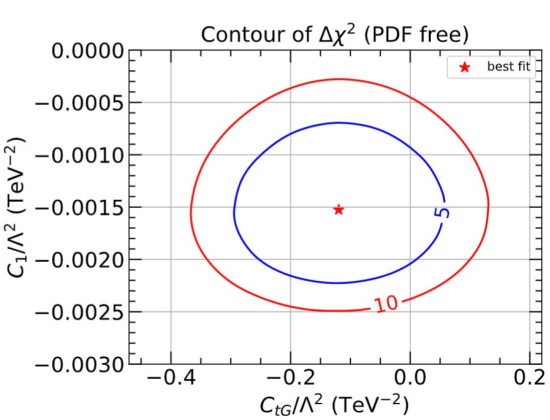
Impact of $t\bar{t}$ and jet data on the joint determination of PDFs and EFT parameters

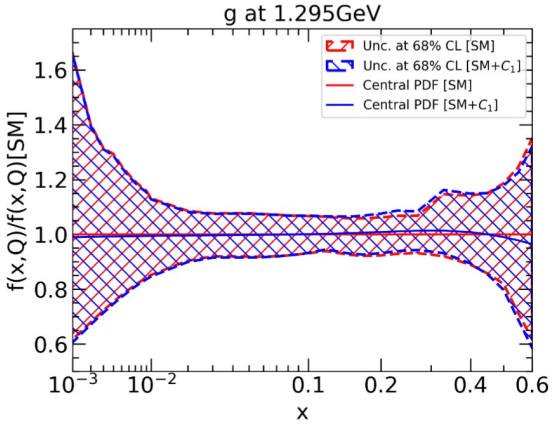
published, May 2023: JHEP05 (2023) 003

Tim Hobbs

Argonne NATIONAL LABORATORY

CTEQ



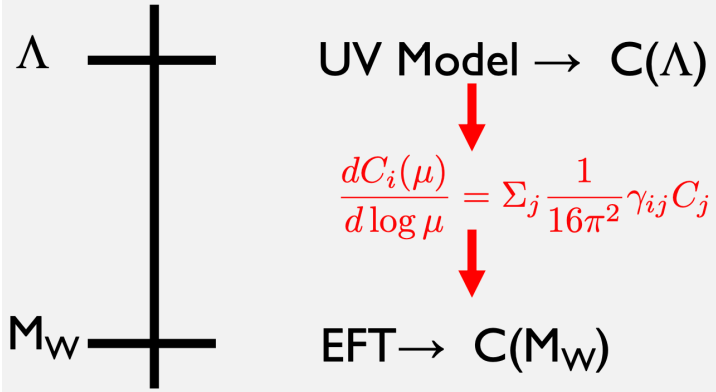


S. Dawson

with J. Gao, M. Gao, D. Liu, and X. Shen

and members of the

CTEQ-TEA (Tung Et. Al.) working group



PDF global analyses have matured to high level of precision

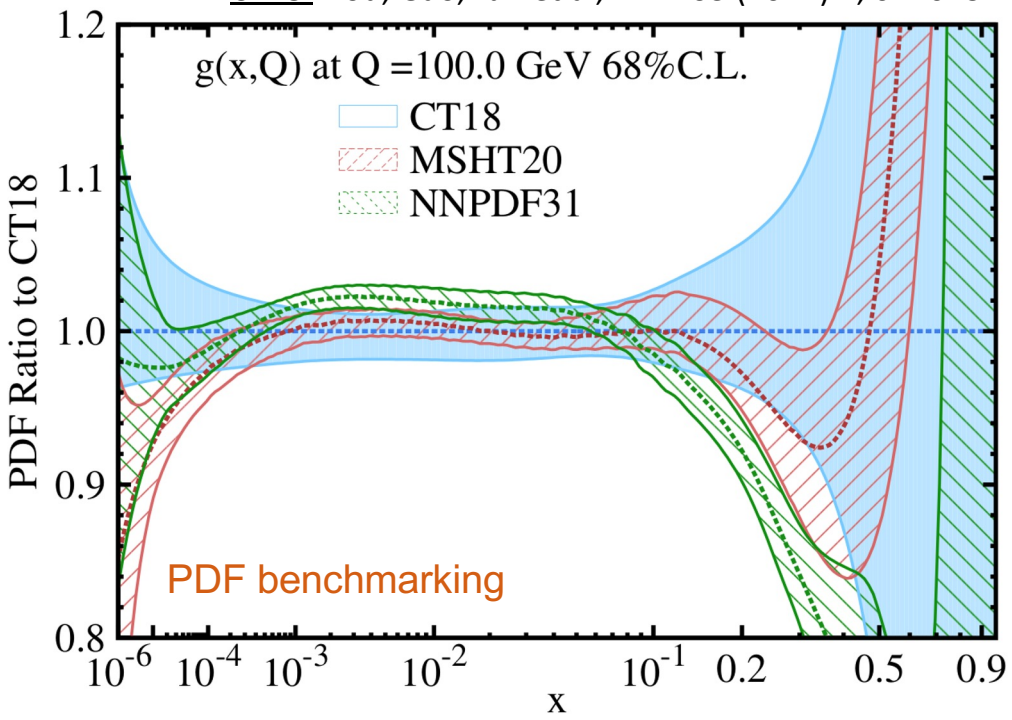
see talk, Keping Xie

BSM searches, upcoming programs → reductions to PDF uncertainties

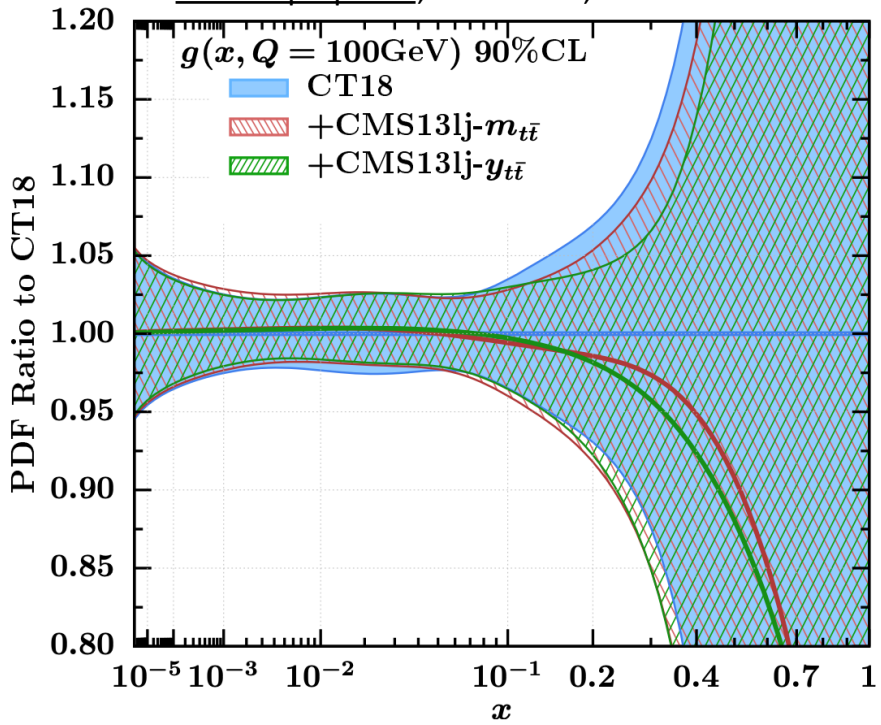
→ analyses are large and challenging: (N)NNLO; NLO EW; **many subtleties**

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

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



CT18 top update, Ablat et al; arXiv: 2307.11153



→ significant interest in **top data**; connections to gluon PDF

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

EFT motivation

- strong recent interest: model-independent BSM analyses

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

  many talks at this meeting → similarities to ideology of PDF-fitting

- 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)

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

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

  → examine possible PDF-SMEFT correlations

large body of recent EFT analyses

- numerous SMEFT fitting frameworks and studies in recent years

  `EFTFitter`[1], `FitMaker`[2], `HEPFit`[3], `SFitter`[4], `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 talks: yesterday, today

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

  Boughezal et al.: 1907.00997, 2306.05564, …

- 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

  → findings in this talk are representative of current PDF-SMEFT situation

basics of (SM)EFT

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

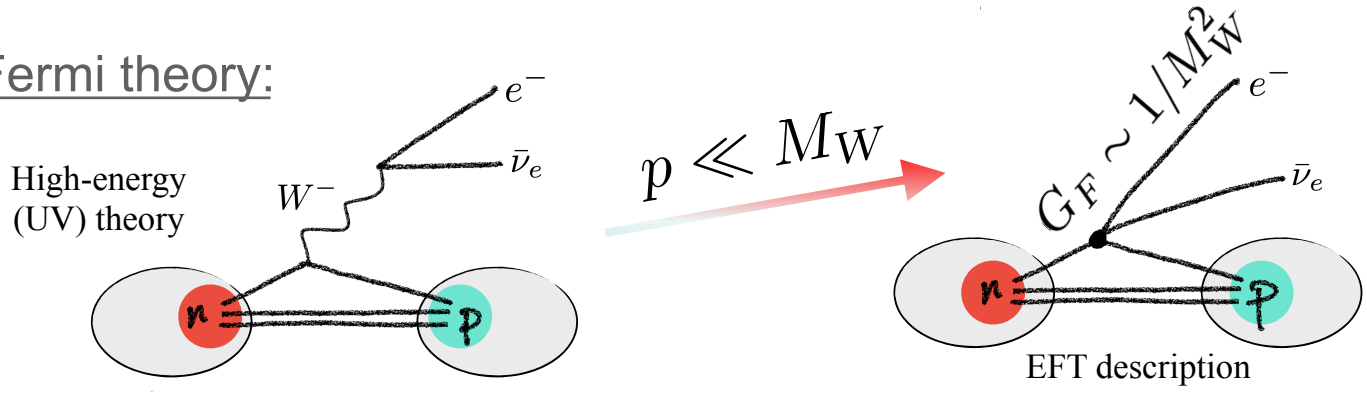
  → explicit non-standard degrees-of-freedom may be integrated away

  → leaves behind basis of higher-dimensional operators

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

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

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

top production

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

Warsaw operator basis

- 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:

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

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

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

- 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

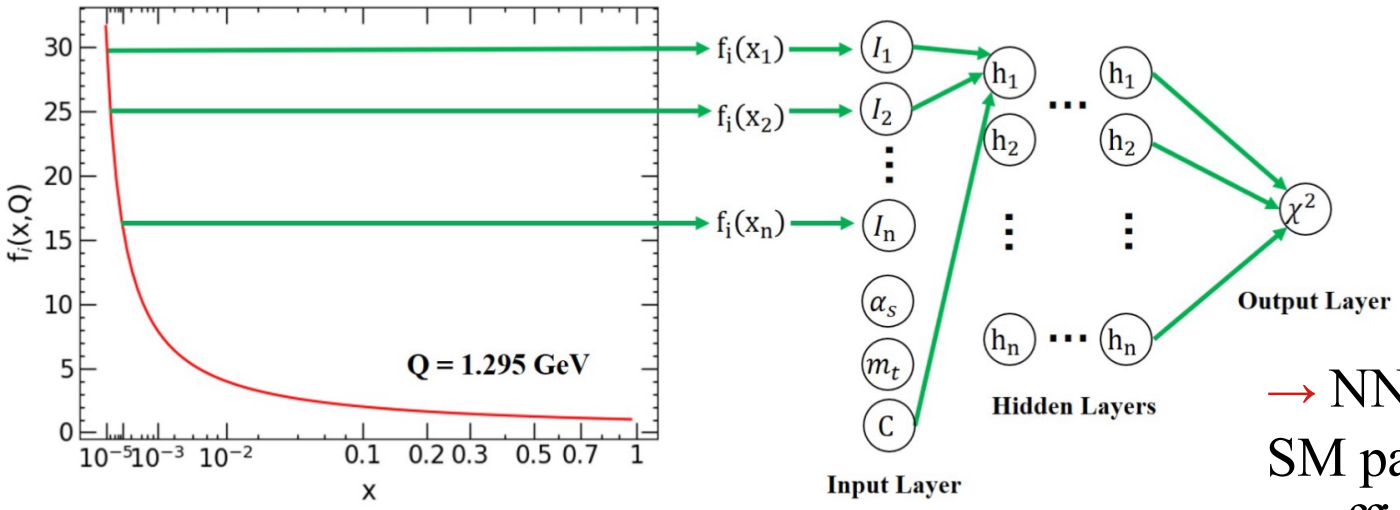
learning log-likelihoods through neural net training

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

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

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

Liu, Sun, and Gao; arXiv: 2201.06586

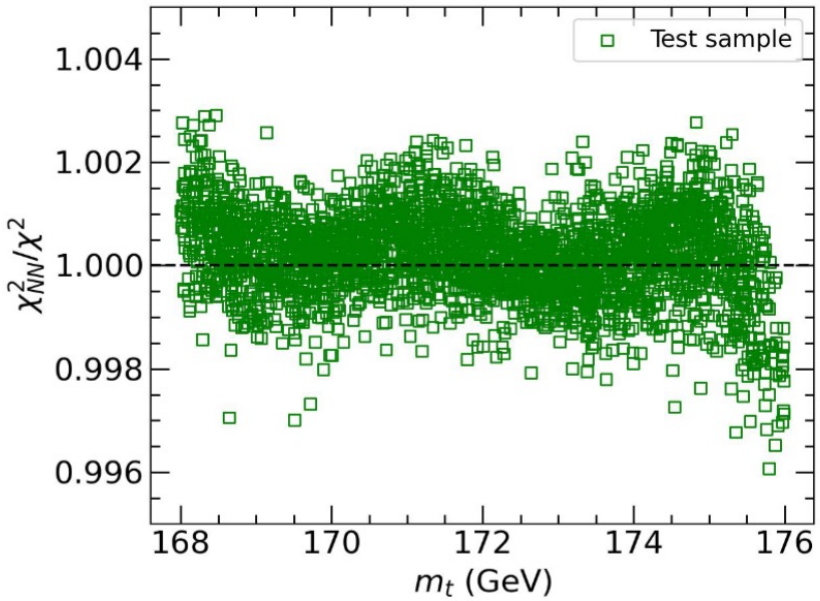


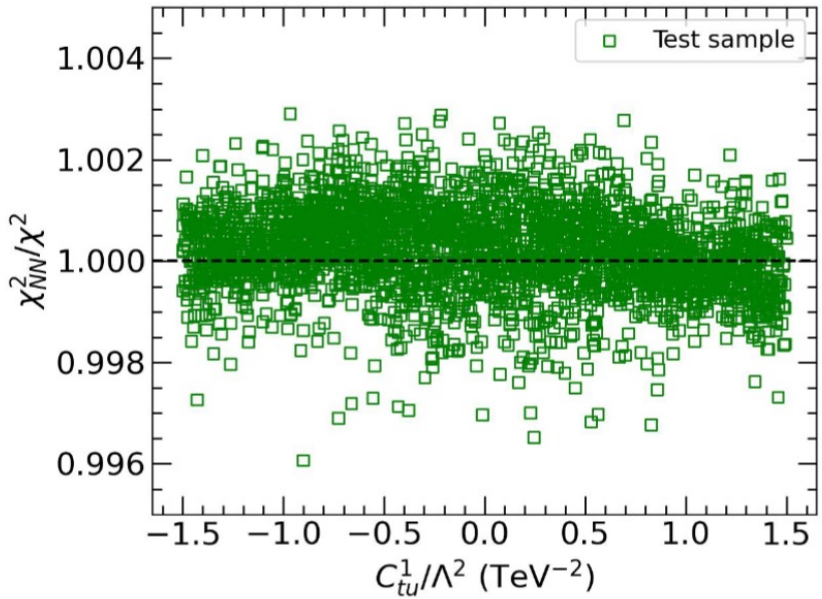
→ NN associates PDFs, SM parameters, SMEFT coefficients with χ^2

NNs effectively learn (PDF-SMEFT) likelihood function

- generate $1.2\text{x}10^4$ replicas over PDFs, SM parameters, SMEFT coeffs.

  → validate performance on $4\text{x}10^3$ test set





→ strong, permille-level agreement achieved!

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

- 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

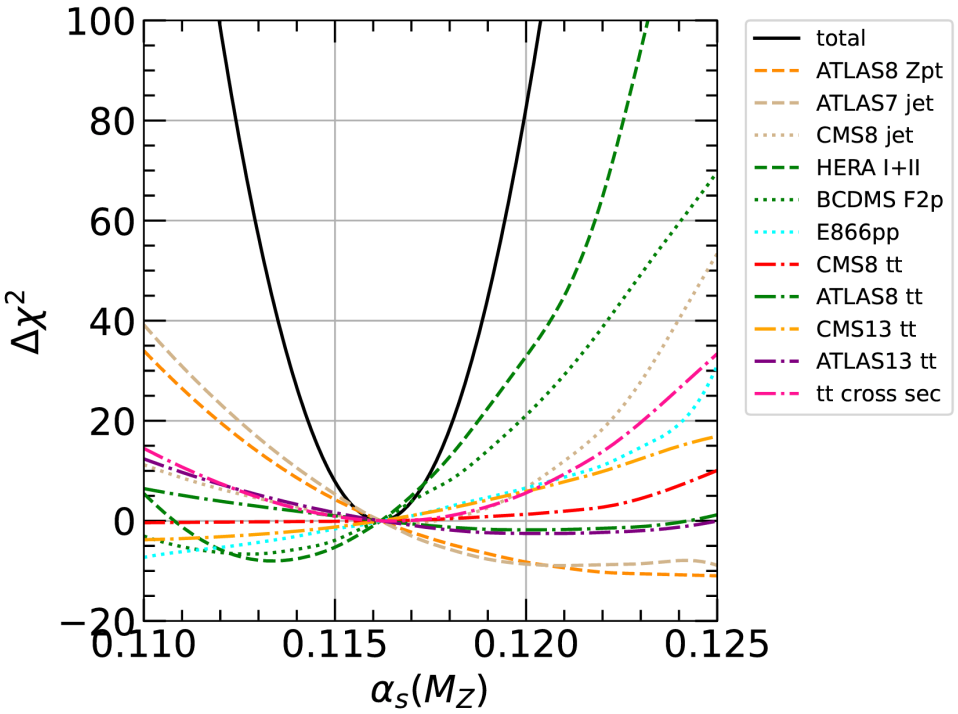
Experiments	$\sqrt{s}$(TeV)	$\mathcal{L}$(fb^{-1})	observable	$N_{\rm pt}$
*† LHC(Tevatron)	7/8/13(1.96)	—	$t\bar{t}$ total cross section	8
*† ATLAS $t\bar{t}$	8	20.3	1D dis. in $p_{T,t}$ or $m_{t\bar{t}}$	15
*† CMS $t\bar{t}$	8	19.7	2D dis. in $p_{T,t}$ and y_t	16
CMS $t\bar{t}$	8	19.7	1D dis. in $m_{t\bar{t}}$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

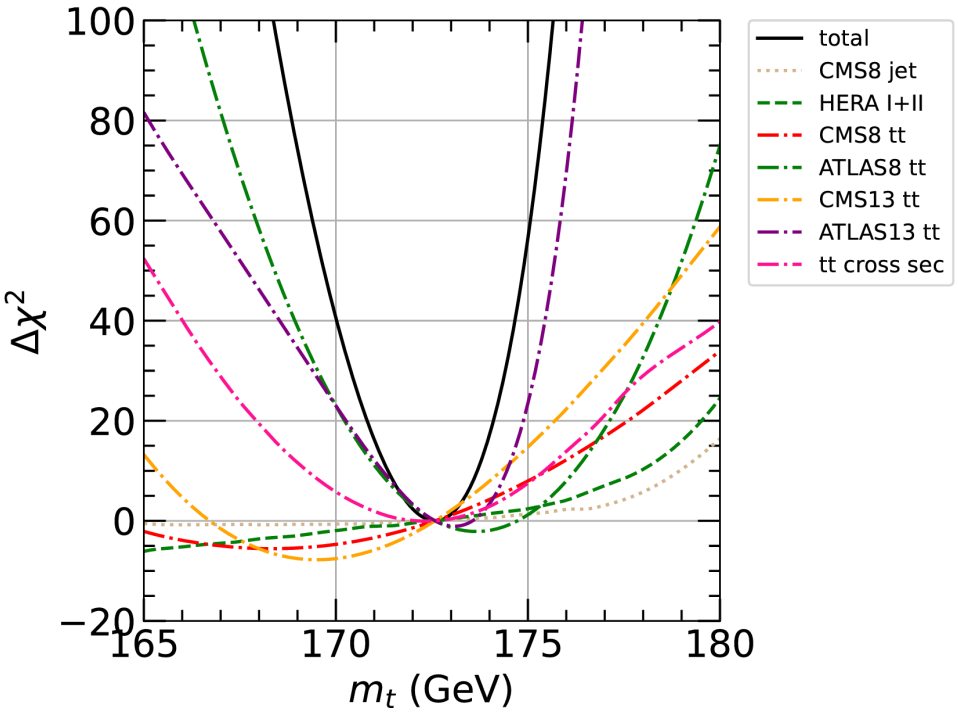
*(in nominal top fits); †(in nominal jet fits)

- 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





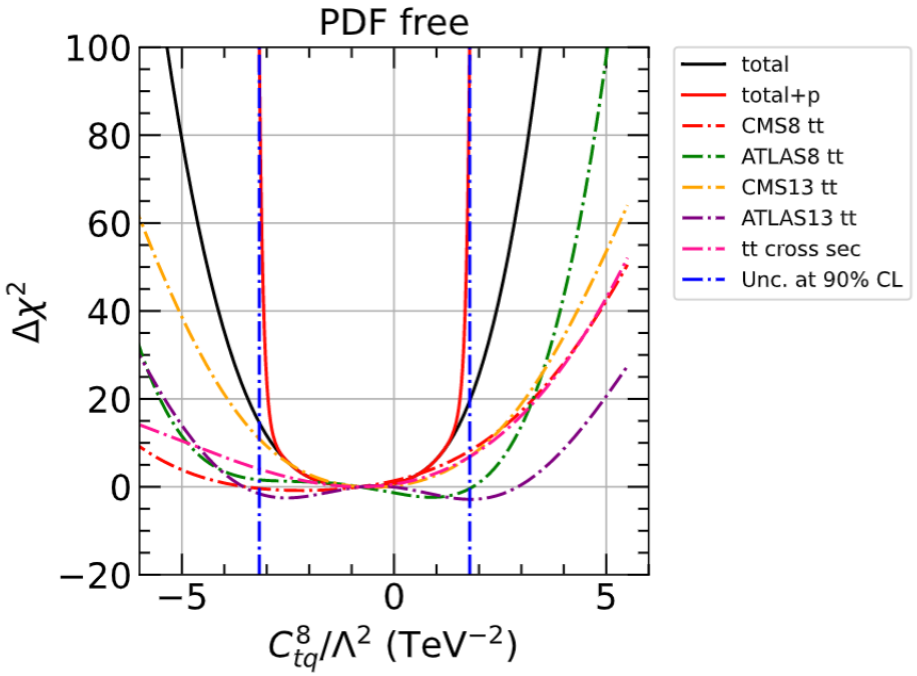
- combined jet and $t\bar{t}$ data significantly constrain strong coupling, top mass

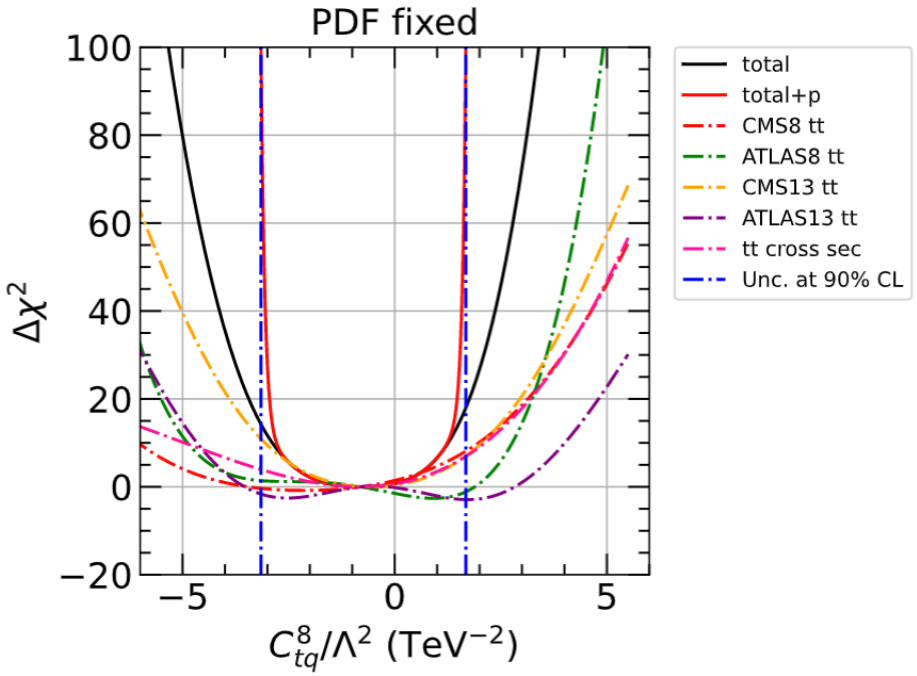
→ competing pulls between CMS, ATLAS $t\bar{t}$ data; lower vs higher top mass, resp.

(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_{tq}^8/Λ^2

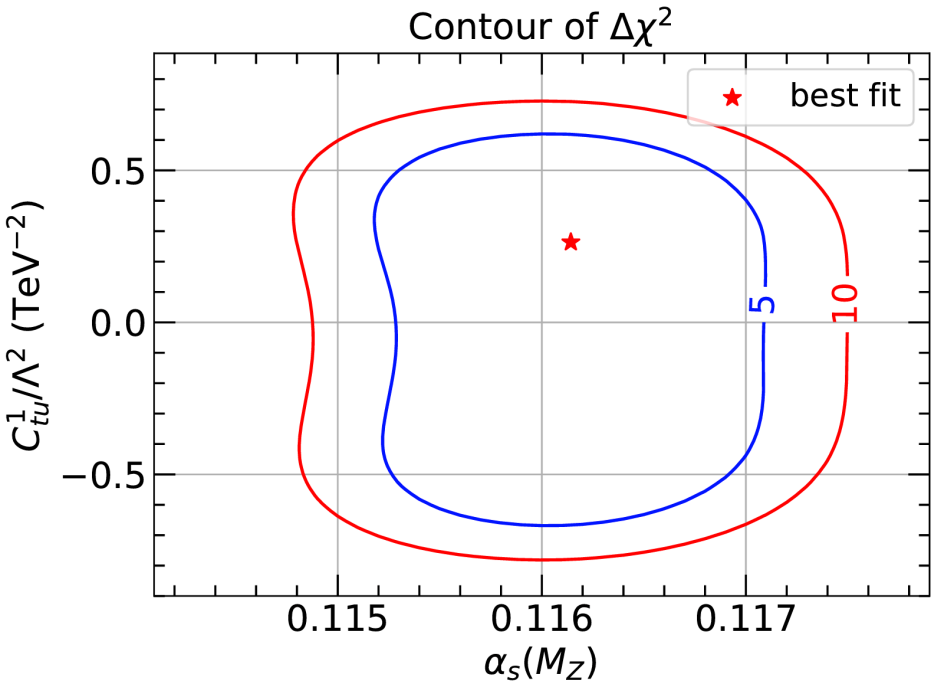
- modest increase in uncertainty when co-fitted with PDFs
- predominantly *quartic* shapes for $\Delta\chi^2$ reflect pure SMEFT contributions $\sim\frac{1}{\Lambda^4}$

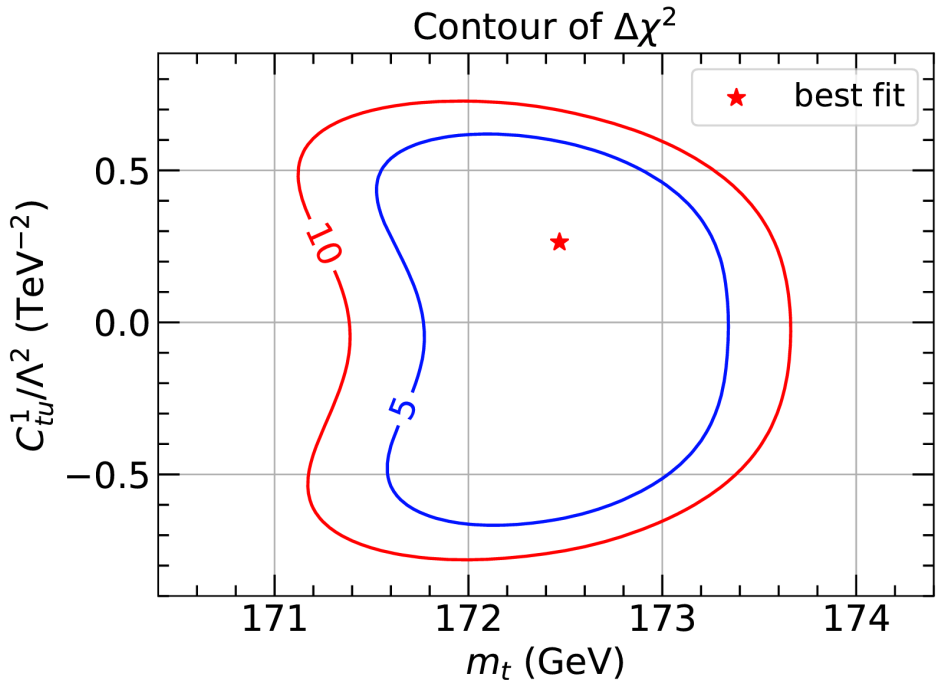
… *i.e.*, importance of quadratic EFT terms in limit-setting

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

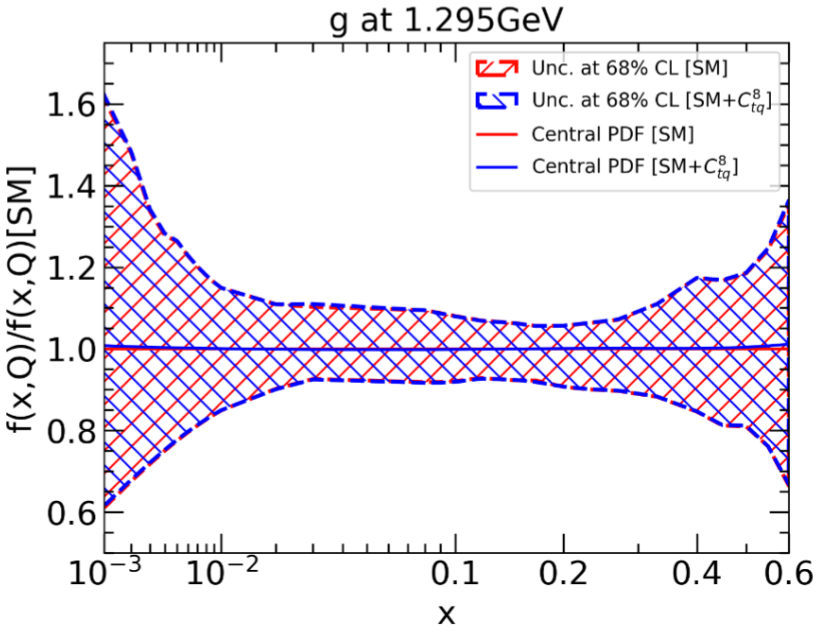
  → still, some deviation from rotational symmetry: non-quadratic profiles for χ^2

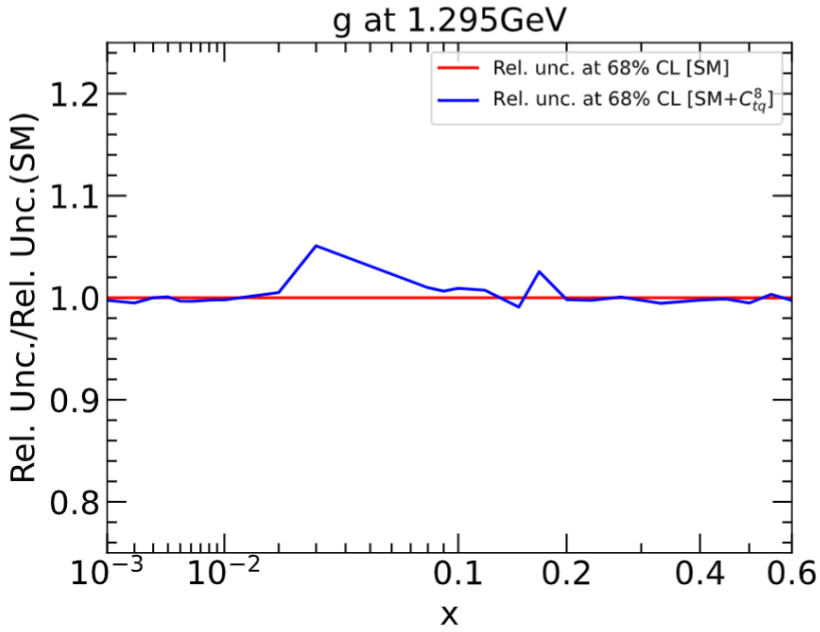
joint fits: very weak correlations with PDFs' *x* dependence

- SMEFT coefficient uncertainties depend on active fitting of PDFs:

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

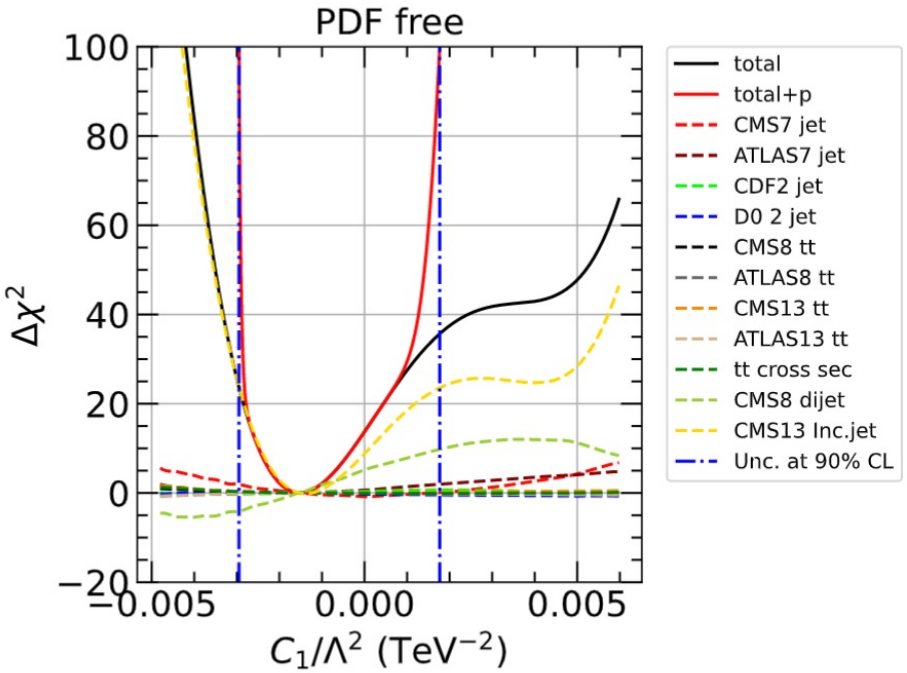
- small variations in gluon PDF, unc. from co-fitting SMEFT:

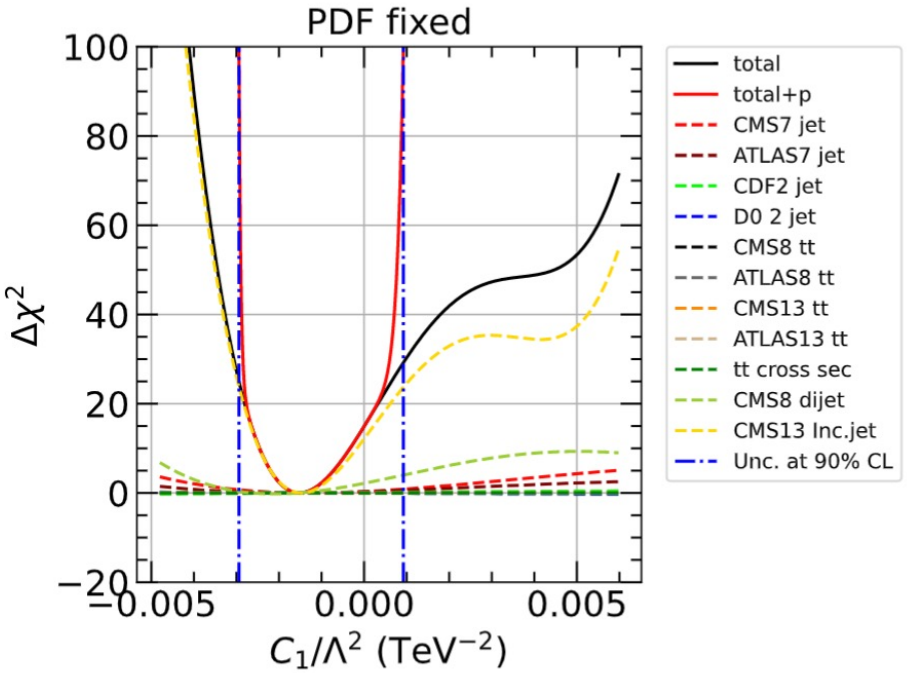




analogous joint fits: jet data and contact interaction

- fitted jet data modestly sensitive to C_1:



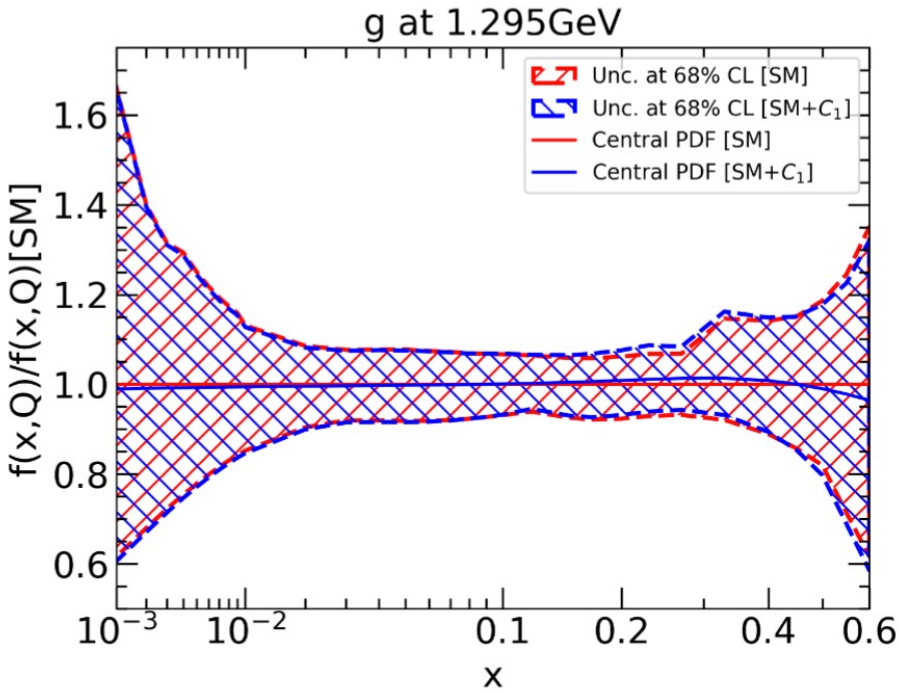


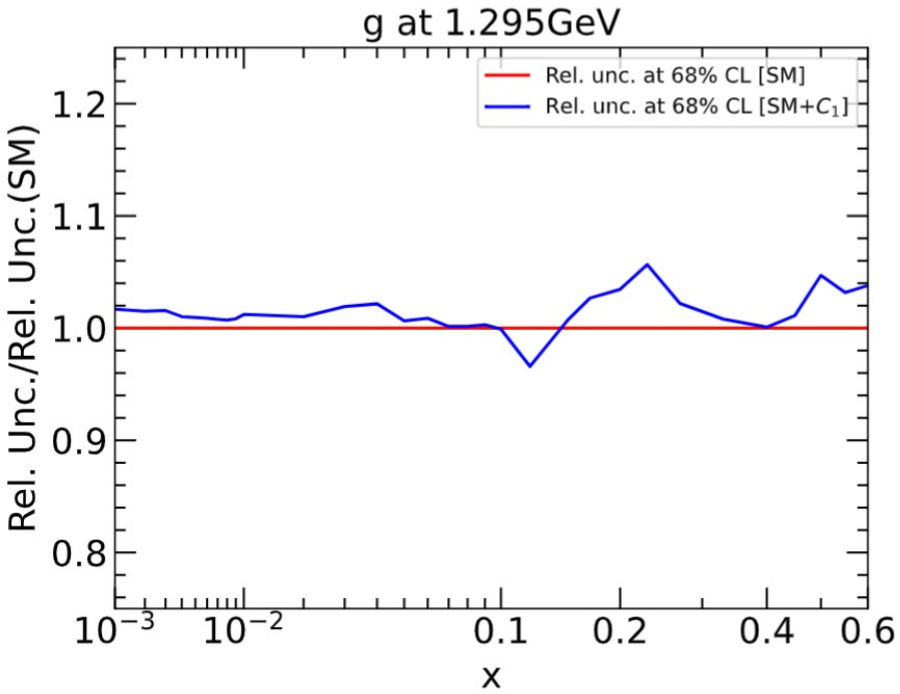
- leading SMEFT sensitivity from CMS: 13 TeV incl. jet, 8 TeV dijet data
- fixing PDFs: (slightly) larger uncertainty underestimate

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

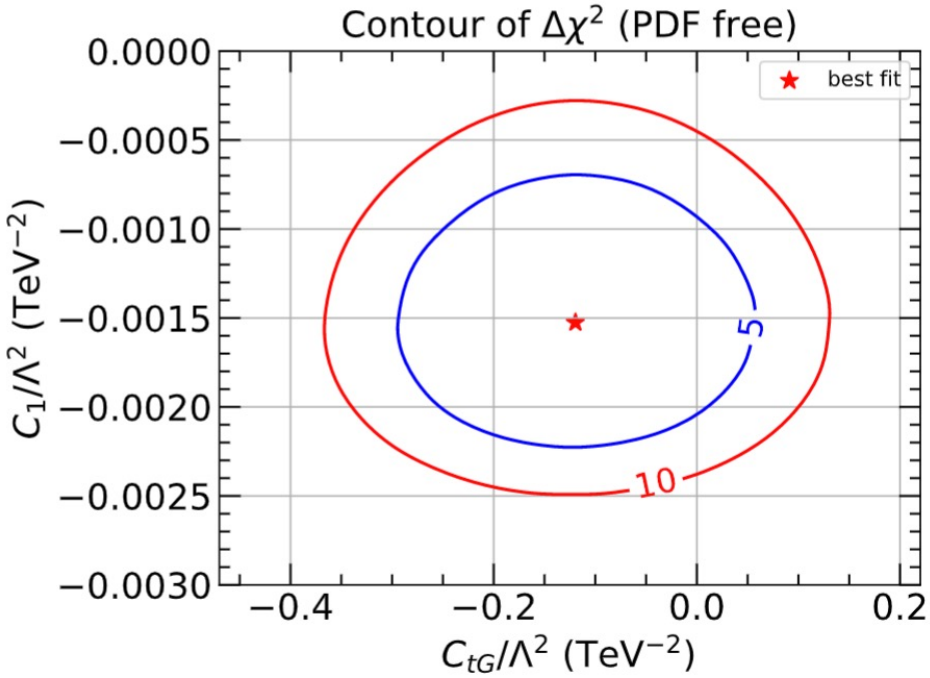


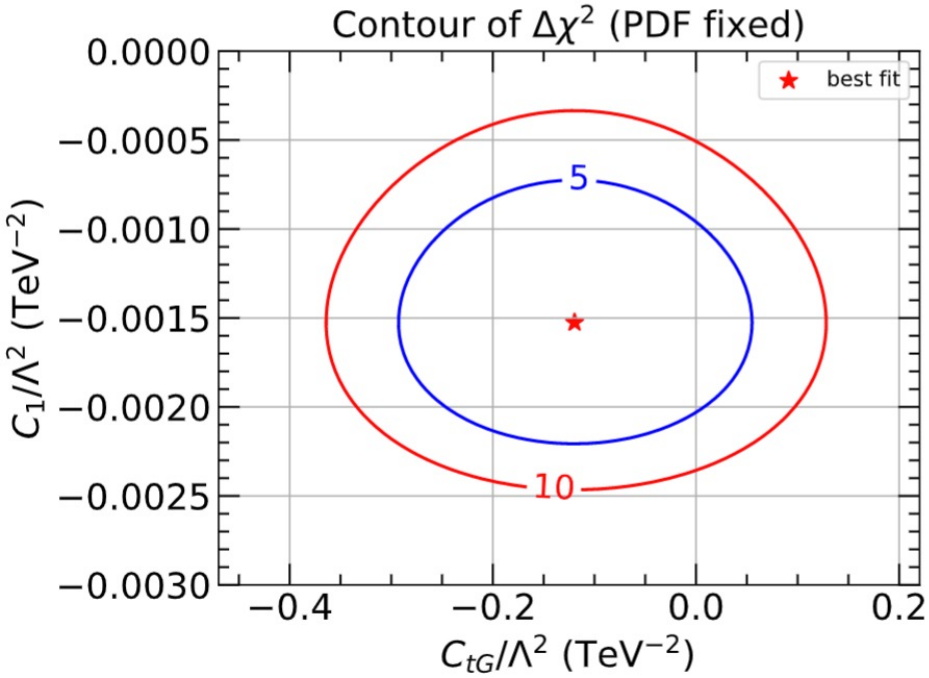


- effect somewhat greater at large $x > 0.1$
- suggests slightly stronger correlation of gluon PDF with C_1

correlations between SMEFT coefficients are mild

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



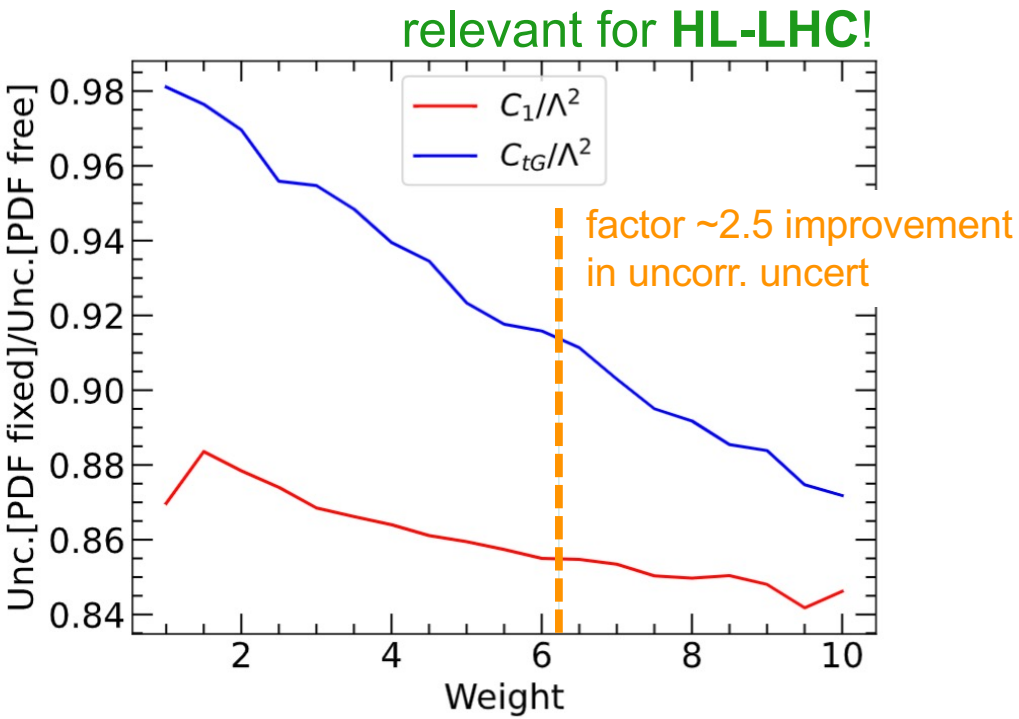


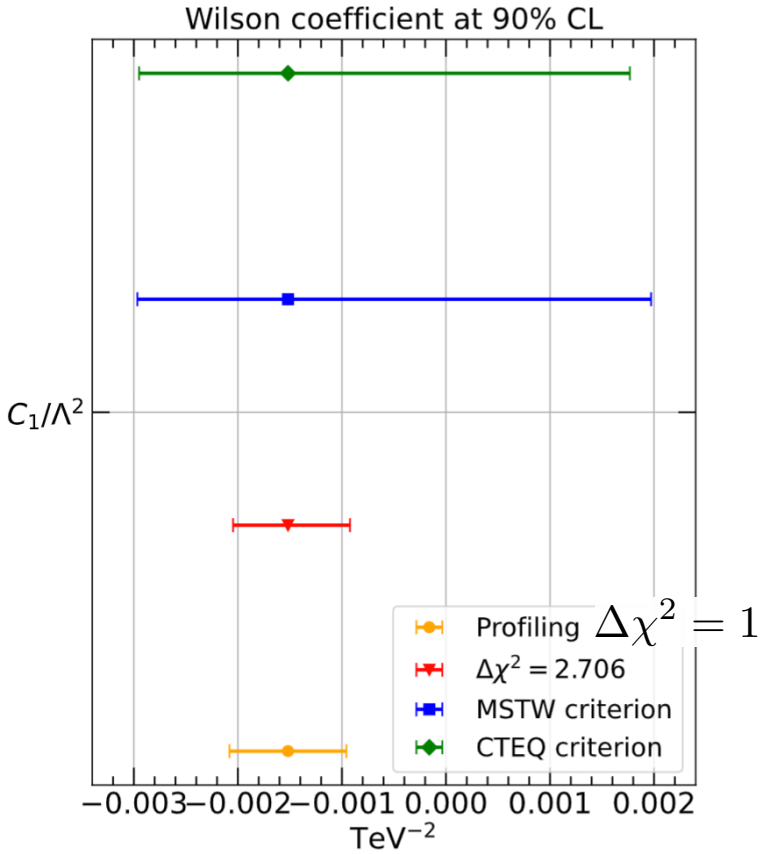
→ strongly rotationally-symmetric $\Delta\chi^2$ contours imply very weak correlations

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





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

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

summary and outlook

- growing interest in EFT global fits, joint analyses with PDFs
- completed first simultaneous PDF-SMEFT fit within CT framework

---

- explore jet and $t\bar{t}$ data as a demonstration study; examine correlations

  → relatively weak PDF-SMEFT correlations

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

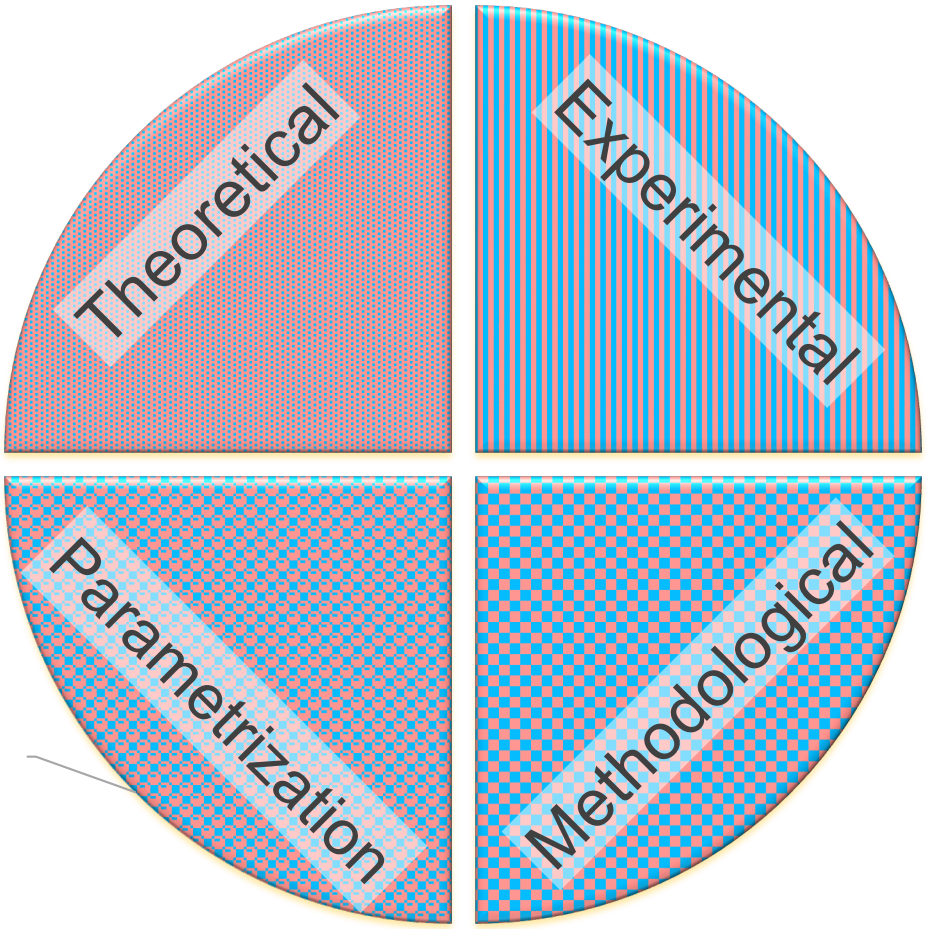
  → these will increase with growing expt precision; e.g., at HL-LHC

  → need further theory development; more operator combinations, …

- ML-based framework; scalable to larger SMEFT parameter space

Backup

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)

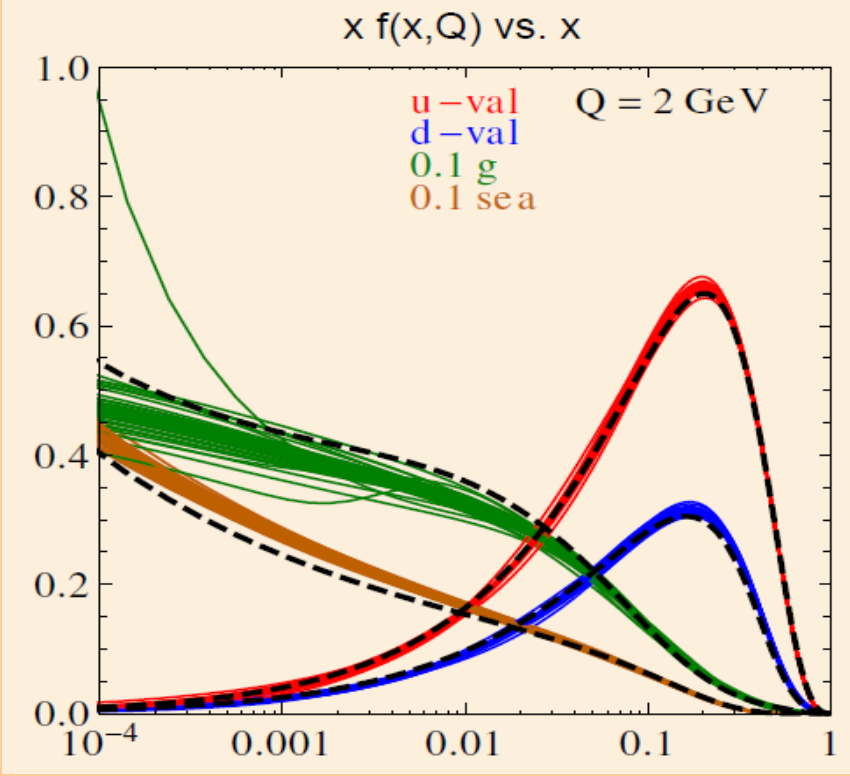
NEW AND IMPORTANT

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

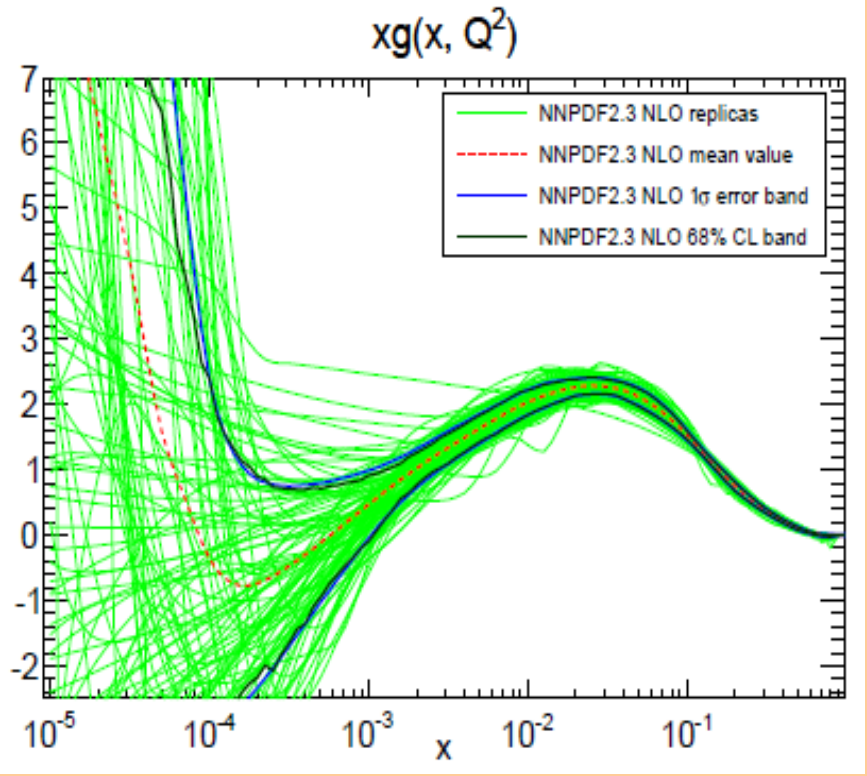
Two types of modern error PDFs

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



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

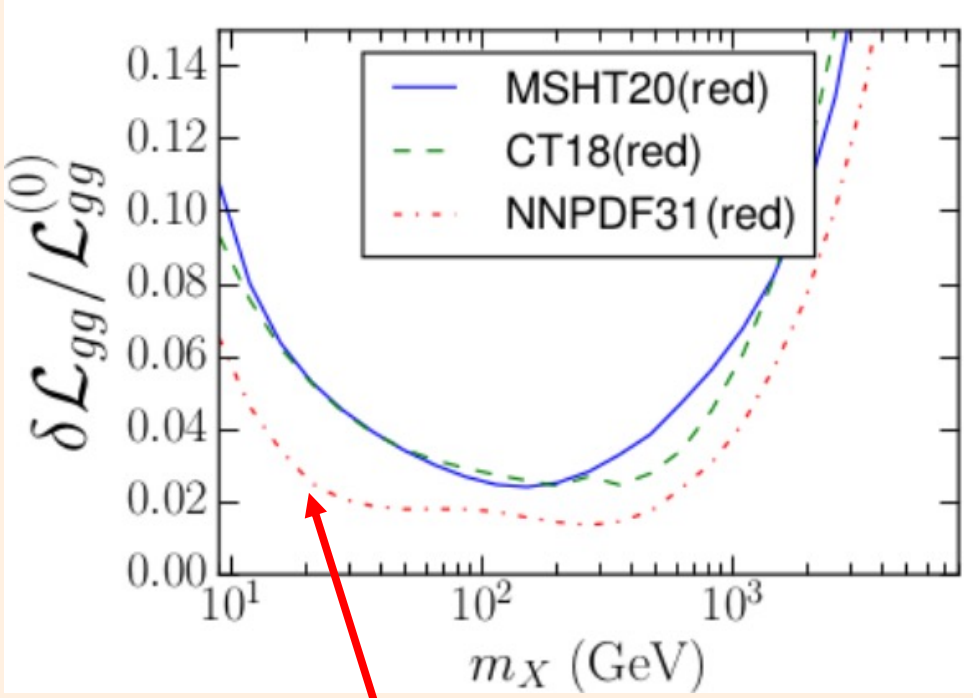
Neural network parameterizations
+ Monte Carlo PDF replicas
(NNPDF)



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

differing PDF errors explained by epistemic uncertainties

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



×1.5 – 2 difference

While the fitted data sets are identical or similar in several such analyses, the differences in uncertainties can be explained by methodological choices adopted by the PDF fitting groups.

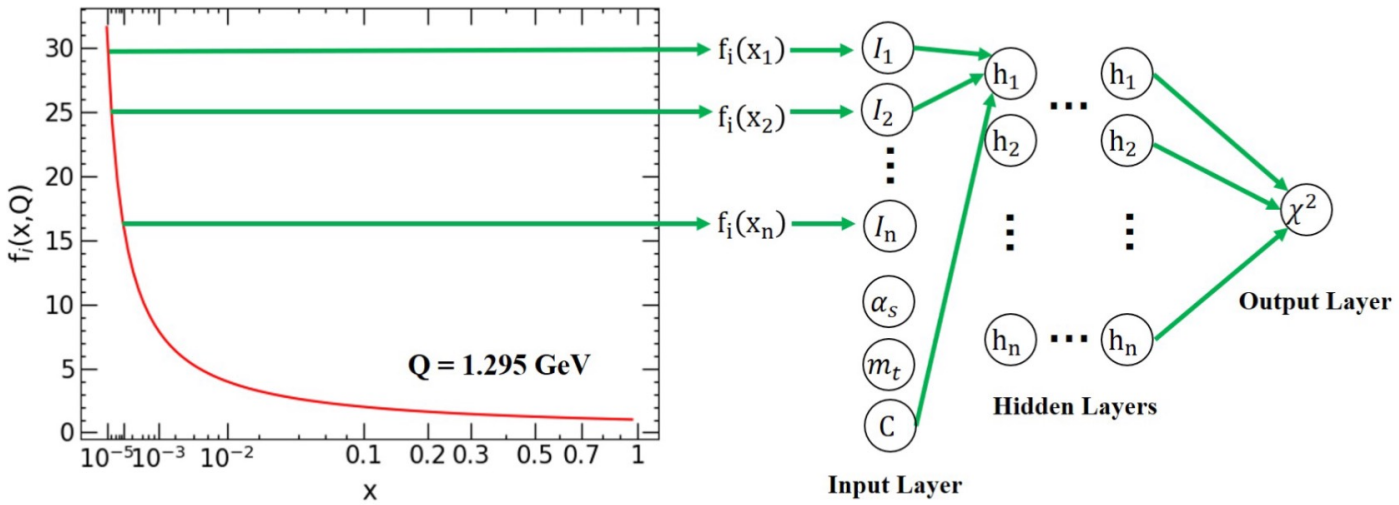
NNPDF3.1' and especially 4.0 (based on the NN's+ MC technique) tend to give smaller nominal uncertainties in data-constrained regions than CT18 or MSHT20

Epistemic uncertainties explain many such differences

Details in **arXiv:2203.05506**, **arXiv:2205.10444**

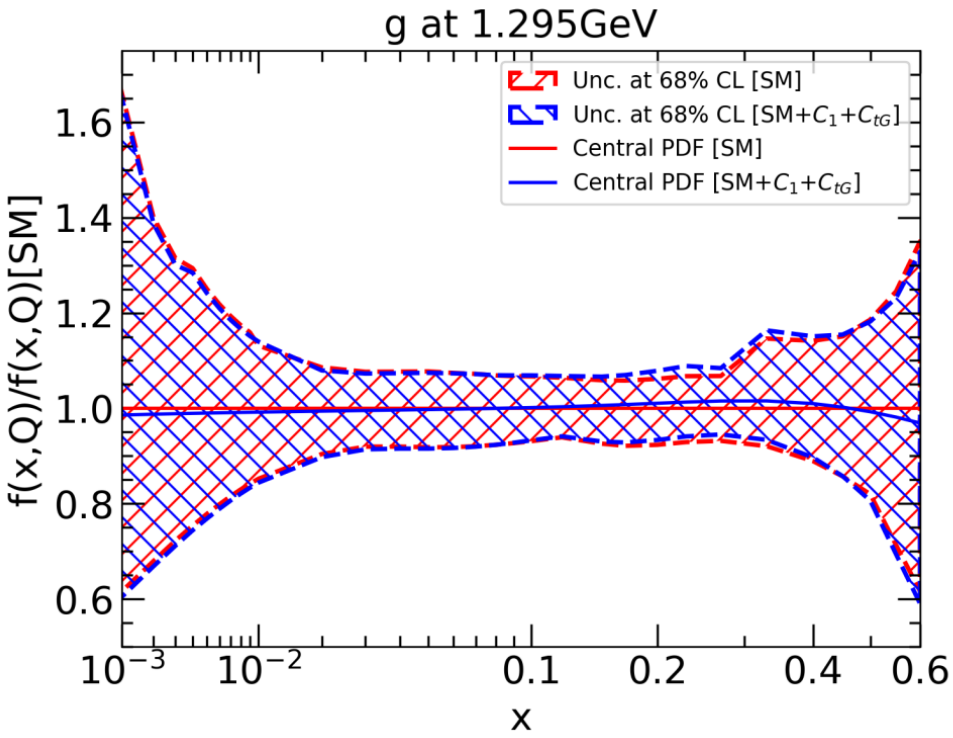
neural network architecture

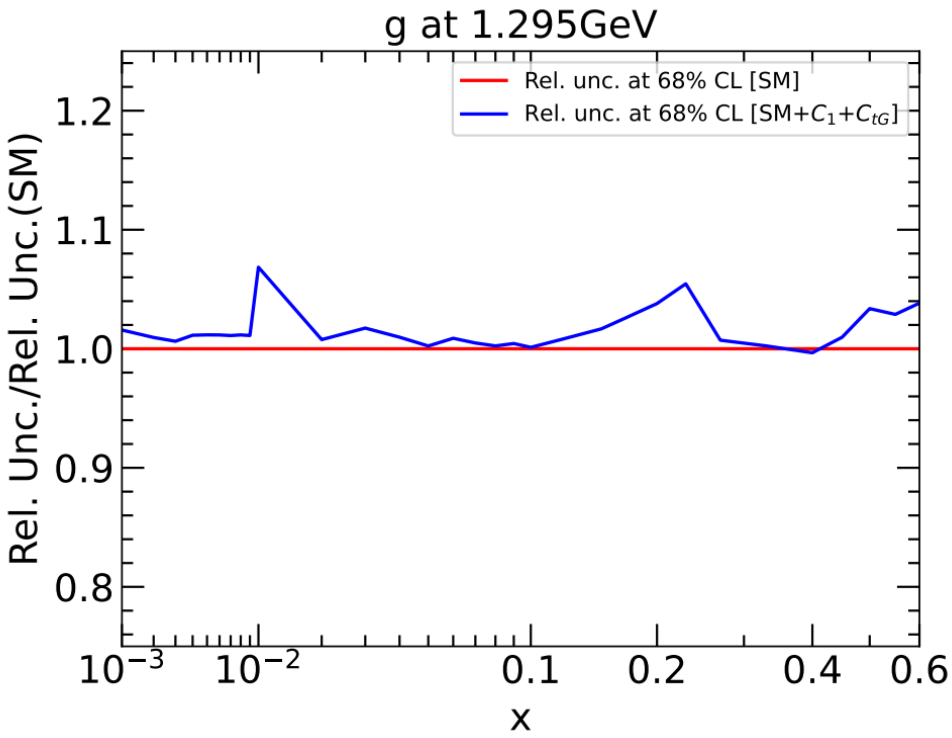
Process (No. of data sets)	Inputs	Architecture	Activation functions for each layer	No. of total params.
$t\bar{t}$ production (6)	{PDFs, α_s, m_t, C^1_{tu} (C^8_{tq}, C_{tG})}	87-60-40-40-1	tanh, (x^2+2), (x^2+2), linear	9401
jet production (7)	{PDFs, α_s, C_1}	86-60-40-40-1	tanh, (x^2+2), (x^2+2), linear	9341
Others (32)	{PDFs, α_s}	85-60-40-1	tanh, (x^2+2), linear	7641



gluon PDF in 2-parameter SMEFT joint fits

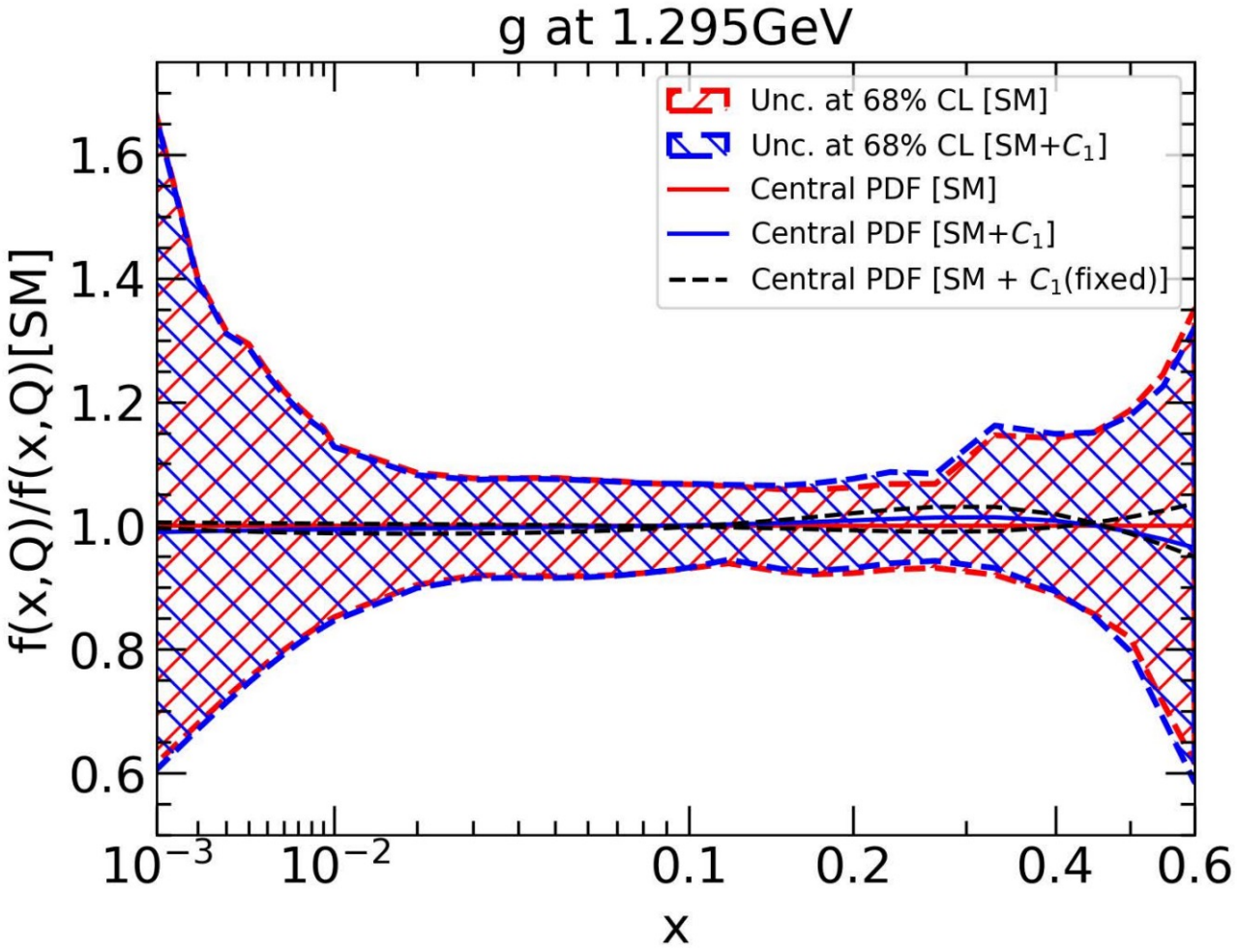
- substantially similar to 1-parameter fit involving C_1





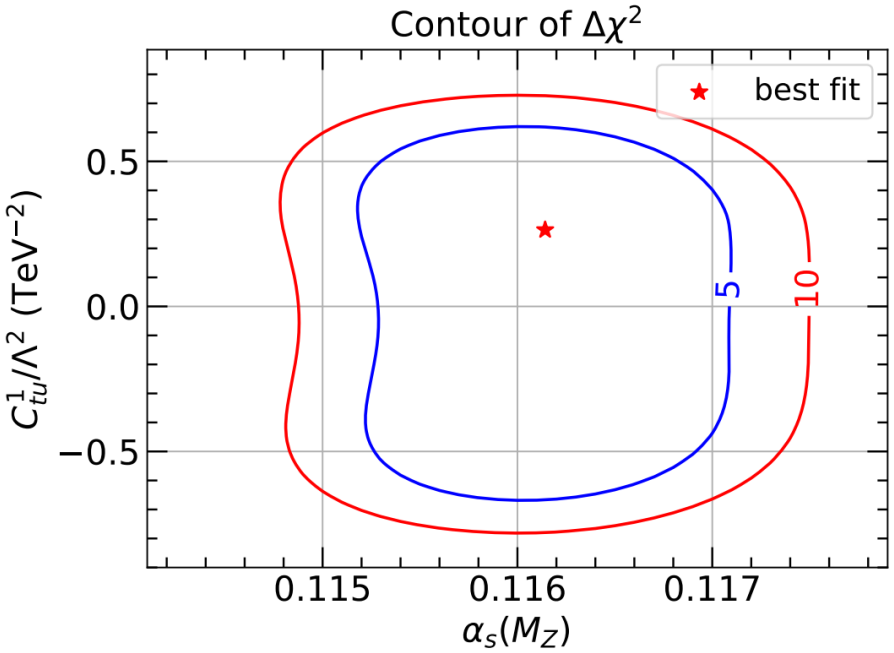
significant nonzero SMEFT can shift PDFs

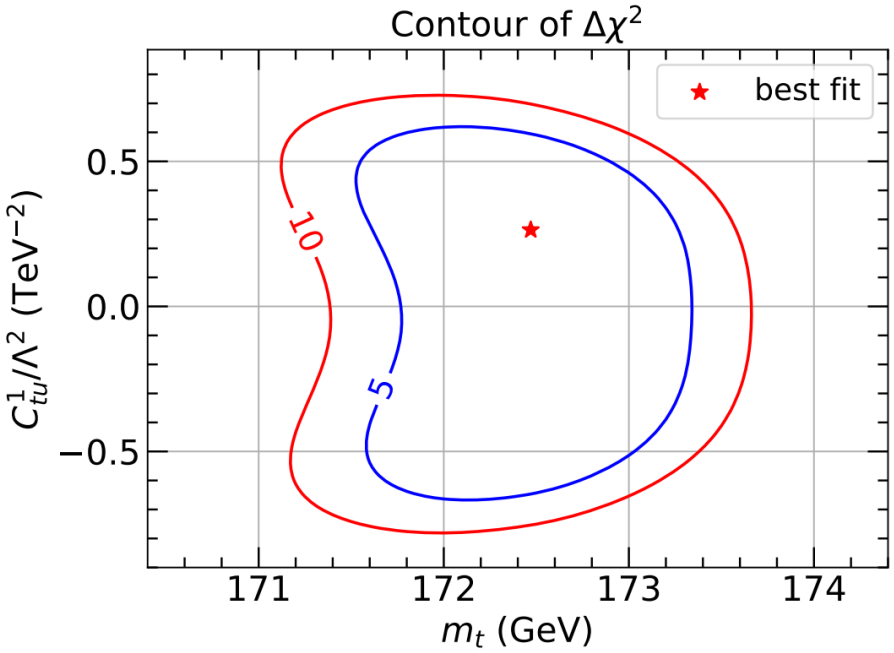
- fitted gluon for extremal values of fixed $C_1 \neq 0$ [black curves]



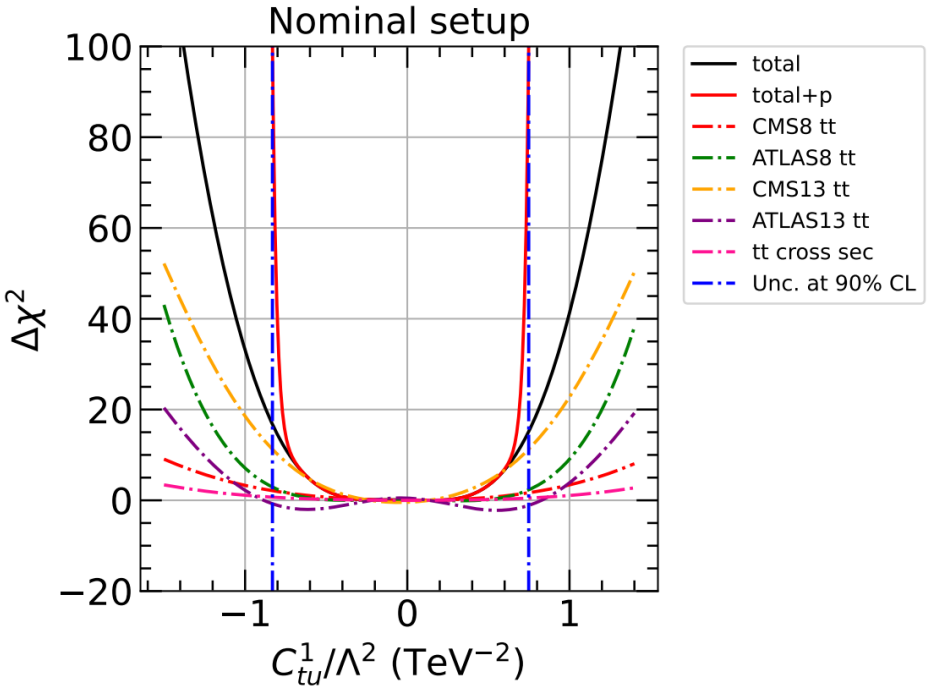
correlations with QCD SM parameters

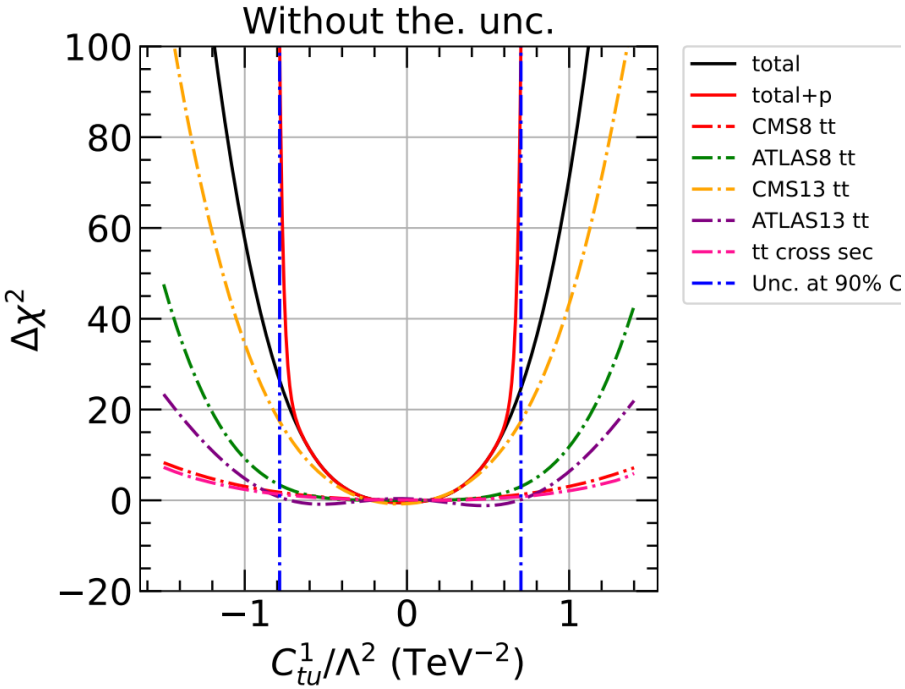
- as with the PDFs, weak correlations between SMEFT and SM parameters

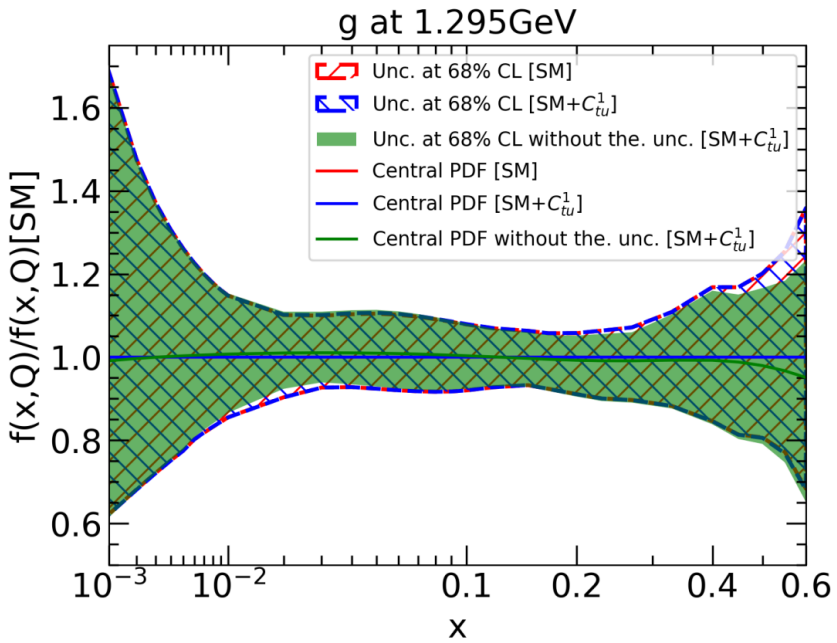


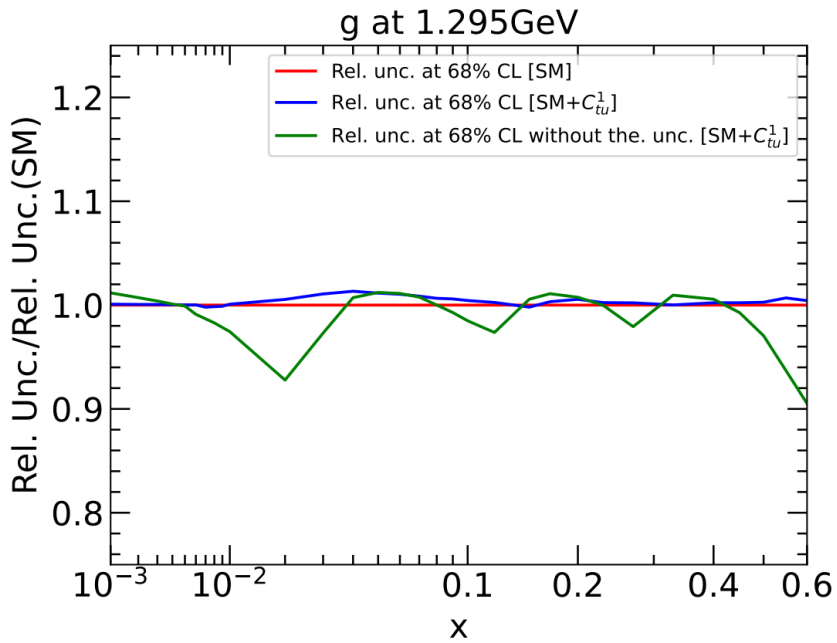


effect of including theory (scale) uncertainties

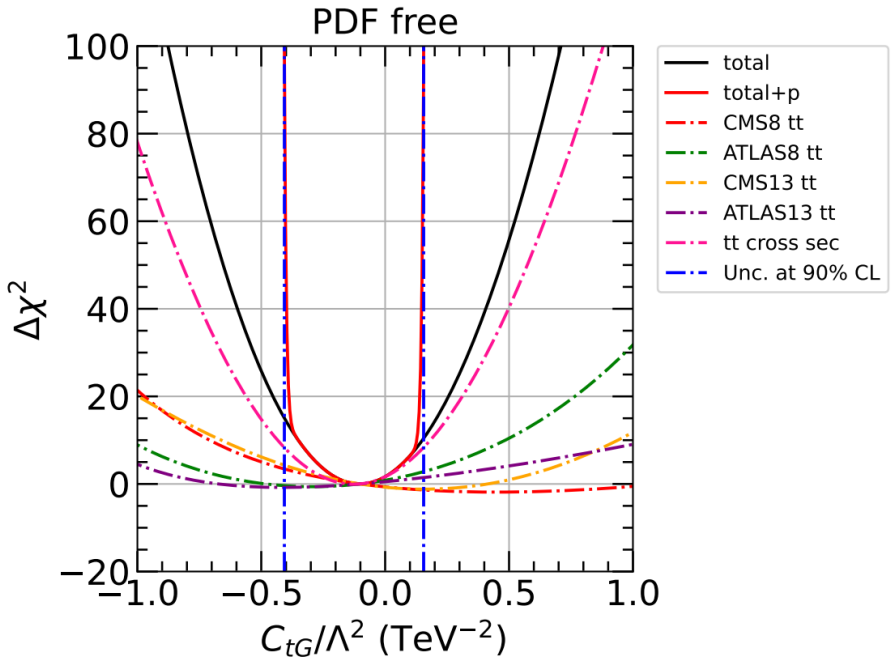


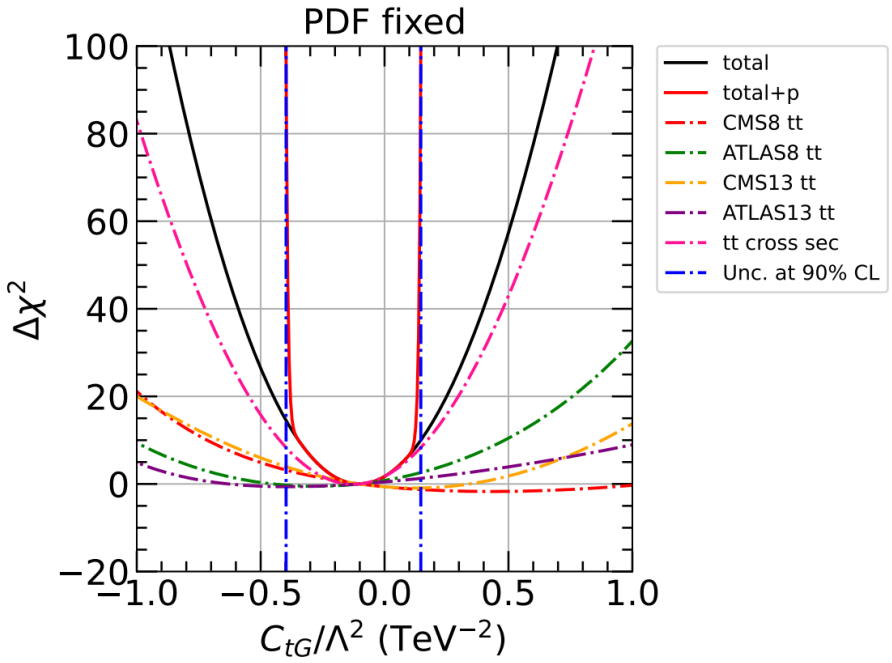


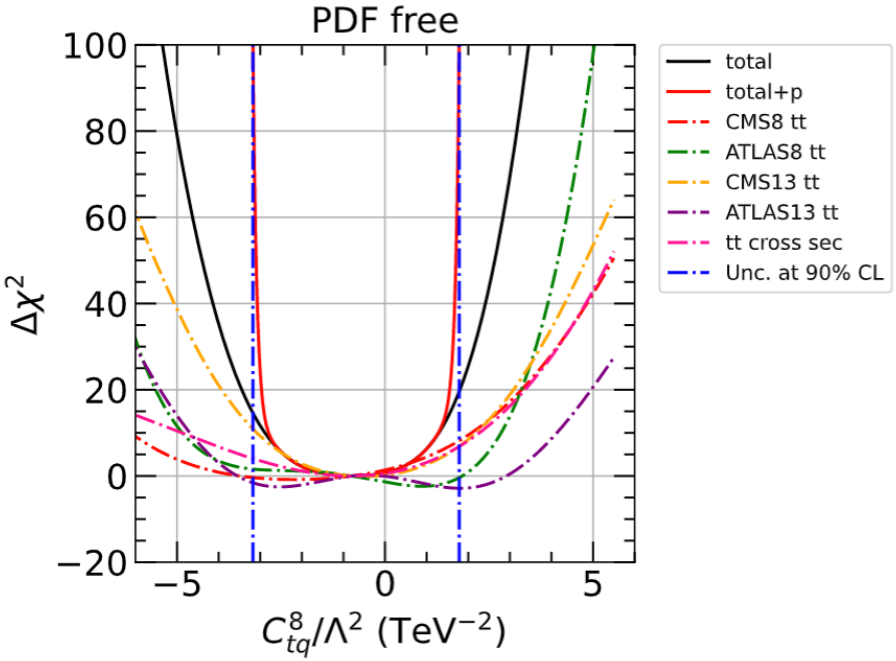


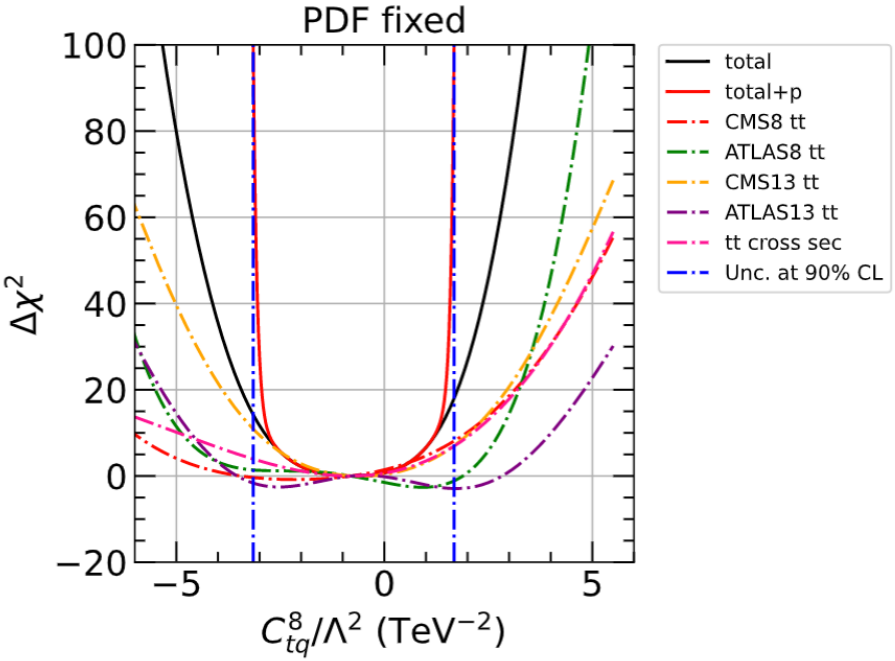


top constraints to gluonic SMEFT operator









inclusion of 8 TeV inclusive or dijet data (CMS)