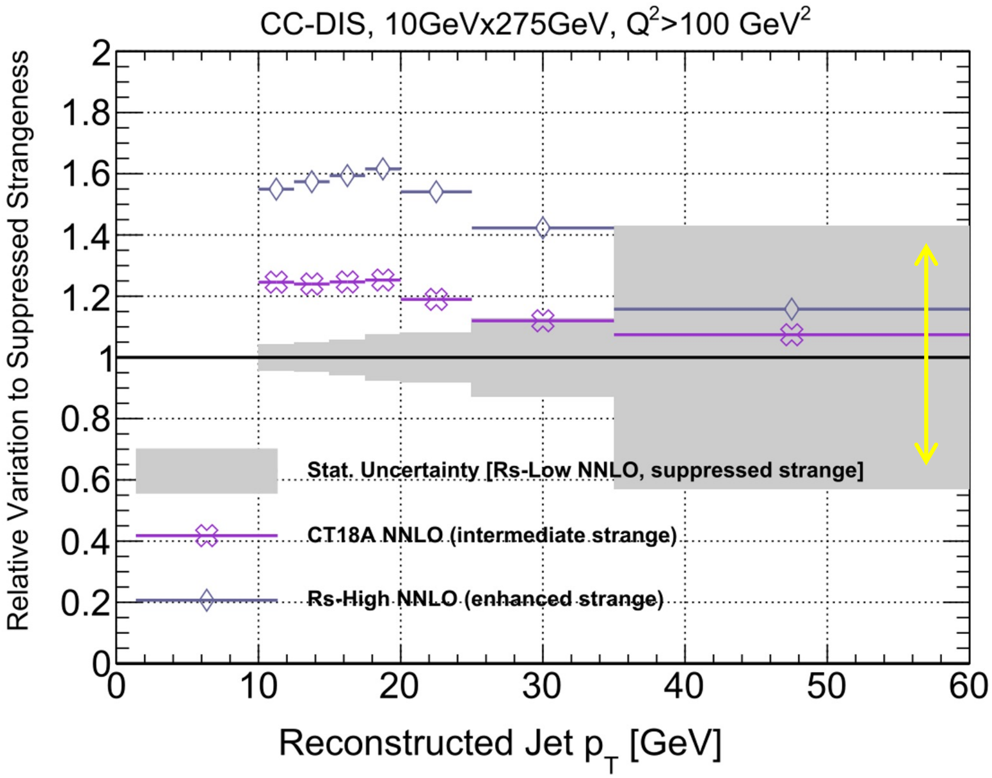
final-state tagging provides lever arm for flavor separation (here, strangeness)

n.b.: event generation, detector sim from PYTHIA8 + DELPHES; FASTJET reconstruction

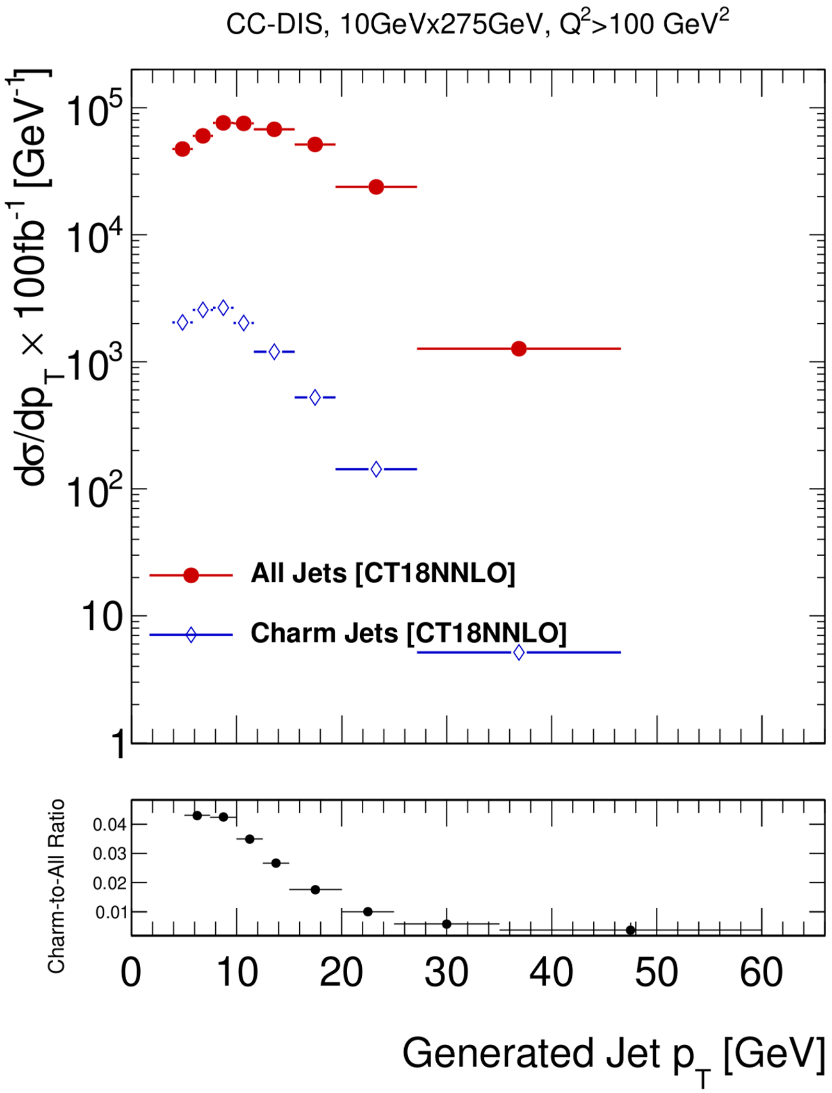
→ analogous jet measurements might be extended to nonperturbative heavy flavor

precision QCD through jet and heavy-flavor production

challenging measurement: final-state flavor tagging; Jacquet-Blondel reconstruction



Arratia, Furletova, TJH, Olness, Sekula; PRD **103** (2021) 7, 074023.



reduced δ_{stat} : enhance knowledge of p_T spectrum

→ appropriate for ePIC follow-up

- EIC PDF sensitivity will derive from multi-channel data; inclusive, tagging, …

→ *indirect* BSM impacts through refined SM baselines

meanwhile at colliders: BSM through EFT fits

- strong recent interest: model-independent BSM analyses

EFT-based parametrizations: *e.g.*, SM effective field theory (SMEFT)

conceptually similar to PDF-fitting ideology

- EFT global analyses often assumed *fixed* SM calculations

→ PDFs not actively fitted alongside SMEFT parameters

→ could potentially bias resulting SMEFT analysis

- solution: develop **joint SMEFT/PDF fits** (field at early stage)

Gao, Gao, TJH, Liu, and Shen; JHEP05 (2023) 003

→ this work: example in context of CTEQ-TEA (CT) framework

→ demonstration study focusing on select data: $\mathrm{jet}, t\bar{t}\ \mathrm{production}$

→ examine possible PDF-SMEFT correlations

large body of recent EFT analyses

- numerous SMEFT fitting frameworks and studies in recent years

$\texttt{EFTFitter}^1$, $\texttt{FitMaker}^2$, $\texttt{HEPFit}^3$, $\texttt{SFitter}^4$, $\texttt{SMEFiT}^5$, . . .

[1]Castro et al, 1605.05585; [2]Ellis et al, 2012.02779; [3]De Blas et al, 1910.14012;
[4]Brivio et al, 1910.03606; [5]Hartland et al, 1901.05965; …

→ dedicated EFT fits within CMS, ATLAS

→ studies related to current, future expts; *e.g.*, DIS colliders, LHC, …

multiple talks at this meeting

⚠

- disclaimer: this talk → *one example* of recent PDF-SMEFT activity

→ *e.g.*, recent fits within the NNPDF-SMEFiT frameworks

Kassabov et al.: top-quark joint fit, 2303.06159

SIMUnet public code: Costantini et al, 2402.03308

→ findings here are **representative** of current PDF-SMEFT situation

basics of (SM)EFT

see talks, Boughezal, Cirigliano, …

- presume BSM scale above the electroweak, $\Lambda \gg M_{W,Z}$

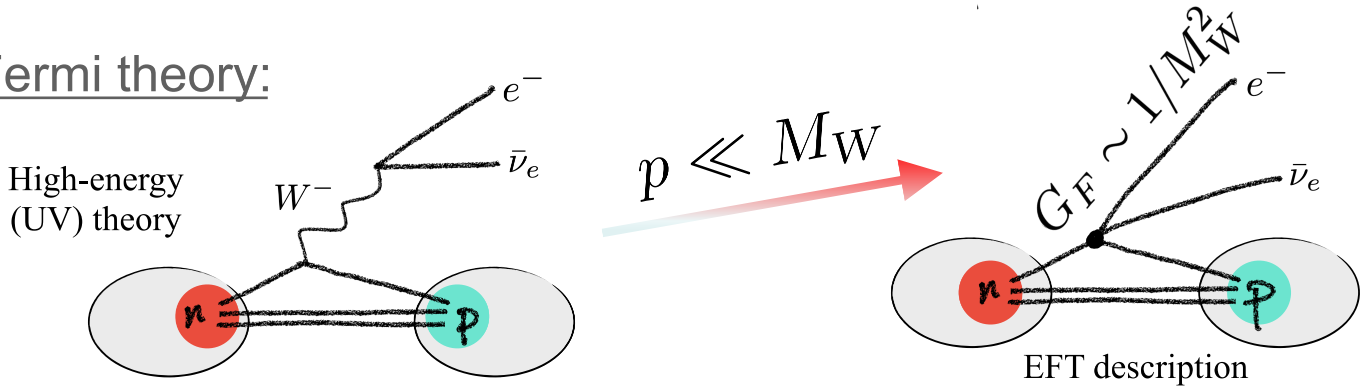
  → explicit non-standard degrees-of-freedom may be integrated away

  → leaves behind basis of higher-dimensional operators

$$\mathcal{L}_{\rm SMEFT} = \mathcal{L}_{\rm SM} + \sum_i \frac{C_i O_i^{(6)}}{\Lambda^2} + \ldots$$

built from SM field content!

analogous EFT – Fermi theory:



- BSM quantified via nonzero Wilson coefficients, $C_i \neq 0$

→ extract from global fit alongside PDFs

selecting dominant SMEFT operators

- this study: dim-6 operators only
  - → dim-8 contributions small (may be relevant for future precision!)
  - → consider several SMEFT operators associated with jet, $t\bar{t}$

jet production: contact interaction

$$O_1 = 2\pi \left(\sum_{i=1}^{3} \bar{q}_{Li} \gamma_\mu q_{Li} \right) \left(\sum_{j=1}^{3} \bar{q}_{Lj} \gamma^\mu q_{Lj} \right)$$

Warsaw operator basis

top production

$$\begin{aligned} O^1_{tu} &= \sum_{i=1}^{2} (\bar{t}\gamma_\mu t)\,(\bar{u}_i \gamma^\mu u_i)\ , \\ O^1_{td} &= \sum_{i=1}^{3} (\bar{t}\gamma^\mu t)\,(\bar{d}_i \gamma_\mu d_i)\ , \\ O_{tG} &= i g_s (\bar{Q}_{L,3} \tau^{\mu\nu} T^A t) \tilde{\varphi} G^A_{\mu\nu} + \text{h.c.}\,, \\ O^8_{tq} &= \sum_{i=1}^{2} (\bar{Q}_i \gamma_\mu T^A Q_i)(\bar{t}\gamma^\mu T^A t)\,, \end{aligned}$$

- have imposed multiple symmetries on SMEFT space

theory calculation setup

- nonzero Wilson coeffs.: finite SMEFT contributions to X-sections

  → pure SM, pure dim-6 SMEFT, and *interference* pieces:

$$\frac{d\sigma}{d\hat{O}} = \frac{d\sigma_{\rm SM}}{d\hat{O}} + \sum_i \frac{d\tilde{\sigma}_i}{d\hat{O}}\frac{C_i}{\Lambda^2} + \sum_{i,j}\frac{d\tilde{\sigma}_{ij}}{d\hat{O}}\frac{C_iC_j}{\Lambda^4}$$

  → relevant for interference term, SMEFT-QCD computed to NLO

  → constrain Wilson coefficients, $C(\mu_c)/\Lambda^2$; for $\mu_c = 1\,\mathrm{TeV}$

- status of theory calculations, uncertainties for all processes:

observable	μ_0	SM QCD	SM EW	SMEFT QCD	th. unc.
$t\bar{t}$ total	m_t	NNLO+NNLL	no	NLO	$\mu_{F,R}$ var.
$t\bar{t}$ p_T dist.	$m_T/2$	NNLO	NLO	NLO	$\mu_{F,R}$ var.
$t\bar{t}$ $m_{t\bar{t}}$ dist.	$H_T/4$	NNLO(+NLP)	NLO	NLO	$\mu_{F,R}$ var.
$t\bar{t}$ 2D dist.	$H_T/4$	NNLO	no	NLO	no
inc. jet	$p_{T,j}$	NNLO	NLO	NLO	0.5% uncor.
dijet	m_{jj}	NNLO	NLO	NLO	0.5% uncor

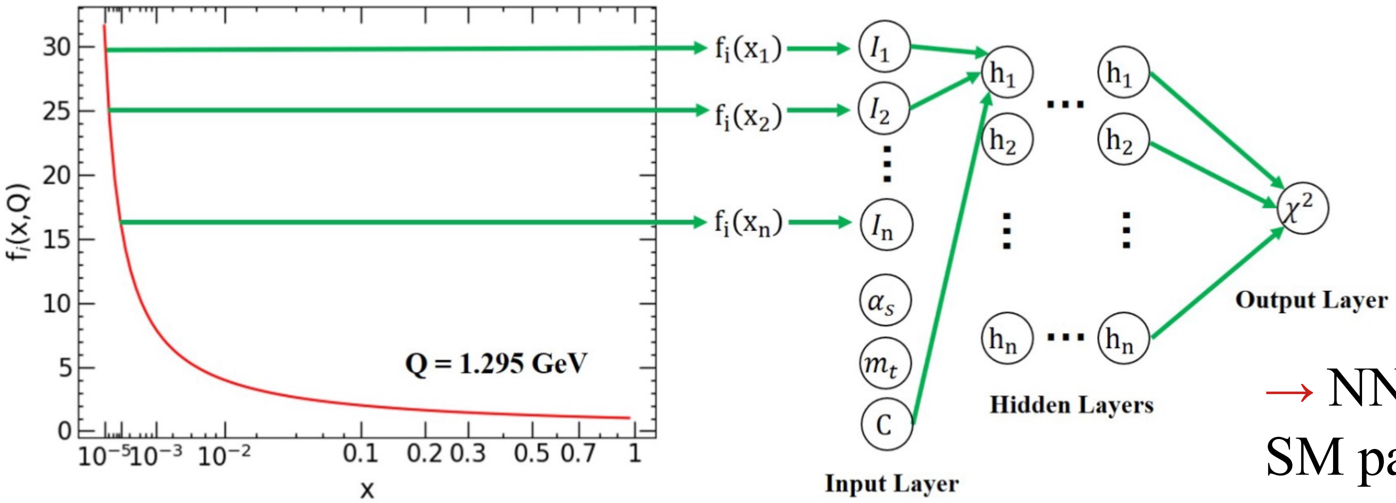
learning log-likelihoods through neural net training

- quantify agreement of theory/data through χ^2 :

$$\chi^2(\{a_\ell\},\{\lambda\}) = \sum_{k=1}^{N_{\mathrm{pt}}} \frac{1}{s_k^2}\left(D_k - T_k(\{a_\ell\}) - \sum_{\alpha=1}^{N_\lambda}\beta_{k,\alpha}\lambda_\alpha\right)^2 + \sum_{\alpha=1}^{N_\lambda}\lambda_\alpha^2$$

→ train a feed-forward neural network (NN) on PDF replicas

Liu, Sun, and Gao; arXiv: 2201.06586

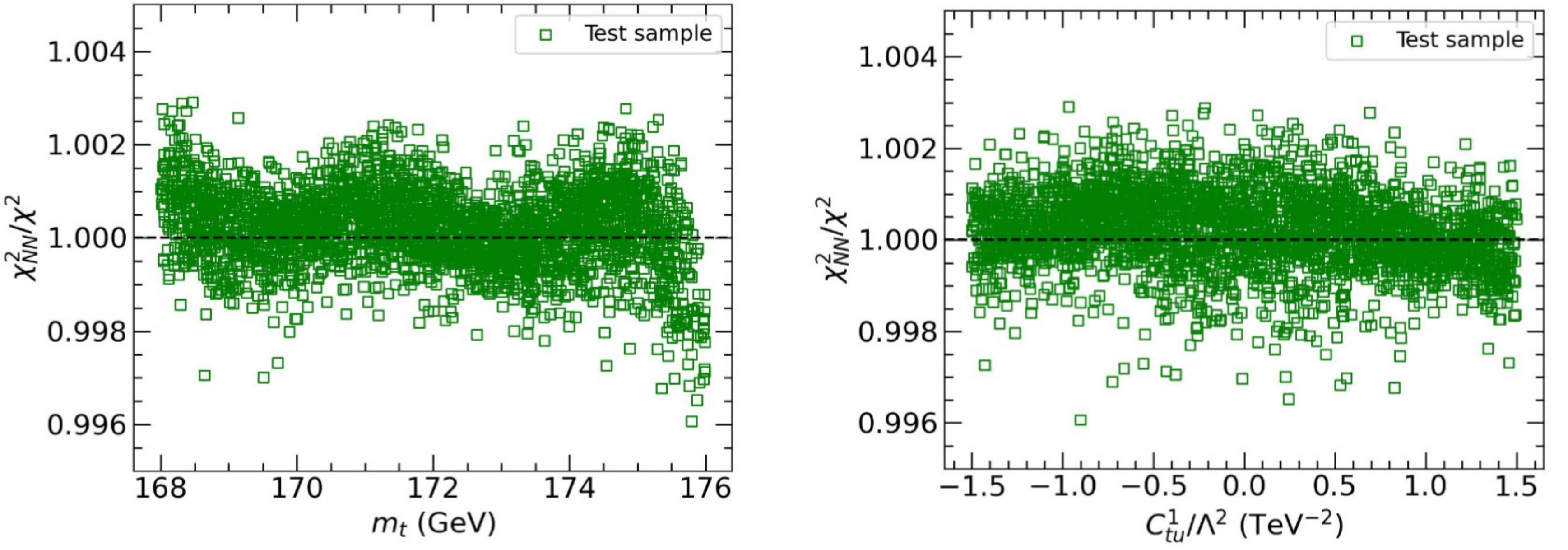


→ NN associates PDFs, SM parameters, SMEFT coefficients with χ^2

NNs effectively learn (PDF-SMEFT) likelihood function

- generate 1.2×10^4 replicas over PDFs, SM parameters, SMEFT coeffs.

  → validate performance on 4×10^3 test set



  → strong, permille-level agreement achieved!

(NB: perfect agreement corresponds to $\chi^2_{\rm NN}/\chi^2 = 1$)

- allows *rapid* exploration of combined PDF-SMEFT uncertainties

explore constraints from range of LHC expts

- included on top of default CT18 fitted experiments

→ nominally fit ~112 fb^{-1} of top data; ~67 fb^{-1} for jet production

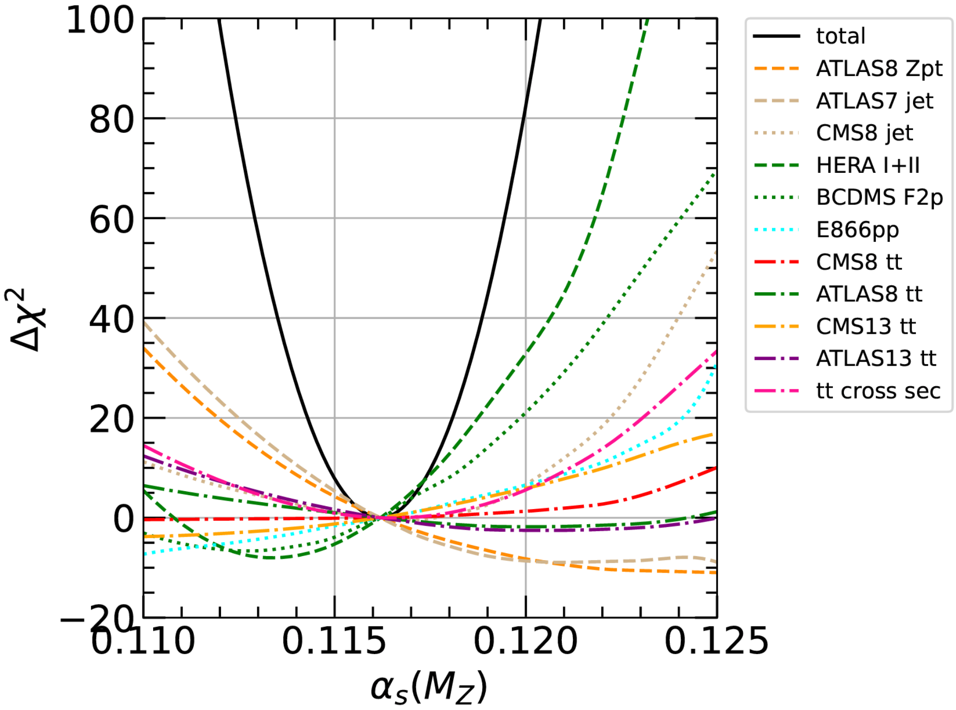
Experiments	$\sqrt{s}$(TeV)	$\mathcal{L}(\text{fb}^{-1})$	observable	N_{pt}
$*^\dagger$ LHC(Tevatron)	7/8/13(1.96)	—	$t\bar{t}$ total cross section	8
$*^\dagger$ ATLAS $t\bar{t}$	8	20.3	1D dis. in $p_{T,t}$ or $m_{t\bar{t}}$	15
$*^\dagger$ CMS $t\bar{t}$	8	19.7	2D dis. in $p_{T,t}$ and y_t	16
CMS $t\bar{t}$	8	19.7	1D dis. in $m_{t\bar{t}}$	7
$*^\dagger$ ATLAS $t\bar{t}$	13	36	1D dis. in $m_{t\bar{t}}$	7
$*^\dagger$ CMS $t\bar{t}$	13	35.9	1D dis. in $m_{t\bar{t}}$	7
$*^\dagger$ CDF II inc. jet	1.96	1.13	2D dis. in p_T and y	72
$*^\dagger$ D0 II inc. jet	1.96	0.7	2D dis. in p_T and y	110
$*^\dagger$ ATLAS inc. jet	7	4.5	2D dis. in p_T and y	140
$*^\dagger$ CMS inc. jet	7	5	2D dis. in p_T and y	158
$*$ CMS inc. jet	8	19.7	2D dis. in p_T and y	185
† CMS dijet	8	19.7	3D dis. in $p_T^{ave.}$, y_b and y^*	122
† CMS inc. jet	13	36.3	2D dis. in p_T and y	78

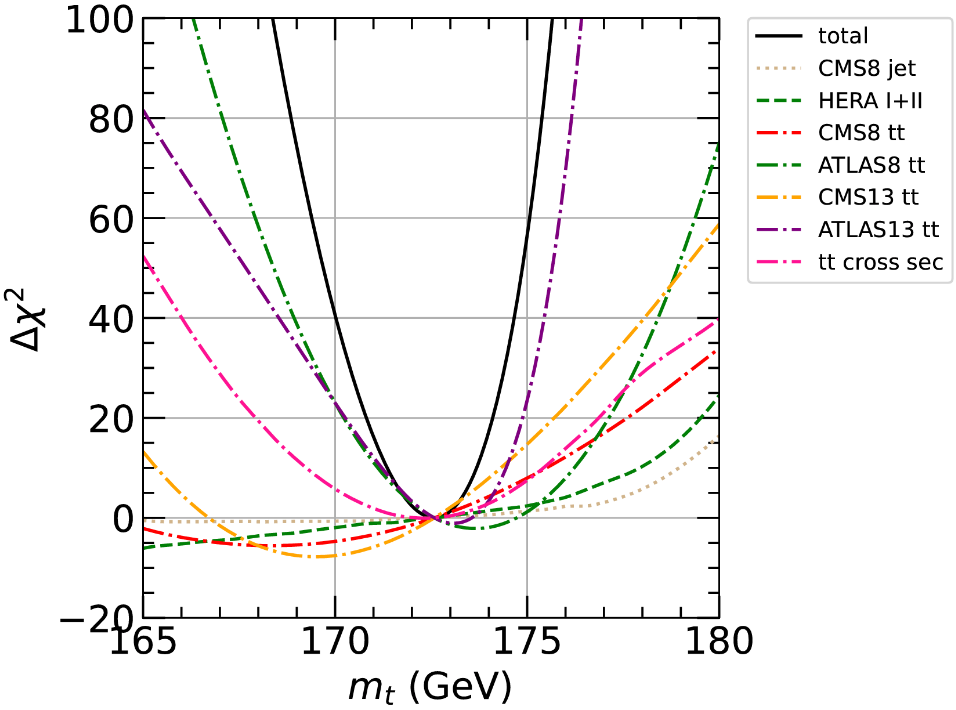
*(in nominal top fits); †(in nominal jet fits)

- for top data, correlated theory errors included via nuisance parameters; uncorr. for jets

first: sensitivity to SM QCD parameters

- simultaneously fit α_s and m_t in the absence of nonzero SMEFT coefficients





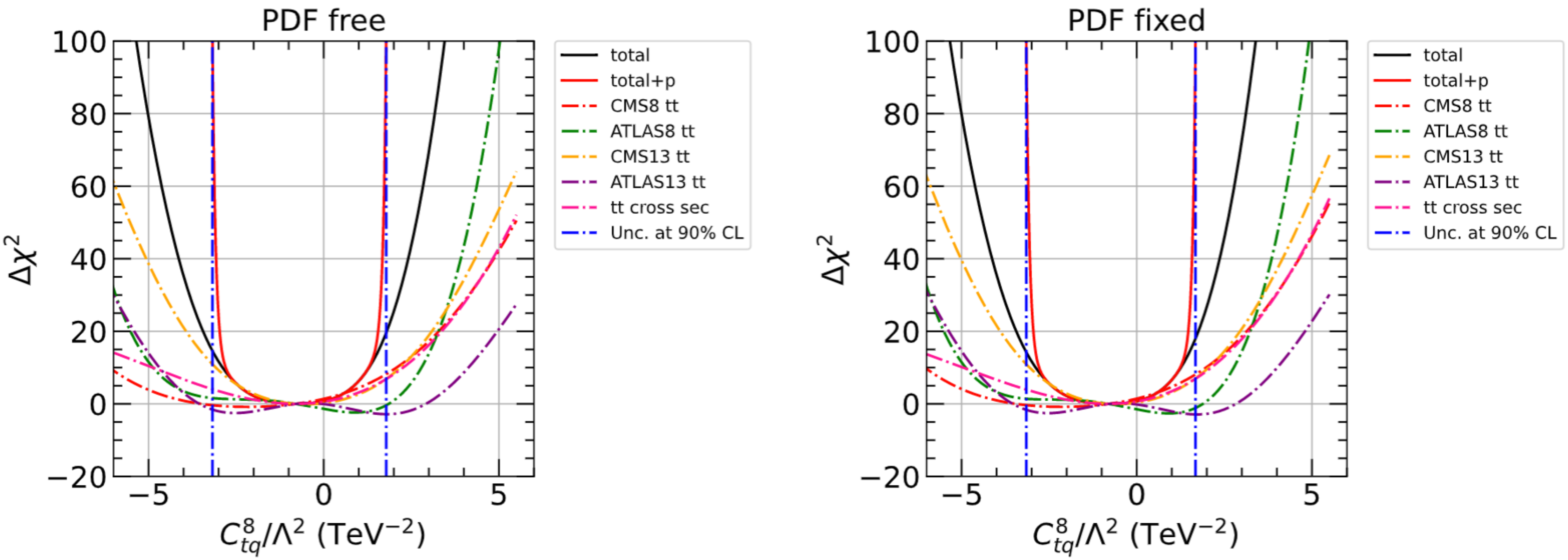
- combined jet and $t\bar{t}$ data significantly constrain strong coupling, top mass

→ competing pulls between CMS, ATLAS $t\bar{t}$ data; lower vs higher top mass, resp.

(consistent with constraints from first near-threshold $m_{t\bar{t}}$ bin)

examine SMEFT uncertainties in joint PDF fit $t\bar{t}$ data

- quantify SMEFT uncert. through Lagrange Multiplier (LM) scans:



→ constraints to top-associated Wilson coefficient, C^8_{tq}/Λ^2

- modest increase in uncertainty when co-fitted with PDFs
- predominantly *quartic* shapes for $\Delta\chi^2$ reflect pure SMEFT contributions $\sim\dfrac{1}{\Lambda^4}$

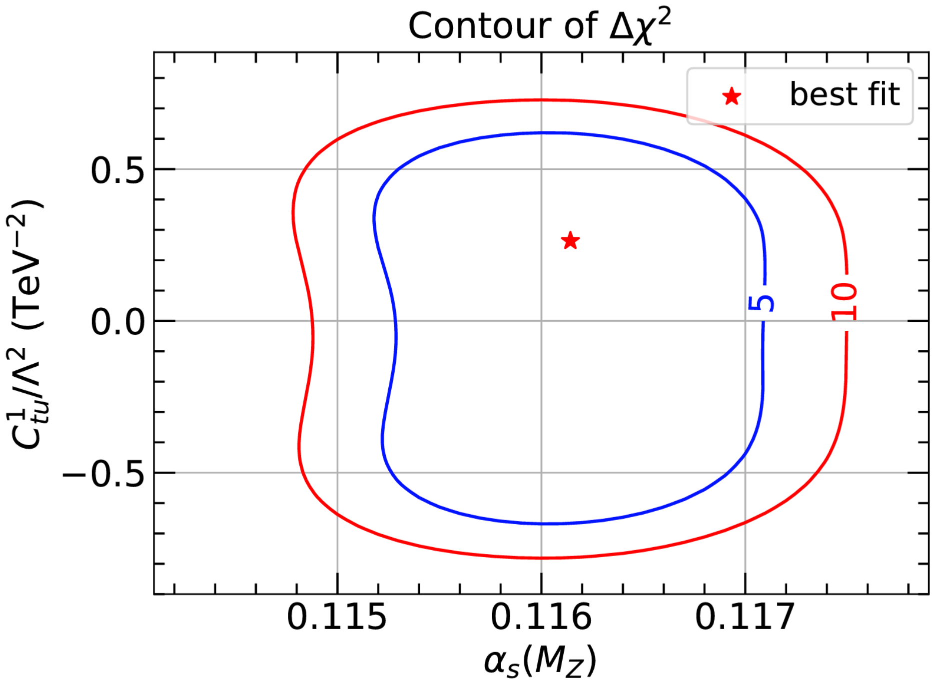
… *i.e.*, importance of quadratic EFT terms in limit-setting

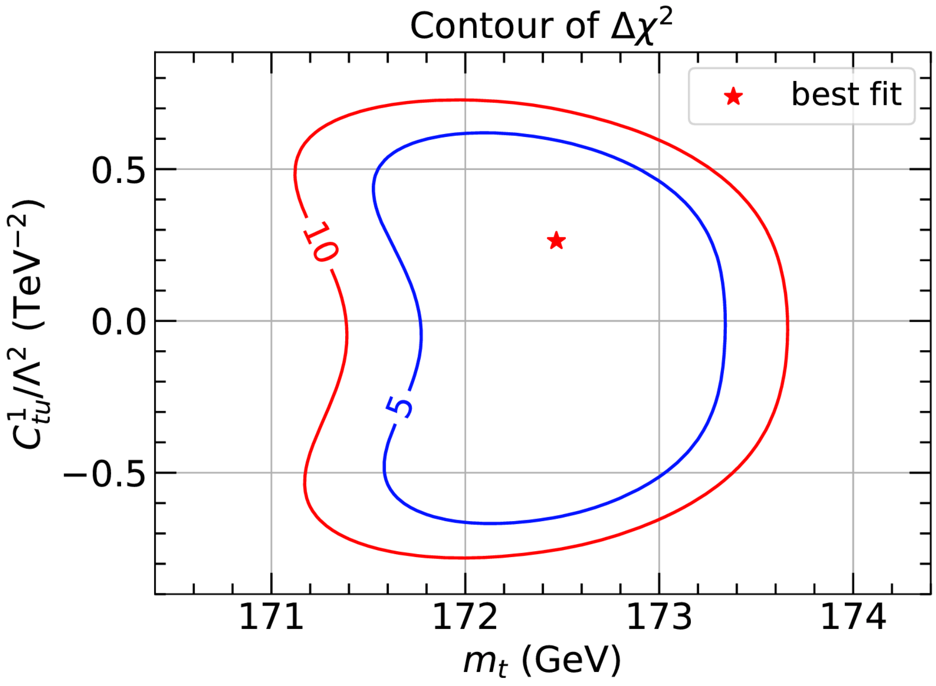
pQCD correlations with top EFT coefficients

$t\bar{t}$ data

→ 2D Lagrange multiplier scans reveal correlations between fitted quantities; minimize:

$$\Psi(\lambda_1, \lambda_2, \{a_\ell\}) = \chi^2(\{a_\ell\}) + \lambda_1 X_1(\{a_\ell\}) + \lambda_2 X_2(\{a_\ell\})$$





- best fit coupling, mass essentially unshifted when fitted alongside EFT

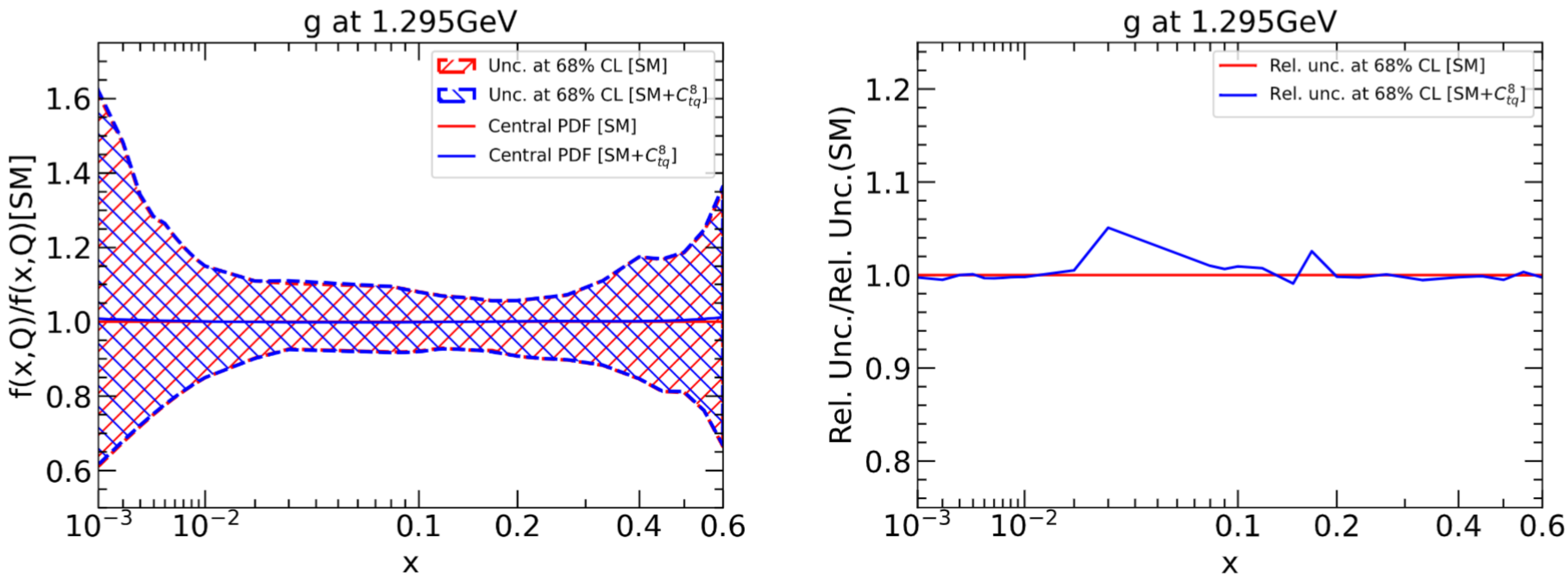
  → still, some deviation from rotational symmetry: non-quadratic profiles for χ^2

joint fits: very weak correlations with PDFs’ *x* dependence

- SMEFT coefficient uncertainties depend on active fitting of PDFs:

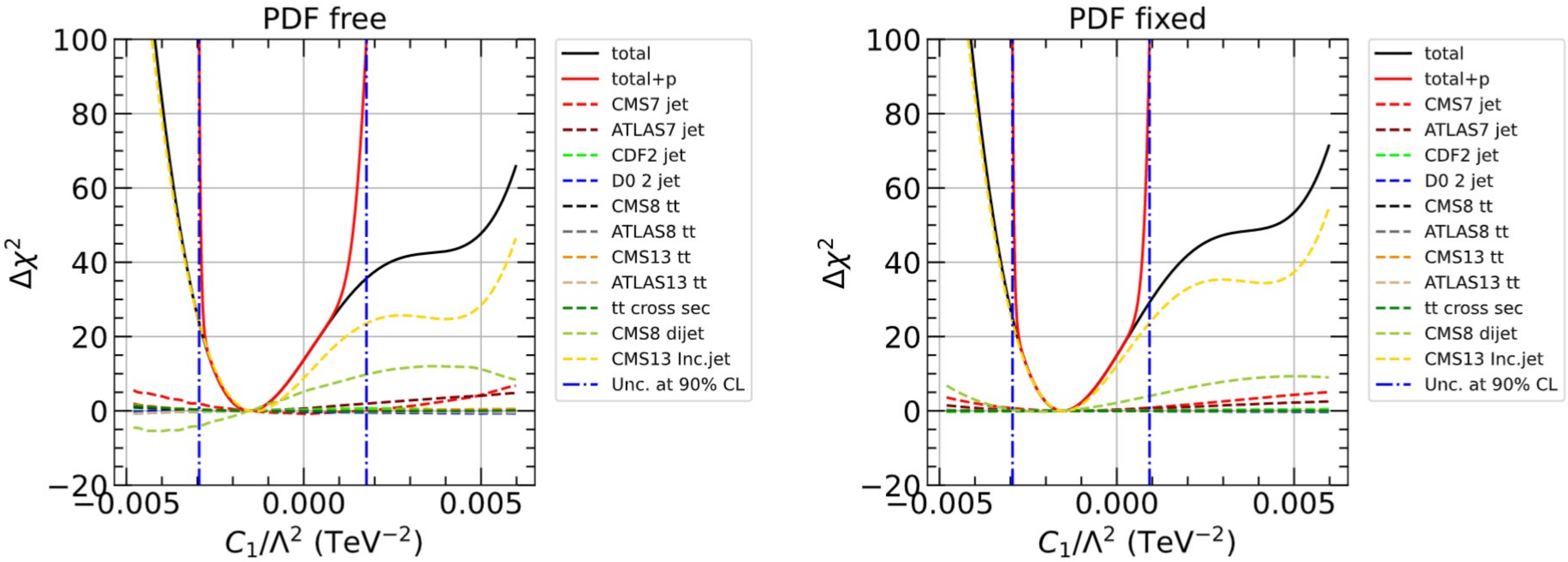
TeV^{-2}	nominal	PDF fixed	no the. unc.
C^{1}_{tu}/Λ^2	$0.14^{+0.61}_{-0.97}$	$0.14^{+0.60}_{-0.95}$	$0.14^{+0.57}_{-0.92}$
C^{8}_{tq}/Λ^2	$-0.80^{+2.58}_{-2.38}$	$-0.80^{+2.48}_{-2.35}$	-
C_{tG}/Λ^2	$-0.10^{+0.26}_{-0.30}$	$-0.10^{+0.25}_{-0.30}$	-

- small variations in gluon PDF, unc. from co-fitting SMEFT:



analogous joint fits: jet data and contact interaction

- fitted jet data modestly sensitive to C_1:

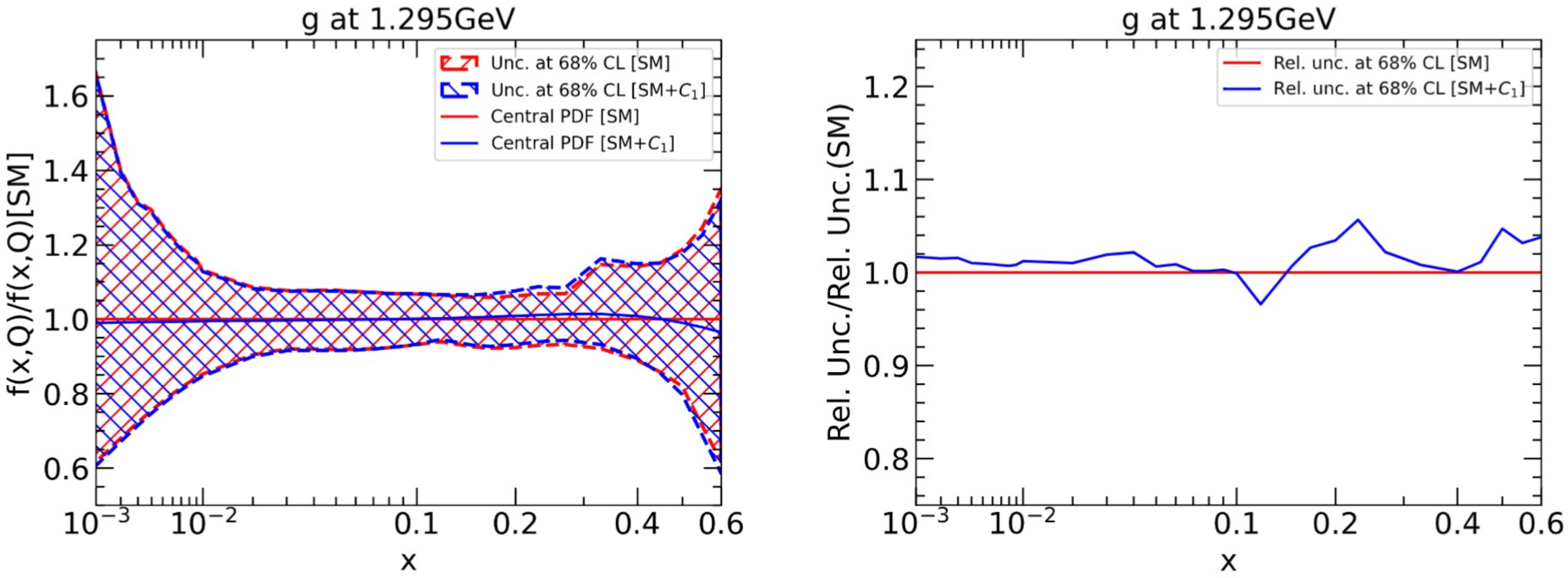


- leading SMEFT sensitivity from CMS: 13 TeV incl. jet, 8 TeV dijet data
- fixing PDFs: (slightly) larger uncertainty underestimate

TeV^{-2}	nominal	CMS 8 dijet	CMS 8 jet	CMS 13 jet
PDF free	$-0.0015^{+0.0033}_{-0.0014}$	$-0.0022^{+0.0187}_{-0.0054}$	$-0.0009^{+0.0138}_{-0.0045}$	$-0.0013^{+0.0059}_{-0.0016}$
PDF fixed	$-0.0015^{+0.0024}_{-0.0014}$	$-0.0022^{+0.0180}_{-0.0051}$	$-0.0009^{+0.0131}_{-0.0049}$	$-0.0013^{+0.0026}_{-0.0015}$

more pronounced PDF correlation for jet data

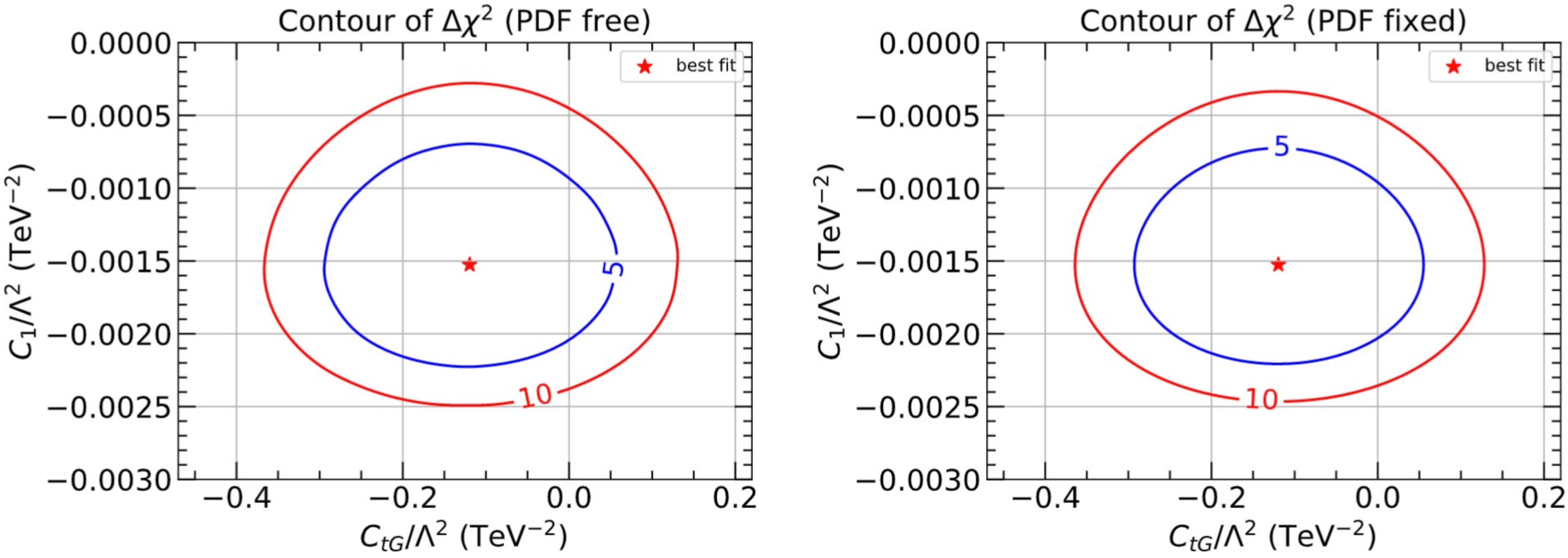
- jointly fitting contact interaction to jet prod. shifts gluon PDF



- effect somewhat greater at large $x > 0.1$
- suggests slightly stronger correlation of gluon PDF with C_1

correlations between SMEFT coefficients are mild

- co-varying top-, jet-associated coeffs. minimally effects uncertainties

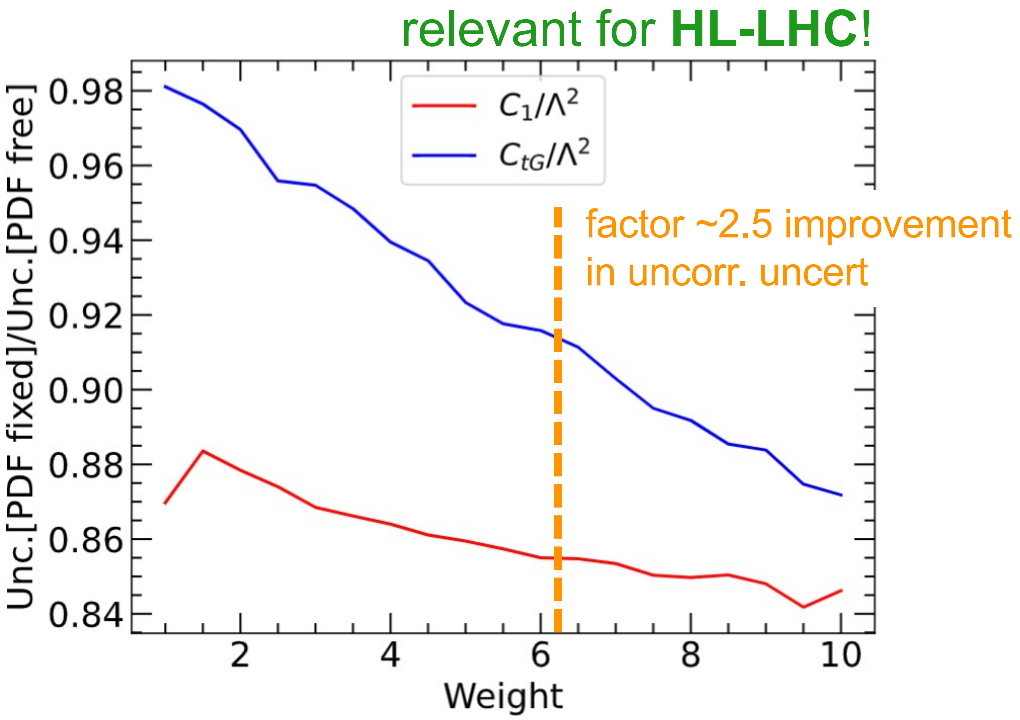


→ strongly rotationally-symmetric $\Delta\chi^2$ contours imply very weak correlations

TeV^{-2}	C_1, C_{tG} free	fix C_1	fix C_{tG}
C_1/Λ^2	$-0.0015^{+0.0033}_{-0.0014}$	0	$-0.0015^{+0.0033}_{-0.0014}$
C_{tG}/Λ^2	$-0.120^{+0.248}_{-0.309}$	$-0.117^{+0.247}_{-0.309}$	0

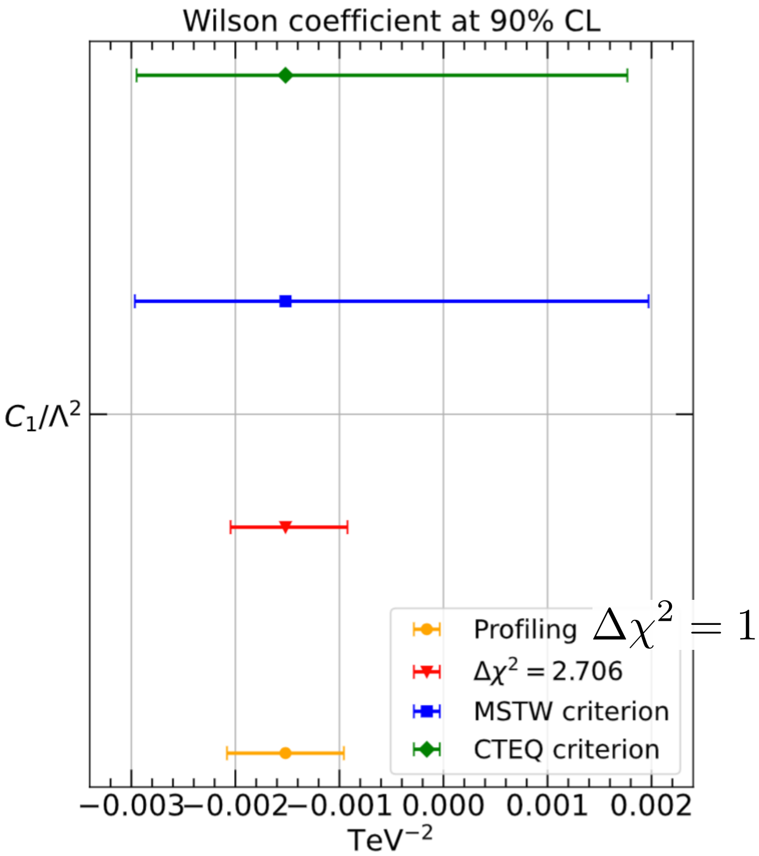
correlations may strengthen with future expts

- increasing Weight (expt. precision) enhances SMEFT coeff. uncertainty dependence on co-fitted PDFs



relevant for **HL-LHC**!

factor ~2.5 improvement in uncorr. uncert



Wilson coefficient at 90% CL

- in addition, extracted SMEFT unc. depends on PDF error (tolerance) conventions

→ both points suggest a **growing need for further investigation**

summary and outlook

- EIC will enter complex landscape of BSM-motivated EW experiments
- extends BSM sensitivity through PDF improvements (& QCD/EW theory)

- jet and $t\bar{t}$ data as CT PDF-EFT demonstration study; examined correlations

→ robust separation between PDFs, co-fitted EFT coefficients

(potentially reassuring for contemporary EFT-only analyses)

→ evidence of correlation between high-x gluon, contact interaction

→ these will increase with growing expt precision; e.g., at HL-LHC

→ need further theory development; more operator combinations, …

- EIC valuable arena to explore PDF-EFT interplay; UQ essential