Parton Distribution Functions at the LHC

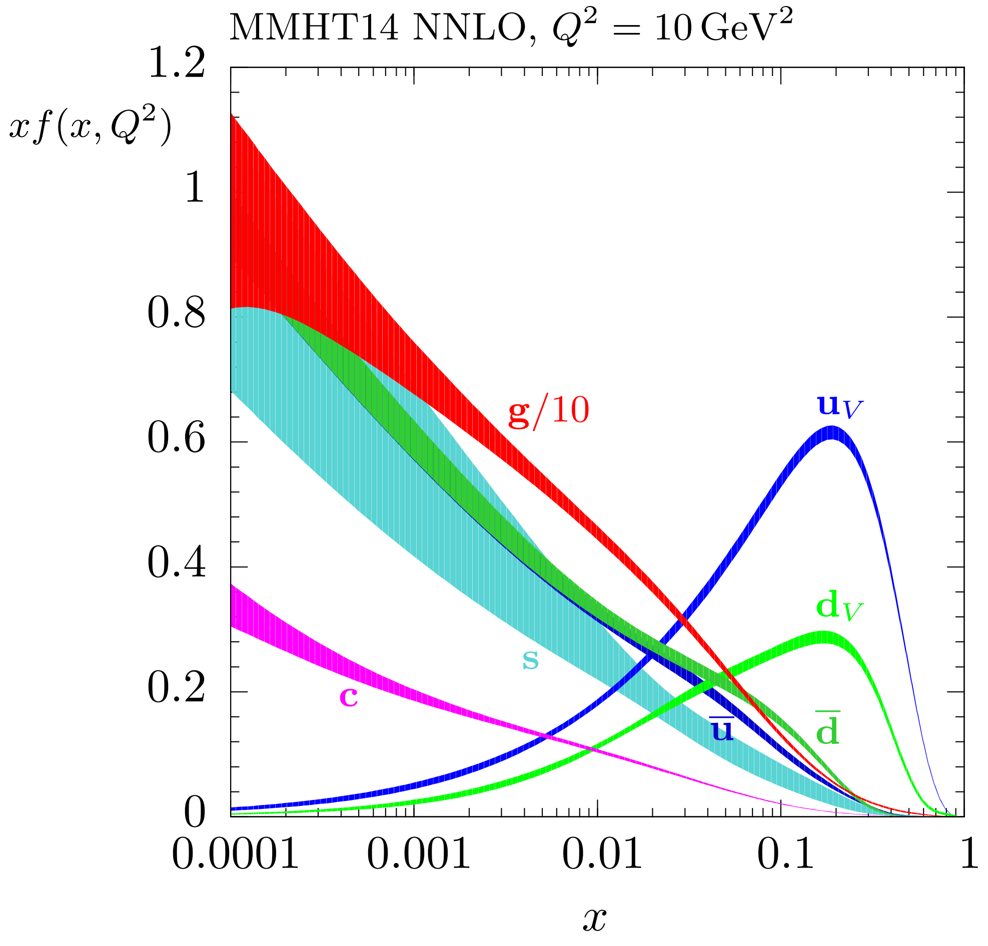
Lucian Harland-Lang, University of Oxford

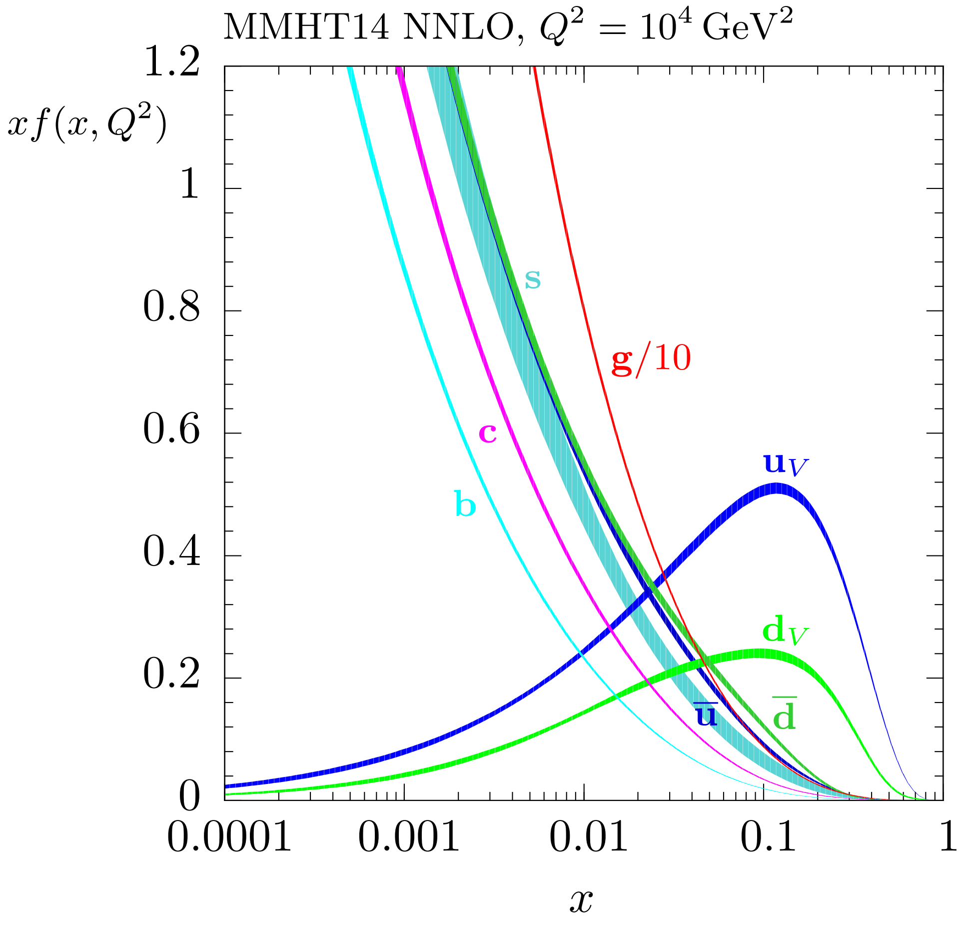
Elementary Particle Physics Seminar, University of Warwick, 9 Jan 2020

UNIVERSITY OF
OXFORD

Outline

- ★ What are PDFs? Why are they important? How do we extract them?
- ★ Opportunities and challenges for PDF determination in high precision LHC era.

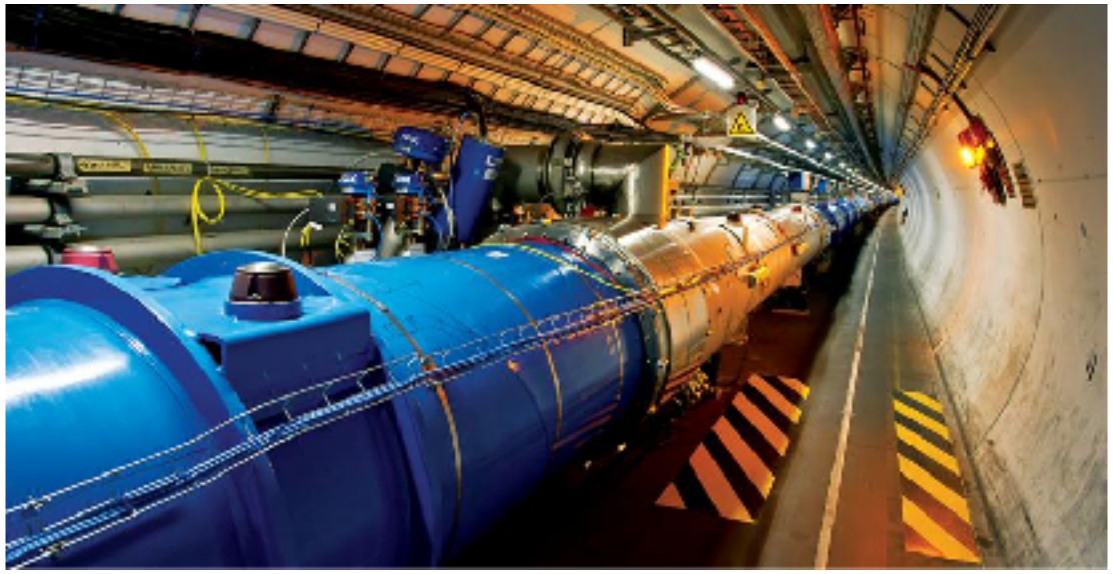

MMHT14 NNLO, $Q^2 = 10\,\mathrm{GeV}^2$

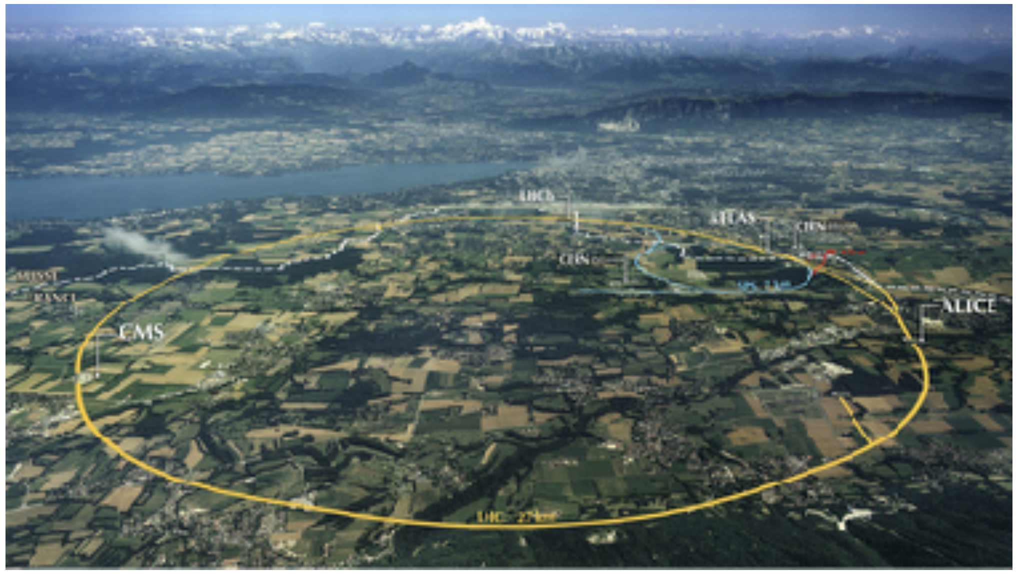

MMHT14 NNLO, $Q^2 = 10^4\,\mathrm{GeV}^2$

An Introduction to PDFs

The LHC: a proton-proton collider

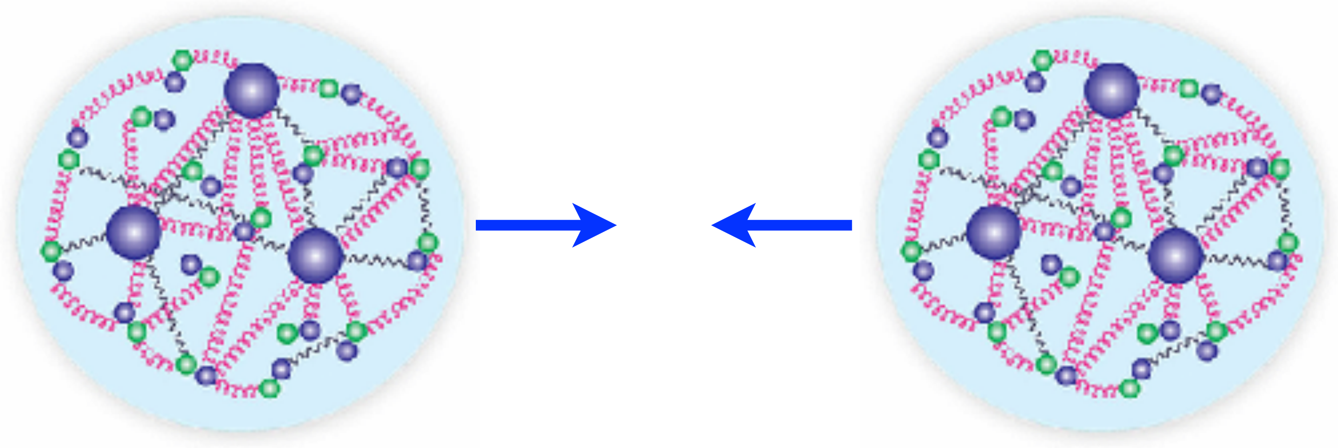
- The **LHC** works by colliding proton beams head on at high energy.
- We examine the debris of these interactions for signs of the Higgs, BSM and to understand the SM better.
- Before doing any of that that: we need to understand what we are colliding: the **proton.**



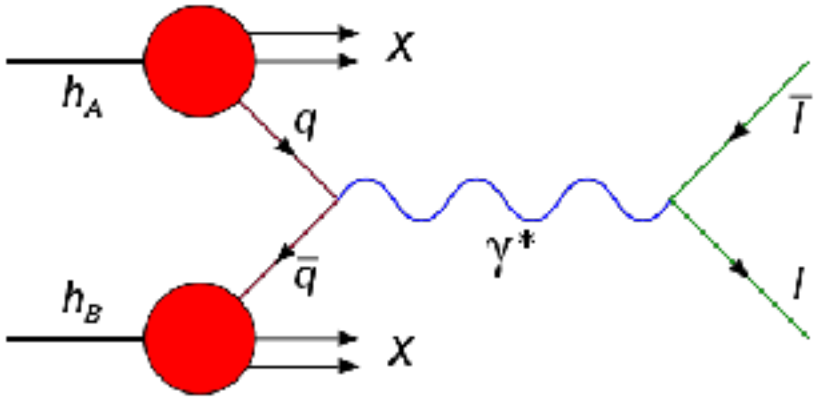


Modeling LHC Collisions

- Proton is composite - LHC collision involves quarks/gluons:



- **Basic idea**: recast proton-proton collisions in terms of more fundamental quark/gluon collision.



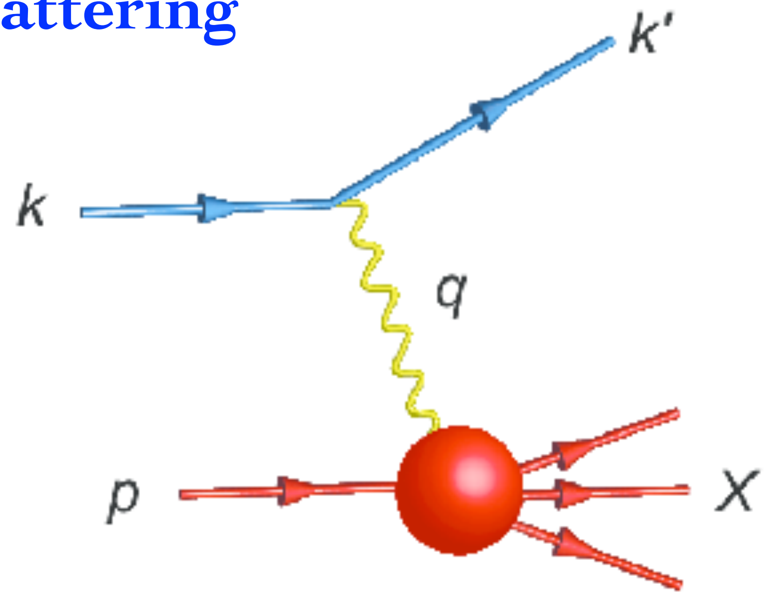
$$\sigma^{DY} \sim \sigma(q\bar{q} \to e^+e^-) \otimes \underbrace{q(x_1) \otimes \bar{q}(x_2)}_{\textbf{PDFs}}$$

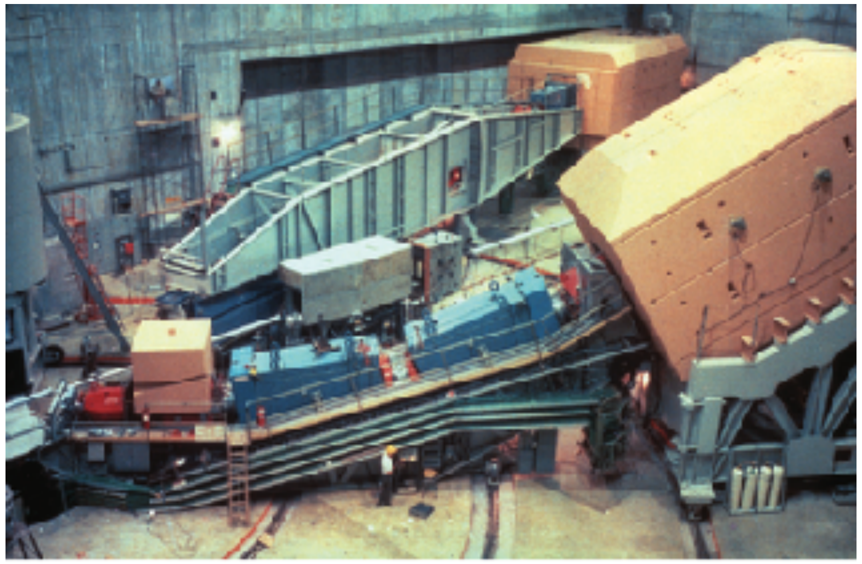
- **Parton distributions functions** (PDFs) encode the binding of the quark/gluons within the proton. In more detail…

Where do PDFs come from?

- Idea goes right back to discovery of quarks at **SLAC** in 1967.
- High* energy beam of electrons collided with proton target. Scattering inelastic, breaking up proton.
- Electron scattering rate vs. angle/energy ($\leftrightarrow x_B, Q^2$) measured, with surprising result…

'Deep Inelastic Scattering'





$$Q^2 = -(k-k')^2 = 4EE' \sin^2\left(\frac{\theta}{2}\right)$$

$$x_B = \frac{Q^2}{2p\cdot q} = \frac{Q^2}{2m_p(E-E')}$$

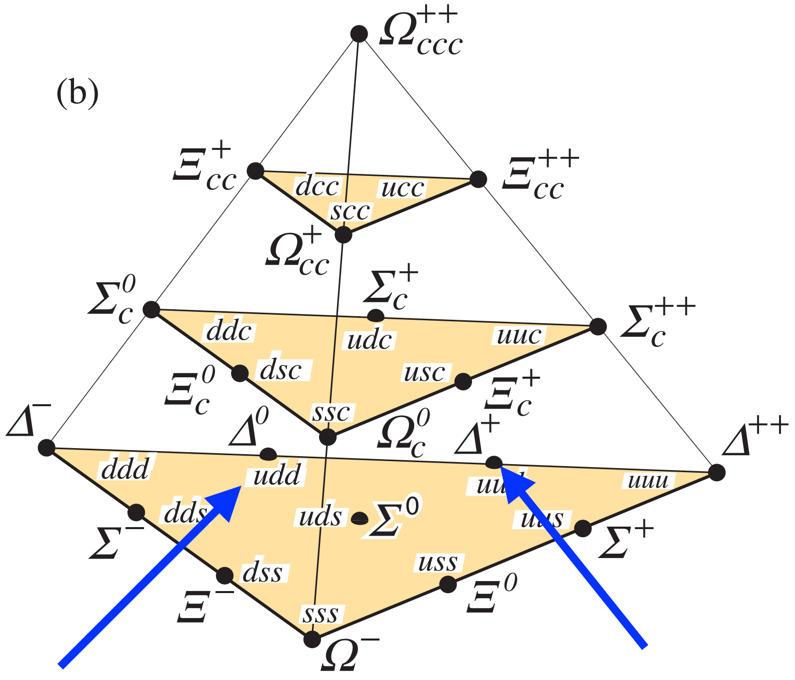
*By standards of the day!

SLAC measurement

- Cross section found to follow expectation for scattering off **point-like** objects within proton.

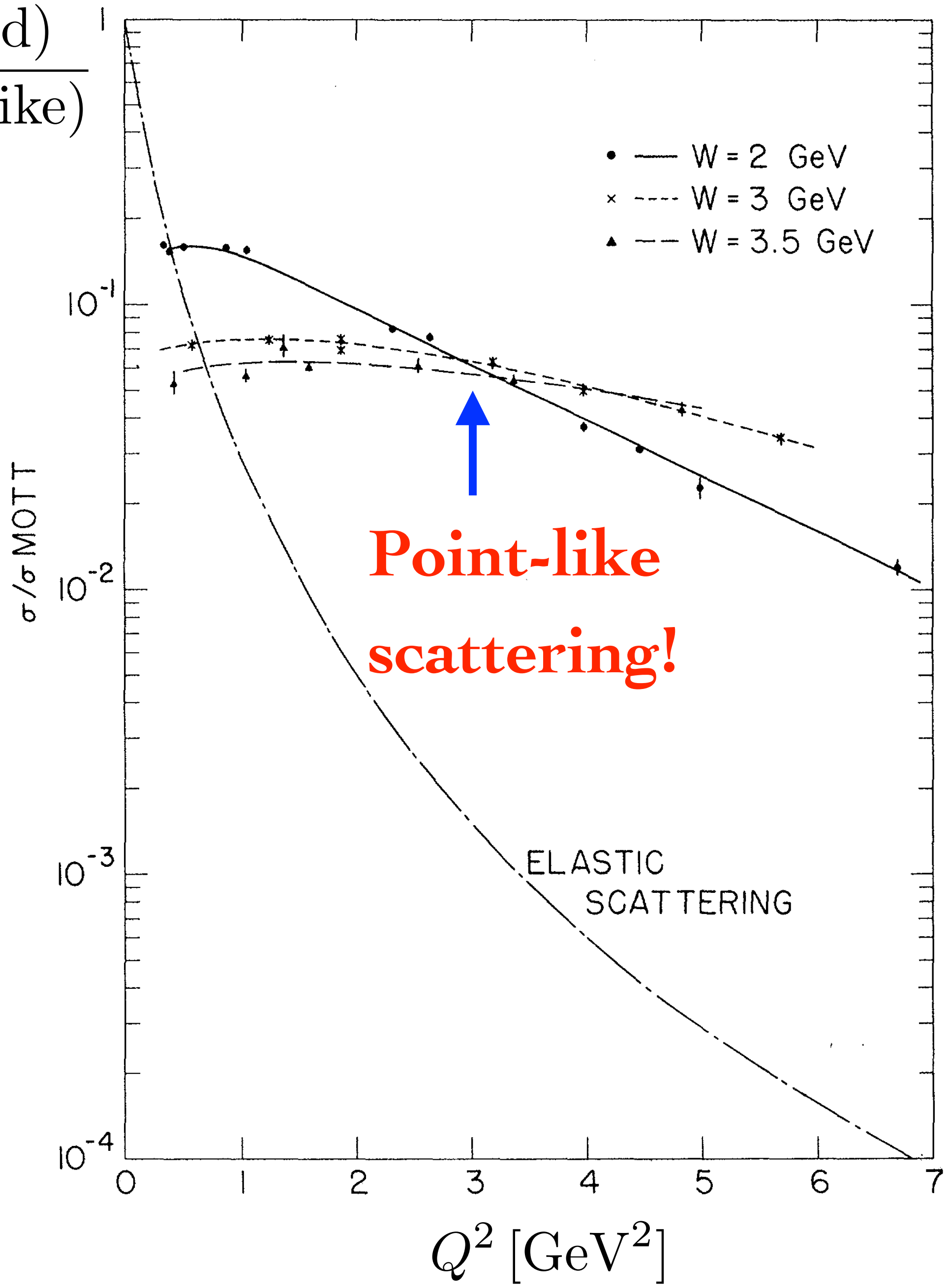
- These objects, first given agnostic label of '**partons**' by Feynman.

- Soon identified with Gell-Mann/Zweig's **quarks** introduced to explain the hadron zoo.



Neutron **Proton**

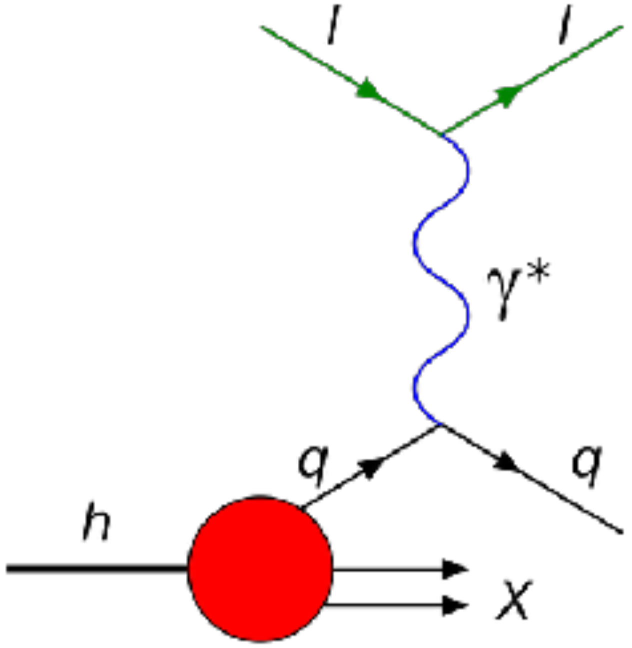
$$\frac{\sigma(\text{observed})}{\sigma(\text{point}-\text{like})}$$



Point-like scattering!

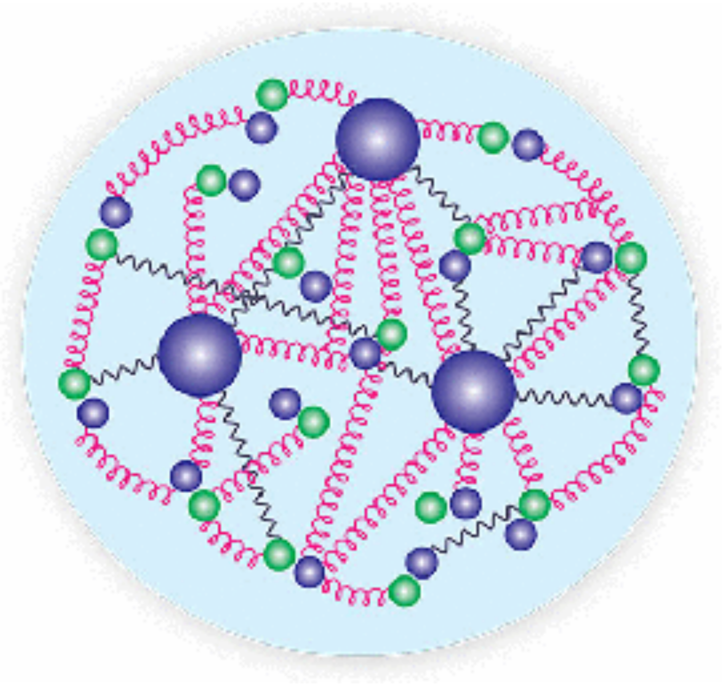
Quark/gluon binding?

- **Experimental fact**: potentially complex **electron-proton** interaction really due to simple scattering between point-like **electron** and **quark** within proton, with rest of proton playing **no role**.



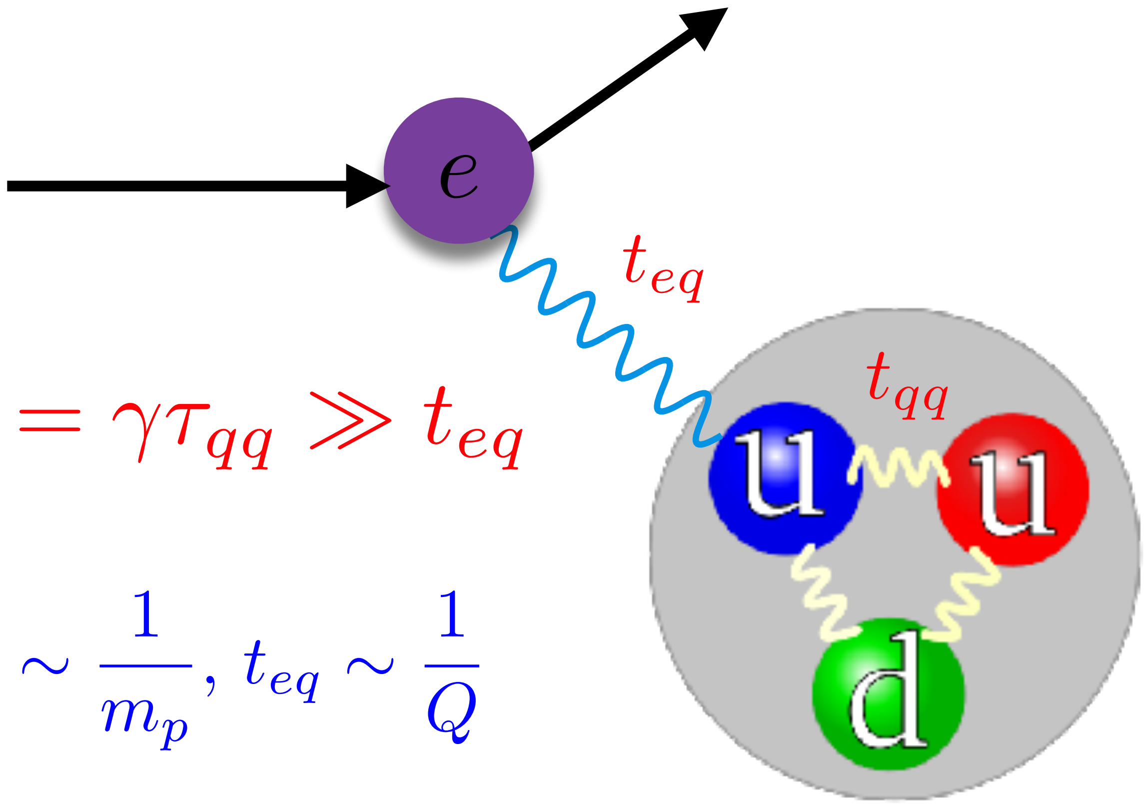
$$\sigma(ep) \sim \sigma(eq)$$

- But how can this be? Quarks are part of a complex and **self-interacting bound** state: the proton. Does this not affect the scattering process?



Proton 'Snapshot'

- Relativity comes to rescue. Proton at **rest**: complex system of interacting quarks ($\tau_{qq} \approx 10^{-24} s$).
- However we are interested in very **high energy** proton collisions. Proton has velocity **v ~ c**, and **relativity** comes into the game.
- What does electron 'see'? **Time dilation**: proton 'clock' much slower than when at rest, electron only sees a ~ **static snapshot** of the proton!



$$t_{qq} = \gamma \tau_{qq} \gg t_{eq}$$

$$t_{qq} \sim \frac{1}{m_p}, \; t_{eq} \sim \frac{1}{Q}$$

- Electron-quark interaction time ≪ timescale of internal quark interactions!

- Note separation of timescales/distances at heart of all QCD 'factorization'.

Colliding Protons

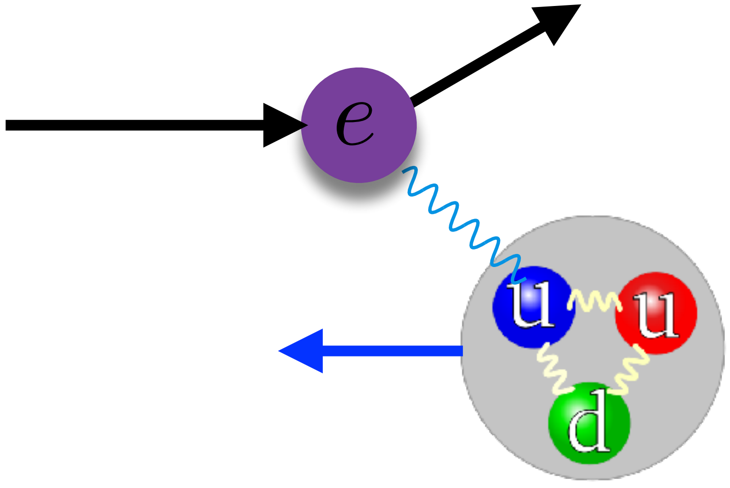
- Electron scatters off a quark within the **static** proton **snapshot**. The **quark interactions** within the proton are **frozen** and can be ignored.
- Valid to consider in terms of **free** quark-electron scattering.

$$\sigma(ep) \sim \sigma(eq)$$ ✓

- Final element: what does the frozen distribution of quarks look like?

Relevant degree of freedom: amount of proton's energy carried by quark.

- Introduce **new variable**: $x = \frac{E_{\text{quark}}}{E_{\text{proton}}}$ $\qquad 0 < x < 1$



$p_p = (E, 0, 0, E)$

$p_q = x(E, 0, 0, E)$

c.m.s. frame

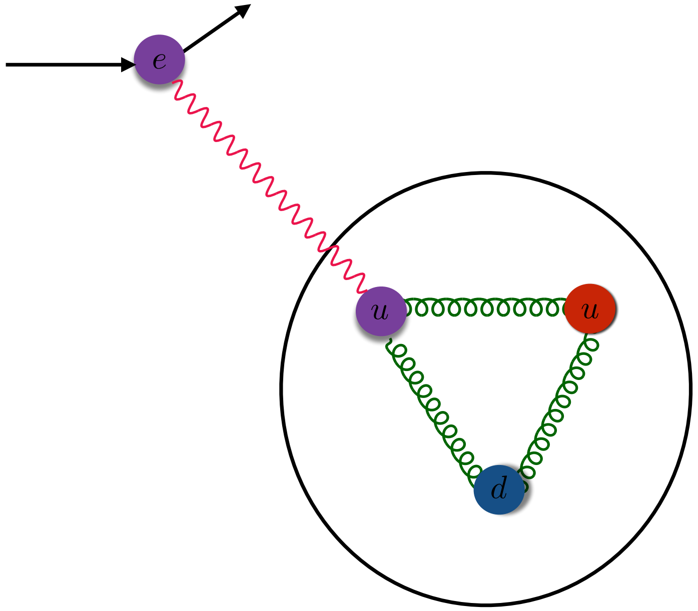
$m_p \ll E$

Mapping out the Proton

- Collision by collision the proton snapshot will be different, but expect it to have some **statistical distribution**.

- Statistical distribution known as '**Parton Distribution Function**' (**PDF**).

$f(x)$: **Probability of finding a quark with energy fraction x in proton snapshot.**



Electron-proton scattering rate:

$$\sigma_{ep} = \int_0^1 \mathrm{d}x \, f_q(x) \, \sigma_{eq}(x)$$

- In DIS: electron energy and scattering angle directly related to quark x and photon Q^2 . By scanning over these, can map out distribution.

- What do they look like?

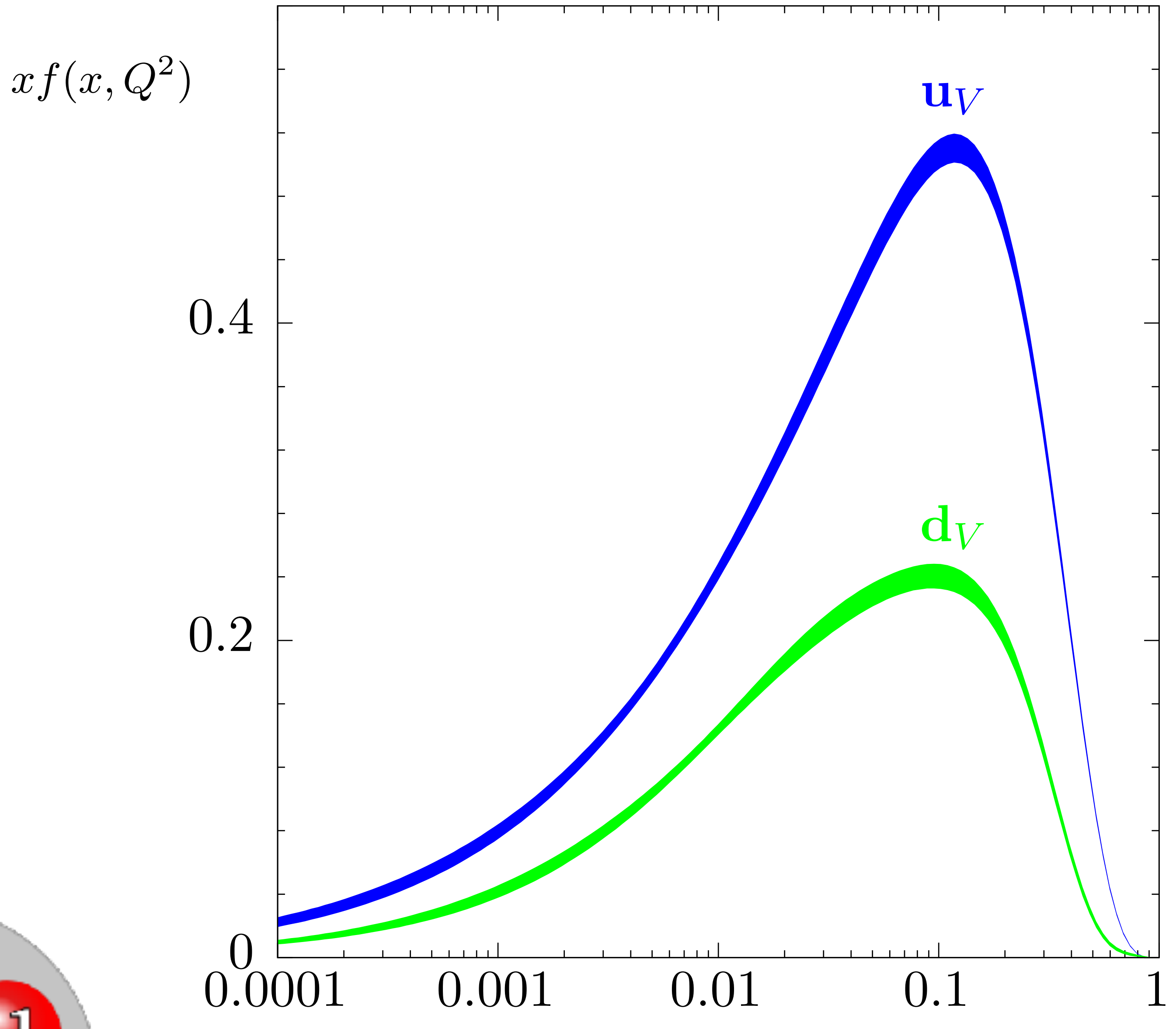
At LO in QCD: $x = x_B = \dfrac{Q^2}{2p \cdot q}$

The Proton Backbone

- Without going into details, some broad features can be picked out:

$u_V = u - \overline{u} \qquad d_V = d - \overline{d}$

$\mu_F \sim M_Z$

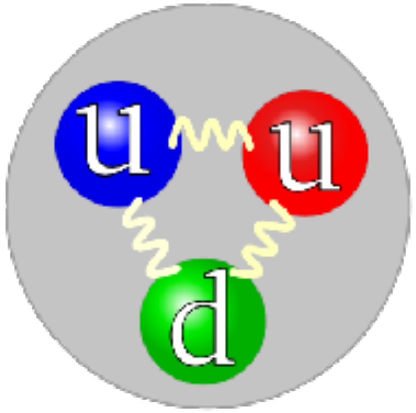


- 'Valence' up and down quark structure consistent with basic **uud** picture.

- $u_V \approx 2d_V$

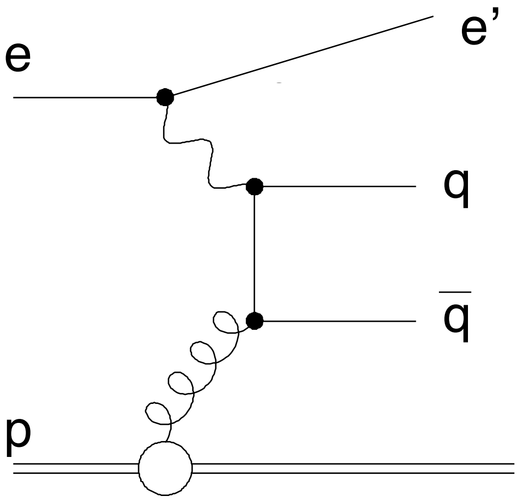
- Peaked at $x \approx \frac{1}{3}$.

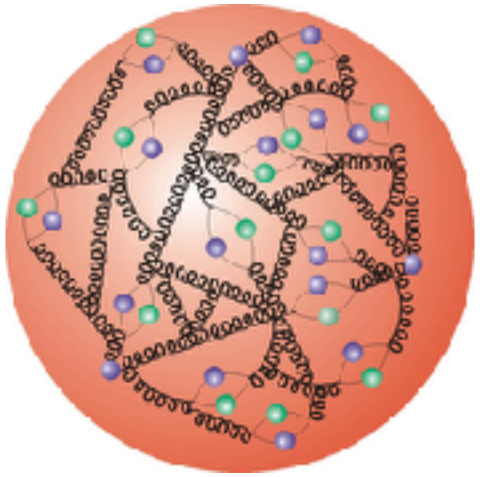
- Not exactly $x = \frac{1}{3}$ as quark interactions can redistribute momenta.

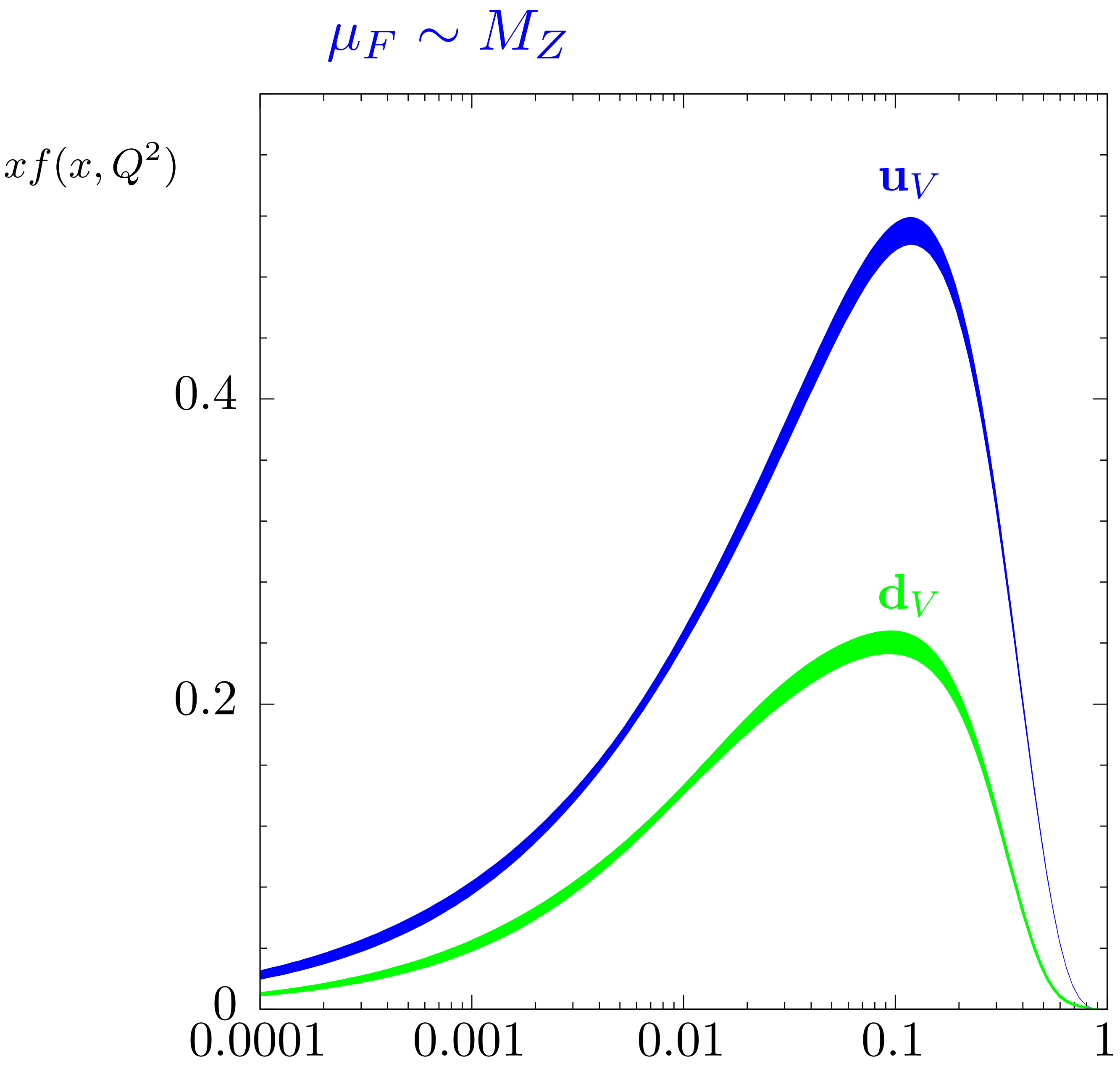


Gluons

- What happens when we add **gluons** into the picture? Can also scatter off these.

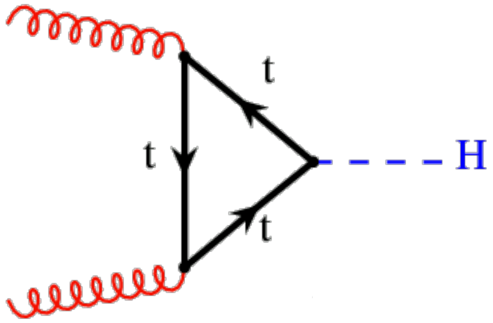


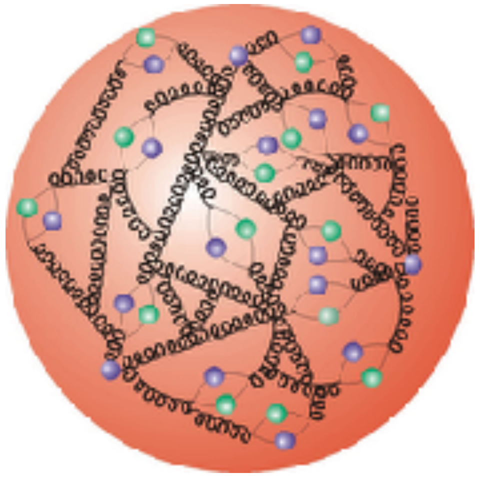


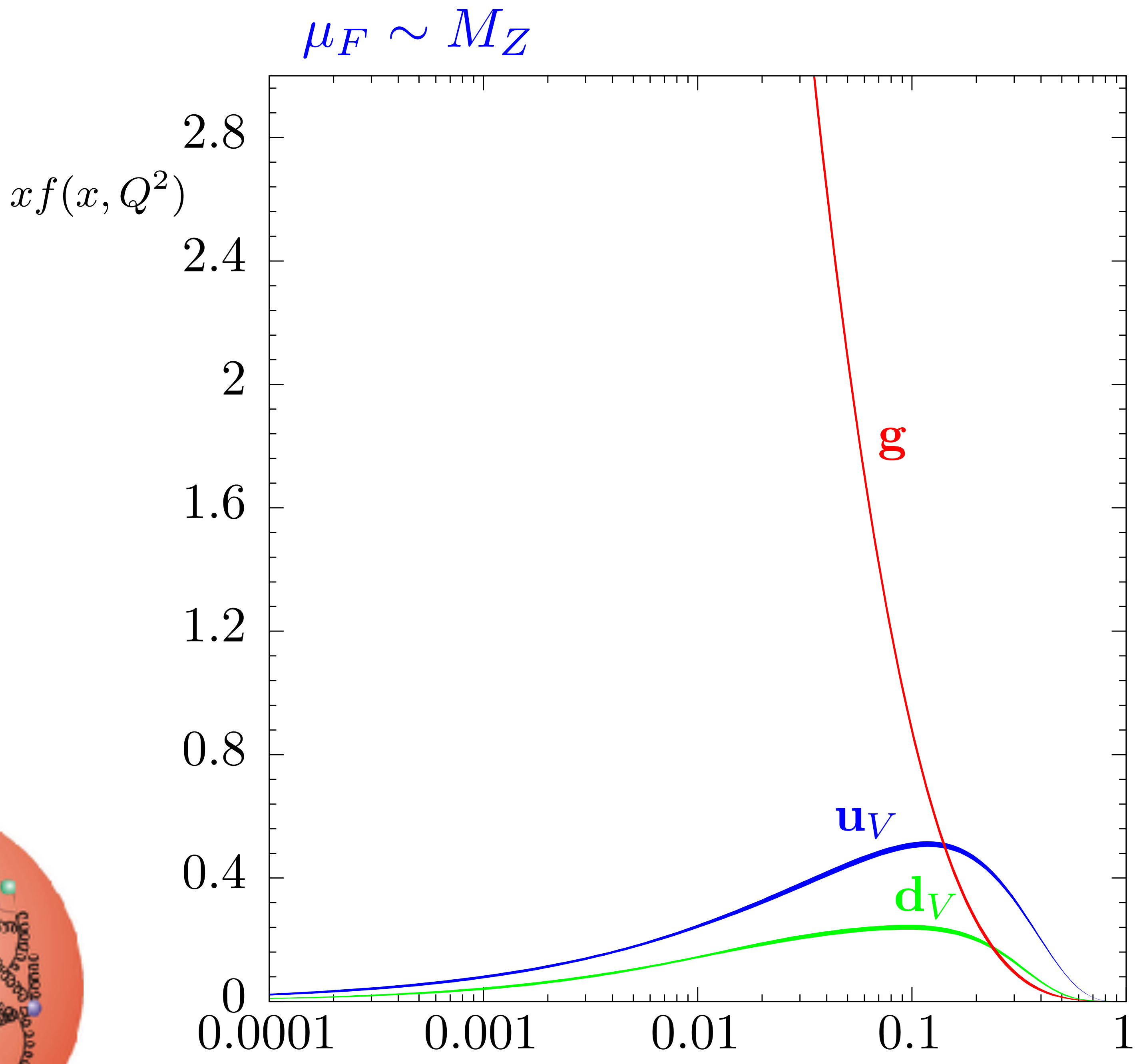


Gluons!

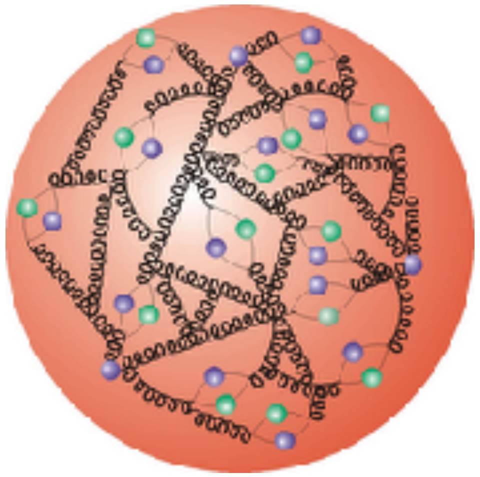
- Contribution from gluons is huge.
- In fact, roughly **50%** of proton energy carried by gluons.
- LHC a gluon-gluon collider: crucial for e.g. **Higgs** physics.

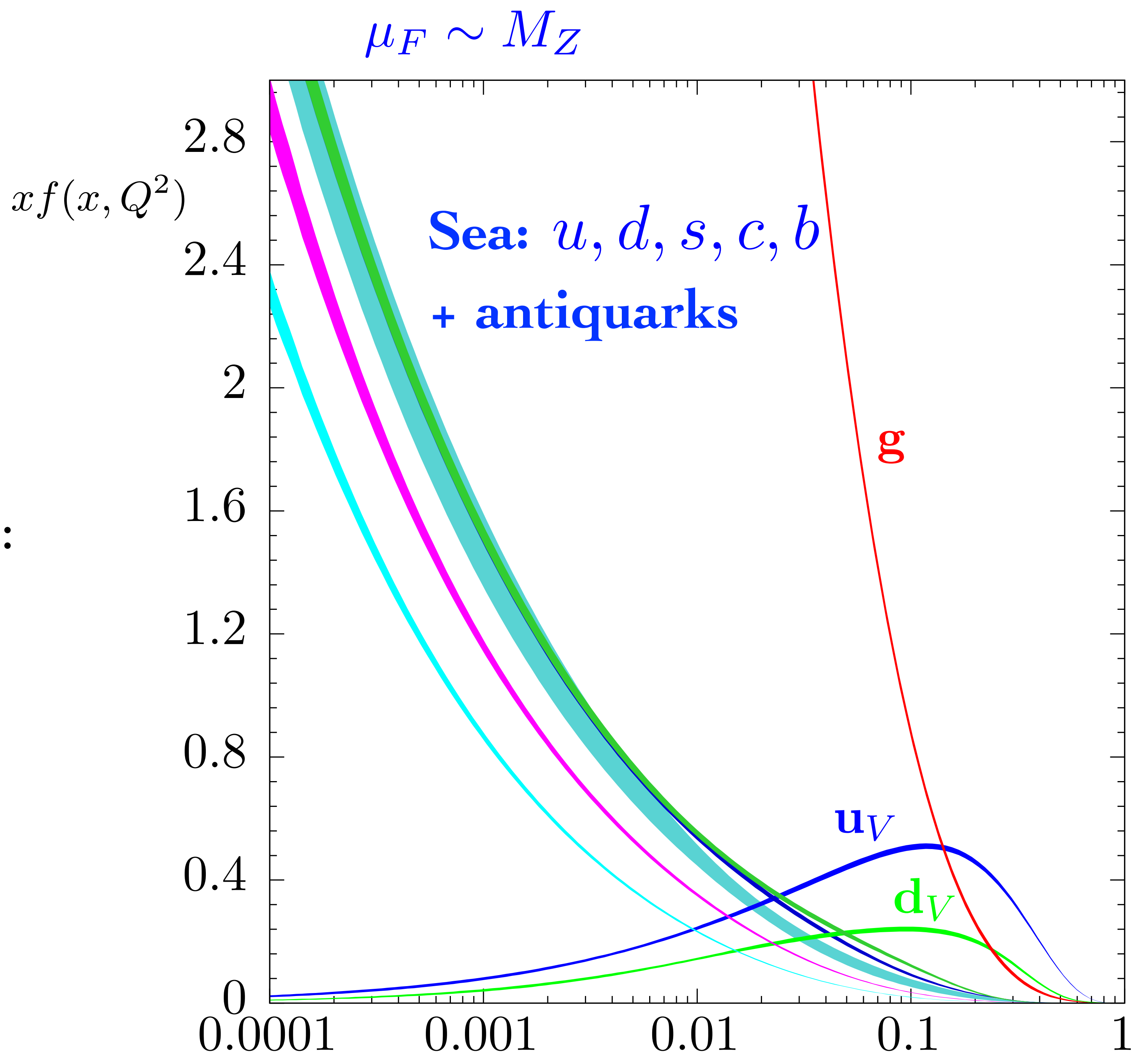






The Quark Sea

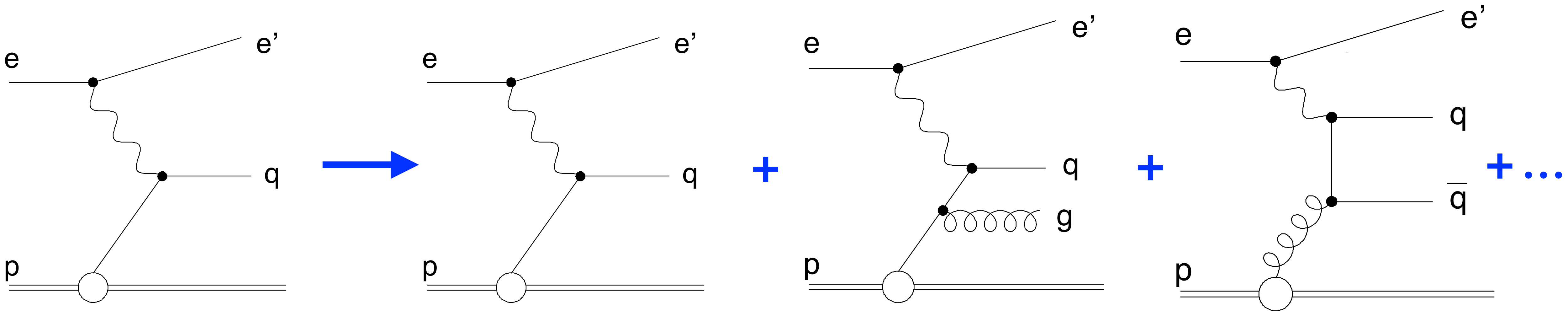




- Quark content of proton not just due to **uud** valence: virtual $q\bar{q}$ pairs contribute.
- This quark '**sea**' completely dominant over valence in many regions.

Adding QCD

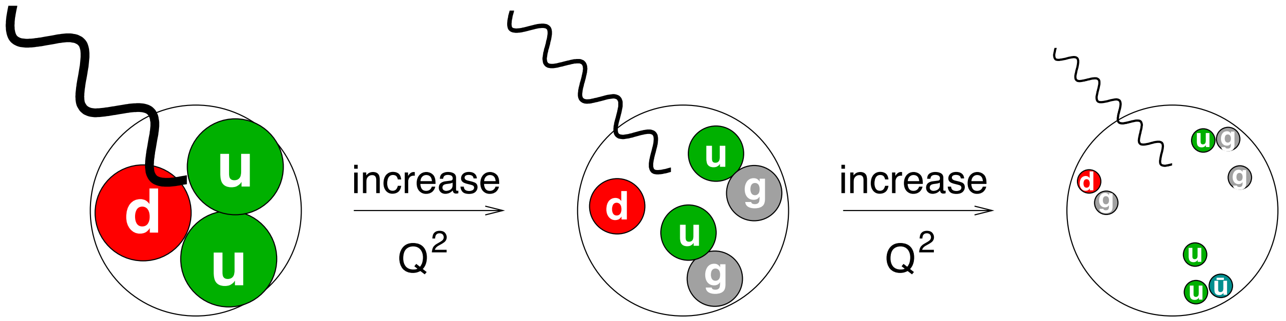
- Simple picture somewhat complicated by inclusion of **interacting QCD**: quarks and gluons like to **radiate**:



- Amount of radiation depends on scale $\Rightarrow$ so do PDFs:

$$\sigma^{lp} \sim \sigma^{lq}(\mu_F) \otimes q(x,\mu_F) \quad \mu_F : \text{ factorization scale} \quad \text{DIS: } \mu_F^2 = Q^2$$

- Physically $Q^2 \sim$ **resolution** at which we resolve proton:

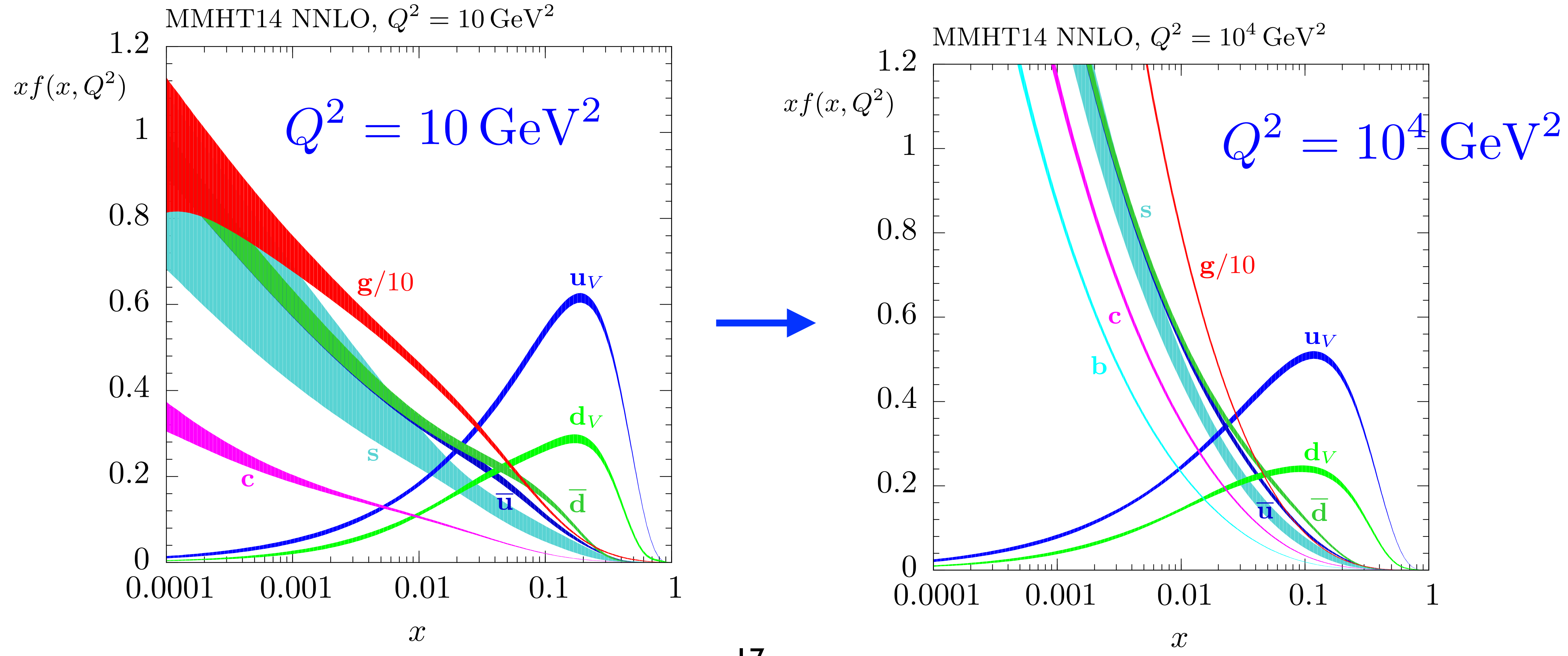


- Note this effect means we lose the interpretation in terms of a classical probability distribution, but still intuitive to think of PDFs in those terms.

- Dependence of PDFs on resolution scale predicted by QCD, via ‘**DGLAP**’ equation:

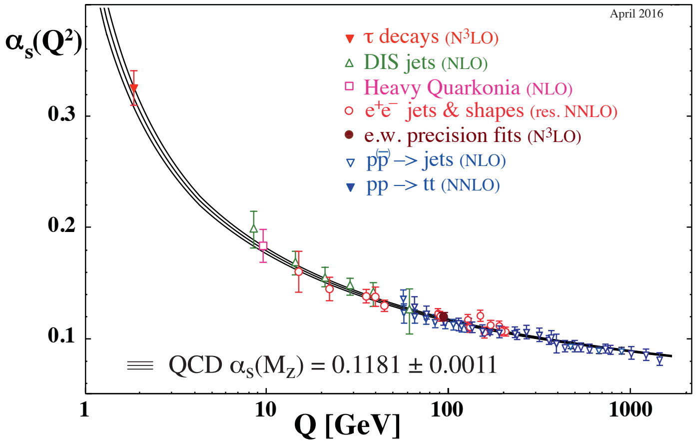
$$\frac{\partial q(x,\mu_F^2)}{\partial \mu_F} = P_{qq} \otimes q(x,\mu_F) + P_{qg} \otimes g(x,\mu_F)$$

- Specifics rather complicated, basic impact simple: higher $Q^2 \Rightarrow$ more $q, \overline{q}, g$ at low x, less at high x, due to radiation ($q \to qg, g \to q\overline{q}, g \to gg...$).



Extracting PDFs

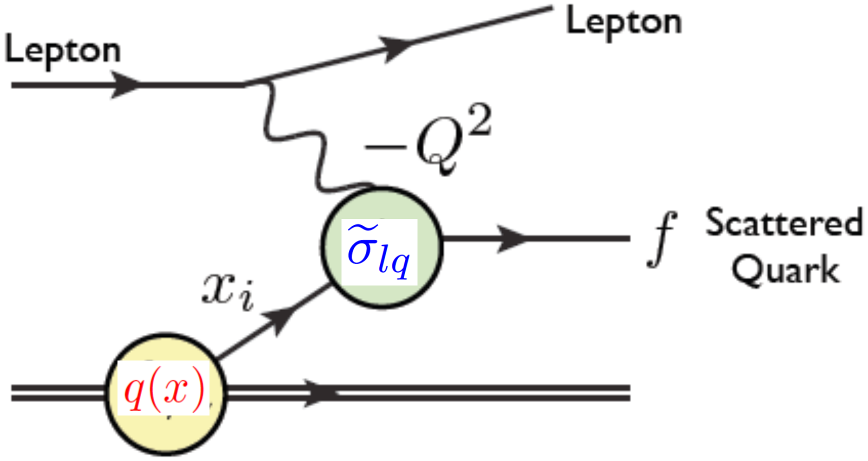
Extracting PDFs

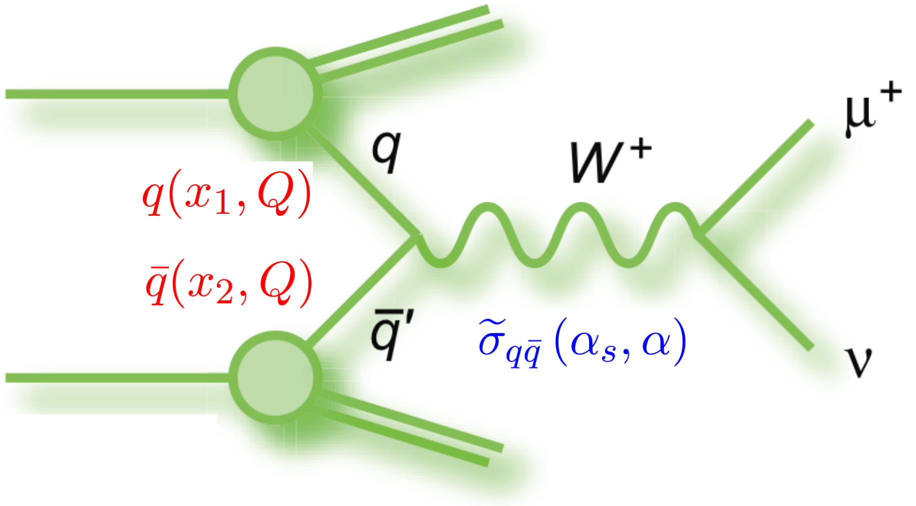


- Binding of quark/gluons in proton due to low-energy QCD $\Rightarrow$ **cannot** use perturbation theory.

- However **PDFs** are **universal**: same quark (antiquark) PDFs enter DIS and Drell-Yan cross sections.

$$\sigma_{lp} \simeq \widetilde{\sigma}_{lq}\,(\alpha_s, \alpha) \otimes q(x, Q) \qquad \sigma_{pp} \simeq \widetilde{\sigma}_{q\bar{q}}\,(\alpha_s, \alpha) \otimes q(x_1, Q) \otimes \bar{q}(x_2, Q)$$





$$\text{Factorization} \Rightarrow q_{DIS}(x, Q^2) \equiv q_{DY}(x, Q^2)$$

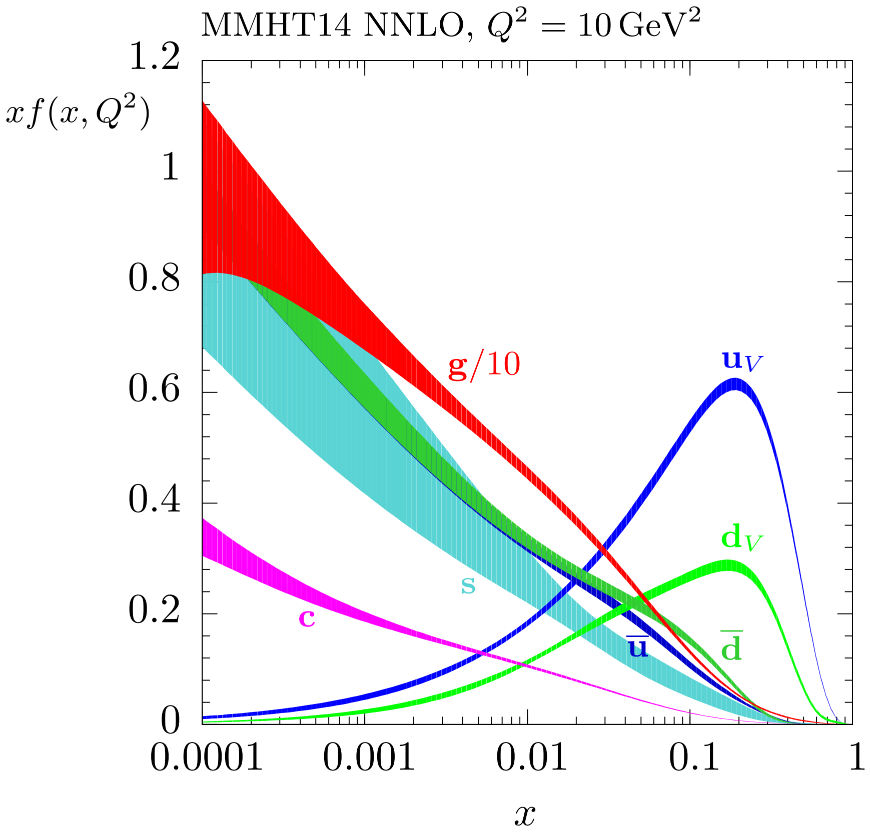
$\longrightarrow$ **Fit** PDFs to one dataset (e.g. DIS) and use to make prediction for another (e.g. DY).

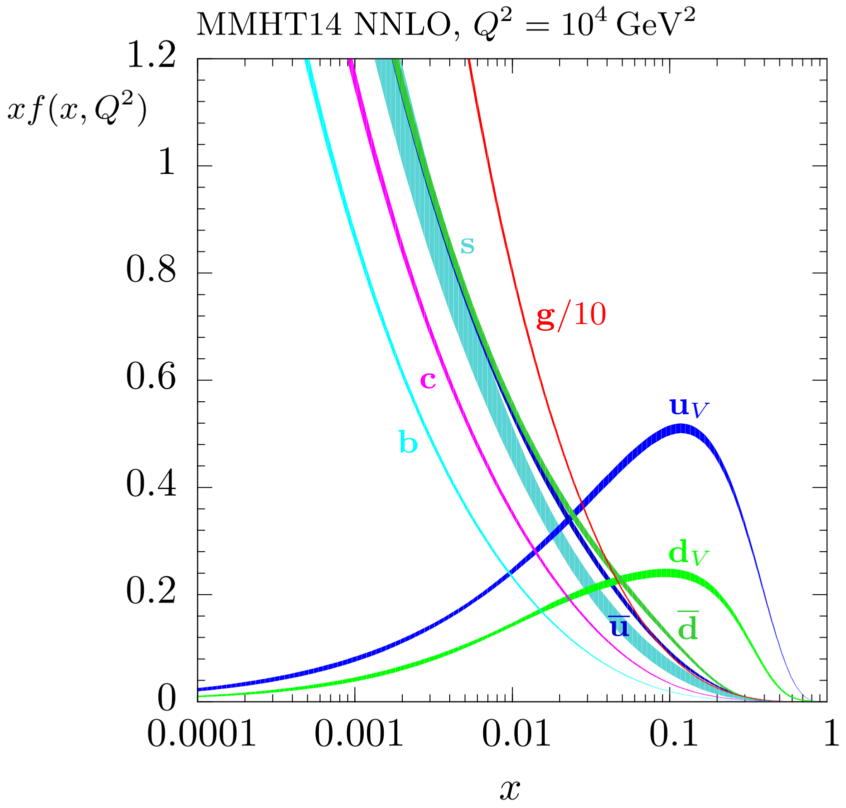
PDF Fits

- For LHC (and elsewhere) aim to constrain PDFs to high precision for all flavours (q, $\bar{q}$, g …) over a wide x region. To achieve this: performs **global PDF fits** to wide range of data.

- Various major global fitting collaboration (**ABM, CT, MMHT, NNPDF**), each taking different approach to this.

- Also various specialised PDF sets: **CJ** (focus on high x), **HERAPDF** (fit to HERA data alone), while **ATLAS/CMS** also performing fits to their data.





PDF Fits: Work Flow

$$O \sim f \otimes \sigma$$

Basic idea: **measure (data)** ↗ O **fit** ↑ f **predict (pQCD)** ↖ σ

In detail…

Parameterise PDFs at low scale Q_0

↓

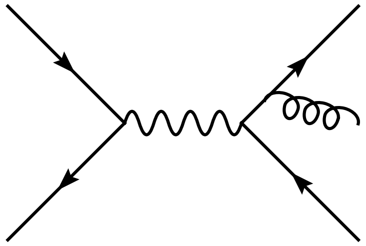
$$f_i(x, Q_0) : \ A_f x^{a_f} (1-x)^{b_f} \times \xrightarrow{} \sum_{i=1}^{n} \alpha_{f,i} P_i(y(x)) \, ,$$ **CT, MMHT…**

$$\searrow \ \mathrm{NN}_i(x)$$ **NNPDF**

Select data & theory

↓

Perturbative QCD for parton-level theory



+ …

Perform fit

↓

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

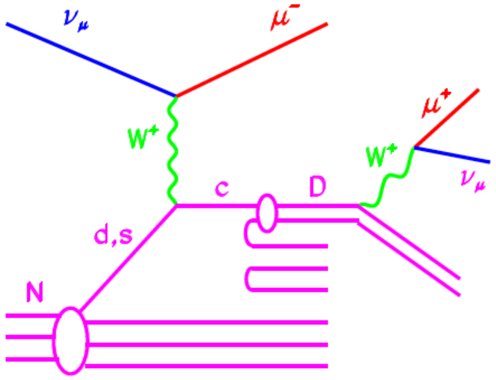
DGLAP: $f(x, Q_0) \to f(x, \mu_{\mathrm{data}})$

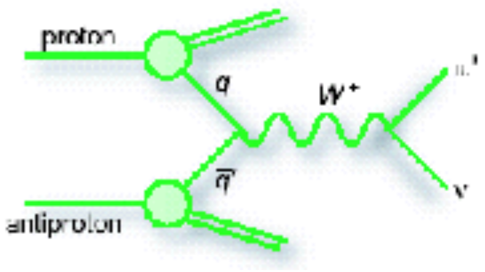
Output PDFs

$$f(x,\mu) \pm \Delta(x,\mu)$$

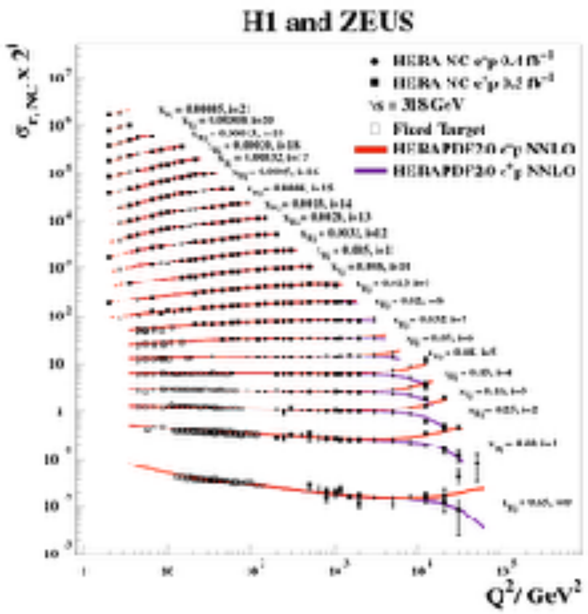
↖ **Due to error in data in fit**

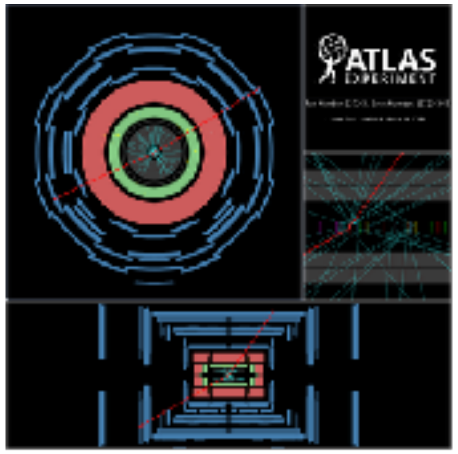
Global Fits: Datasets

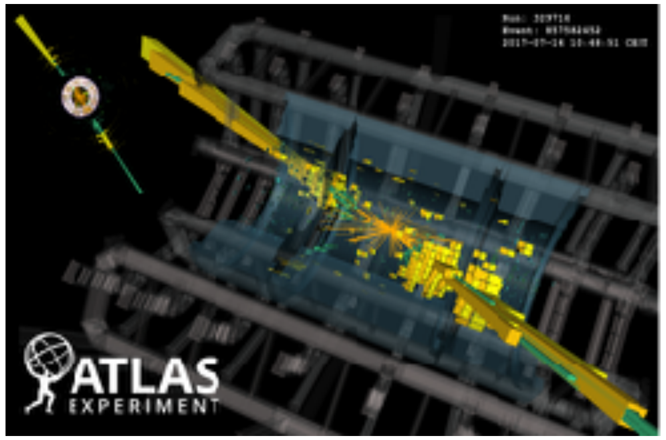




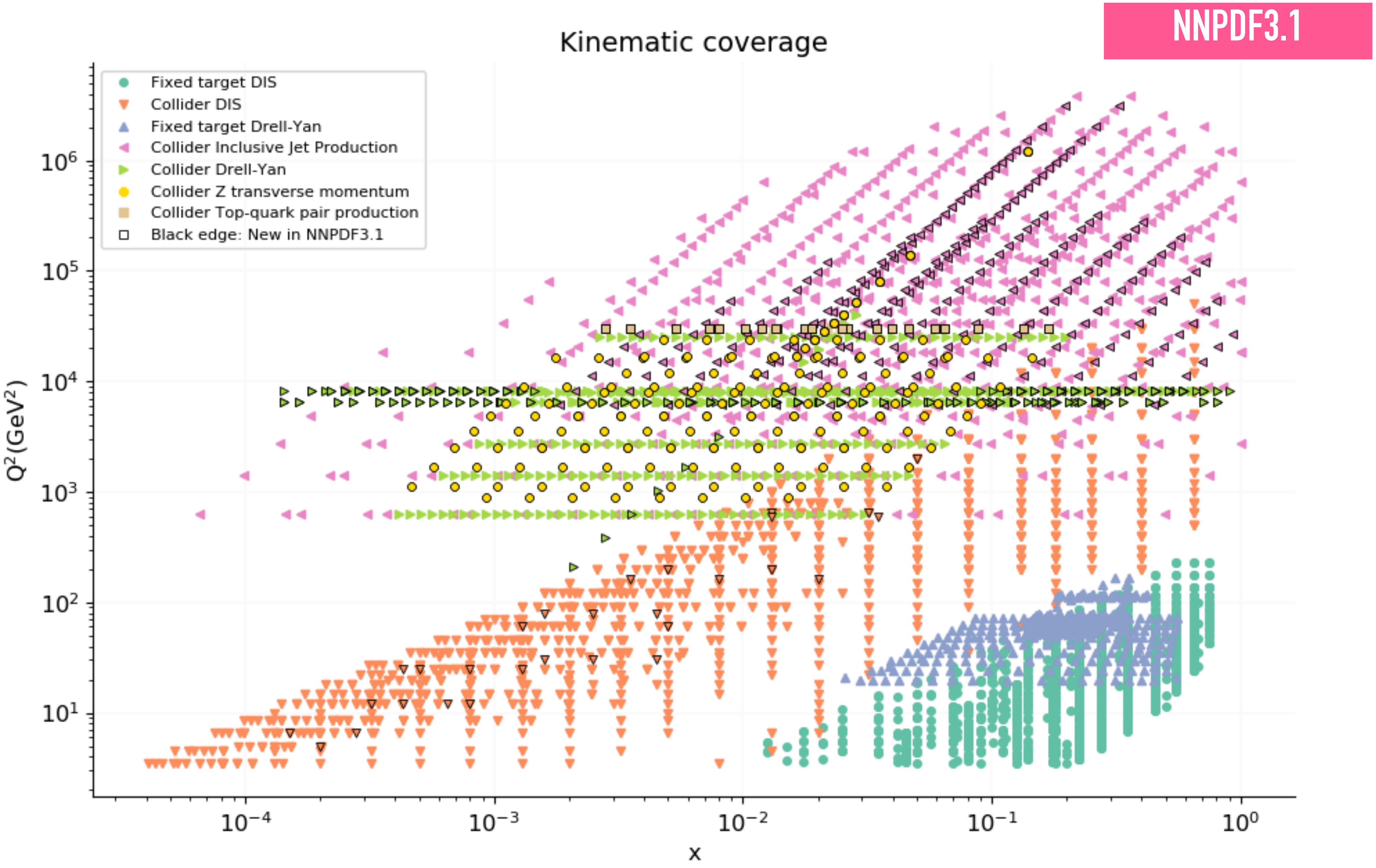
	Process	Subprocess	Partons	x range
Fixed Target	$\ell^\pm\{p,n\} \to \ell^\pm + X$	$\gamma^* q \to q$	$q, \bar{q}, g$	$x \gtrsim 0.01$
	$\ell^\pm n/p \to \ell^\pm + X$	$\gamma^* d/u \to d/u$	d/u	$x \gtrsim 0.01$
	$pp \to \mu^+\mu^- + X$	$u\bar{u}, d\bar{d} \to \gamma^*$	$\bar{q}$	$0.015 \lesssim x \lesssim 0.35$
	$pn/pp \to \mu^+\mu^- + X$	$(u\bar{d})/(u\bar{u}) \to \gamma^*$	$\bar{d}/\bar{u}$	$0.015 \lesssim x \lesssim 0.35$
	$\nu(\bar{\nu})\,N \to \mu^-(\mu^+) + X$	$W^* q \to q'$	$q, \bar{q}$	$0.01 \lesssim x \lesssim 0.5$
	$\nu\,N \to \mu^-\mu^+ + X$	$W^* s \to c$	s	$0.01 \lesssim x \lesssim 0.2$
	$\bar{\nu}\,N \to \mu^+\mu^- + X$	$W^* \bar{s} \to \bar{c}$	$\bar{s}$	$0.01 \lesssim x \lesssim 0.2$
Collider DIS	$e^\pm p \to e^\pm + X$	$\gamma^* q \to q$	$g, q, \bar{q}$	$0.0001 \lesssim x \lesssim 0.1$
	$e^+ p \to \bar{\nu} + X$	$W^+\{d, s\} \to \{u, c\}$	d, s	$x \gtrsim 0.01$
	$e^\pm p \to e^\pm c\bar{c} + X$	$\gamma^* c \to c, \gamma^* g \to c\bar{c}$	c, g	$10^{-4} \lesssim x \lesssim 0.01$
	$e^\pm p \to e^\pm b\bar{b} + X$	$\gamma^* b \to b, \gamma^* g \to b\bar{b}$	b, g	$10^{-4} \lesssim x \lesssim 0.01$
	$e^\pm p \to \text{jet} + X$	$\gamma^* g \to q\bar{q}$	g	$0.01 \lesssim x \lesssim 0.1$
Tevatron	$p\bar{p} \to \text{jet} + X$	$gg, qg, qq \to 2j$	g, q	$0.01 \lesssim x \lesssim 0.5$
	$p\bar{p} \to (W^\pm \to \ell^\pm \nu) + X$	$ud \to W^+, \bar{u}\bar{d} \to W^-$	$u, d, \bar{u}, \bar{d}$	$x \gtrsim 0.05$
	$p\bar{p} \to (Z \to \ell^+\ell^-) + X$	$uu, dd \to Z$	u, d	$x \gtrsim 0.05$
	$p\bar{p} \to t\bar{t} + X$	$qq \to t\bar{t}$	q	$x \gtrsim 0.1$
LHC	$pp \to \text{jet} + X$	$gg, qg, q\bar{q} \to 2j$	g, q	$0.001 \lesssim x \lesssim 0.5$
	$pp \to (W^\pm \to \ell^\pm \nu) + X$	$u\bar{d} \to W^+, d\bar{u} \to W^-$	$u, d, \bar{u}, \bar{d}, g$	$x \gtrsim 10^{-3}$
	$pp \to (Z \to \ell^+\ell^-) + X$	$q\bar{q} \to Z$	$q, \bar{q}, g$	$x \gtrsim 10^{-3}$
	$pp \to (Z \to \ell^+\ell^-) + X,\ p_\perp$	$gq(\bar{q}) \to Zq(\bar{q})$	$g, q, \bar{q}$	$x \gtrsim 0.01$
	$pp \to (\gamma^* \to \ell^+\ell^-) + X$, Low mass	$q\bar{q} \to \gamma^*$	$q, \bar{q}, g$	$x \gtrsim 10^{-4}$
	$pp \to (\gamma^* \to \ell^+\ell^-) + X$, High mass	$q\bar{q} \to \gamma^*$	$\bar{q}$	$x \gtrsim 0.1$
	$pp \to W^+\bar{c}, W^-c$	$sg \to W^+c, \bar{s}g \to W^-\bar{c}$	$s, \bar{s}$	$x \sim 0.01$
	$pp \to t\bar{t} + X$	$gg \to t\bar{t}$	g	$x \gtrsim 0.01$
	$pp \to D, B + X$	$gg \to c\bar{c}, b\bar{b}$	g	$x \gtrsim 10^{-6}, 10^{-5}$
	$pp \to J/\psi, \Upsilon + pp$	$\gamma^*(gg) \to c\bar{c}, b\bar{b}$	g	$x \gtrsim 10^{-6}, 10^{-5}$
	$pp \to \gamma + X$	$gq(\bar{q}) \to \gamma q(\bar{q})$	g	$x \gtrsim 0.005$







Global Fits: Kinematic Coverage



- Global fits achieve **broad coverage** from low to high x , and over many orders of magnitude in Q^2.

Fit Quality

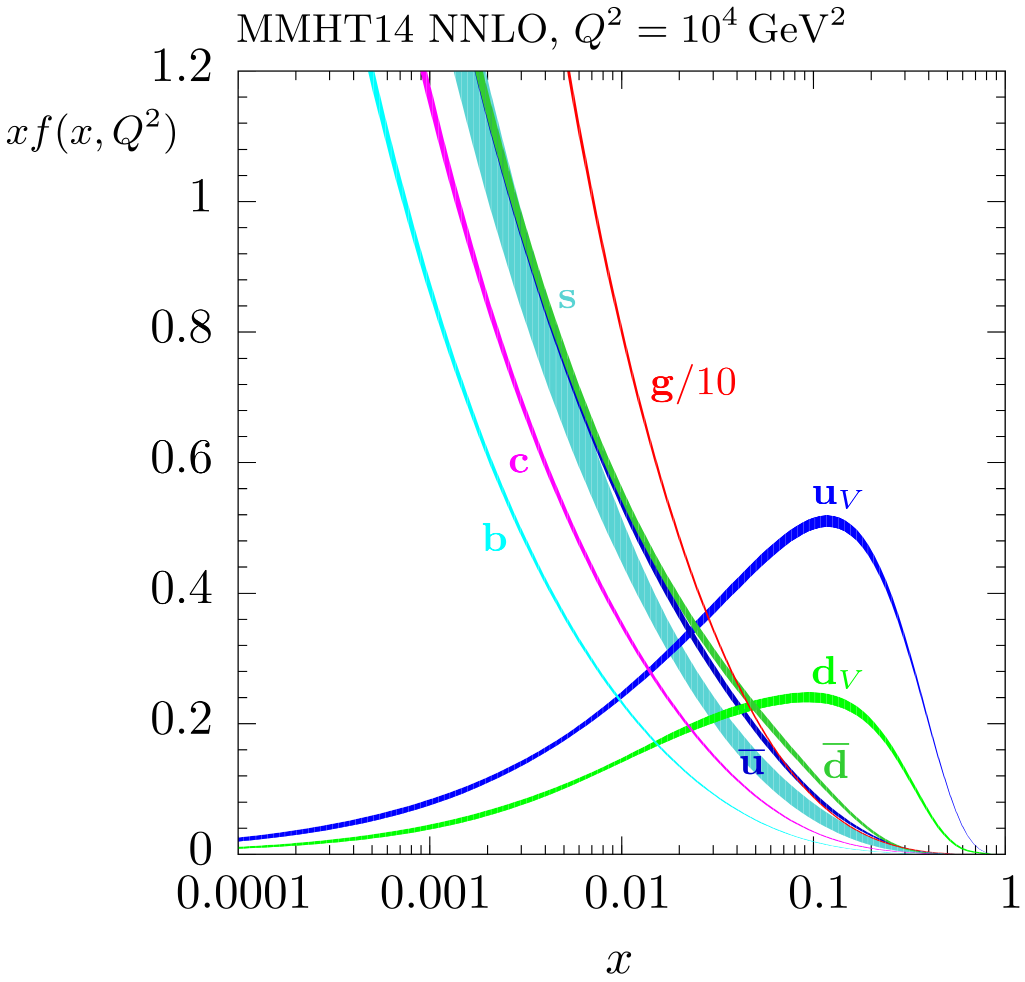
- Fits to wide range of data from different colliders/experiments. Is a good/ reliable fit possible from this? **Yes**!

$\chi^2/\mathrm{dof} \sim 1$

$\Rightarrow$ **Non-trivial check of QCD.**

Data set	LO	NLO	NNLO
BCDMS μp F_2 [125]	162 / 153	176 / 163	173 / 163
BCDMS μd F_2 [19]	140 / 142	143 / 151	143 / 151
NMC μp F_2 [20]	141 / 115	132 / 123	123 / 123
NMC μd F_2 [20]	134 / 115	115 / 123	108 / 123
NMC $\mu n/\mu p$ [21]	122 / 137	131 / 148	127 / 148
E665 μp F_2 [22]	59 / 53	60 / 53	65 / 53
E665 μd F_2 [22]	52 / 53	52 / 53	60 / 53
SLAC ep F_2 [23, 24]	21 / 18	31 / 37	31 / 37
SLAC ed F_2 [23, 24]	13 / 18	30 / 38	26 / 38
NMC/BCDMS/SLAC/HERA F_L [20, 125, 24, 63, 64, 65]	113 / 53	68 / 57	63 / 57
E866/NuSea pp DY [88]	229 / 184	221 / 184	227 / 184
E866/NuSea pd/pp DY [89]	29 / 15	11 / 15	11 / 15
NuTeV νN F_2 [29]	35 / 49	39 / 53	38 / 53
CHORUS νN F_2 [30]	25 / 37	26 / 42	28 / 42
NuTeV νN xF_3 [29]	49 / 42	37 / 42	31 / 42
CHORUS νN xF_3 [30]	35 / 28	22 / 28	19 / 28
CCFR $\nu N \to \mu\mu X$ [31]	65 / 86	71 / 86	76 / 86
NuTeV $\nu N \to \mu\mu X$ [31]	53 / 40	38 / 40	43 / 40
HERA e^+p NC 820 GeV[61]	125 / 78	93 / 78	89 / 78
HERA e^+p NC 920 GeV[61]	479 /330	402 /330	373/ 330
HERA e^-p NC 920 GeV [61]	158/ 145	129/ 145	125 /145
HERA e^+p CC [61]	41 / 34	34 / 34	32 / 34
HERA e^-p CC [61]	29 / 34	23 / 34	21 / 34
HERA ep F_2^{charm} [62]	105 /52	72 / 52	82 / 52
H1 99–00 e^+p incl. jets [126]	77 / 24	14 / 24	—
ZEUS incl. jets [127, 128]	140/60	45 / 60	—
DØ II $p\bar{p}$ incl. jets [119]	125 / 110	116 / 110	119 / 110
CDF II $p\bar{p}$ incl. jets [118]	78 / 76	63 / 76	59 / 76
CDF II W asym. [66]	55 / 13	32 / 13	30 / 13
DØ II $W \to \nu e$ asym. [67]	47 / 12	28 / 12	27 / 12
DØ II $W \to \nu\mu$ asym. [68]	16 / 10	19 / 10	21 / 10
DØ II Z rap. [90]	34 / 28	16 / 28	16 / 28
CDF II Z rap. [70]	95 / 28	36 / 28	40 / 28
ATLAS W^+, W^-, Z [10]	94/30	38/30	39/30
CMS W asymm $p_T > 35$ GeV [9]	10/11	7/11	9/11
CMS asymm $p_T > 25$ GeV, 30 GeV[77]	7/24	8/24	10/24
LHCb $Z \to e^+e^-$[79]	76/9	13/9	20/9
LHCb W asymm $p_T > 20$ GeV[78]	27/10	12/10	16/10
CMS $Z \to e^+e^-$ [84]	46/35	19/35	22/35
ATLAS high-mass Drell-Yan [83]	42/13	21/13	17/13
CMS double diff. Drell-Yan [86]	—	372/132	149/132
Tevatron, ATLAS, CMS $\sigma_{t\bar{t}}$ [91]–[97]	53/13	7/13	8/13
ATLAS jets (2.76 TeV+7 TeV)[108, 107]	162/116	106/116	—
CMS jets (7 TeV) [106]	150/133	138/133	—

All data sets	**3706 / 2763**	**3267 / 2996**	**2717 / 2663**
	LO	**NLO**	**NNLO**



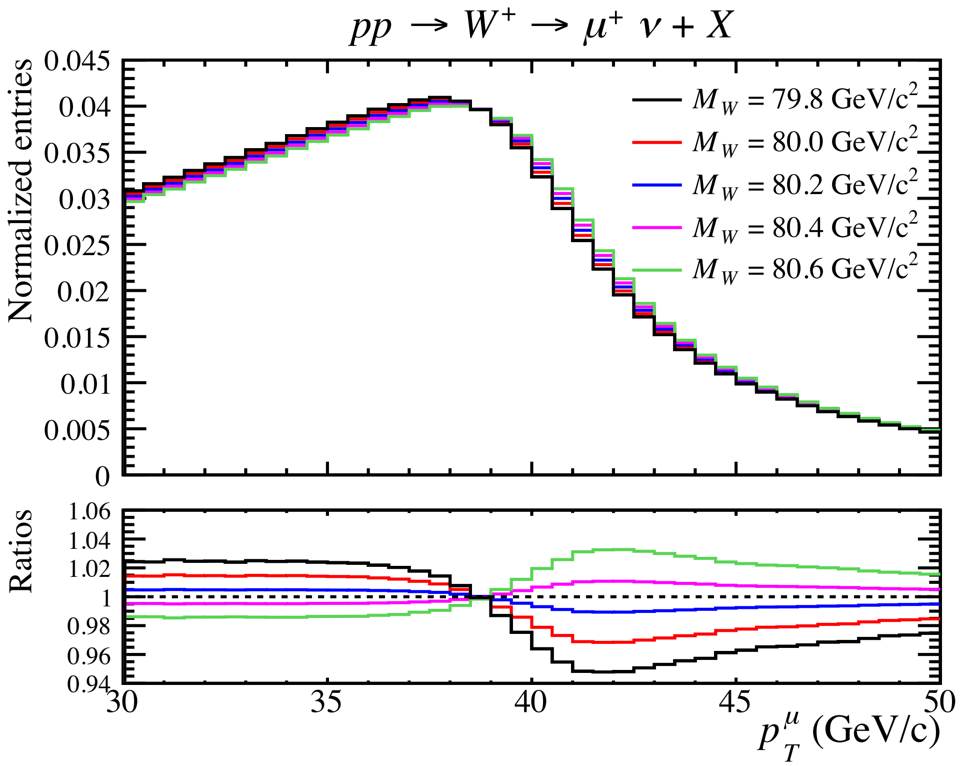
LHL et al., Eur. Phys. J. C75 (2015) no.5 204

PDFs at the LHC

Precise PDFs for the LHC

- Why do we care about PDFs at LHC?

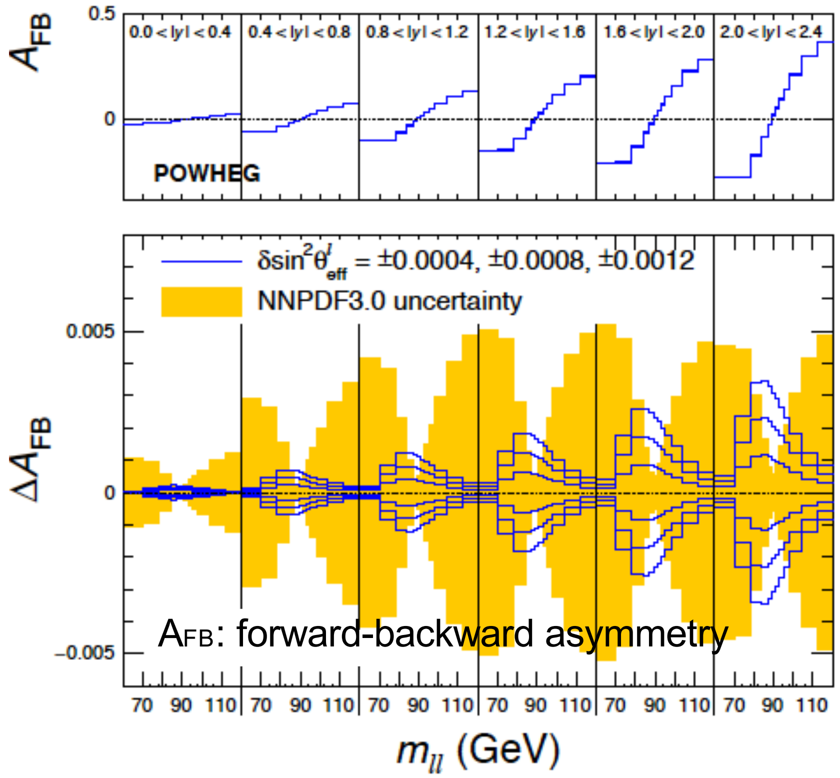
★ **Precision SM** measurements.



S. Farry et al., EPJC79 (2019) no.6, 497

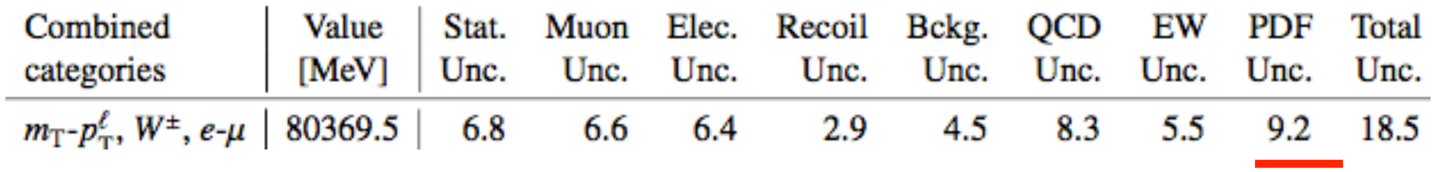
- W mass: fit to lepton ($W \to l\nu$) kinematics.

CMS collab., EPJC78 (2018) no.9, 701



- Weak mixing angle θ_W : lepton decay distribution ($Z \to l^+l^-$) w.r.t. initial quark.

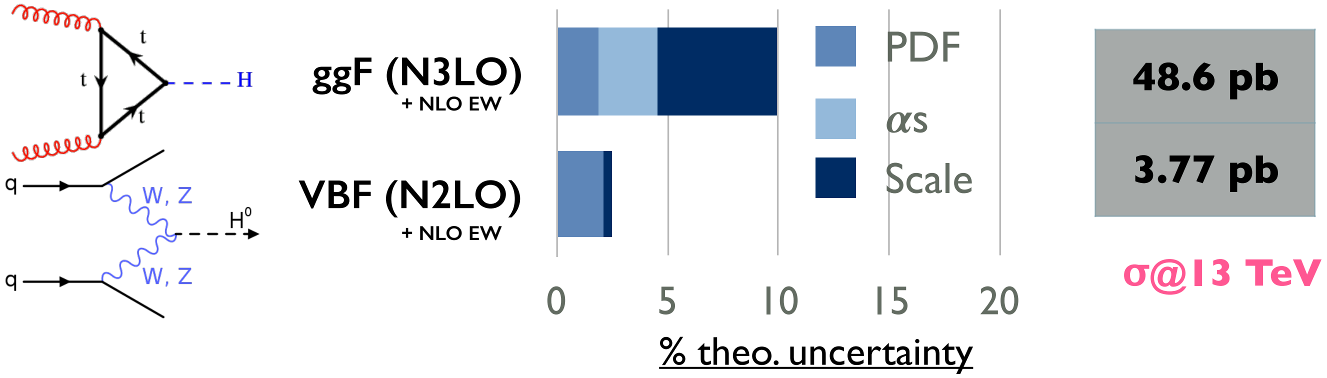
ATLAS collab., EPJC78 (2018) no.2, 110

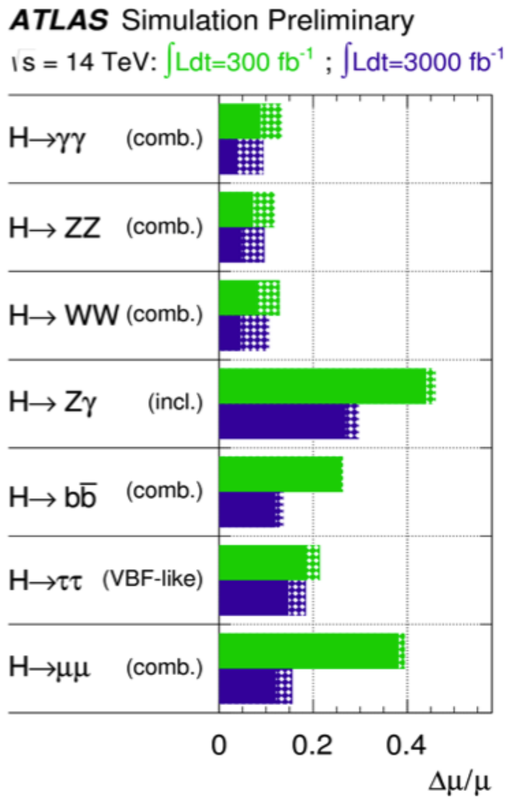


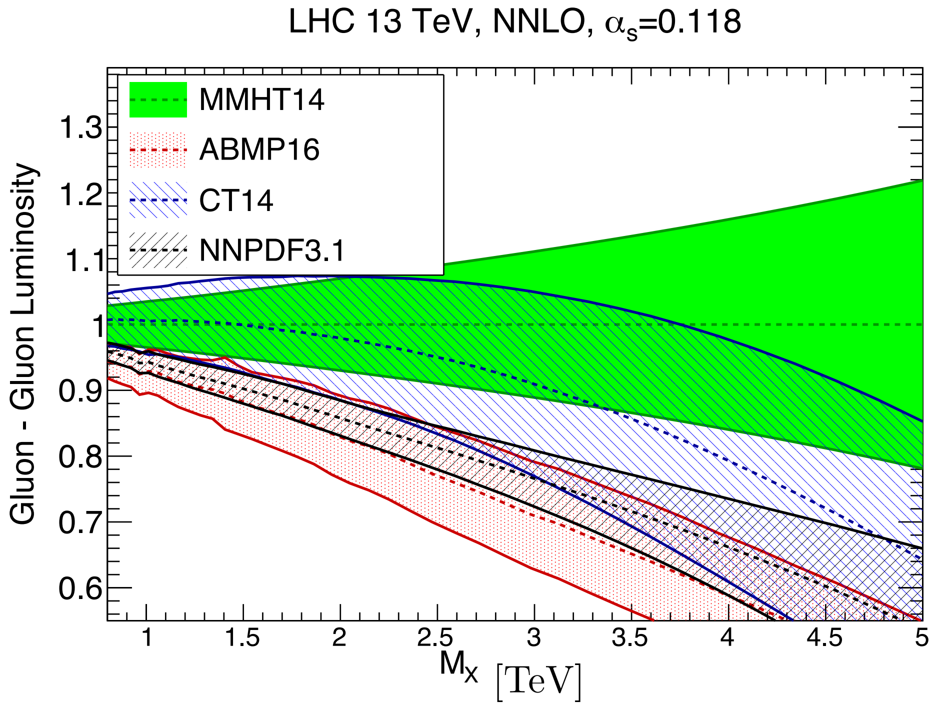
Combined categories	Value [MeV]	Stat. Unc.	Muon Unc.	Elec. Unc.	Recoil Unc.	Bckg. Unc.	QCD Unc.	EW Unc.	PDF Unc.	Total Unc.
m_T-p_T^ℓ, $W^\pm$, e-μ	80369.5	6.8	6.6	6.4	2.9	4.5	8.3	5.5	9.2	18.5

- Both approaching level of indirect EW determination, but strongly sensitive to **PDF uncertainties** via underlying $q\bar{q} \to W, Z$ process.

★ **Higgs couplings** → need to model SM production precisely.







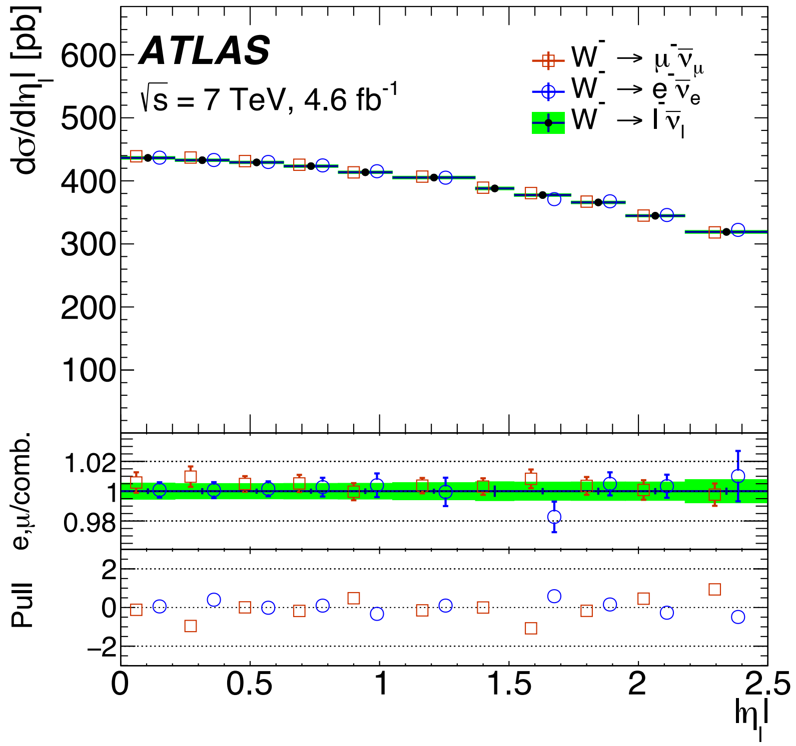
★ **High mass searches** for new resonances/contact interaractions - PDFs in high x region. Currently constraints poor.

⟶ Ultimate reach of LHC limited by knowledge of PDFs.

Fits Today

- Current fits very much aiming for (and in some cases achieving) high precision (~ **1% level**) PDF determination in some regions. Key ingredients:

Extremely precise LHC data



ATLAS collab., Eur. Phys. J C77 (2017) 367

NNLO QCD calculations ‘standard’

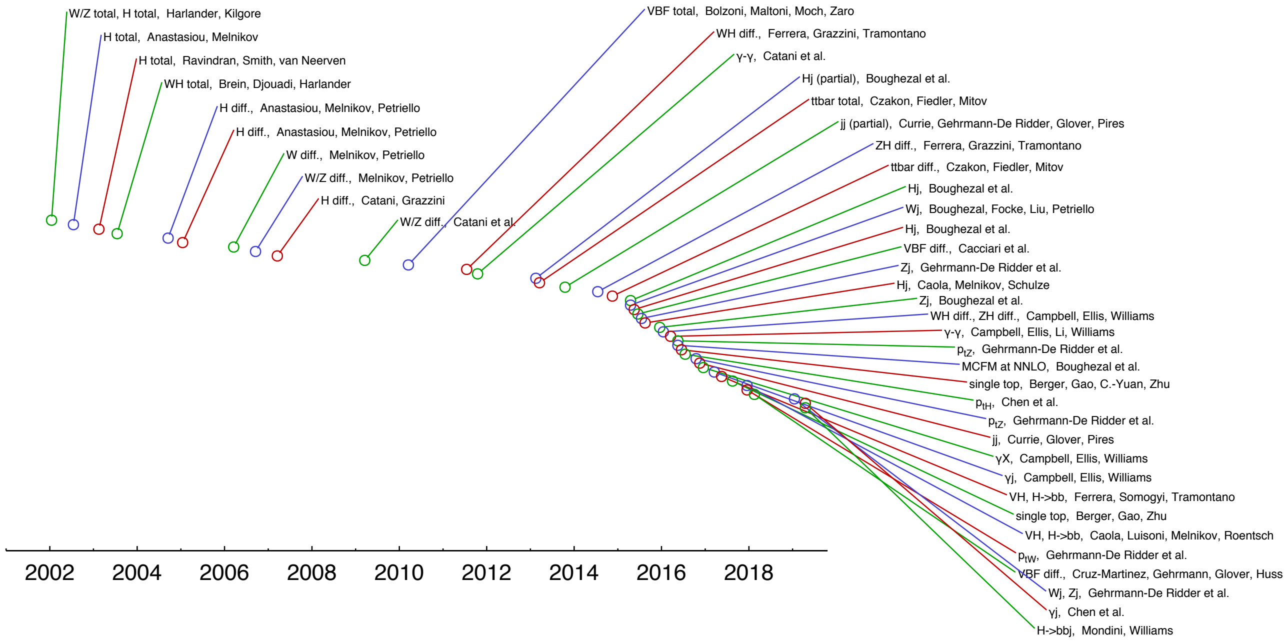
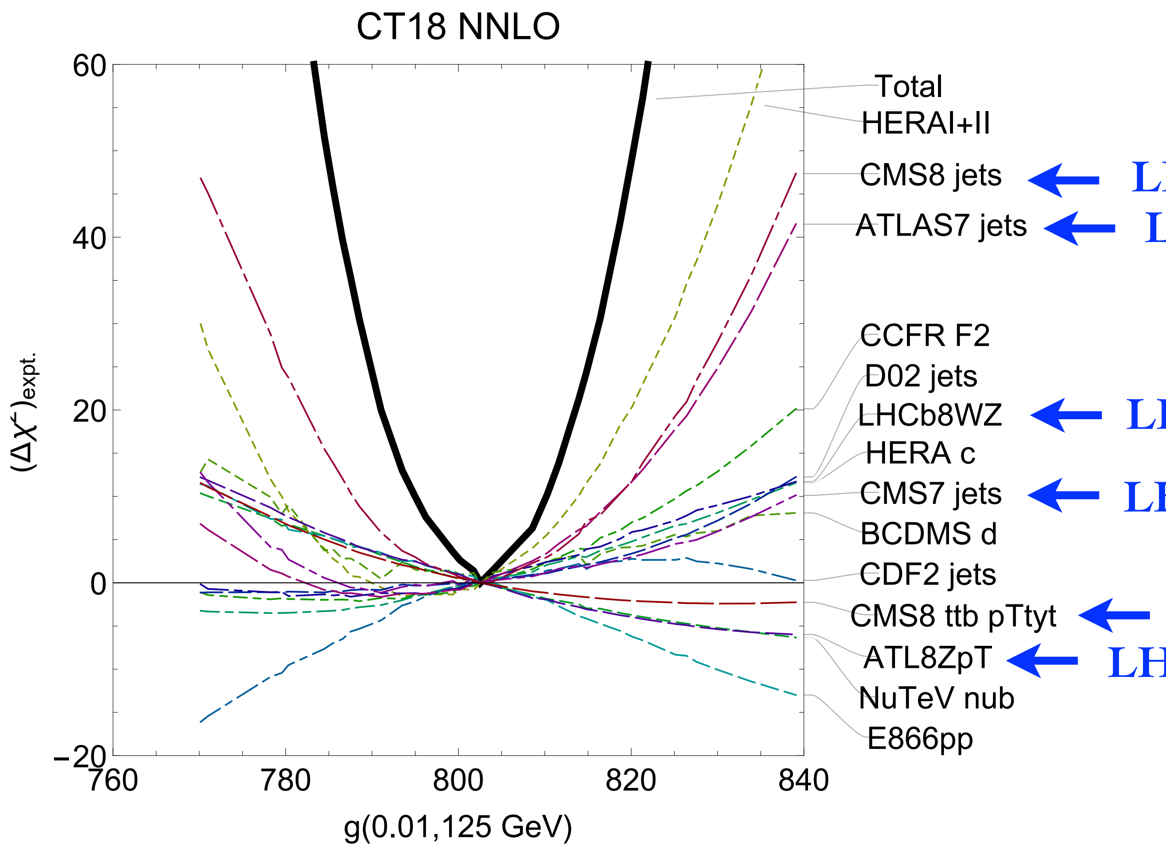
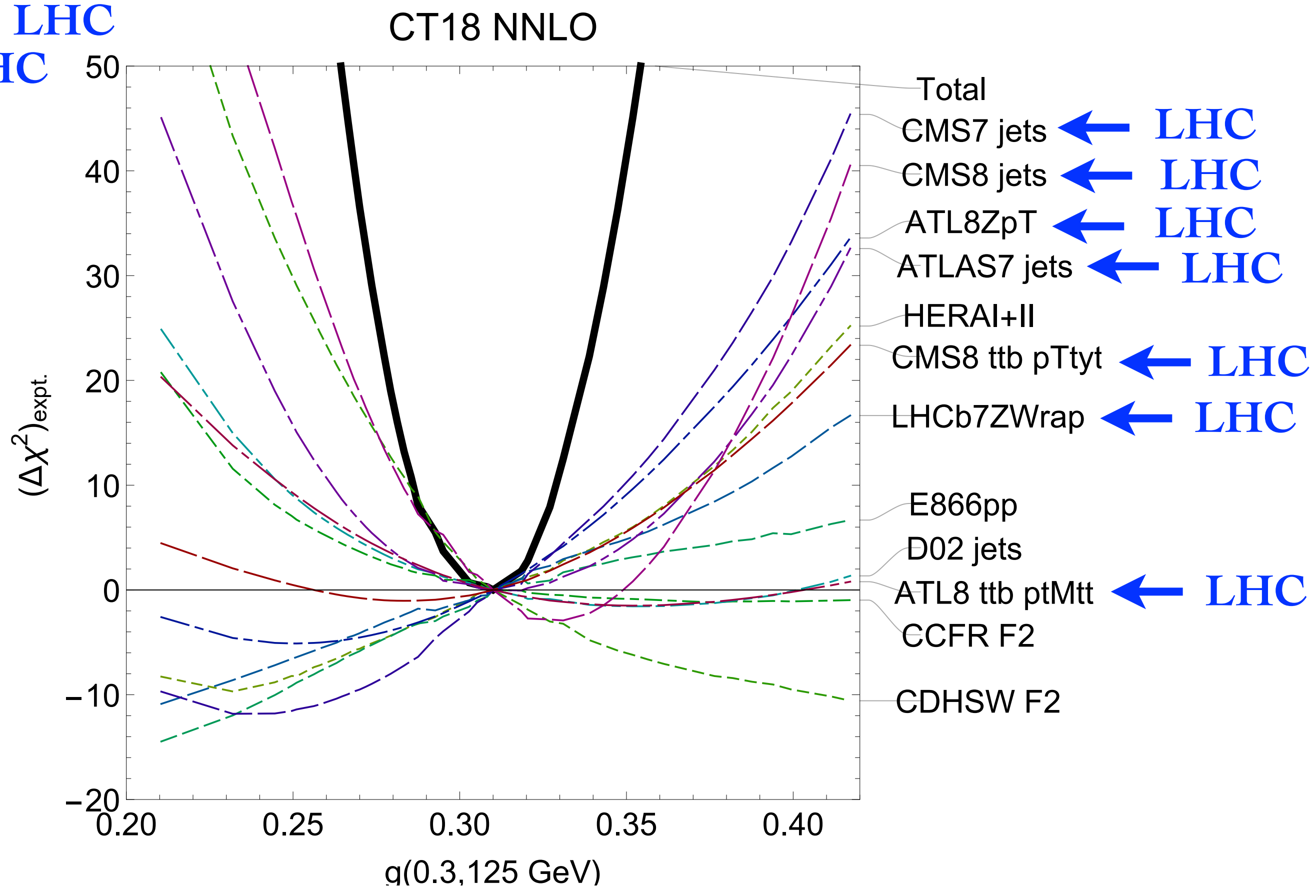


Image Credit: Gavin Salam

- **LHC data** now playing a key role in all fits.

- Example from recent **CT18** fit. Lagrange multiplier scans determining constraints on gluon at different χ^2values:

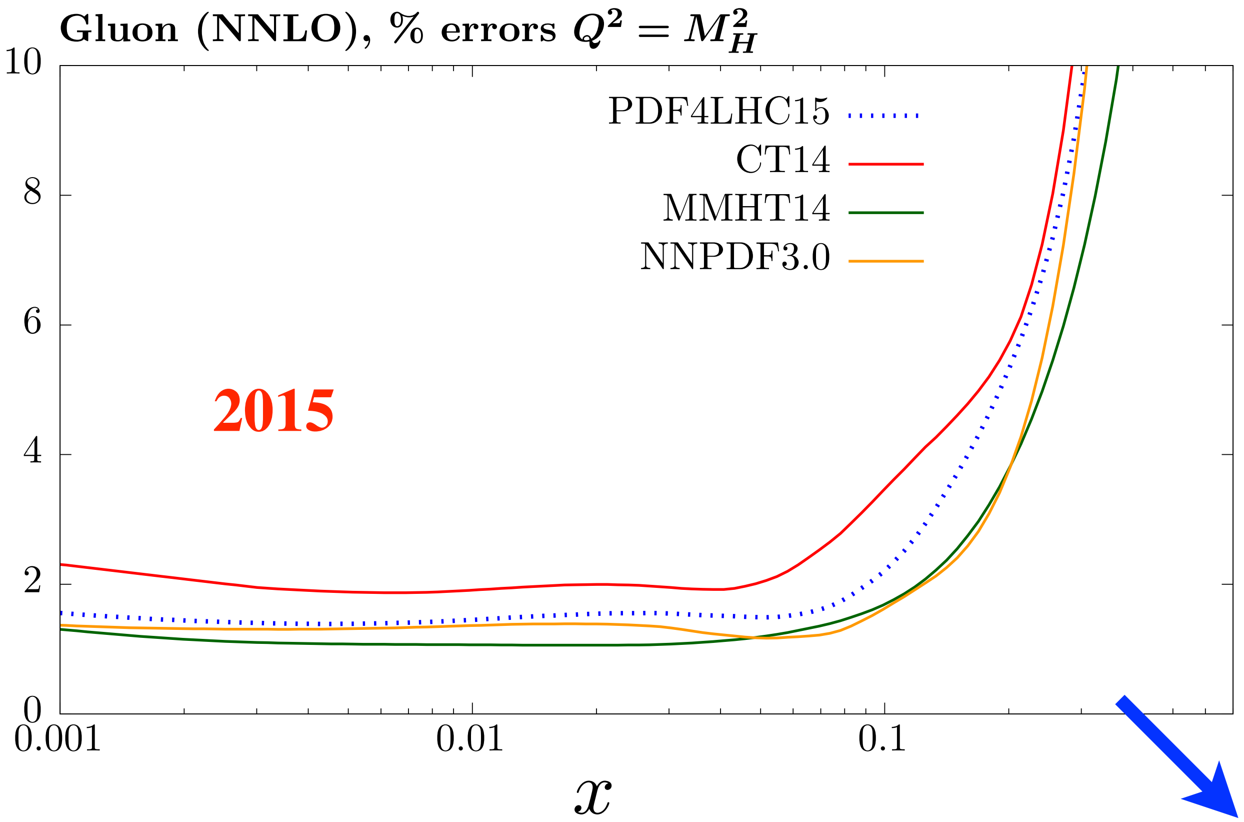




- Plenty of LHC data driving fits!

T-J Hou et al., arXiv:1908.11238

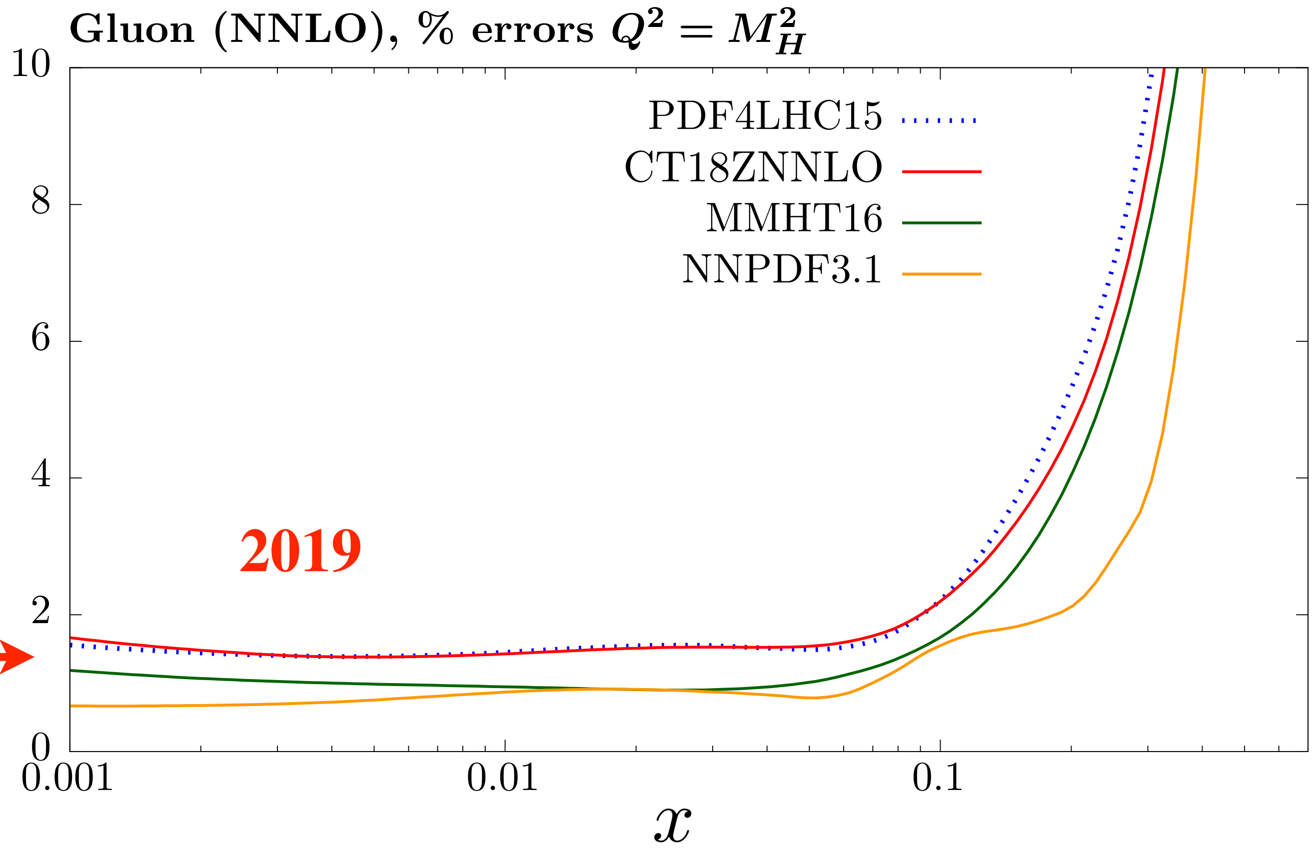
PDF Errors Today



- Consider e.g. **gluon PDF**.
- Clear reduction in **individual errors** in fits from 2019 vs. 2015.

- Driven by new **LHC data**.

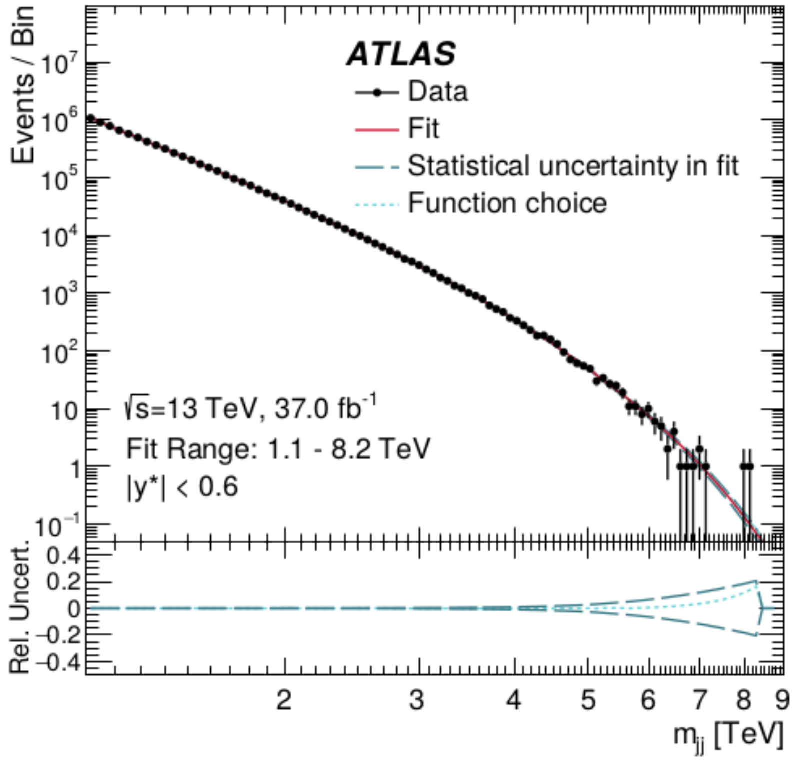
% level uncertainty



Opportunities

Example 1 - The Gluon

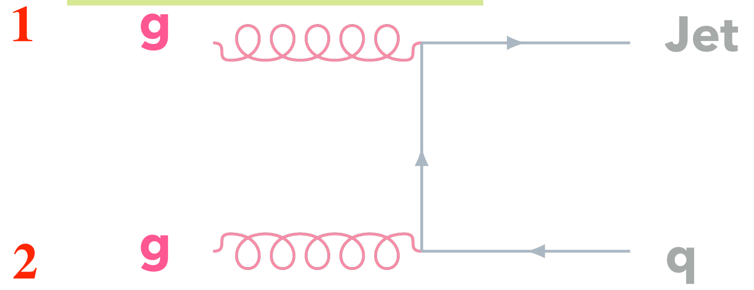
- Gluon at high x is both important for **BSM searches** and quite **poorly constrained** from DIS.
- LHC data such plays crucial role in constraining this.



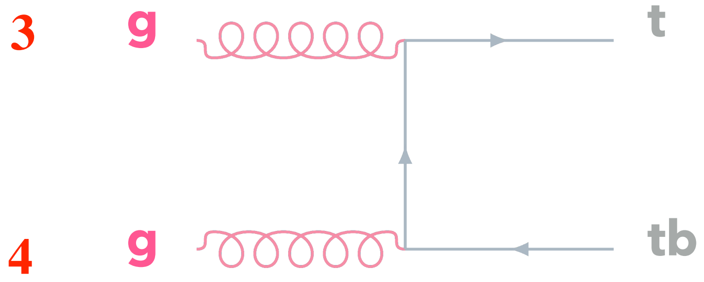
- Generically achieved by looking for gluon-initiated processes at high system transverse momentum/invariant mass/rapidity.
- Three textbook candidates at LHC:

$$x_{1,2} = \frac{p_\perp}{\sqrt{s}}(e^{\pm y_3} + e^{\pm y_4})$$

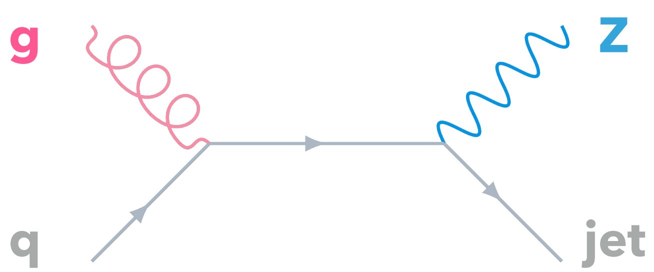
INCLUSIVE JETS



TOP PAIR

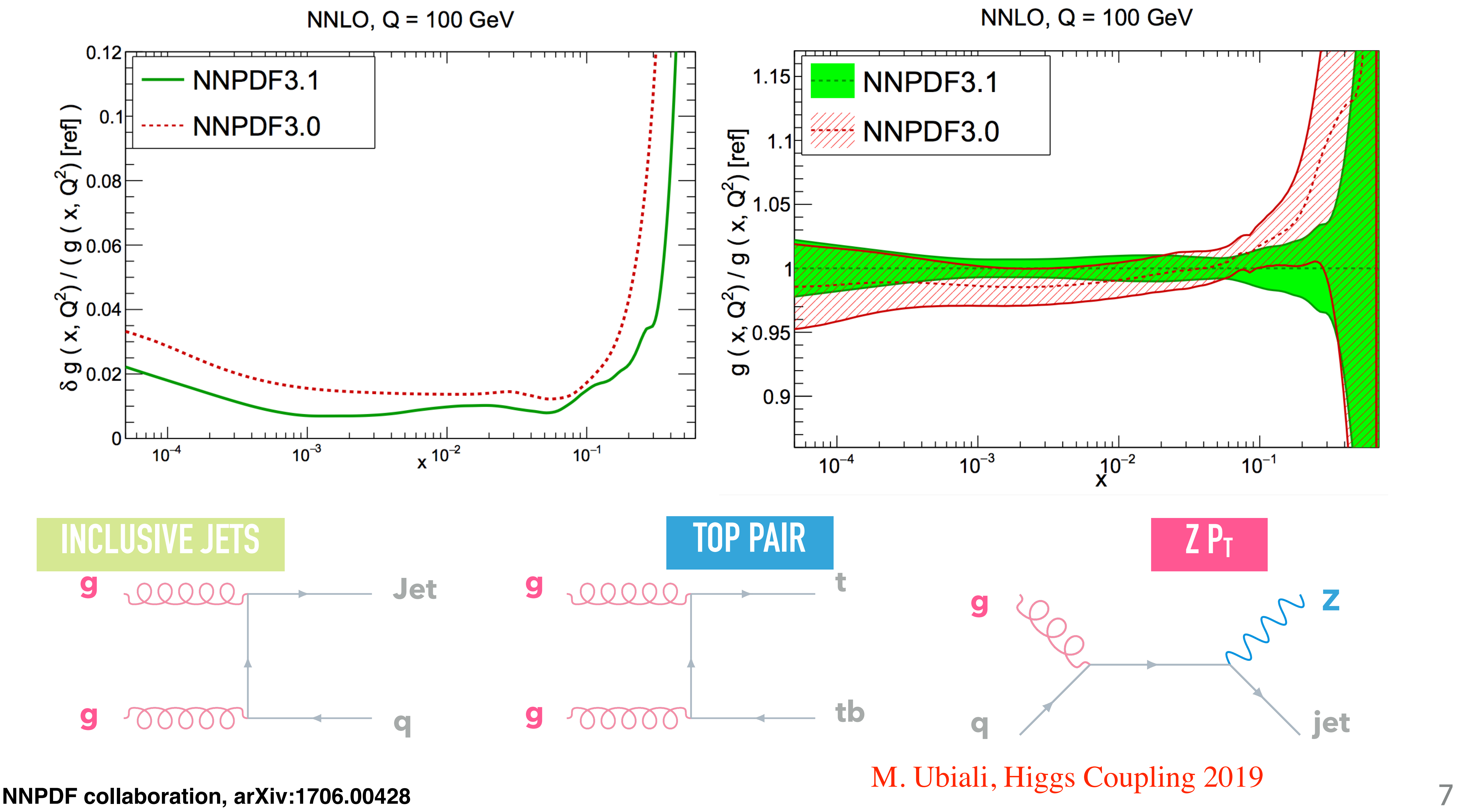


Z P_T



M. Ubiali, Higgs Coupling 2019

Example 1 - The Gluon



NNPDF collaboration, arXiv:1706.00428

M. Ubiali, Higgs Coupling 2019

- Impact of most recent LHC data (red → green) **significant**, with percent level uncertainties across wide range of x .

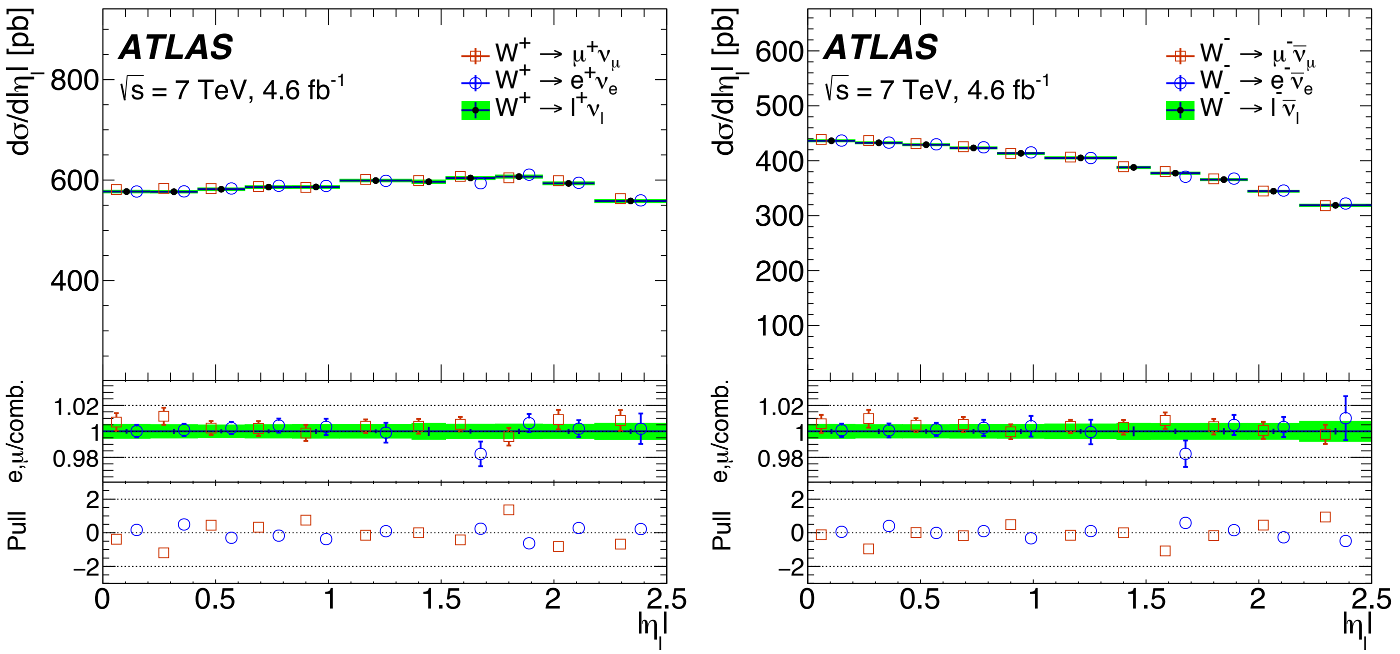
Example 2 - Proton Strangeness

- Vector boson ($W,\ Z$) production proceeds via range of channels.

$$u\bar{d},\ c\bar{s} \quad (u\bar{s},\ c\bar{d}) \to W^+\ ,$$
$$d\bar{u},\ s\bar{c} \quad (s\bar{u},\ d\bar{c}) \to W^-\ ,$$
$$q\bar{q} \to Z/\gamma^*\ ,$$

- Least constrained involves initial state $s,\ \bar{s}$ (no valence s) $\longrightarrow$ sensitive to **proton strangeness**.
- Only in principle: small contribution, requires **precise data** to pin down.

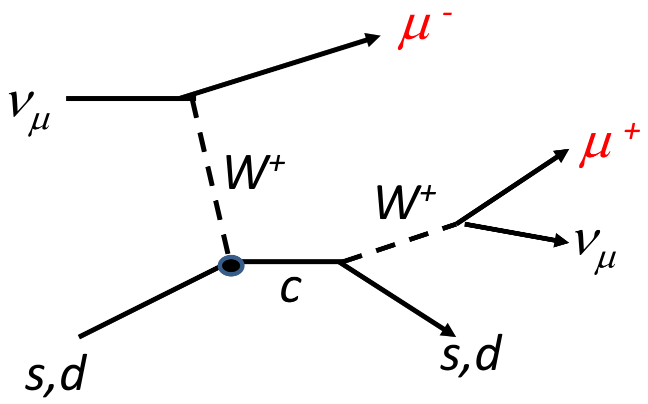
- Now available - highest ever precision measurement of W, Z production by **ATLAS.**
- Data uncertainties at the sub-% level. Statistical errors negligible completely dominated by systematics.



ATLAS collab., Eur. Phys. J C77 (2017) 367

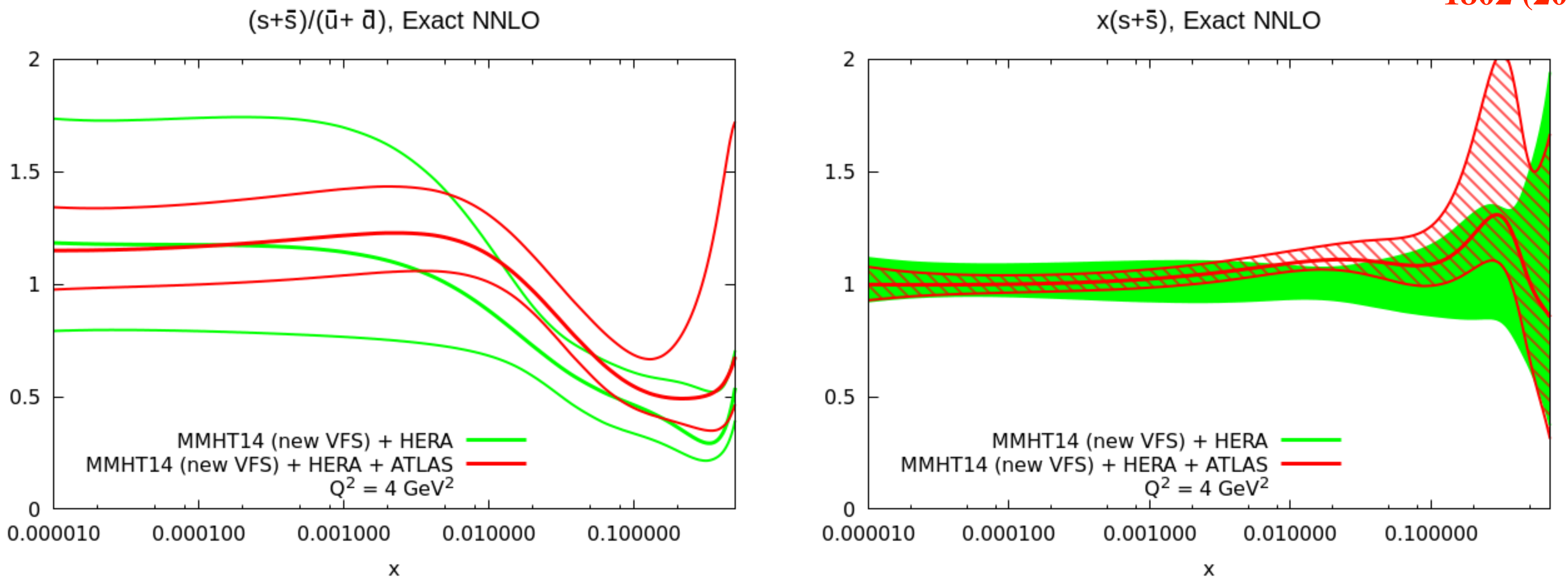
Example 2 - Proton Strangeness

- Impact of ATLAS data significant. Most notably: prefers **larger** strangeness than global fits, where previous constraints from neutrino-induced DIS ($\overline{\nu} s \to lc$).



- However global fits can safely accommodate both (rather distinct) datasets. Key ingredient: new **NNLO** calculation of DIS process.

J. Gao, JHEP 1802 (2018) 026



R. Thorne, DIS19

Summary: LHC Data For PDF Fits

- **Wide range** of LHC data available and included in/to be included in PDF fits. No time to cover everything, but in summary…

Single, double, triple differential

+ jets, W/Z $p_\perp$ distribution

Electroweak Boson Production

Low/High Mass, Central/Forward

+ charm/bottom

Inclusive Jet

Jets and Photons

Dijets (double, triple differential)

Isolated Photon

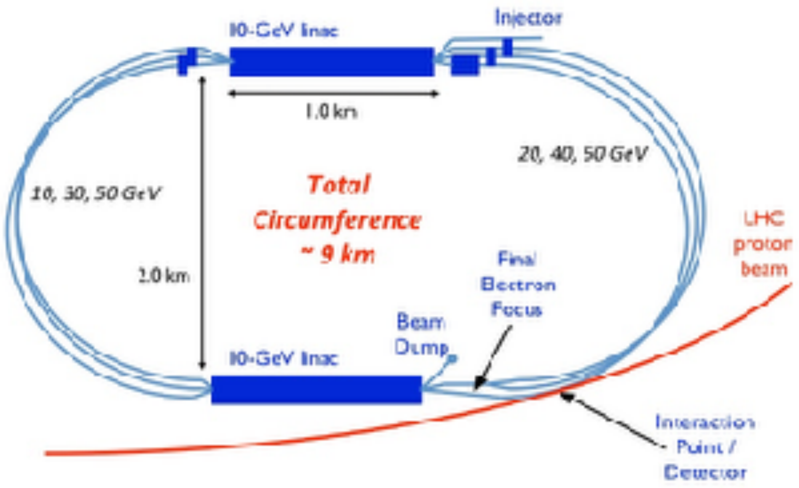
Pair production: total, single, double differential

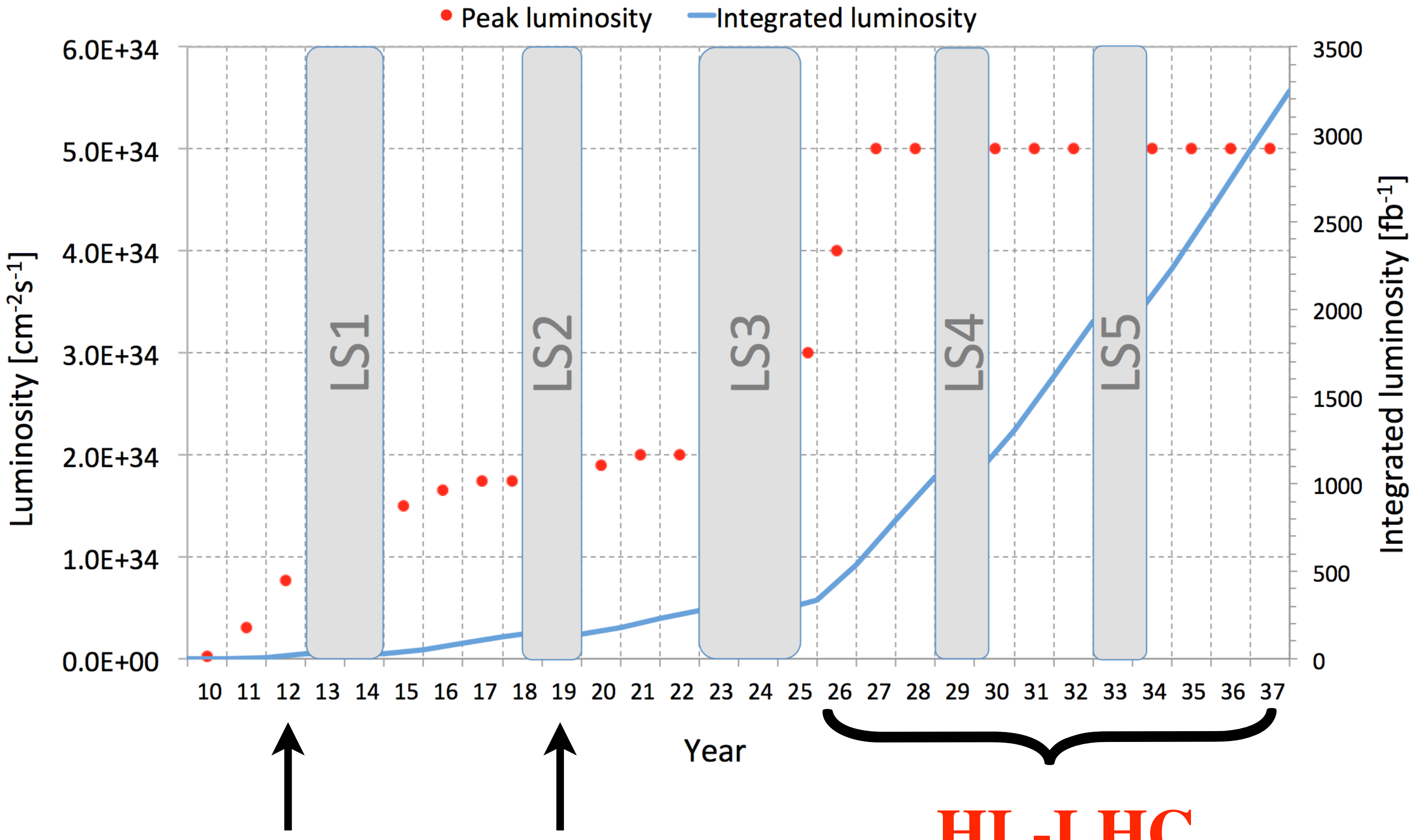
Top Quark

Single top

LHC: The Future

- At very **early stage** in LHC: so far only a few percent of the final projected data sample to be collected during High Luminosity **(HL)-LHC** running.





Higgs

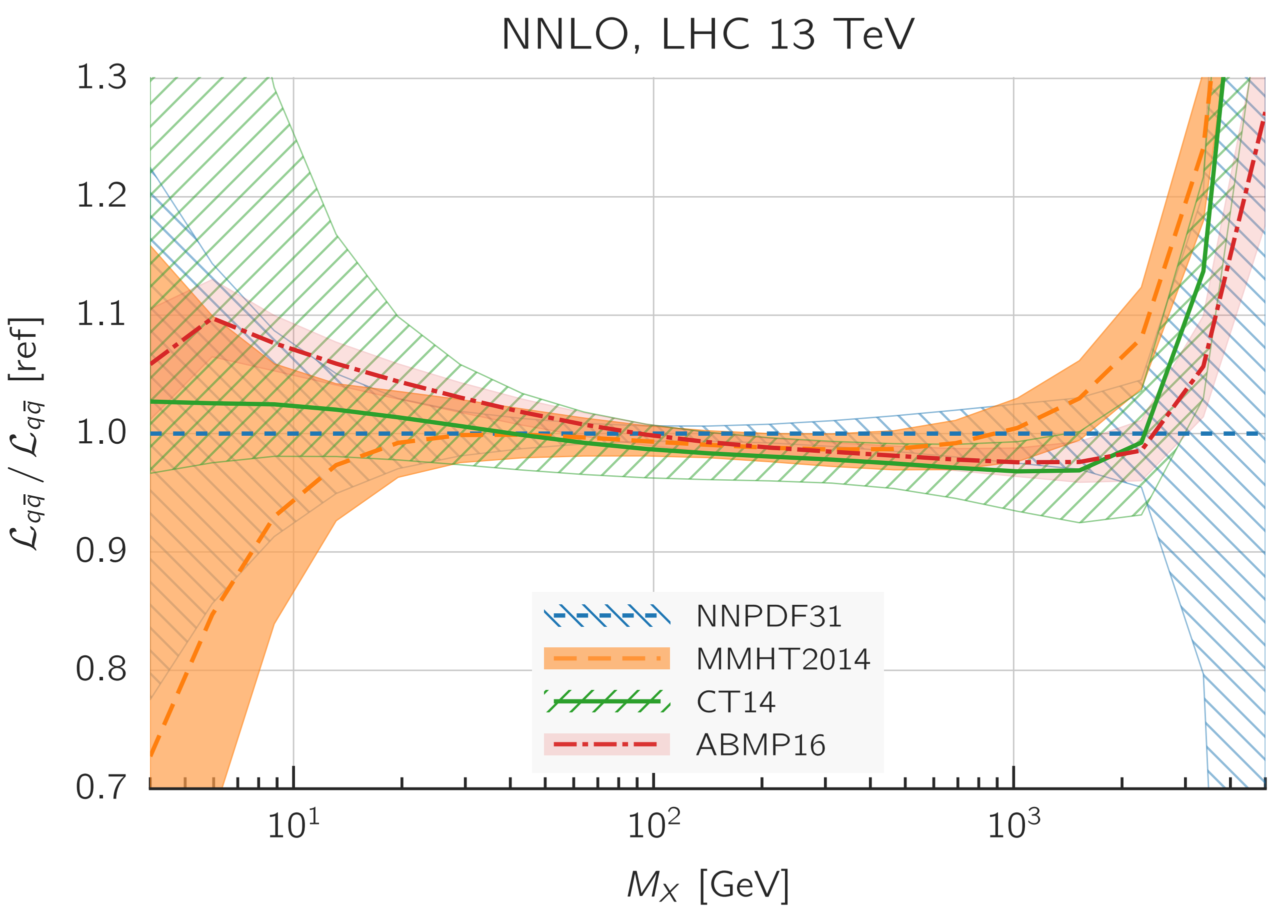
We are here

HL-LHC
(LHeC?)

- In addition exciting upgrade possibility of Large Hadron Electron Collider (**LHeC**): colliding lepton beam with LHC protons. Providing unprecedented high precision DIS data on proton structure.

Ultimate PDFs - Motivation

- Both HL-LHC and LHeC (if approved) will provide a vast range of data with a direct impact on the PDFs.
- **Question:** what exactly can we expect that impact to be?
- Collaborative effort to produce ‘**Ultimate**’ PDF set.



?

R. Abdul Khalek, S. Bailey, J. Gao, LHL, J. Rojo. Eur.Phys.J. C78 (2018) no.11, 962 & SciPost Phys. 7, 051 (2019)

Basic Idea

Produce theory predictions for relevant processes, in kinematic region probed by HL-LHC and LHeC

↓

Produce pseudodata - binned predictions, provided with corresponding statistical + systematic errors

↓

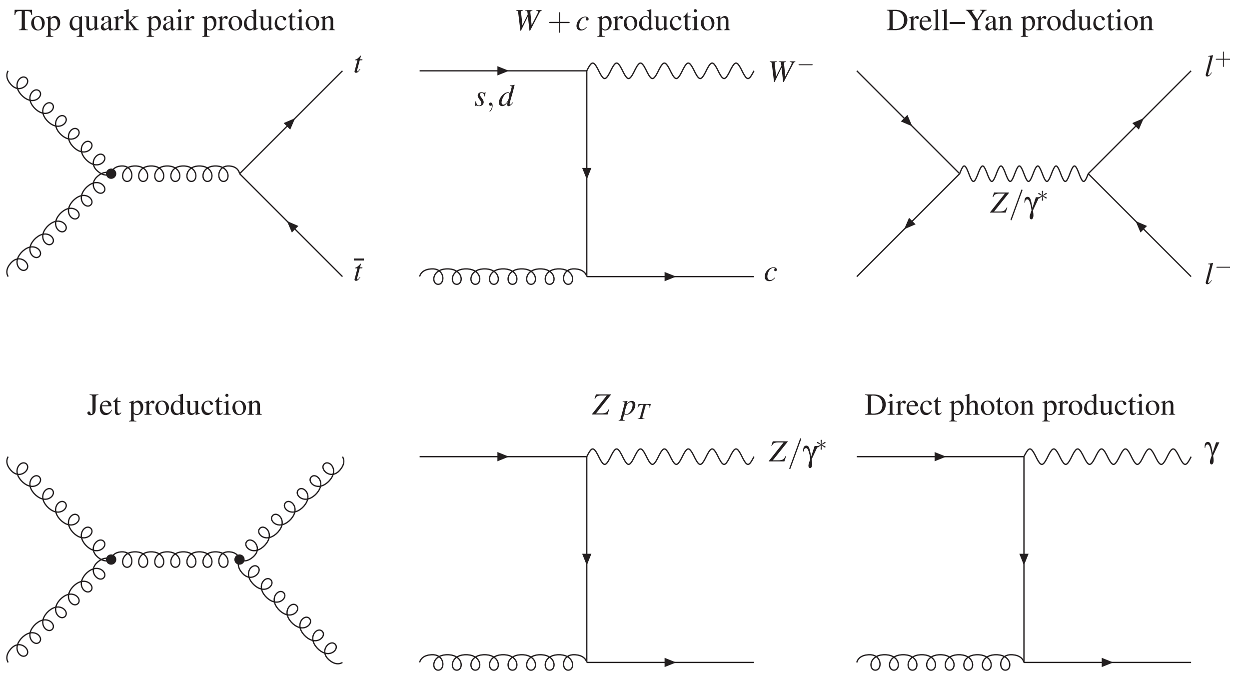
Perform PDF fit to this pseudodata

↓

Evaluate impact on PDF uncertainties

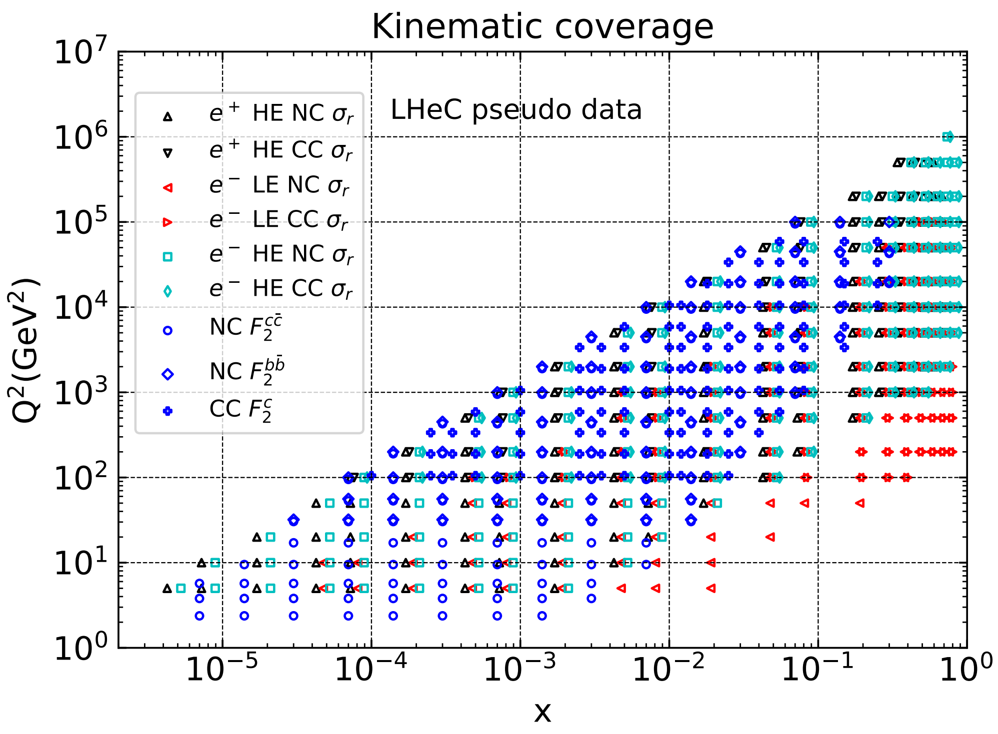
- **HL-LHC** processes: emphasis on high x region + measurements not already limited by systematic uncertainties.
- **LHeC**: range of DIS processes, according to official projections.
- Improvement in **statistical** uncertainties: straightforward extrapolation.
- Improvement in **systematics**: for HL-LHC take conservative/aggressive scenarios (little dependence in the end). For LHeC take official projections.
- Generate pseudo-data for these using **PDF4LHC** set: demonstrate expected improvement w.r.t. ~ current state of the art fits.

R. Abdul Khalek, S. Bailey, J. Gao, LHL, J. Rojo. Eur.Phys.J. C78 (2018) no.11, 962

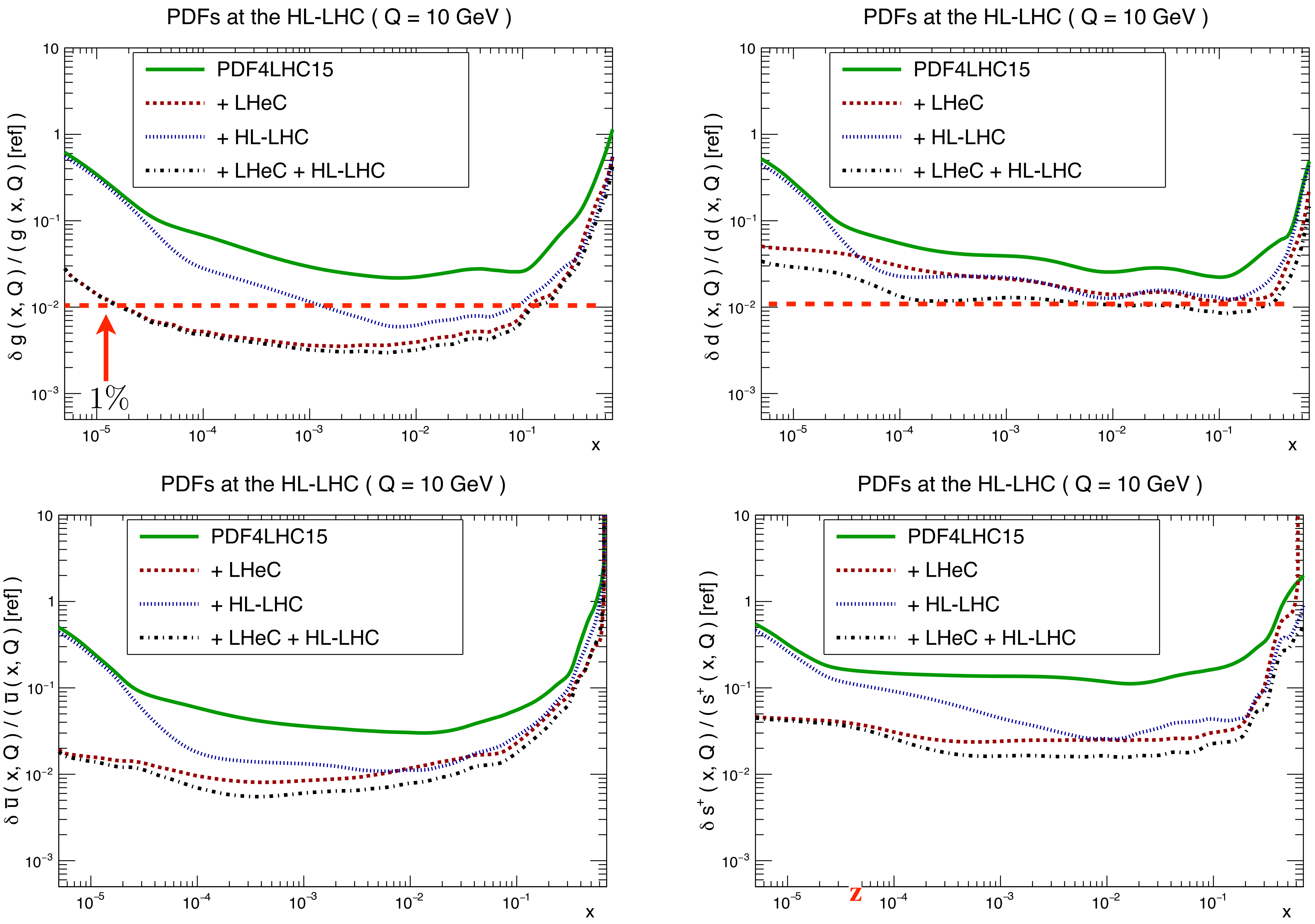


HL-LHC

R. Abdul Khalek, S. Bailey, J. Gao, LHL, J. Rojo. SciPost Phys. 7, 051 (2019)

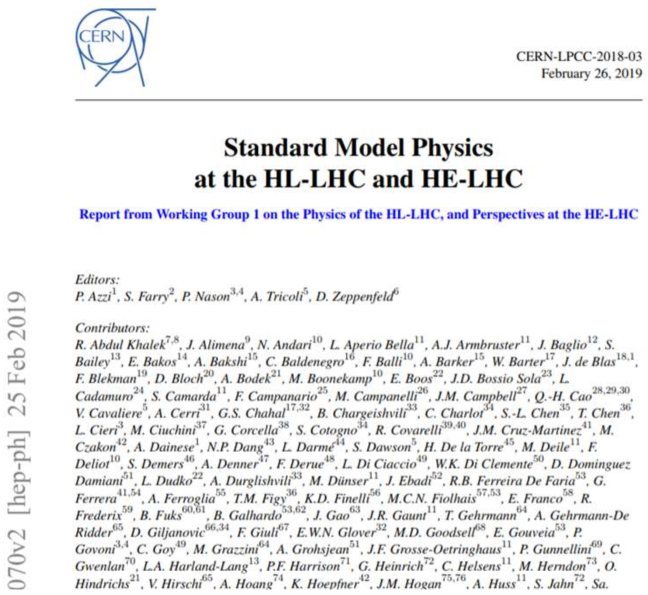


LHeC



- **Sub percent level** uncertainty in e.g. gluon in some x regions. Impressive constraints out to rather high x in general.
- LHeC placing very clean constraints across x range.

- PDFs used in corresponding HL-LHC + HE-LHC **yellow report** document.
- Clear reduction in PDF uncertainty for e.g. W mass and dijet measurements.



070v2 [hep-ph] 25 Feb 2019

CERN-LPCC-2018-03
February 26, 2019

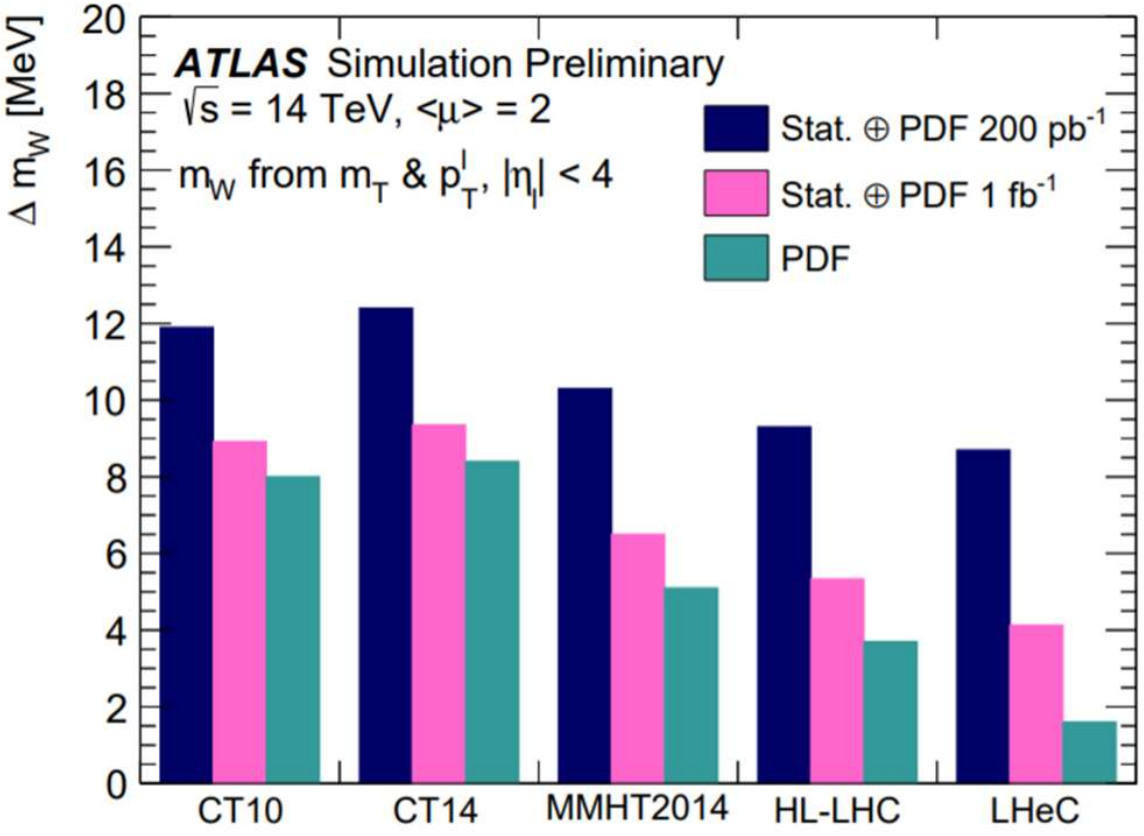
Standard Model Physics
at the HL-LHC and HE-LHC

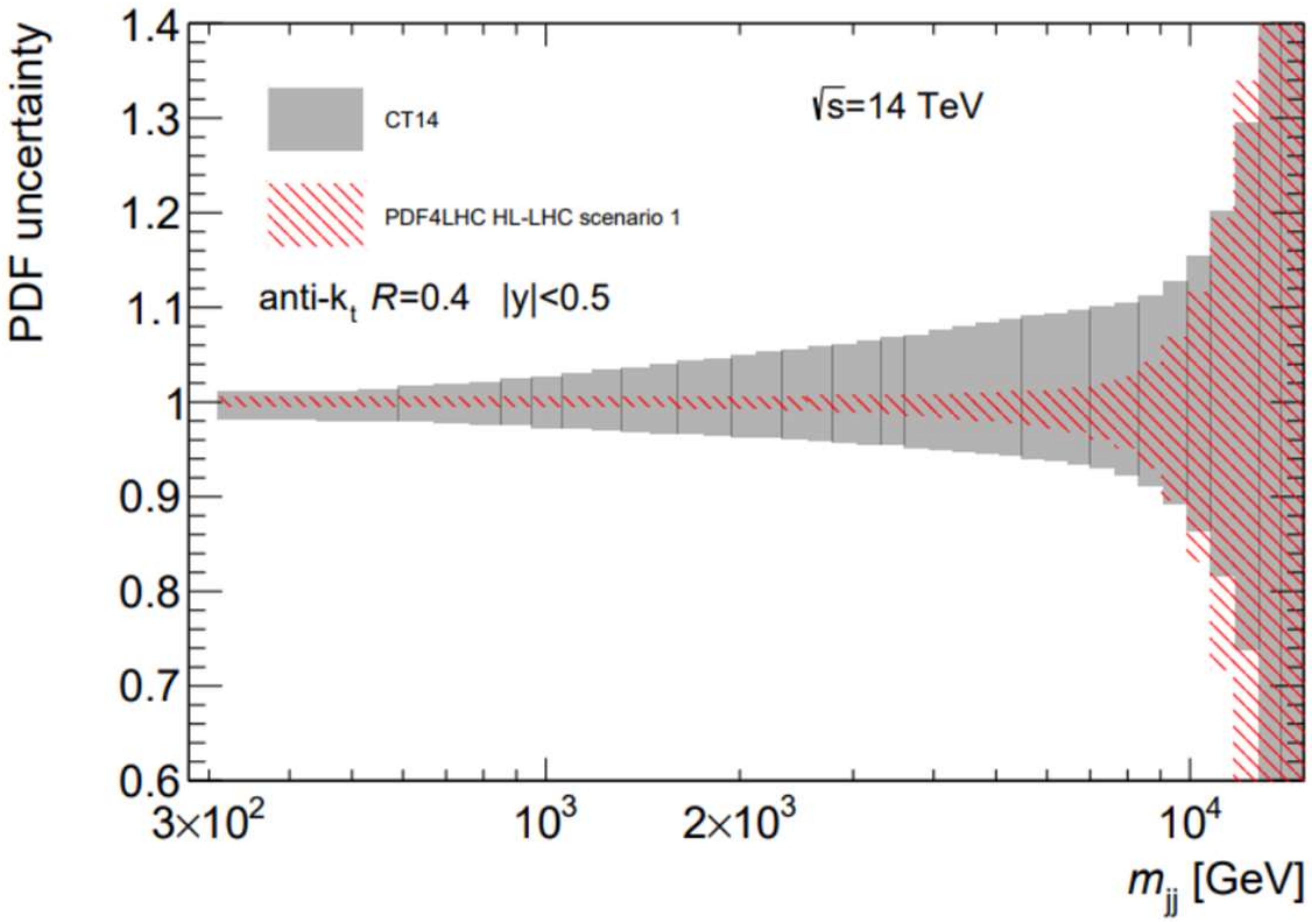
Report from Working Group 1 on the Physics of the HL-LHC, and Perspectives at the HE-LHC

Editors:
P. Azzi[1], S. Farry[2], P. Nason[3,4], A. Tricoli[5], D. Zeppenfeld[6]

Contributors:
R. Abdul Khalek[7,8], J. Alimena[9], N. Andari[10], L. Aperio Bella[11], A.J. Armbruster[11], J. Baglio[12], S. Bailey[13], E. Bakos[14], A. Bakshi[15], C. Baldenegro[16], F. Balli[10], A. Barker[15], W. Barter[17], J. de Blas[18,1], F. Blekman[19], D. Bloch[20], A. Bodek[21], M. Boonekamp[10], E. Boos[22], J.D. Bossio Sola[23], L. Cadamuro[24], S. Camarda[11], F. Campanario[25], M. Campanelli[26], J.M. Campbell[27], Q.-H. Cao[28,29,30], V. Cavaliere[5], A. Cerri[31], G.S. Chahal[17,32], B. Chargeishvili[33], C. Charlot[34], S.-L. Chen[35], T. Chen[36], L. Cieri[3], M. Ciuchini[37], G. Corcella[38], S. Cotogno[34], R. Covarelli[39,40], J.M. Cruz-Martinez[41], M. Czakon[42], A. Dainese[1], N.P. Dang[43], L. Darmé[44], S. Dawson[5], H. De la Torre[45], M. Deile[11], F. Deliot[10], S. Demers[46], A. Denner[47], F. Derue[48], L. Di Ciaccio[49], W.K. Di Clemente[50], D. Dominguez Damiani[51], L. Dudko[22], A. Durglishvili[33], M. Dünser[11], J. Ebadi[52], R.B. Ferreira De Faria[53], G. Ferrera[41,54], A. Ferroglia[55], T.M. Figy[36], K.D. Finelli[56], M.C.N. Fiolhais[57,53], E. Franco[58], R. Frederix[59], B. Fuks[60,61], B. Galhardo[53,62], J. Gao[63], J.R. Gaunt[11], T. Gehrmann[64], A. Gehrmann-De Ridder[65], D. Giljanovic[66,34], F. Giuli[67], E.W.N. Glover[32], M.D. Goodsell[68], E. Gouveia[53], P. Govoni[3,4], C. Goy[49], M. Grazzini[64], A. Grohsjean[51], J.F. Grosse-Oetringhaus[11], P. Gunnellini[69], C. Gwenlan[70], L.A. Harland-Lang[13], P.F. Harrison[71], G. Heinrich[72], C. Helsens[11], M. Herndon[73], O. Hindrichs[21], V. Hirschi[65], A. Hoang[74], K. Hoepfner[42], J.M. Hogan[75,76], A. Huss[11], S. Jahn[72], Sa.

CERN Yellow Rep. Monogr. 7 (2019) 1-220

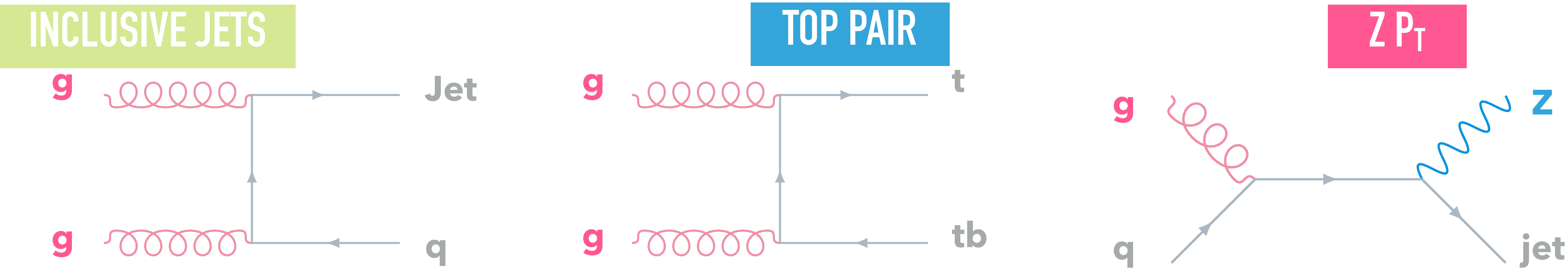




Challenges

Confronting Precise Data

- **Good news**: LHC has already had significant impact on PDFs and HL-LHC has potential to improve on this even further.
- However in a number of cases we are seeing **difficulty** in confronting such high precision data in PDF fits.
- Occurs in three ‘textbook’ LHC processes for PDF determination:



M. Ubiali, Higgs Coupling 2019

- Will consider one case (ATLAS 8 TeV $t\bar{t}$ production) in detail, but one of many datasets where significant issues seen: CMS top (single/double differential), ATLAS/CMS $Z\,p_\perp$, ATLAS isolated photon….

What Drives Fit Quality?

- The χ^2 can in presence of correlated errors can be written as:

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

$$D_k \to D_k - \sum_{\alpha=1}^{N_\lambda}\beta_{k,\alpha}\hat{\lambda}_\alpha$$

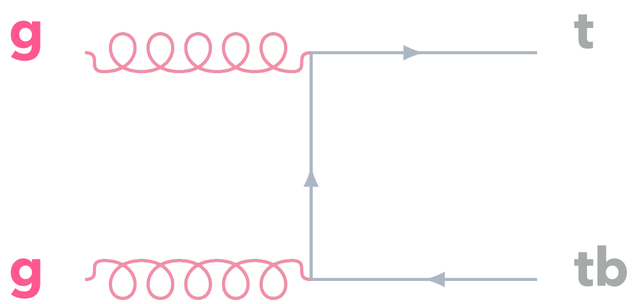
Uncorrelated errors **Correlated errors** **Penalty for shifts**

- The set of N_λ nuisance parameters λ_α take values so as to minimise χ^2, effectively shifting data points D_k.

- At LHC we are increasingly in the $s_k \to 0$ regime. Dominance of:
  - ★ **Experimental systematic** errors, $\beta_{k,\alpha}$.
  - ★ **Theoretical uncertainties** (in particular missing higher orders).

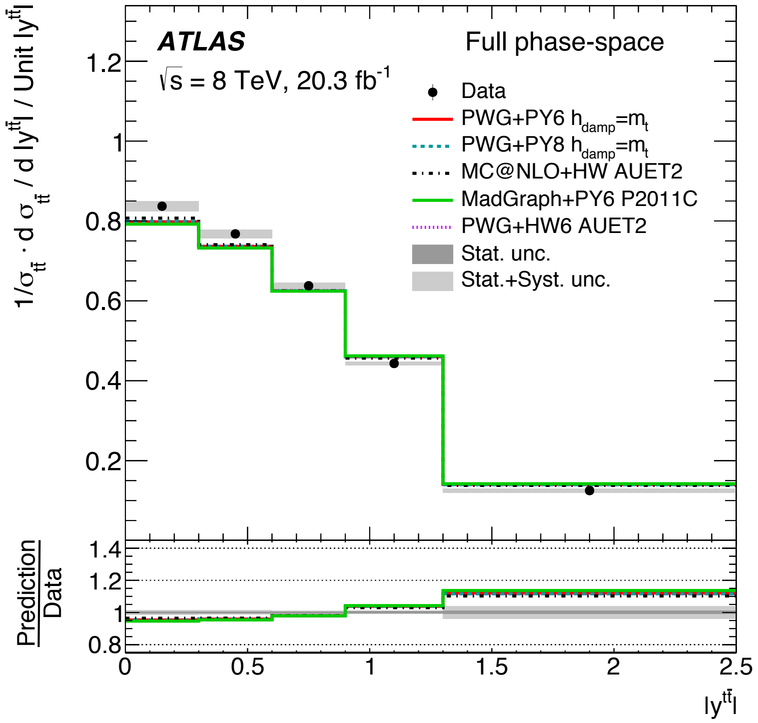
- In both cases strong sensitivity not just to the size of the errors but to their **correlation**. Complicates interpretation of fit quality greatly.

Top Quark Production



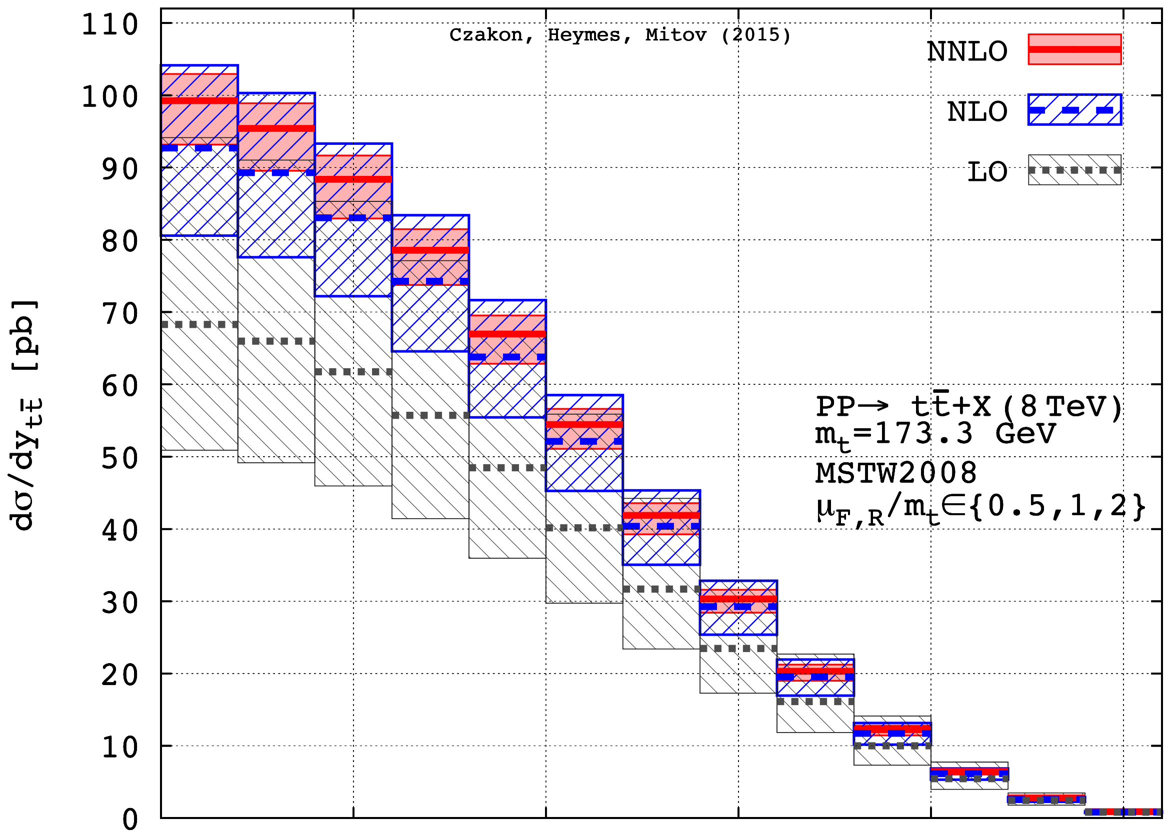
- **In principle** ideal candidate for precision PDF determination: parton-level theory known to **NNLO** in QCD, while precise data provided multi-differentially in various observables.

$$y_t,\ y_{t\bar{t}},\ p_\perp^t,\ M_{t\bar{t}}$$



ATL-PHYS-PUB-2018-017

Lepton + jet channel



M. Czakon, D. Heymes, A. Mitov, PRL 116 (2016) no.8, 082003

- Recent study: attempt to fit this dataset…

S. Bailey & LHL, arXiv:1909.10541

- ...find **terrible fit** quality! What is going on?

$$\chi^2/N_{\rm pts}\ (N_{\rm pts}^{\rm tot}=25)$$

p_T	0.53
y_t	3.12
y_{tt}	3.51
M_{tt}	0.70
$p_T + M_{tt}$	5.73
Combined	$\underline{7.00}$

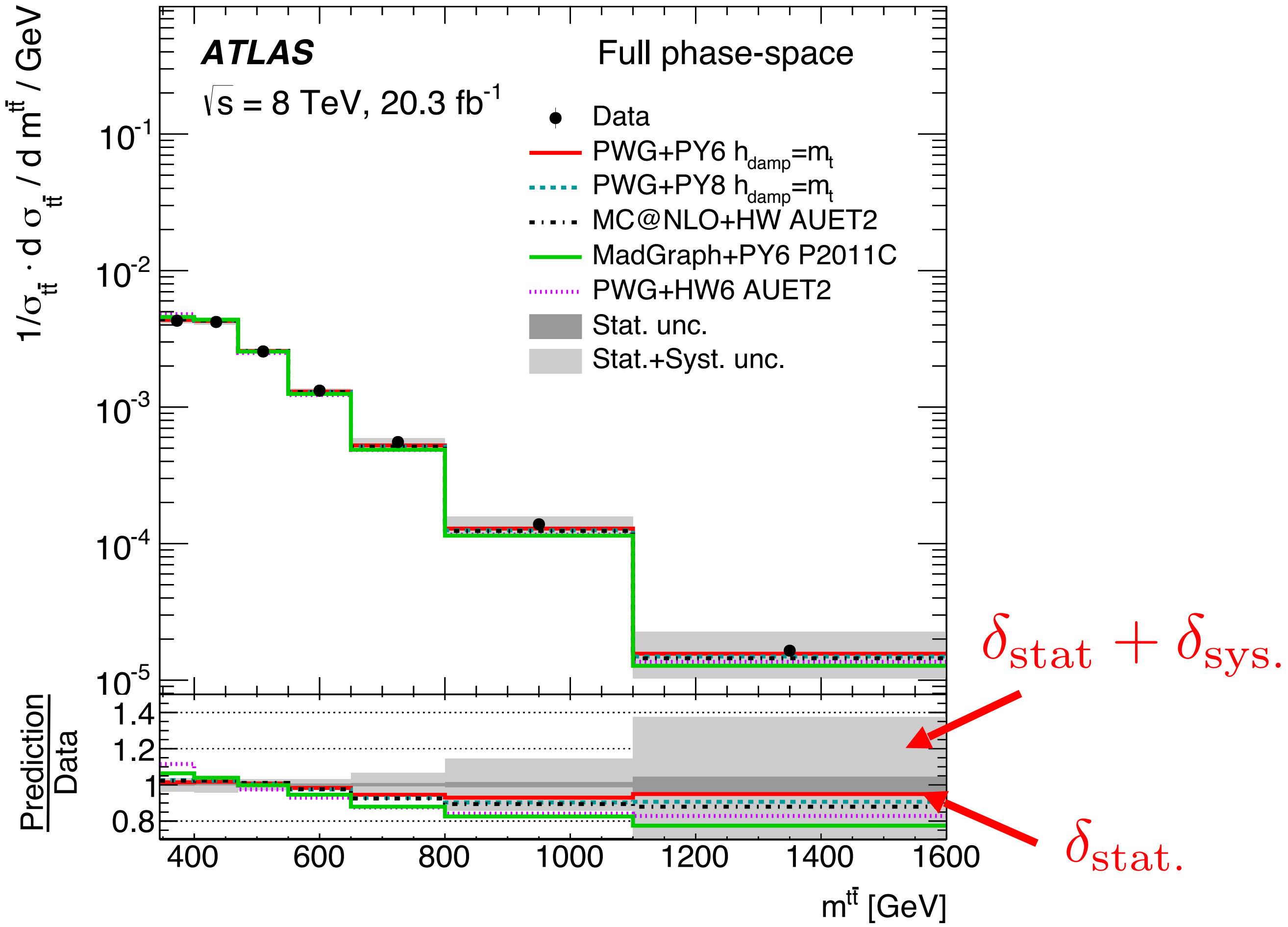
S. Bailey & LHL, arXiv:1909.10541

- **Clue**: look at relative size of statistical vs. systematic errors.

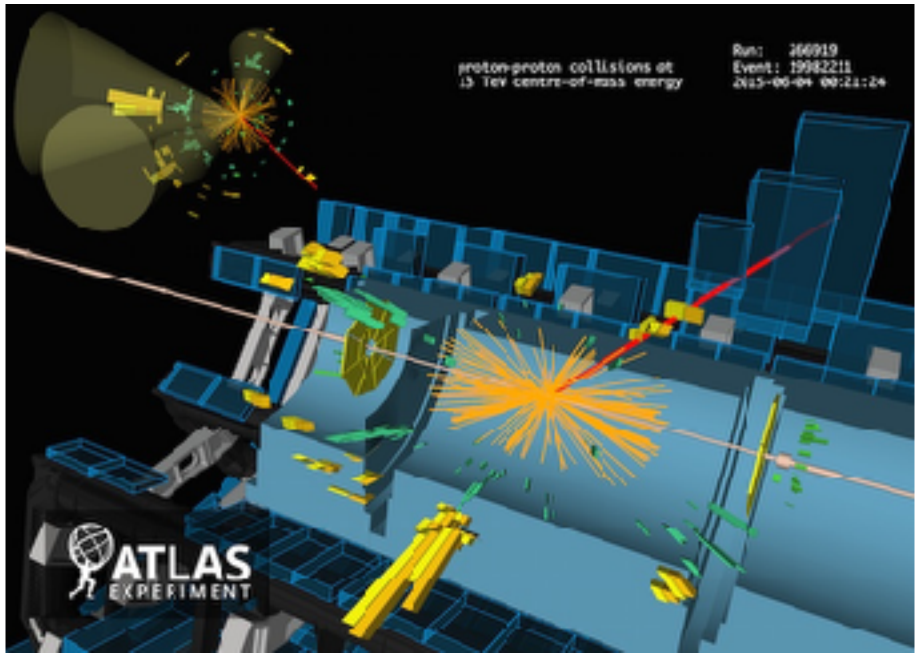
- Systematics completely **dominant**.

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

- These are in many cases highly correlated across y_t, $y_{t\bar{t}}$, $p_\perp^t$, $M_{t\bar{t}}$.

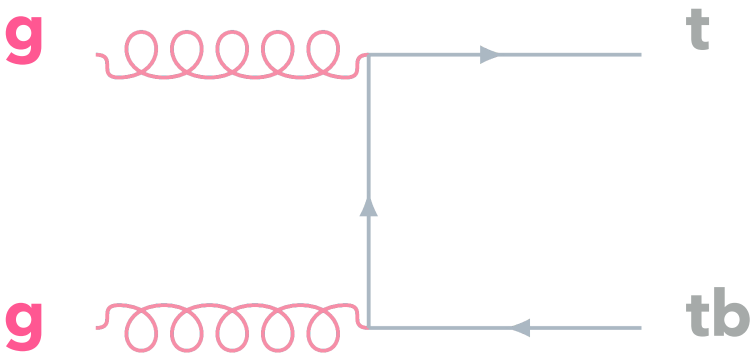


- Many sources of systematics, but by far the largest related to **unfolding** from detector back to top quark level:



Measurement

→

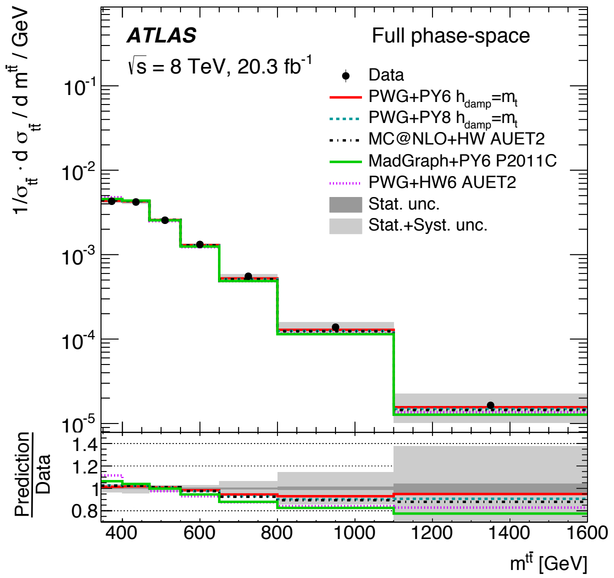


PDF Fit

- Requires understanding of top quark production/decay and subsequent showering/hadronization. All of this needs theory (**Monte Carlo**) input.
- If we get correction (→) wrong the top-quark level data will be wrong.
- Uncertainty due to this? Take event sample with second MC, apply correction (→) derived with baseline MC. Difference between this and true result gives uncertainty.

- These **two-point** MC uncertainties by far the largest:
  - ★ Parton Shower: `POWHEG` + `Herwig` vs. `POWHEG` + `Pythia`
  - ★ Hard Scattering: `MC@NLO` + `Herwig` vs. `POWHEG` + `Herwig`
  - ★ ISR/FSR: `POWHEG` + `Pythia`(1) vs. `POWHEG` + `Pythia`(2)

- Uncertainty **and** correlation effectively given by envelope of two MCs ~ reasonable as correction should be smooth between bins. But will not capture the full correlation ⇒ fit quality sensitive to it.



- Our study*: try some reasonable **loosening** of the assumed **correlation**:

$$\sum_{\alpha=1}^{N_\lambda} \beta_{k,\alpha} \to \sum_{\alpha=1}^{\tilde{N}_\lambda} \tilde{\beta}_{k,\alpha}$$

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

S. Bailey & LHL, arXiv:1909.10541

- Note input (clear breakdown of errors) from experiment **essential** here: which systematics can we do this to?
- Taking the above two-point MC errors and applying this **decorrelation**…

***Similar more limited study in ATL-PHYS-PUB-2018-017**

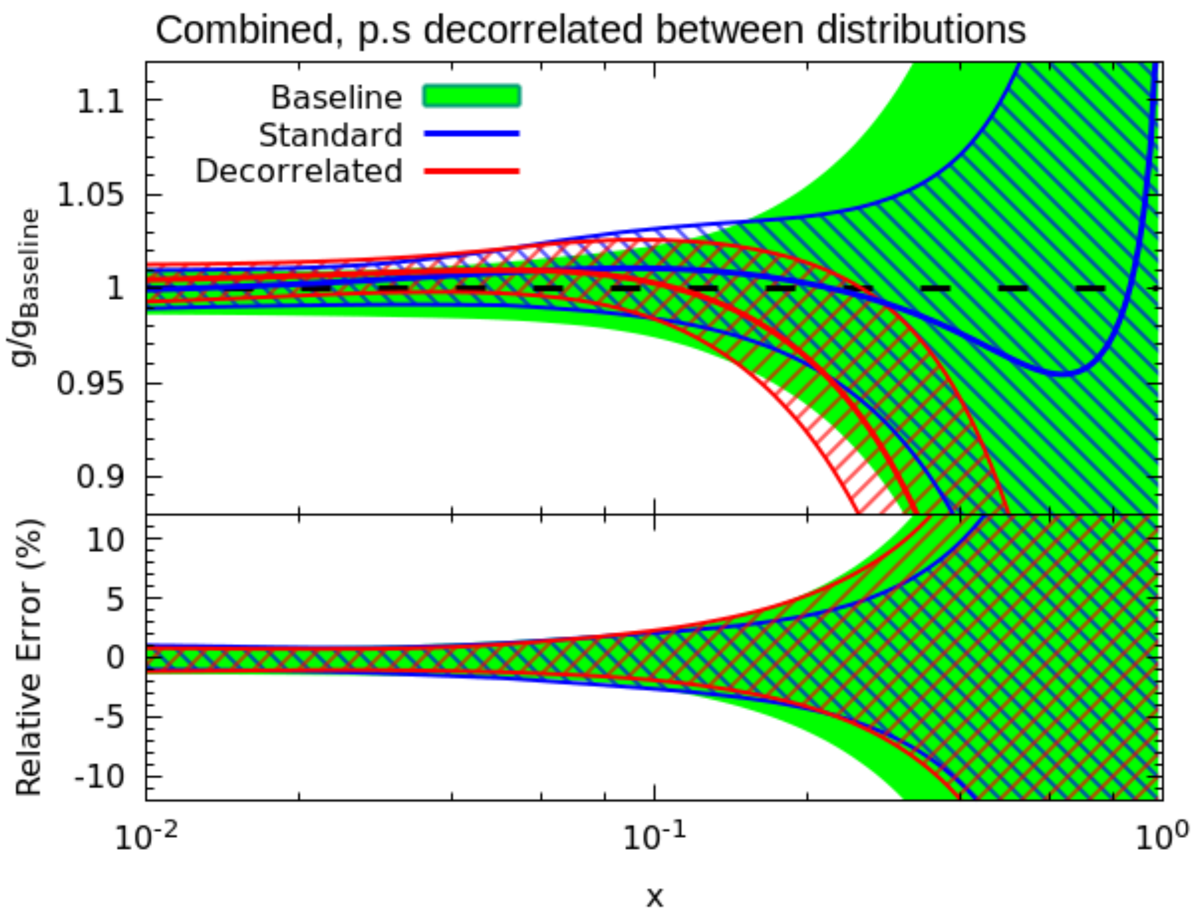
- Consider decorrelation of just one error source, e.g. '**parton shower**' error (other two similar).
- Gives **huge improvement** in fit quality!

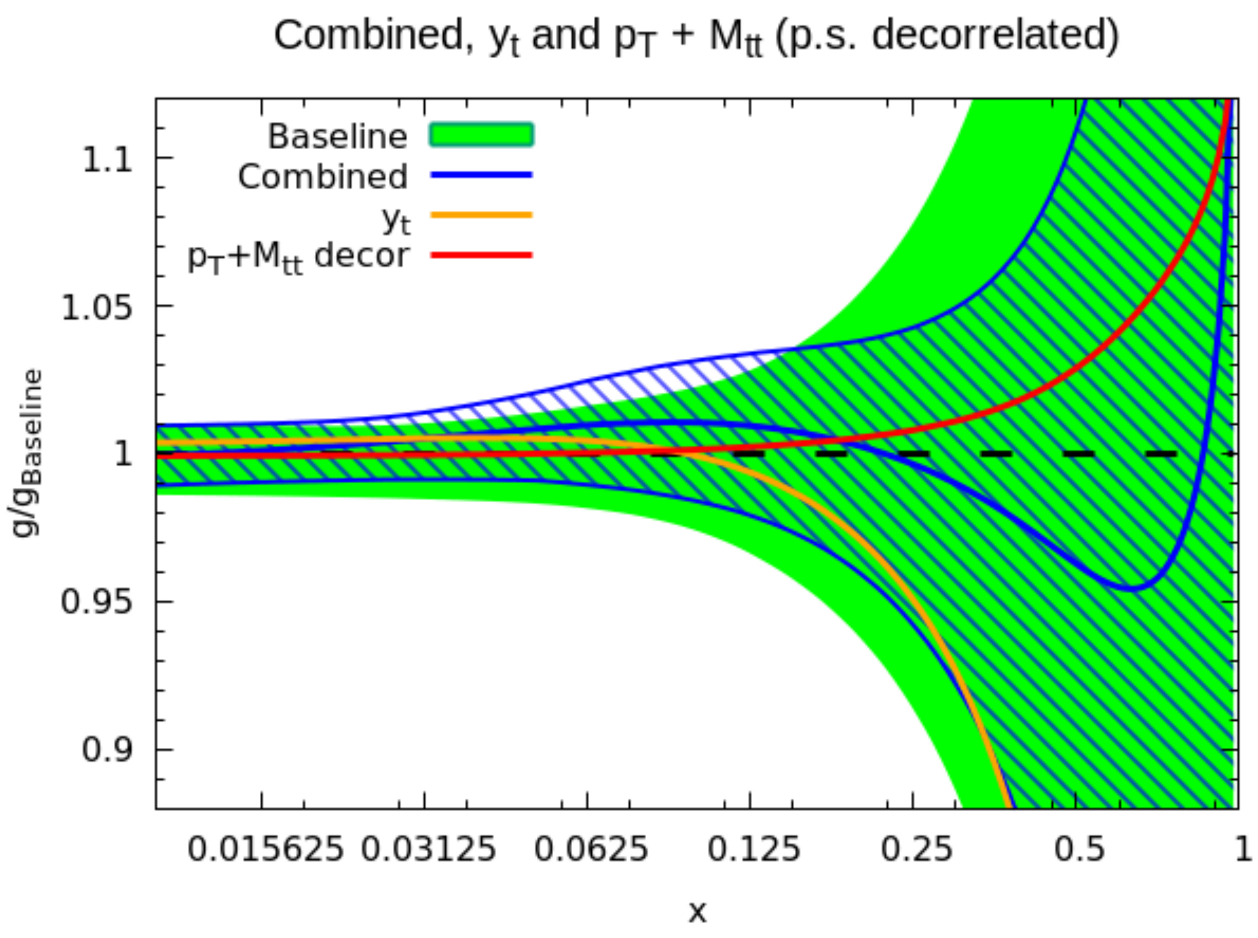
Allow independent variation across $y_t, y_{t\bar{t}}, p_\perp^t, M_{t\bar{t}}$ but keep correlation within.

$y_t, y_{t\bar{t}}, p_\perp^t, M_{t\bar{t}}$

Distribution	p.s. correlated	p.s. decorrelated
Combined	7.00	1.80
$p_\perp^t + M_{tt}$	5.73	0.66

- Has **significant** impact on gluon: larger than NLO vs. NNLO theory difference and not same as picking one distribution (e.g. y_t).

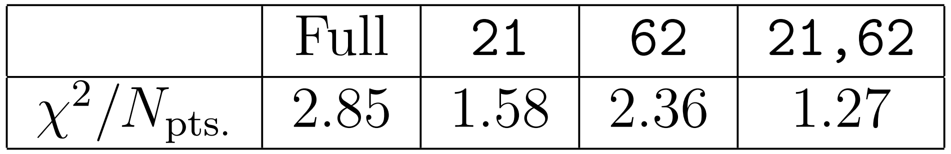

Combined, p.s decorrelated between distributions


Combined, y_t and p_T + M_{tt} (p.s. decorrelated)

ATLAS Jets

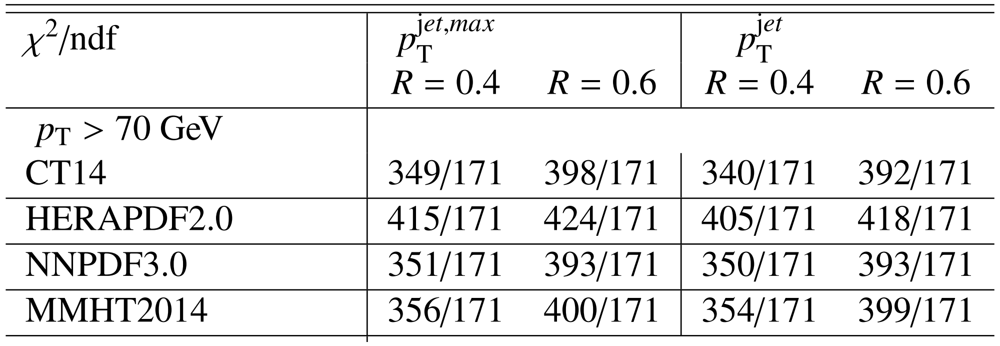
LHL, R.S. Thorne, A.D. Martin, EPJC78 (2018) no.3, 248

- ATLAS **Jet data**: again systematics dominated, and fit quality highly sensitive to correlations.
- Again identify systematics with potentially too strong assumption about correlations. **Decorrelating** again has **large impact** on fit quality.
- Detailed **ATLAS study**: identifies those error sources that can be decorrelated and by how much.
- However also find that including **theory uncertainties** from missing higher order in pQCD theory improves fit.



	Full	21	62	21,62
$\chi^2/N_{\rm pts.}$	2.85	1.58	2.36	1.27

ATLAS Collab., JHEP 09 (2017) 020



χ^2/ndf	$p_T^{jet,max}$		p_T^{jet}	
	$R = 0.4$	$R = 0.6$	$R = 0.4$	$R = 0.6$
$p_T > 70$ GeV				
CT14	349/171	398/171	340/171	392/171
HERAPDF2.0	415/171	424/171	405/171	418/171
NNPDF3.0	351/171	393/171	350/171	393/171
MMHT2014	356/171	400/171	354/171	399/171

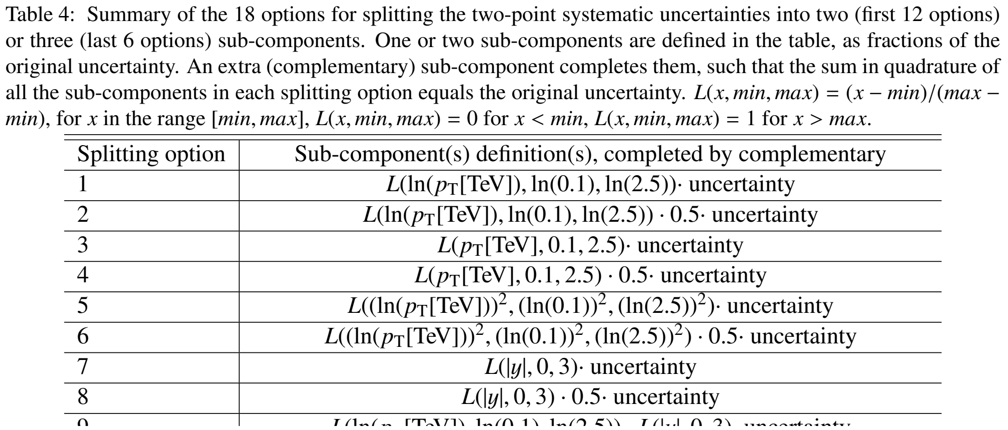


Table 4: Summary of the 18 options for splitting the two-point systematic uncertainties into two (first 12 options) or three (last 6 options) sub-components. One or two sub-components are defined in the table, as fractions of the original uncertainty. An extra (complementary) sub-component completes them, such that the sum in quadrature of all the sub-components in each splitting option equals the original uncertainty. $L(x, min, max) = (x - min)/(max - min)$, for x in the range $[min, max]$, $L(x, min, max) = 0$ for $x < min$, $L(x, min, max) = 1$ for $x > max$.

Splitting option	Sub-component(s) definition(s), completed by complementary
1	$L(\ln(p_T[\text{TeV}]), \ln(0.1), \ln(2.5))\cdot$ uncertainty
2	$L(\ln(p_T[\text{TeV}]), \ln(0.1), \ln(2.5)) \cdot 0.5\cdot$ uncertainty
3	$L(p_T[\text{TeV}], 0.1, 2.5)\cdot$ uncertainty
4	$L(p_T[\text{TeV}], 0.1, 2.5) \cdot 0.5\cdot$ uncertainty
5	$L((\ln(p_T[\text{TeV}]))^2, (\ln(0.1))^2, (\ln(2.5))^2)\cdot$ uncertainty
6	$L((\ln(p_T[\text{TeV}]))^2, (\ln(0.1))^2, (\ln(2.5))^2) \cdot 0.5\cdot$ uncertainty
7	$L(\lvert y\rvert, 0, 3)\cdot$ uncertainty
8	$L(\lvert y\rvert, 0, 3) \cdot 0.5\cdot$ uncertainty

Theory Uncertainties?

- Focus in previous discussion on experimental systematics, but not the end of the story.
- Consider fit quality again:

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

- Even if experimental systematics perfectly accounted for, in $s_k \to 0$ limit the theory T_k will not by default match the data D_k, and $\chi^2 \to \infty$.
- Why? Because T_k given by (fixed order) pQCD, and uncertainty on this due to **missing higher orders** (MHOs) not generally included.

→ Essential to include measure of this if we are to have reasonable/viable interpretation of **fit quality** at high precision, in particular if default poor. Without this may be biasing fit.

→ Additional motivation, to give estimation of **uncertainty** in extracted PDFs due to MHOs in fit.

Theory Uncertainties

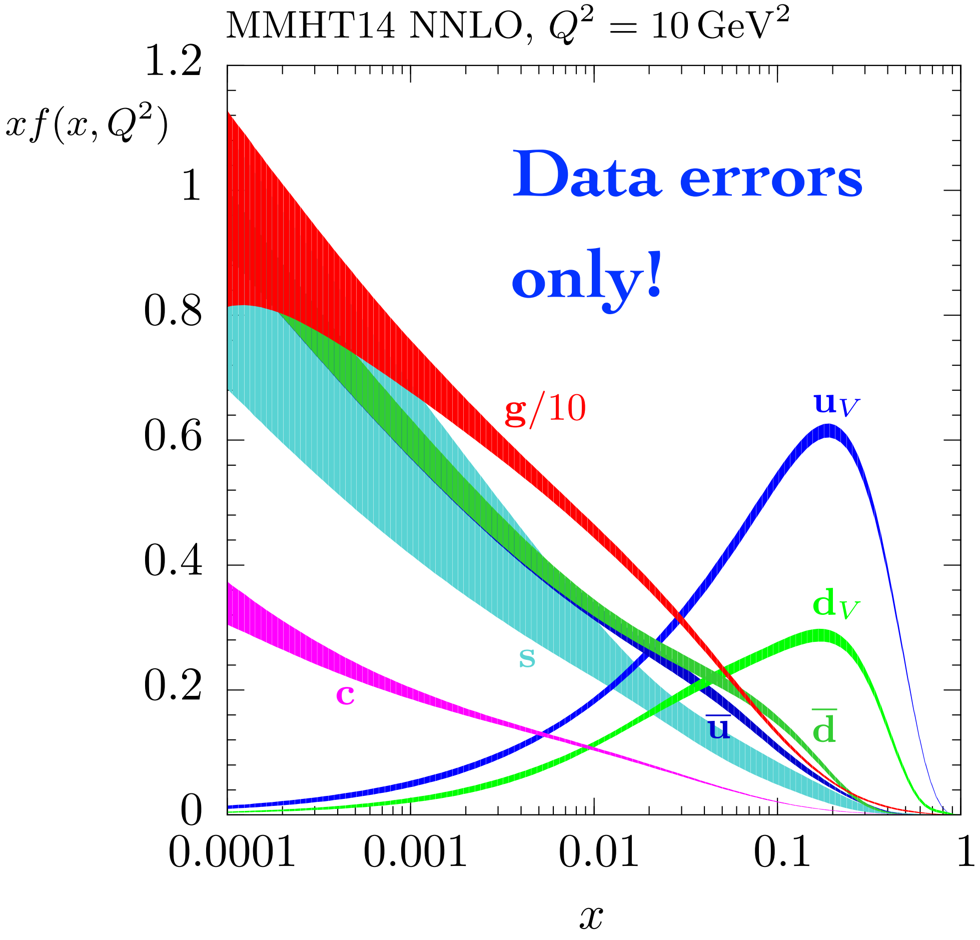
- PDF fit schematically given by inverting:

Dataset $$O \sim f \otimes \sigma \sim f \otimes \left(\sigma^{(0)} + \alpha_S \sigma^{(1)} + \cdots\right)$$

- Until recently only PDF errors corresponding to data errors in fit included.
- However in principle not only error source. Also that due to **missing higher orders** (the '…') in theory, from truncation of pert. expansion.

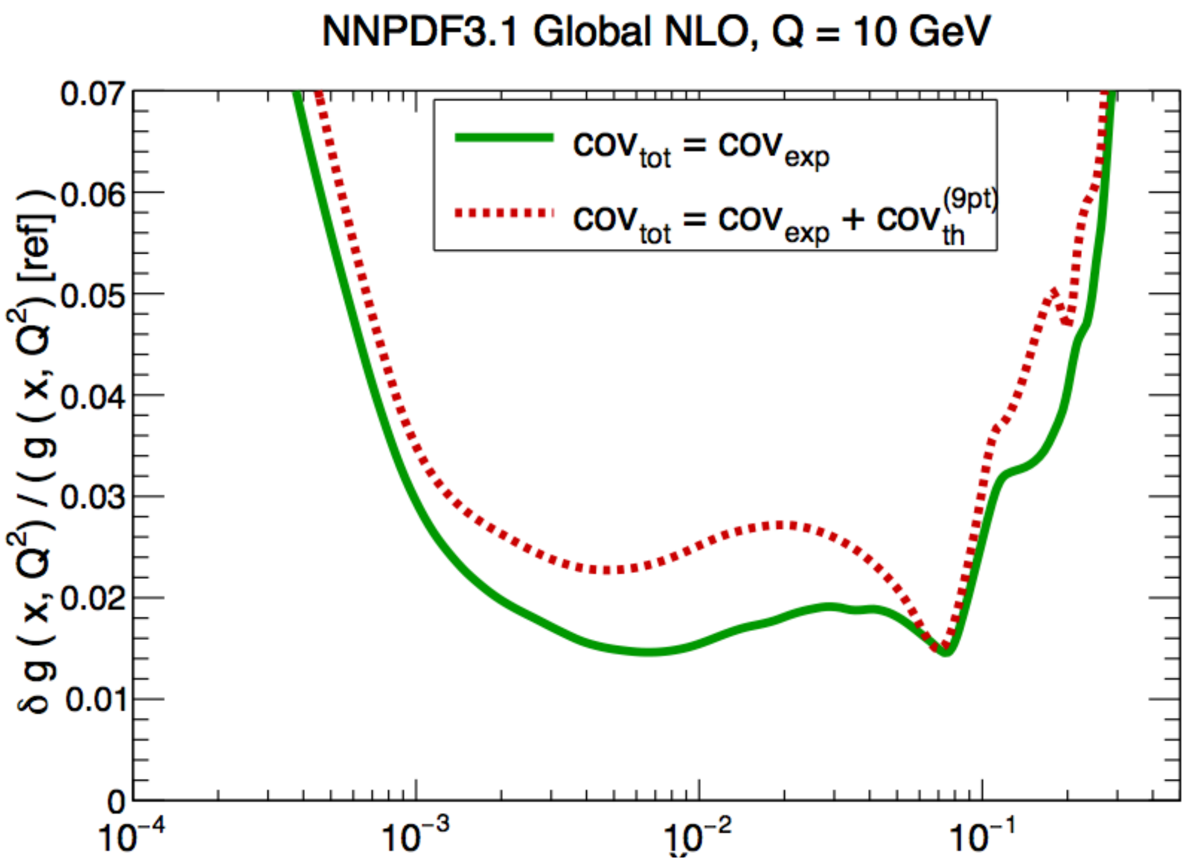
- In truth impossible to know the size of these '…' but various ways to estimate.
- Standard method: factorization/ renormalization **scale variations.**

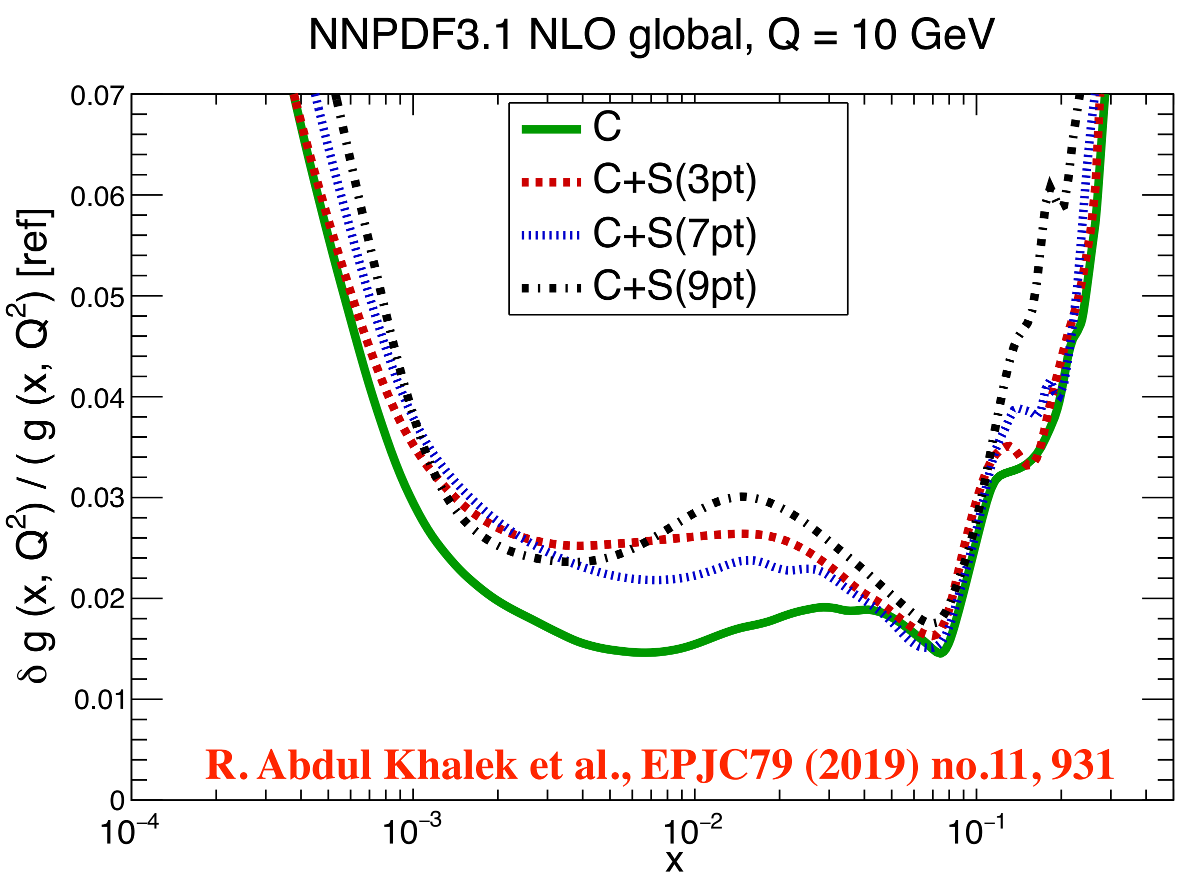
$$\delta O(\mu_F, \mu_R, \mu_0): \quad \mu_{F,R} \in \left(k\mu_0, \frac{\mu_0}{k}\right)$$



Theory Uncertainties

- Recent work: MHO uncertainties via **scale variations** in **NLO NNPDF** fit. Impact on PDF uncertainties not negligible (will be less at NNLO).





R. Abdul Khalek et al., EPJC79 (2019) no.11, 931

- Important **open questions**:

★ Significant source of ambiguity: how one treats the correlations (between/across datasets) for these errors.

★ Are scale variations the best way to estimate MHOs?

★ Risk of **double counting** with MHO uncertainty already accounted for when making predictions via PDFs?

Recent work: yes!

- ★ We already include MHO uncertainty (by scale variation) when predicting observables with PDFs. Risk of **double counting**?
- ★ **Simplified study**: recast PDF fit as direct relationship between fit and predicted observables.

LHL and R. S. Thorne, EPJC79 (2019), no.1, 39

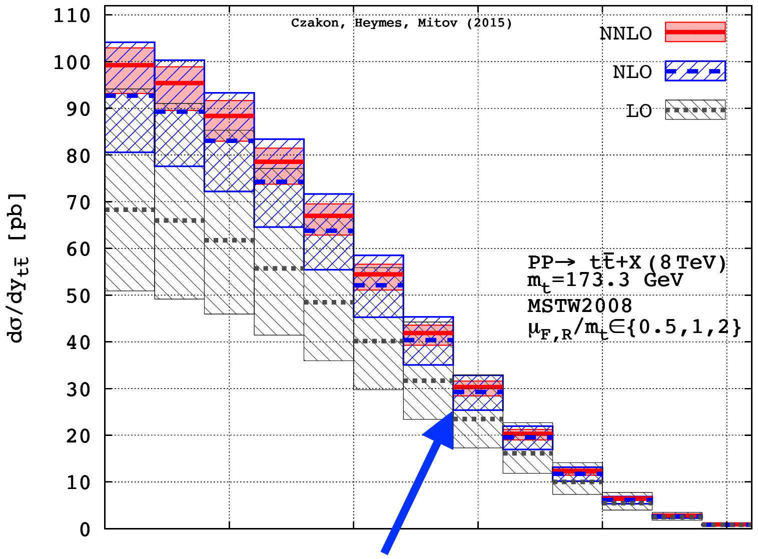
$$\begin{tikzcd}
\textbf{Fit} & O_{\rm fit} \arrow[dd, "A"] & \sim f_i(\mu^2) \otimes \sigma_i(\mu^2) \sim f_i(\mu^2) \otimes \left(\sigma_i^{(0)}(\mu^2) + \alpha_S \, \sigma_i^{(1)'}(\mu^2) + \cdots \right) \arrow[d, "B"] \\
& & f_i \arrow[d, "C"] \\
\textbf{Prediction} & O_{\rm pred} & \sim f_i(\mu^2) \otimes \sigma_i'(\mu^2) \sim f_i(\mu^2) \otimes \left(\sigma_i^{(0)'}(\mu^2) + \alpha_S \, \sigma_i^{(1)'}(\mu^2) + \cdots \right)
\end{tikzcd}$$

i : PDF type

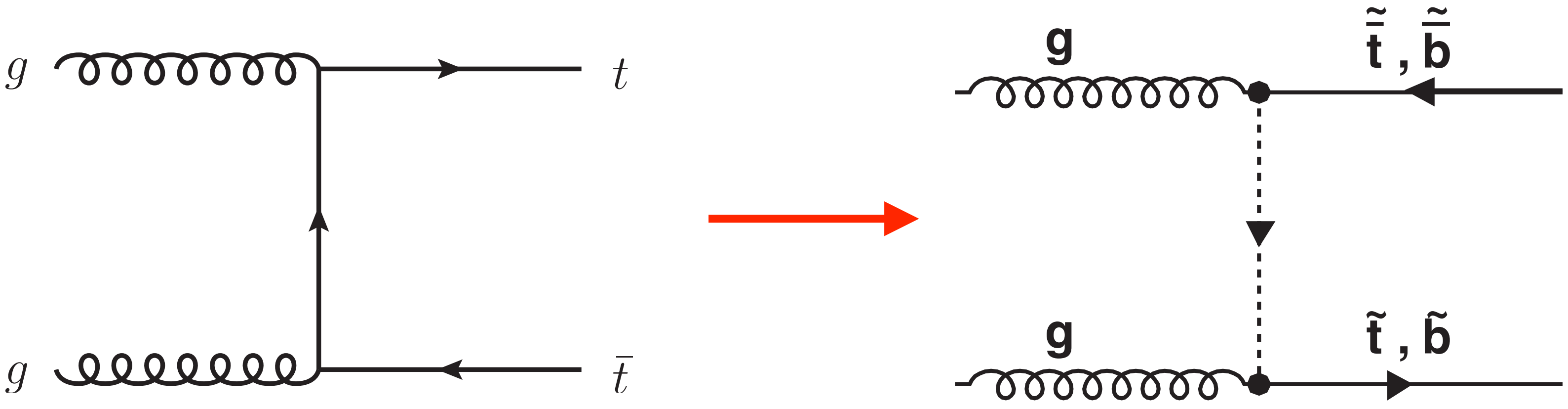
- Can propagate MHO uncertainty (scale variation) through to PDFs, but will then include such an uncertainty **again** in prediction.
- '**Theory uncertainty**': that inherent in expressing predicted quantity in terms of measured one. Varying at both **B** and **C** not obviously right.
- Recasting in terms of $O_1 \leftrightarrow O_2$ via **A** makes this concrete.

Details in paper!

- **Cutting to chase**: in certain cases including MHO uncertainty in PDFs and in prediction risks double counting, i.e. **overestimating error** on predictions.
- **Standard example**: fit process sensitive to high x gluon (e.g. top quark pair production) and predicts gluon-initiated BSM production.



Scale variation



★ Bottom line: as well as correlation in theory uncertainty between/within processes in fit, also have those between fit and prediction.

★ Not easy to do this in standard PDF fit, but indication that more work is needed here. Studies **ongoing** in this direction…stay tuned!

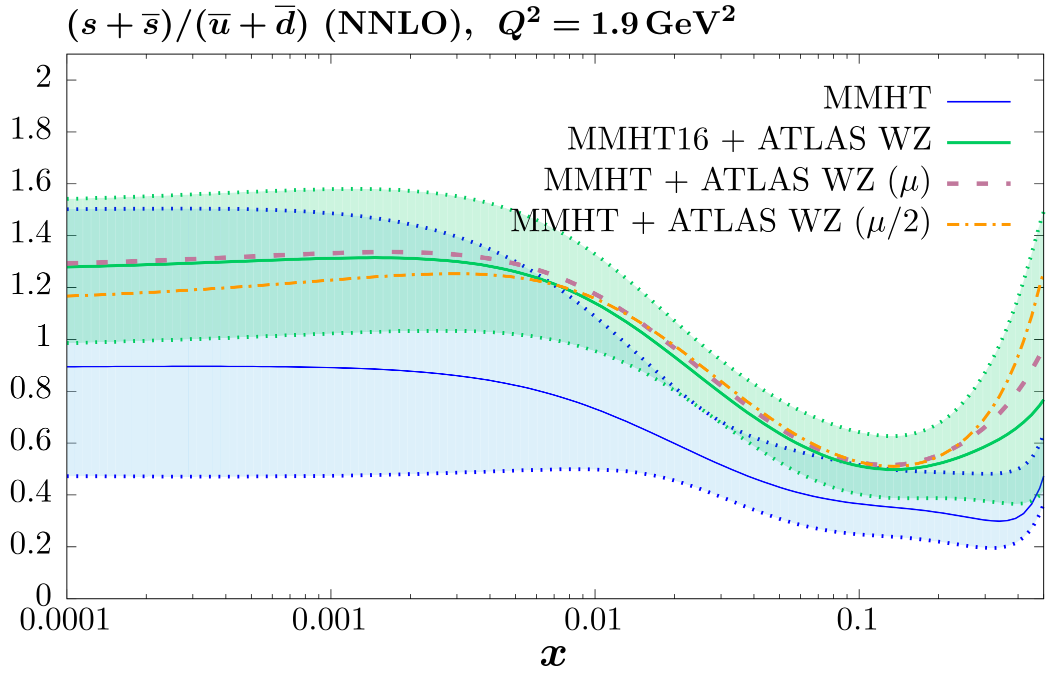
Interpreting Fit Quality: Summary

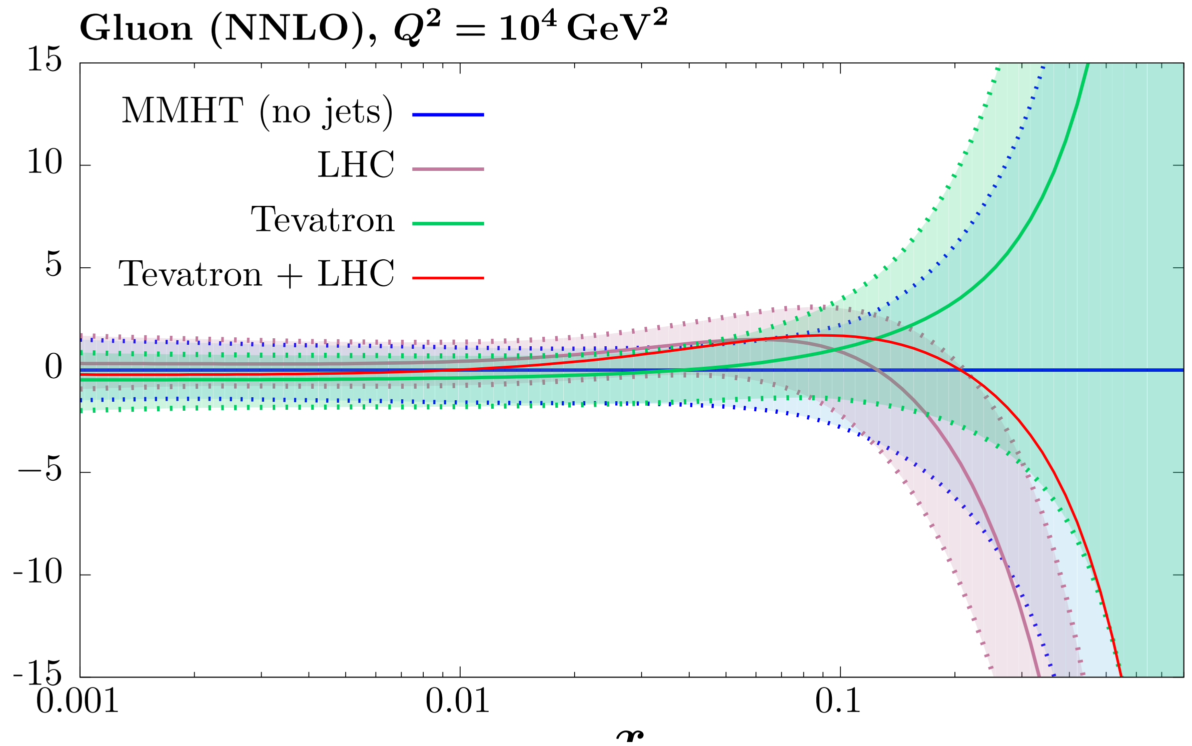
- In high precision LHC era, interpreting fit quality not straightforward. Depends sensitively on **experimental systematic** and **theoretical uncertainties**, and their **correlations**.
- Miss out/get wrong and very bad $\chi^2/N_{\rm pt}$ occurs, with unreliable PDF fit.
  - ★ **Cannot** simply sweep these issues under carpet, by e.g. post-hoc choice of fitting subset of data (individual distributions etc).
  - ★ **Full breakdown** of systematics and indication of uncertainty on default correlations seems essential (far from the current default).
  - ★ Including theoretical (MHO) uncertainty **mandatory**.
- But still many **open questions**: how much can we decorrelate experimental systematics? How do we best account for MHO uncertainties and their correlations?
- At what point is a bad $\chi^2/N_{\rm pt}$ due to new physics? Clearly question not just of relevance to PDF fitting!

Towards 'MMHT' 19

- Global groups busily updating fits to include new LHC data. '**MMHT'19** on its way. Include:
  - ★ Full LHC Run I dataset & latest updates from Tevatron/HERA.
  - ★ Extended parameterisation.
  - ★ Photon 'a la LUX' as standard.

- Much LHC Run-I data included already, with encouraging results!





Summary/Outlook

- ⋆ PDFs a key ingredient in LHC physics.
- ⋆ Precision LHC era: significant opportunity for PDF determination.
- ⋆ Projections for future: still encouraging potential for greatly improved PDF constraints from LHC.
- ⋆ But significant challenges before us: confronting high precision data in fits, dealing with tensions, poor fit quality, including theory uncertainties effectively etc.
- ⋆ Not simply question of adding ever more data to PDF fits. Much work ahead to make sense of what we are seeing…

Thank you for listening!