HUGS2023

May 30 - June 16, 2023 • Newport News, VA

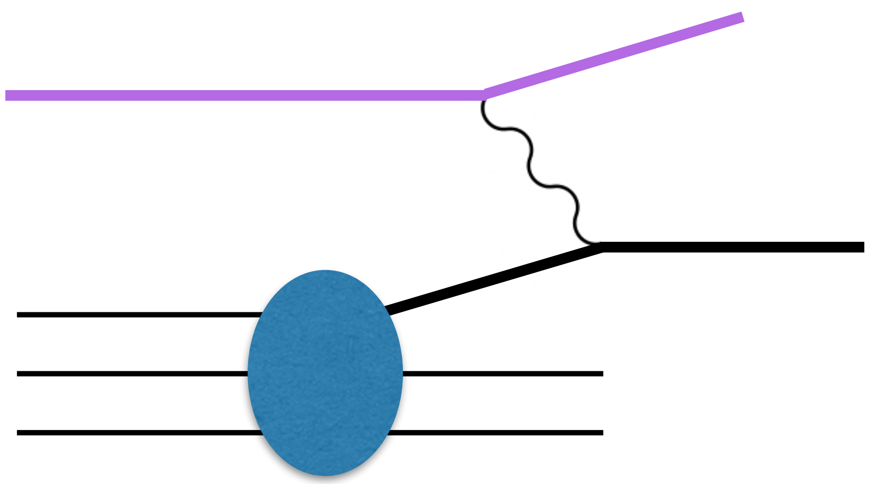
GLOBAL PDF FITS: CONNECTING LOW TO HIGH ENERGY PHYSICS

LECTURE III & IV

Maria Ubiali

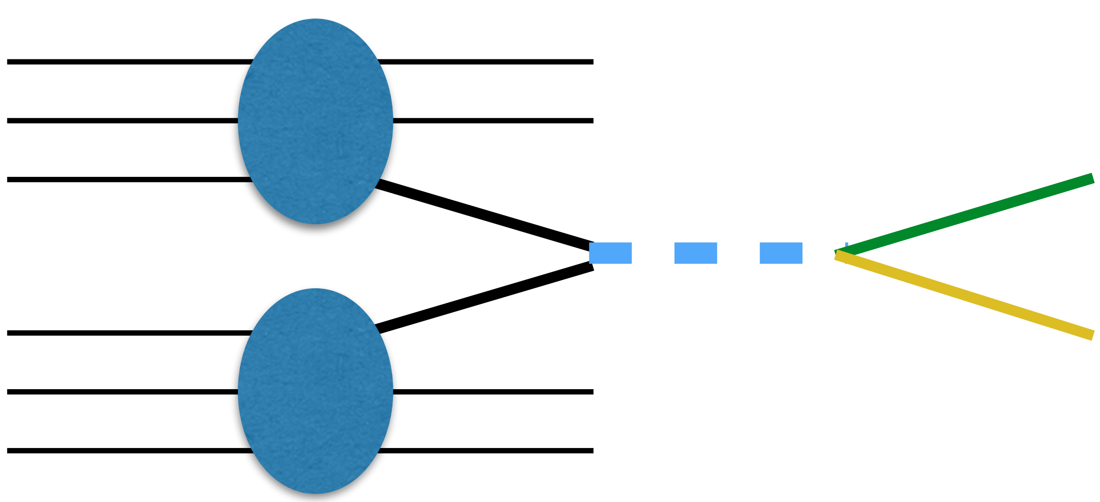
University of Cambridge

Highlights from yesterday



- Collinear factorisation picture
- Universality of PDFs
- DGLAP evolution of PDFs predicted by perturbative QCD

$$\frac{d\sigma_H^{ep\to ab}}{dX} = \sum_{i=-n_f}^{+n_f} \int_{x_B}^{1} \frac{dz}{z}\, f_i(z,\mu_F)\, \frac{d\hat{\sigma}_i^{ei}}{dX}(zS, \alpha_s(\mu_R), \mu_F) + \mathcal{O}\left(\frac{\Lambda^n}{S^n}\right)$$



- All terms in master equation have an uncertainty associated that is a component of the uncertainty of theory predictions

$$\frac{d\sigma_H^{pp\to ab}}{dX} = \sum_{i,j=-n_f}^{+n_f} \int_{\tau_0}^{1} \frac{dz_1}{z_1}\frac{dz_2}{z_2}\, f_i(z_1,\mu_F)\, f_j(z_2,\mu_F)\, \frac{d\hat{\sigma}_i^{ij}}{dX}(zS, \alpha_s(\mu_R), \mu_F) + \mathcal{O}\left(\frac{\Lambda^n}{S^n}\right)$$

Outline

- First two lectures (yesterday)
  - Motivation:
    the high energy big picture
  - Parton Model and QCD
  - Collinear Factorisation
- **Third and fourth lecture (today)**
  - Ingredients of a PDF global fits
  - Experimental input
  - Methodological aspects
  - Theoretical aspects
- Fourth lecture (tomorrow)
  - New frontiers and challenges

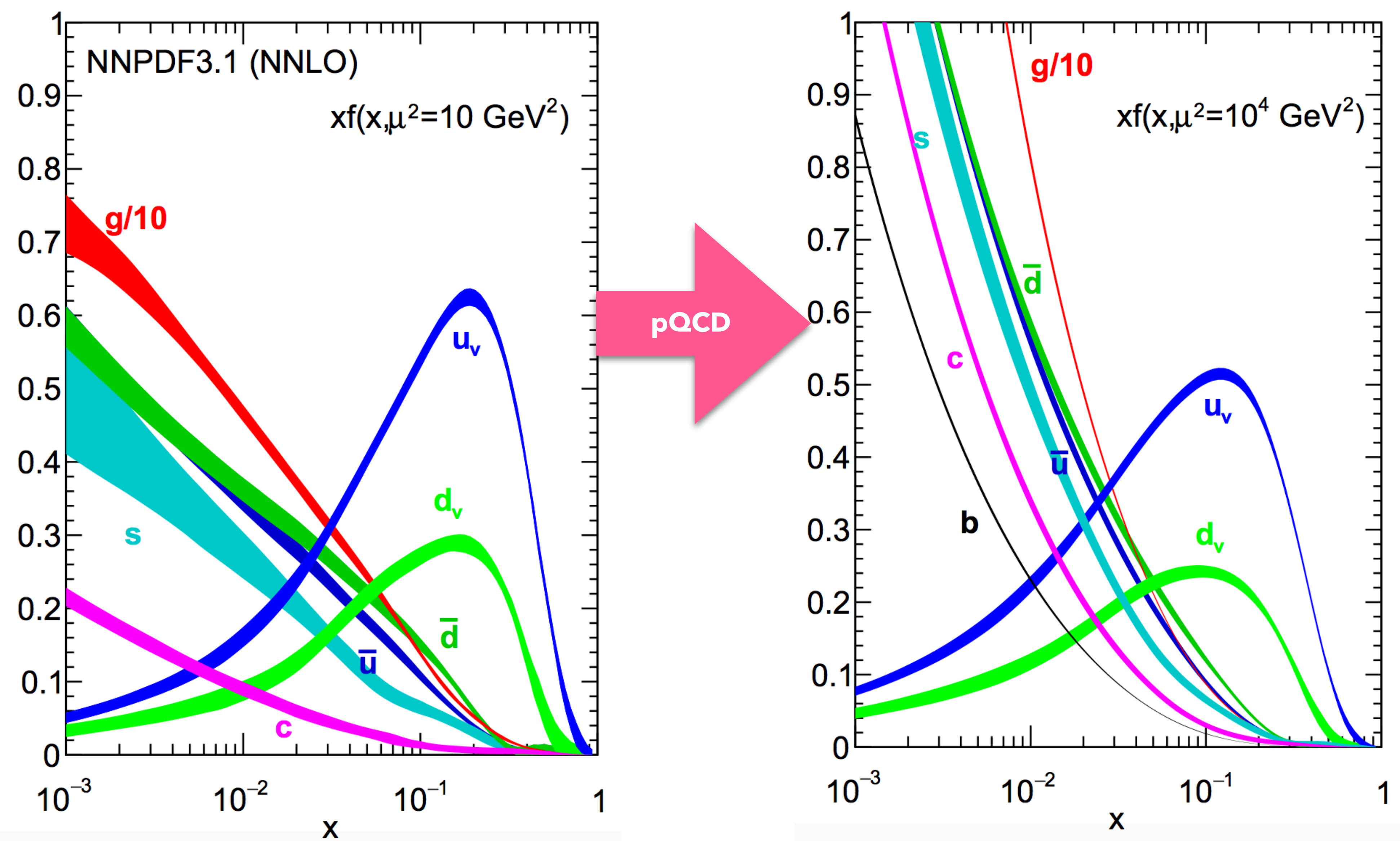
PDF determination

$$f_i(x, \mu)$$

Data

Perturbative QCD

Hadronic scale:
global fit of PDFs



High scale:
input to the LHC

PDF determination

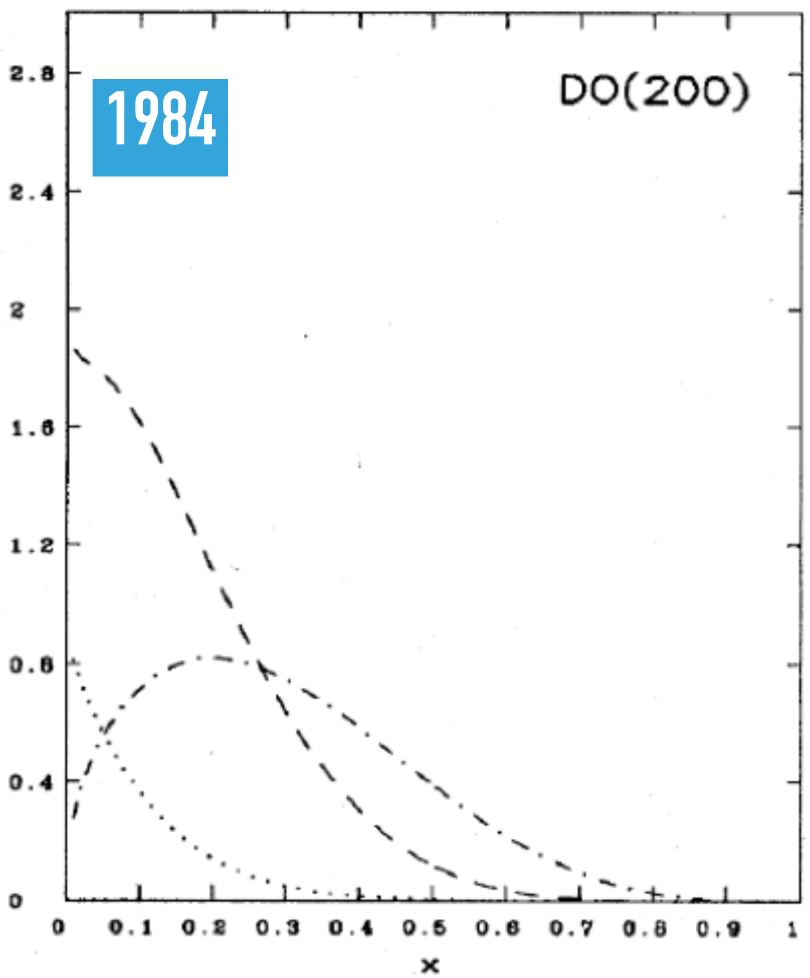
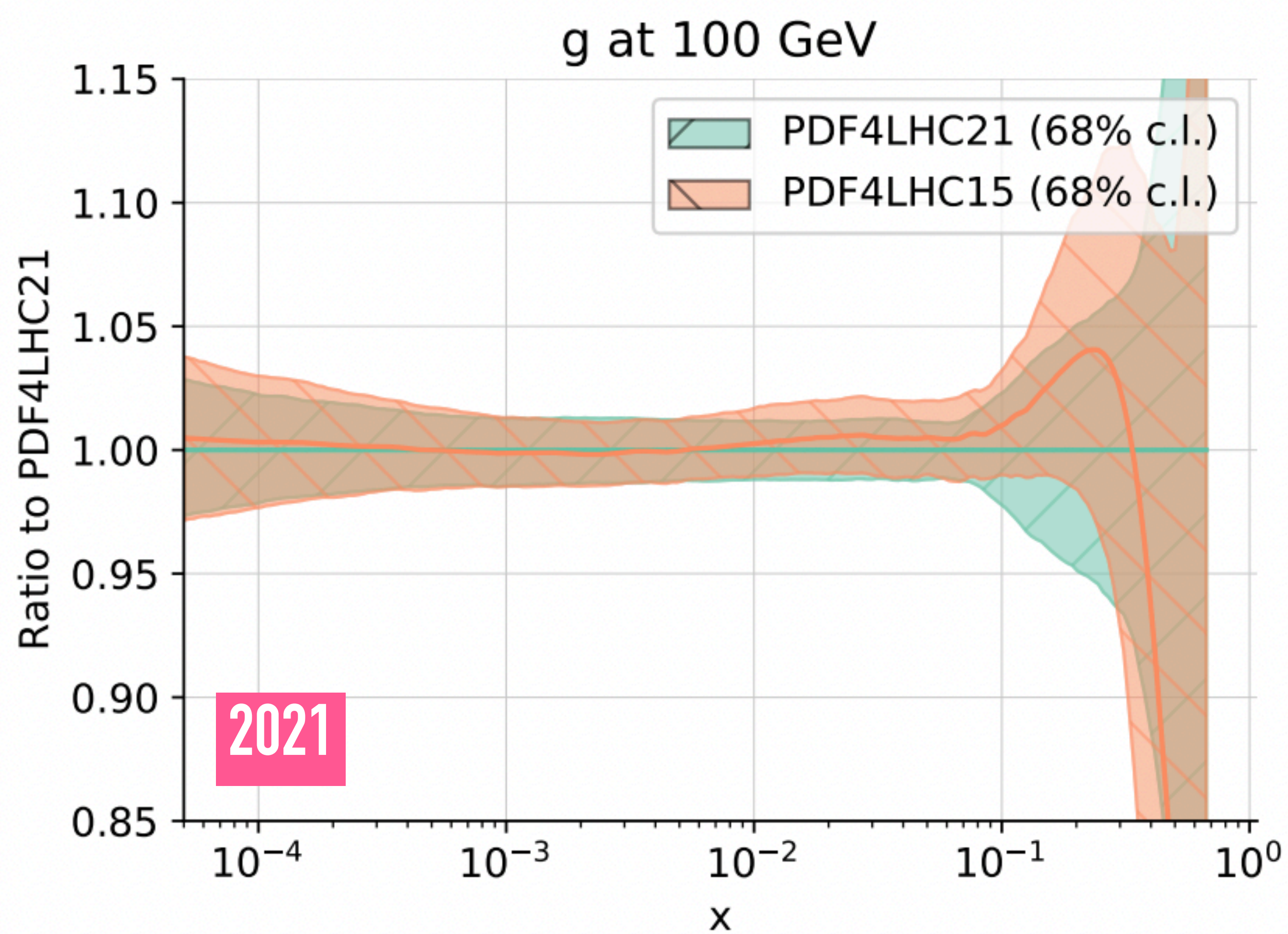


FIG. 27. "Soft-gluon" ($\Lambda=200$ MeV) parton distributions of Duke and Owens (1984) at $Q^2=5$ GeV^2: valence quark distribution $x[u_v(x)+d_v(x)]$ (dotted-dashed line), $xG(x)$ (dashed line), and $q_s(x)$ (dotted line).

Rev.. Mod. Phys. 1984



PDF4LHC21

★ 30 years of steady progress in PDF community have produced a huge impact on understanding of proton structure and precision physics

Ingredients of a PDF global fits

The ingredients

- Choose **experimental data** to fit and include all info on correlations
- **Theory settings**: perturbative order, heavy quark mass scheme, EW corrections, intrinsic heavy quarks, α_s, quark masses value and scheme
- Choose a starting scale Q_0 where pQCD applies
- **Parametrise** independent quarks and gluon distributions at the starting scale
- Solve **DGLAP equations** from initial scale to scales of experimental data and build up observables
- **Fit** PDFs to data
- Provide **error sets** to compute PDF uncertainties

The ingredients

- Choose **experimental data** to fit and include all info on correlations
- **Theory settings**: perturbative order, heavy quark mass scheme, EW corrections, intrinsic heavy quarks, α_s, quark masses value and scheme
- Choose a starting scale Q_0 where pQCD applies
- **Parametrise** independent quarks and gluon distributions at the starting scale
- Solve **DGLAP equations** from initial scale to scales of experimental data and build up observables
- **Fit** PDFs to data
- Provide **error sets** to compute PDF uncertainties

The ingredients

- Choose **experimental data** to fit and include all info on correlations
- **Theory settings**: perturbative order, heavy quark mass scheme, EW corrections, intrinsic heavy quarks, α_s, quark masses value and scheme
- Choose a starting scale Q_0 where pQCD applies
- **Parametrise** independent quarks and gluon distributions at the starting scale
- Solve **DGLAP equations** from initial scale to scales of experimental data and build up observables
- **Fit** PDFs to data
- Provide **error sets** to compute PDF uncertainties

The ingredients

- Choose **experimental data** to fit and include all info on correlations
- **Theory settings**: perturbative order, heavy quark mass scheme, EW corrections, intrinsic heavy quarks, α_s, quark masses value and scheme
- Choose a starting scale Q_0 where pQCD applies
- **Parametrise** independent quarks and gluon distributions at the starting scale
- Solve **DGLAP equations** from initial scale to scales of experimental data and build up observables
- **Fit** PDFs to data
- Provide **error sets** to compute PDF uncertainties

The ingredients

- Choose **experimental data** to fit and include all info on correlations
- **Theory settings**: perturbative order, heavy quark mass scheme, EW corrections, intrinsic heavy quarks, α_s, quark masses value and scheme
- Choose a starting scale Q_0 where pQCD applies
- **Parametrise** independent quarks and gluon distributions at the starting scale
- Solve **DGLAP equations** from initial scale to scales of experimental data and build up observables
- **Fit** PDFs to data
- Provide **error sets** to compute PDF uncertainties

The ingredients

- Choose **experimental data** to fit and include all info on correlations
- **Theory settings**: perturbative order, heavy quark mass scheme, EW corrections, intrinsic heavy quarks, α_s, quark masses value and scheme
- Choose a starting scale Q_0 where pQCD applies
- **Parametrise** independent quarks and gluon distributions at the starting scale
- Solve **DGLAP equations** from initial scale to scales of experimental data and build up observables
- **Fit** PDFs to data
- Provide PDF **error sets** to compute PDF uncertainties

The ingredients

$$\sigma_{\mathcal{F}} = \left(\sum_{k=1}^{N_{\rm set}} \left(\mathcal{F}[\{f^{(k)}\}] - \mathcal{F}[\{f^{(0)}\}] \right)^2 \right)^{1/2}$$

error sets
mem > 1

central set
mem = 0

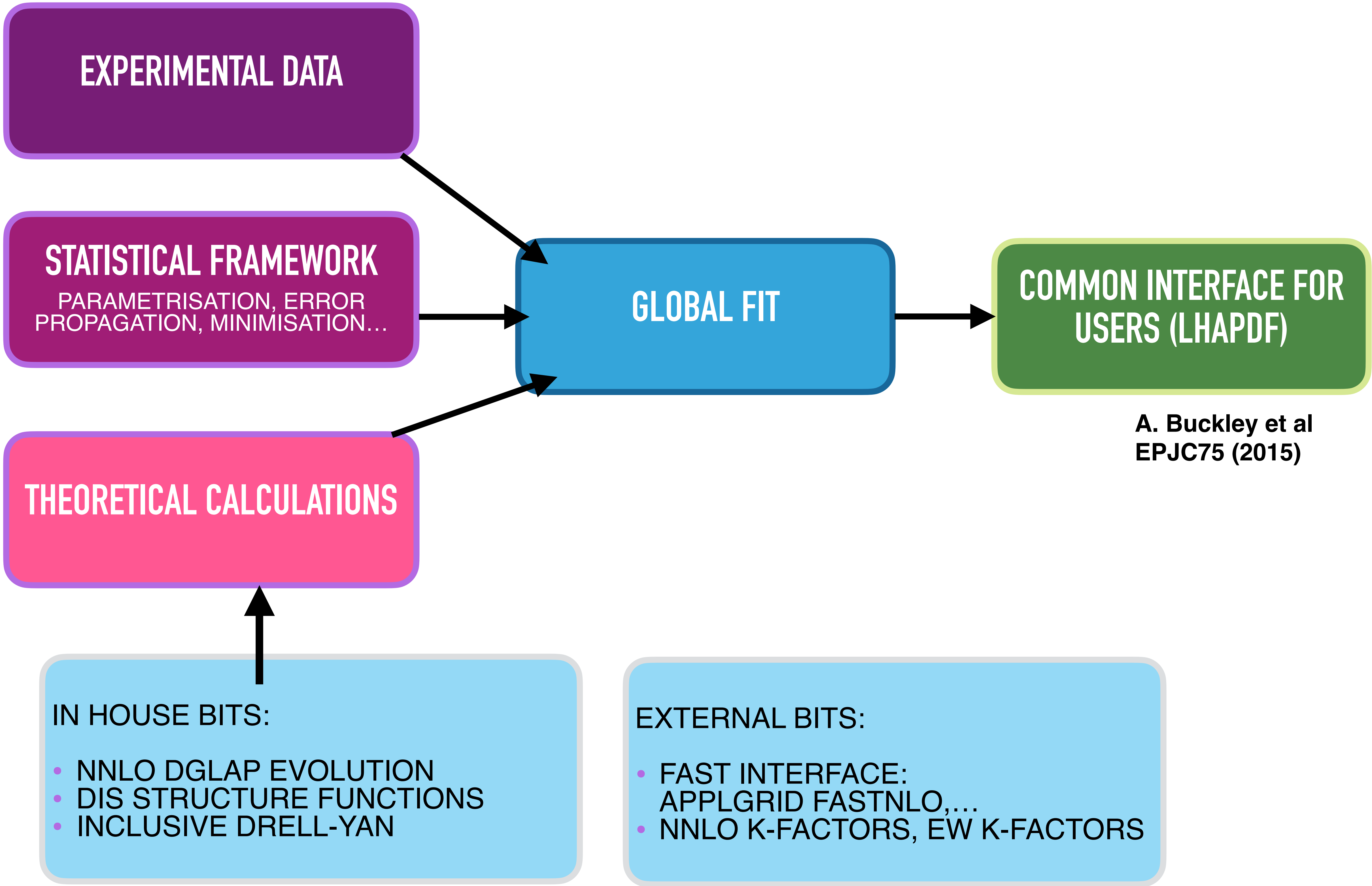
```
call InitPDF(mem)

call evolvePDF(x,Q,f)
```

LHAPDF interface
http://lhapdf.hepforge.org

- Provide PDF **error sets** to compute PDF uncertainties

	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6
Parton	tbar	bbar	cbar	sbar	ubar	dbar	g	d	u	s	c	b	t

A complex machinery



Experimental input

Experimental data

- PDFs are not measurable, we measure observables that convolute PDFs with partonic cross sections

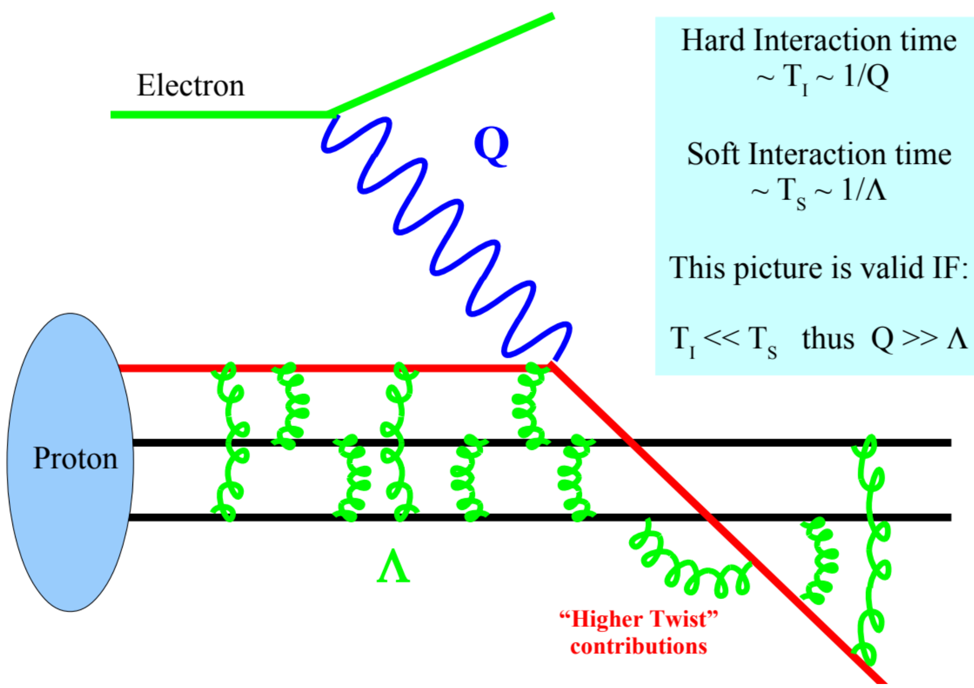
$$\frac{d\sigma_H^{ep\to ab}}{dX} = \sum_{i=-n_f}^{+n_f} \int_{x_B}^{1} \frac{dz}{z}\, f_i(z,\mu_F)\, \frac{d\hat{\sigma}_i^{ei}}{dX}(zS, \alpha_s(\mu_R), \mu_F) + \mathcal{O}\left(\frac{\Lambda^n}{S^n}\right)$$

$$\frac{d\sigma_H^{pp\to ab}}{dX} = \sum_{i,j=-n_f}^{+n_f} \int_{\tau_0}^{1} \frac{dz_1}{z_1}\frac{dz_2}{z_2}\, f_i(z_1,\mu_F)\, f_j(z_2,\mu_F)\, \frac{d\hat{\sigma}_i^{ij}}{dX}(zS, \alpha_s(\mu_R), \mu_F) + \mathcal{O}\left(\frac{\Lambda^n}{S^n}\right)$$

- Most fits exclude regions where factorisation fails to apply (low Q2 and large x). Typically

$$Q^2_{\min} = 2\,\mathrm{GeV}^2$$

$$W^2_{\min} = \left(Q^2\frac{1-x}{x}\right)_{\min} = 12.5\,\mathrm{GeV}^2$$



Experimental data

- PDFs are not measurable, we measure observables that convolute PDFs with partonic cross sections

$$\frac{d\sigma_H^{ep\to ab}}{dX} = \sum_{i=-n_f}^{+n_f} \int_{x_B}^{1} \frac{dz}{z}\, f_i(z,\mu_F)\, \frac{d\hat{\sigma}_i^{ei}}{dX}(zS, \alpha_s(\mu_R), \mu_F) + \mathcal{O}\left(\frac{\Lambda^n}{S^n}\right)$$

$$\frac{d\sigma_H^{pp\to ab}}{dX} = \sum_{i,j=-n_f}^{+n_f} \int_{\tau_0}^{1} \frac{dz_1}{z_1}\frac{dz_2}{z_2}\, f_i(z_1,\mu_F)\, f_j(z_2,\mu_F)\, \frac{d\hat{\sigma}_i^{ij}}{dX}(zS, \alpha_s(\mu_R), \mu_F) + \mathcal{O}\left(\frac{\Lambda^n}{S^n}\right)$$

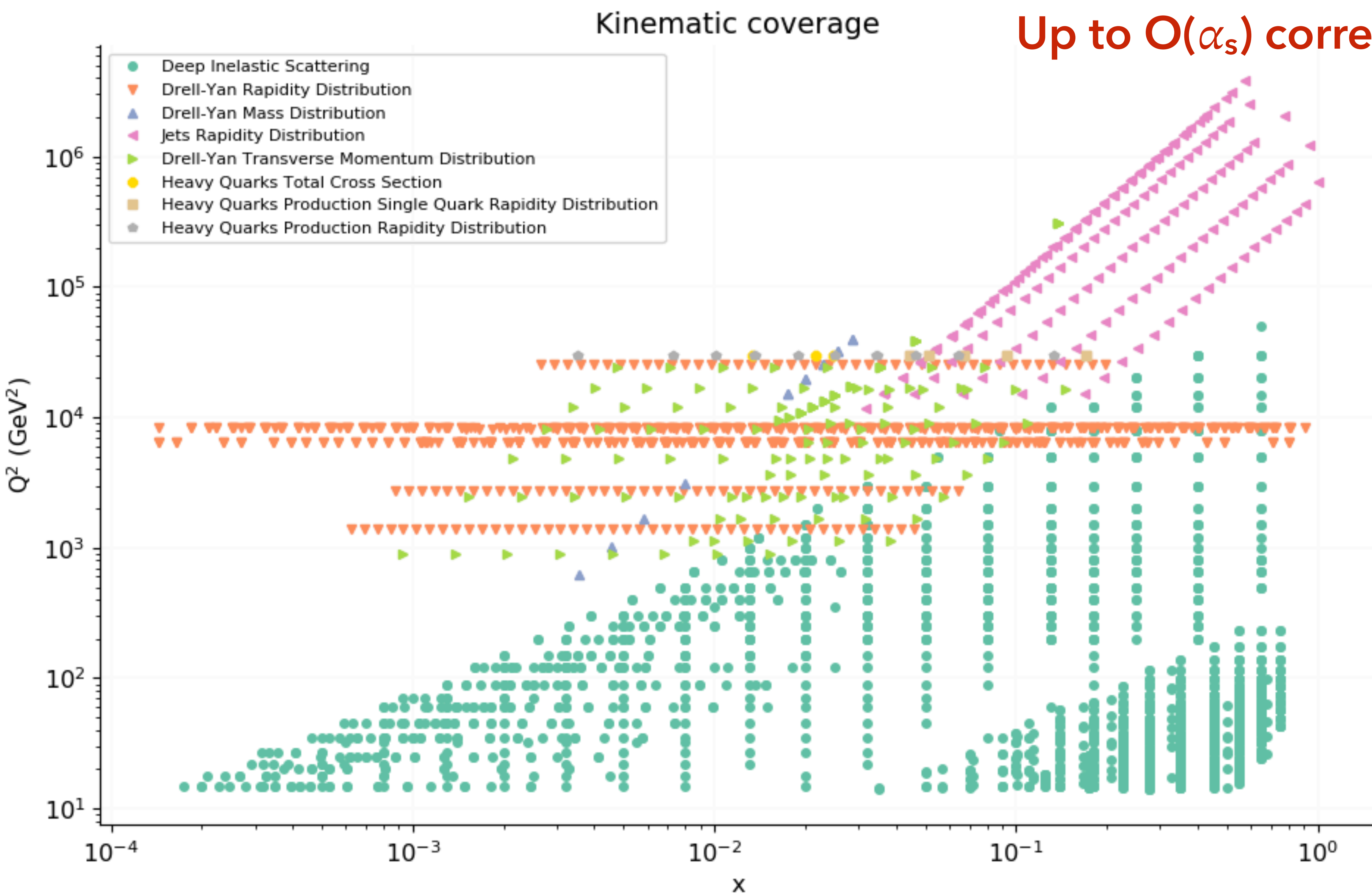
- Different data constrain different PDF combinations in different regions

    ➡ DIS data on proton abundant and precise (HERA)
    ➡ In principle F_2, F_3 CC provide 4 light quark combinations
    F_2, F_3 NC provide 2 extra light quark combinations
    ➡ HERA data only determine four combinations of PDFs
    ➡ Old DIS and Drell-Yan data still used because of isospin symmetry
    ➡ W,Z boson final state provide lot of information, gluon from scale dependence
    ➡ Processes with jets and/or heavy quark in final states direct handle on the gluon

Disentangling PDFs

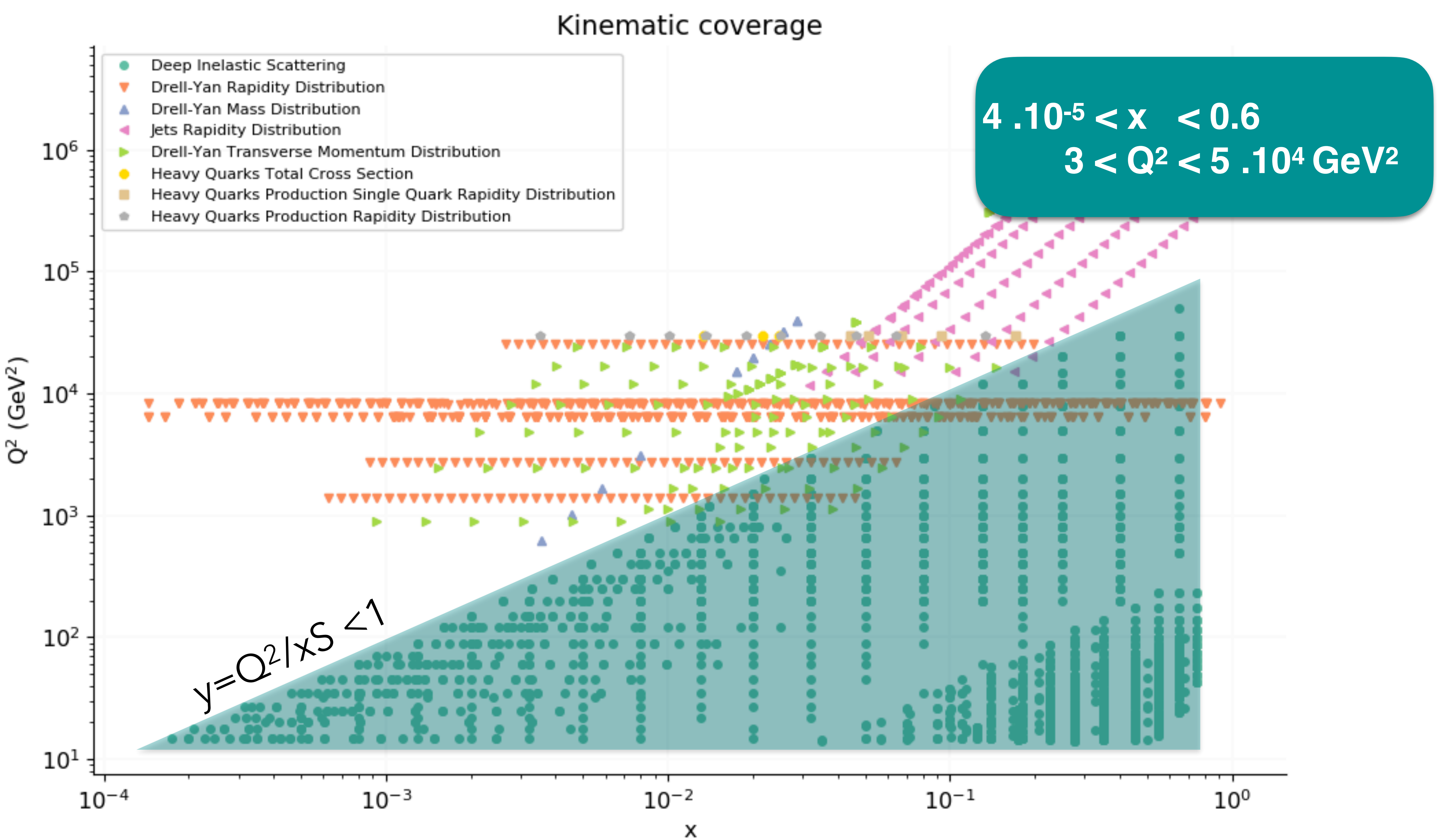
Up to $O(\alpha_s)$ corrections

Q-dependence: pert. theory



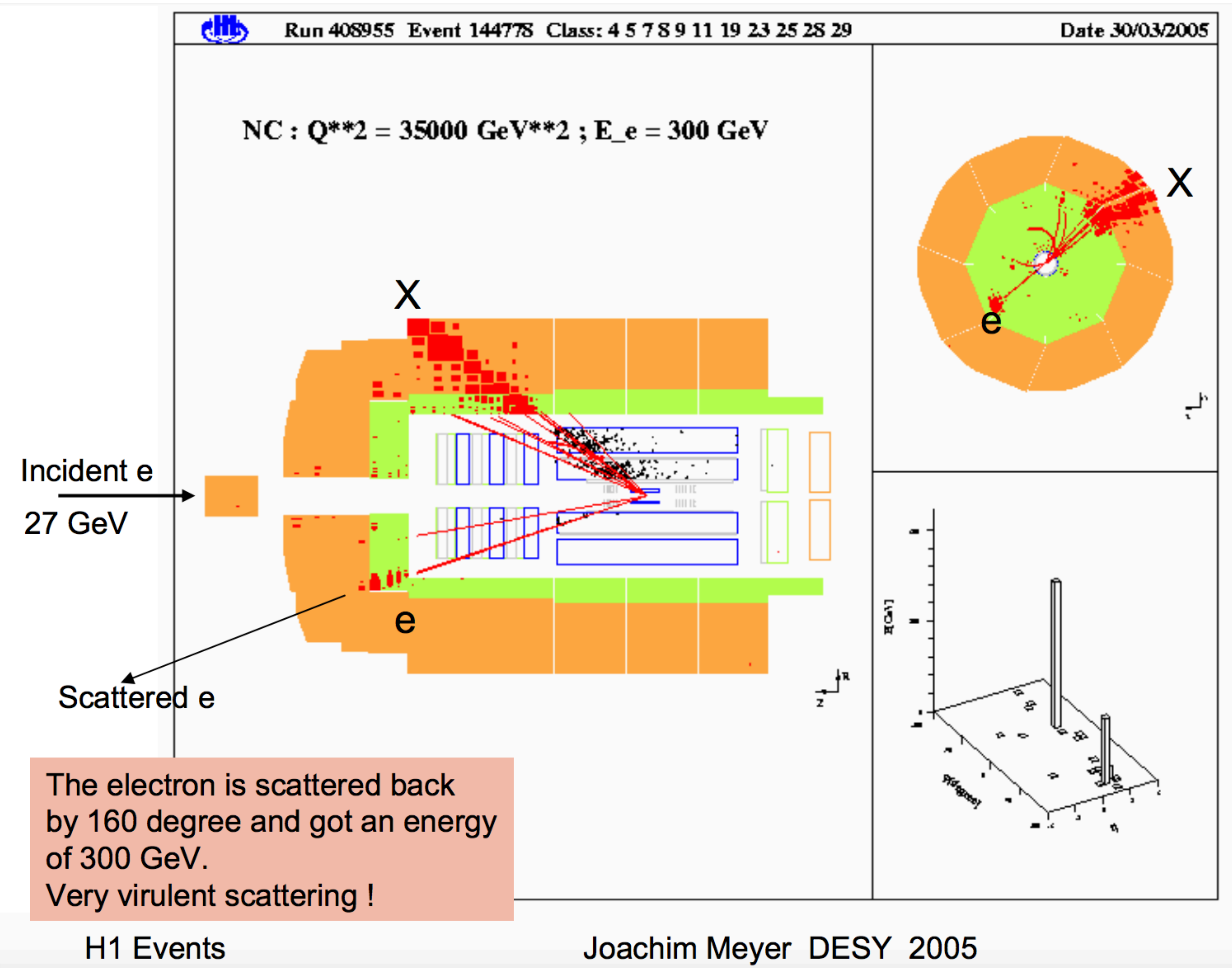
x-dependence: from data

HERA data



Kinematic coverage
Deep Inelastic Scattering
Drell-Yan Rapidity Distribution
Drell-Yan Mass Distribution
Jets Rapidity Distribution
Drell-Yan Transverse Momentum Distribution
Heavy Quarks Total Cross Section
Heavy Quarks Production Single Quark Rapidity Distribution
Heavy Quarks Production Rapidity Distribution
Q^2 (GeV^2)
x
$y=Q^2/xS<1$
$4 .10^{-5} < x < 0.6$
$3 < Q^2 < 5 .10^4$ GeV^2

HERA data



Neutral Current event

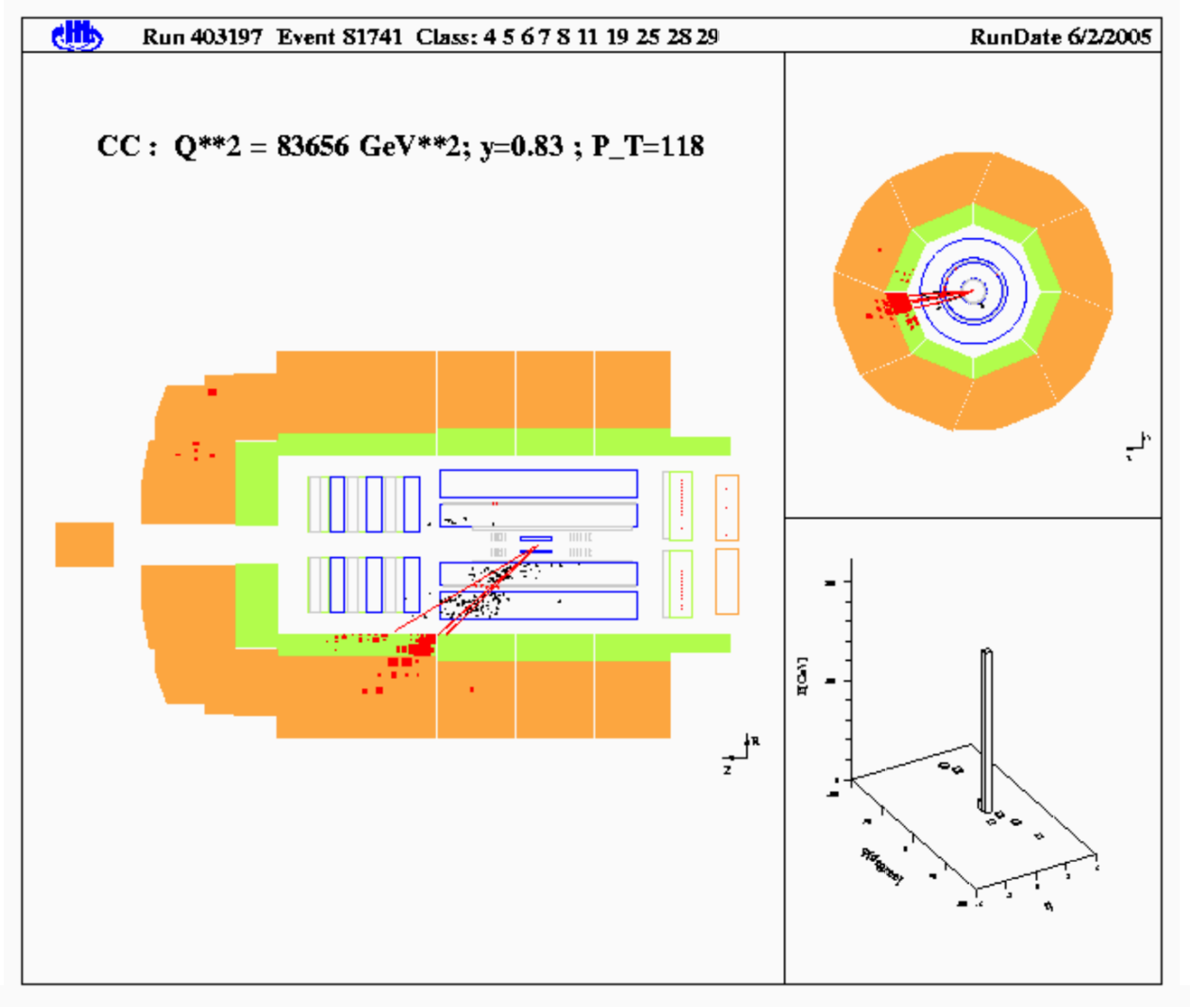
$$ep \to e\,X$$

$$Q = 180\,\mathrm{GeV}$$

$$y = 0.66$$

$$x_B = 0.47$$

HERA data



Charged Current event

$$ep \to \nu X$$

$$Q = 289\,\text{GeV}$$

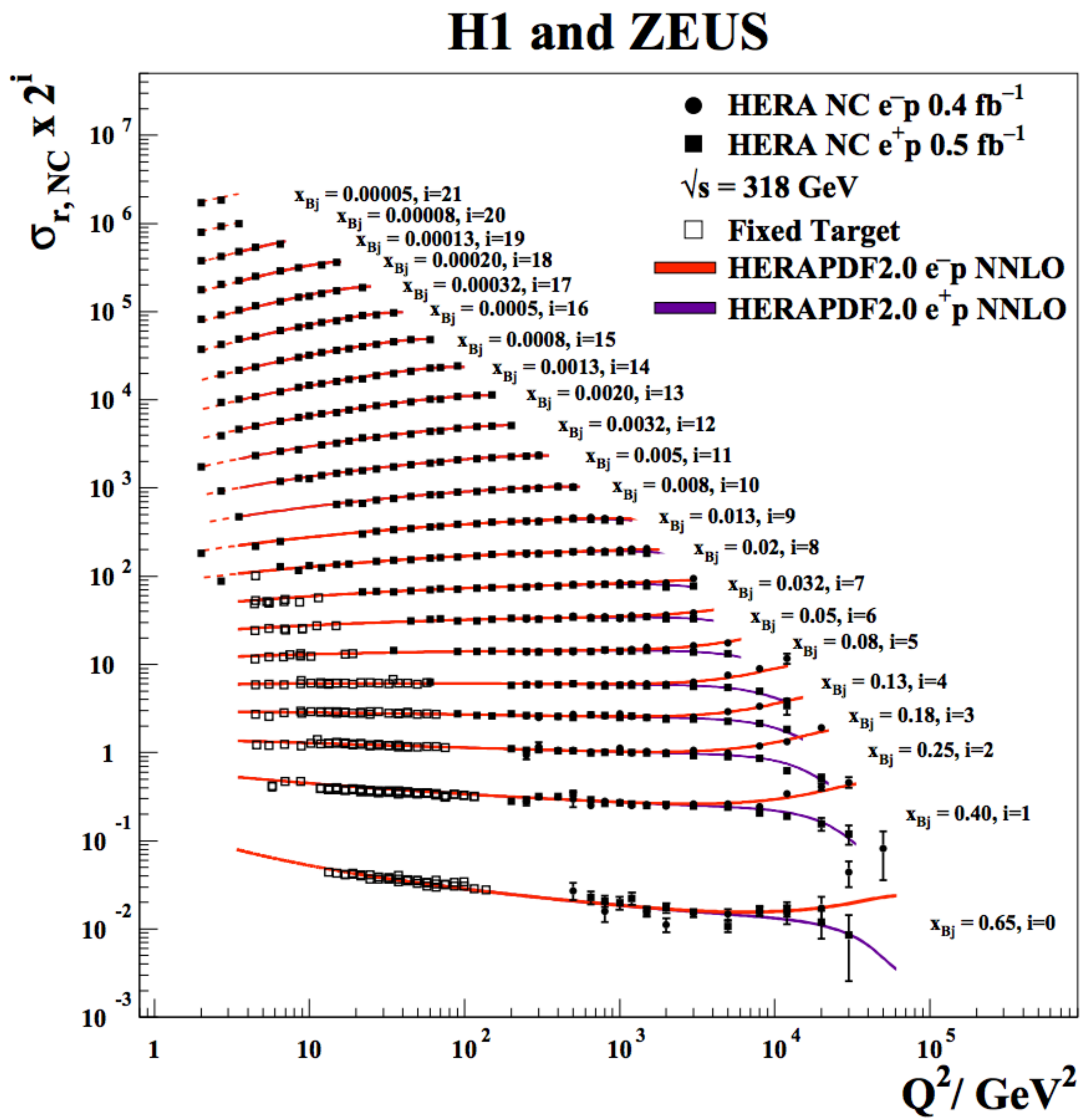
$$y = 0.83$$

$$x_B = 0.91$$

Joachim Meyer DESY 2005

HERA data

- Combination of Run I + Run II data led to very precise measurements of reduced xsec
- F3 contribution visible at larger x and Q~Mz



arXiv: 1506.06042

HERA data

Neutral Current

$$\left[F_2^{\gamma}, F_2^{\gamma Z}, F_2^{Z}\right] = x \sum_{i=1}^{n_f} \left[e_i^2, 2e_i g_V^i, (g_V^i)^2 + (g_A^i)^2\right] (q_i + \bar{q}_i)$$

3 4

$$\left[F_3^{\gamma}, F_3^{\gamma Z}, F_3^{Z}\right] = x \sum_{i=1}^{n_f} \left[0, 2e_i g_A^i, 2g_V^i g_A^i\right] (q_i - \bar{q}_i)$$

Charged Current

1

$$F_2^{W^-} = 2x\left(u + \bar{d} + \bar{s} + c\right),$$

$$F_3^{W^-} = 2x\left(u - \bar{d} - \bar{s} + c\right),$$

2

$$F_2^{W^+} = 2x\left(d + \bar{u} + \bar{c} + s\right),$$

$$F_3^{W^+} = 2x\left(d - \bar{u} - \bar{c} + s\right),$$

$$\boxed{\frac{d^2\sigma}{dx dQ^2} \propto Y_+ F_2(x, Q^2) \mp Y_- x F_3(x, Q^2) - y^2 F_L(x, Q^2)}$$

Longitudinal Structure function

$$F_L(x, Q^2) = \frac{\alpha_s(Q^2)}{\pi} \left[\frac{4}{3} \int_x^1 \frac{dy}{y} \left(\frac{x}{y}\right)^2 F_2(y, Q^2) + 2 \sum_i e_i^2 \int_x^1 \frac{dy}{y} \left(\frac{x}{y}\right)^2 \left(1 - \frac{x}{y}\right) g(y, Q^2)\right]$$

Altarelli, Martiinelli Phys. Lett. 76B (1978)

HERA data

Neutral Current

$$\left[F_2^{\gamma}, F_2^{\gamma Z}, F_2^{Z}\right] = x \sum_{i=1}^{n_f} \left[e_i^2, 2e_i g_V^i, (g_V^i)^2 + (g_A^i)^2\right](q_i + \bar{q}_i)$$

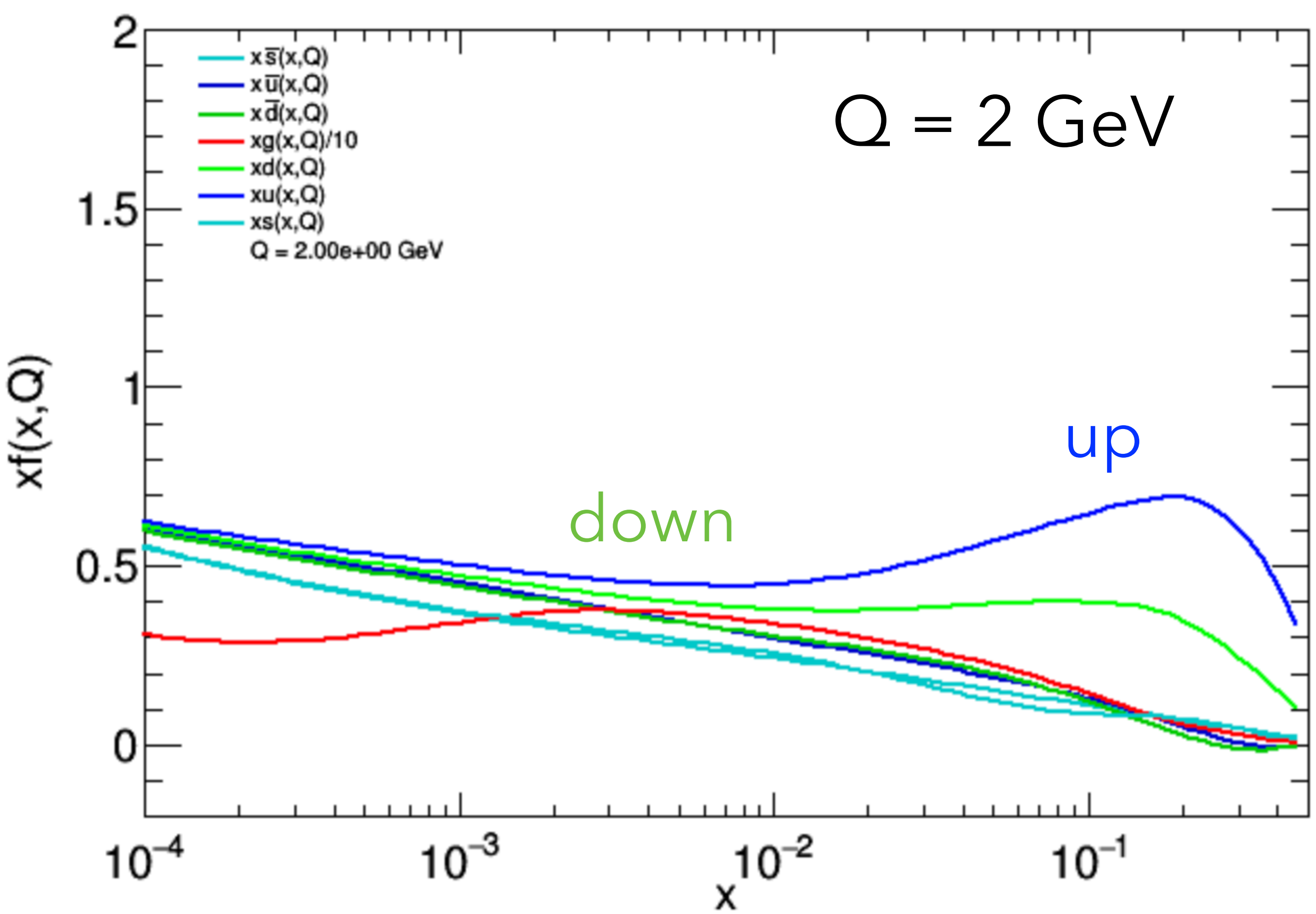
3 4

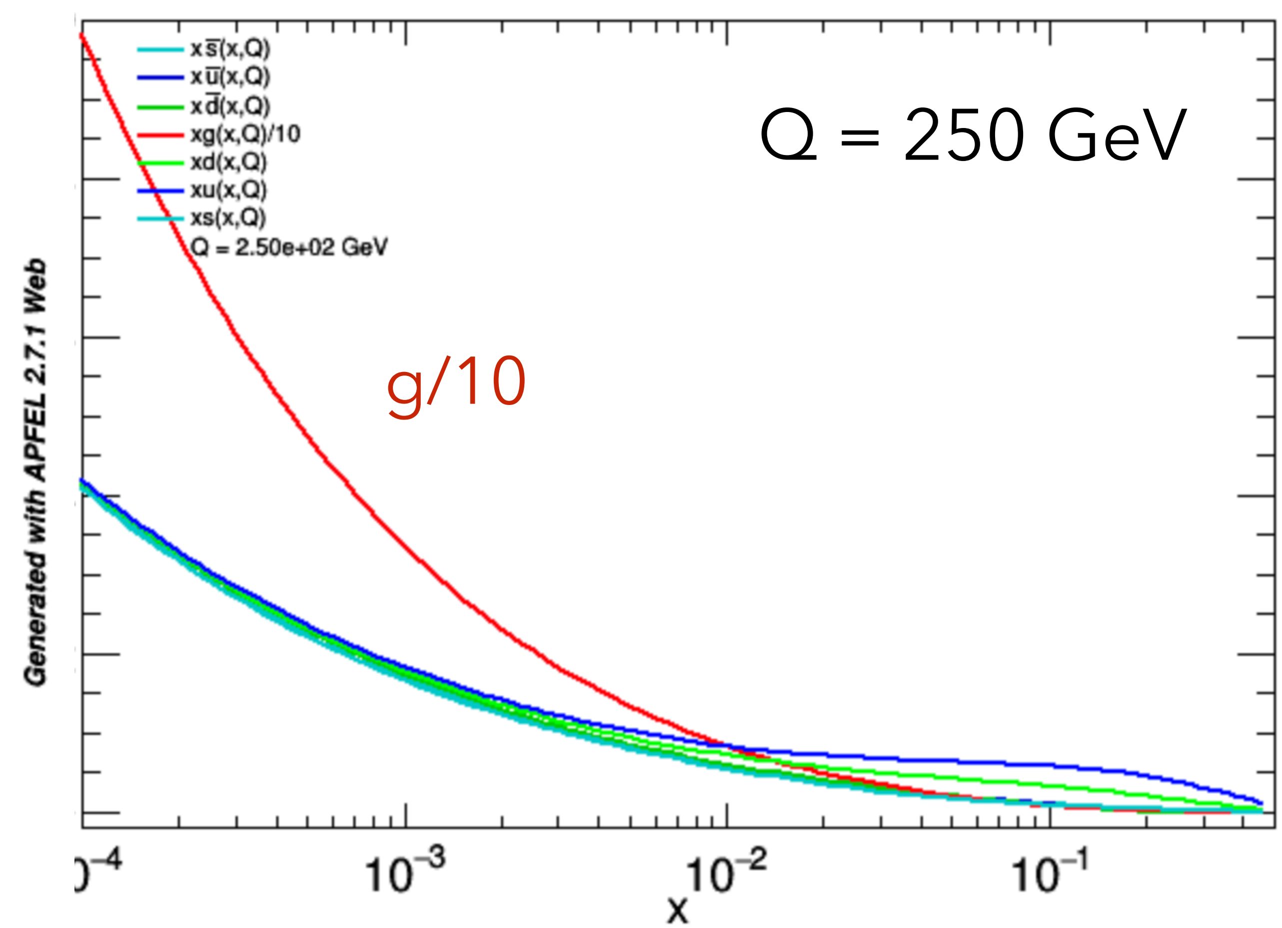
$$\left[F_3^{\gamma}, F_3^{\gamma Z}, F_3^{Z}\right] = x \sum_{i=1}^{n_f} \left[0, 2e_i g_A^i, 2g_V^i g_A^i\right](q_i - \bar{q}_i)$$

Charged Current

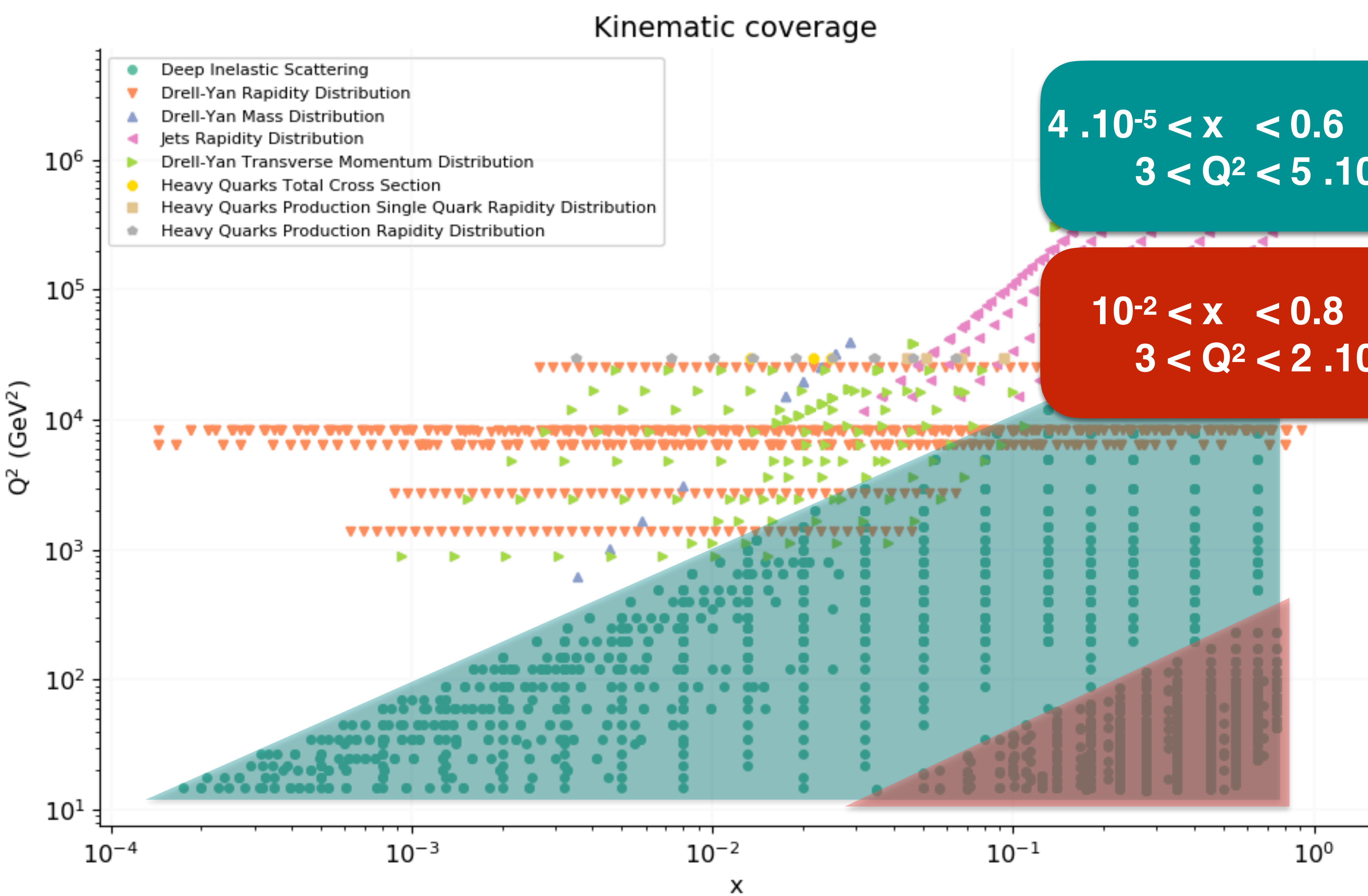
1
$$F_2^{W^-} = 2x\left(u + \bar{d} + \bar{s} + c\right),$$
$$F_3^{W^-} = 2x\left(u - \bar{d} - \bar{s} + c\right),$$

2
$$F_2^{W^+} = 2x\left(d + \bar{u} + \bar{c} + s\right),$$
$$F_3^{W^+} = 2x\left(d - \bar{u} - \bar{c} + s\right),$$





Fixed target DIS data



$4 .10^{-5} < x < 0.6$
$3 < Q^2 < 5 .10^4$ GeV^2

$10^{-2} < x < 0.8$
$3 < Q^2 < 2 .10^2$ GeV^2

Fixed target DIS data

- Experimentally measured is deuteron structure function

$$F_2^d = (F_2^p + F_2^n)/2$$

- Assumption (SU(2) isospin): neutron is just like proton with u ⇔ d

proton = uud

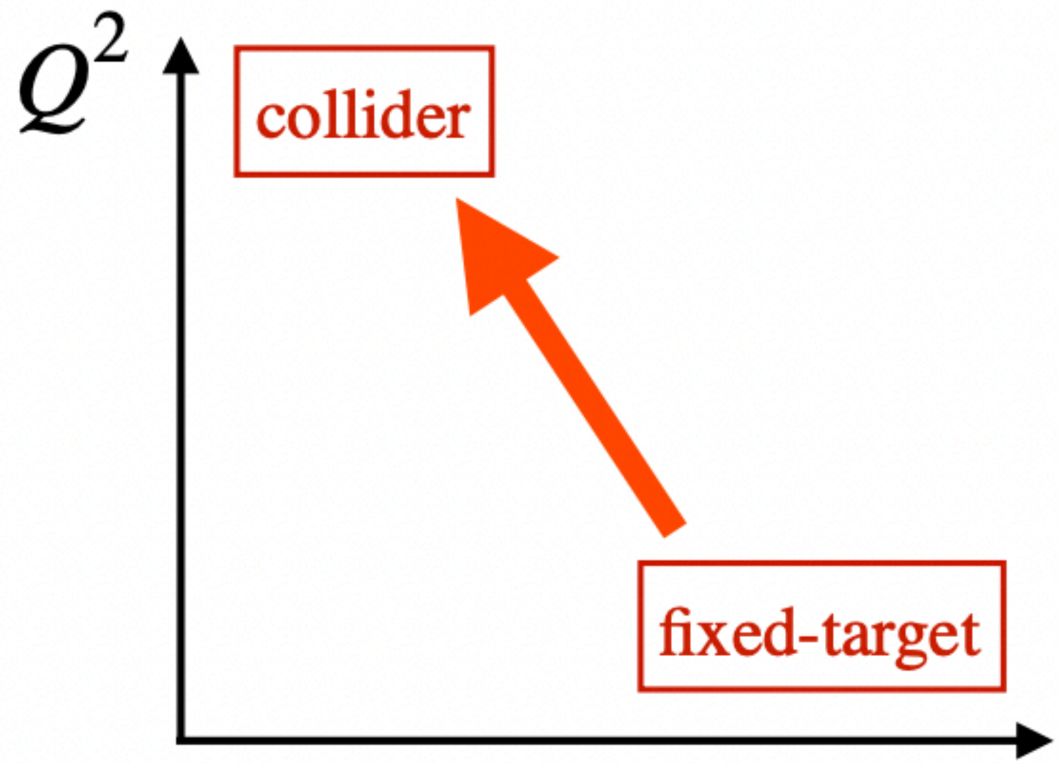
neutron = ddu

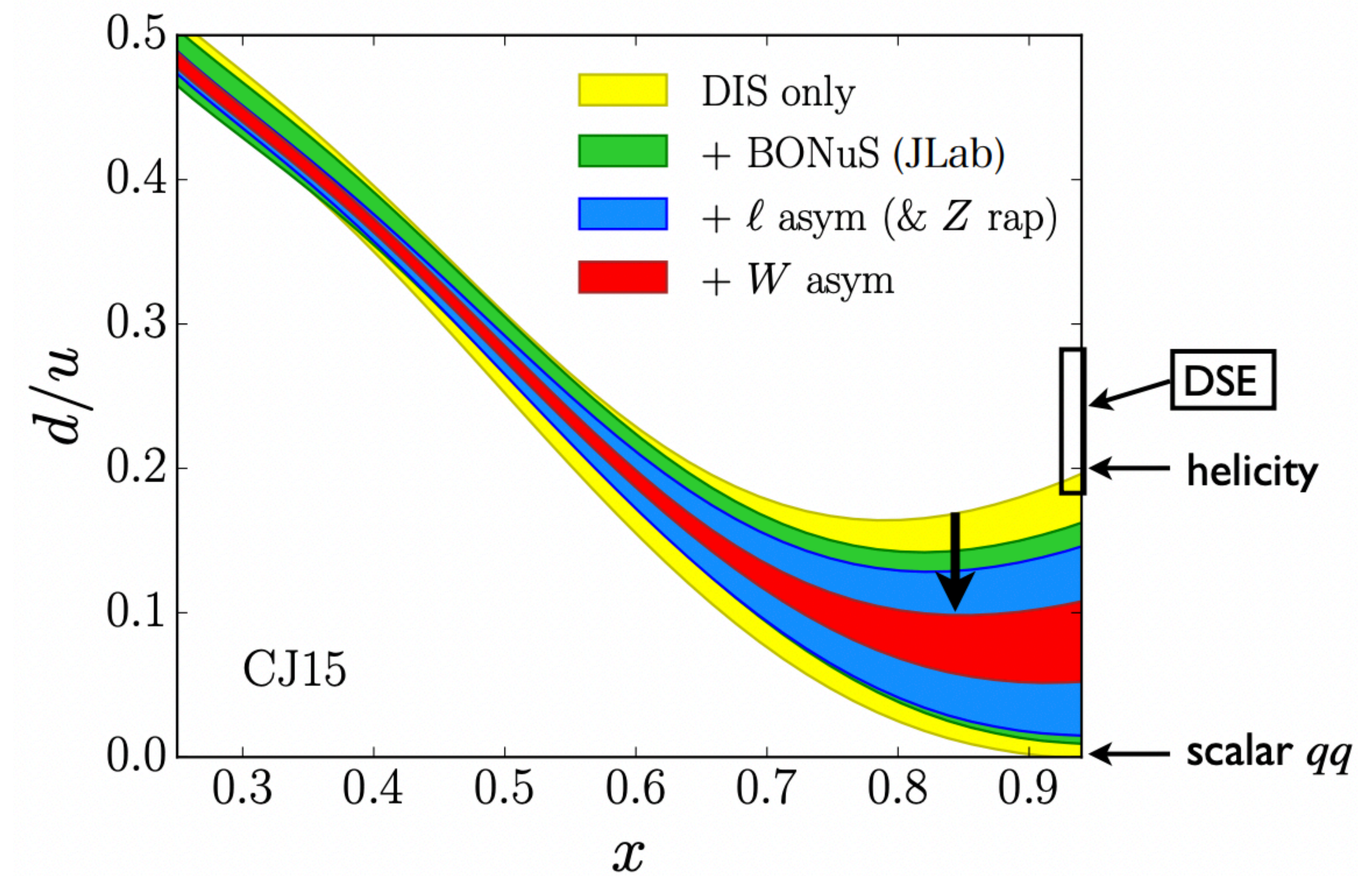
$$\Rightarrow \mathbf{u_n(x)=d_p(x)} \text{ and } \mathbf{d_n(x) = u_p(x)}$$

- Linear combinations of F_2^p and F_2^n give separately $u_p(x) \equiv u(x)$ and $d_p(x) \equiv d(x)$,

$$\boxed{\begin{gathered} F_2^p(x,Q^2) - F_2^d(x,Q^2) = \frac{1}{3}(u + \bar{u} - d - \bar{d}) \\ \frac{F_2^d(x)}{F_2^p(x)} \sim \frac{u}{d} \end{gathered}}$$

Fixed target DIS data





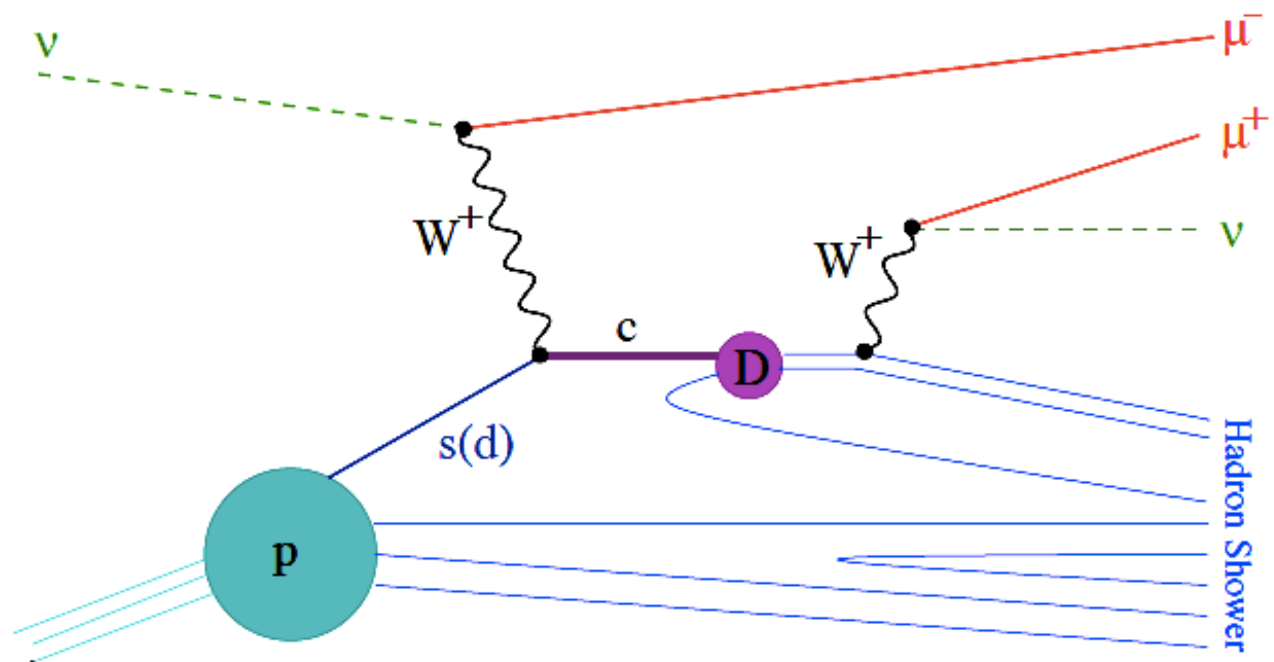
- Valence d/u ratio at high-x accessible thanks to low-Q2 fixed-target experiments (SLAC, BCDMS, NMC, CHORUS, NuTeV, JLAB)
- Testing ground for nucleon models in the x → 1 limit
- At high-x nuclear corrections are very important (deuteron target)
- Tevatron W asymmetry data and JLab tagged neutron help constraining d/u ratio up to large values of x ~ 0.85

Fixed target DIS neutrino

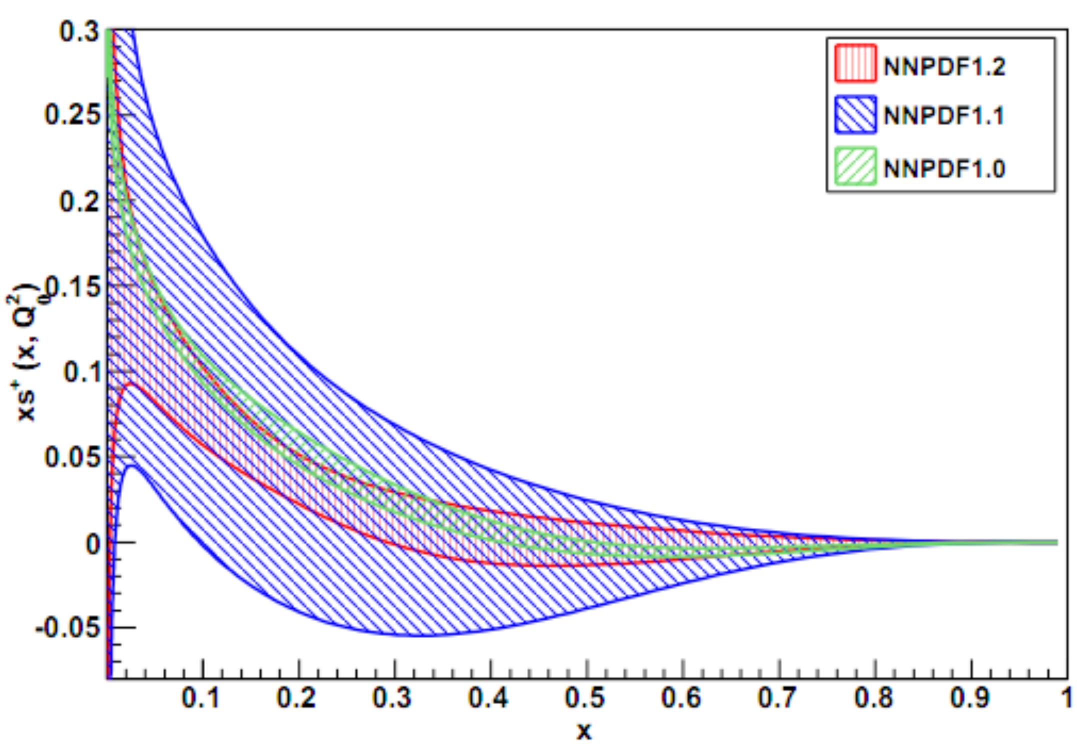
$$\tilde{\sigma}^{\nu(\bar{\nu}),c} \propto \left(F_2^{\nu(\bar{\nu}),c}, F_3^{\nu(\bar{\nu}),c}, F_L^{\nu(\bar{\nu}),c}\right)$$

$$F_2^{\nu,c} = x\left[C_{2,q}\otimes 2|V_{cs}|^2 s + \frac{1}{n_f}C_{2,g}\otimes g\right]$$

$$F_2^{\bar{\nu},c} = x\left[C_{2,q}\otimes 2|V_{cs}|^2 \bar{s} + \frac{1}{n_f}C_{2,g}\otimes g\right]$$



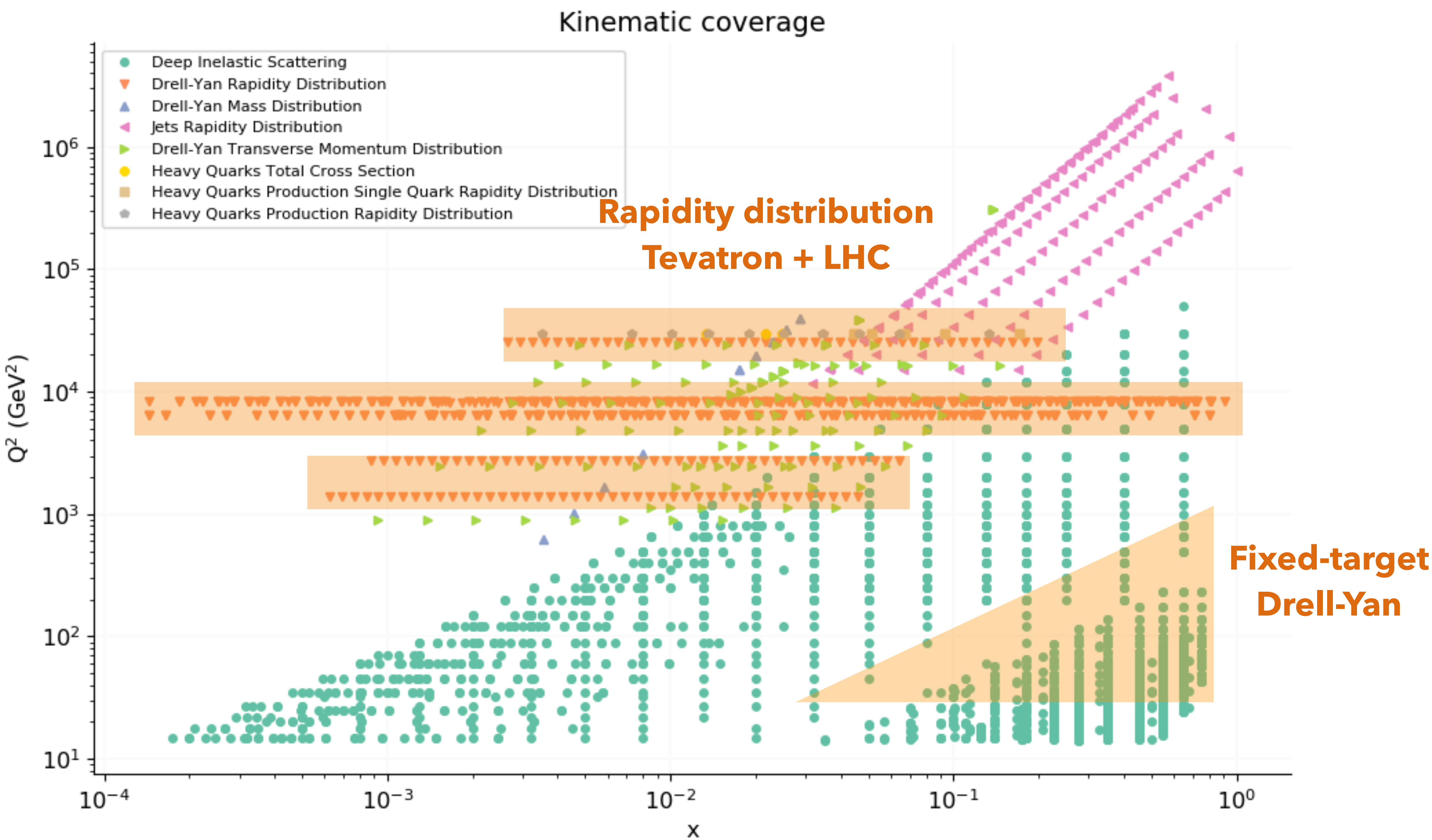
V_{cs} enhancement



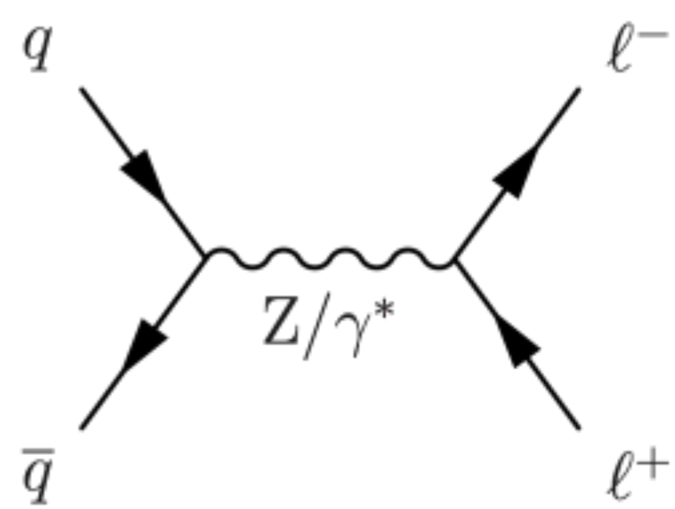
$$x = \frac{Q^2}{2M_n E_\nu y}$$

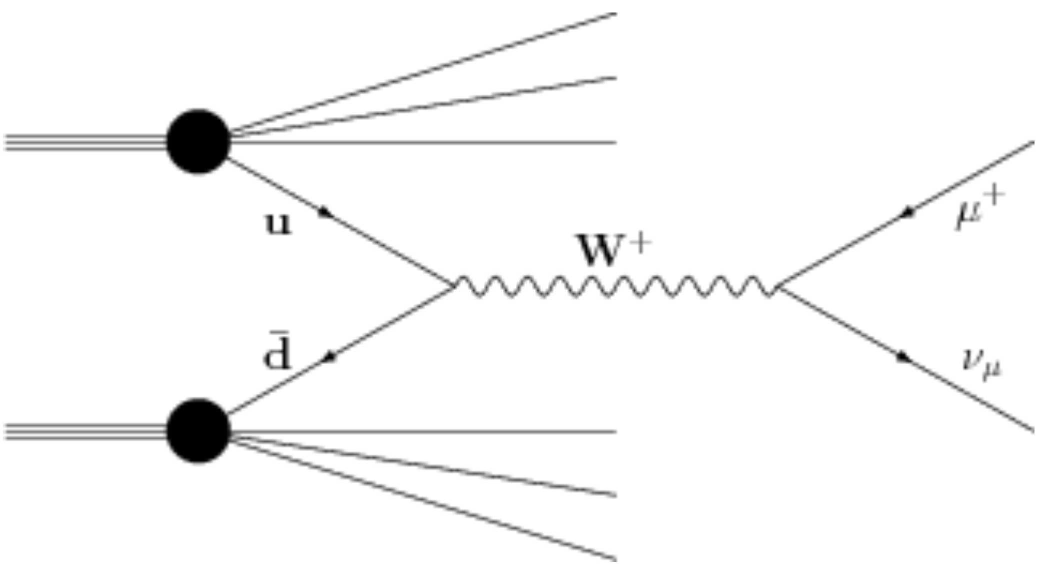
- NuTeV structure function data on isoscalar iron target provide strong constraints on strangeness inside the proton
- Some (mild) tension between fixed target data and W+c data at the LHC

Drell-Yan/V production data



Drell-Yan/V production data





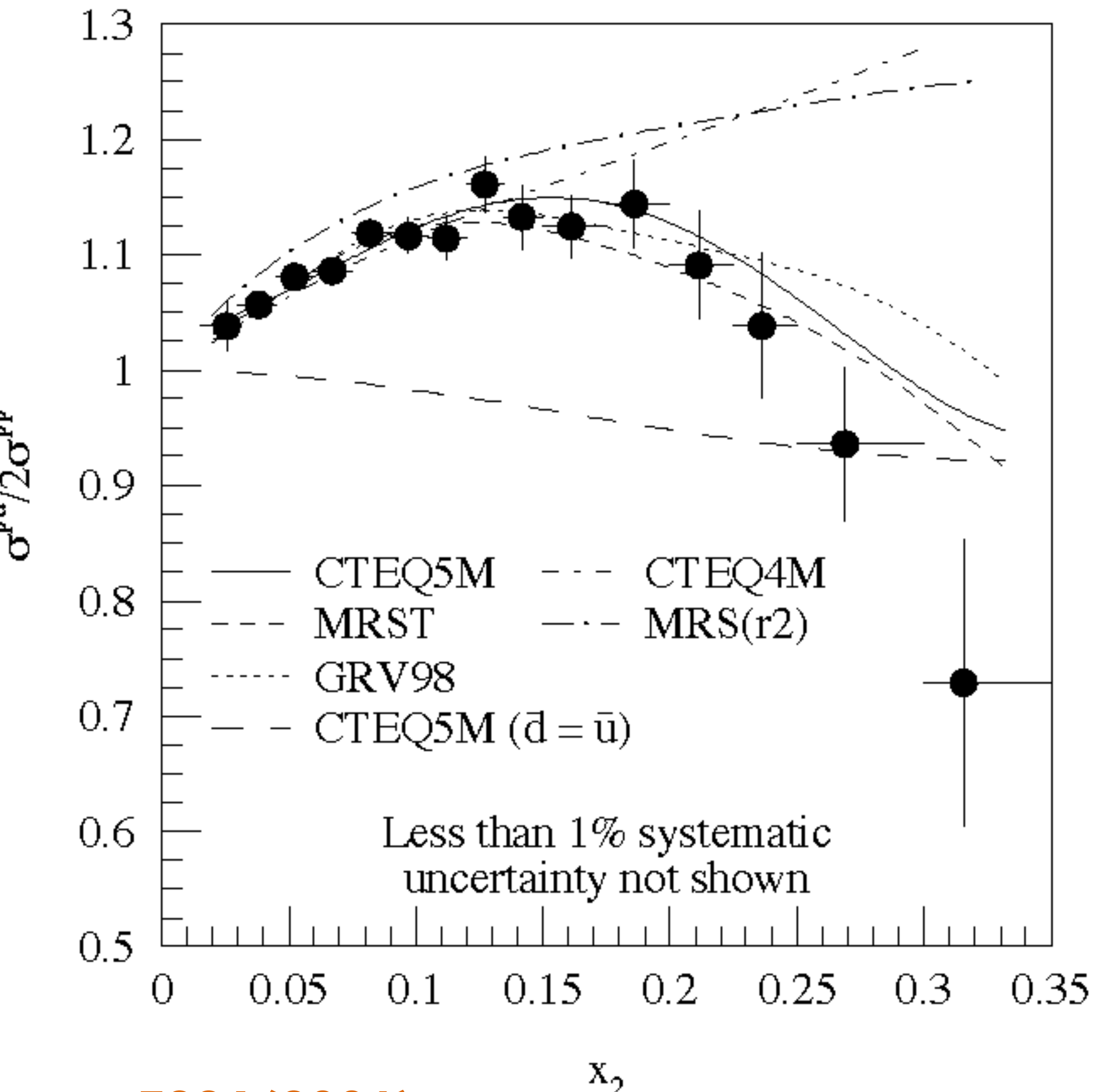
$$L_{ij}(x_1, x_2) = q_i(x_1)\bar{q}_j(x_2)$$

$$\gamma^* : \quad \frac{d\sigma}{dydM^2} = \frac{4\pi\alpha^2}{9M^2S}\sum_i e_i^2 L_{ij}(x_1, x_2)$$

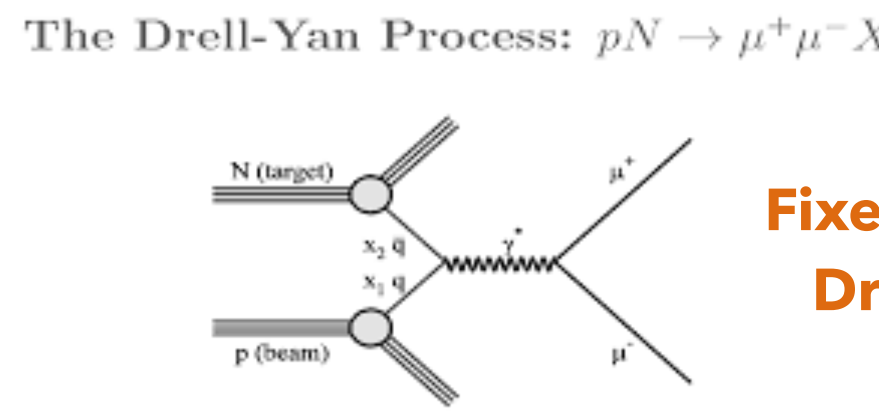
$$Z : \quad \frac{d\sigma}{dy} = \frac{\pi G_F M_V^2 \sqrt{2}}{3S}\sum_i (v_{iZ}^2 + a_{iZ}^2) L_{ij}(x_1, x_2)$$

$$W : \quad \frac{d\sigma}{dydM^2} = \frac{\pi G_F M_V^2 \sqrt{2}}{3S}\sum_{ij} |V_{ij}^{\mathrm{CKM}}|^2 L_{ij}(x_1, x_2)$$

Drell-Yan data

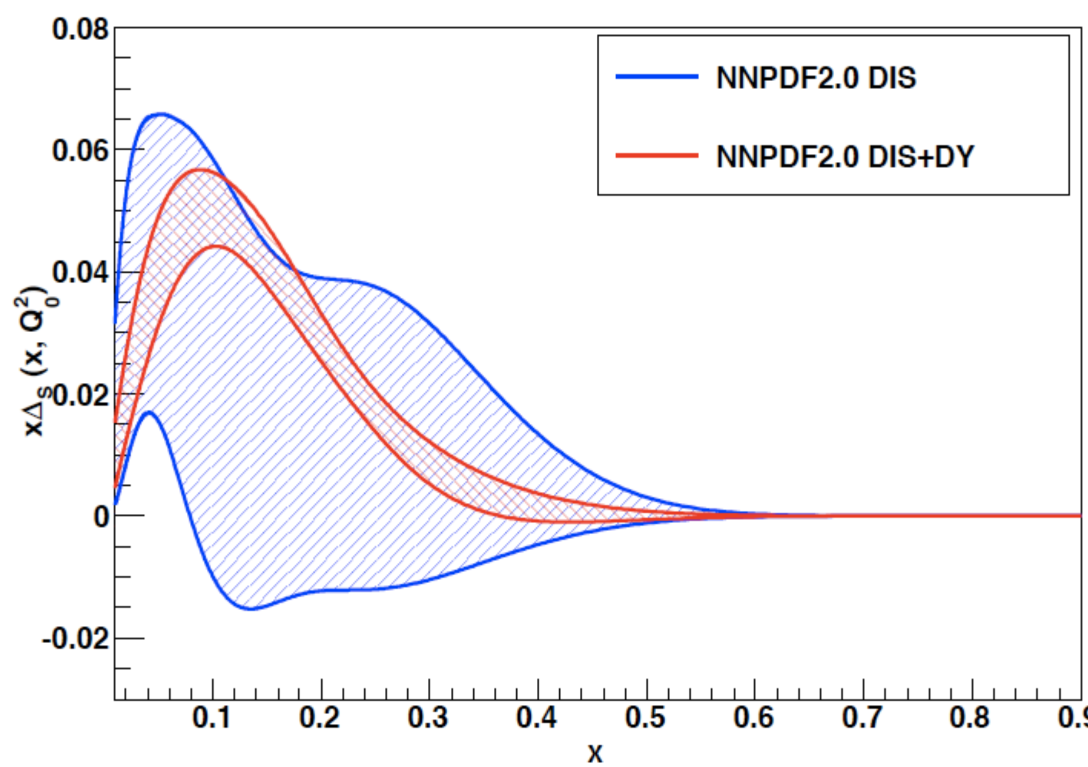


E886 (2001)

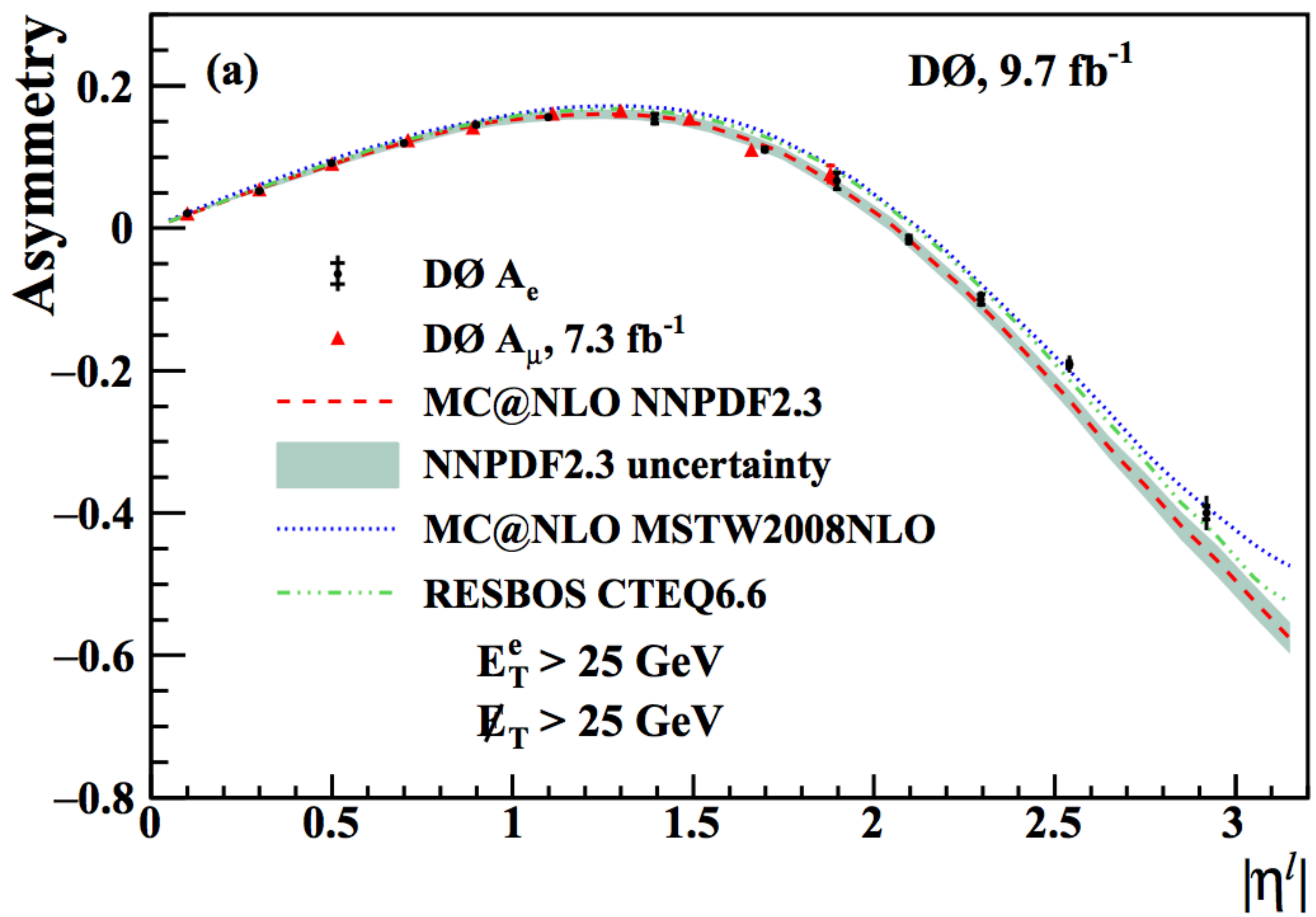


Fixed-target Drell-Yan

$$\frac{\sigma(pd \to \mu^+\mu^-)}{\sigma(pp \to \mu^+\mu^-)} = \frac{\frac{4}{9}u\bar{d} + \frac{1}{9}d\bar{u}}{\frac{4}{9}u\bar{u} + \frac{1}{9}d\bar{d}} \sim \frac{\bar{d}}{\bar{u}}$$



Z/W production data



W asymmetry at Tevatron

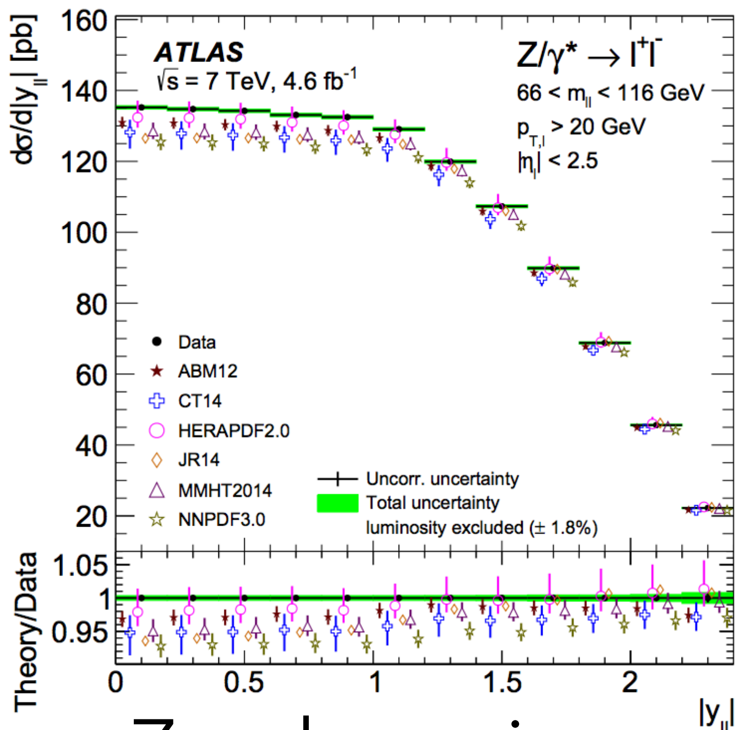
$$\boxed{\begin{aligned} u^{\bar p} &= \bar u^p \\ d^{\bar p} &= \bar d^p \end{aligned}}$$

Charge conjugation

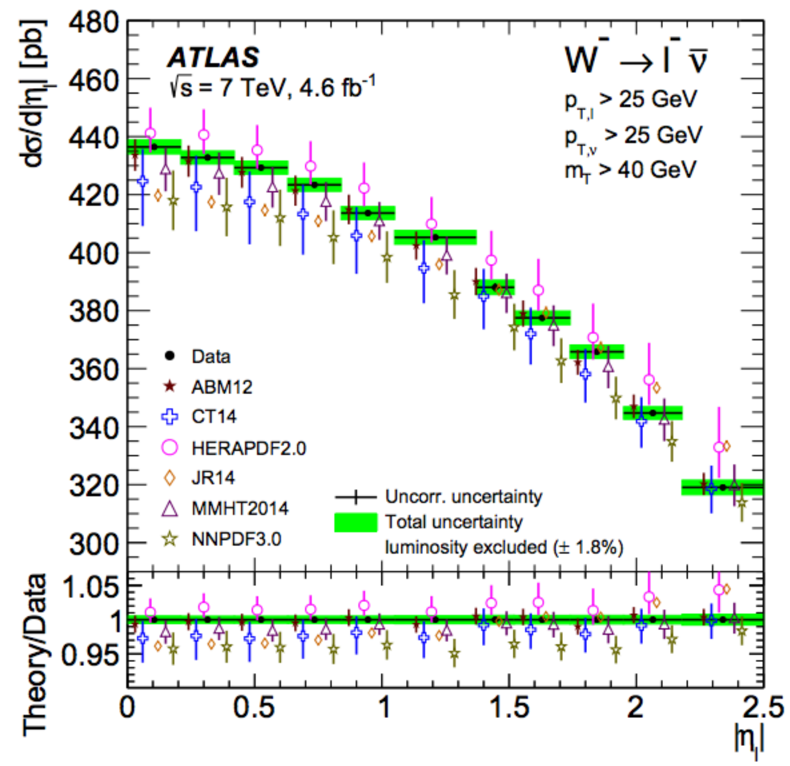
$$\frac{\sigma(p\bar p \to W^+)}{\sigma(p\bar p \to W^-)} = \frac{u(x_1)d(x_2) + \bar u(x_1)\bar d(x_2)}{d(x_1)u(x_2) + \bar d(x_1)\bar u(x_2)} \sim \frac{u}{d}(x_1)\frac{u}{d}(x_2)$$

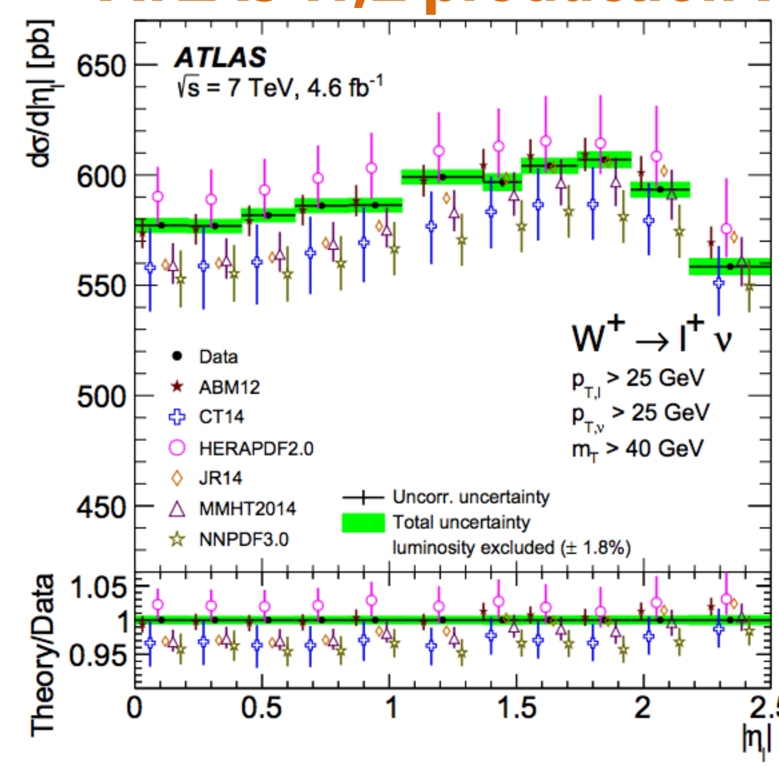
Z/W production data

ATLAS W,Z production 7 TeV



Z pole region





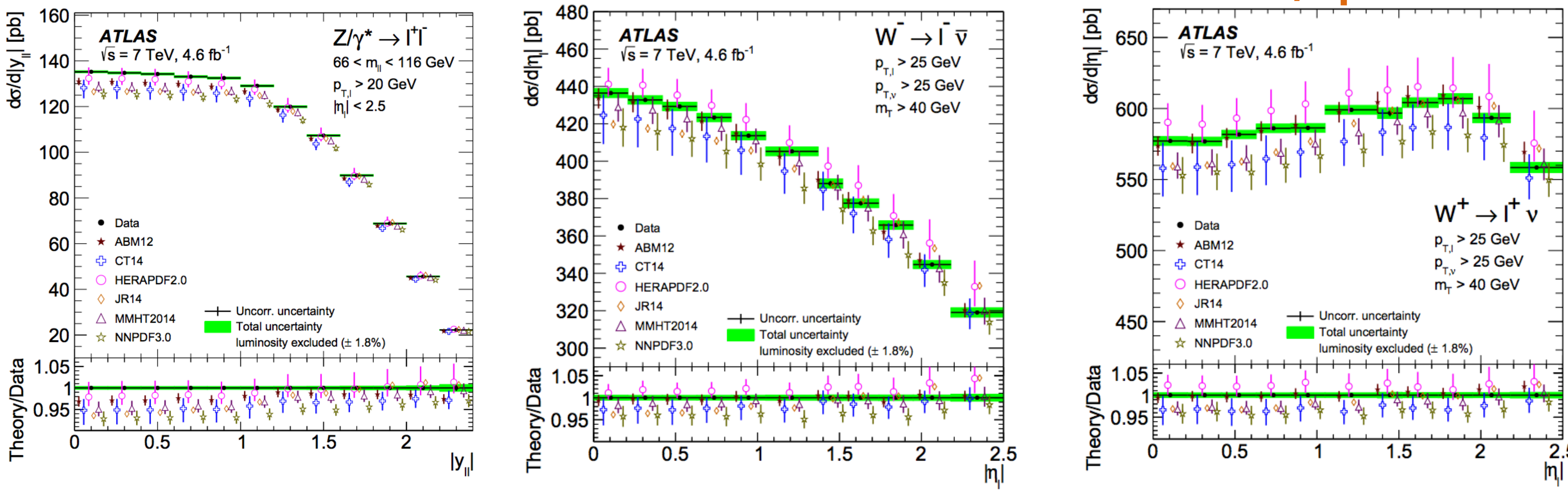
$$\sigma(pp \to Z) = u\bar{u} + d\bar{d} + s\bar{s}$$

$$\sigma(pp \to W^+) = u\bar{d} + c\bar{s}$$

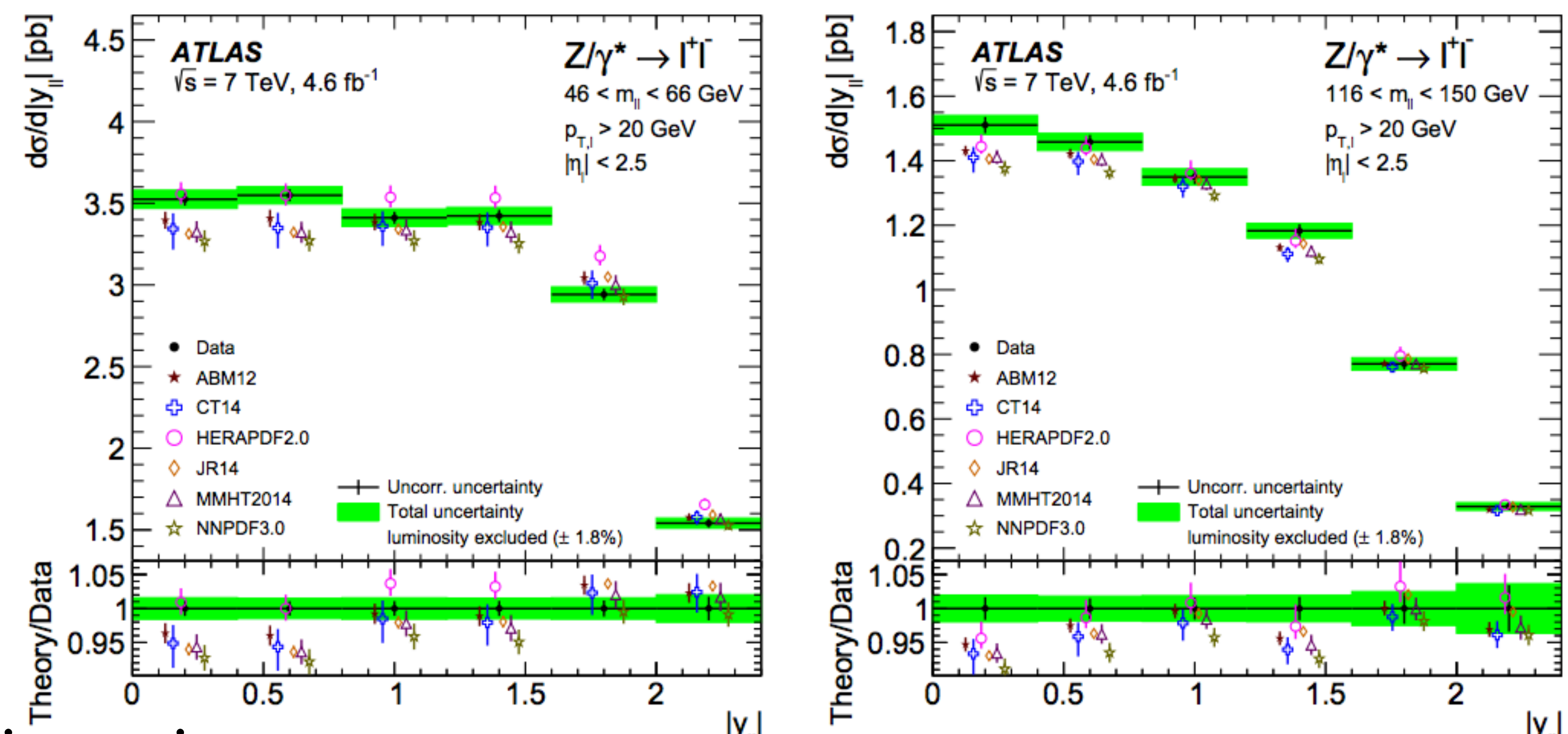
$$\sigma(pp \to W^-) = d\bar{u} + s\bar{c}$$

Z/W production data

ATLAS W,Z production 7 TeV

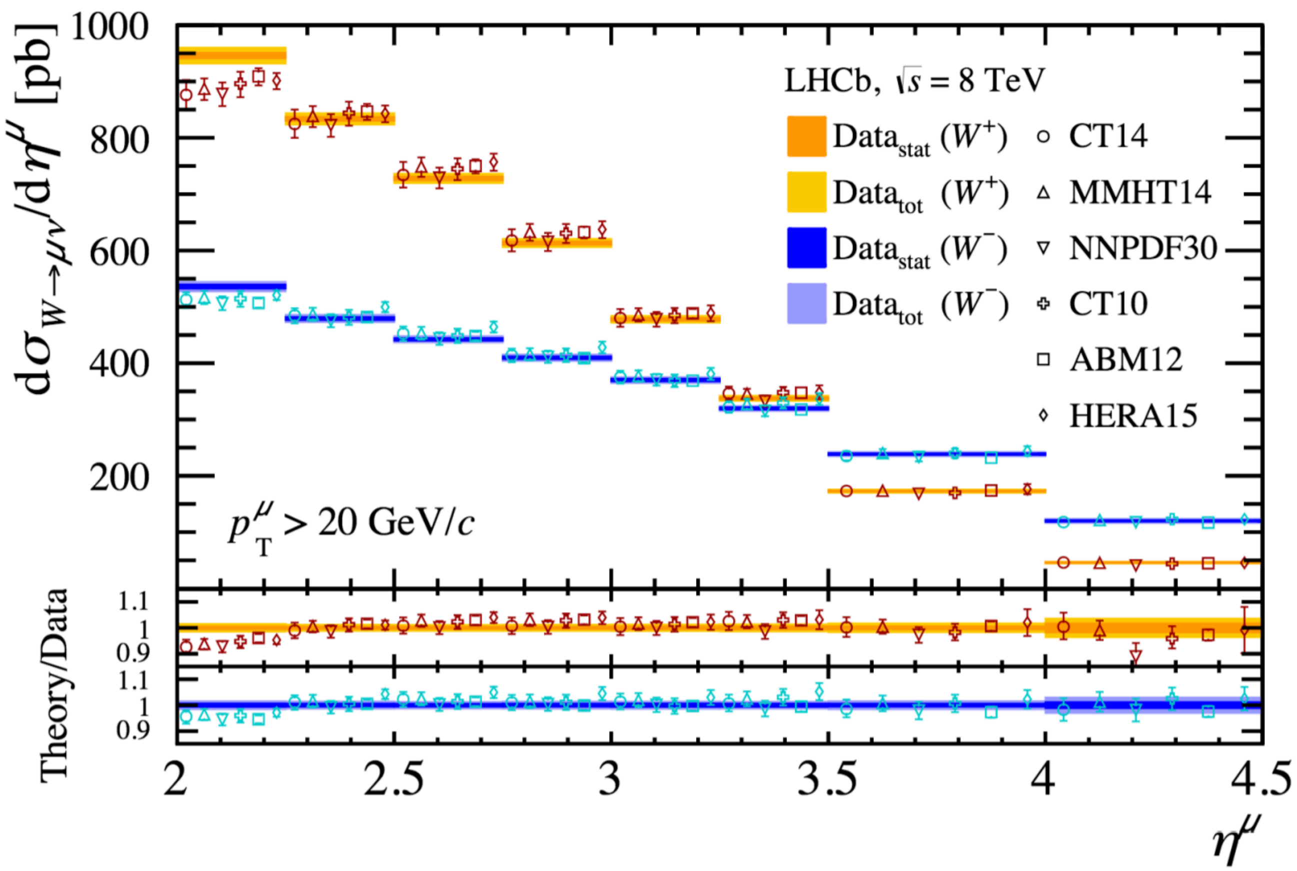


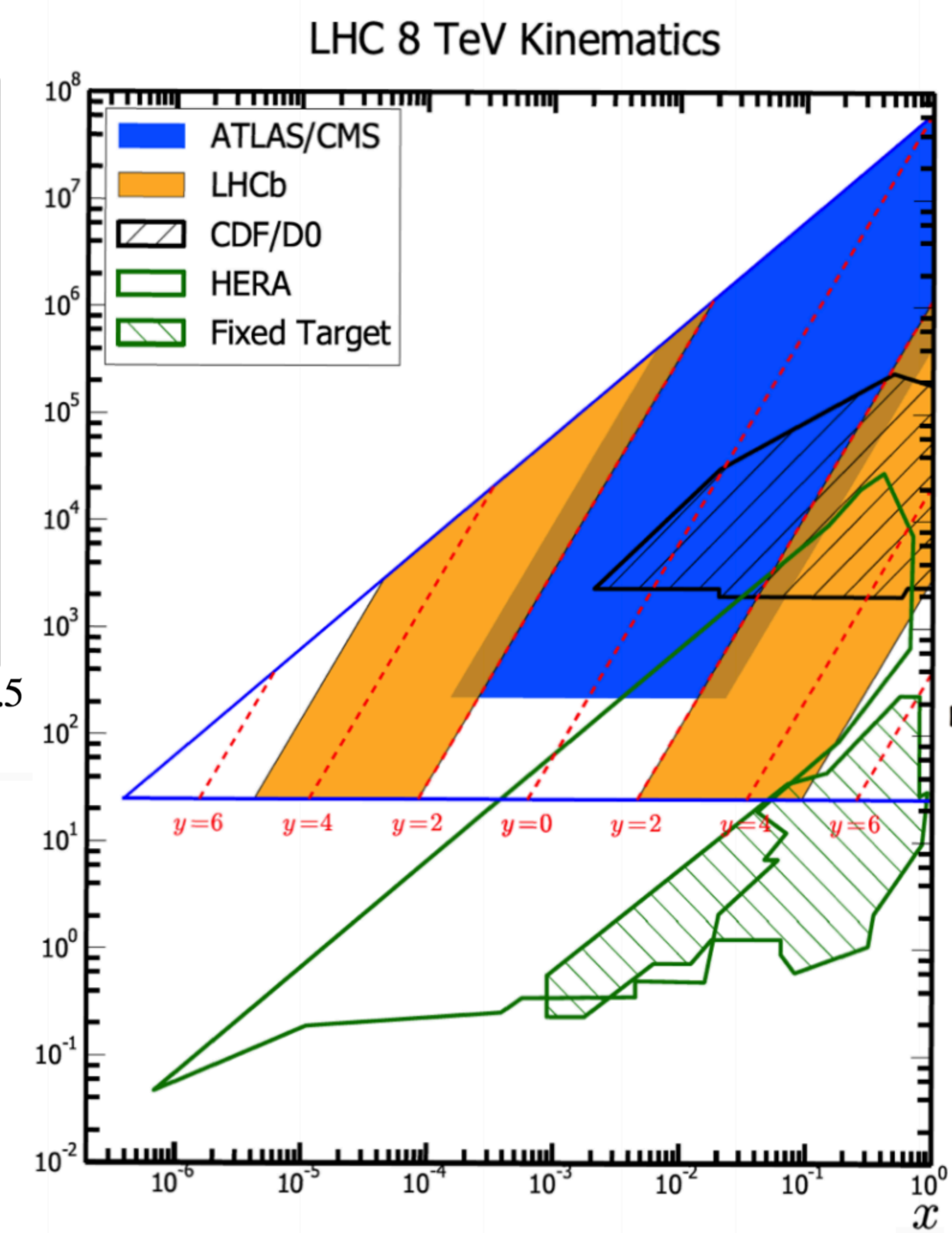
$$x_{1,2} = \frac{M^2}{s} e^{\pm y}$$



high/low invariant mass

Z/W production data

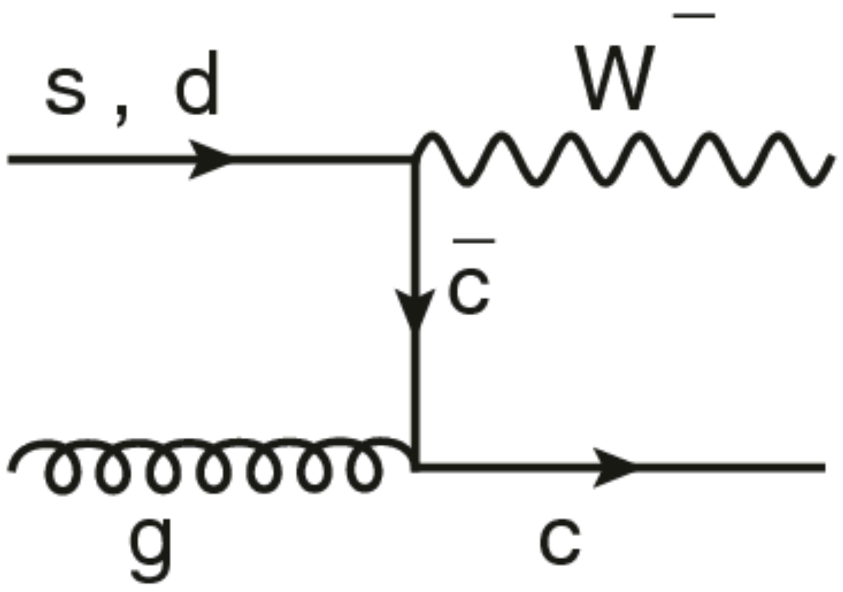


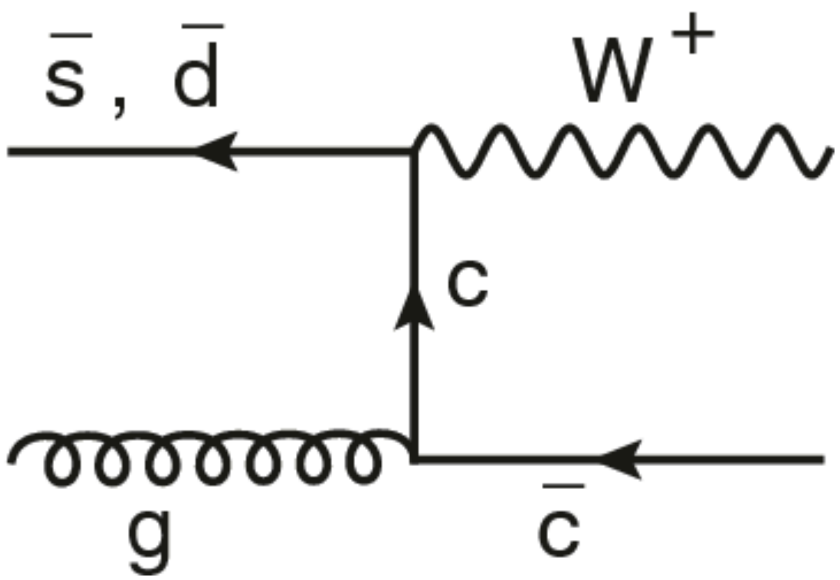


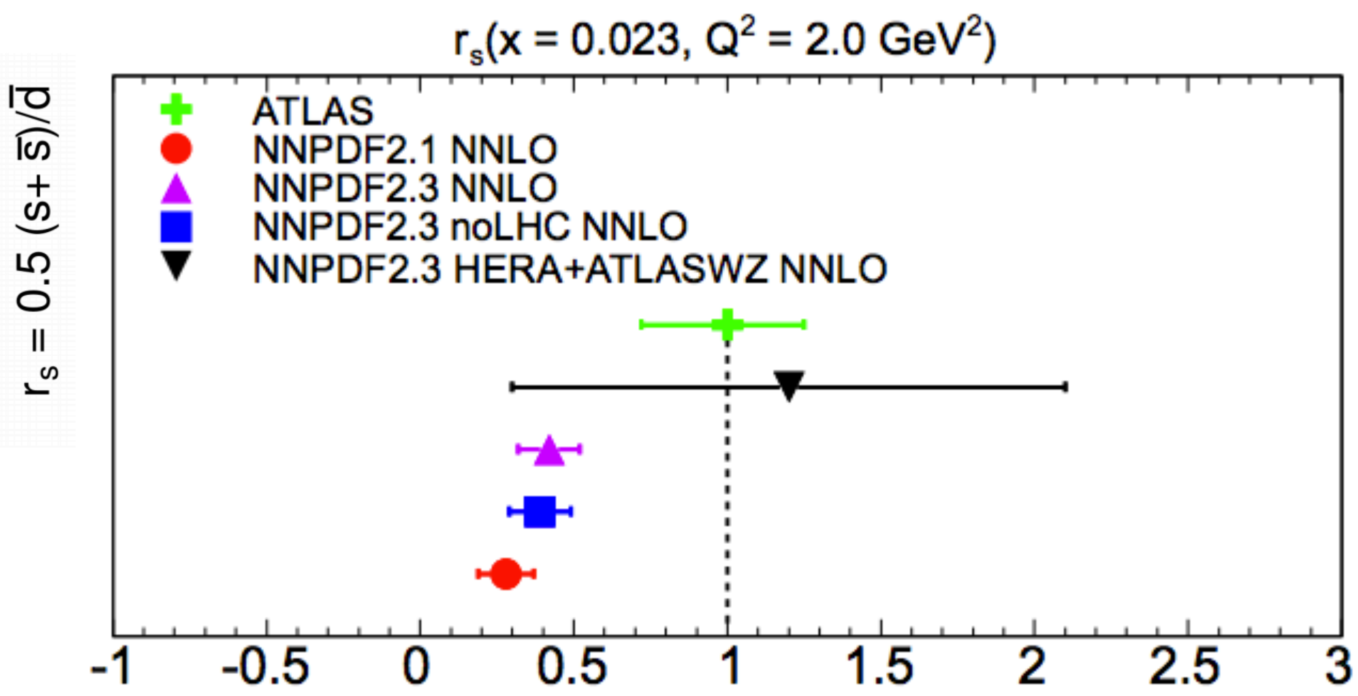
$$x_{1,2} = \frac{M^2}{s} e^{\pm y}$$

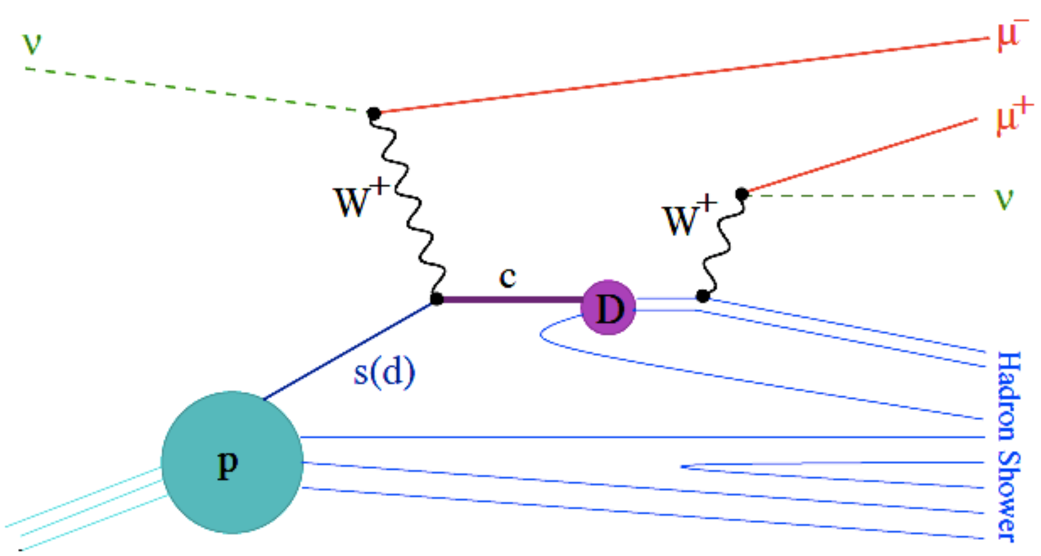
LHCb: forward region

W+charm data


s , d
W$^-$
$\bar{c}$
g
c


$\bar{s}$, $\bar{d}$
W$^+$
c
g
$\bar{c}$

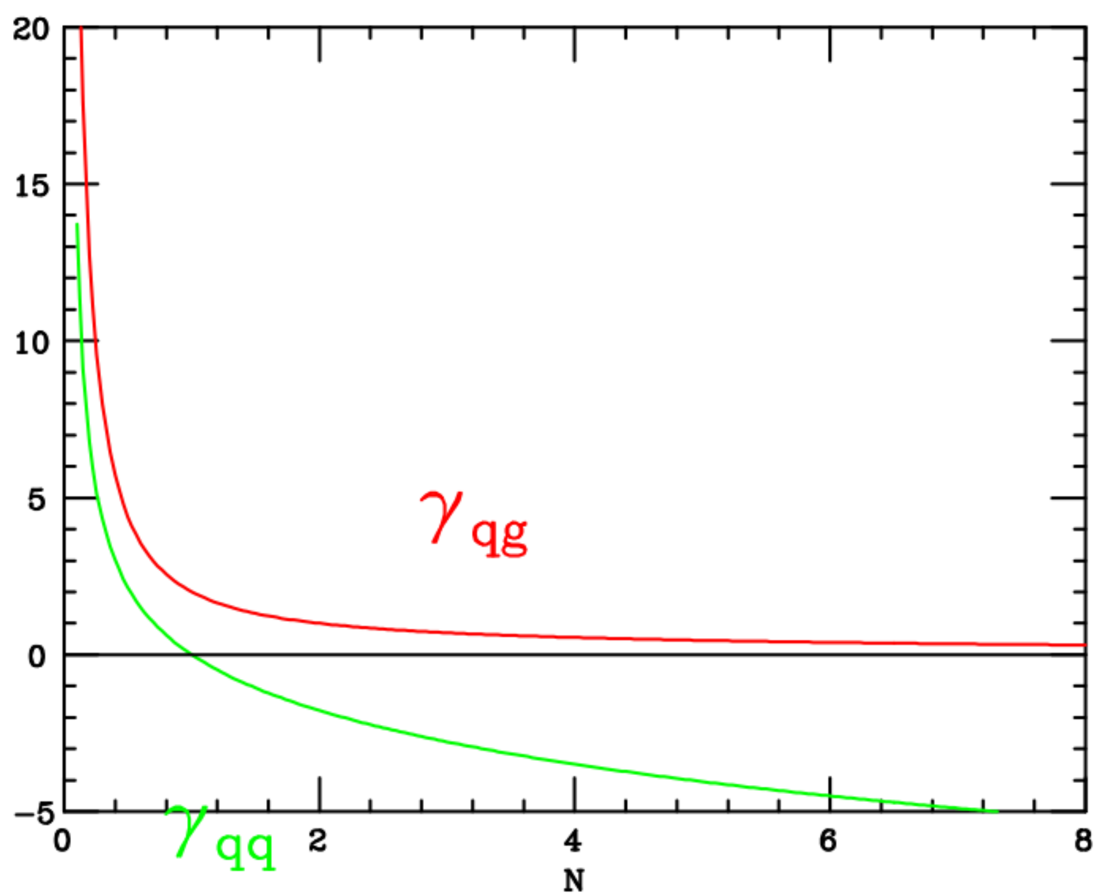

$r_s(x = 0.023, Q^2 = 2.0\ \mathrm{GeV}^2)$
$r_s = 0.5\,(s+\bar{s})/\bar{d}$
ATLAS
NNPDF2.1 NNLO
NNPDF2.3 NNLO
NNPDF2.3 noLHC NNLO
NNPDF2.3 HERA+ATLASWZ NNLO
-1
-0.5
0
0.5
1
1.5
2
2.5
3


ν
μ⁻
μ⁺
W⁺
W⁺
ν
c
D
s(d)
p
Hadron Shower

Gluon: indirect handle

- Gluon is partially determined by scale dependence of DIS structure functions and Drell-Yan/Vector Boson production

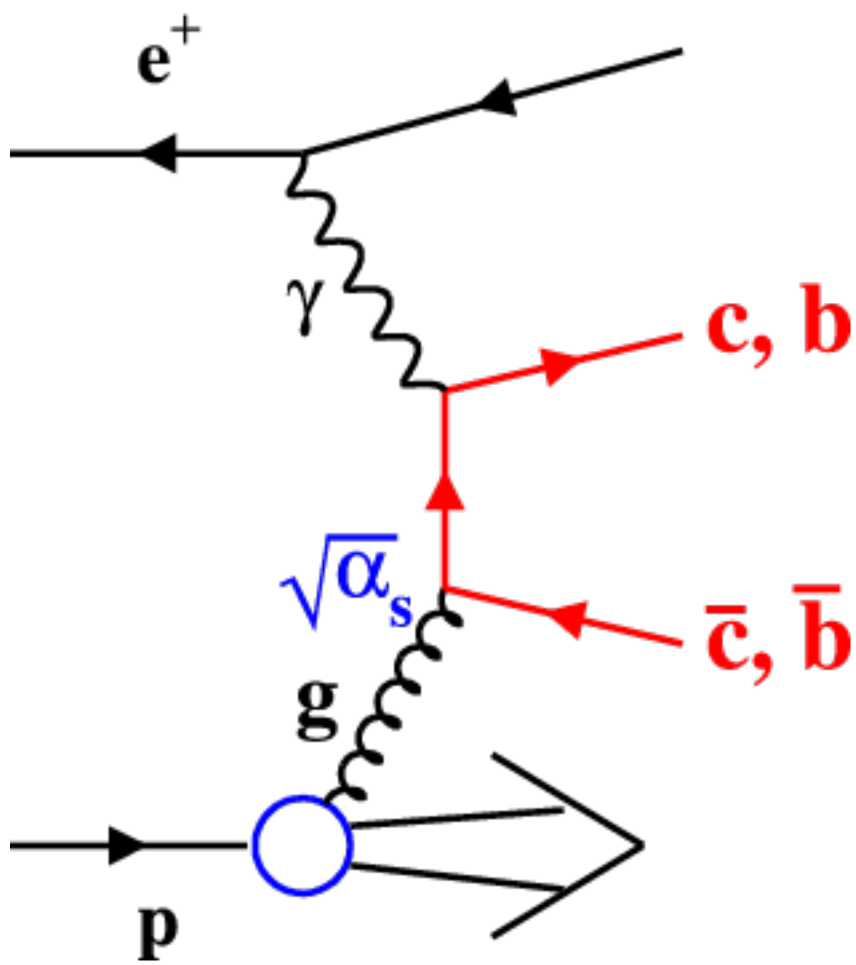
$$\frac{d}{d\log\mu^2}F_2(x,\mu^2)=\frac{\alpha_s(\mu^2)}{2\pi}\left[P_{qq}\otimes F_2(x,\mu^2)+2n_f P_{qg}\otimes g(x,\mu^2)\right]$$

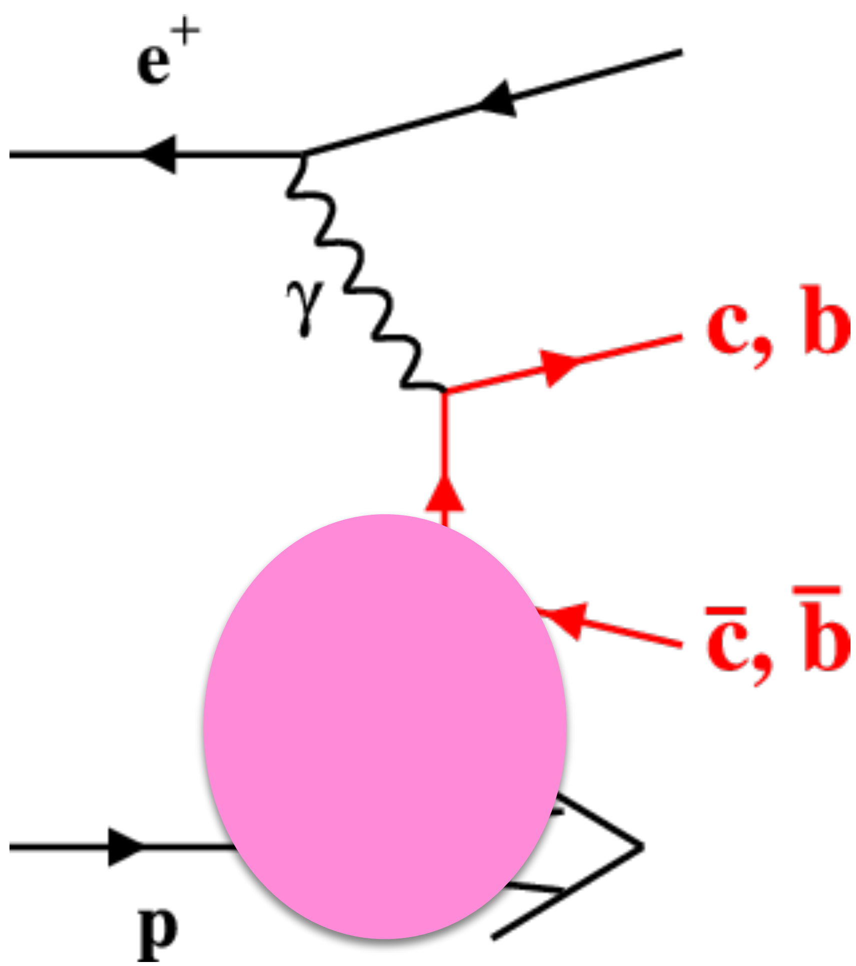


- Mostly determine small-x gluon, large-x gluon hard to determine from DIS+DY only data

Gluon: indirect handle

- Heavy quarks are produced at threshold inside proton
- Heavy quark production process (at ep and pp colliders) probe gluon
- Dependence on heavy flavour scheme adopted in PDF fitting



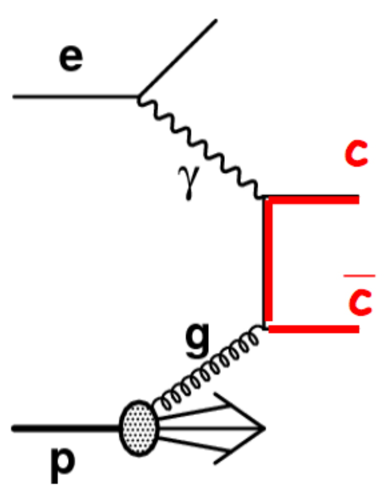


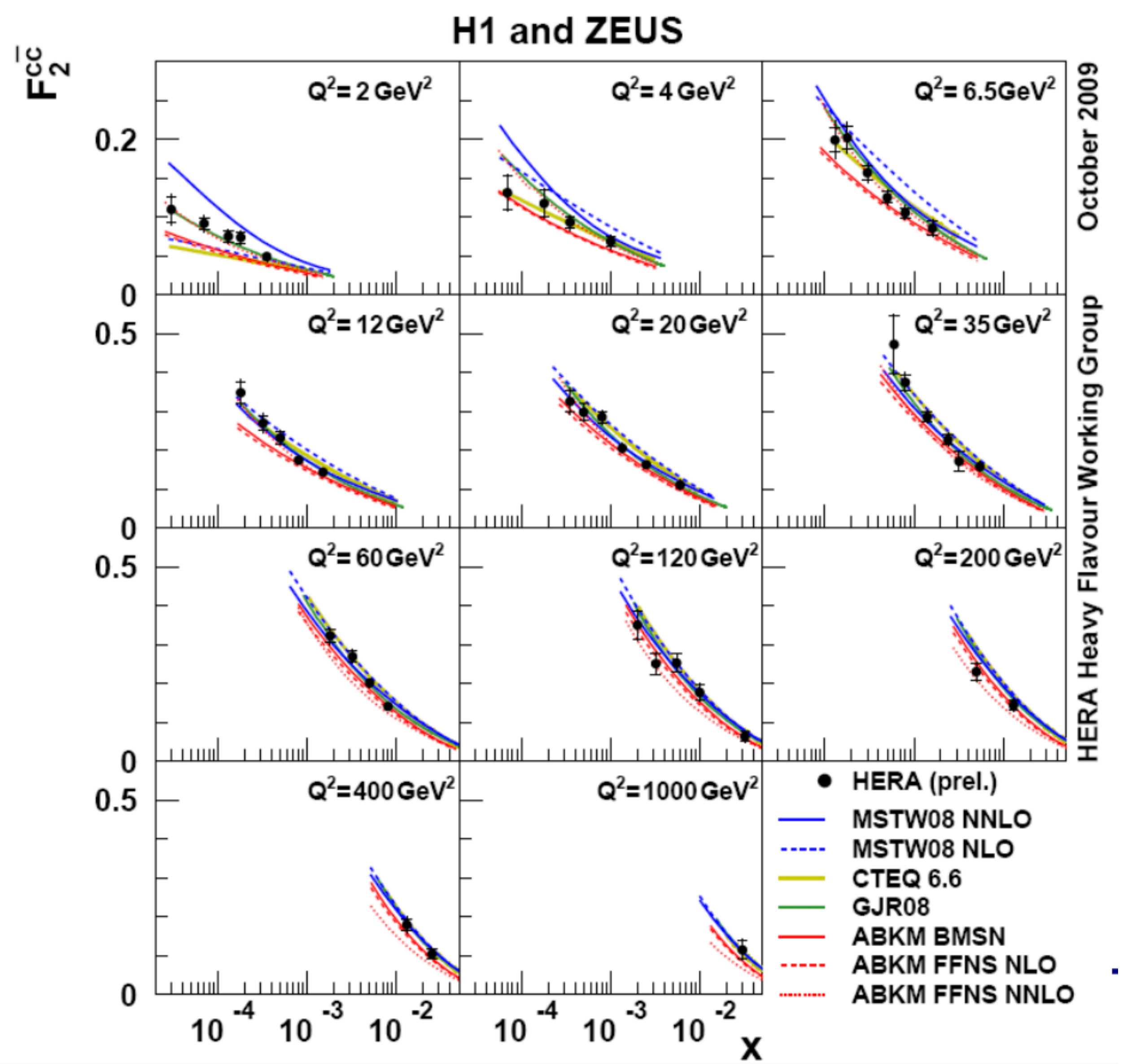
Nf = 3,4

Nf = 5

Gluon: indirect handle

- Heavy quarks are produced at threshold inside proton
- Heavy quark production process (at ep and pp colliders) probe gluon
- Dependence on heavy flavour scheme adopted in PDF fitting





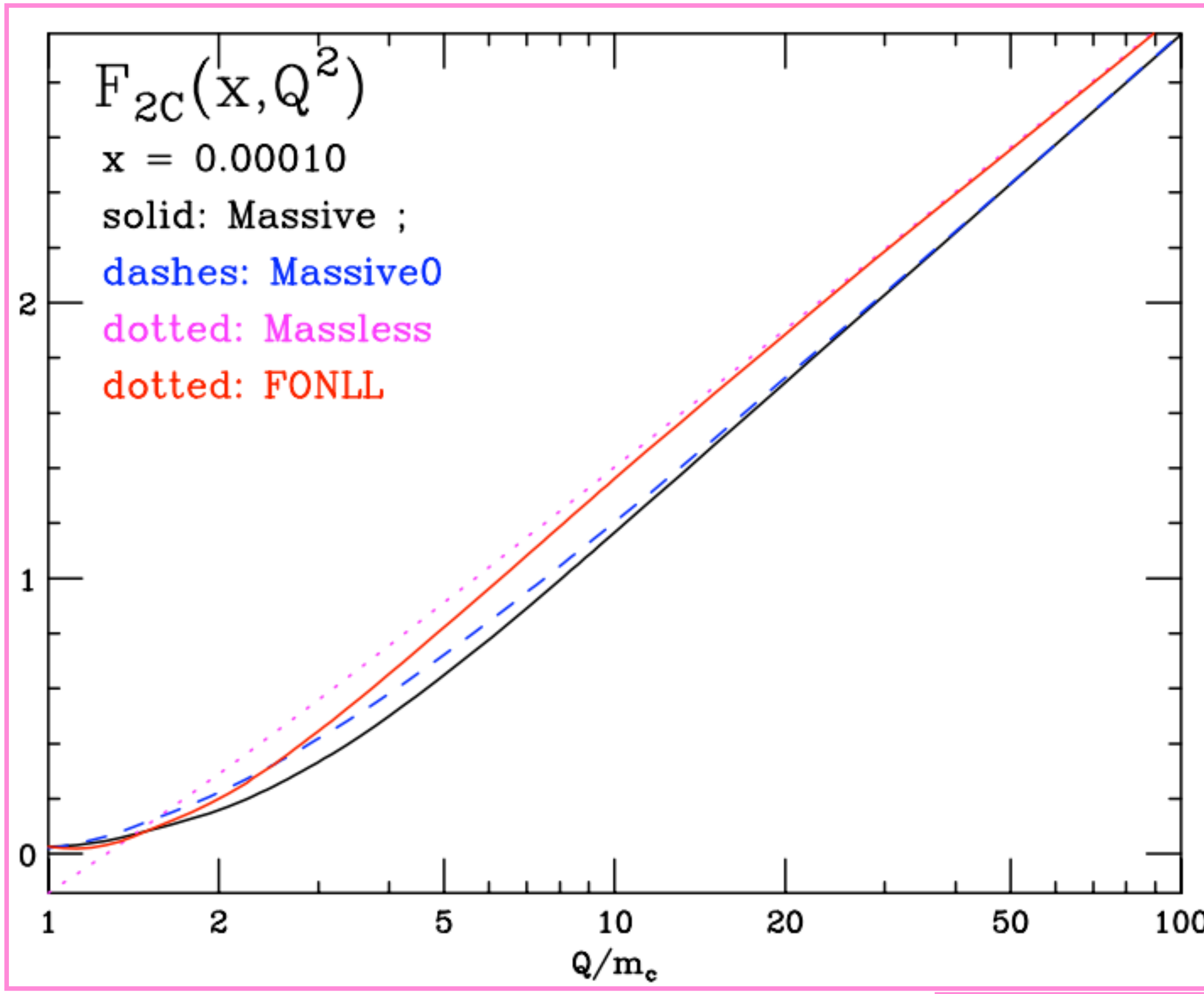
Intermission: heavy flavour schemes

- Charm, Bottom and Top have mass >> Λ_{QCD} - heavy quarks (HQ)
- The presence of a new scale, m_Q, makes pert QCD calculations more challenging
- Two well understood schemes:
    - Assume heavy quark effectively massless for Q > m_Q
      HQ becomes active massless parton above threshold
    - Heavy quarks retain their mass for all Q
      HQ is not a parton, it is a final state particle
- However in PDF fits we have all scales. General-Mass Variable-Flavor-Number schemes allow to match between the zero-mass and the massive scheme
- Many schemes available

e.g. FONLL

$$
\begin{aligned}
\sigma^{(FONLL)} &= \sigma^{(4)} + \sigma^{(5)} - \text{ double counting} \\
&= \mathcal{L}_{ij}(x_1, x_2, \mu^2) \otimes \sum_p^N \left(\alpha_S^{(5)}(\mu^2)\right)^p \\
&\times \left\{ \mathcal{B}_{ij}^{(p)}\left(x_1, x_2, \frac{\mu^2}{m_b^2}\right) + \sum_{k=0}^{\infty} \mathcal{A}_{ij}^{(p),(k)}(x_1, x_2)\left(\alpha_S^{(5)}(\mu^2) L\right)^k \right\} \\
&- \text{ double counting}
\end{aligned}
$$

Intermission: heavy flavour schemes



- heavy quarks (HQ)

D calculations more challenging

less for Q > mQ
above threshold

Q2
article

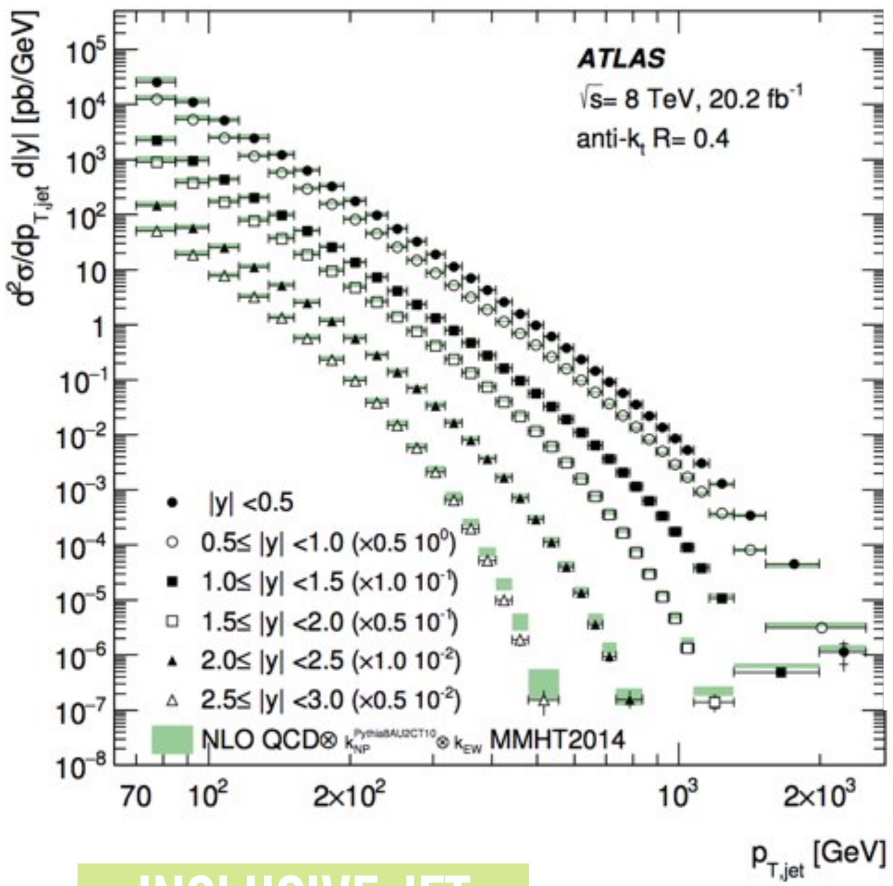
l-Mass Variable-Flavor-Number
and the massive scheme

- Many schemes available

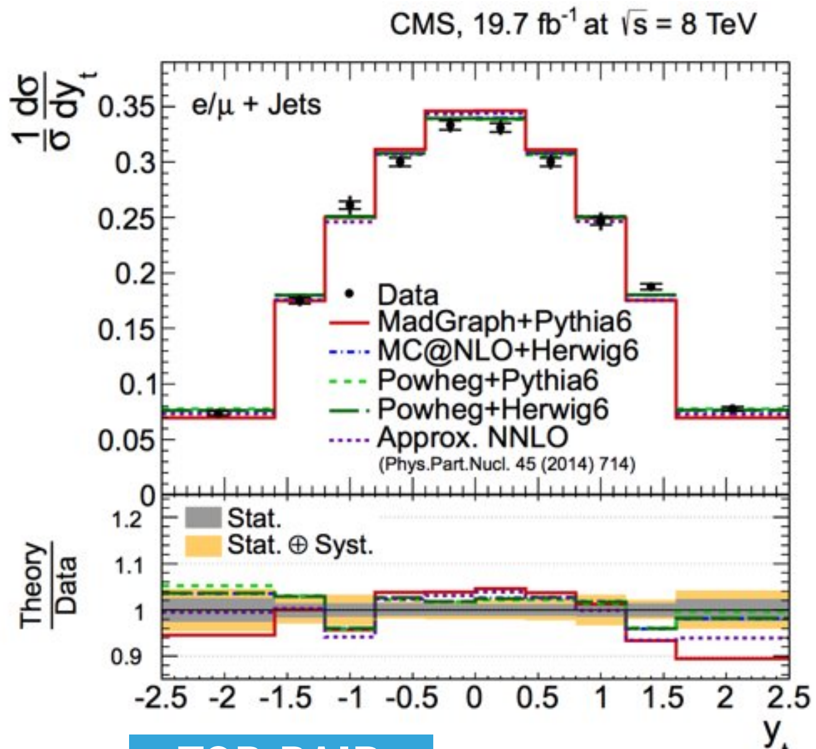
e.g. FONLL

$$\sigma^{(FONLL)} = \sigma^{(4)} + \sigma^{(5)} - \text{double counting}$$
$$= \mathcal{L}_{ij}(x_1, x_2, \mu^2) \otimes \sum_{p}^{N} \left(\alpha_S^{(5)}(\mu^2)\right)^p$$
$$\times \left\{ \mathcal{B}_{ij}^{(p)}\left(x_1, x_2, \frac{\mu^2}{m_b^2}\right) + \sum_{k=0}^{\infty} \mathcal{A}_{ij}^{(p),(k)}(x_1, x_2)\left(\alpha_S^{(5)}(\mu^2) L\right)^k \right\}$$
$$- \text{double counting}$$

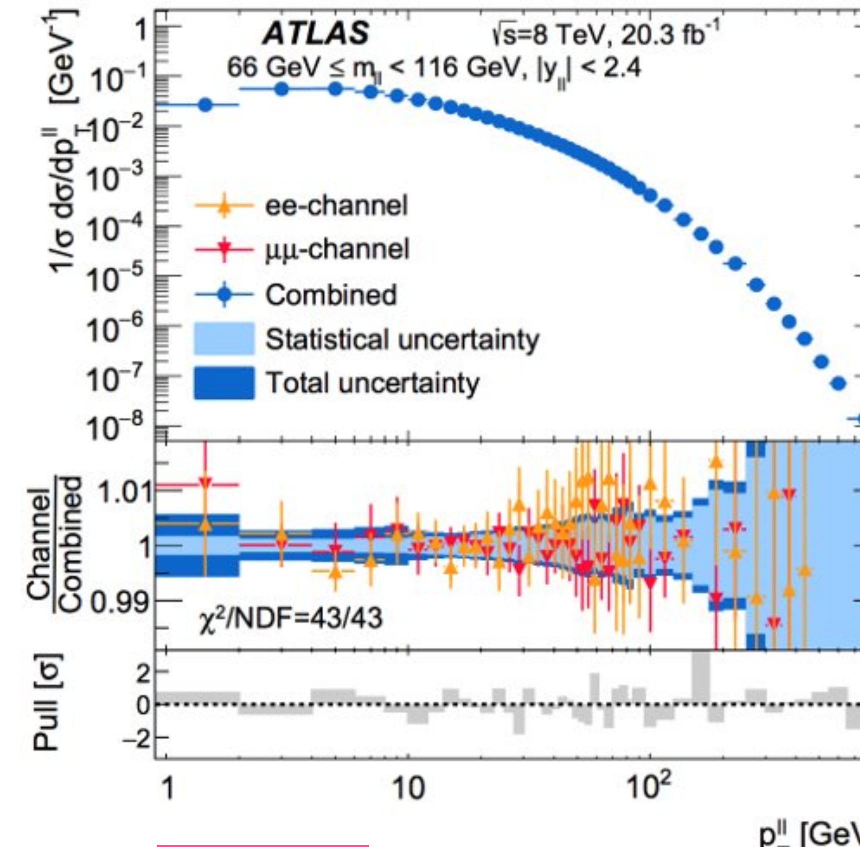
Gluon: direct handle



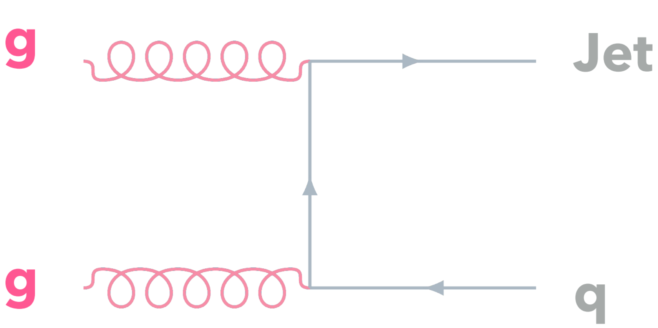
INCLUSIVE JET

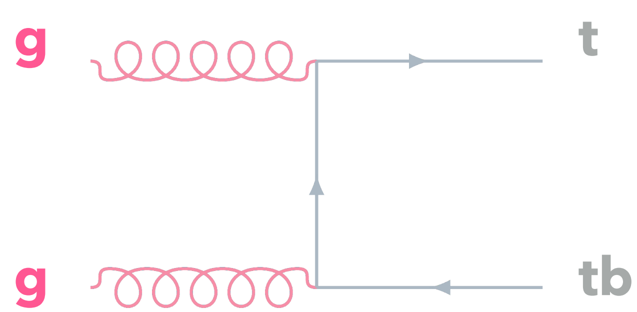


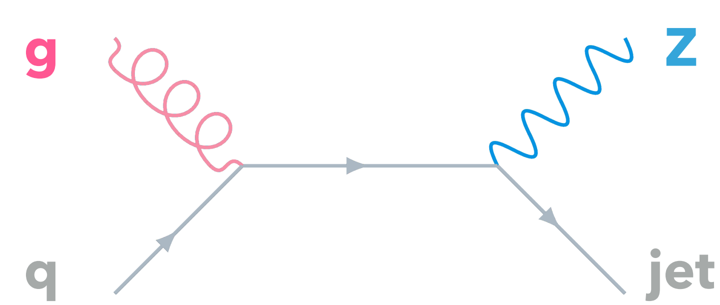
TOP PAIR



Z P_T

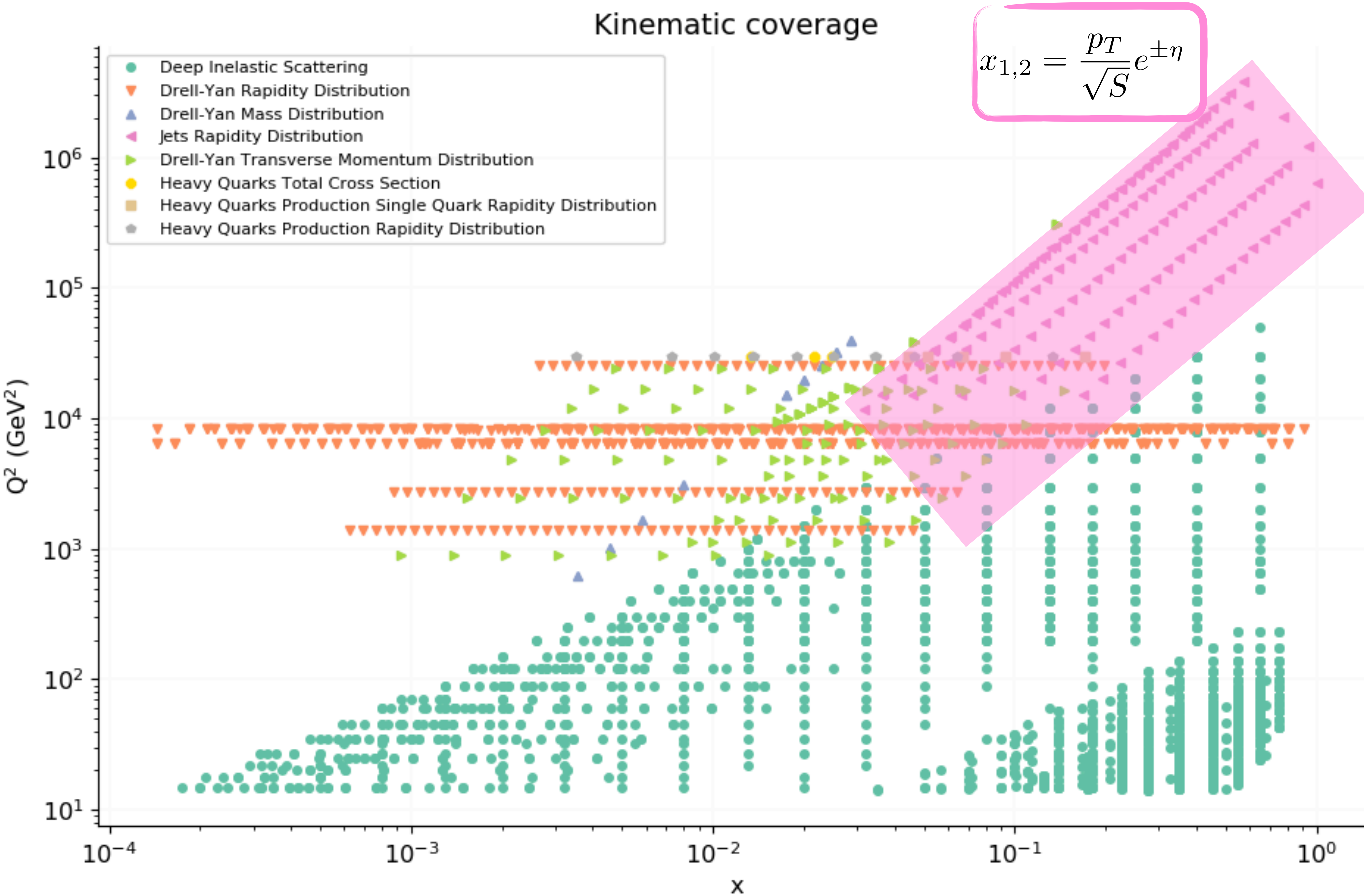




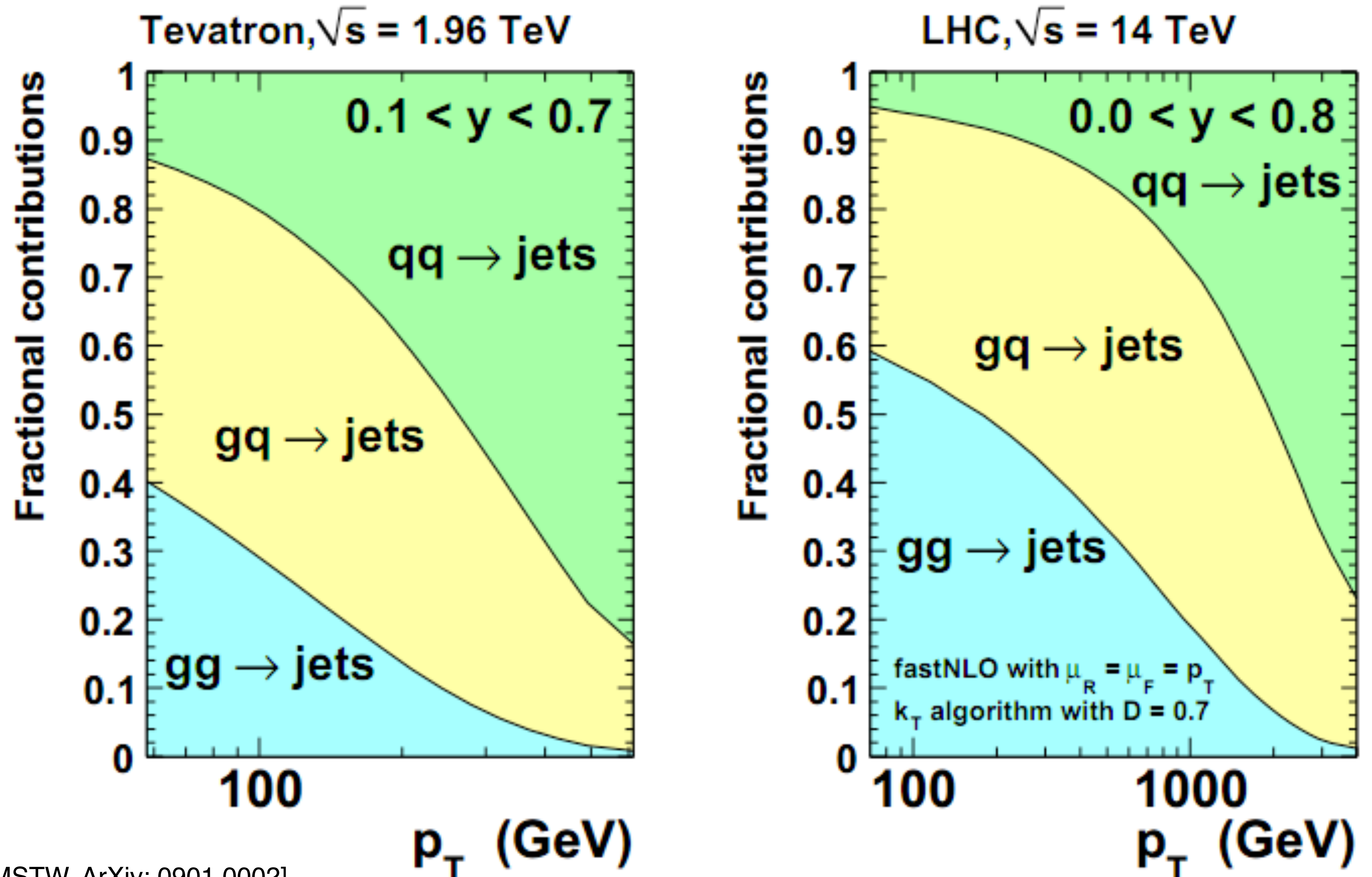


Gluon: jets data

$$\boxed{x_{1,2} = \frac{p_T}{\sqrt{S}} e^{\pm\eta}}$$



Gluon: jets data

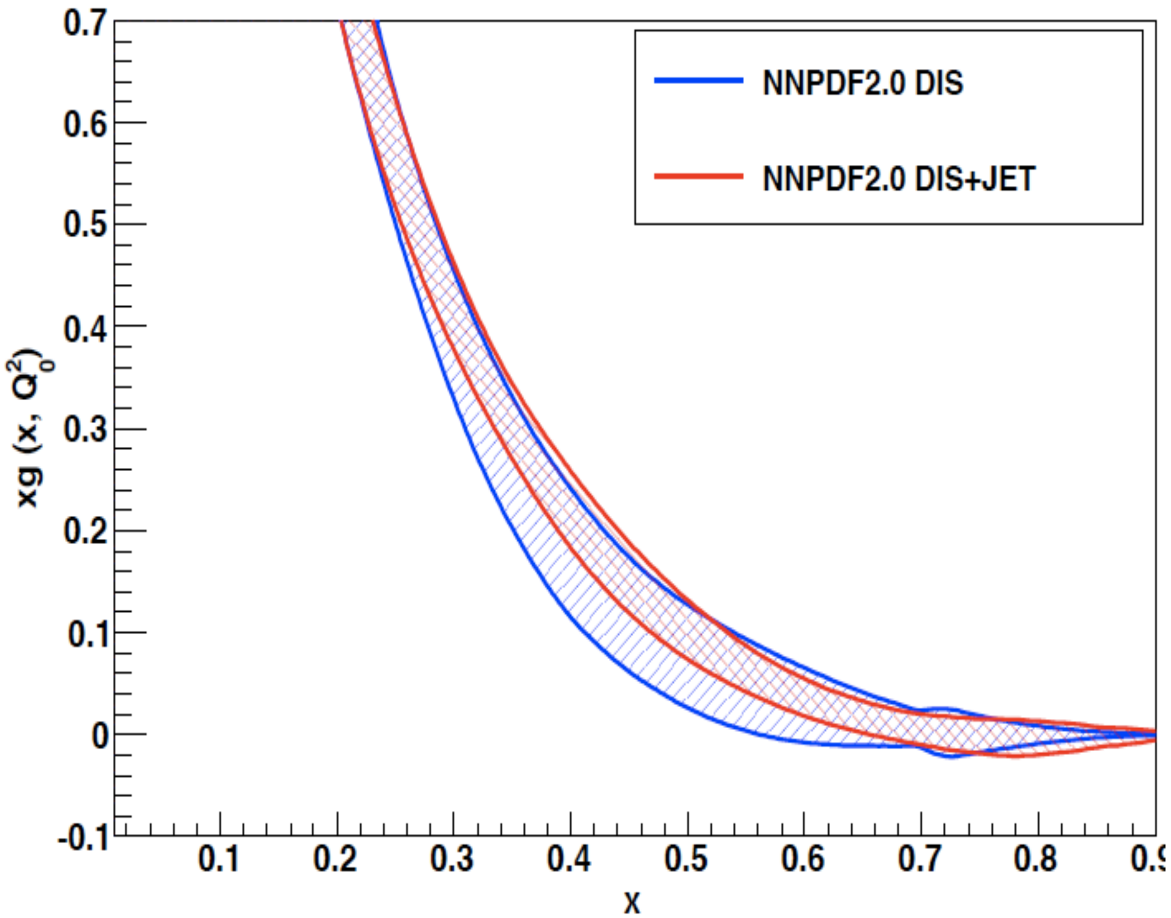


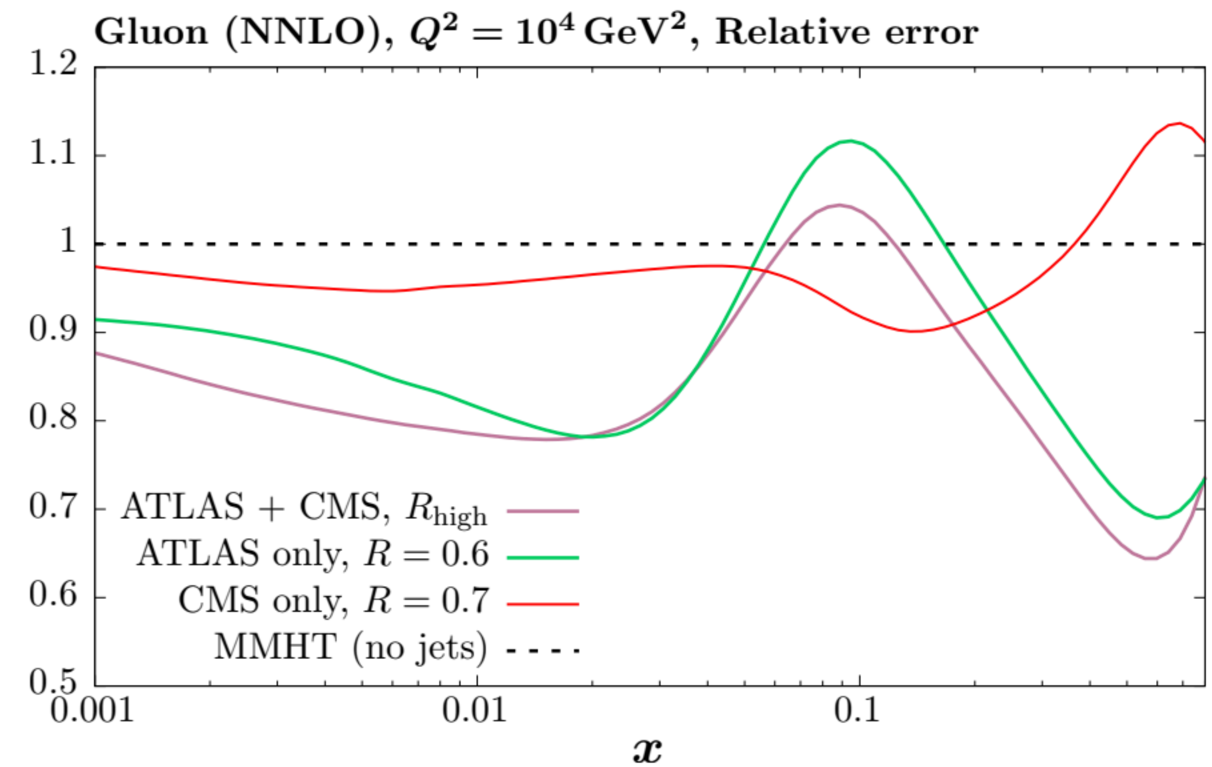
[MSTW, ArXiv: 0901.0002]

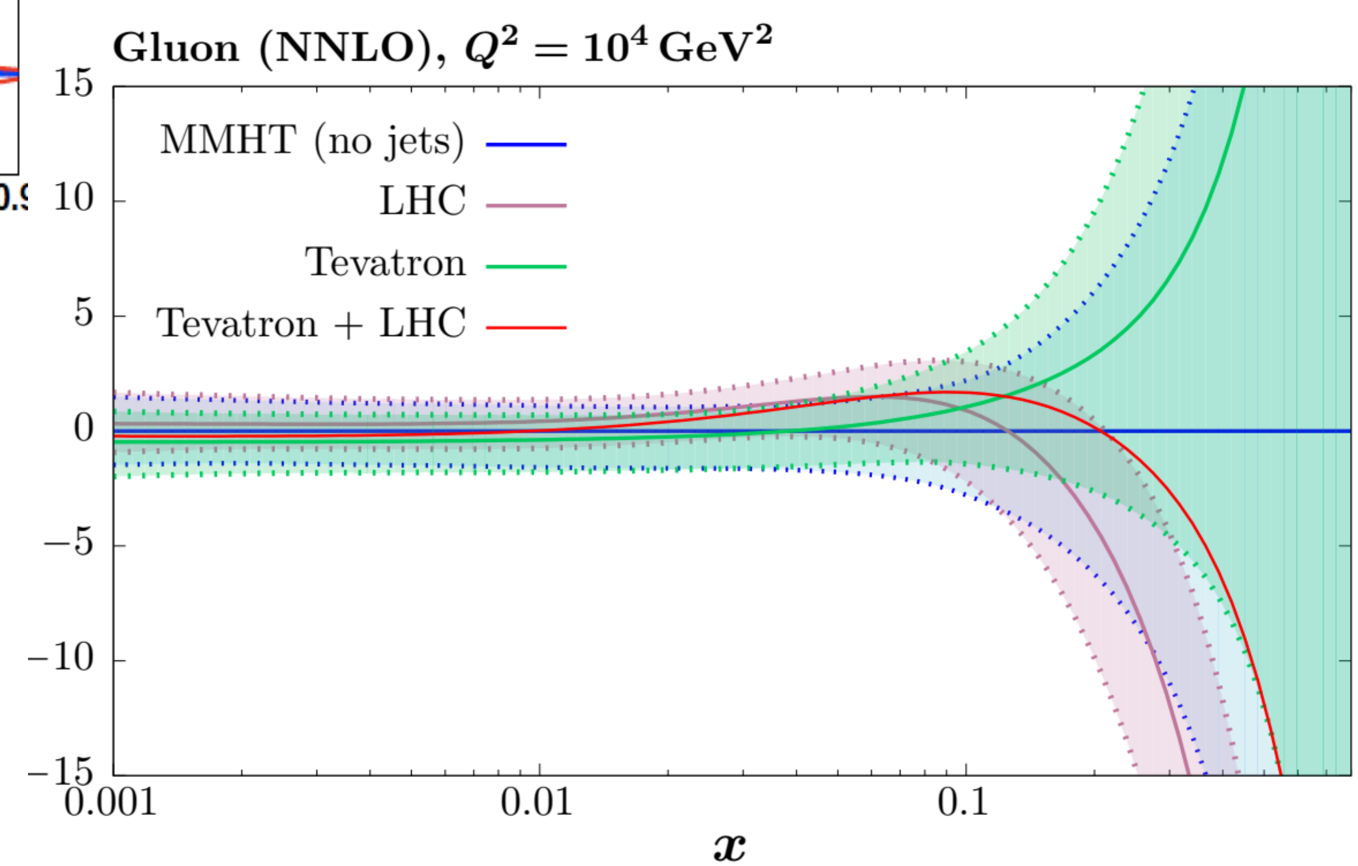
Gluon: jets data

LHC jet data

Tevatron jet data





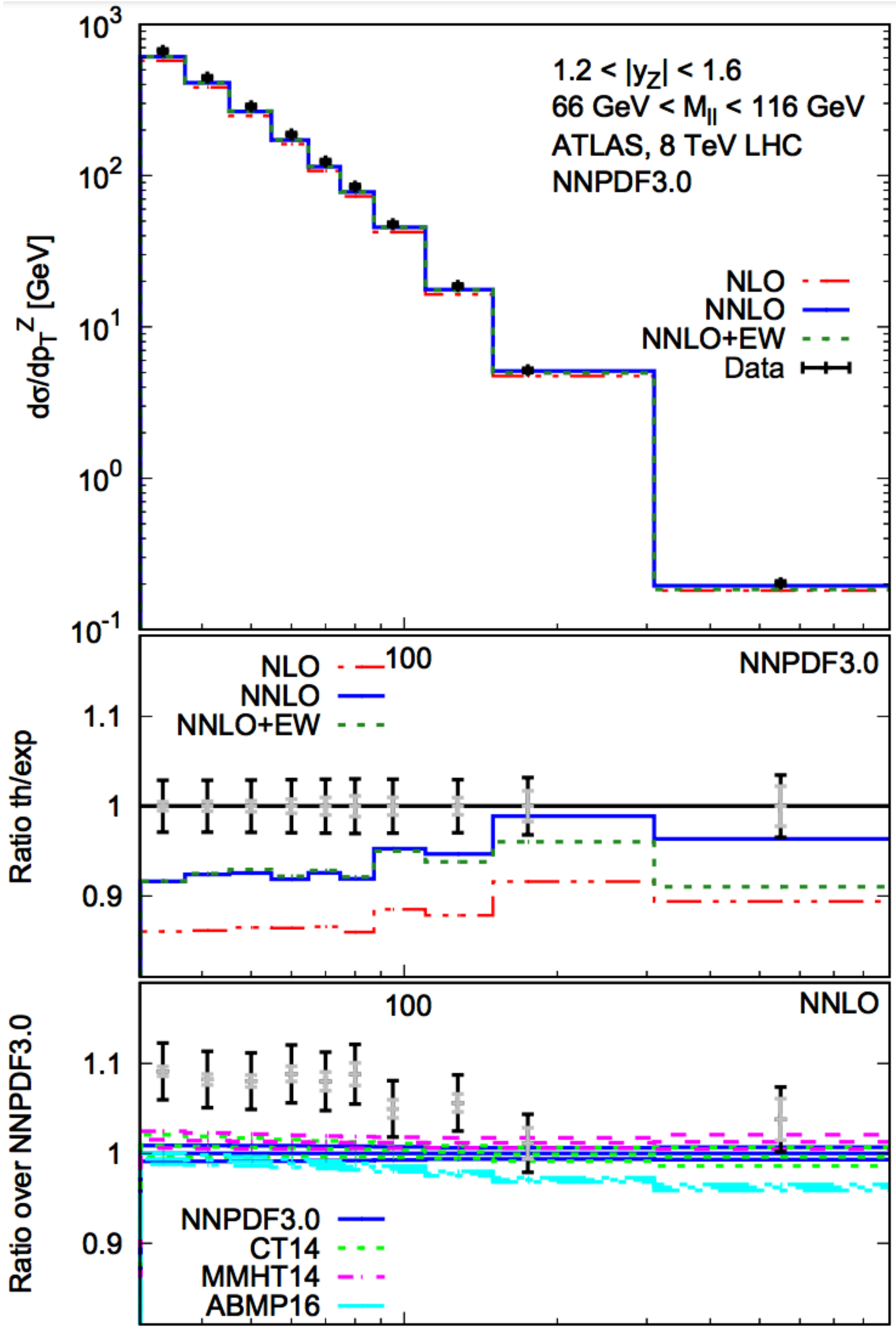


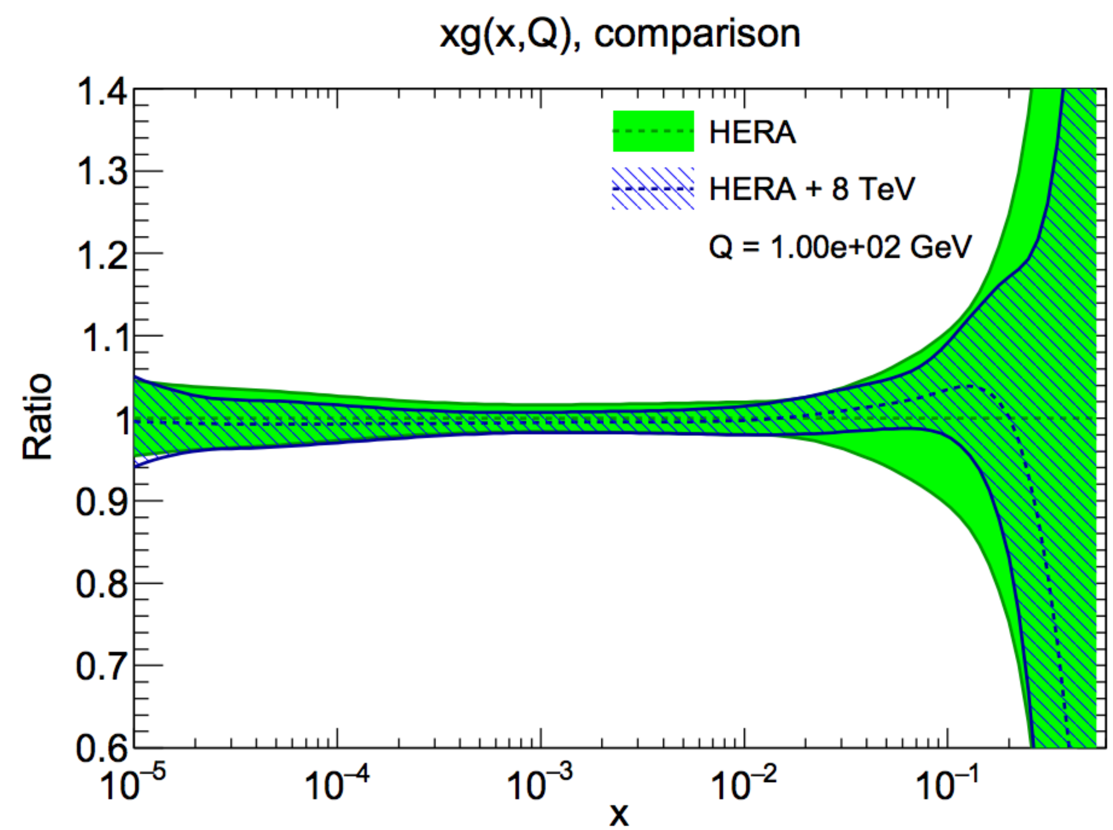
[Ball et al, ArXiv: 1002.4407]

[Harland-Lang et al, ArXiv: 1711.05757]

Gluon: Z transverse momentum

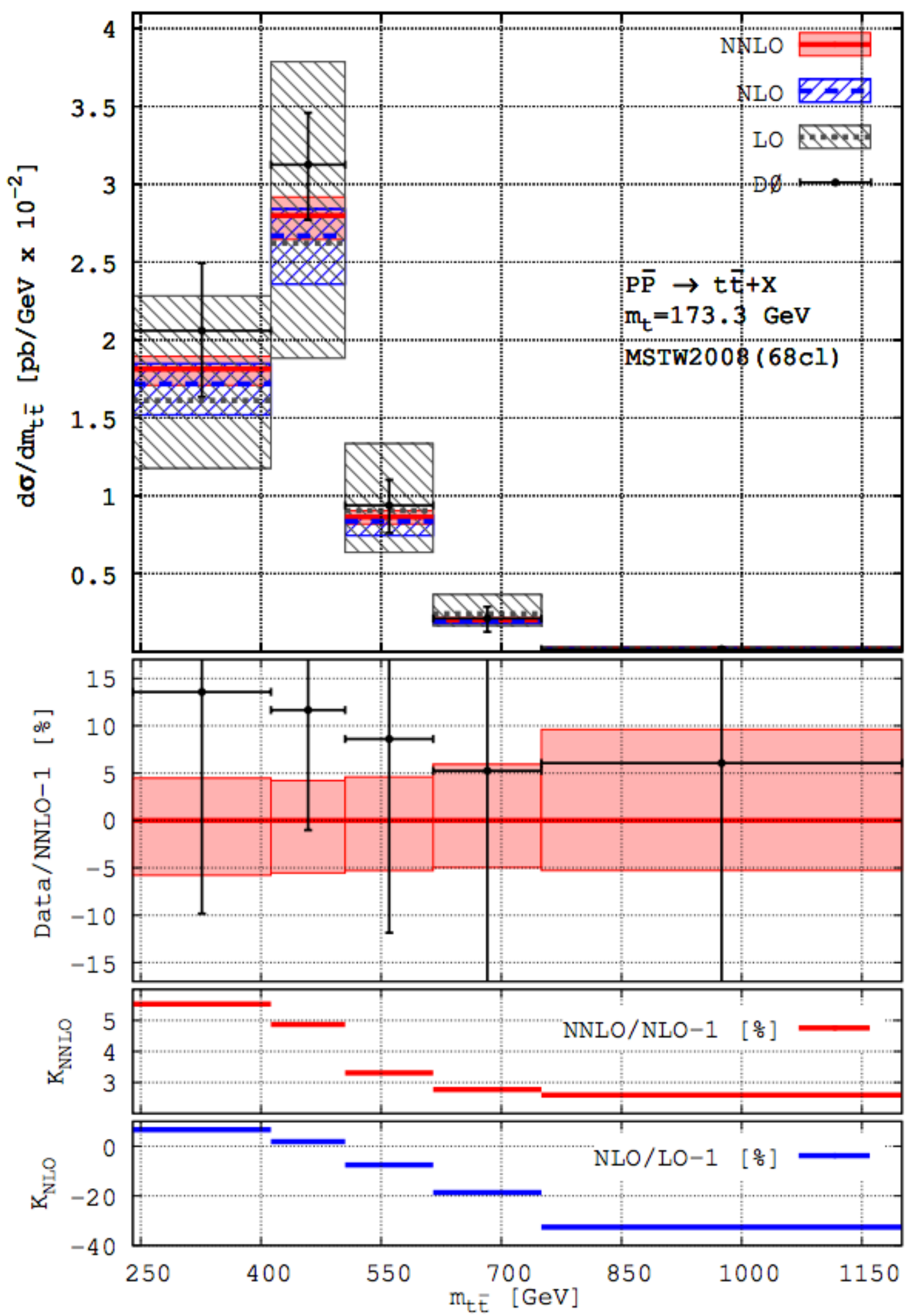
- Experimental precision < 1% up to pT~200 GeV
- Data hugely dominate by correlated systematic uncertainties
- Interesting case-study to probe current theory-experiment frontier

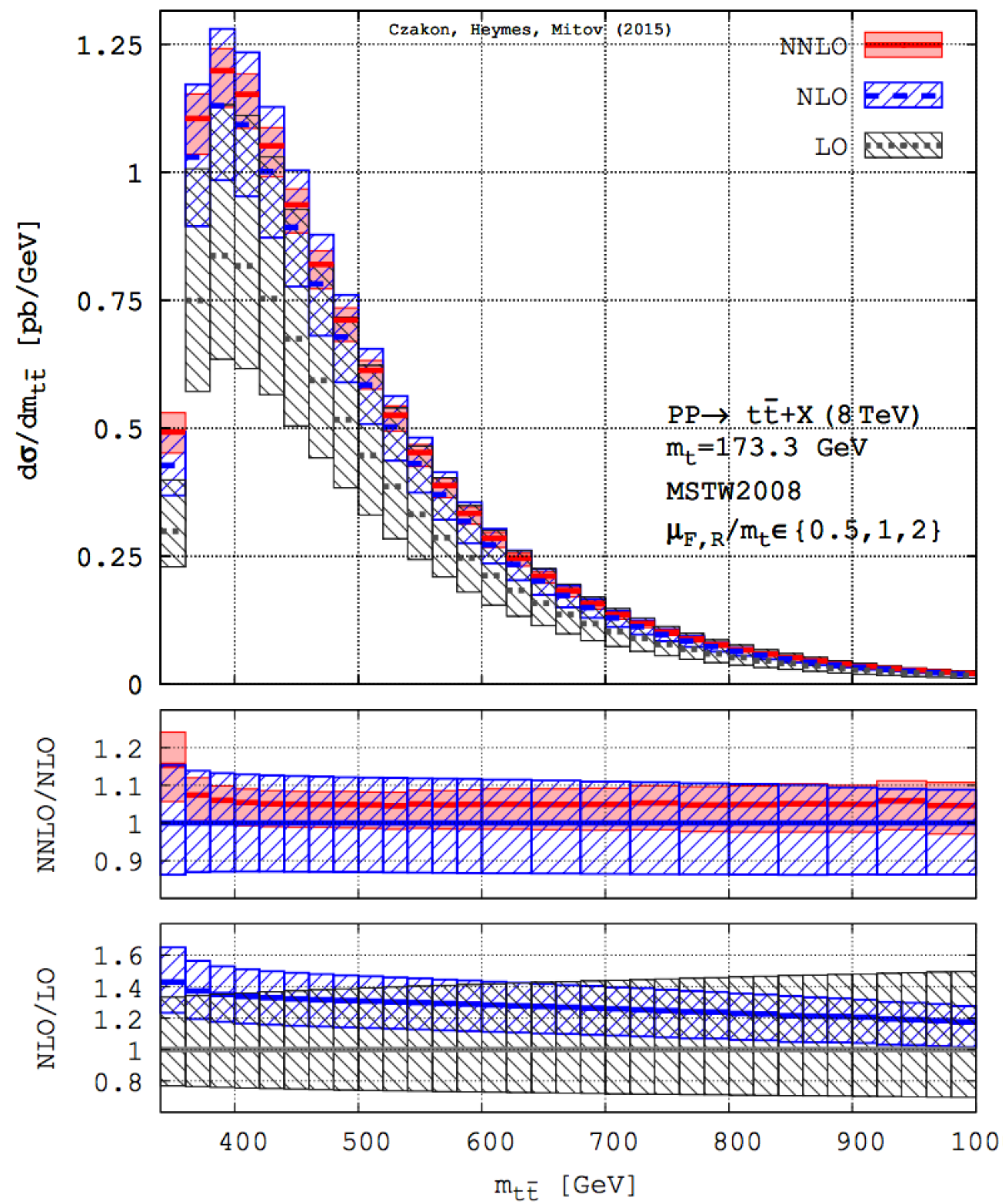




- Data/Theory comparison not so intuitive for correlation-dominated data
- Fluctuation in NNLO predictions (0.5 - 1%) had to be accounted for as extra nuisance parameter to get a good fit of such precise data

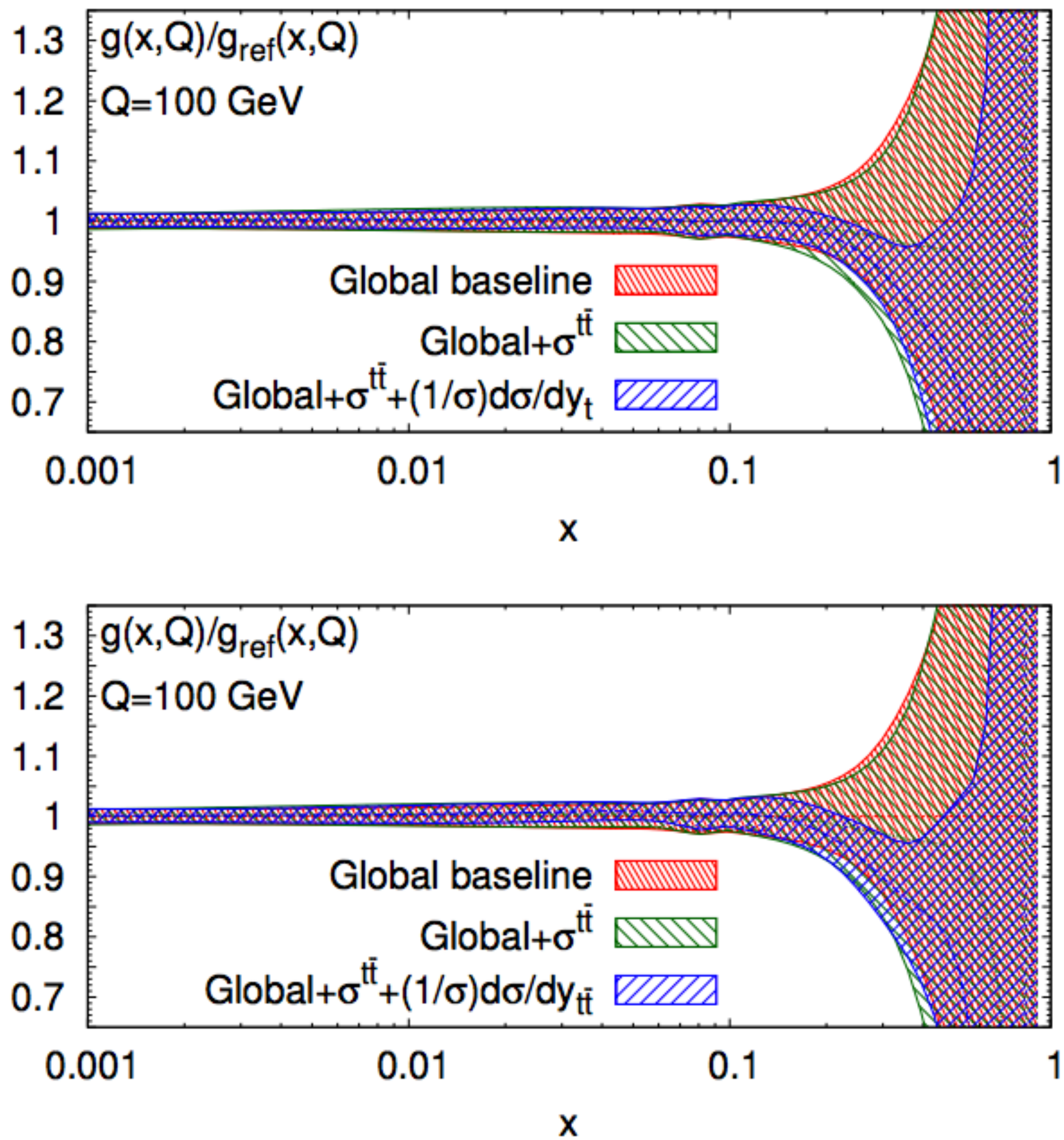
Gluon: top pair production



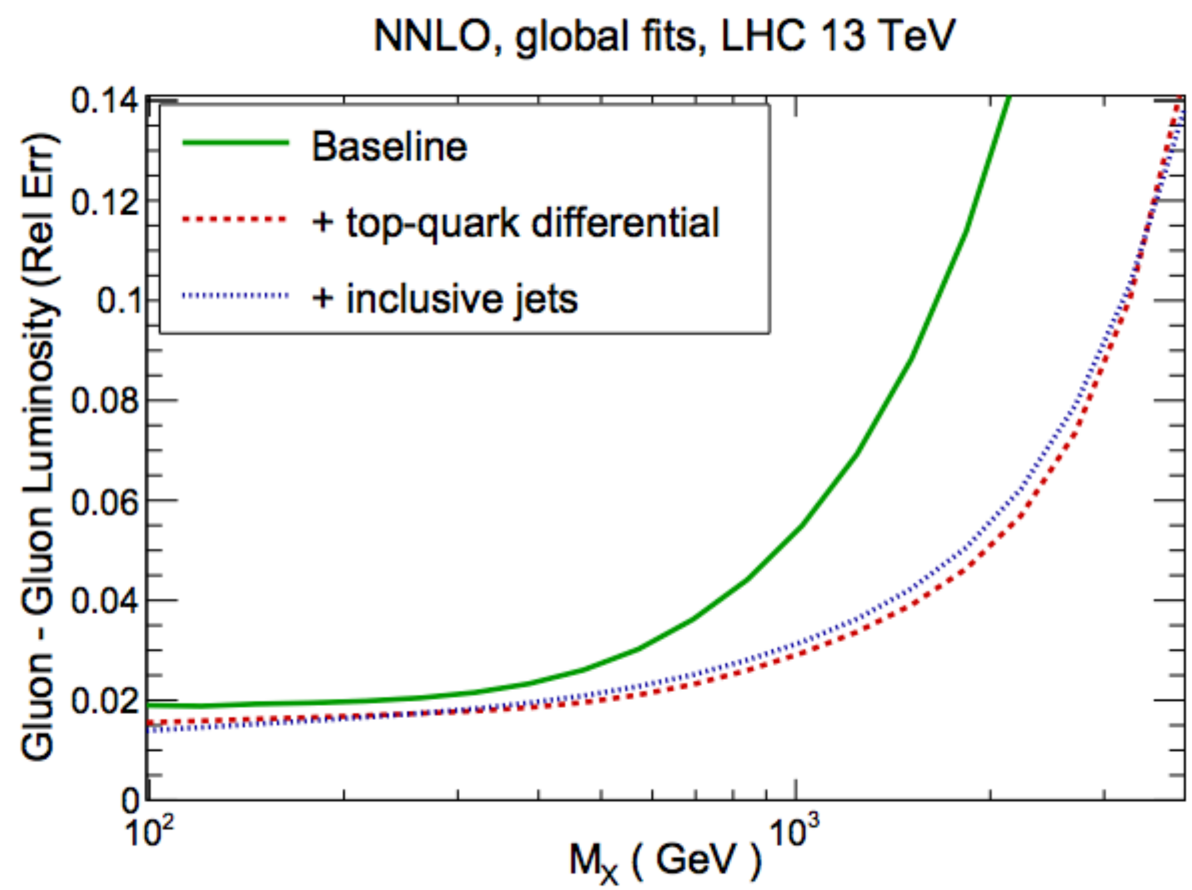


Czakon, Fiedler, Mitov [PRL 116(2016) 082003]

Gluon: top pair production



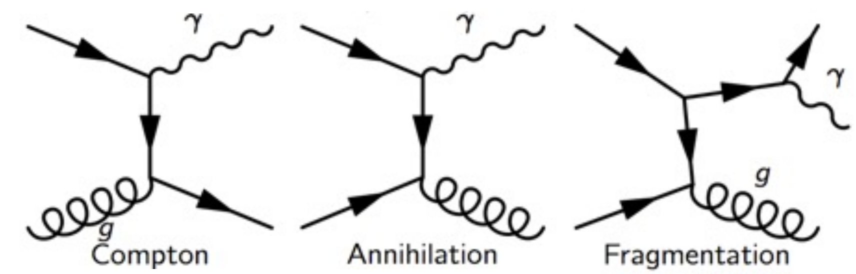
Czakon, Hartland, Mitov, Nocera and Rojo, arXiv: 1611.08609

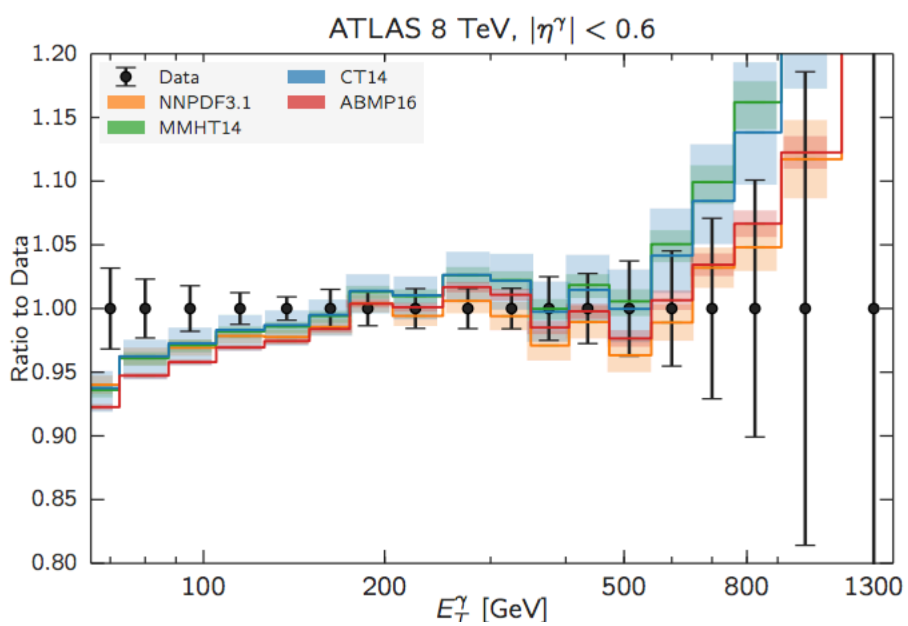


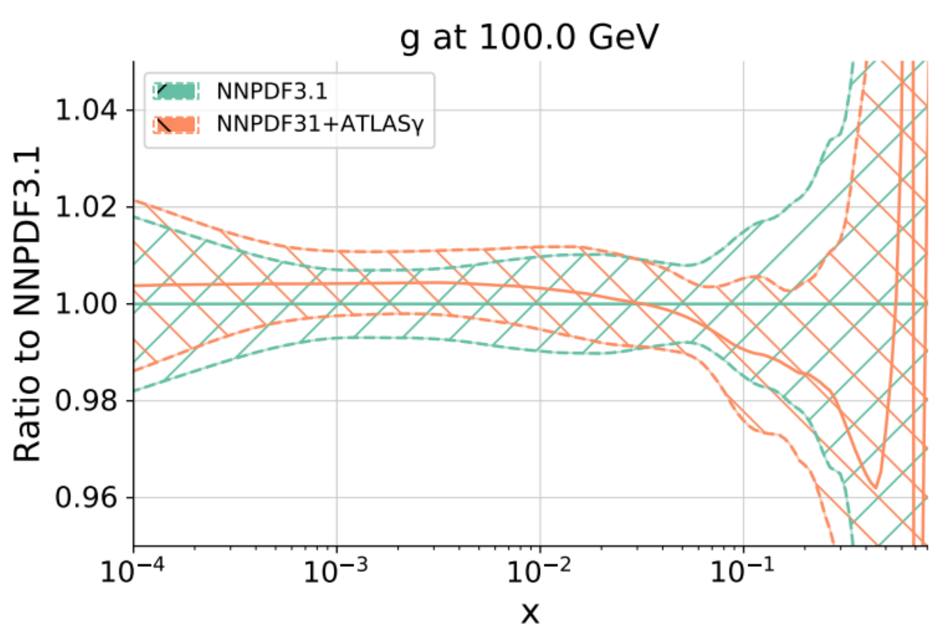
- Most constraining is inclusion of y_t list from ATLAS and y_{tt} from CMS jointly with total xsec
- Competitive reduction of gluon uncertainty with jets measurement
- Slight tension between ATLAS and CMS in NNPDF3.1 ($\chi^2_{ATLAS} \sim 1.6$, $\chi^2_{CMS} \sim 0.9$)

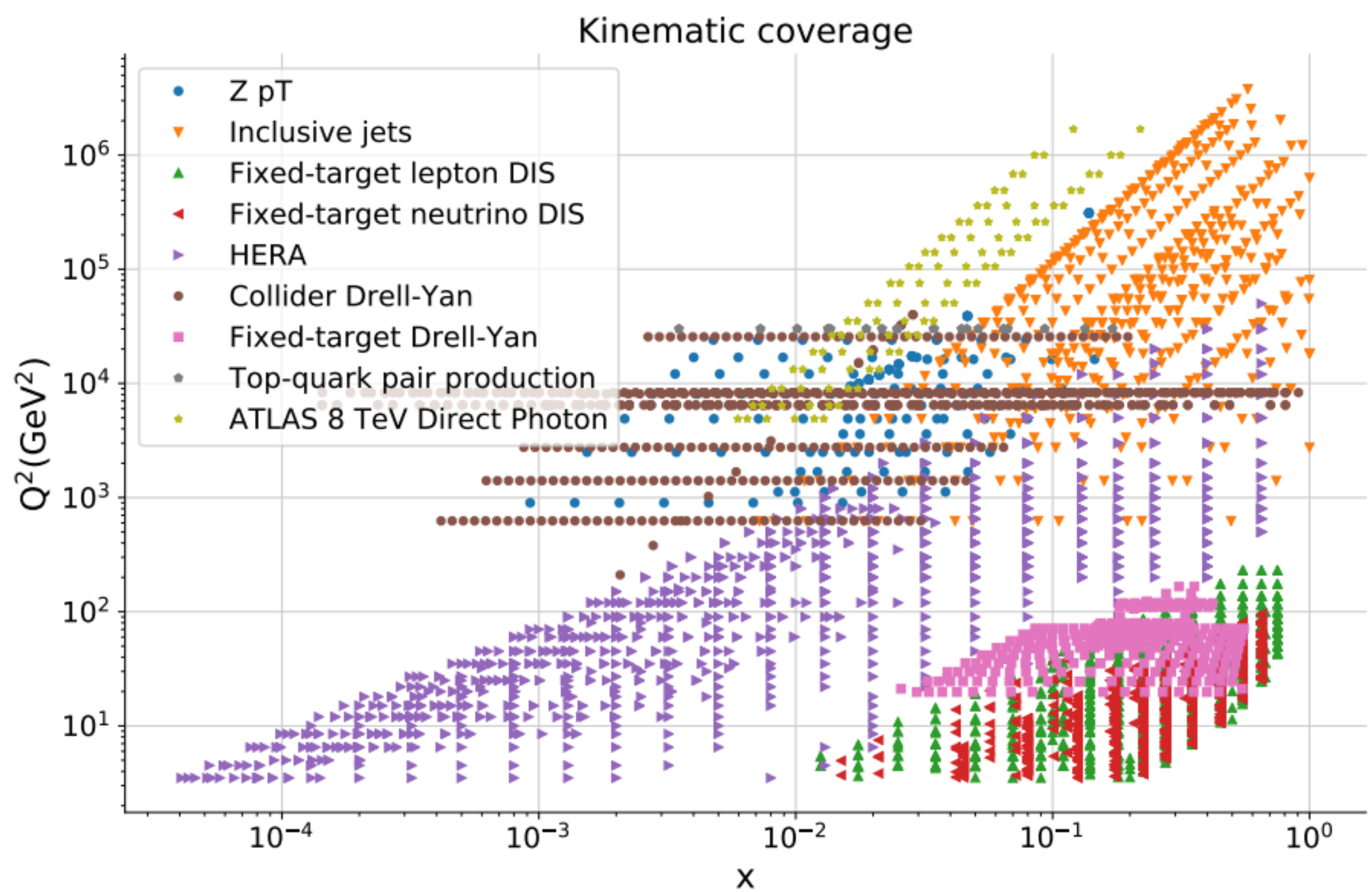
Gluon: direct photon production

- Prompt photon production directly sensitive to the gluon-quark luminosity via Compton scattering
- Isolated prompt photon data known at NNLO [Campbell et al 1612.04333] and accurately measured by ATLAS









Campbell et al 1802.03021

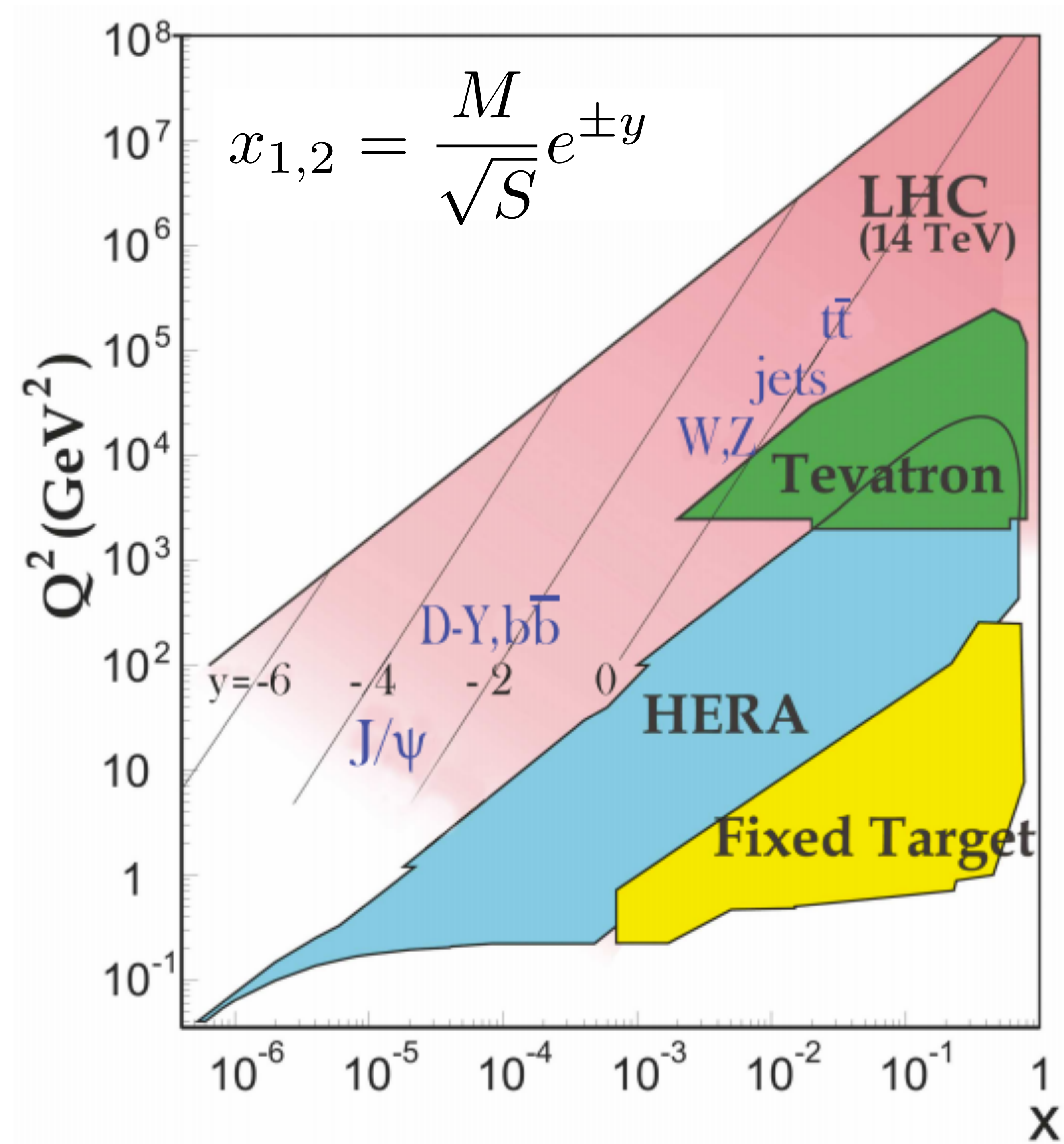
To summarise

GLUON

- Inclusive jets and dijets **(medium/large x)**
- Isolated photon and γ+jets **(medium/large x)**
- Top pair production **(large x)**
- High p_T V(+jets) distribution **(medium x)**

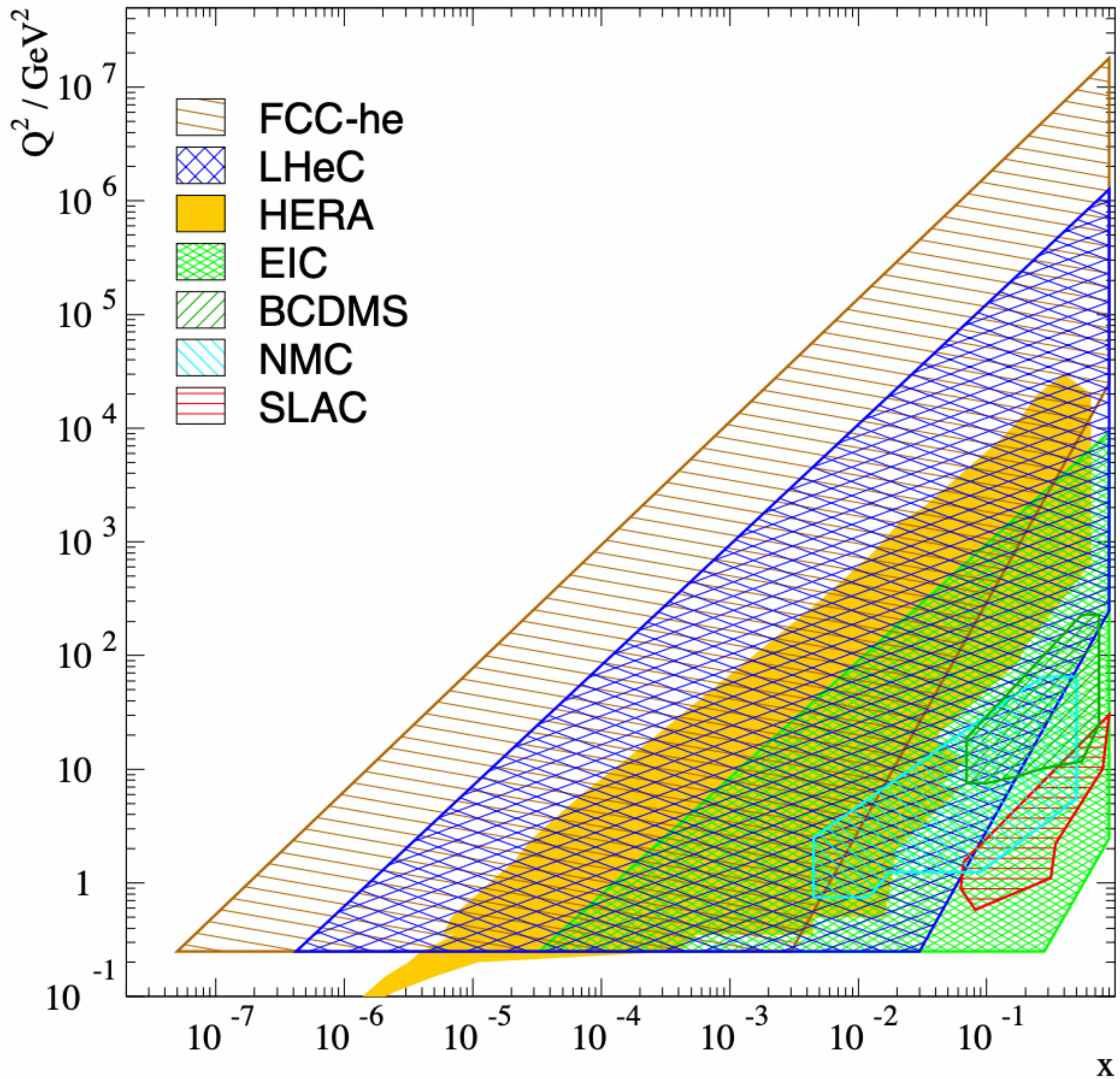
QUARKS

- High p_T V(+jets) ratios **(medium x)**
- W and Z production **(medium x)**
- Low and high mass Drell-Yan **(small and large x)**
- Wc **(strangeness at medium x)**



PDG 2016

Looking forward



T. Hobbs, SciPost(2022)

Parton Luminosities

- A quick and easy way to assess the mass and the collider dependence of production cross sections at hadron-hadron colliders is to use Parton Luminosities
- At leading order in QCD (parton model)

$$\begin{aligned}\hat{\sigma}_{ab\to X} &= C_X \delta(x_a x_b S - M^2) \\ \sigma_{pp\to X} &= \int_0^1 dx_a dx_b f_a(x_a, M^2) f_b(x_b, M^2) \hat{\sigma}_{ab\to X}\end{aligned}$$

- Thus

$$\begin{aligned}\sigma_{pp\to X} &= C_X \int_0^1 dx_a dx_b f_a(x_a, M^2) f_b(x_b, M^2) \delta(x_a x_b S - M^2) \\ &= \frac{C_X}{S} \frac{\partial \mathcal{L}_{ab}}{\partial \tau}\end{aligned}$$

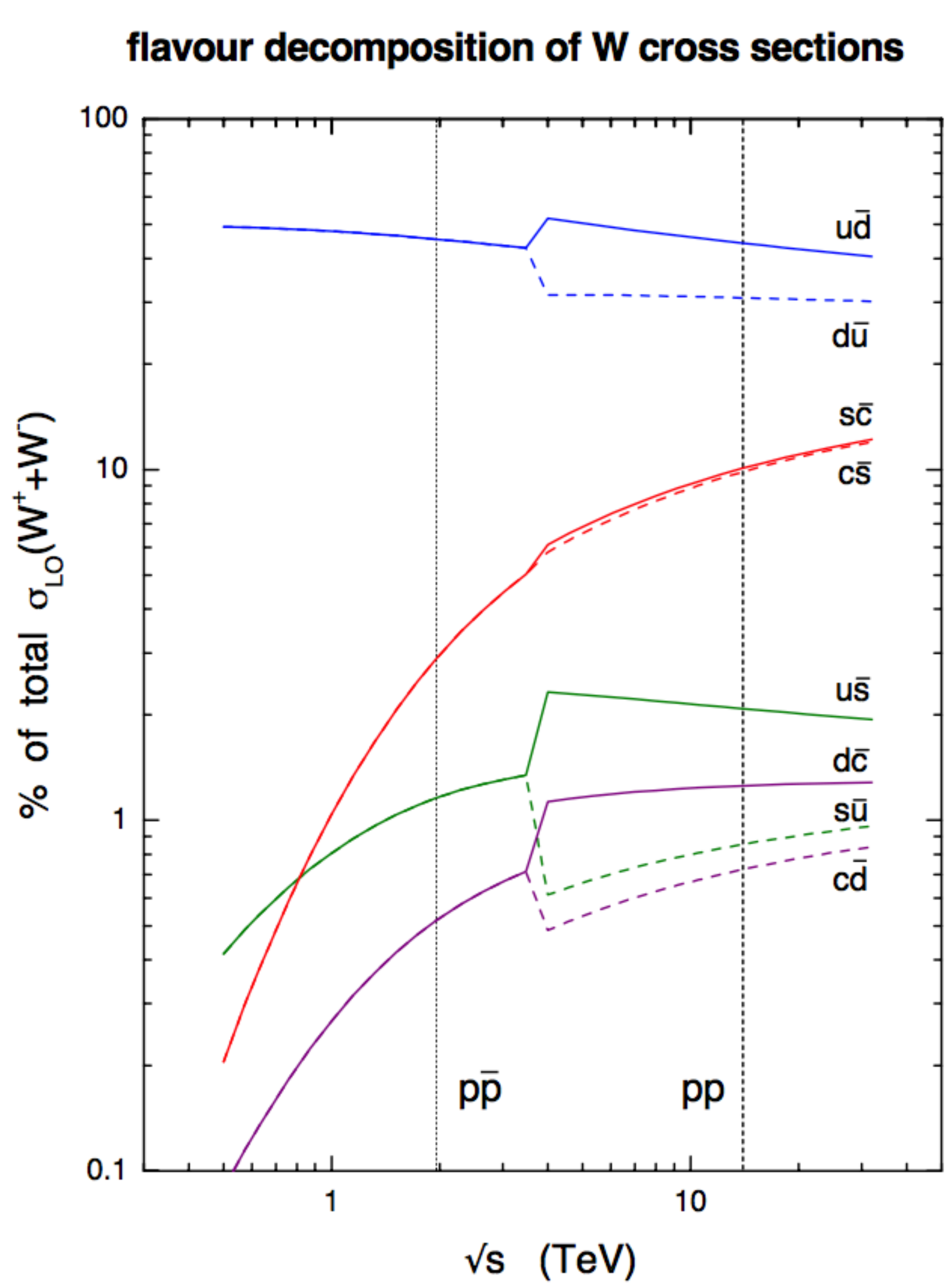
with

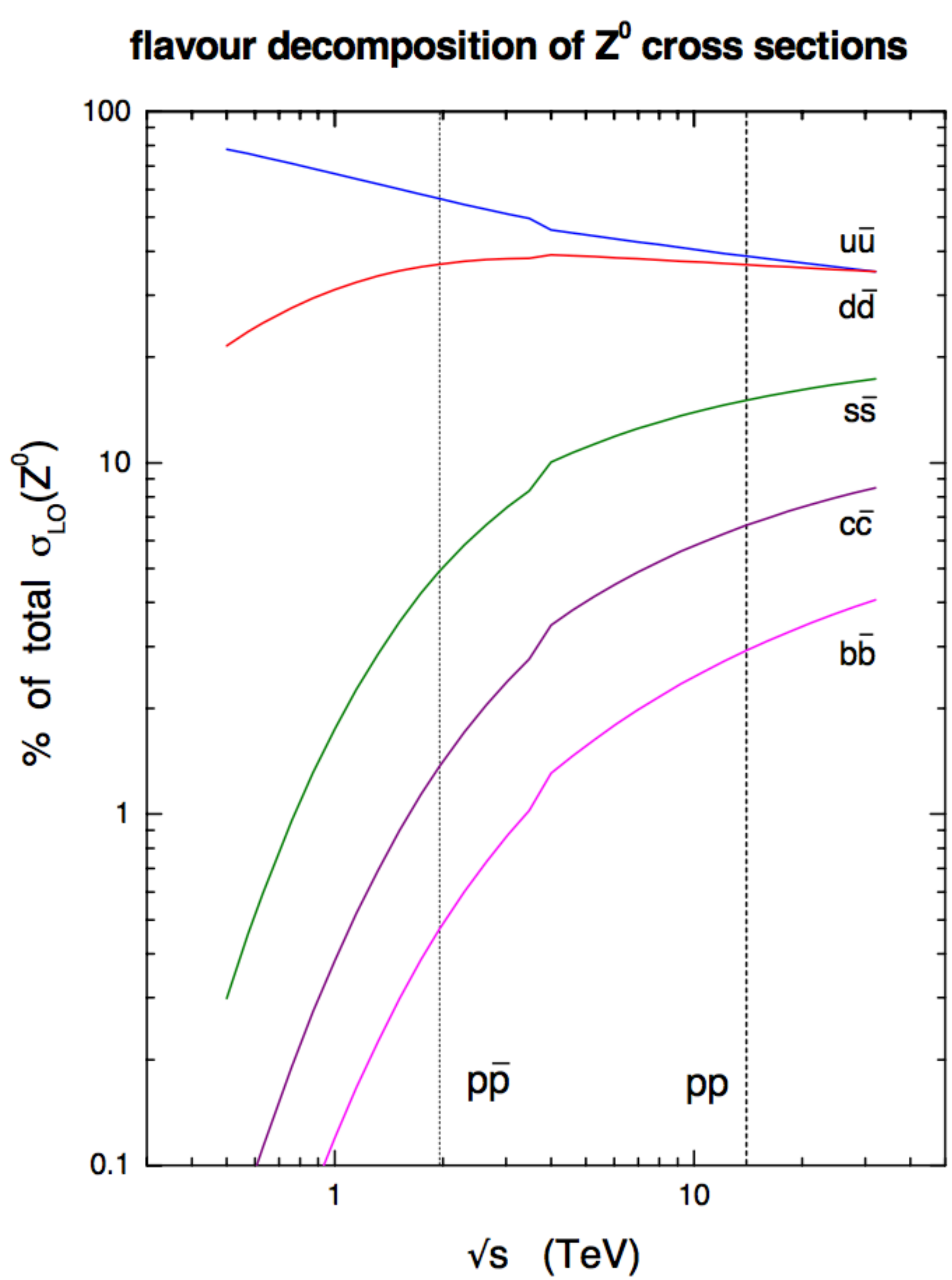
$$\tau = \frac{M^2}{\tau}$$

- Define

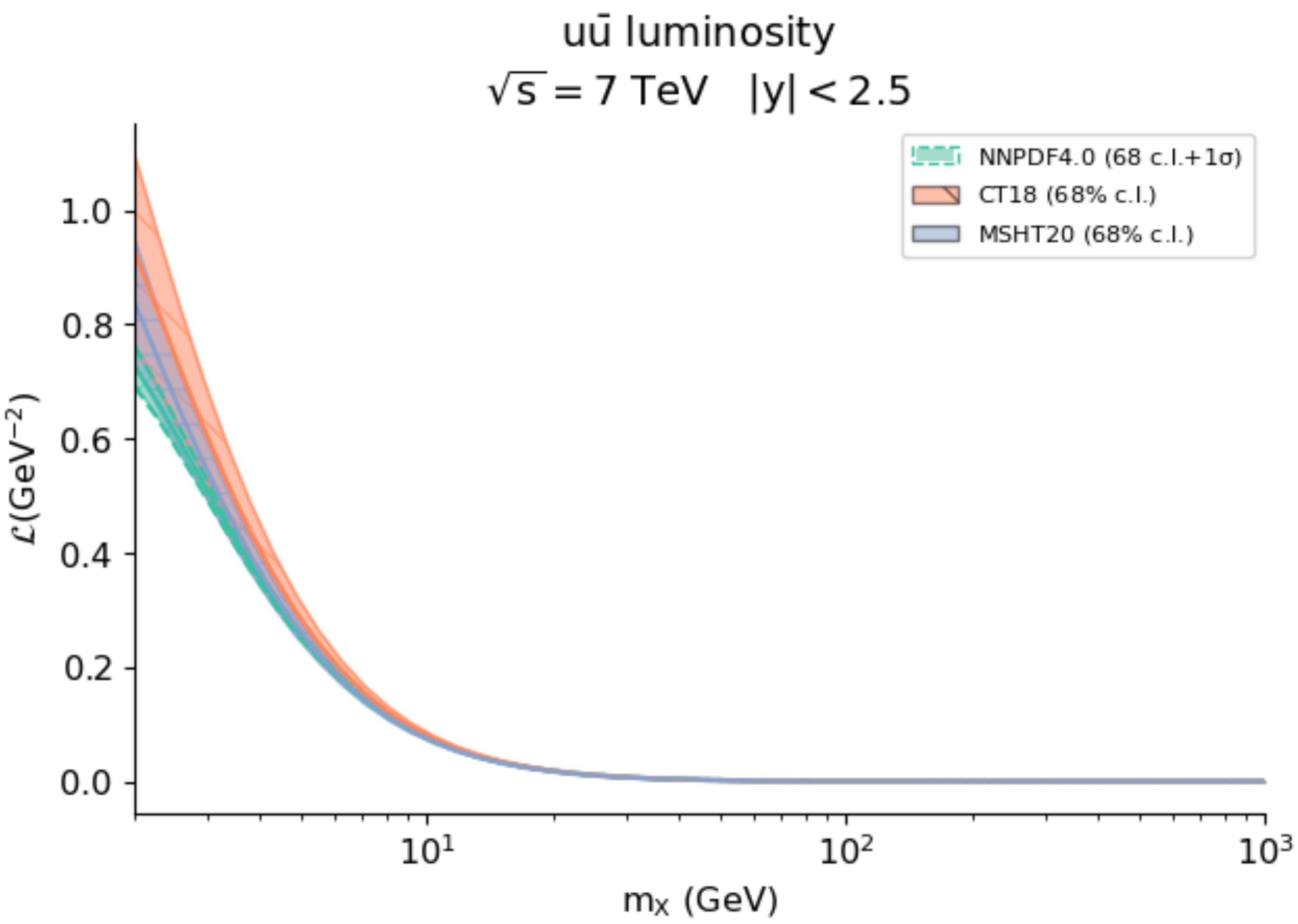
$$\boxed{\begin{aligned}\Phi_{ab}(M^2) &= \frac{\partial \mathcal{L}_{ab}}{\partial \tau} \\ &= \int_0^1 dx_a dx_b f_a(x_a, M^2) f_b(x_b, M^2) \delta(x_a x_b - \tau) \\ &= \frac{1}{S} \int_\tau^1 \frac{dy}{y} f_a(y, M^2) f_b\left(\frac{\tau}{y}, M^2\right)\end{aligned}}$$

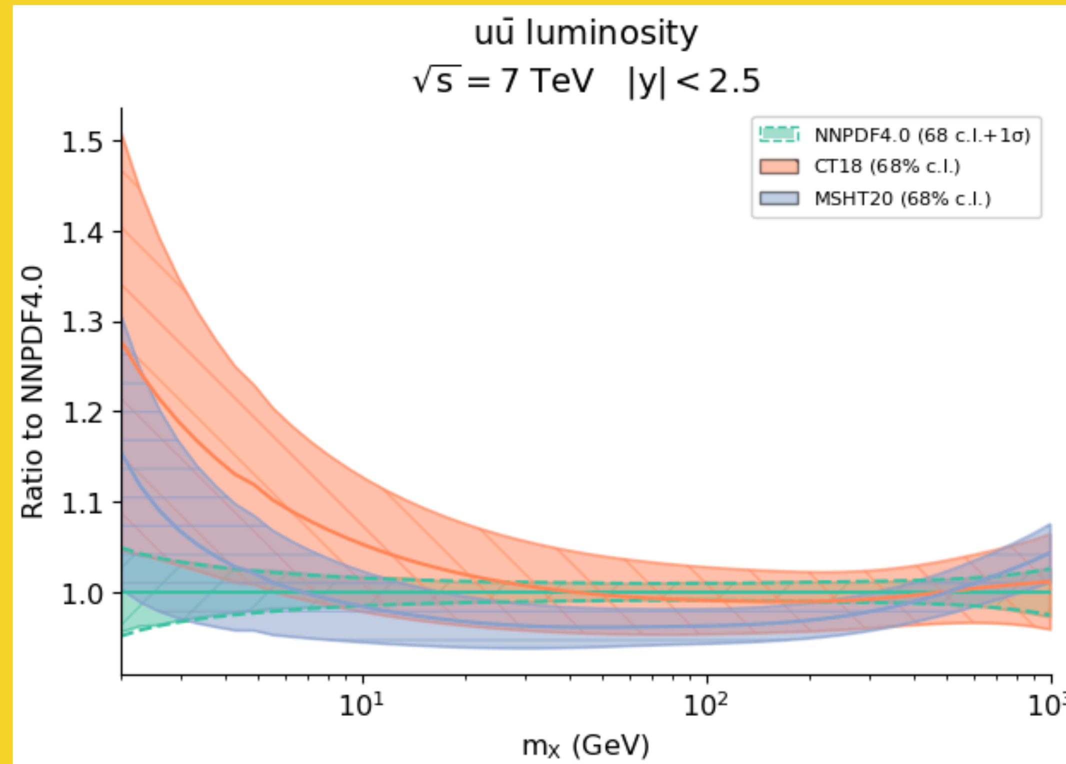
Parton Luminosities





Parton Luminosities





Methodological aspects

A quite complicated game

- A single quantity: 1σ error

- Multiple quantities: 1σ contours

- Functions: 1σ "error band" in the space of functions
  = find the probability density in the space of functions f(x)
  Expectation values are functional integrals

Not as simple as it may look…

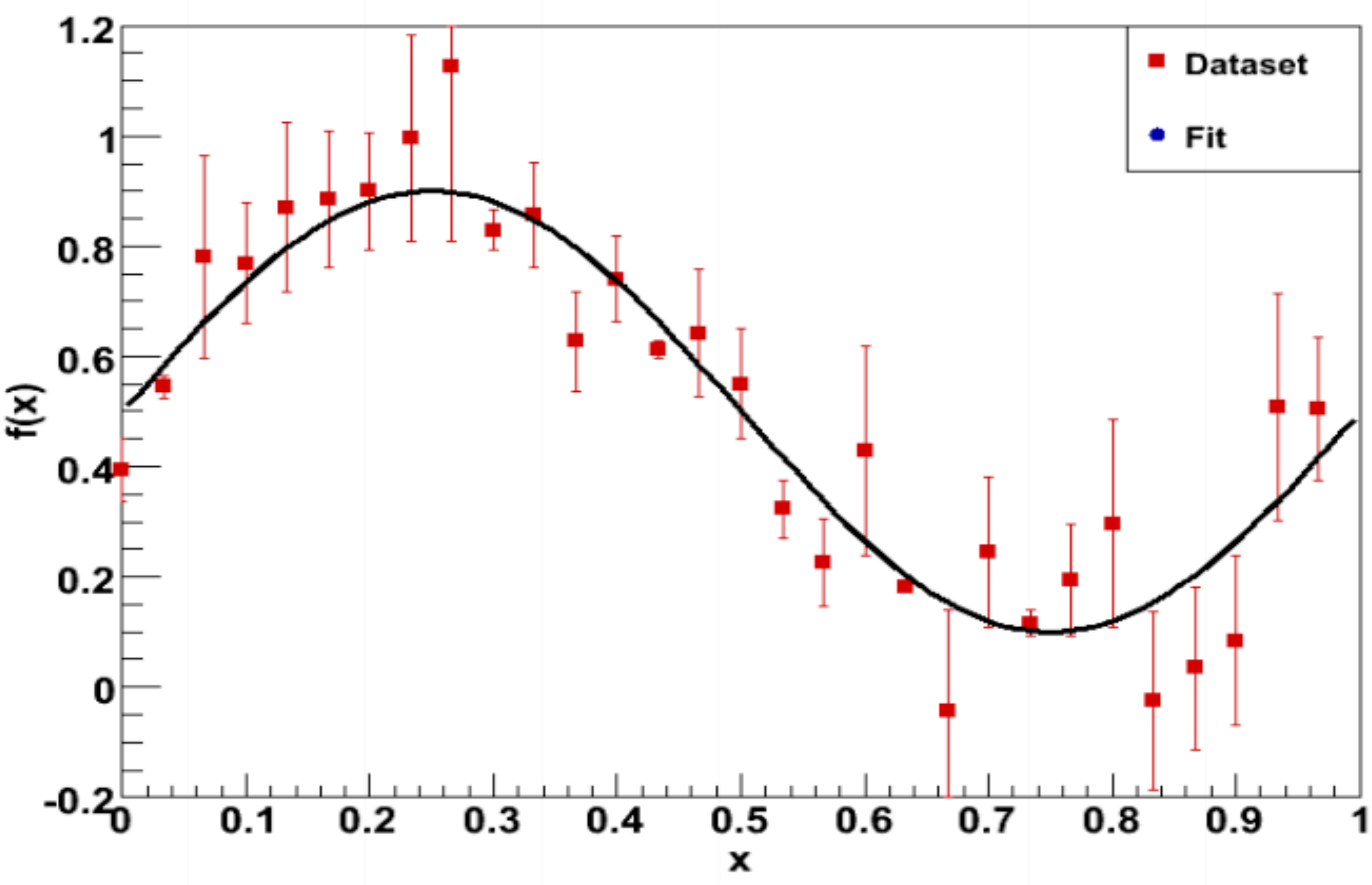
$$\langle \mathcal{O}[\{f\}] \rangle = \int [\mathcal{D}f]\, \mathcal{O}[\{f\}] \mathcal{P}[\{f\}]$$

- Given a finite number of experimental data points want a set of functions
- Want to find a infinite-dimensional object from a finite number of information

A toy model

A toy-model:

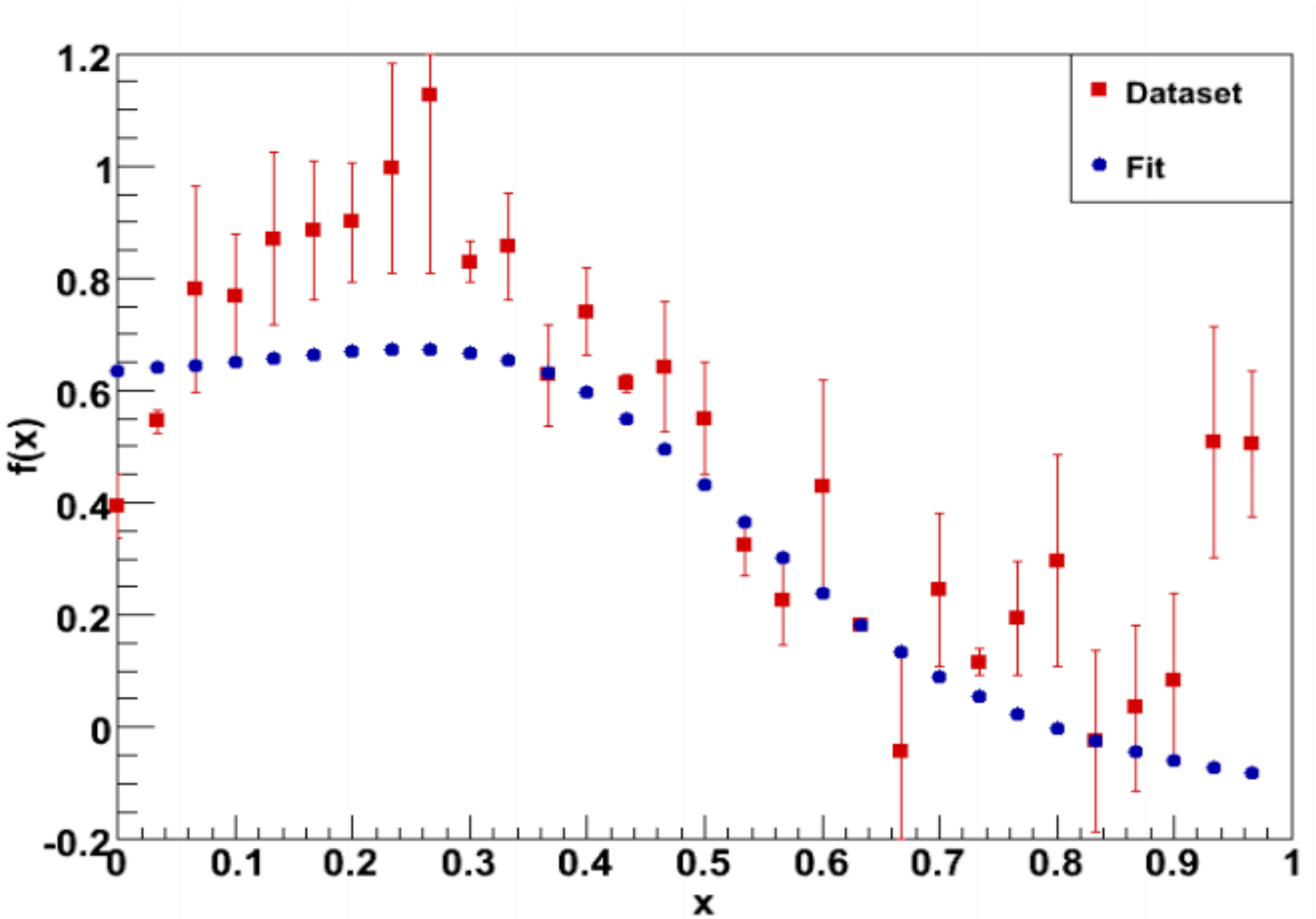
1) Imagine that we have a set of uncorrelated measurements of a quantity f(x) at different x. The underlying law that Nature established for this quantity is a sinusoidal, but we don't know anything about that and try to guess it with a fit.



A toy model

A toy-model:

2) Choose a parametrisation for f(x) and perform a fit by minimising a function, a figure of merit, like the χ^2



$$\chi^2 = \sum_{i=1}^{N_{\rm data}} \frac{(D_i - T_i)^2}{\sigma_i^2}$$

χ^2/d.o.f. » 1

We are not quite there...

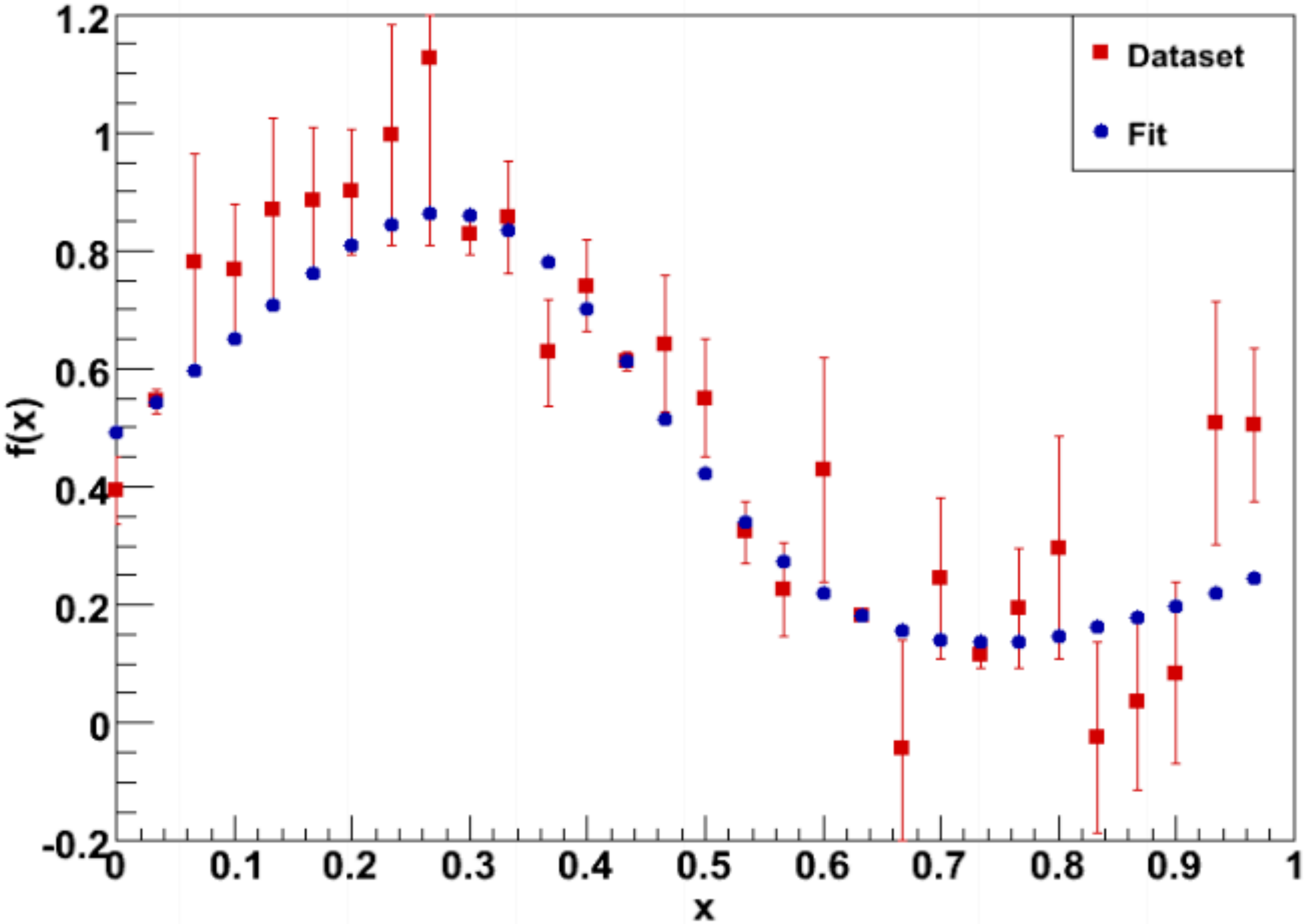
under-learning

A toy model

A toy-model:

2) Choose a parametrisation for f(x) and perform a fit by minimising a function, a figure of merit, like the χ^2

$$\boxed{\chi^2 = \sum_{i=1}^{N_{\rm data}} \frac{(D_i - T_i)^2}{\sigma_i^2}}$$



χ^2/d.o.f. ~ 1

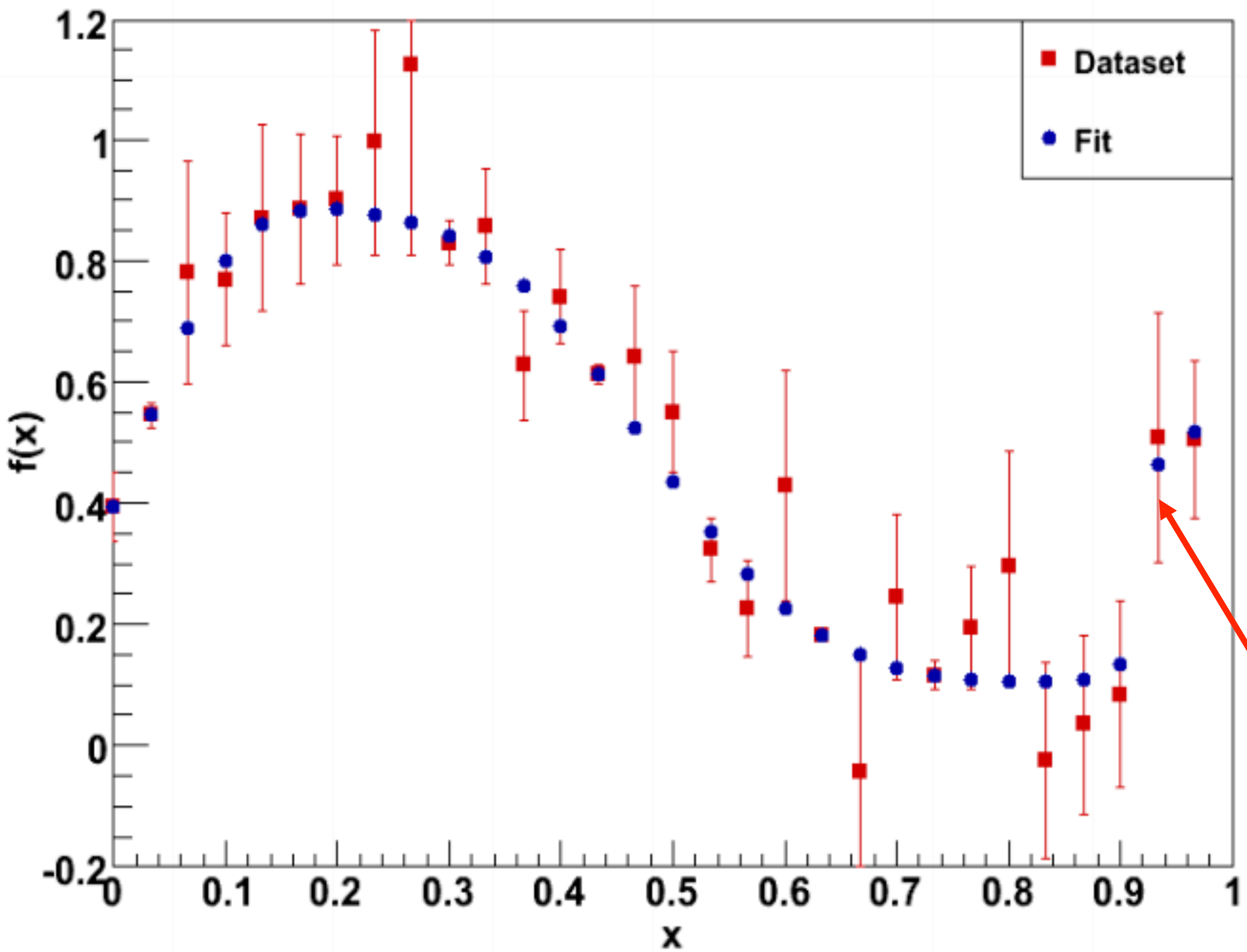
We are there…

proper learning

A toy model

A toy-model:

2) Choose a parametrisation for f(x) and perform a fit by minimising a function, a figure of merit, like the χ^2



$$\chi^2 = \sum_{i=1}^{N_{\text{data}}} \frac{(D_i - T_i)^2}{\sigma_i^2}$$

χ^2/d.o.f. → 0

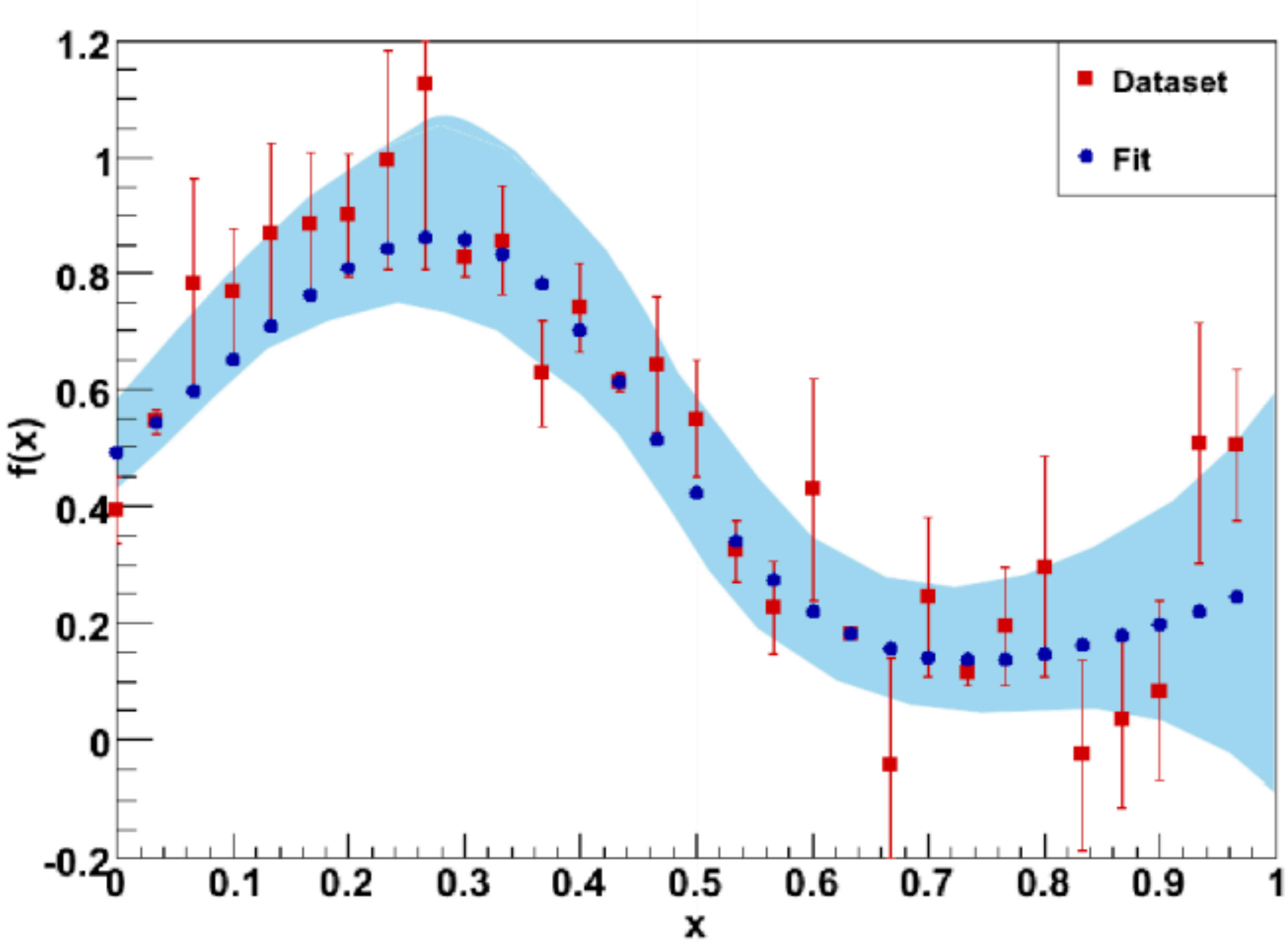
We went too far…

over-learning

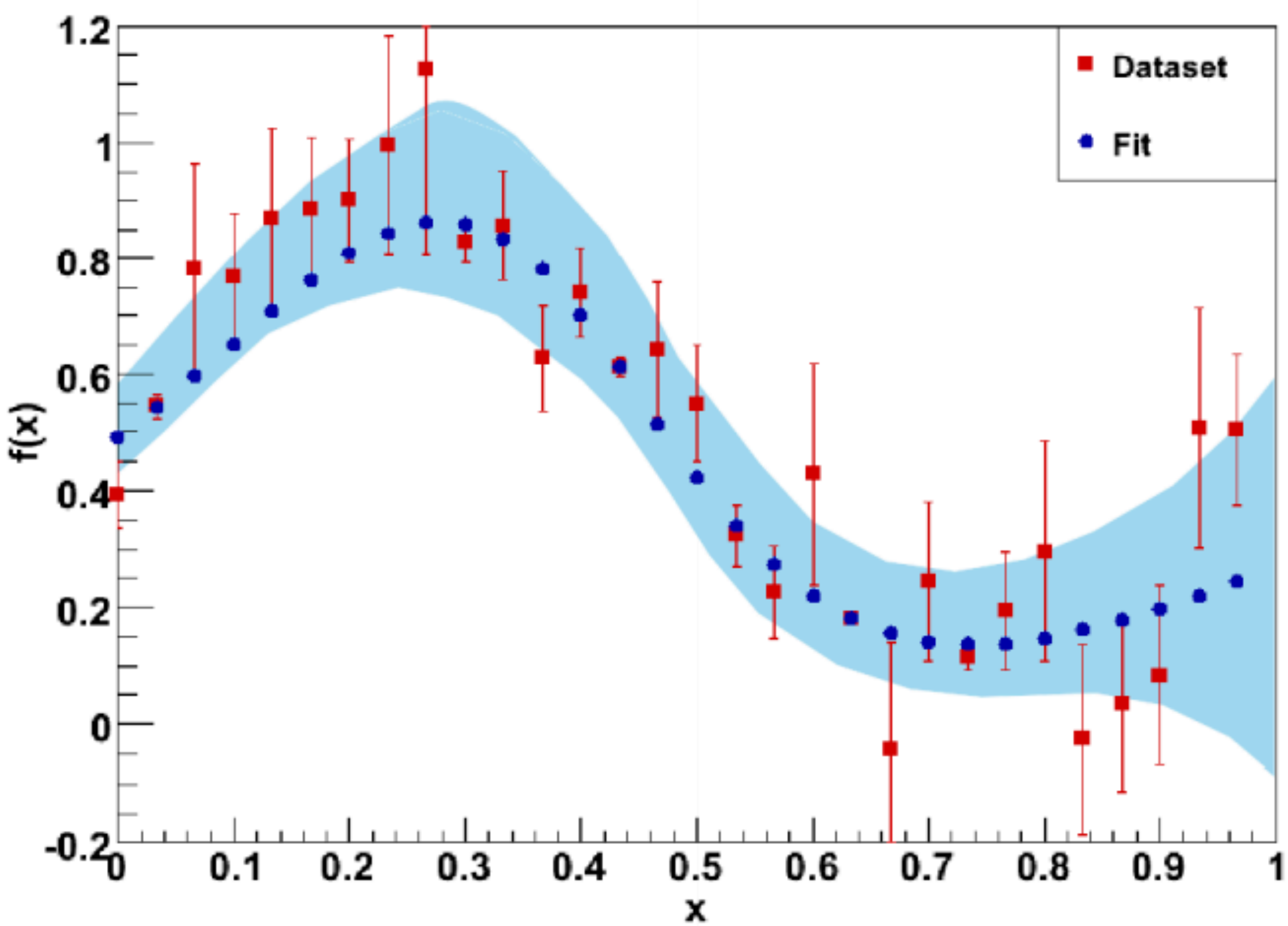
start fitting the statistical noise

A toy model

A toy-model:
3) Determine the error of our fit, which corresponds to the lack of information that the data provide. In the limit of infinite and infinitely-precise and compatible data, the error band tends to 0



The actual game



The actual games is more complicated since we have 6+6+1 functions (actually 3+3+1+1) and errors to determine which are not directly measured. They enter in the measured observables according to different combinations. But still…

- ✓ Need to choose a clever and flexible parametrisation
- ✓ Need a way to stop the fit before over-learning sets in to avoid fitting statistical noise
- ✓ Need a reliable error estimate

Parametrisation

Choice of parametrisation

Usually one parametrises independently the gluon, light quarks and anti-quarks, strange and anti-strange (+ intrinsic charm), while heavy quarks are generated perturbatively from light quarks and gluons*

The ideal parametrisation

- Too rigid

Global fit might not have flexibility to describe data or inadequate small uncertainties where there are no data

- Too flexible

Difficult minimisation and it might develop artefacts driven by statistical fluctuations of the data

Sum rules

- From baryon number conservation → Valence Sum Rules

$$\int_0^1 dx\,(u(x,Q^2) - \bar{u}(x,Q^2)) = 2$$

$$\int_0^1 dx\,(d(x,Q^2) - \bar{d}(x,Q^2)) = 1$$

$$\int_0^1 dx\,(s(x,Q^2) - \bar{s}(x,Q^2)) = 0$$

- From momentum conservation → Momentum Sum Rule

$$\int_0^1 dx\,(x\Sigma(x,Q^2) + xg(x,Q^2)) = 1$$

$$\text{with } \Sigma = \sum_{i=1}^{n_F} q_i + \bar{q}_i$$

Traditional (parametrical) approach

- Introduce a simple functional form with enough free parameters

$$f_i(x, Q_0^2) = a_0\, x^{a_1}\, (1-x)^{a_2} P(x, a_3, a_4, \ldots)$$

- Typically about 20-25 free parameters for 7 independent functions

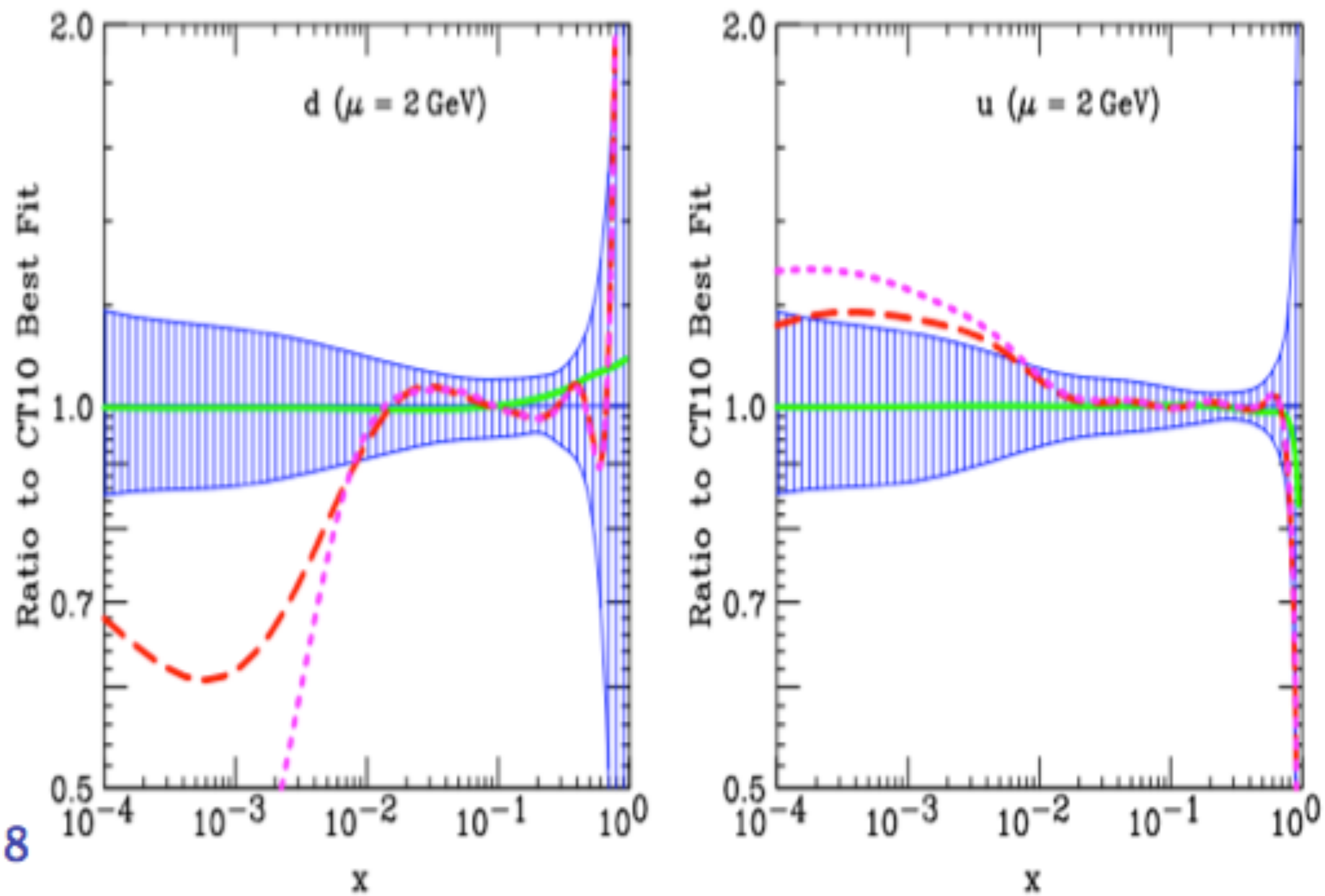
MSTW2008

20 free parameters

$$\begin{aligned}
xu_v(x, Q_0^2) &= A_u\, x^{\eta_1} (1-x)^{\eta_2} (1 + \epsilon_u \sqrt{x} + \gamma_u\, x), \\
xd_v(x, Q_0^2) &= A_d\, x^{\eta_3} (1-x)^{\eta_4} (1 + \epsilon_d \sqrt{x} + \gamma_d\, x), \\
xS(x, Q_0^2) &= A_S\, x^{\delta_S} (1-x)^{\eta_S} (1 + \epsilon_S \sqrt{x} + \gamma_S\, x), \\
x\Delta(x, Q_0^2) &= A_\Delta\, x^{\eta_\Delta} (1-x)^{\eta_S+2} (1 + \gamma_\Delta\, x + \delta_\Delta\, x^2), \\
xg(x, Q_0^2) &= A_g\, x^{\delta_g} (1-x)^{\eta_g} (1 + \epsilon_g \sqrt{x} + \gamma_g\, x) + A_{g'}\, x^{\delta_{g'}} (1-x)^{\eta_{g'}}, \\
x(s+\bar{s})(x, Q_0^2) &= A_+\, x^{\delta_S} (1-x)^{\eta_+} (1 + \epsilon_S \sqrt{x} + \gamma_S\, x), \\
x(s-\bar{s})(x, Q_0^2) &= A_-\, x^{\delta_-} (1-x)^{\eta_-} (1 - x/x_0),
\end{aligned}$$

Traditional (parametrical) approach

- Possible issues:
  What is the error associated to a given functional form?

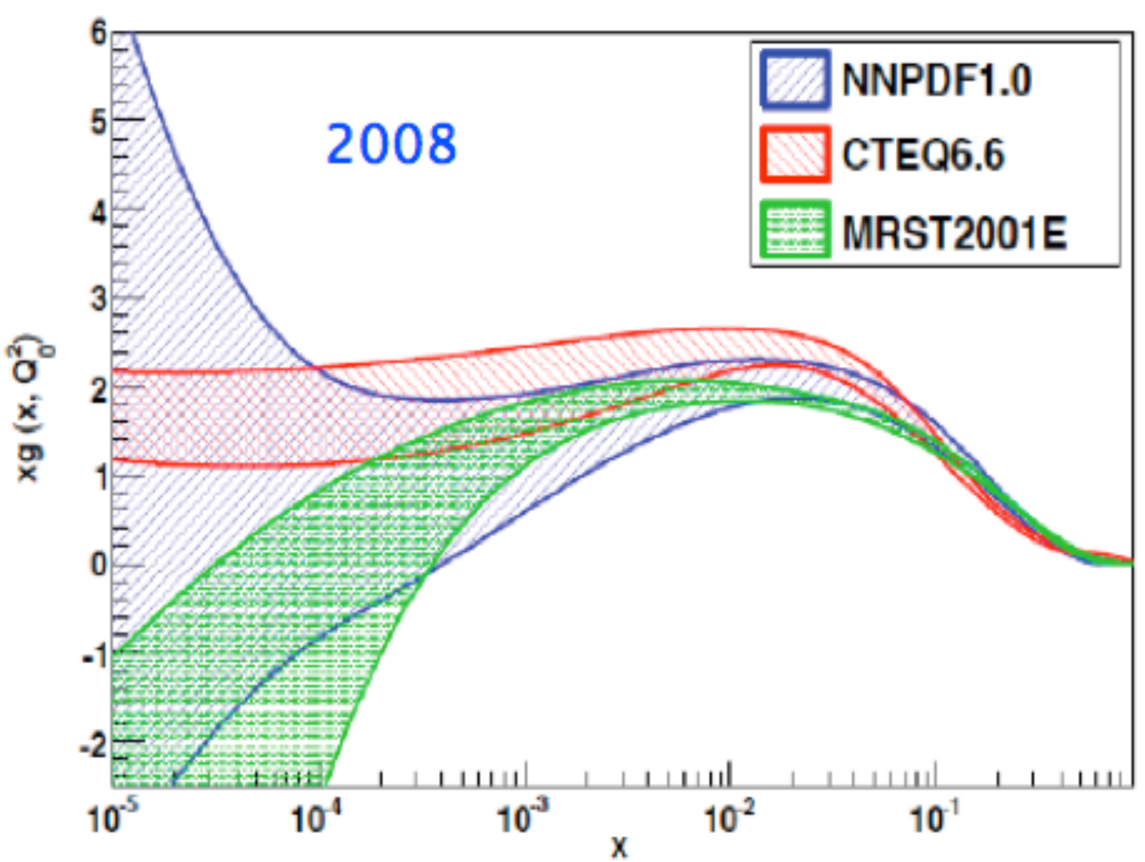


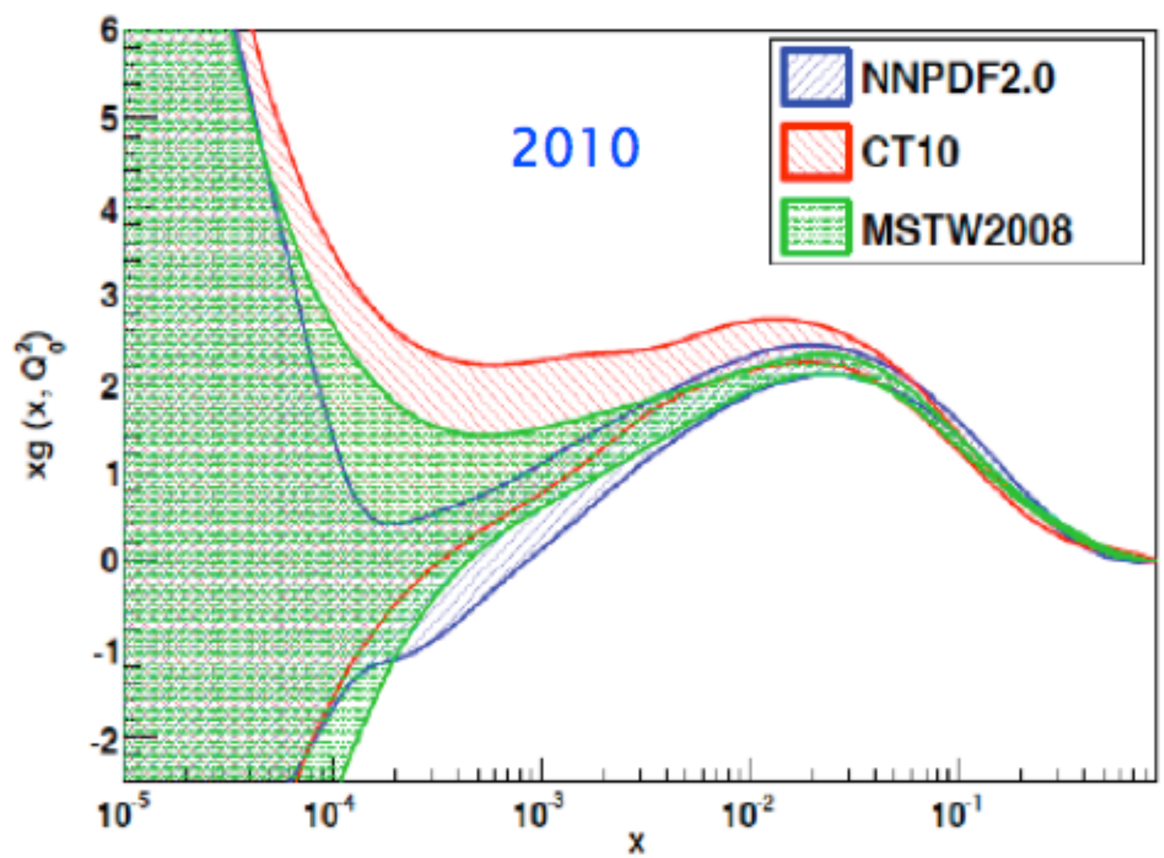
Pink and red curves give same good description of data but outside error bar

J. Pumplin ArXiv:0909.0268

Traditional (parametrical) approach

- Possible issues:
  If functional form not flexible enough PDFs may present unrealistically small errors where data do not constrain PDF uncertainties

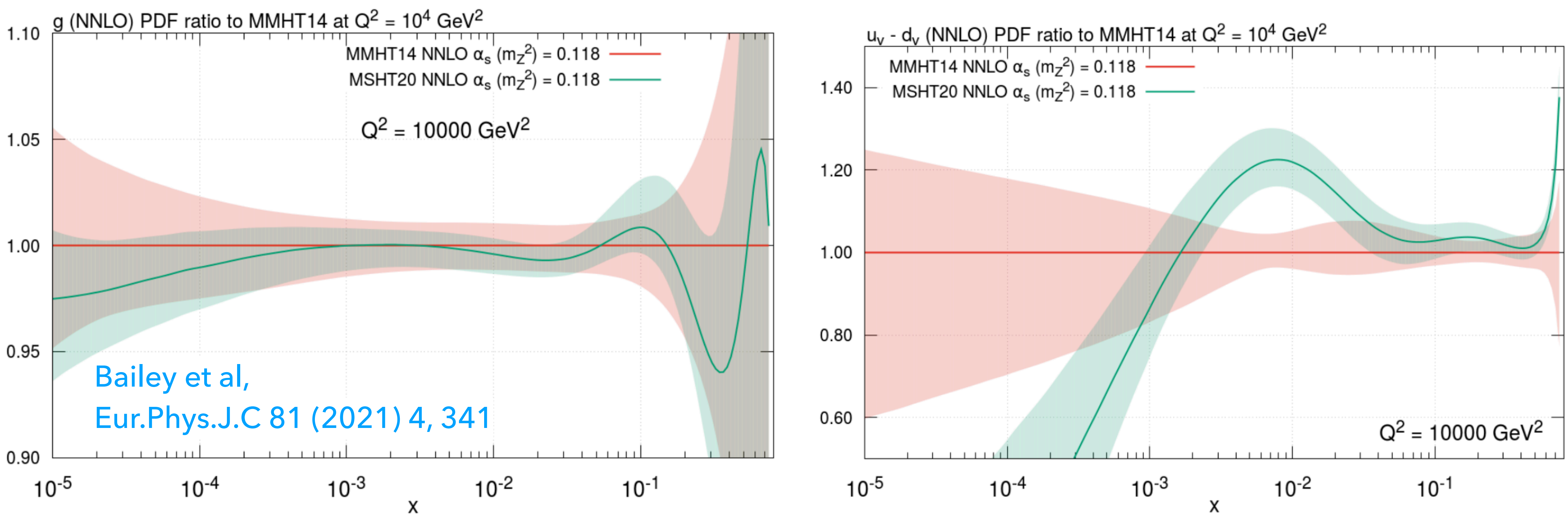




$$xg = A_g\, x^{\delta_g} (1-x)^{\eta_g} (1 + \epsilon_g \sqrt{x} + \gamma_g\, x) + A_{g'}\, x^{\delta_{g'}} (1-x)^{\eta_{g'}}$$

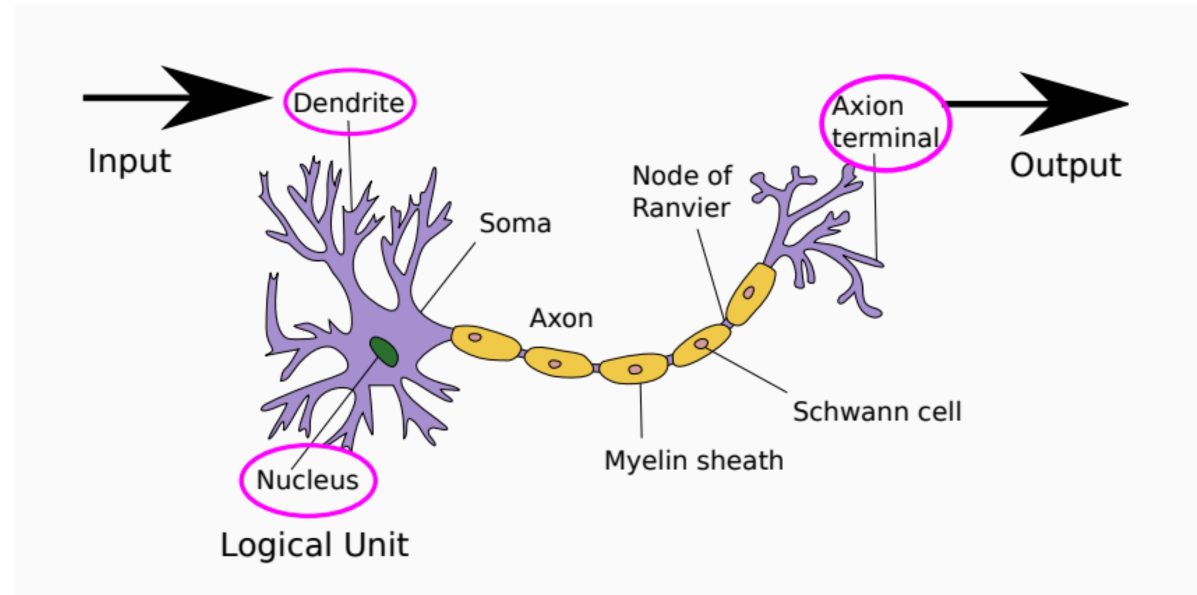
Traditional (parametrical) approach

- Possible issues:
  If functional form not flexible enough PDFs may be not able to adapt to new data



- In recent updates from a global PDF fitting collaborations (MSHT20) the effect of LHC data required big change in the parametrization which makes PDF uncertainty increase (data-driven parametrization)

Neural networks and ML

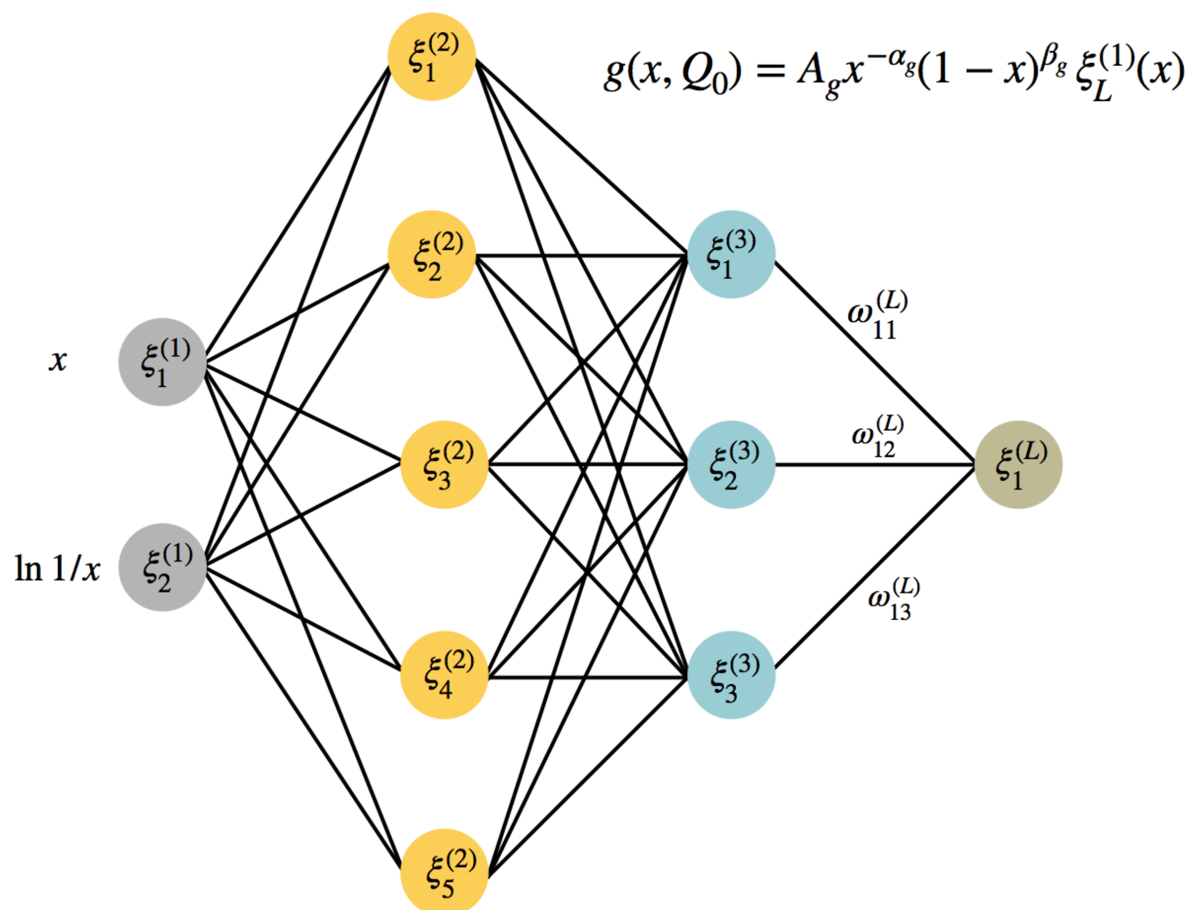


S. Carrazza, Colloquium, S. Paolo, N3PDF

- Artificial neural networks are computer systems inspired by the biological neural networks in the brain
- Data communication pattern
- Currently state-of-the-art for several Machine Learning Applications

Neural network parametrisation

Fully connected multi-layer perceptron



For a 1-2-1 feedforward neural network can write explicitly functional form

$$\xi_1^{(3)}(\xi_1^{(1)}) = \frac{1}{1+e^{\theta_1^{(3)} - \frac{\omega_{11}^{(2)}}{1+e^{\theta_1^{(2)} - \xi_1^{(1)}\omega_{11}^{(1)}}} - \frac{\omega_{12}^{(2)}}{1+e^{\theta_2^{(2)} - \xi_1^{(1)}\omega_{21}^{(1)}}}}}$$

- Neural Networks: all independent PDFs are associated to an unbiased and flexible parametrisation: O(300) parameters versus O(30) in polynomial parametrisation

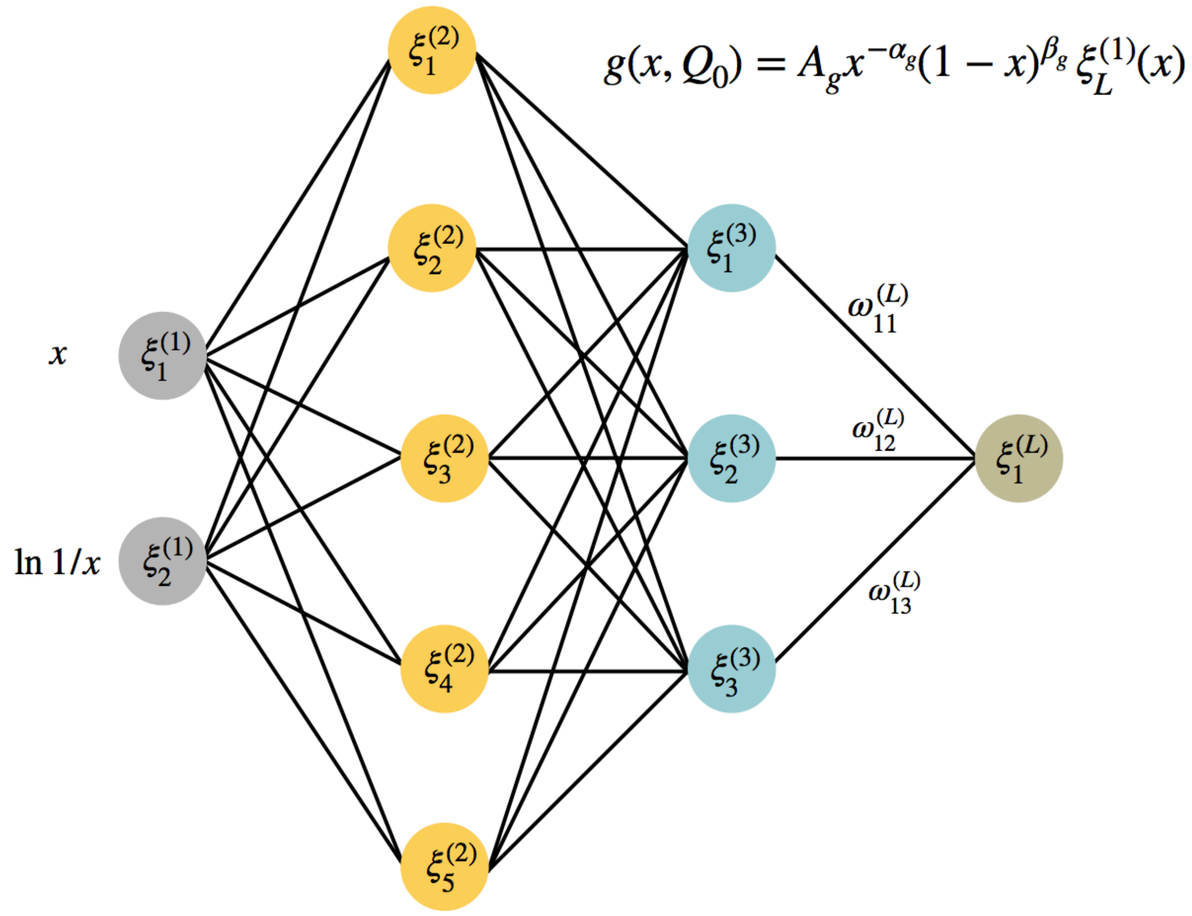
- 2-5-3-1 Neural network associated to each independent PDF (gluon, up, anti-up, down, anti-down, strange, anti-strange and charm)

$$\xi_i = g\left(\sum_j \omega_{ij}\xi_j - \theta_i\right)$$

$$g(x) = \frac{1}{1+e^{-x}}$$

Neural network training

Fully connected multi-layer perceptron



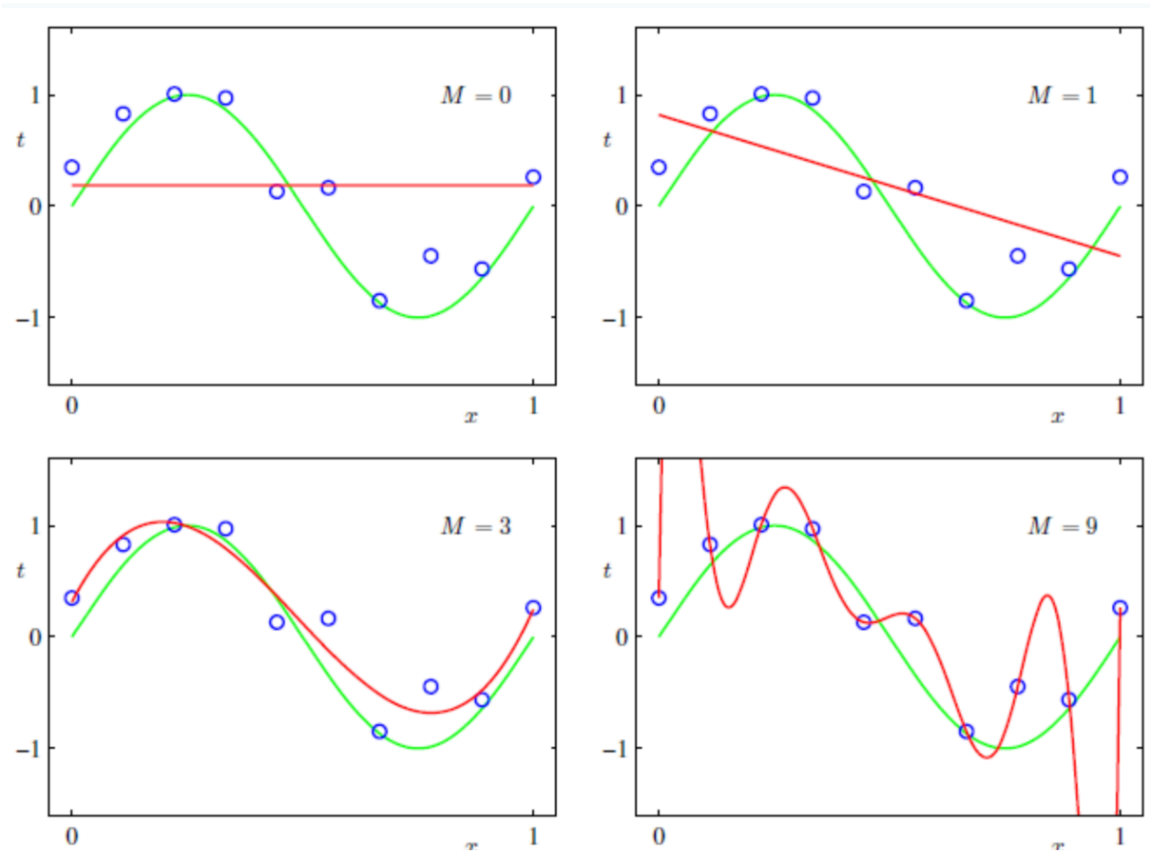
How do we train the 7(8) independent NN?

Minimise the cost function:

$$\chi^2 = \sum_{i,j=1}^{N_{\rm dat}} (D_i - T_i)(\text{cov})^{-1}_{ij}(D_j - T_j)$$

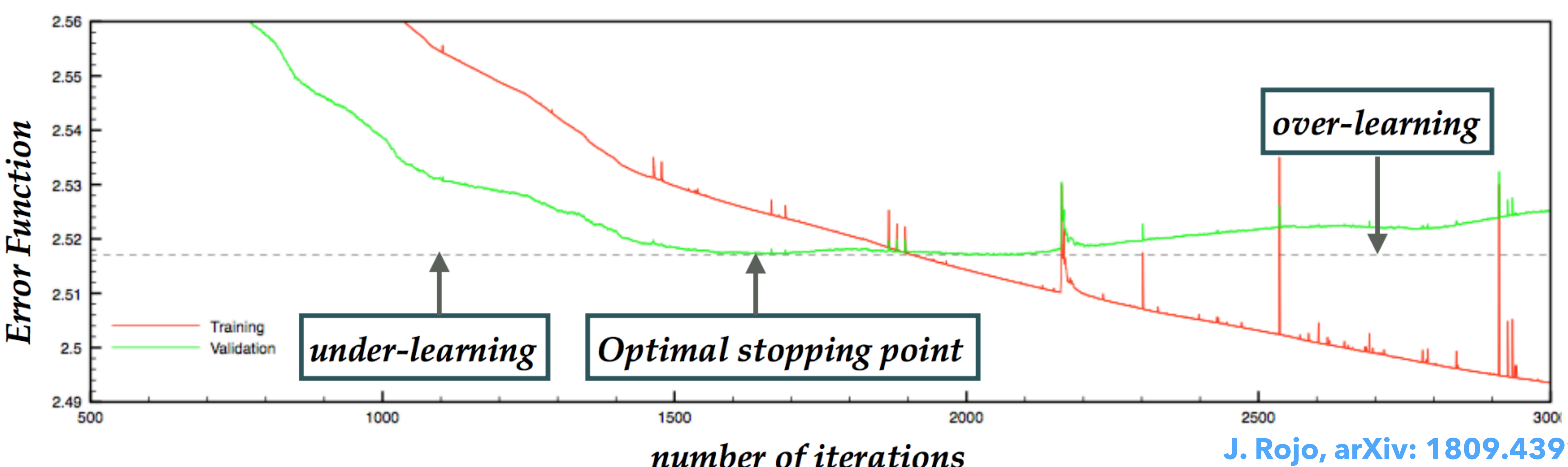
- D_i experimental measurement for the point i
- T_i theoretical prediction for the point i(depending on PDF parameters $\sigma_h = \sigma_{12} \otimes f_1 \otimes (f_2)$)
- $(\text{cov})_{ij}$ is the covariance matrix between point i and j with corrections for normalisation uncertainties
- Supplemented by additional penalty for positive observables

Neural network training



- Large parameter space: need an algorithm that is able to explore it without getting trapped in local minima such as genetic algorithm

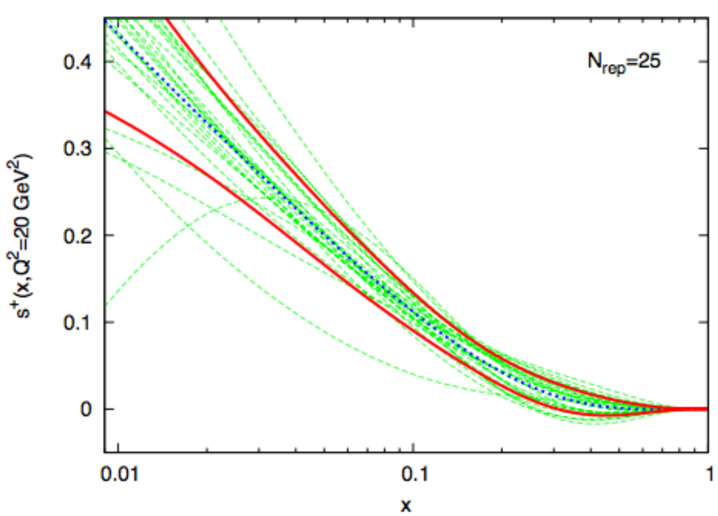
- Redundant parametrization: risk of over-fitting. Cross-validation necessary.



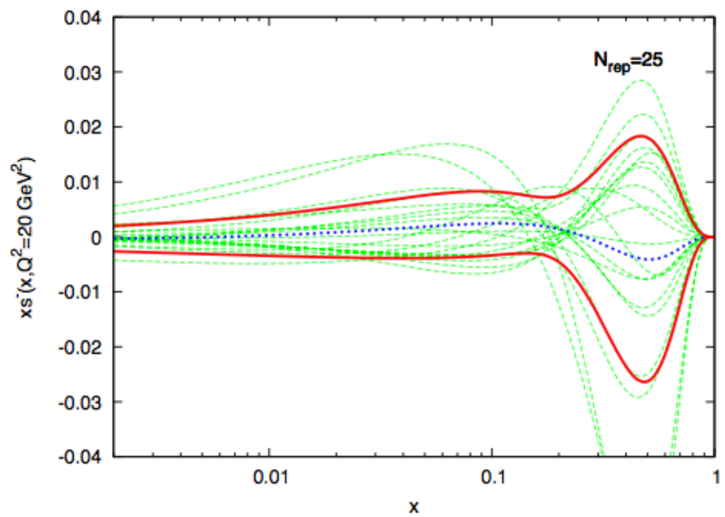
J. Rojo, arXiv: 1809.4392

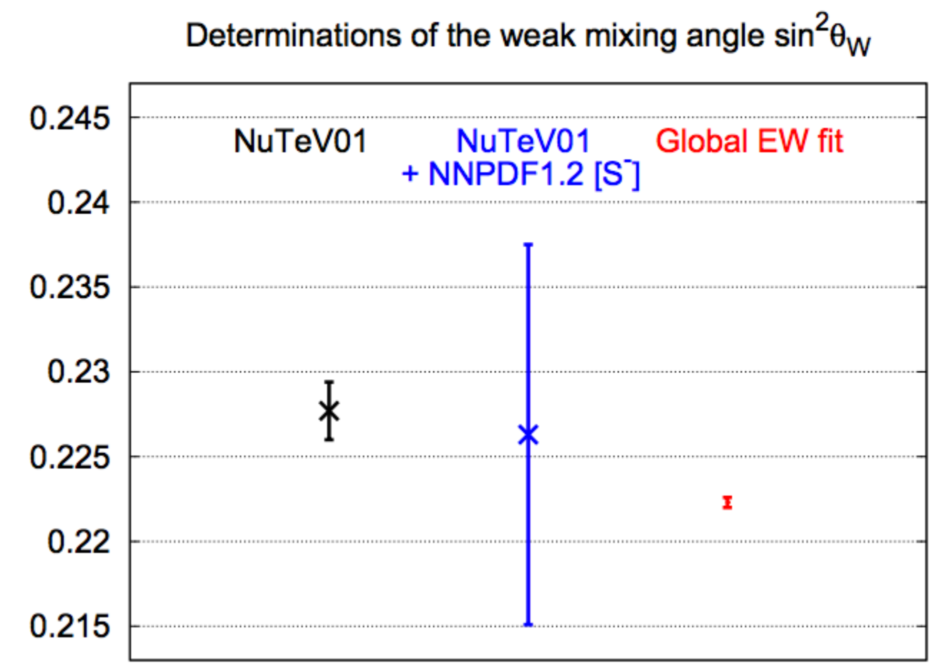
E.g. the NuTeV anomaly

Total strangeness ↓



Strange valence ↓





EW fit

$\sin^2\theta_W = 0.2223 \pm 0.0002$

NuTeV

$\sin^2\theta_W = 0.2276 \pm 0.0014$

$$\sin^2\theta_W\Big|_{\rm NuTeV} - \sin^2\theta_W\Big|_{\rm EW} = 0.0053$$

$$[F] = \int_0^1 dx\, x\, f(x, Q^2).$$

- $>3\sigma$ discrepancy between EW fits and NuTeV measurements
- Unbiased parametrisation of strangeness (2010) solved NuTeV anomaly

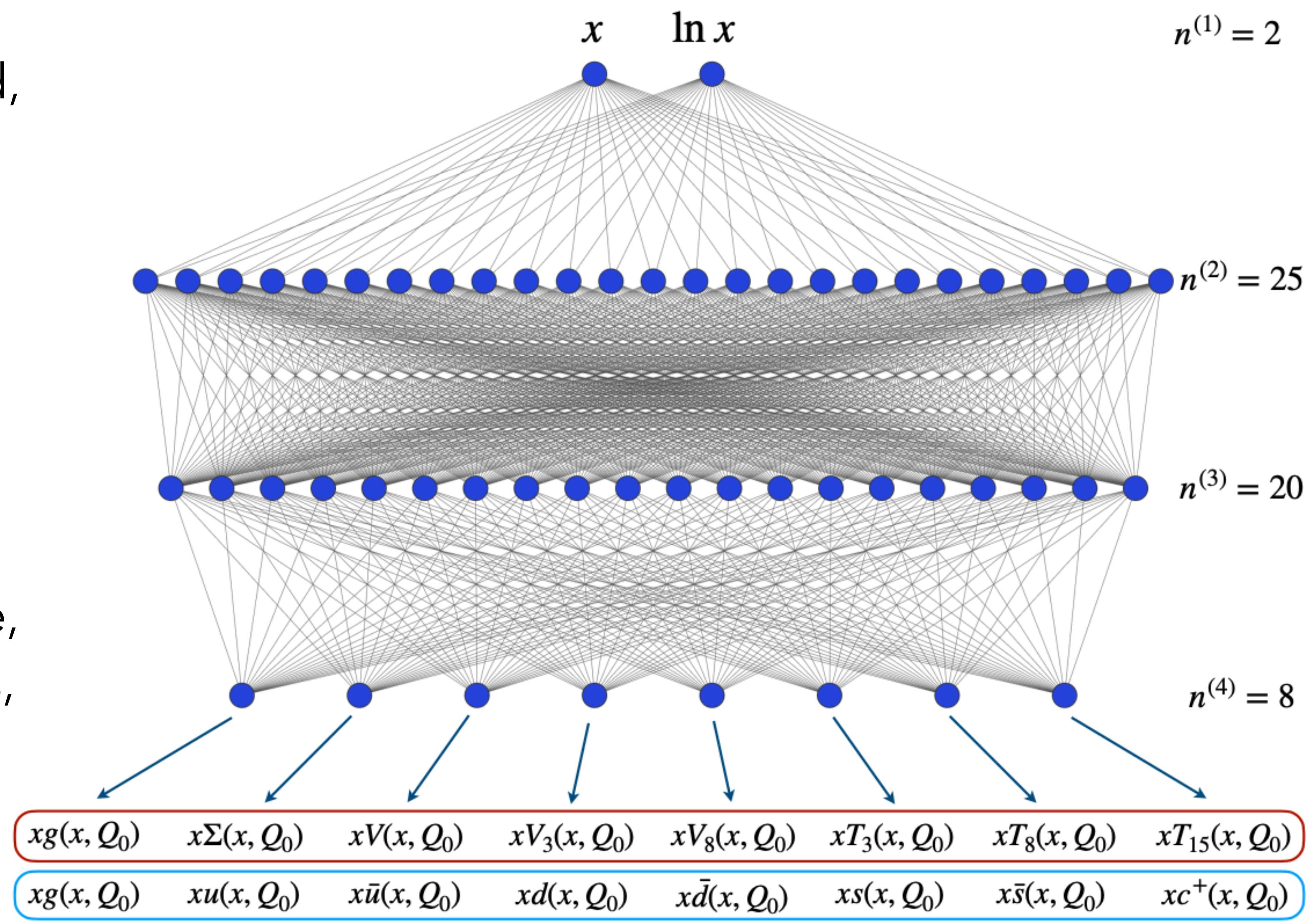
$$\delta_s \sin^2\theta_W \sim -0.240\frac{[S^-]}{[Q^-]}$$

$$\delta_s \sin^2\theta_W = -0.0005 \pm 0.0096^{\rm PDFs} \pm sys$$

Ball et al, 0906.1958

A deep-learning based fit

- Single neural network to parametrise 8 independent PDF combinations (g, u, d, s, u~, d~, s~, c=c~)
- New optimisation strategy based on gradient descent rather than genetic algorithm
- Hyper-optimised methodology: scan of the hyper parameter space to find optimal minimisation settings (optimiser, initialiser, stopping patience, number of layers, learning rate, epochs, activation function) by minimising χ^2_{val}
  [Carrazza et al, Eur.Phys.J.C 79 (2019) 8, 676]
- Statistical validation of PDF uncertainties via closure tests (data region)
  [Del Debbio et al, Eur.Phys.J.C 82 (2022) 4, 330]
  and future test (extrapolation region)
  [J. Cruz-Martinez et al, Acta Phys.Polon.B 52 (2021) 243]



NNPDF4.0, arXiv: 2109.02653

Error propagation

$$\langle \mathcal{O}[\{f\}] \rangle = \int [\mathcal{D}f]\, \mathcal{O}[\{f\}] \mathcal{P}[\{f\}].$$

- Given a finite number of experimental data points want a set of functions
- Want to find a infinite-dimensional object from a finite number of information

Option a) Project into a n-dimensional space of parameters which parametrise PDFs and use linear approximation around minimum χ^2

$$\langle \mathcal{O}[\{f\}] \rangle \simeq \int da_1 da_2 ... da_{N_{par}}\, \mathcal{O}[\vec{a}]\, \mathcal{P}[\vec{a}]$$

Hessian Method

Option b) Choose a parametrisation and perform a Monte Carlo sampling of probability density in functional space

$$\langle \mathcal{O}[\{f\}] \rangle \simeq \frac{1}{N_{\rm rep}} \sum_{i=1}^{N_{\rm rep}} \mathcal{O}[f_i],$$

Monte Carlo Method

Hessian method

➡ Used by most PDF fitters (CTEQ/TEA, MSTW/MMHT, HERAPDF, ABM)

➡ Pick a functional form and project problem in the N_{par}-dimensional space of parameters (typically 15 - 25)

➡ Determine best fit values of parameters $\{\vec{a}_0\}$

➡ Shift $\vec{a} \to \vec{a} - \vec{a}_0$

➡ Determine error on PDFs and any observable depending on PDFs (all denoted by X) by propagation of the error in the parameter space

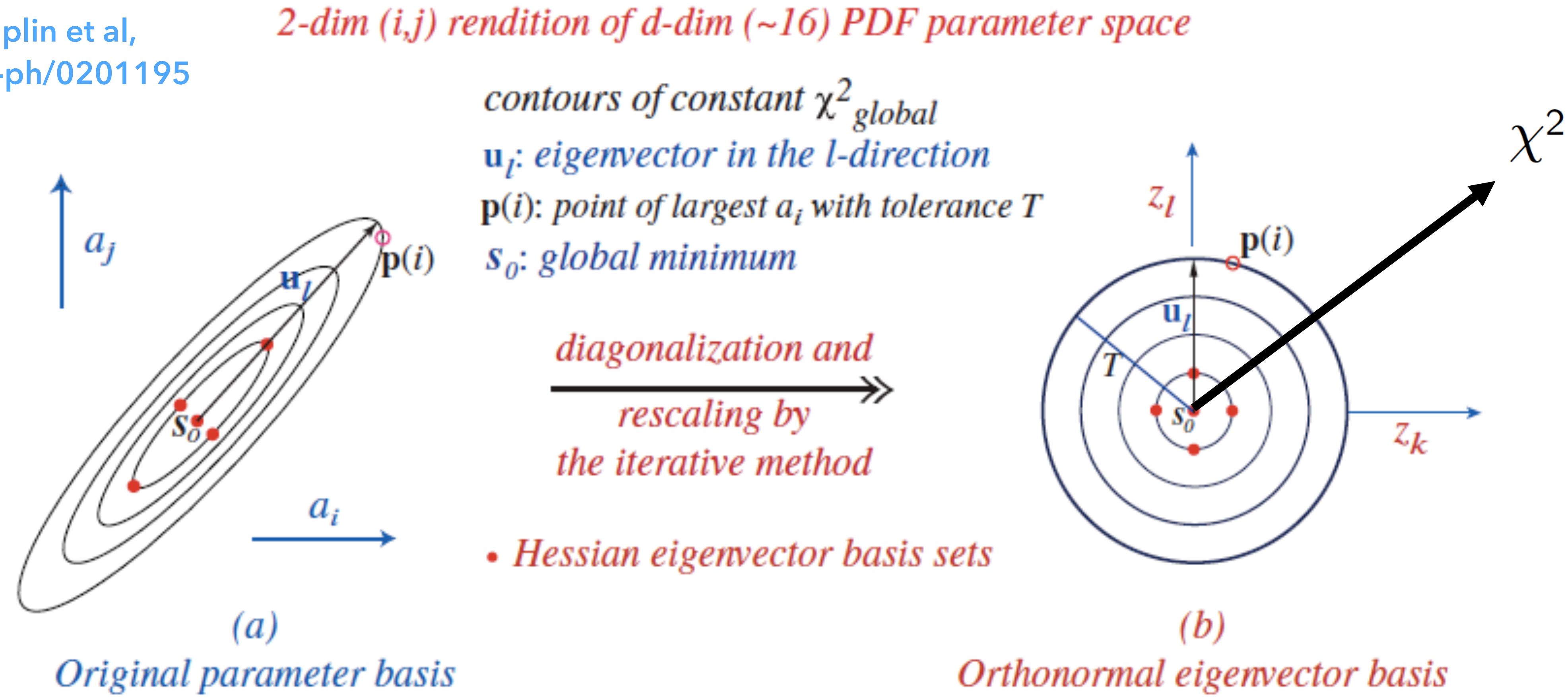
Assuming linear prop: $X(\vec{a}) \simeq X(\vec{0}) + a_i \partial_i X(\vec{a})\big|_{\vec{a}=\vec{0}}$

Variance: $\sigma_X^2 = (\mathrm{cov})_{ij}\, \partial_i X \partial_j X$ $(\mathrm{cov})_{IJ}$ covariance matrix in param, space

Maximum likelihood: $(\mathrm{cov})_{ij} = (H)_{ij} = \left.\dfrac{\partial^2 \chi^2(\vec{a})}{\partial_i a \partial_j a}\right|_{\vec{a}=\vec{0}}$ cov $\Longleftrightarrow$ Hessian at the minimum of χ^2

Hessian method

Pumplin et al,
hep-ph/0201195



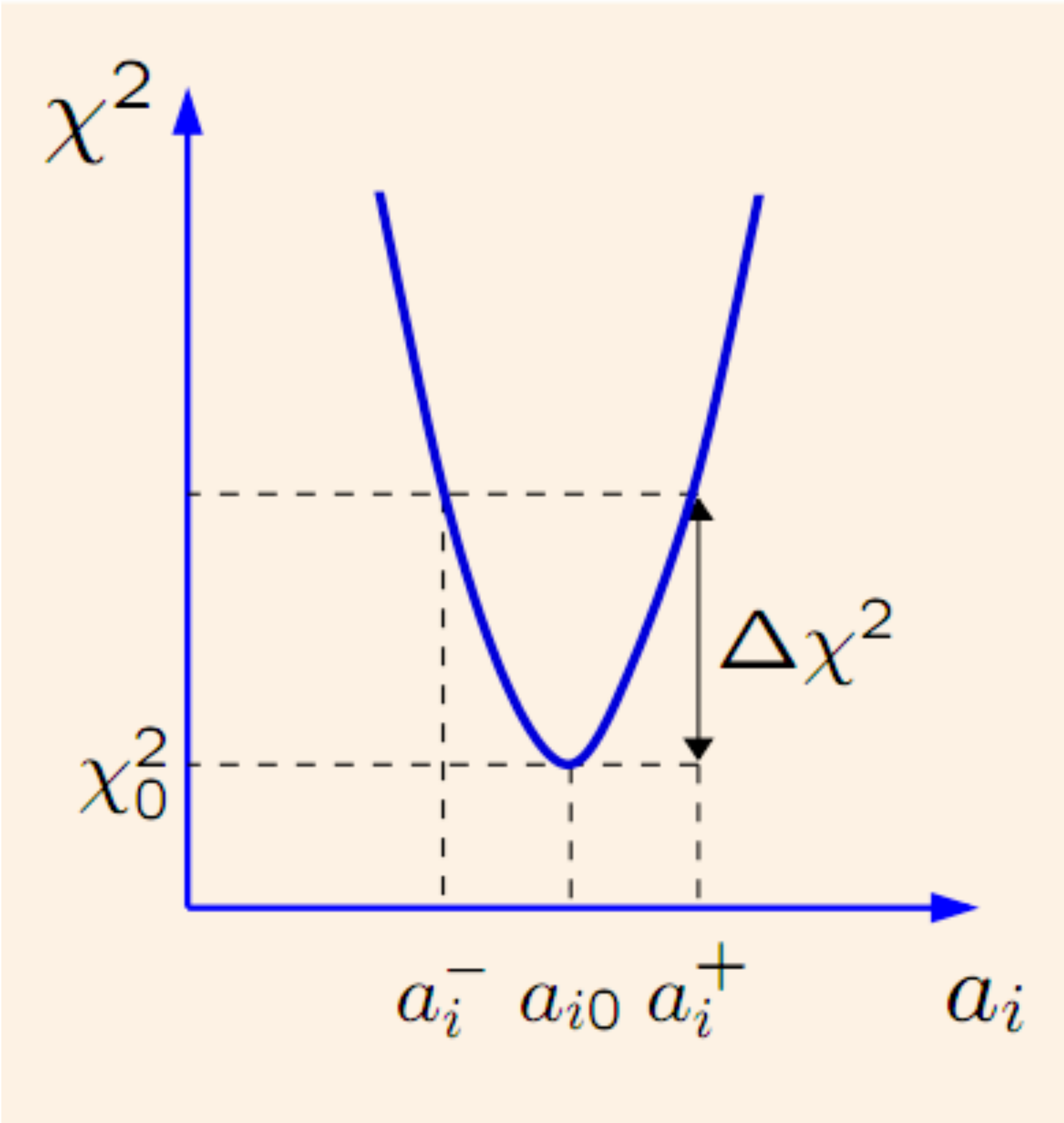
$$\sigma_X^2 = (H)_{ij}\,\partial_i X\,\partial_j X \xrightarrow{\text{diagonalisation}} \sigma_X^2 = |\vec{\nabla} X|^2$$

z_i eigenvectors of H with unit eigenvalues

➡ The total uncertainty is the sum in quadrature of the uncertainties due to each parameter

➡ $\Delta\chi^2 = \sum z_i^2$ the surfaces of constant χ^2 are spheres in the z space of radius $\sqrt{\Delta\chi^2}$

Hessian method



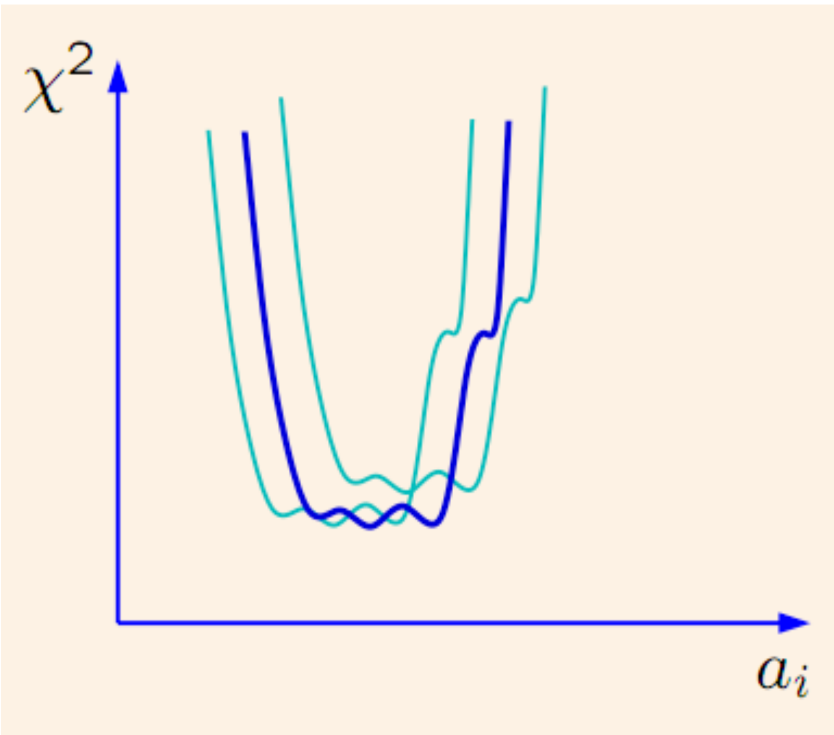
P. Nadolsky, CTEQ summer school 2009

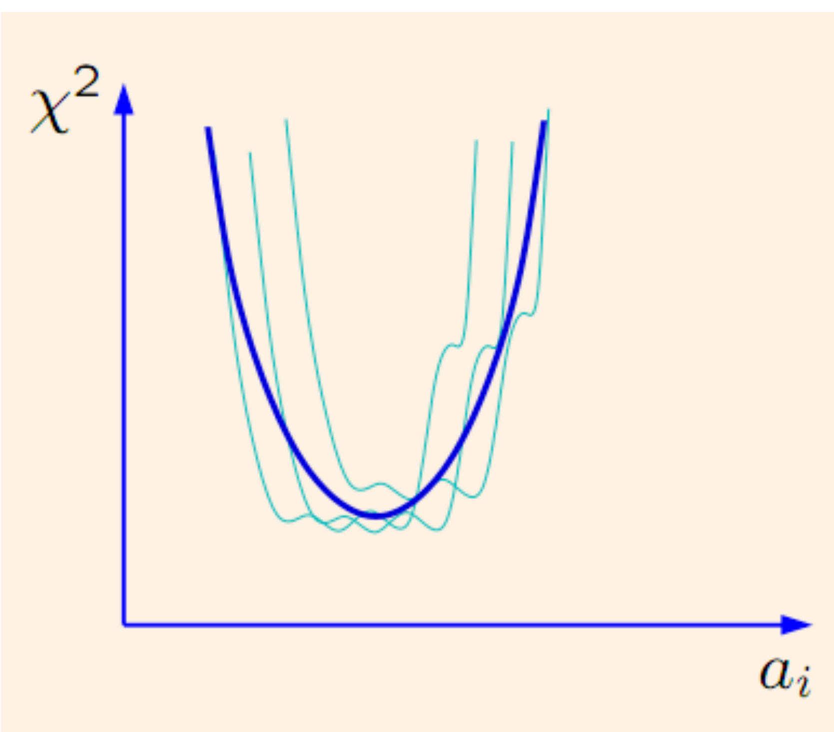
- According to textbook statistics, the 1σ contour in parameter space is given by

$$\Delta\chi^2 = 1$$

- Projection of the radius one sphere would give the uncertainty on parameters and on the PDFs, observables…
- The textbook statistics should work in case of perfectly compatible Gaussian errors
- But in practice, for global fits a tolerance is introduced
- NB: introducing a tolerance corresponds to blow up uncertainties by a factor $\sqrt{\Delta\chi^2}$

Hessian method





The actual χ^2 function displays

- A well pronounced global minimum χ_0^2
- Some tensions between datasets in the vicinity of the minimum
- Some dependence on assumptions about flat directions (= unconstrained combinations of PDF parameters)

The likelihood is approximately described by a quadratic χ^2 with a revised tolerance condition

$$\Delta\chi^2 \leq T^2$$

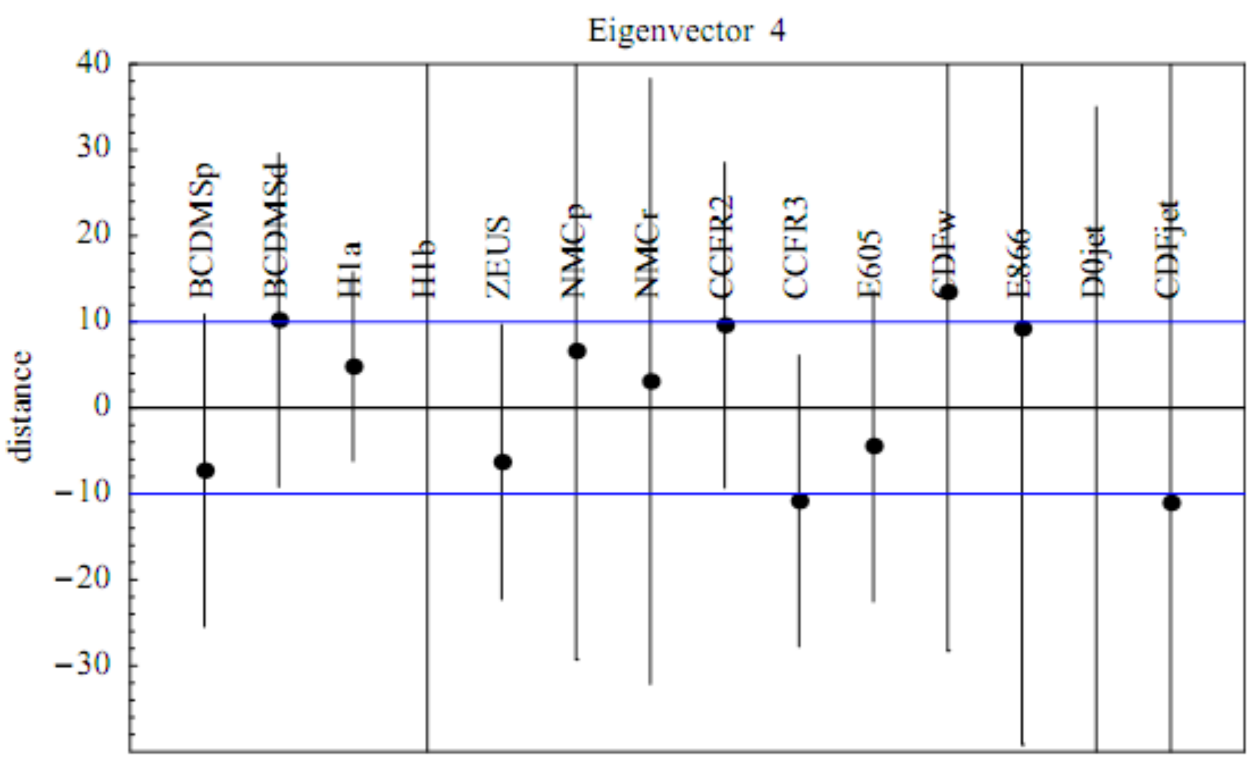
P. Nadolsky, CTEQ summer school 2009

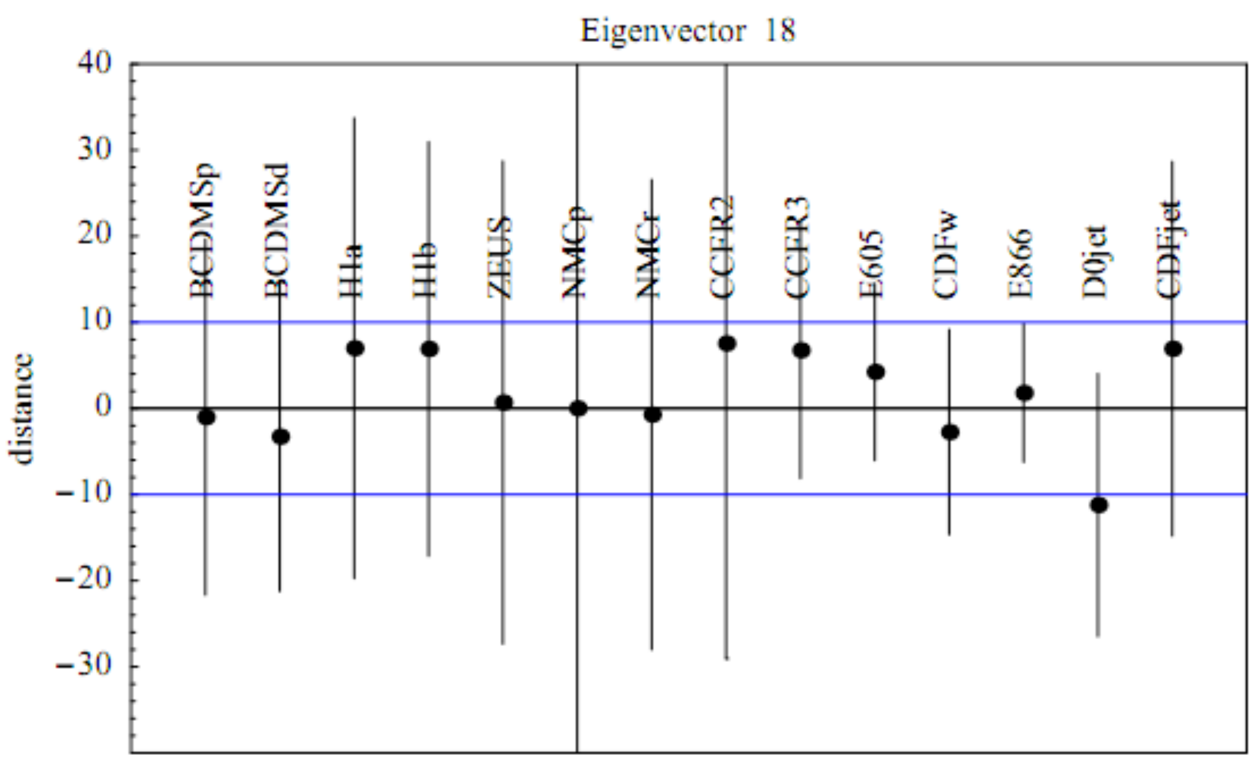
Hessian method

CTEQ6 tolerance criterion

- Acceptable values of PDF parameters must agree at ~ 90% C.L. with all experiments included in the fit, for a plausible range of assumptions about the PDF parametrisation, scale dependence, systematic uncertainty

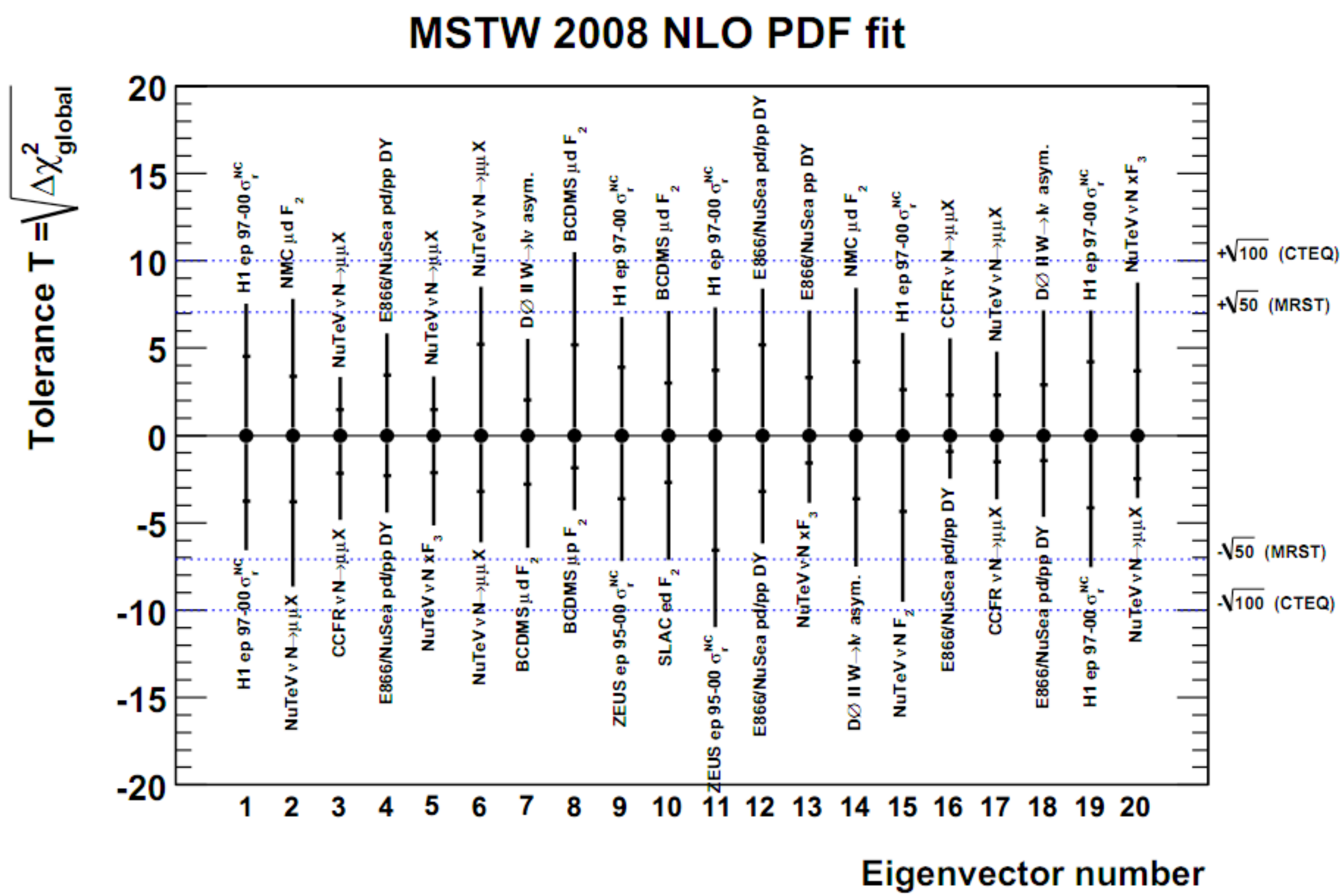
- Can be crudely approximated by assuming T ~ 10 for all PDF parameters





Hessian method

MSTW08 tolerance criterion



A dynamical tolerance, which varies according to the considered parameter

Monte Carlo method

- ➡ First idea by Giele Keller Kosover (hep-ph/0104052)
- ➡ Monte Carlo in parameter space

$$X(\vec{a}) \xrightarrow{\text{MC sampling}} \langle X \rangle = \int d\vec{a} X[\vec{a}]\mathcal{P}[\vec{a}]$$

P probability of parameter values

MC sampling in **parameter** space

$$\langle X \rangle \sim \frac{1}{N_{\rm rep}} \sum_{i=1}^{N_{\rm rep}} X(\vec{a}_i)$$

$$\sigma_X^2 = \langle X^2 \rangle - \langle X \rangle^2$$

Problem

How many replicas are needed?
Three bins per parameter ⇒ $3^{N_{par}}$ bins

E.g. for 23 parameters need more than 10^{11} replicas!!!

Monte Carlo method

- Forte, J. I Latorre, Piccione (hep-ph/0701127)
- First applied to structure functions then to PDFs

$$X(\vec{a})$$

MC sampling

$$\langle X \rangle = \int d\vec{a} X[\vec{a}]\mathcal{P}[\vec{a}]$$

P probability of parameter values

MC sampling in **data** space

$$\langle X \rangle \sim \frac{1}{N_{\rm rep}} \sum_{i=1}^{N_{\rm rep}} X(\vec{a}_i)$$

$$\sigma_X^2 = \langle X^2 \rangle - \langle X \rangle^2$$

Idea

Choose parameters along ∇X $\Longleftrightarrow$
Choose replicas of the data, i.e. work in the space of data and project back into PDF space

How many replicas does one need? 1-dim average of N_{rep} converges to true average with standard deviation $\sigma/\sqrt{N_{rep}}$

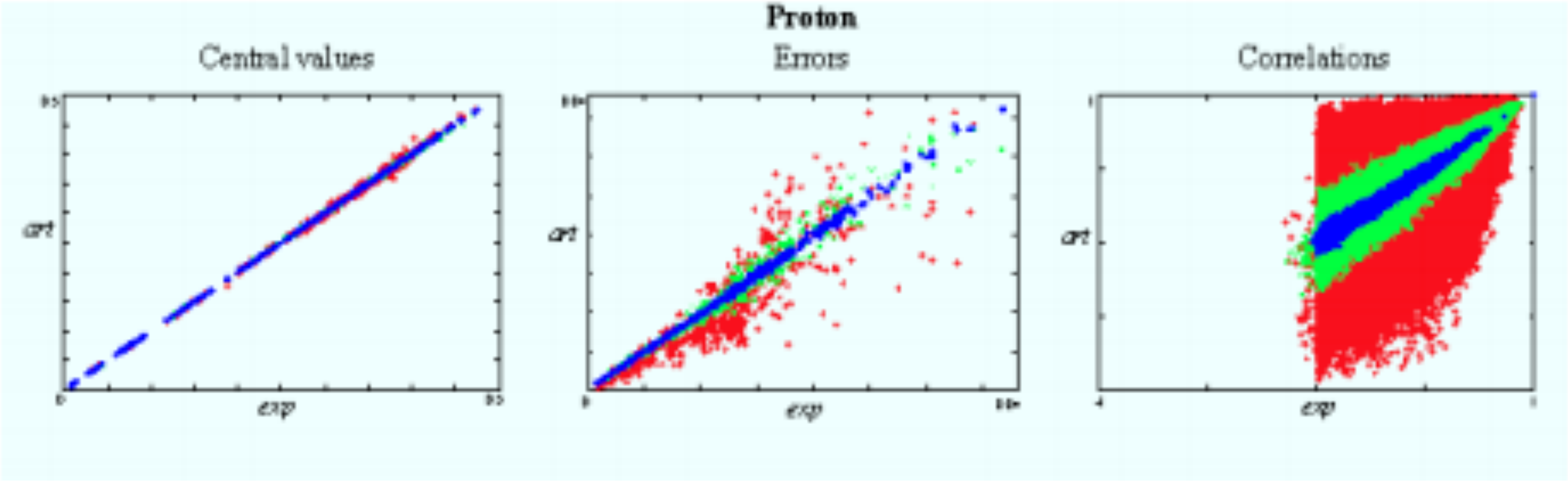
E.g. 10 replicas are enough for getting "true" central value with σ/3 accuracy

Monte Carlo method

- Generate artificial data according to distribution

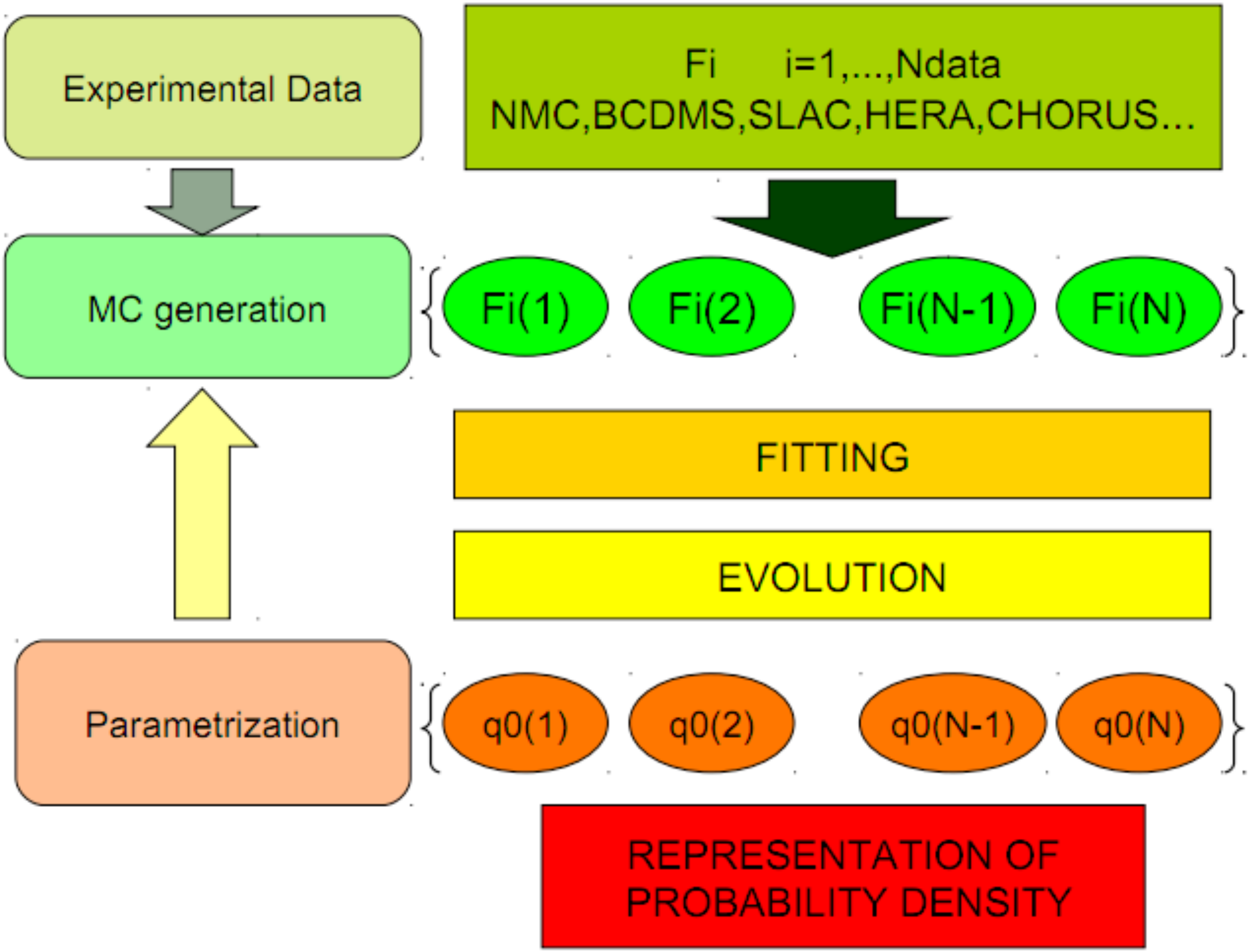
$$F_p^{(\mathrm{art})(k)} = S_{p,N}^{(k)} F_p^{(exp)} \left(1 + \sum_{l=1}^{N_c} r_{p,l}^{(k)} \sigma_{p,l} + r_p^{(k)} \sigma_{p,s} \right)$$

- r_i are univariate Gaussian random numbers such that if two points have correlated systematic uncertainties, they oscillate in the same directions
- S normalisation factors
- Validate Monte Carlo replicas against experimental data

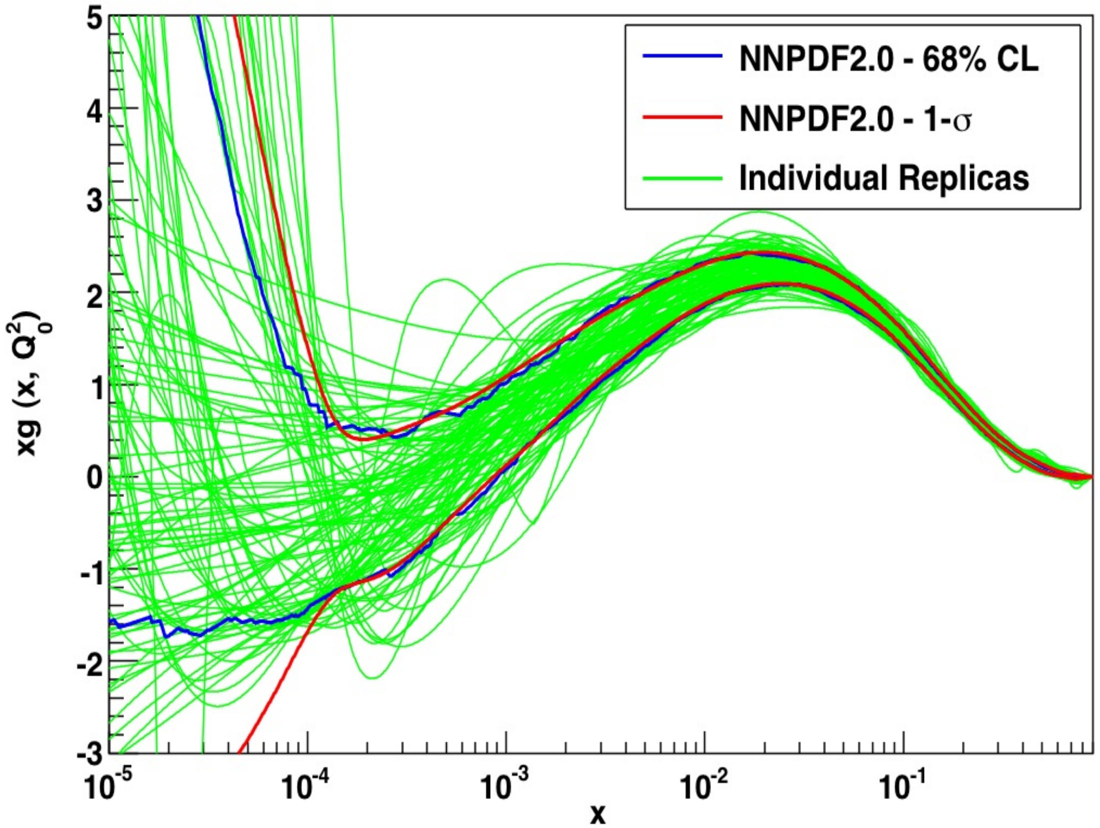


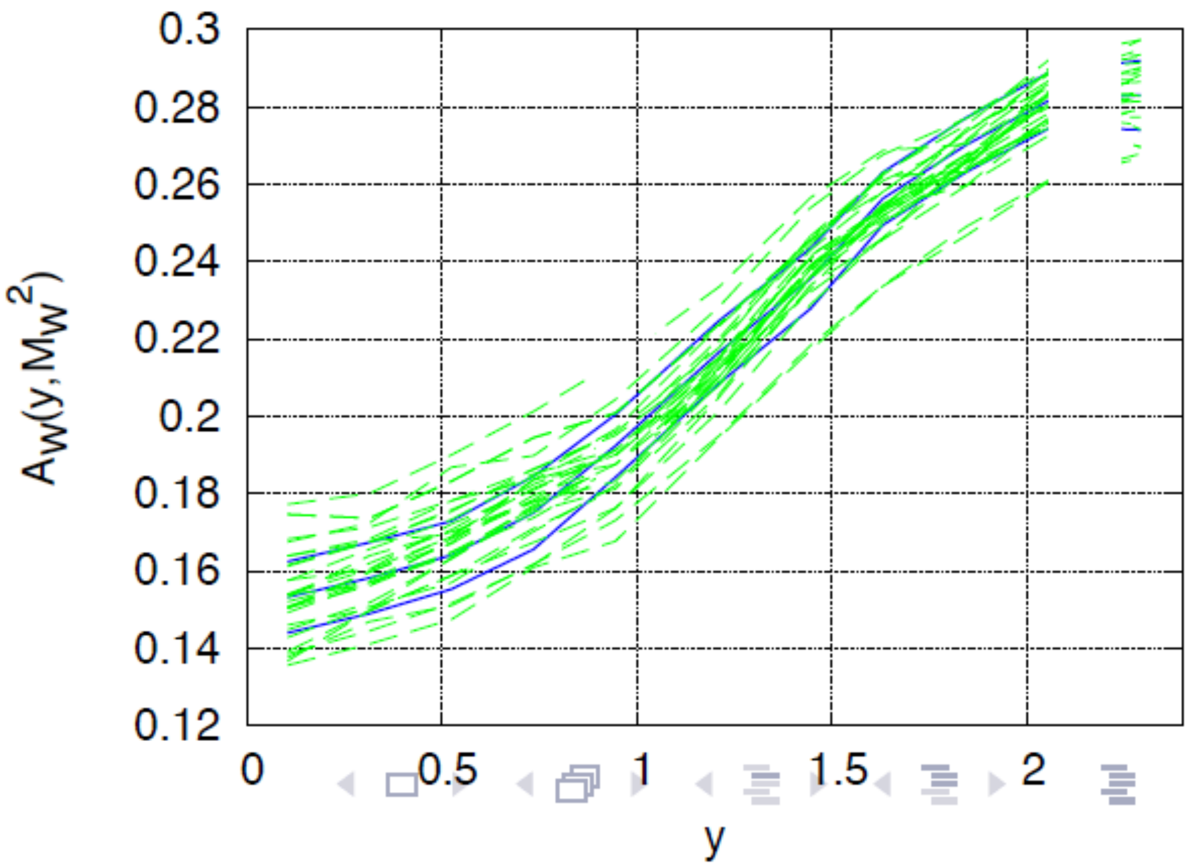
- Convergence rate increases with N_{rep}
- Correlations reproduced to % accuracy with 1000 reps

Monte Carlo method



Monte Carlo method





$$\langle f_J \rangle = \frac{1}{N_{\rm rep}} \sum_{k=1}^{N_{\rm rep}} f_J^{(k)}$$

$$\sigma_X^2 = \langle X^2 \rangle - \langle X \rangle^2$$

$$\langle A(\{f\}) \rangle = \frac{1}{N_{\rm rep}} \sum_{k=1}^{N_{\rm rep}} A(\{f^{(k)}\})$$

Individual replicas may fluctuate significantly, average quantities such as central values and 1σ error bands are smooth inasmuch as stability is reached due to the dimension of the ensemble increasing

The NNPDF solution

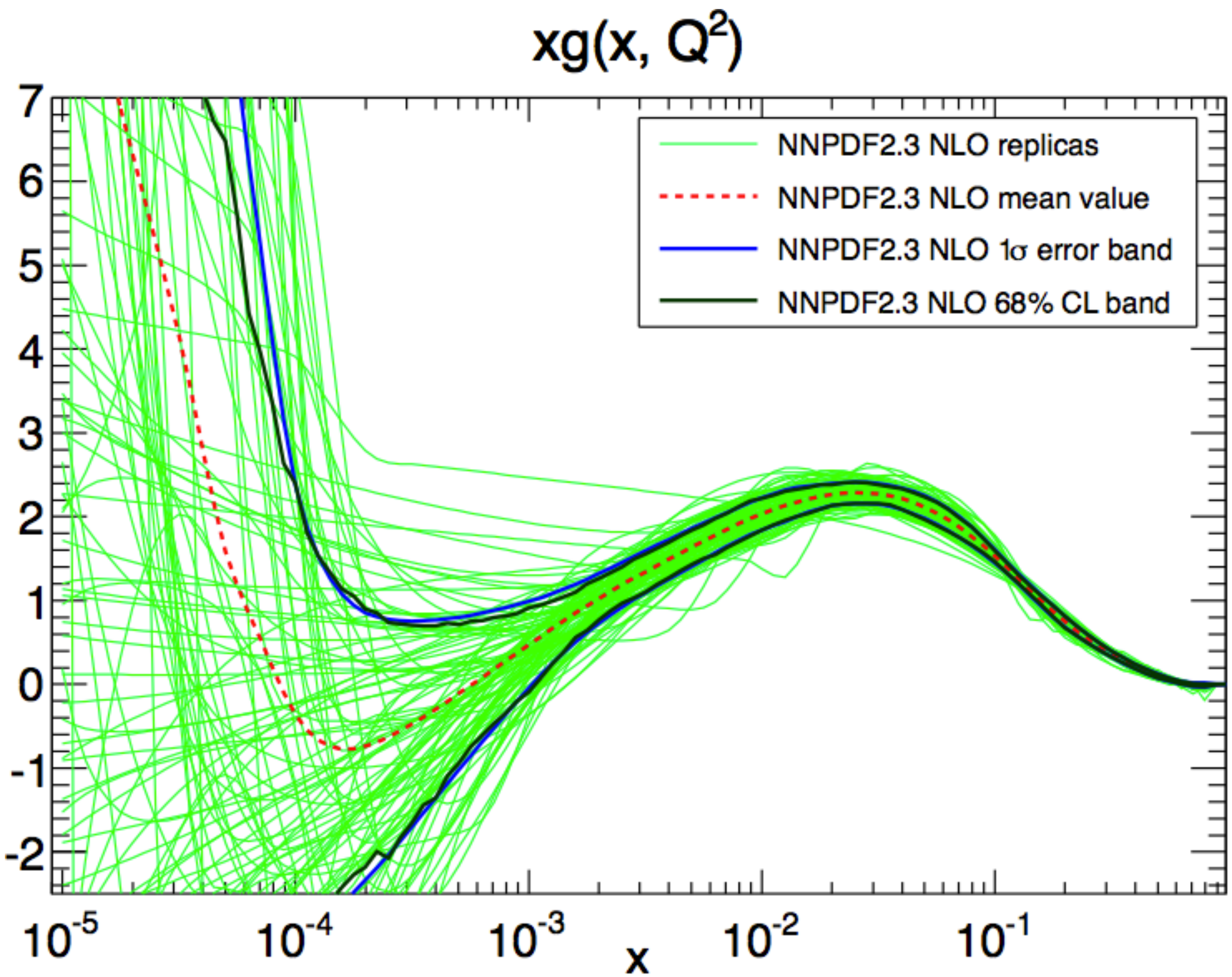
Monte Carlo sampling

Neural Network

http://nnpdf.mi.infn.it

- Fit of structure function **(2005)**
- DIS-only fit of PDFs **(2008)**
- First NNPDF global fit **(2010)**
- First fit including LHC data **(2013)**
- Closure test **(2016)**
- Fitted charm **(2018)**
- …

The NNPDF solution



The N(eural)N(etwork)PDFs:

- Monte Carlo techniques: sampling the probability measure in PDF functional space
- Neural Networks: all independent PDFs are associated to single NN

Summary for the user

Hessian method (CT, CJ, MSTW, ABKM, HERAPDF)

$$\langle \mathcal{F} \rangle = \mathcal{F}[q^{(0)}]$$

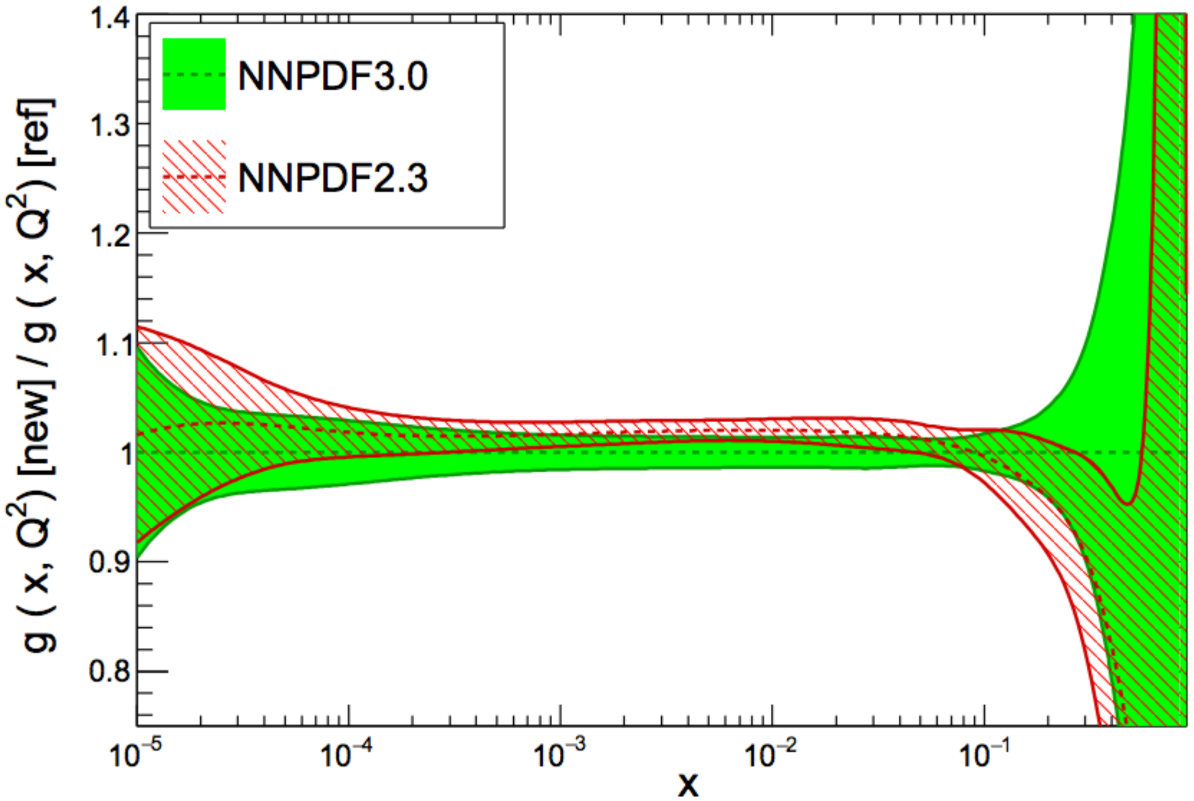
$$\sigma_{\mathcal{F}}^{\mathrm{Hess}} = \frac{1}{2} \left(\sum_{k=1}^{N_{\mathrm{set}/2}} \left(\mathcal{F}[\{q^{(2k-1)}\}] - \mathcal{F}[\{q^{(2k)}\}] \right)^2 \right)^{1/2}$$

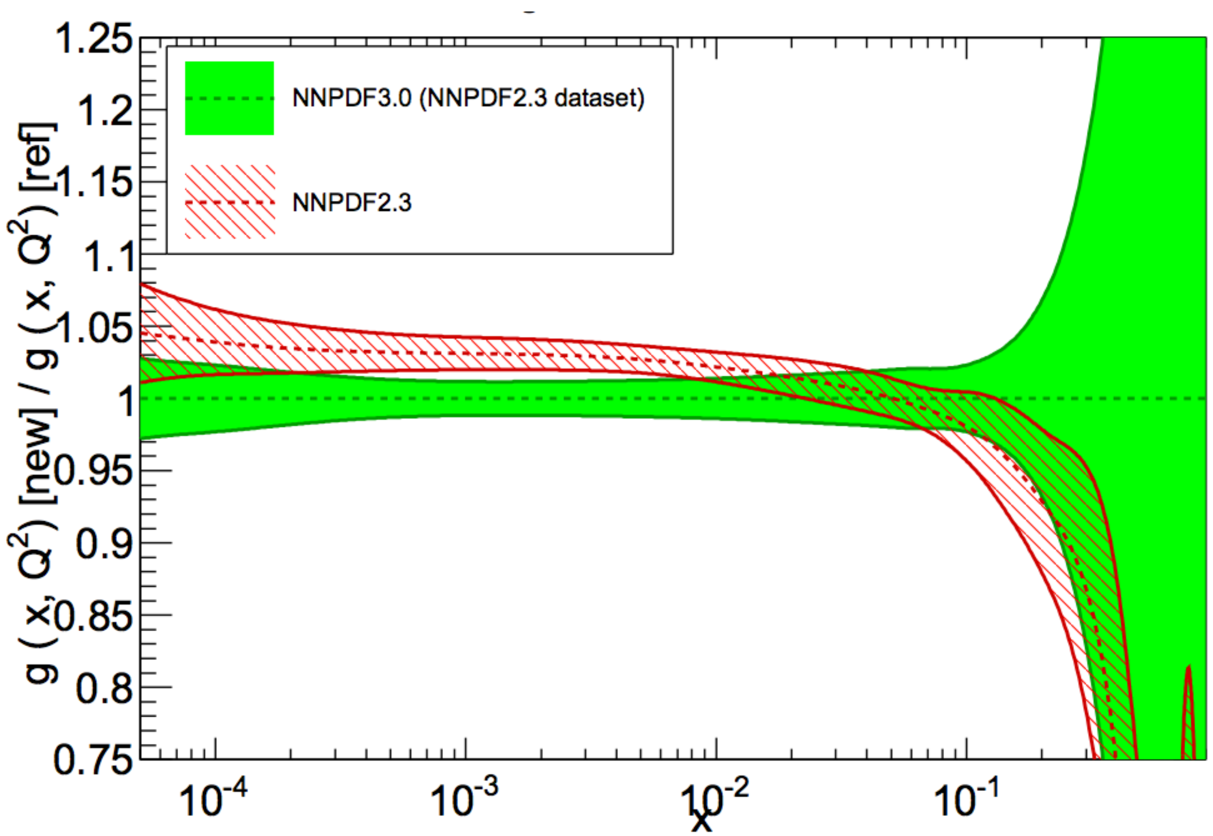
Monte Carlo method (NNPDF)

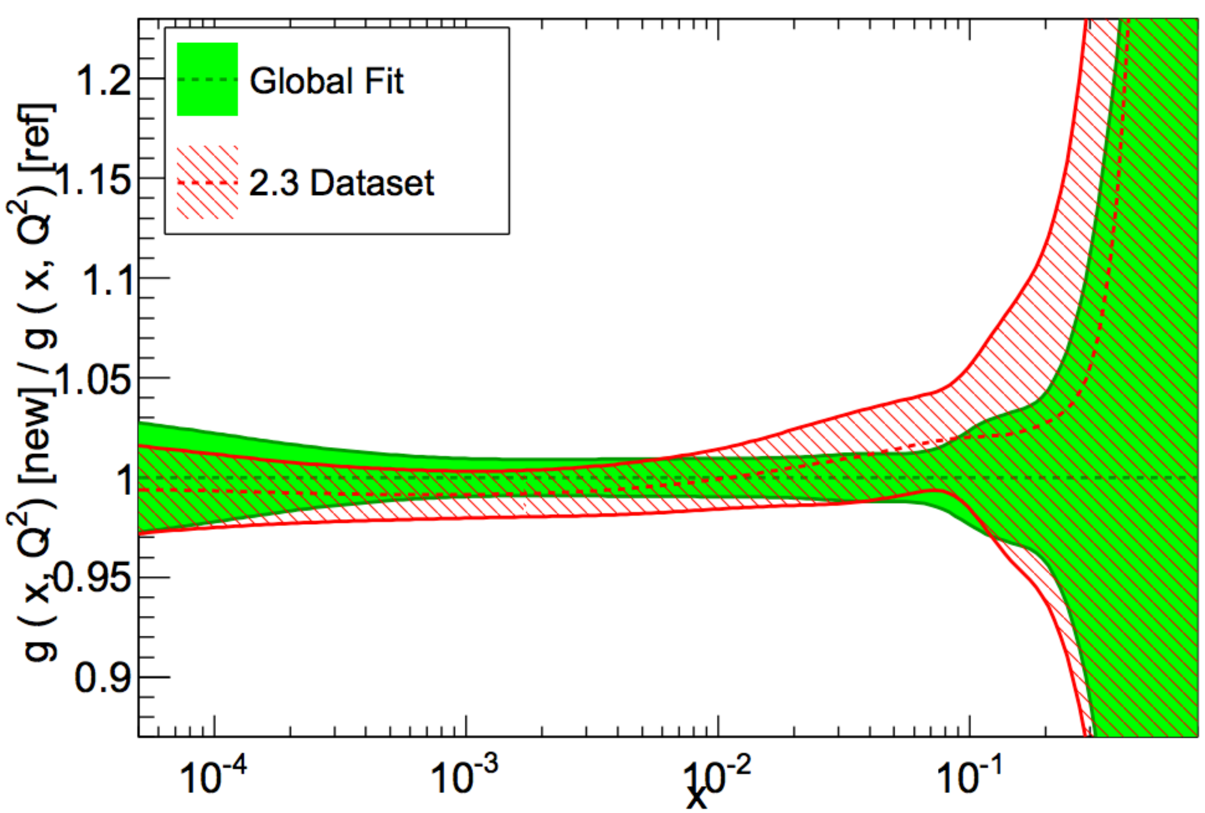
$$\langle \mathcal{F} \rangle = \frac{1}{N_{\mathrm{set}}} \sum_{i=1}^{N_{\mathrm{set}}} \mathcal{F}[q^{(i)}]$$

$$\sigma_{\mathcal{F}}^{\mathrm{MC}} = \left(\frac{1}{N_{\mathrm{N_{set}}}} \sum_{k=1}^{N_{\mathrm{set}}} \left(\mathcal{F}[\{q^{(k)}\}] - \langle \mathcal{F}[\{q\}] \rangle \right)^2 \right)^{1/2}$$

Key issue: methodology





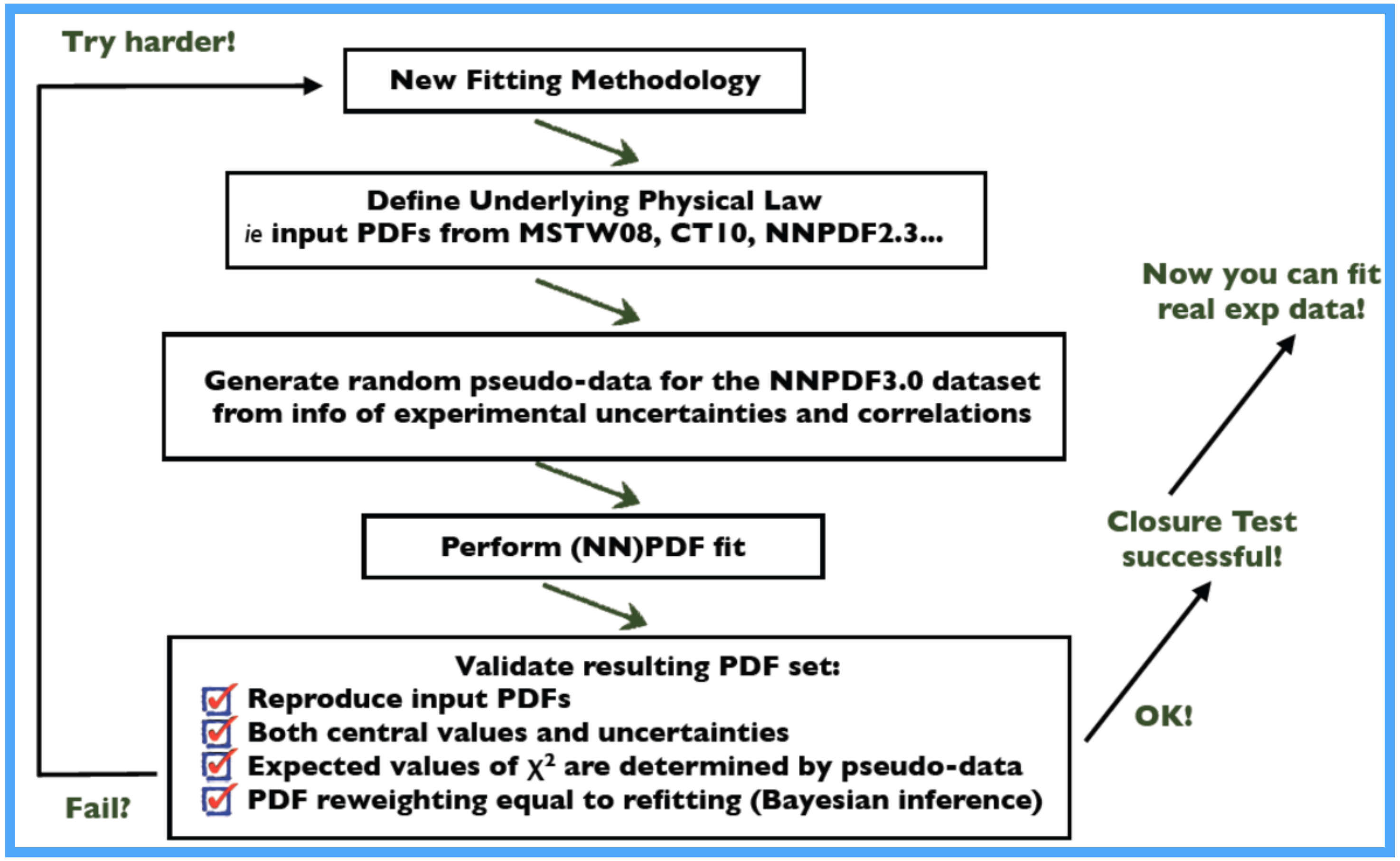


- NNPDF2.3 -> NNPDF3.0: included many new data (LHC and combined HERA) & change in fitting methodology (genetic algorithm and stopping criterion)
- Main changes in the gluon are due to the change in methodology
- How to make sure that we have a "perfect" methodology?

Statistical validation

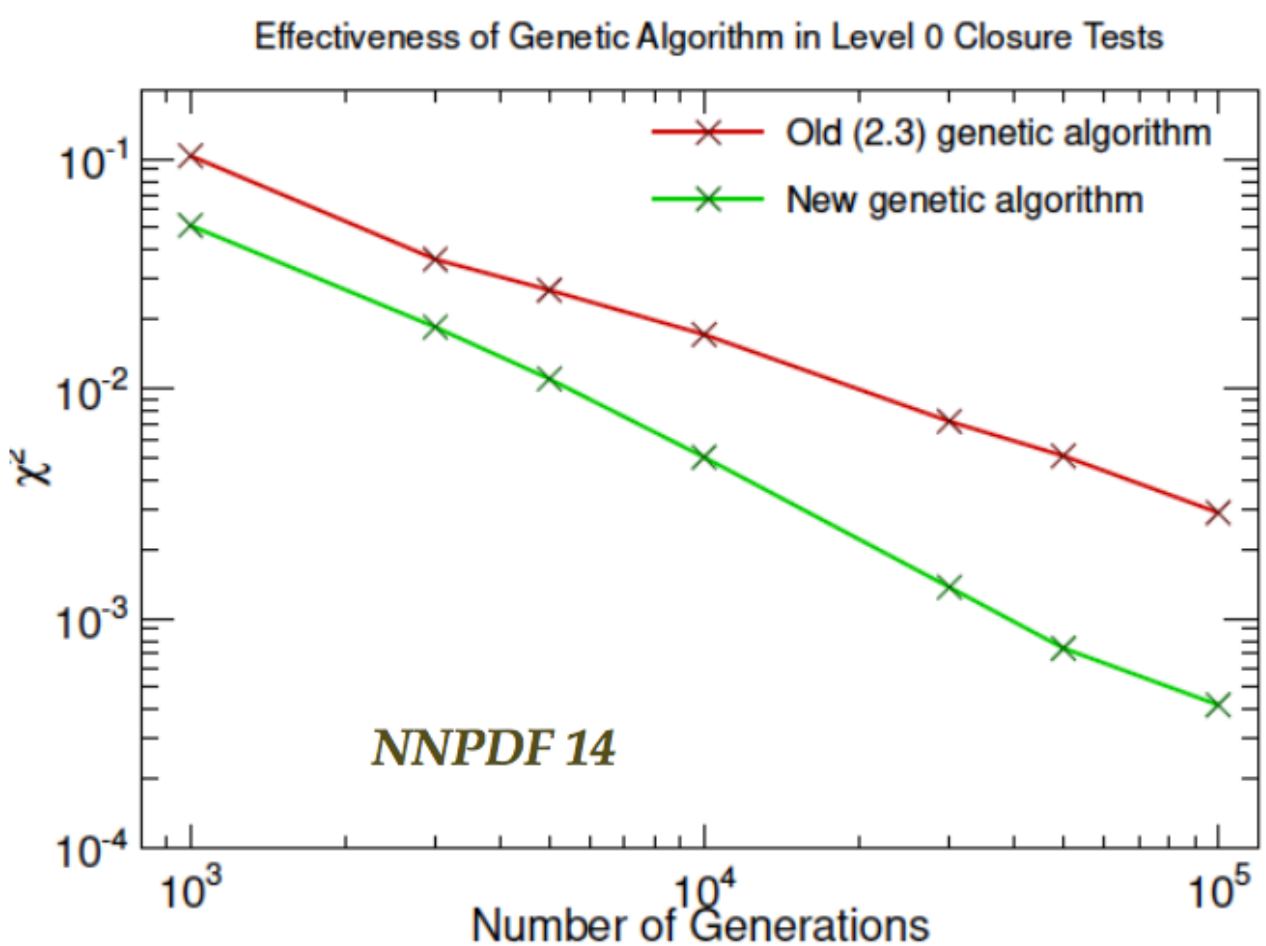
Closure test: the ultimate check of PDF fitting

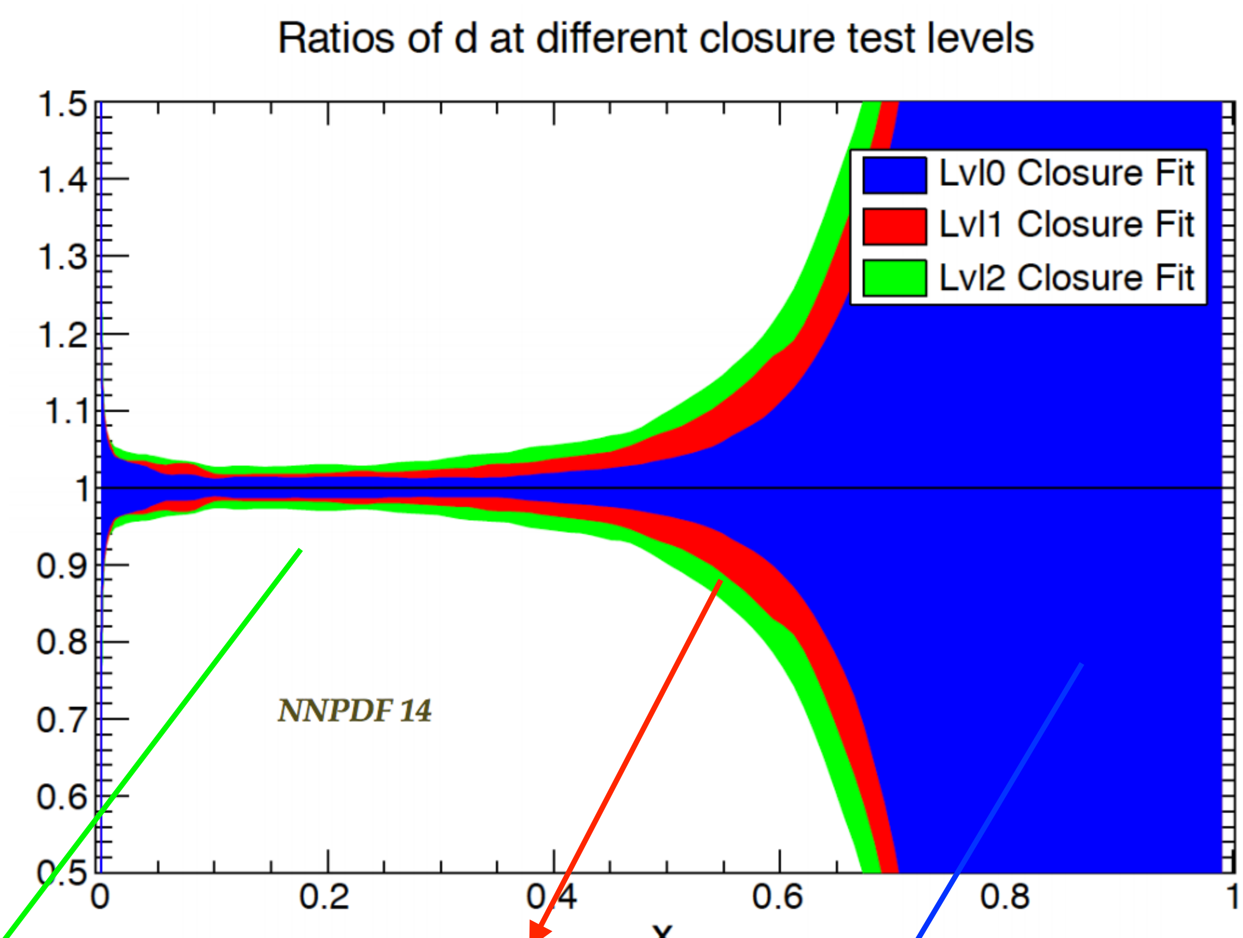
- Assume PDFs known: generate fake experimental data with them and th predictions
- Can decide data uncertainty (zero uncertainty - level 1-2, or as in real data - level 3)
- Fit PDFs to fake data
- Check whether fit reproduces the underlying "truth"
  - ➡ Check whether true values are gaussianly distributed about the fit
  - ➡ Check whether uncertainties are faithful
  - ➡ Trace different sources of uncertainty



Statistical validation

- **Level-0:** if pseudo-data are identical to the input theory, then agreement with theory should be arbitrarily good, i.e. $\chi^2 \to 0$ but PDF uncertainty $\to 0$ only in the region where there are enough data
- **Level-1:** add uncertainty to pseudo data equal to actually experimental uncertainties: replicas fit same data over and over again, then $\chi^2 \to 1$ and test equivalent minima (parametrisation Δ)
- **Level-2:** generate Monte Carlo replicas of pseudo-data with fluctuations, then $\chi^2 \to 2$ (data Δ)





parametrisation uncertainty

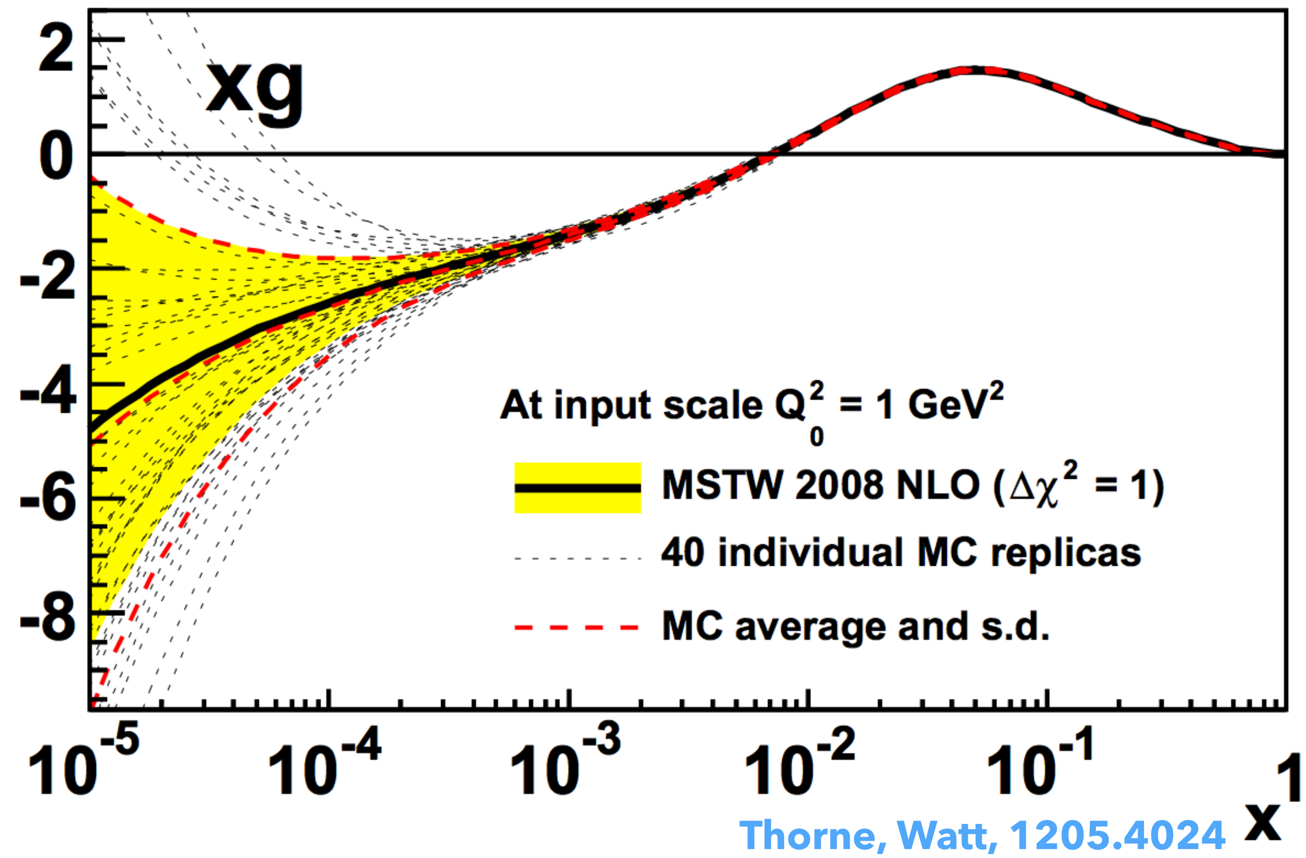
3 Δs comparable in data region

data uncertainty

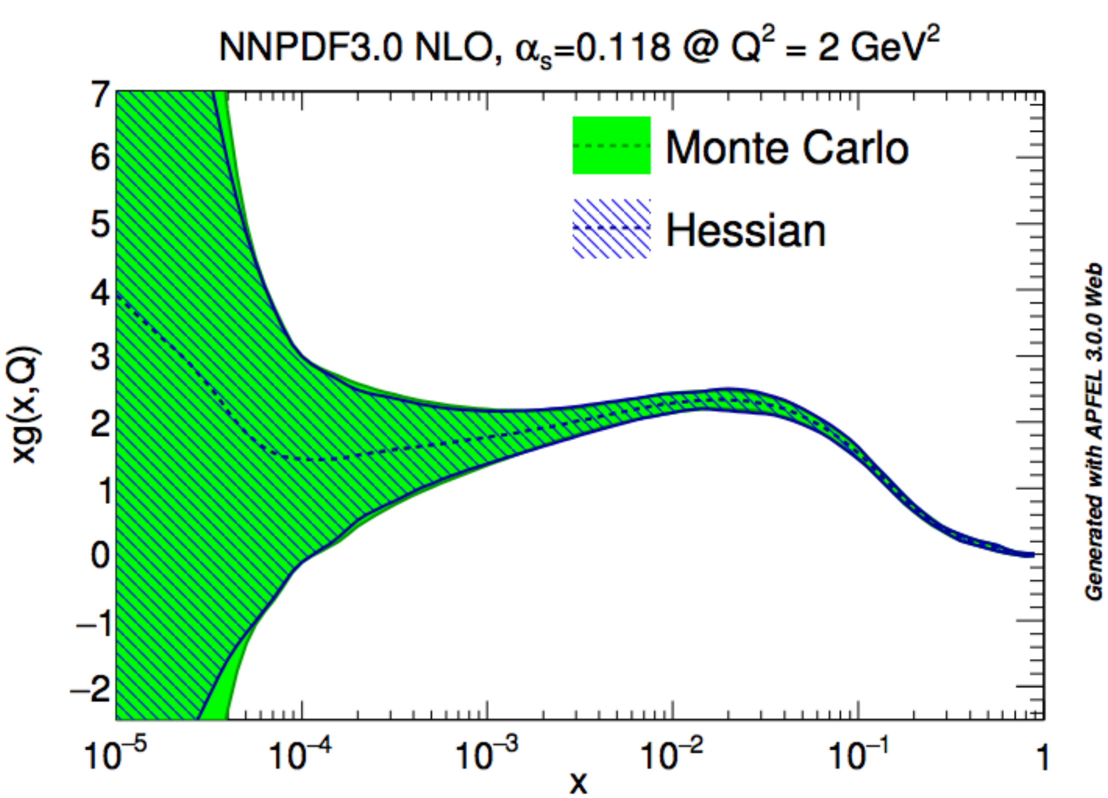
extrapolation uncertainty

Hessian ⟺ Monte Carlo

- To convert Hessian into Monte Carlo, generate multi-gaussian replicas in the fitted parameters space
- Accurate when the number of replicas similar to that that reproduces the data



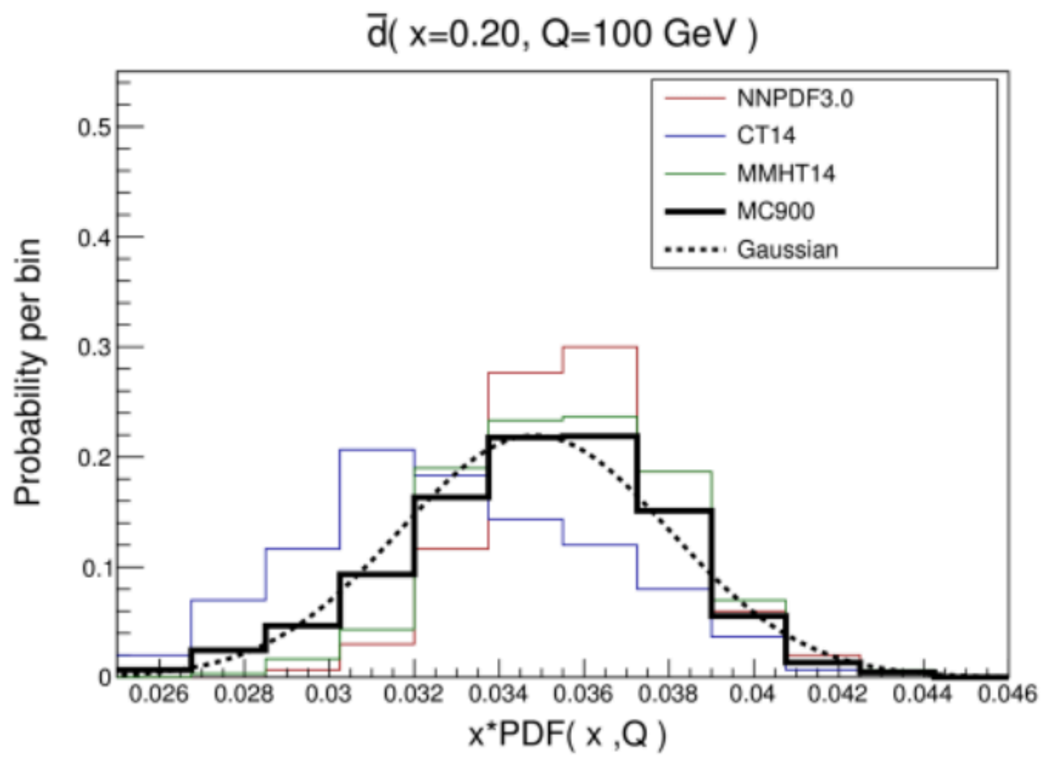
Thorne, Watt, 1205.4024



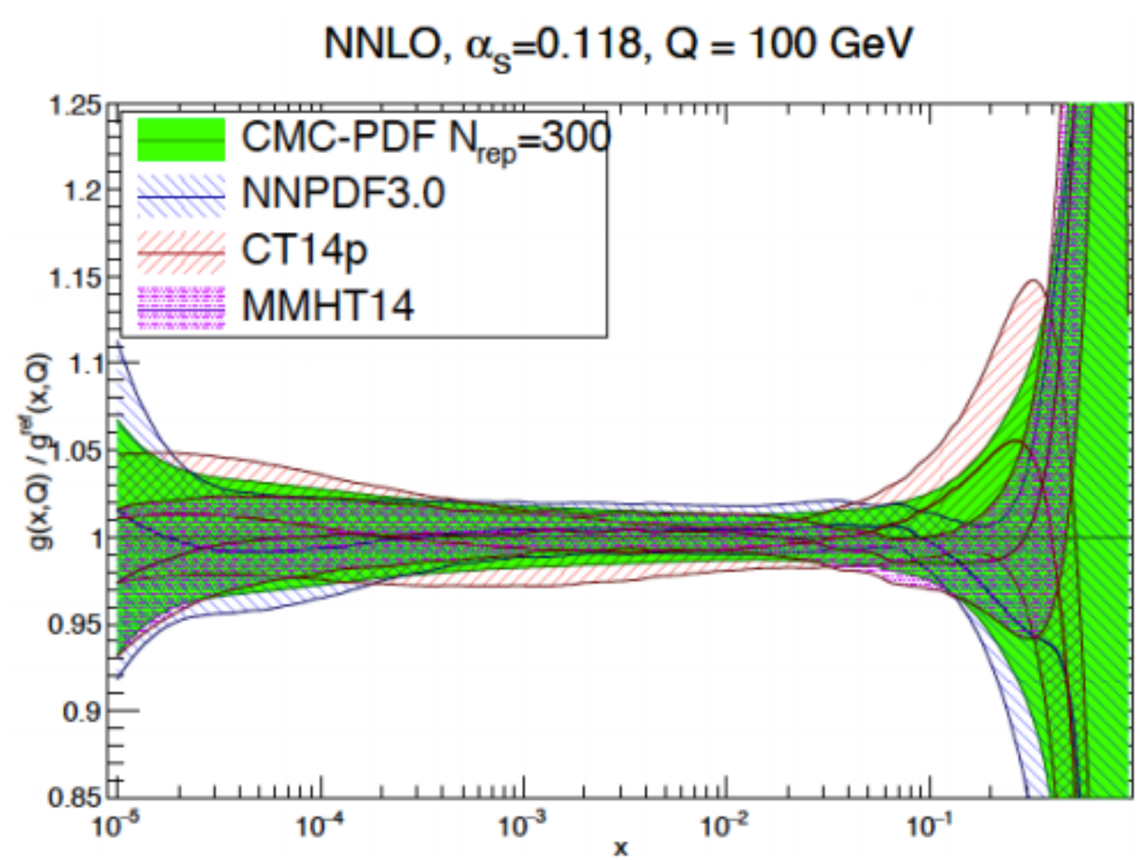
Carrazza et al 1505.06736

- To convert Monte Carlo into Hessian, sample the replicas f(x) at discrete set of points and construct the ensuing covariance matrix
- Eigenvectors of the covariance matrix as a basis in the vector space spanned by the replicas by the singular-value decomposition
- Number of dominant eigenvectors similar to numbers of replicas for accurate representation

Hessian ⟺ Monte Carlo



- Using Monte Carlo conversion of Hessian sets, can combine different PDF sets, combining MC replicas into a single set
- Useful for conservative estimate
- Combined set approximatively Gaussian



PDF4LHC15 recipe

- Monte Carlo combination of most recent global PDF sets **[Forte, Watt]**
- Each replica receives the same weight: uncertainty smaller than in the envelope, as in the latter outliers are given a larger weight
- New compression studies: N=40 replicas are virtually identical to the original 300 replicas from the point of view of correlation, standard deviation, observables **[Carrazza et al.]**

Statistics and methodology summary

➡ PDF determination: Hessian Method
- Simple linear error propagation
- Tolerance required for realistic uncertainties
- Parametrisation bias possible

➡ PDF determination: Monte Carlo method
- Two-step procedure: data MC -> PDF MC
- Very general parametrisation allowed
- Need optimal fit determination method (cross-validation)

➡ PDF representation: Hessian vs Monte Carlo
- Conversion possible either way
- Compression method available either way
- MC very flexible, Hessian very efficient

➡ PDF validation: the closure test
- Performed in the MC approach (so far)
- Interpolation and functional uncertainties significant

Wrap-up

- ➡ The determination of the proton structure and of its uncertainty requires sophisticated statistical techniques. Data are the crucial input and it is highly non trivial to combine thousands of datapoints from many sources that might display incompatibilities

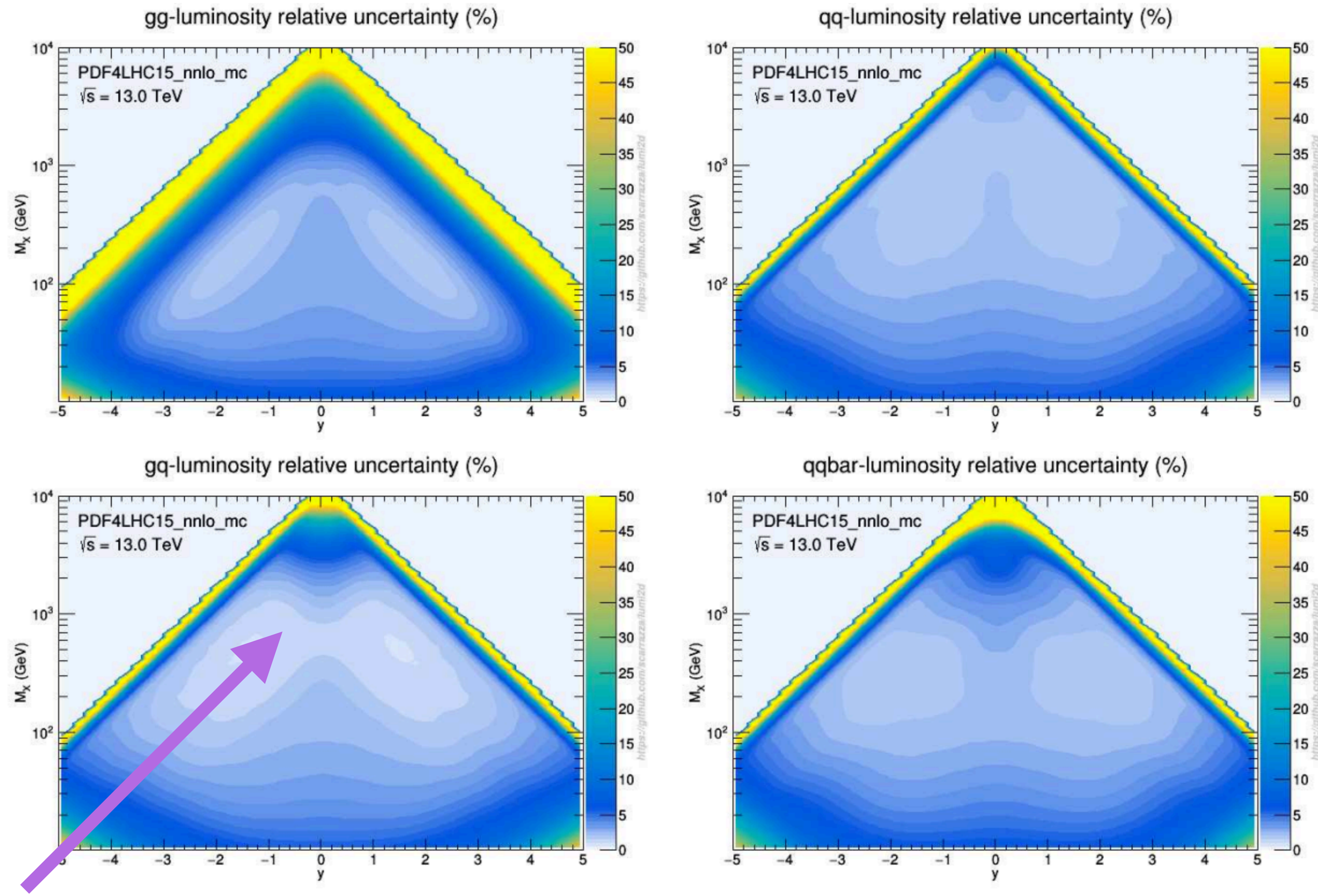
- ➡ Today's lecture

  - ✓ Ingredients of PDF fits
  - ✓ Experimental input
  - ✓ Fitting methodology
  - ✓ Parametrization
  - ✓ Error propagation
  - ✓ Statistical validation

Extra material

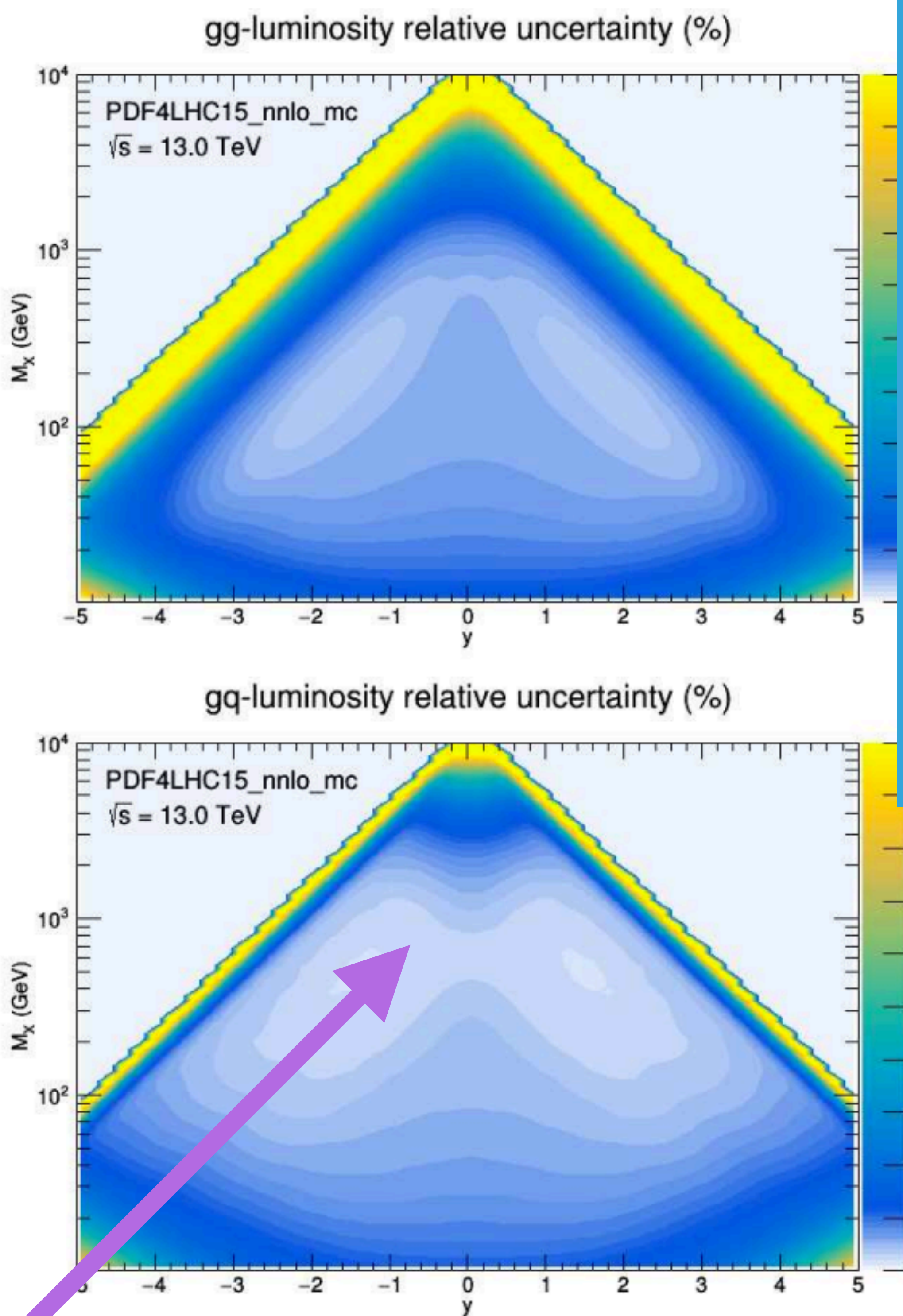
Theoretical aspects

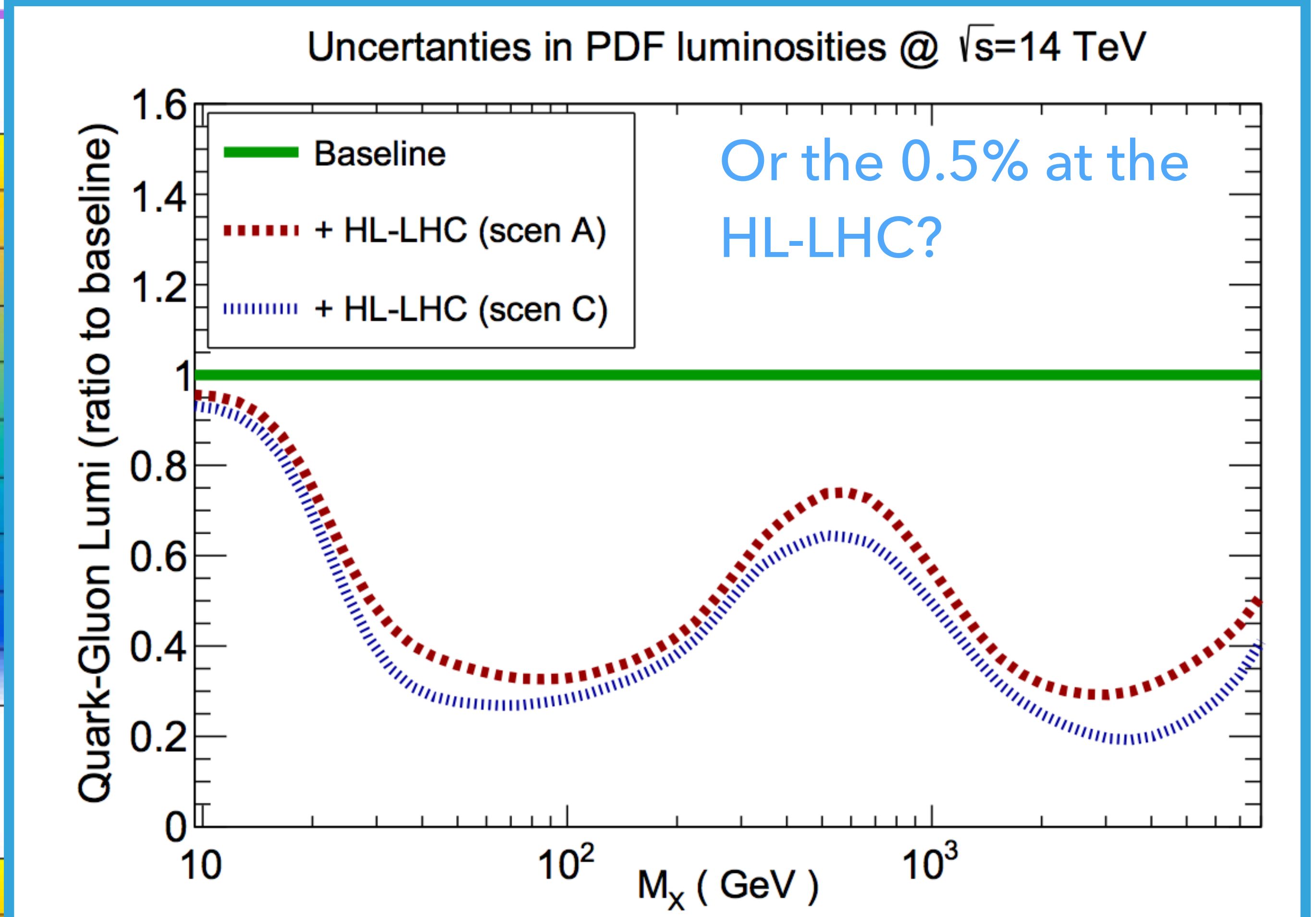
The precision frontier



Can we trust 1% accuracy?

The precision frontier



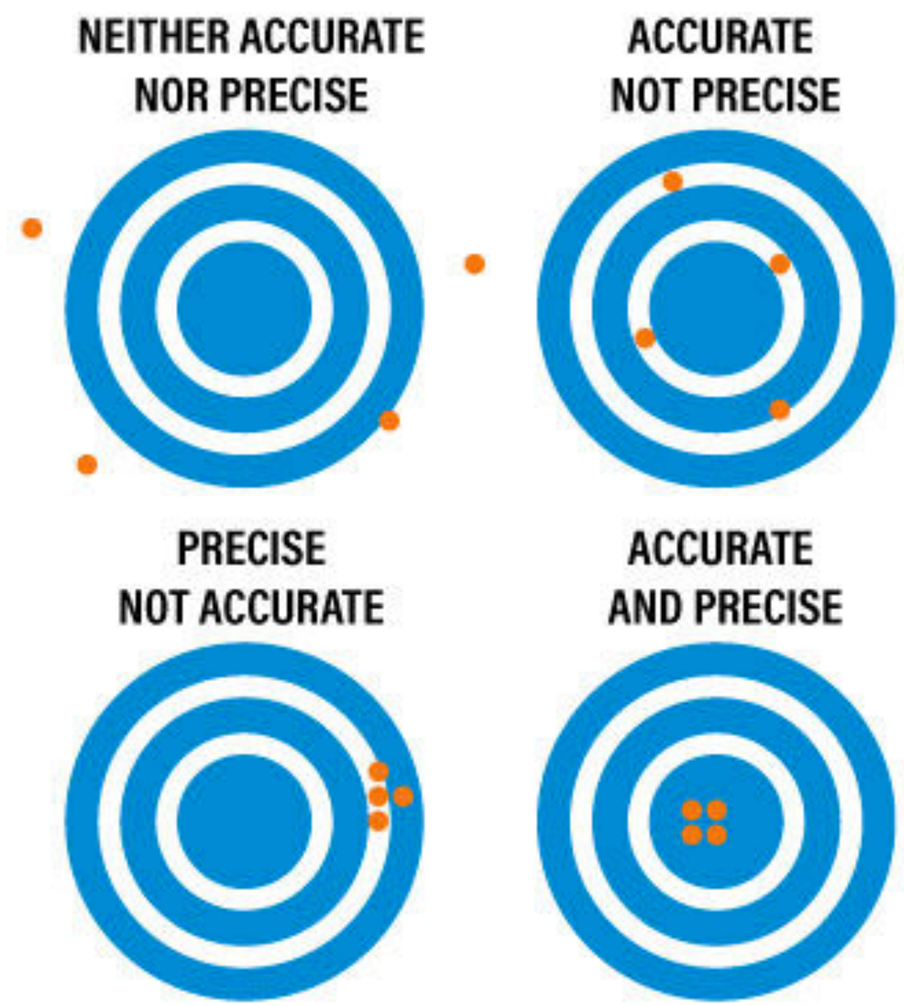


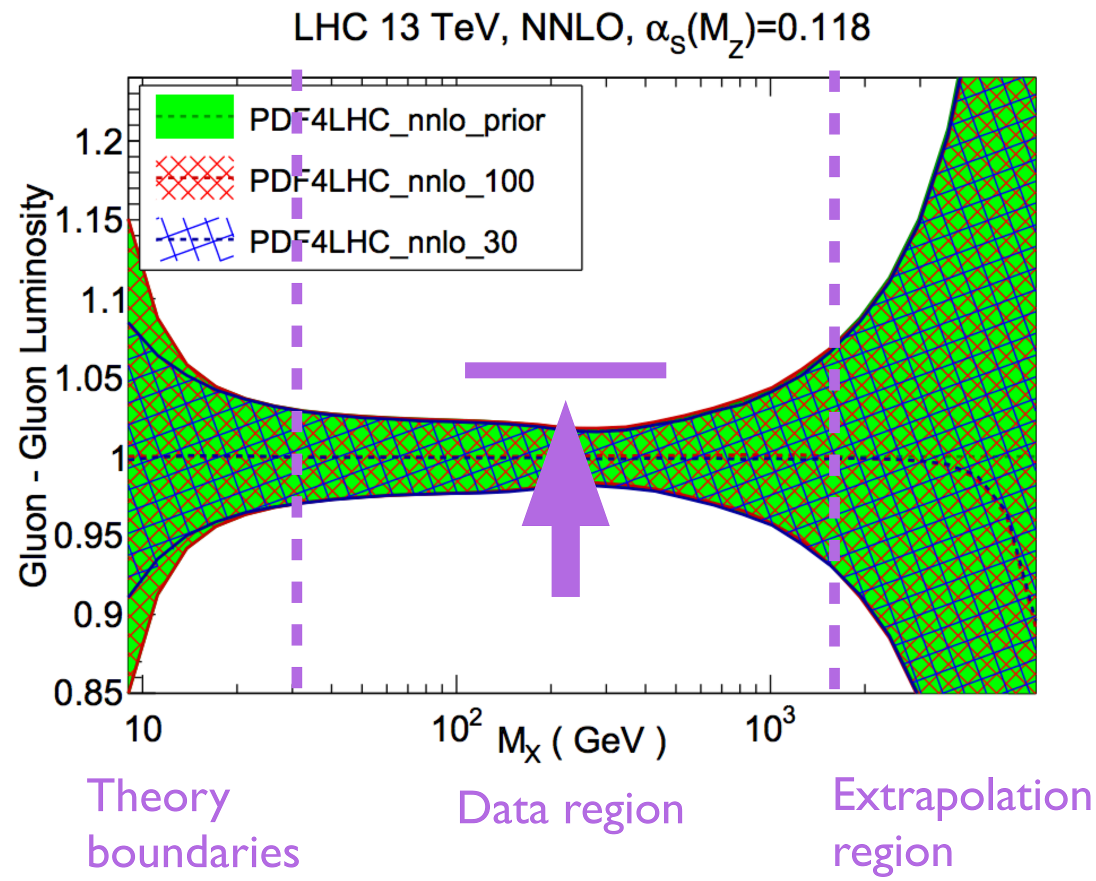
Or the 0.5% at the HL-LHC?

Can we trust 1% accuracy?

Theory uncertainties

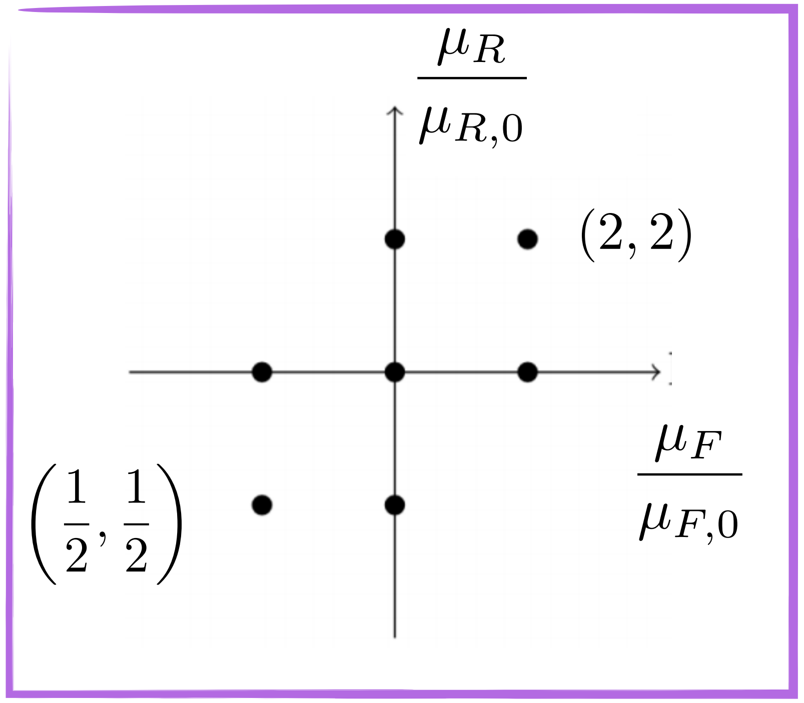
In updated PDF analysis, shift between old and new set may be larger than PDF uncertainties



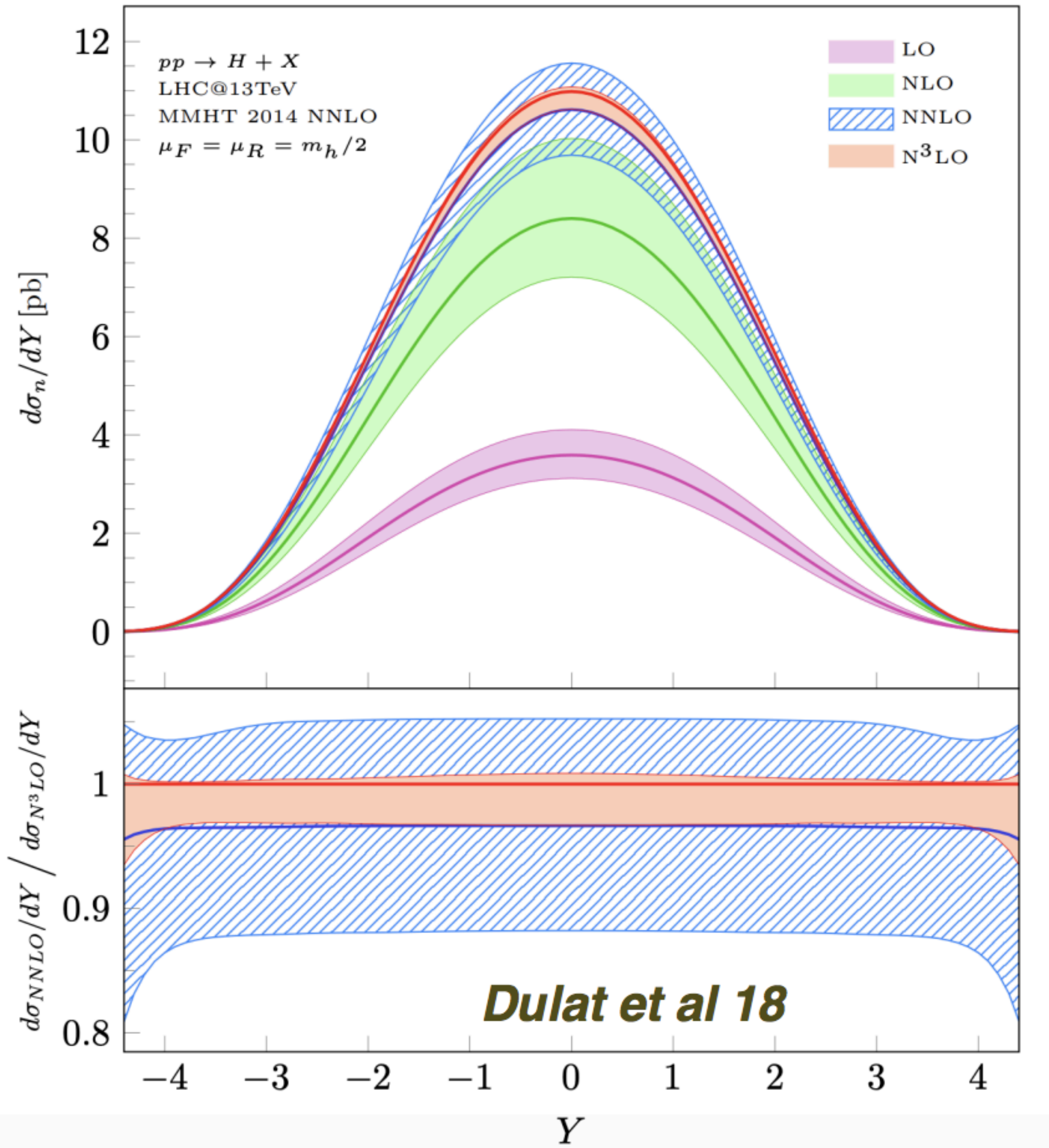


- Inconsistent data → Tolerance/ Statistical estimators
- Updated parametrization
- Differences in fitting methodology/minimisation? → Closure Test
- **<u>Changes in theory?</u>**

MHOU in theoretical predictions

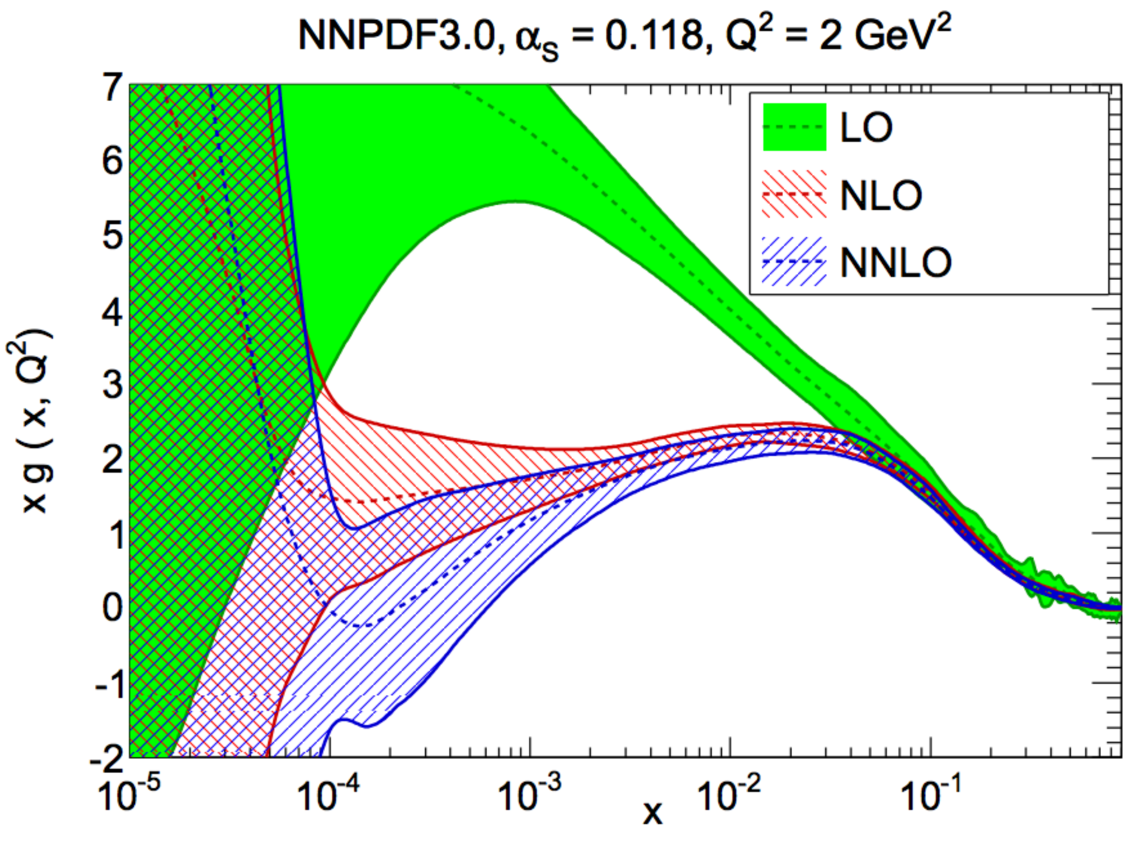


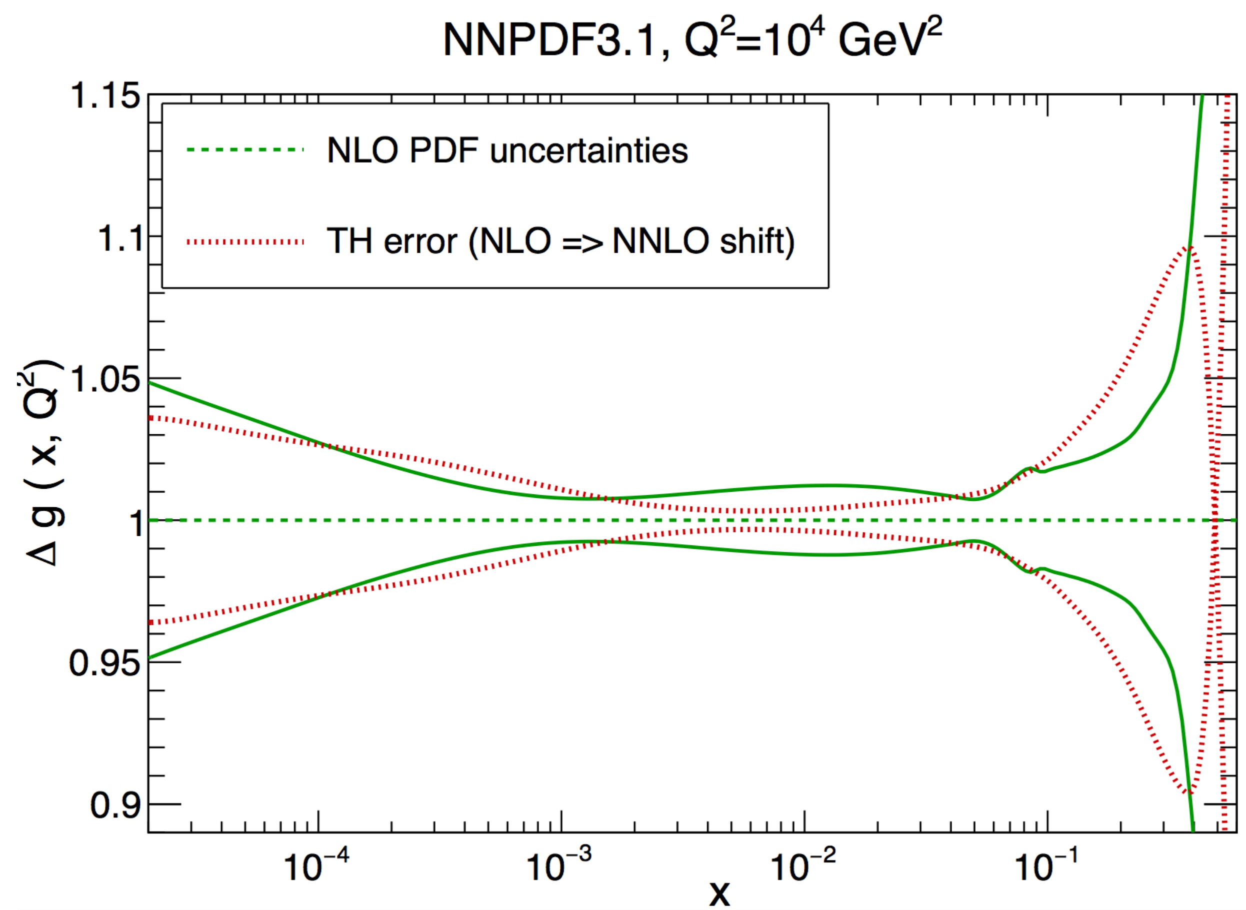
Increasing order in perturbation theory reduced "scale" uncertainty (or MHOU) in theoretical predictions



MHOU in PDF fits

- PDF fits performed at given perturbative order
- PDF uncertainties only reflect lack of information from data
- Theoretical uncertainties (dominated by MHOU) ignored so far
- At NLO PDF uncertainties and MHOU comparable
- Near future: NNLO PDF uncertainties will go down to level of MHOU
- Inclusion of theory uncertainties is the next frontier

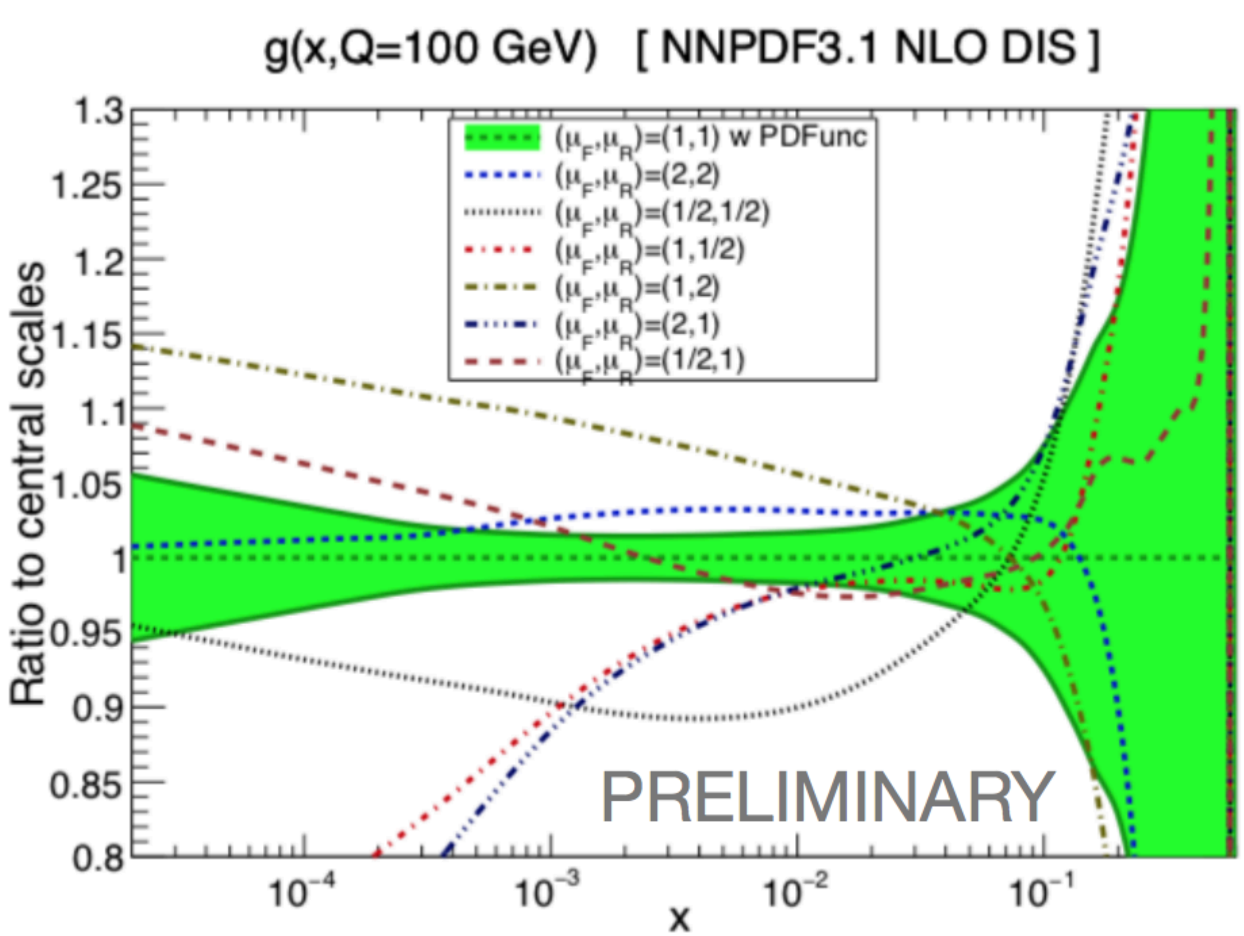


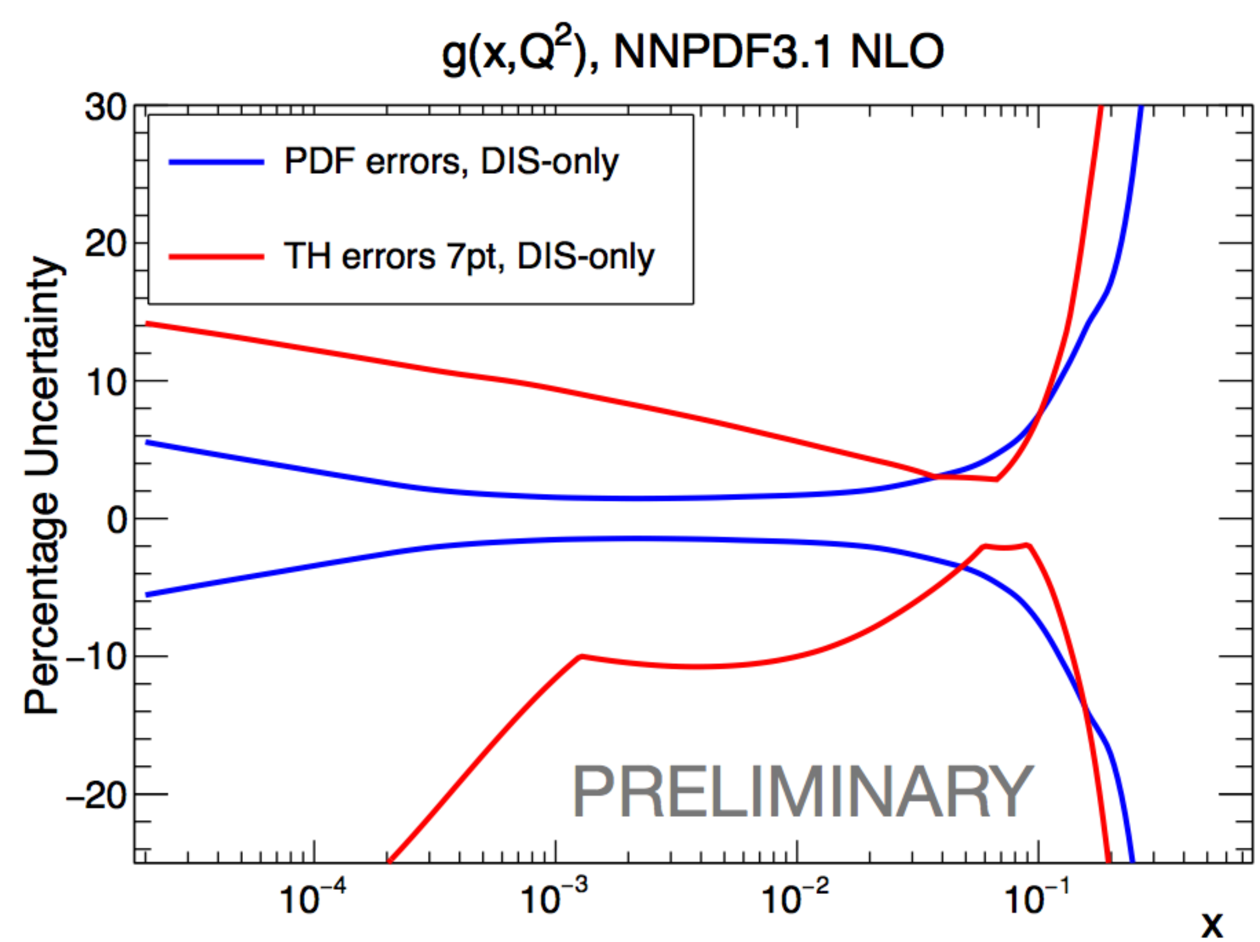


Ball et al, EPJC 77 (2017)

MHOU in PDF fits

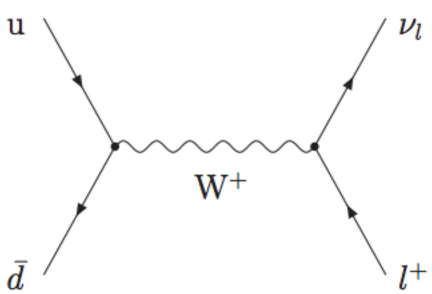
- How to estimate MHOU in PDF fits?
- Compare fits with varied scales
- Useful to have indication on the size of MHOU in PDFs
- A posteriori combination?
- How to include them in the fitting methodology along with other sources of theoretical uncertainty? - see tomorrow's lecture



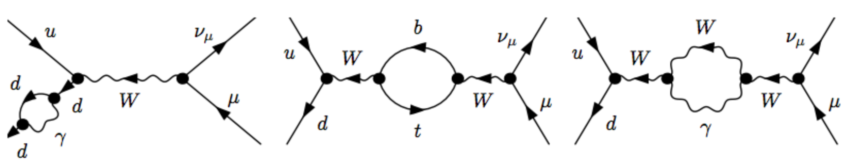


Electroweak corrections

- Because $\alpha(M_Z) \sim \alpha_S(M_Z)/10 \Longrightarrow$ NLO EW corrections ~ NNLO QCD corrections



$$\mathcal{O}(\alpha^2)\,[1]$$



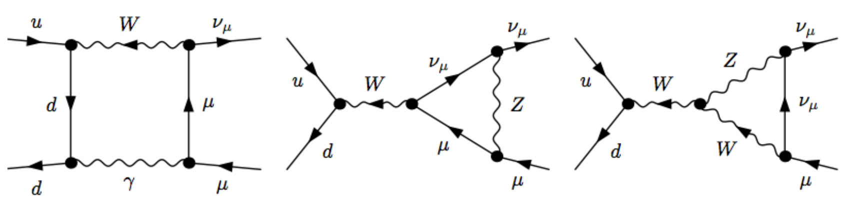
(a) (b) (c)

Virtual corrections

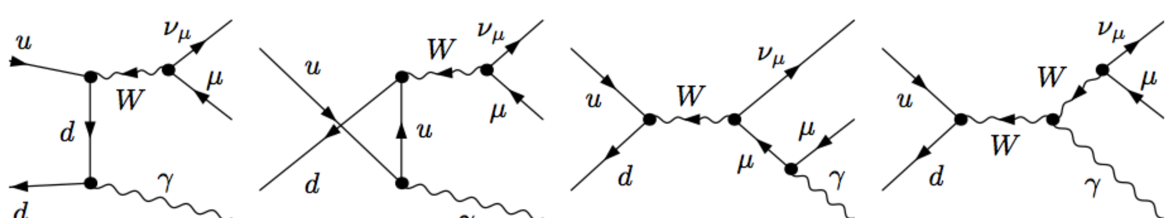
$$\mathcal{O}(\alpha^2)\,[1+\mathcal{O}(\alpha)]$$

$$+\frac{3\alpha}{\pi s_W^2}\ln\left(\frac{s}{M_W^2}\right)$$

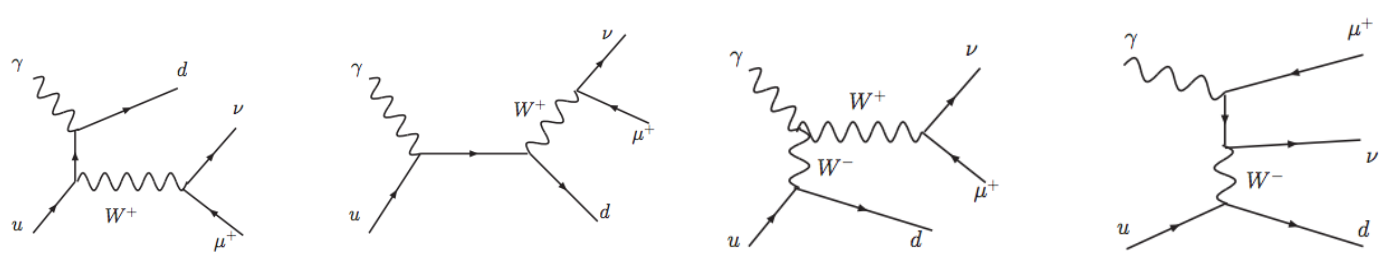
At large s these logs can become large



Real corrections - quark initiated



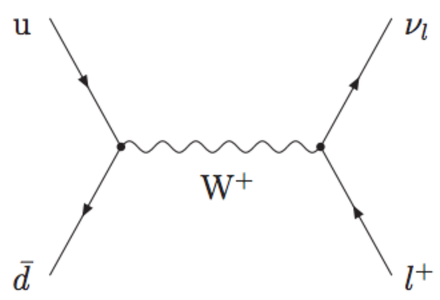
(a) (b) (c) (d)



Real corrections - photon initiated

Electroweak corrections

- Because $\alpha(M_Z) \sim \alpha_S(M_Z)/10 \Longrightarrow$ NLO EW corrections ~ NNLO QCD corrections



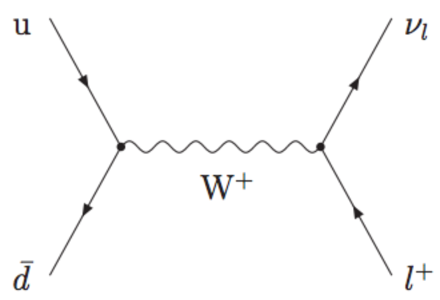
- NLO EW corrections become large in the large pT region of lepton but partially compensated by photon-initiated real corrections

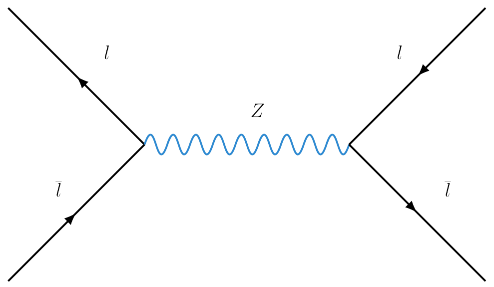
$p_{T,l}/\mathrm{GeV}$	25–∞	50–∞	100–∞	200–∞	500–∞	1000–∞
$\delta_{e^+\nu_e}/\%$	$-5.19(1)$	$-8.92(3)$	$-11.47(2)$	$-16.01(2)$	$-26.35(1)$	$-37.92(1)$
$\delta_{\mu^+\nu_\mu}/\%$	$-2.75(1)$	$-4.78(3)$	$-8.19(2)$	$-12.71(2)$	$-22.64(1)$	$-33.54(2)$
$\delta_{rec}/\%$	$-1.73(1)$	$-2.45(3)$	$-5.91(2)$	$-9.99(2)$	$-18.95(1)$	$-28.60(1)$
$\delta_{\gamma q}/\%$	$+0.071(1)$	$+5.24(1)$	$+13.10(1)$	$+16.44(2)$	$+14.30(1)$	$+11.89(1)$

Dittmaier, Krämer

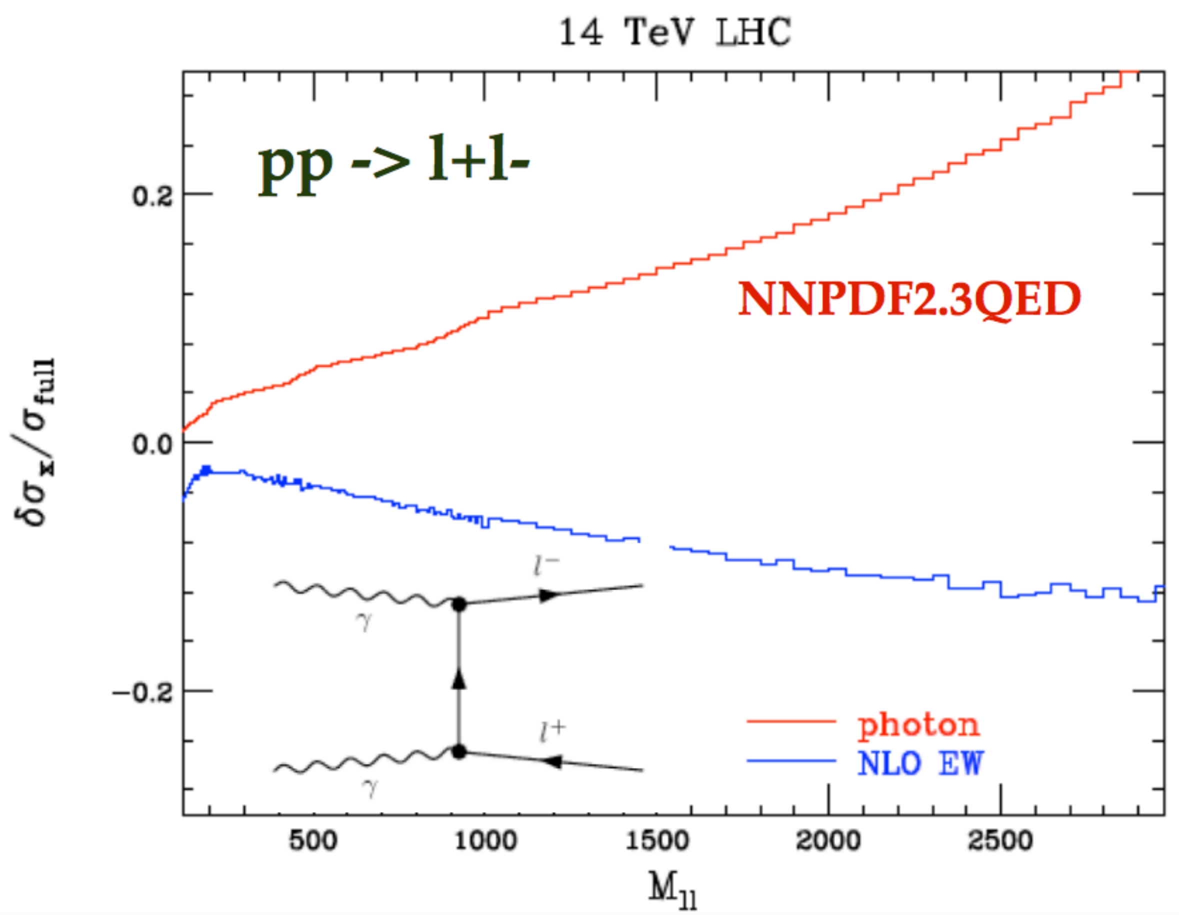
Electroweak corrections

- Because $\alpha(M_Z) \sim \alpha_S(M_Z)/10 \Longrightarrow$ NLO EW corrections ~ NNLO QCD corrections





- NLO EW corrections become large in the large pT region of lepton but partially compensated by photon-initiated real corrections



Boughezal et al Phys.Rev. D89 (2014)3, 034030

Modified DGLAP

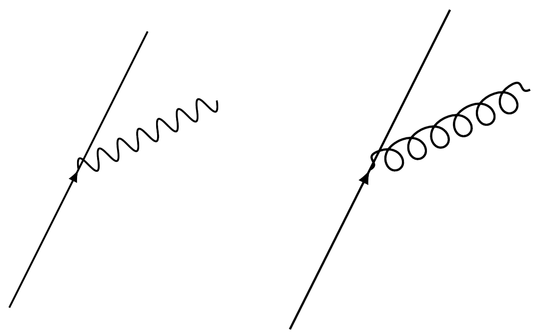
- How are PDFs modified by inclusion of initial photon PDF?

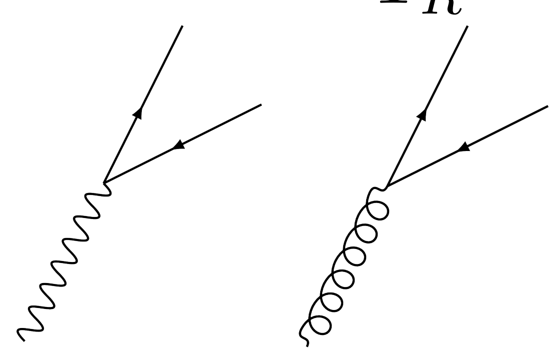
$$
\boxed{
\begin{aligned}
Q^2 \frac{\partial}{\partial Q^2} g(x, Q^2) &= \sum_{q,\bar{q},g} P_{ga}(x, \alpha_s(Q^2)) \otimes f_a(x, Q^2) + P_{g\gamma}(x, \alpha_s(Q^2)) \otimes \gamma(x, Q^2), \\
Q^2 \frac{\partial}{\partial Q^2} q(x, Q^2) &= \sum_{q,\bar{q},g} P_{qa}(x, \alpha_s(Q^2)) \otimes f_a(x, Q^2) + P_{q\gamma}(x, \alpha_s(Q^2)) \otimes \gamma(x, Q^2), \\
Q^2 \frac{\partial}{\partial Q^2} \gamma(x, Q^2) &= P_{\gamma\gamma} \otimes \gamma(x, Q^2) + \sum_{q,\bar{q},g} P_{\gamma a}(x, \alpha_s(Q^2)) \otimes f_a(x, Q^2).
\end{aligned}
}
$$

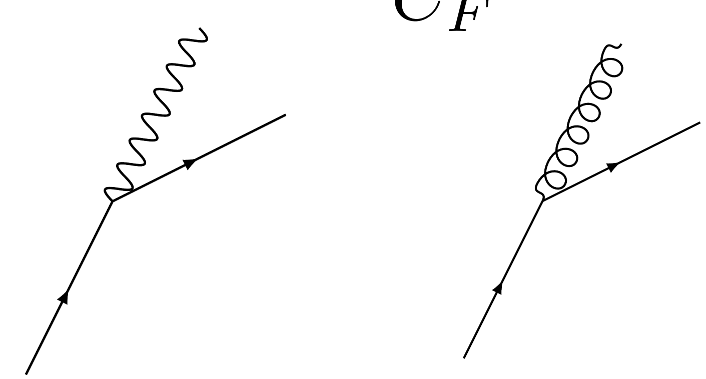
- DGLAP splitting functions expanded in powers of α_s and $\boldsymbol{\alpha}$

$$P_{ij} = \sum_{m,n} \left(\frac{\alpha_S}{2\pi}\right)^m \left(\frac{\alpha}{2\pi}\right)^n P_{ij}^{(m,n)}$$

$$P_{qq}^{(0,1)} = \frac{e_q^2}{C_F} P_{qq}^{(1,0)} \qquad P_{q\gamma}^{(0,1)} = \frac{e_q^2}{T_R} P_{qg}^{(1,0)} \qquad P_{\gamma q}^{(0,1)} = \frac{e_q^2}{C_F} P_{gq}^{(1,0)}$$

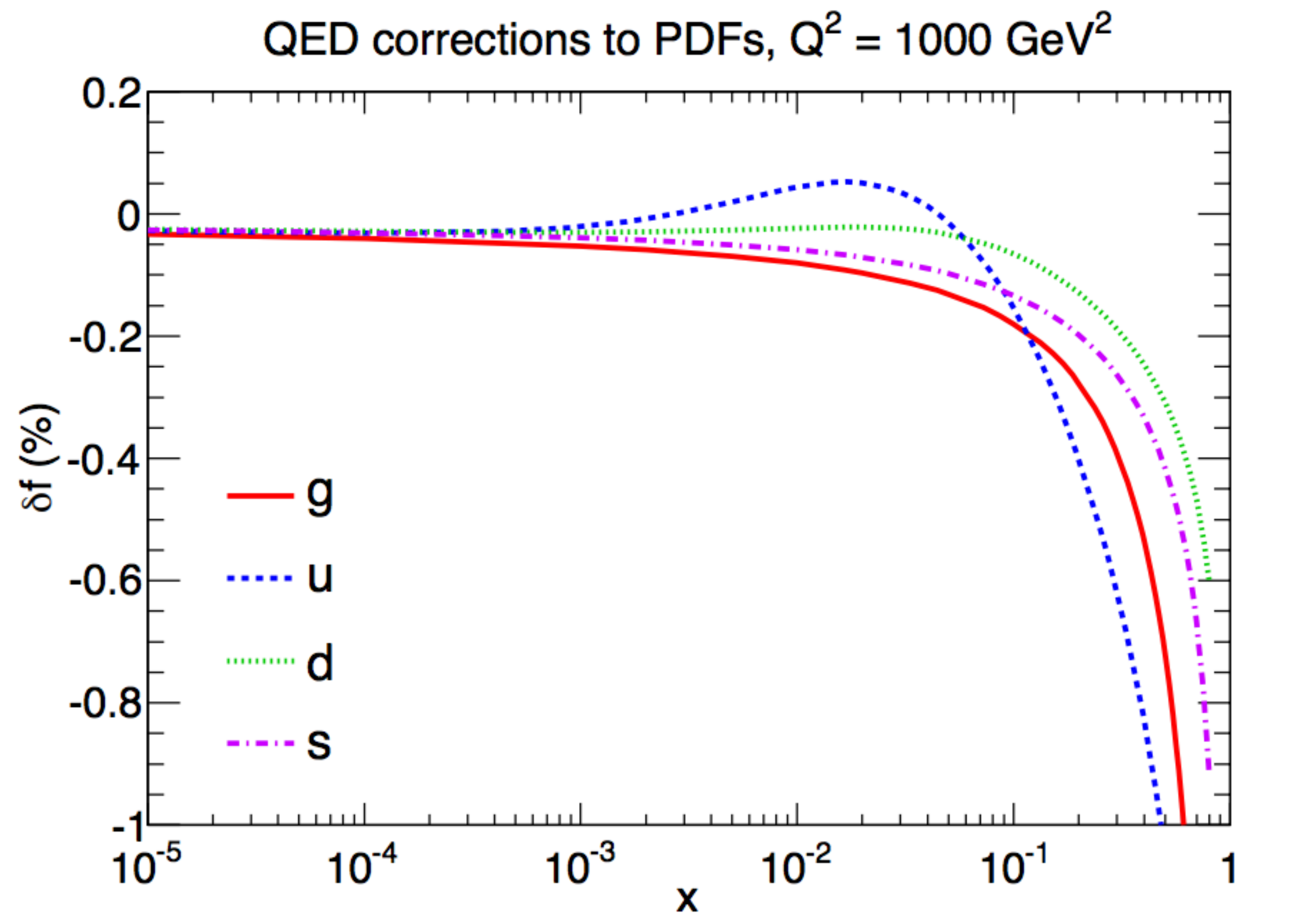






Modified DGLAP

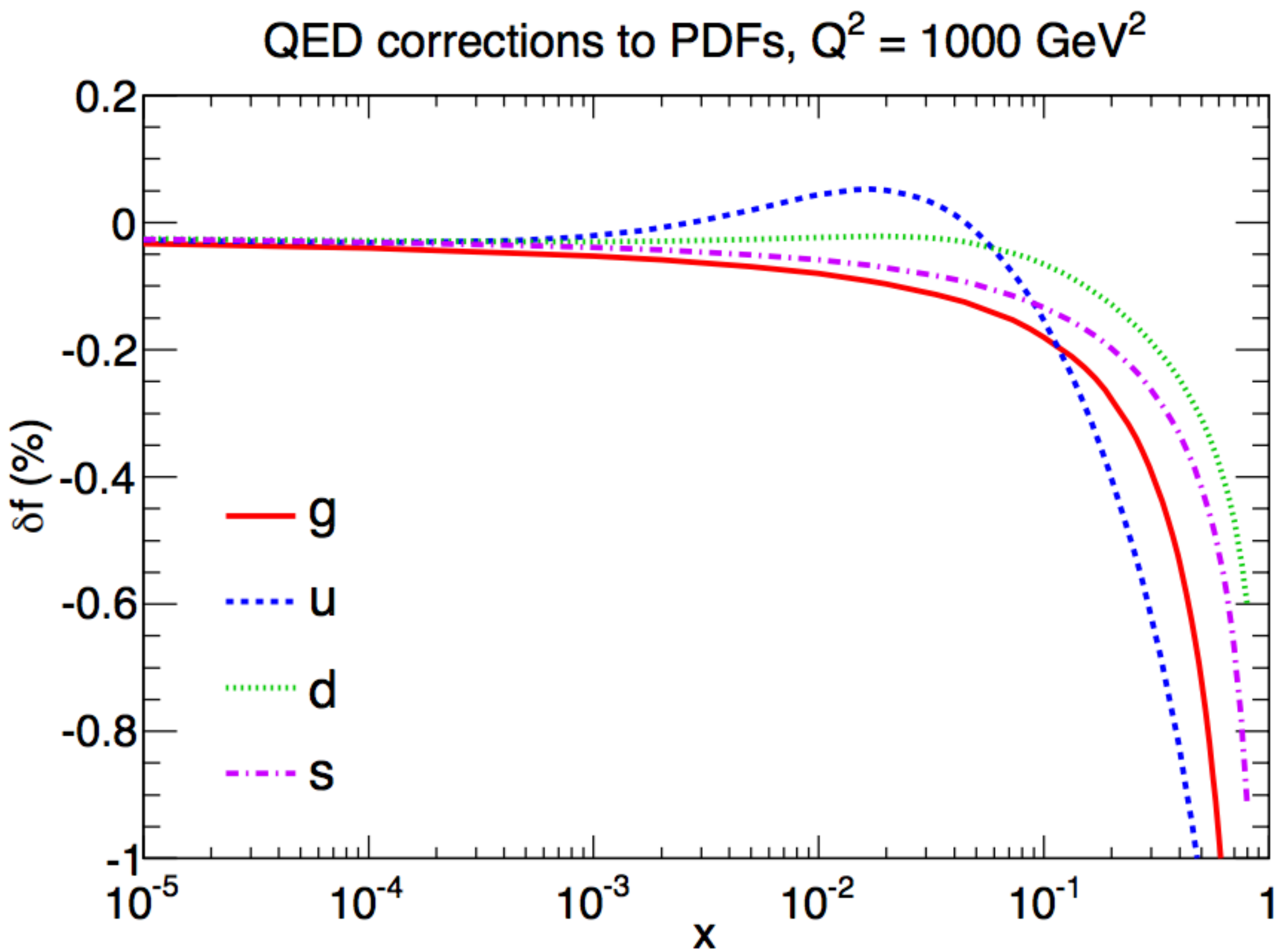
- Quark and gluon PDFs change up to 1% at large x



Ball et al, Nucl.Phys. B877 (2013) 290-320

Modified DGLAP

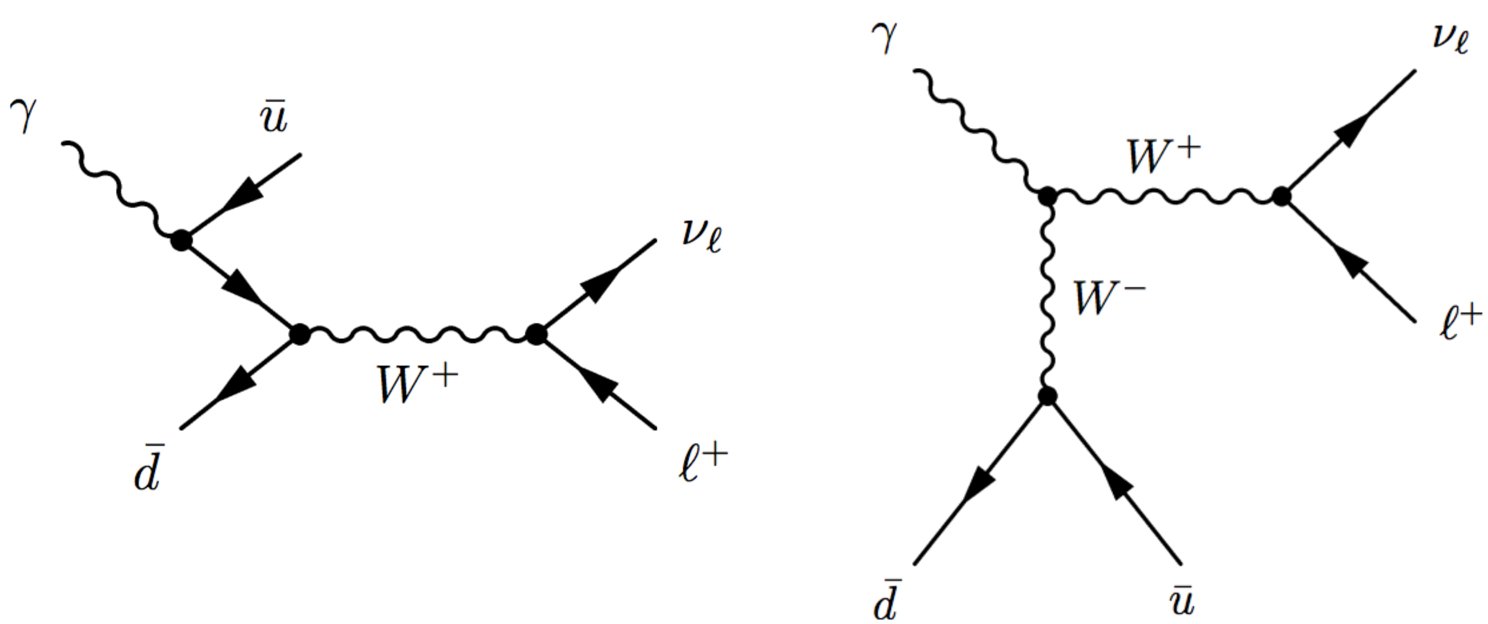
- Quark and gluon PDFs change up to 1% at large x
- How do we determine the photon PDF?
- Two ways in the next slides: from data or from theory
- In the best possible world: theory input and data input together



Ball et al, Nucl.Phys. B877 (2013) 290-320

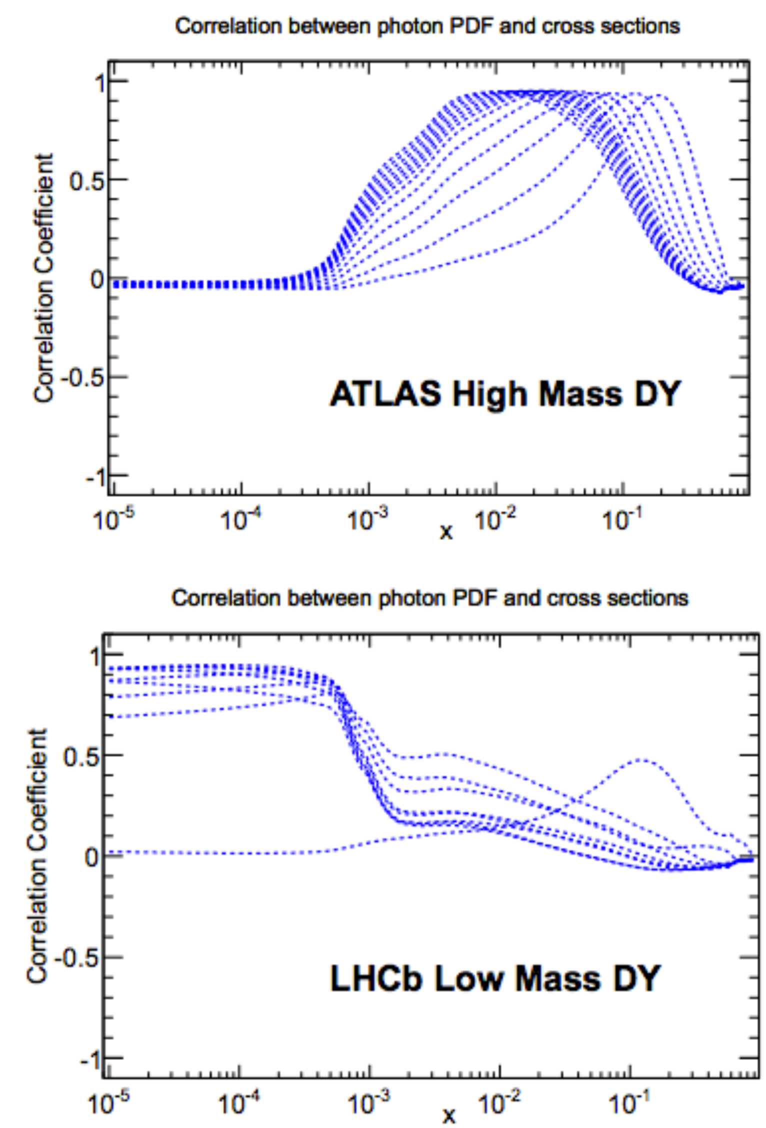
Photon PDFs

- Largest correlations between photon PDFs and pp cross sections are for Drell-Yan processes, but also for top pair production and VV production



Photon-induced Drell-Yan

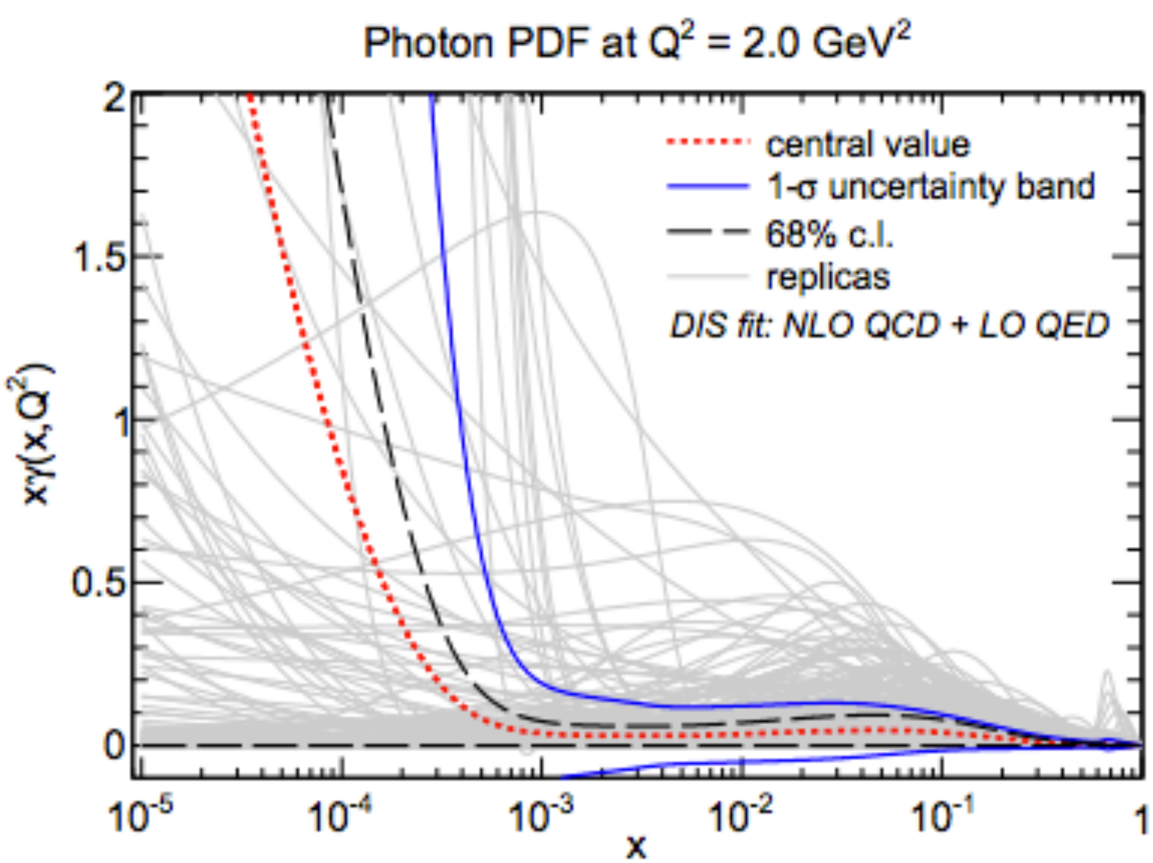
$Q = 100\,\mathrm{GeV}$

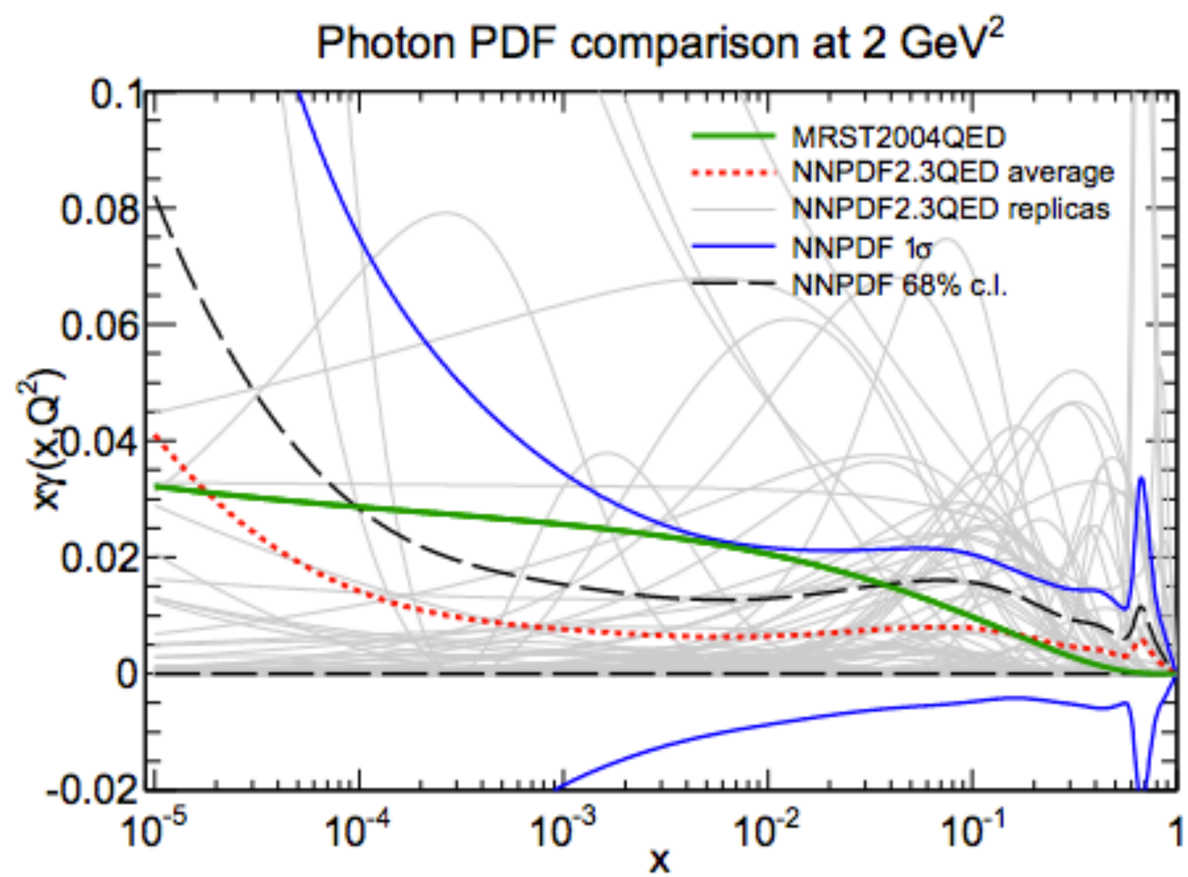


Ball et al, Nucl.Phys. B877 (2013) 290-320

Photon PDFs

Data-driven knowledge



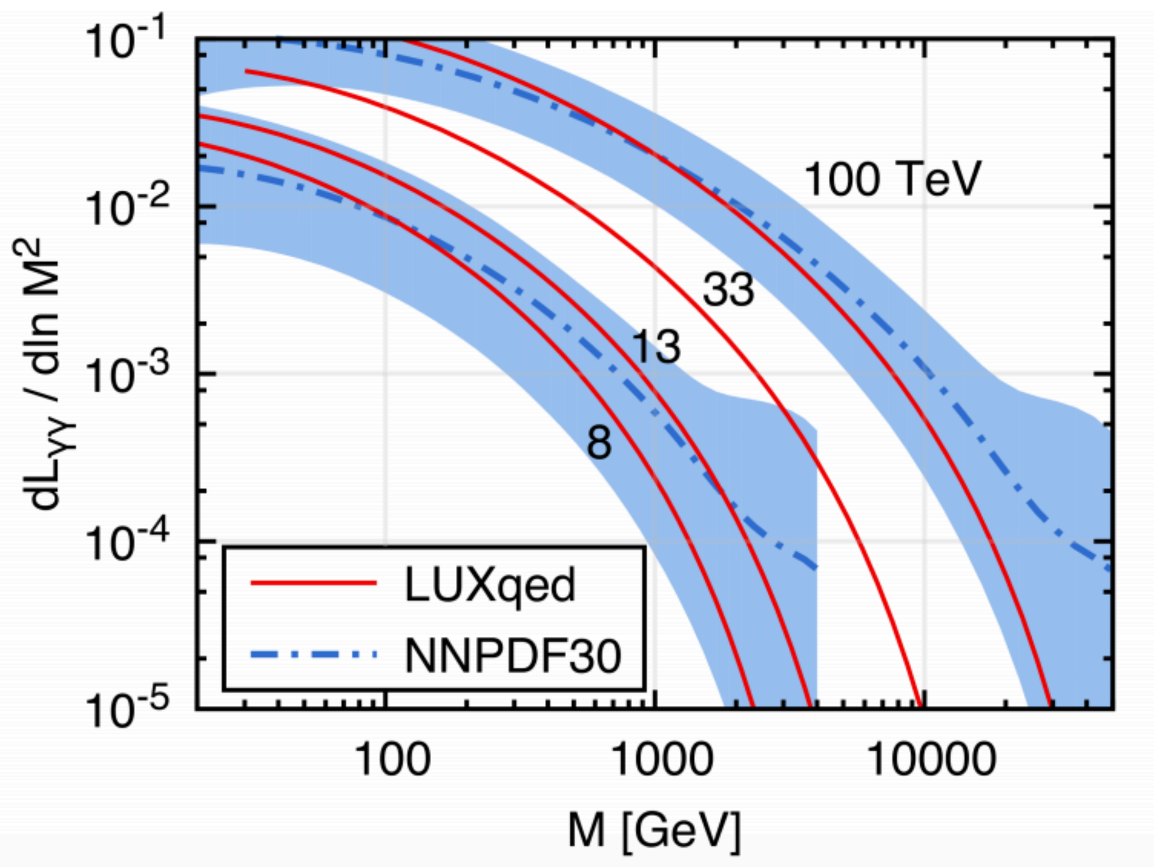


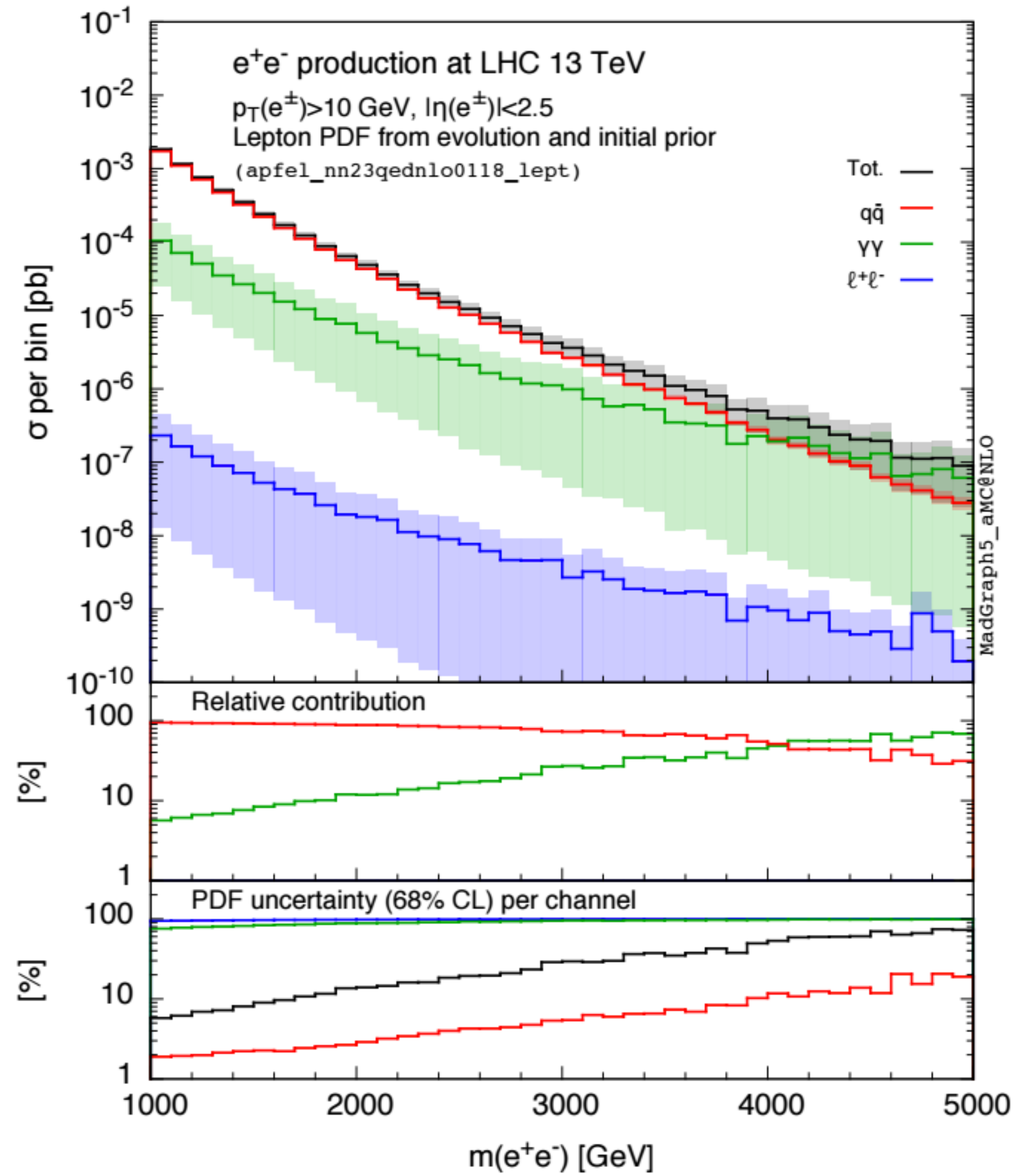
Ball et al, Nucl.Phys. B877 (2013) 290-320

Photon PDFs

- Data-driven approach associated with a large uncertainty on photon PDF

- Theory breakthrough: LUX PDF [Manohar, Nason, Salam, Zanderighi, 1607.04266]





Bertone et al, 1508.07002

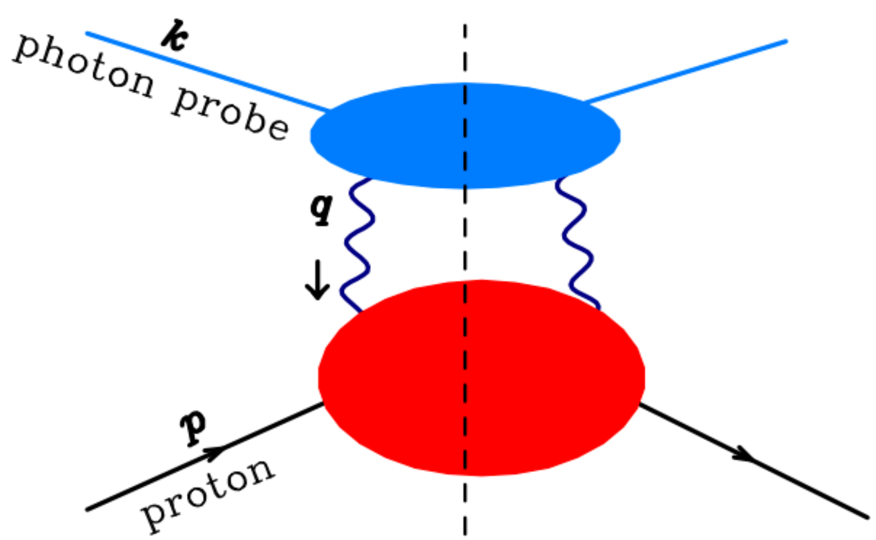
Photon PDFs

- QED is perturbative down to low scales $\Longrightarrow$ The photon must be computable is the input mark substructure is known
- Manohar et al: write the cross section for a chosen BSM process, e.g. production of heavy supersymmetric lepton L in ep collision (Drees, Zeppenfeld 1989)

$$\sigma = \frac{1}{4p\cdot k}\int \frac{d^4q}{(2\pi)^4 q^4} e_{\rm ph}^2(q^2)\left[4\pi W_{\mu\nu}(p,q)\, L^{\mu\nu}(k,q)\right]\; 2\pi\delta((k-q)^2-M^2)$$

$$l(k) + p(p) \to L(k') + X$$

$$\sigma = c_0 \sum_a \int_x^1 \frac{dz}{z}\,\hat{\sigma}_a(z,\mu^2)\,\frac{M^2}{zs} f_{a/p}\left(\frac{M^2}{zs},\mu^2\right)$$



$$\sigma = \frac{c_0}{2\pi}\int_x^{1-\frac{2xm_p}{M}} \frac{dz}{z}\int_{Q^2_{\rm min}}^{Q^2_{\rm max}} \frac{dQ^2}{Q^2}\alpha^2_{\rm ph}(-Q^2)\Bigg[\Bigg(2-2z+z^2 + \frac{2x^2m_p^2}{Q^2} + \frac{z^2Q^2}{M^2} - \frac{2zQ^2}{M^2} - \frac{2x^2Q^2m_p^2}{M^4}\Bigg)F_2(x/z,Q^2) + \left(-z^2 - \frac{z^2Q^2}{2M^2} + \frac{z^2Q^4}{2M^4}\right)F_L(x/z,Q^2)\Bigg], \quad (3)$$

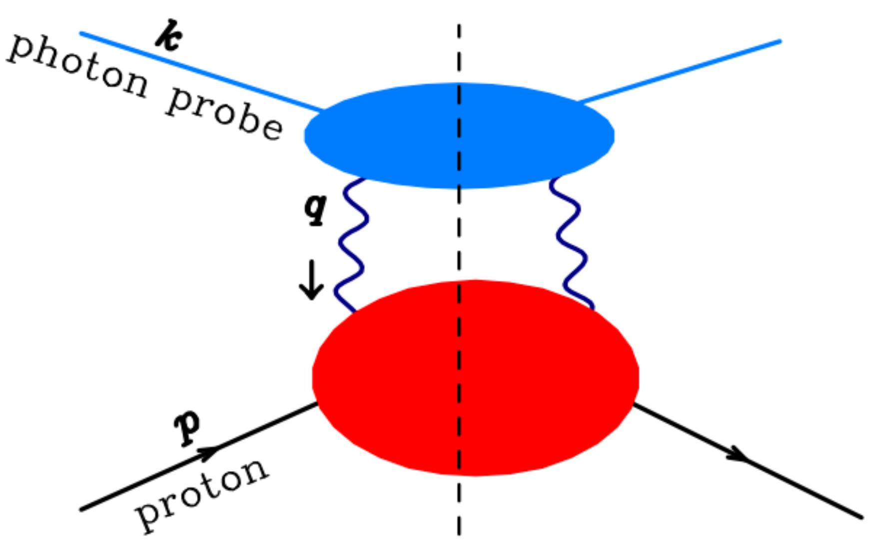
Manohar et al 1607.04266

<u>Theory-driven knowledge</u>

Photon PDFs

- QED is perturbative down to low scales $\Longrightarrow$ The photon must be computable is the input mark substructure is known
- Manohar et al: write the cross section for a chosen BSM process, e.g. production of heavy supersymmetric lepton L in ep collision (Drees, Zeppenfeld 1989)
- Equate the two expressions and find analytically the PDF of the photon

$\Rightarrow$ PDFs expressed in terms of the structure functions integrated over all scales, including elastic form factors (in the x $\rightarrow$1 region)

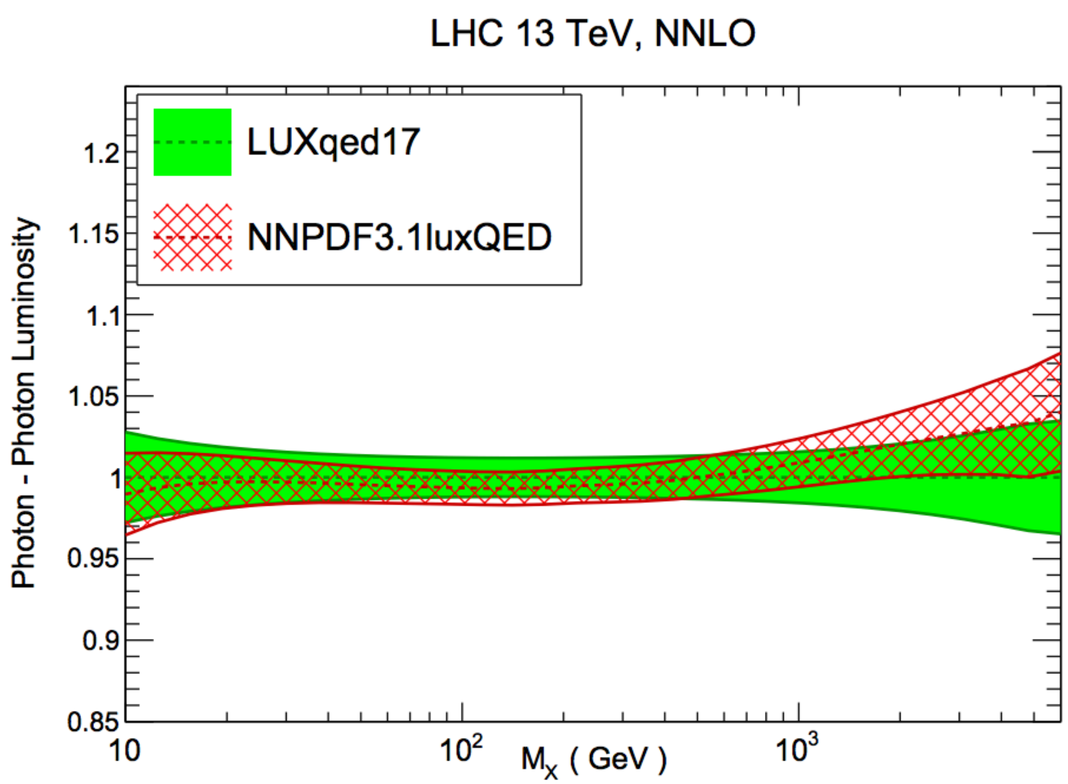


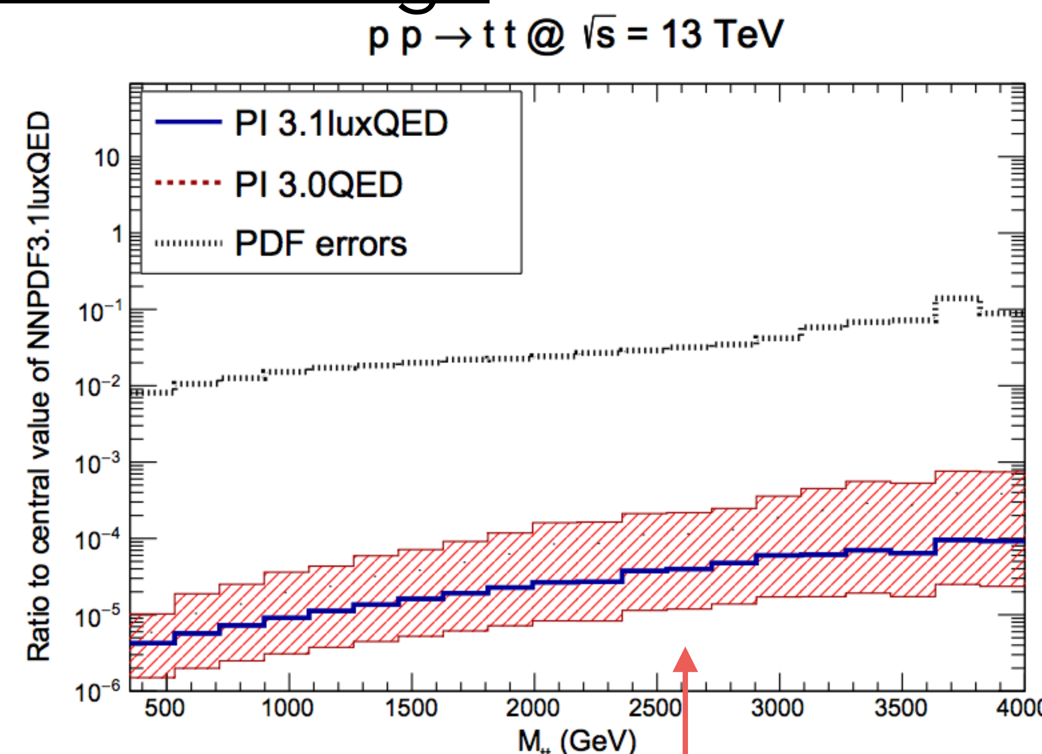
$$x f_{\gamma/p}(x,\mu^2) = \frac{1}{2\pi\alpha(\mu^2)} \int_x^1 \frac{dz}{z} \Bigg\{ \int_{\frac{x^2 m_p^2}{1-z}}^{\frac{\mu^2}{1-z}} \frac{dQ^2}{Q^2} \alpha^2(Q^2) \left[\left(z p_{\gamma q}(z) + \frac{2x^2 m_p^2}{Q^2} \right) F_2(x/z, Q^2) - z^2 F_L\left(\frac{x}{z}, Q^2\right) \right] - \alpha^2(\mu^2) z^2 F_2\left(\frac{x}{z}, \mu^2\right) \Bigg\},$$

<u>Theory-driven knowledge</u>

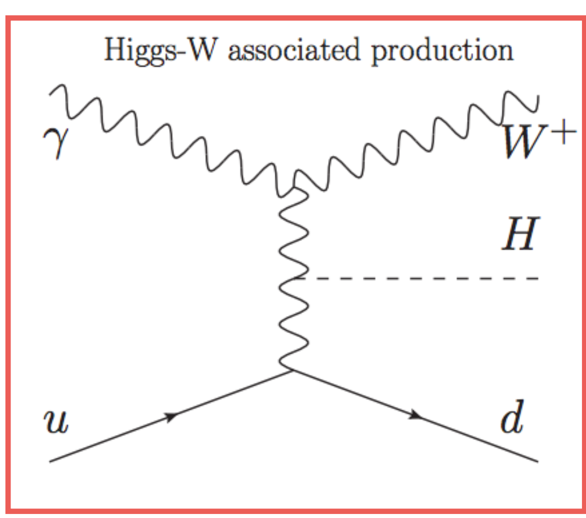
Photon PDFs

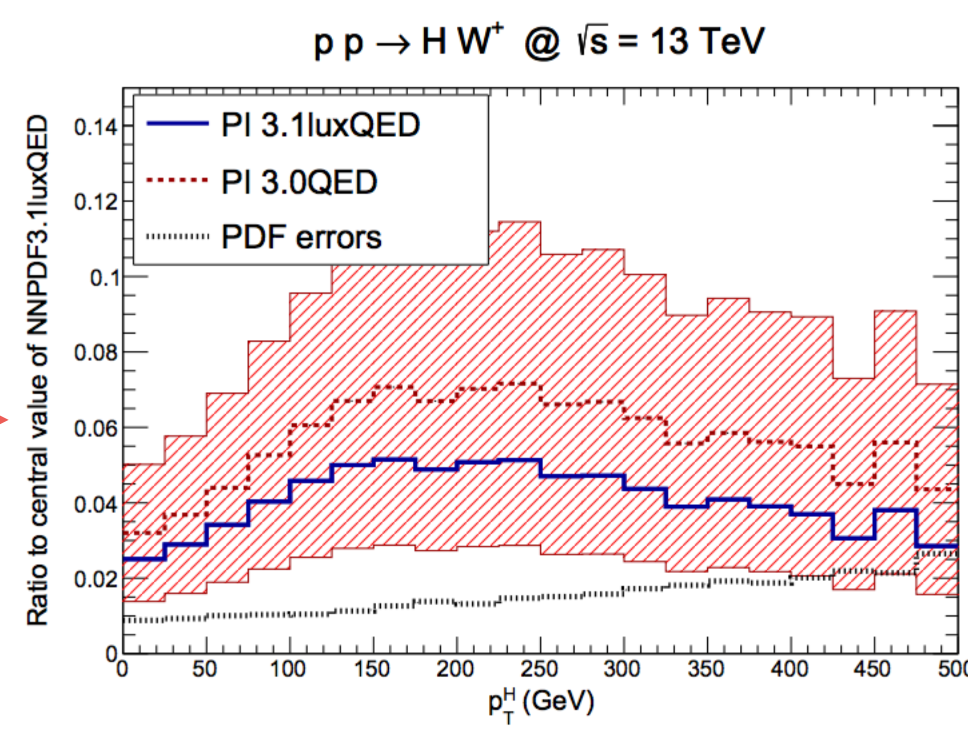
(Data+Theory)-driven knowledge

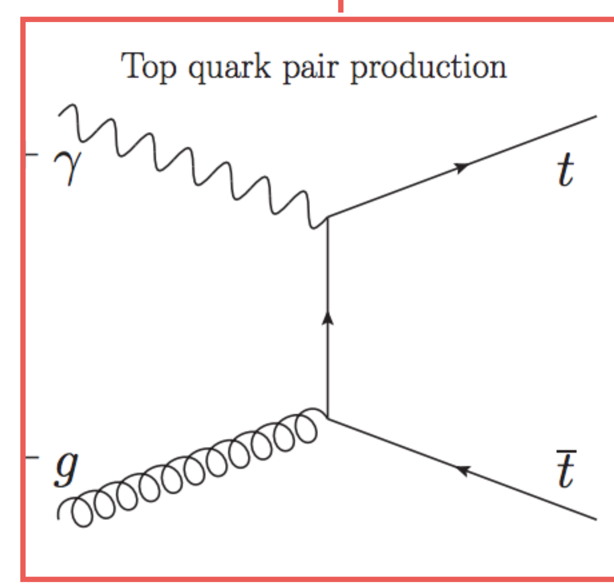




Bertone et al, 1712.07053



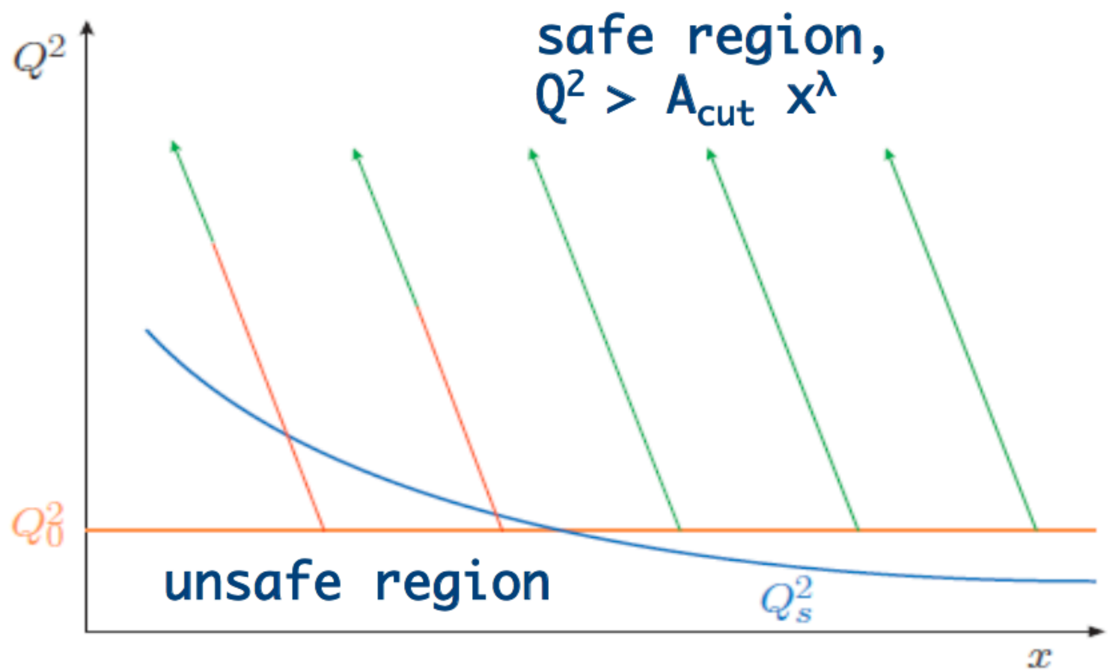




Beyond Collinear Factorisation

Beyond DGLAP

- In DGLAP formalism there is an implicit approximation: the transverse momentum of the emitted partons in the initial state is much smaller than hard scale
- It works well for inclusive processes with one hard scale and for not-too-small x



Possible effects beyond DGLAP

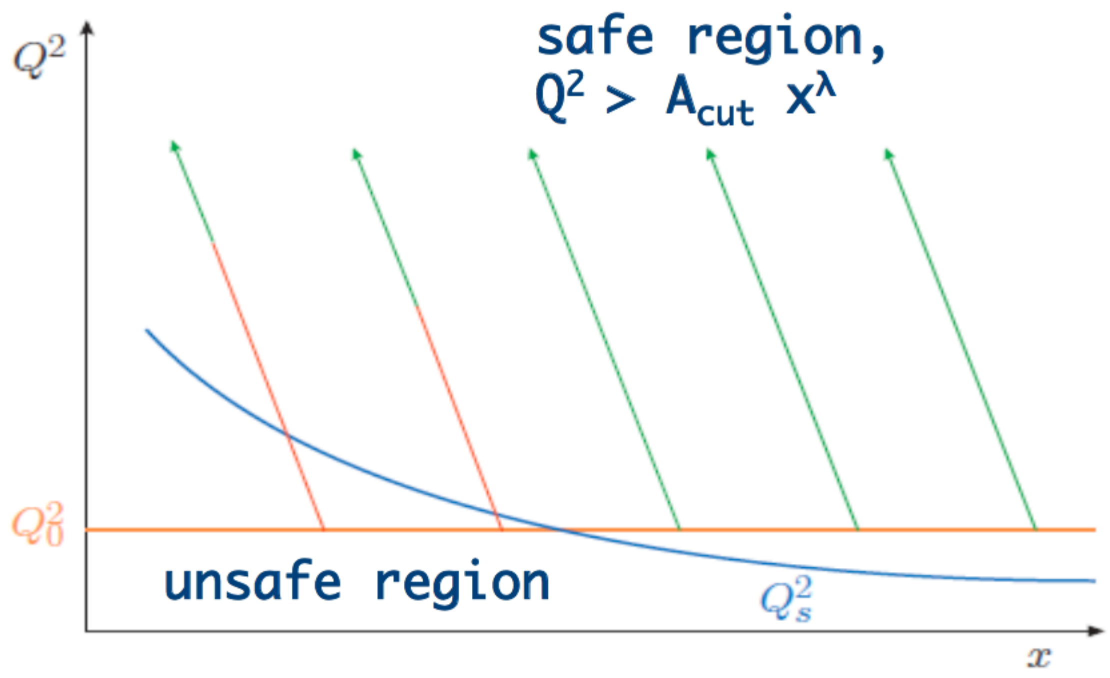
(i) Leading-twist small-x perturbative resummation

(ii) Non-linear evolution and saturation

(iii) Higher twist effects

Beyond DGLAP

- In DGLAP formalism there is an implicit approximation: the transverse momentum of the emitted partons in the initial state is much smaller than hard scale
- It works well for inclusive processes with one hard scale and for not-too-small x



Possible effects beyond DGLAP

(i) Leading-twist small-x perturbative resummation

(ii) Non-linear evolution and saturation

(iii) Higher twist effects

(i) Small-x resummation

- In DGLAP formalism there is an implicit approximation: the transverse momentum of the emitted partons in the initial state is much smaller than hard scale
- It works well for inclusive processes with one hard scale and for not-too-small x
- If s >> M^2 (the high-energy limit or small-x limit) then there are enhanced small-x logarithms in the DGLAP Pqg and Pgg splitting functions that spoil the perturbative expansion in α_s
- These large logs are resumed by BFKL evolution equations

DGLAP Evolution in Q^2

$$\frac{\partial}{\partial \ln Q^2} f_i(x, Q^2) = \int_x^1 \frac{dz}{z} P_{ij}\left(\frac{x}{z}, \alpha_s(Q^2)\right) f_j(z, Q^2)$$

BFKL Evolution in x

$$\frac{\partial}{\partial \ln 1/x} f_+(x, Q^2) = \int_0^\infty \frac{d\nu^2}{\nu^2} K\left(\frac{Q^2}{\nu^2}, \alpha_s(Q^2)\right) f_+(x, \nu^2)$$

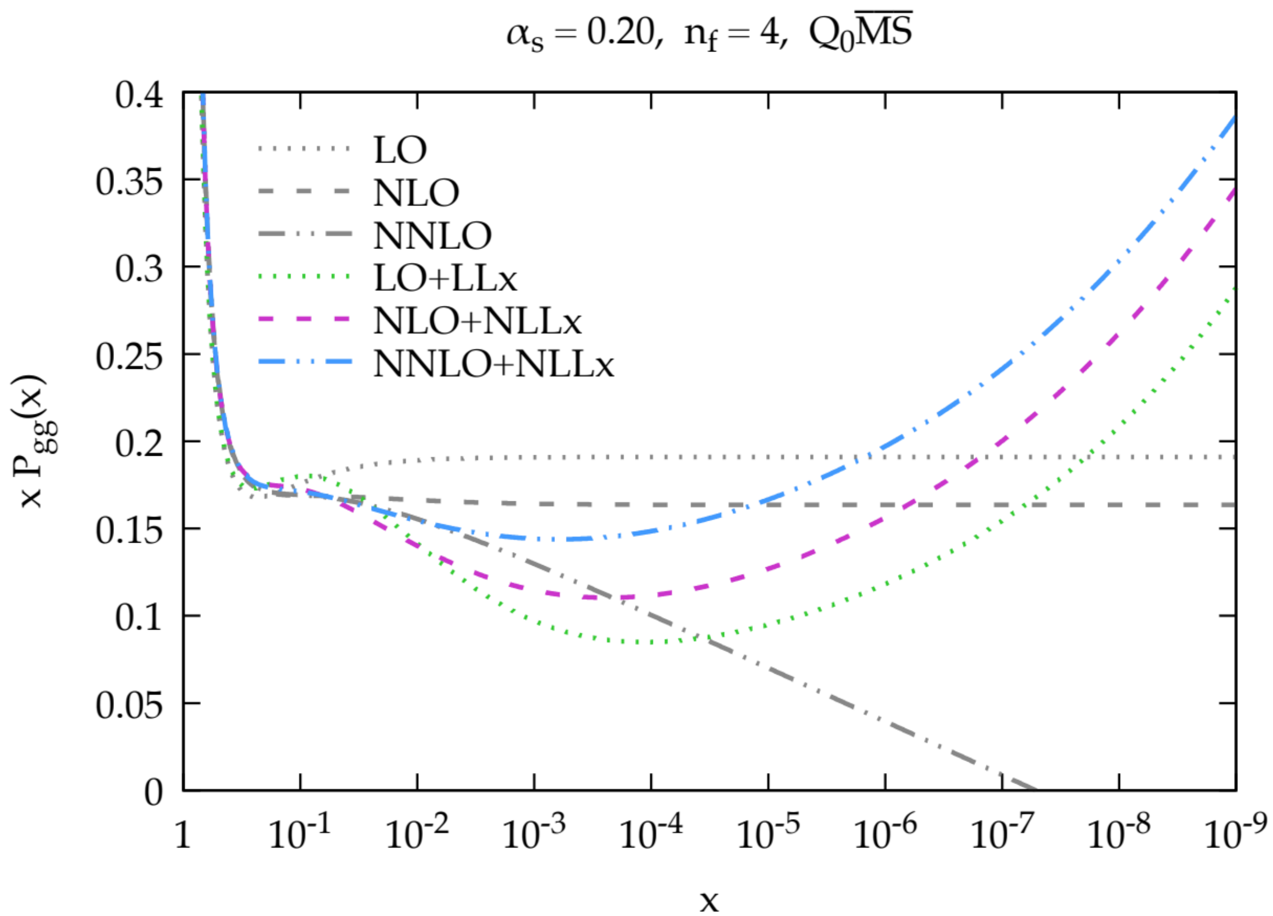
Valid only at small-x

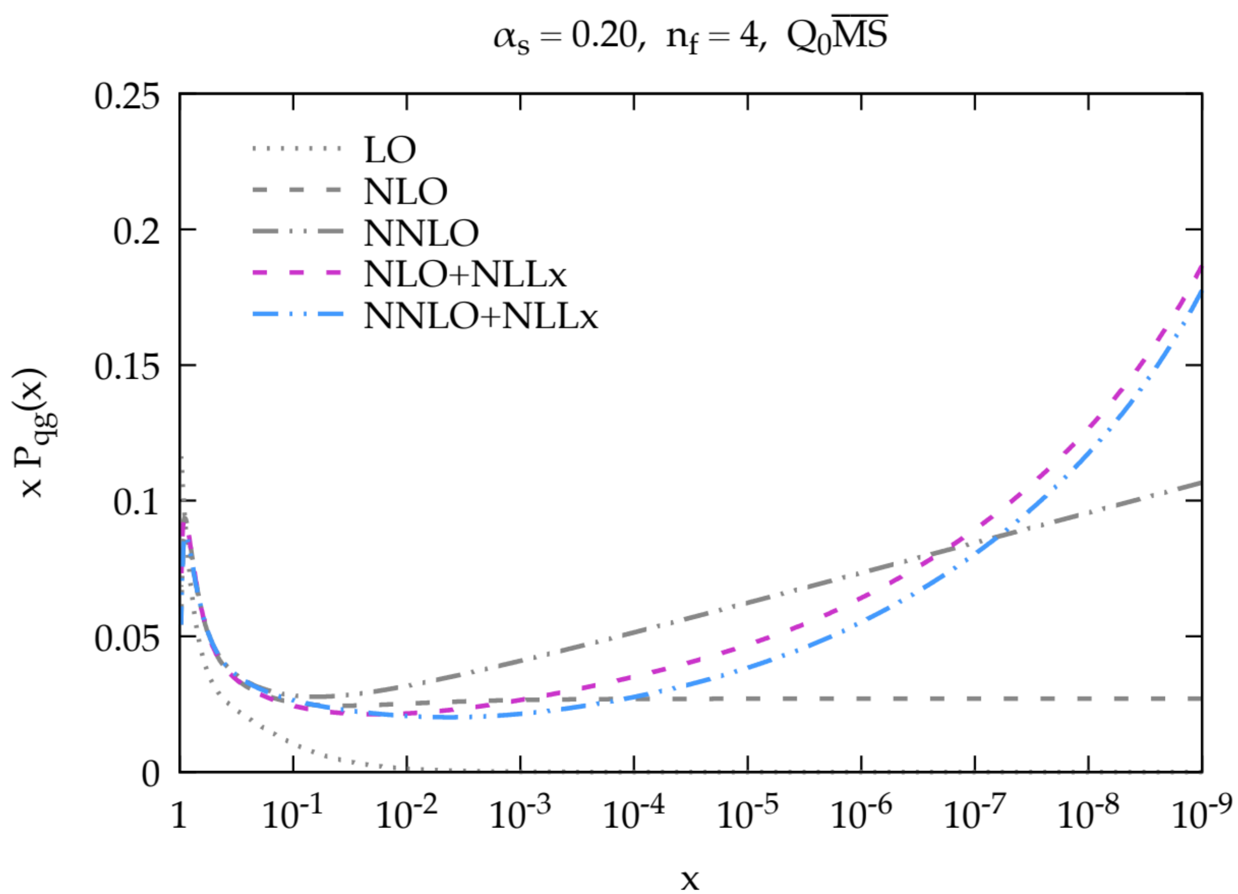
- There are way of combining DGLAP & BFKL - what are the effects?

Balitsky Fadin Kuraev Lipatov 1978, Lipatov Fadin 1998

(i) Small-x resummation

- Small-x resummation stabilises splitting function behaviour at small x

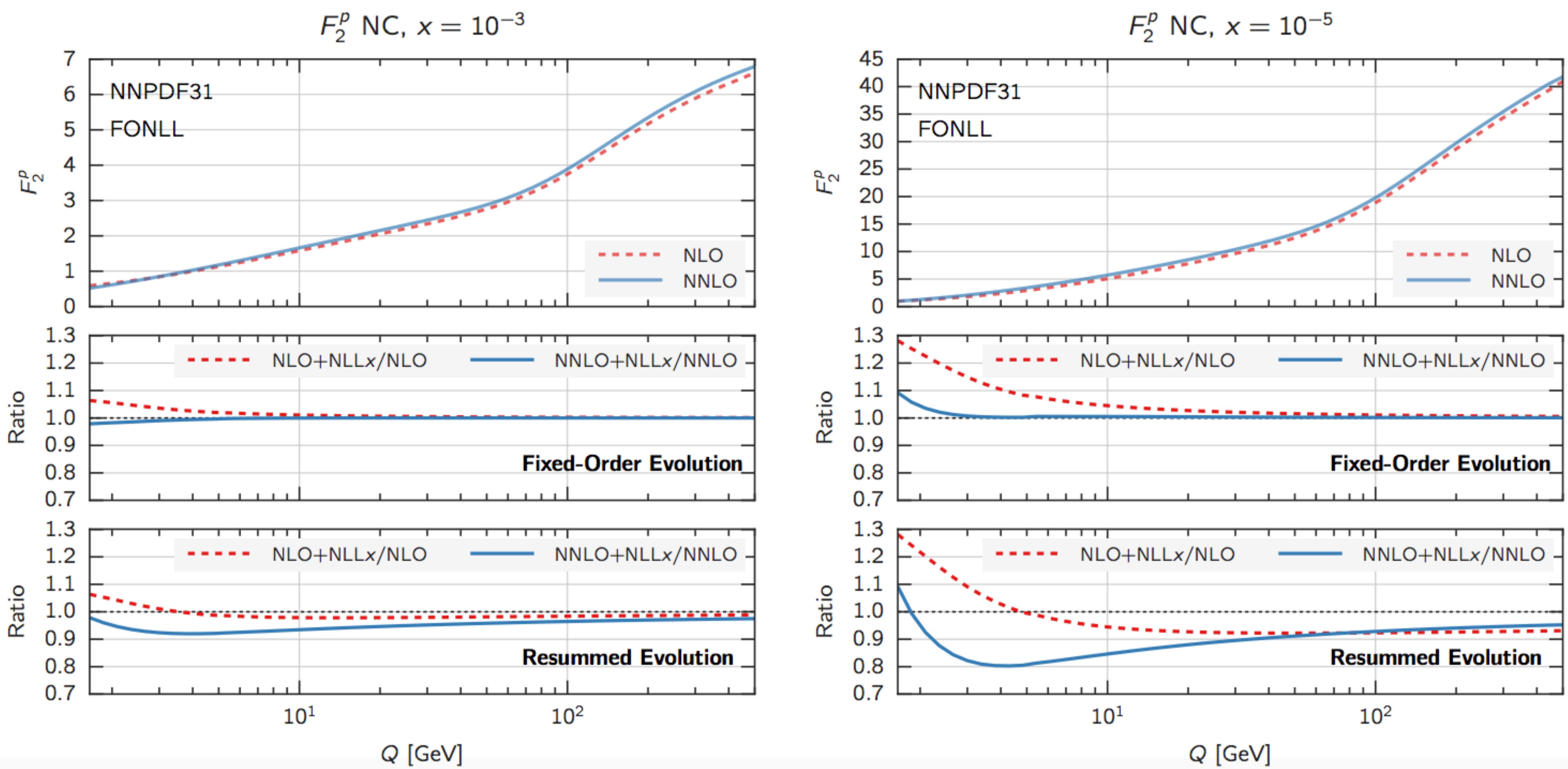




Ball et al, 1710.05935

(i) Small-x resummation

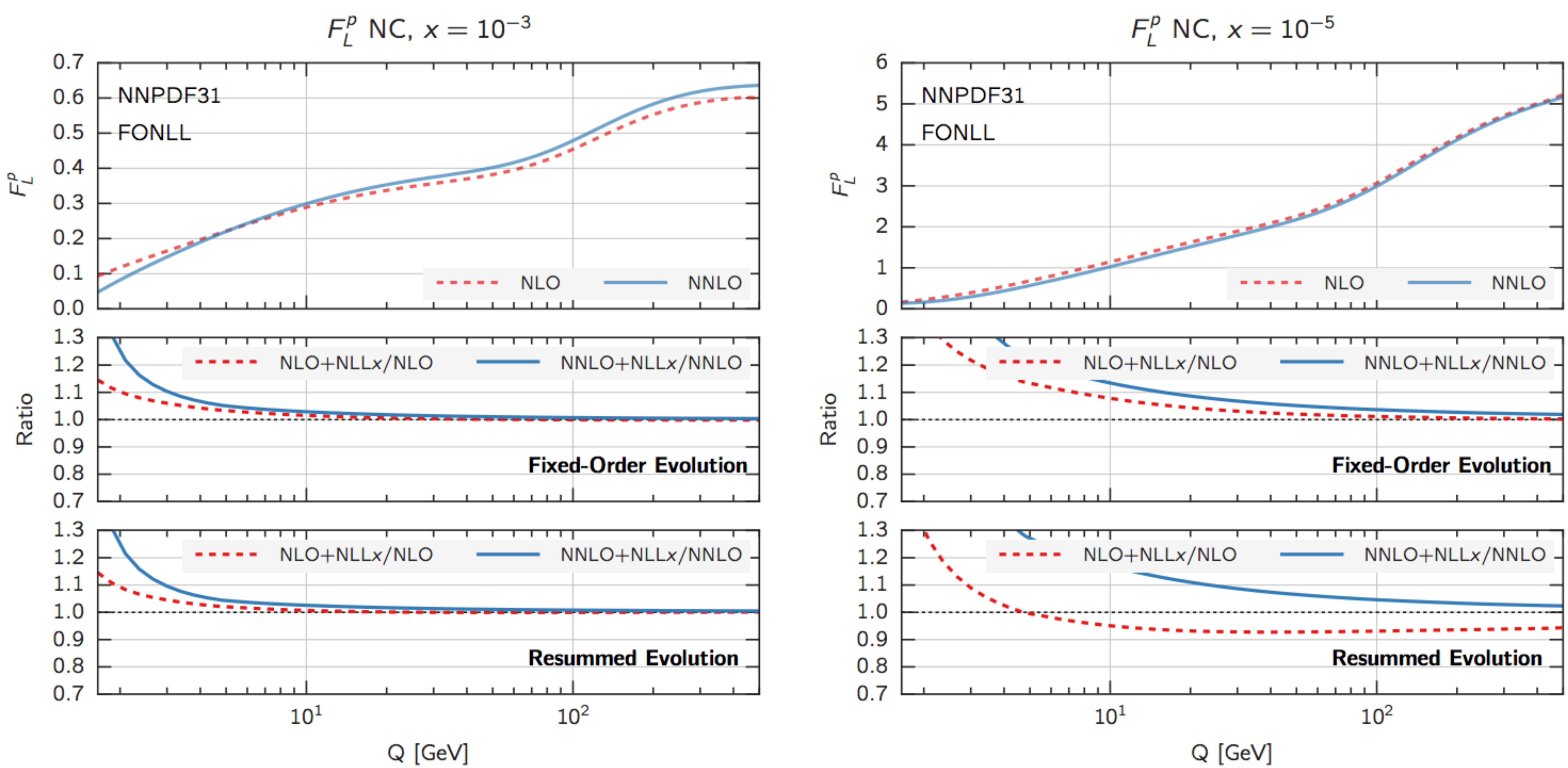
- Large corrections at small Q2 and small-x



Ball et al, 1710.05935

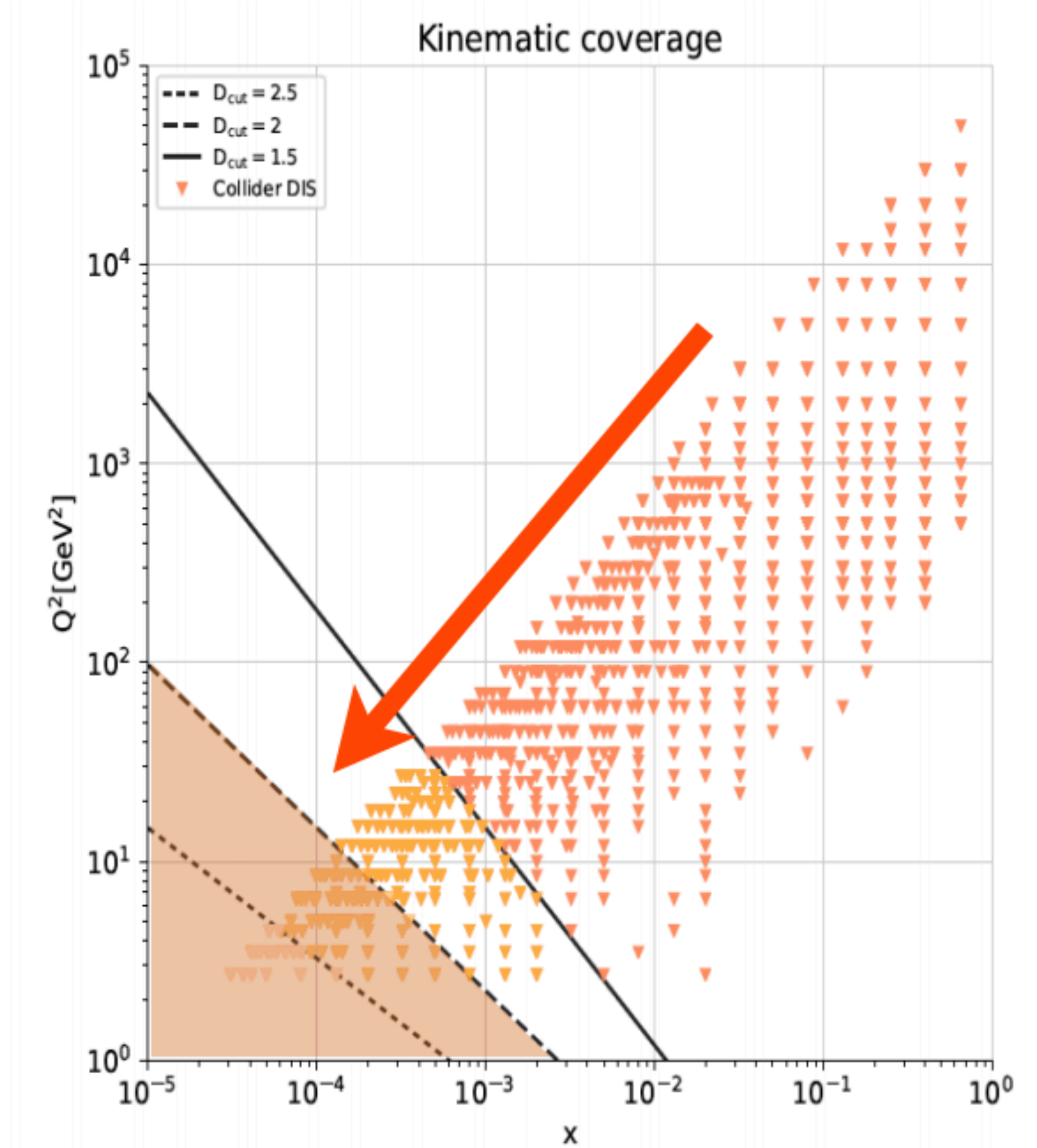
(i) Small-x resummation

- Large corrections at small Q2 and small-x, especially for FL

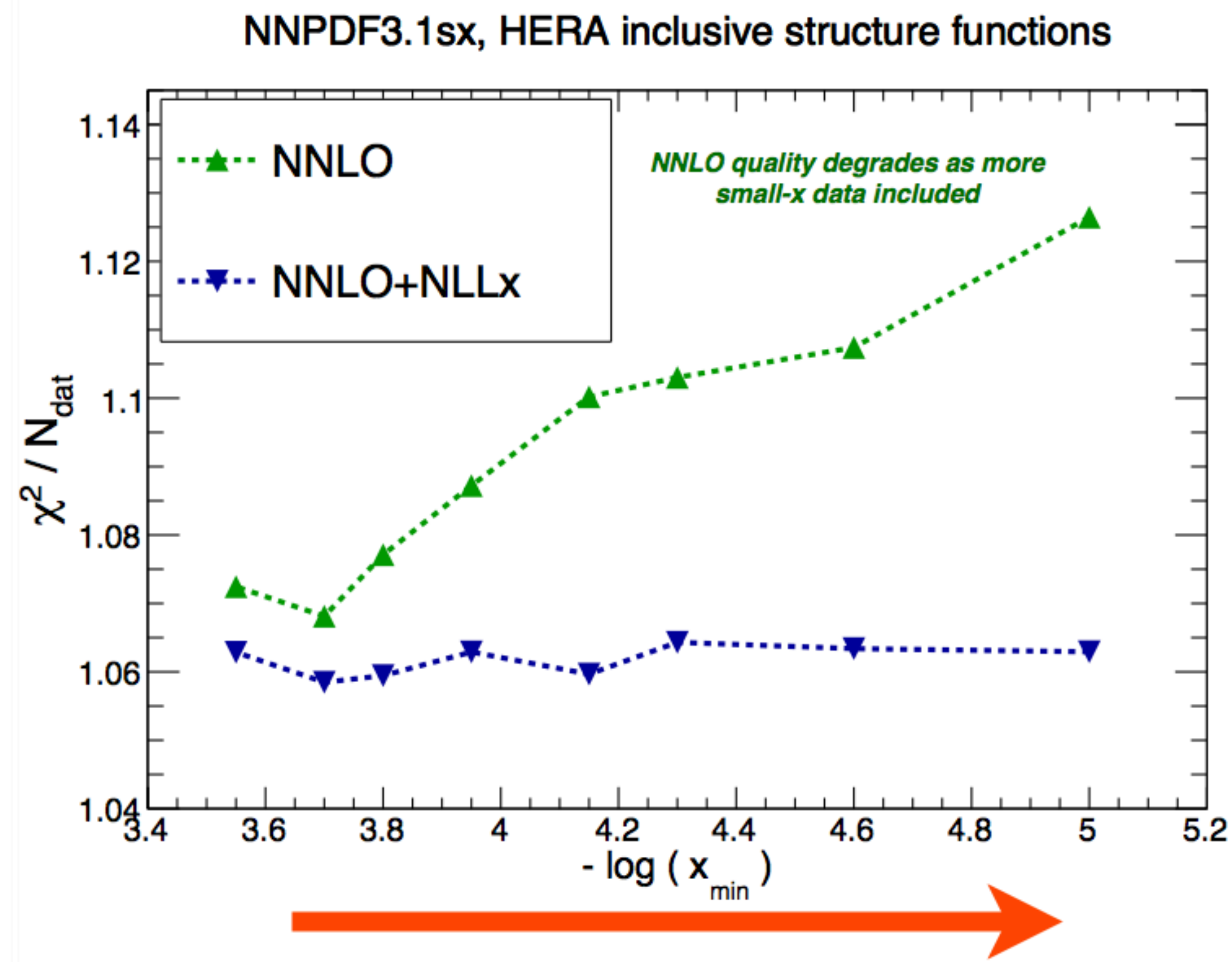


Ball et al, 1710.05935

(i) Small-x resummation



Monitor the **fit quality** as one includes more data from the **small-*x* region**



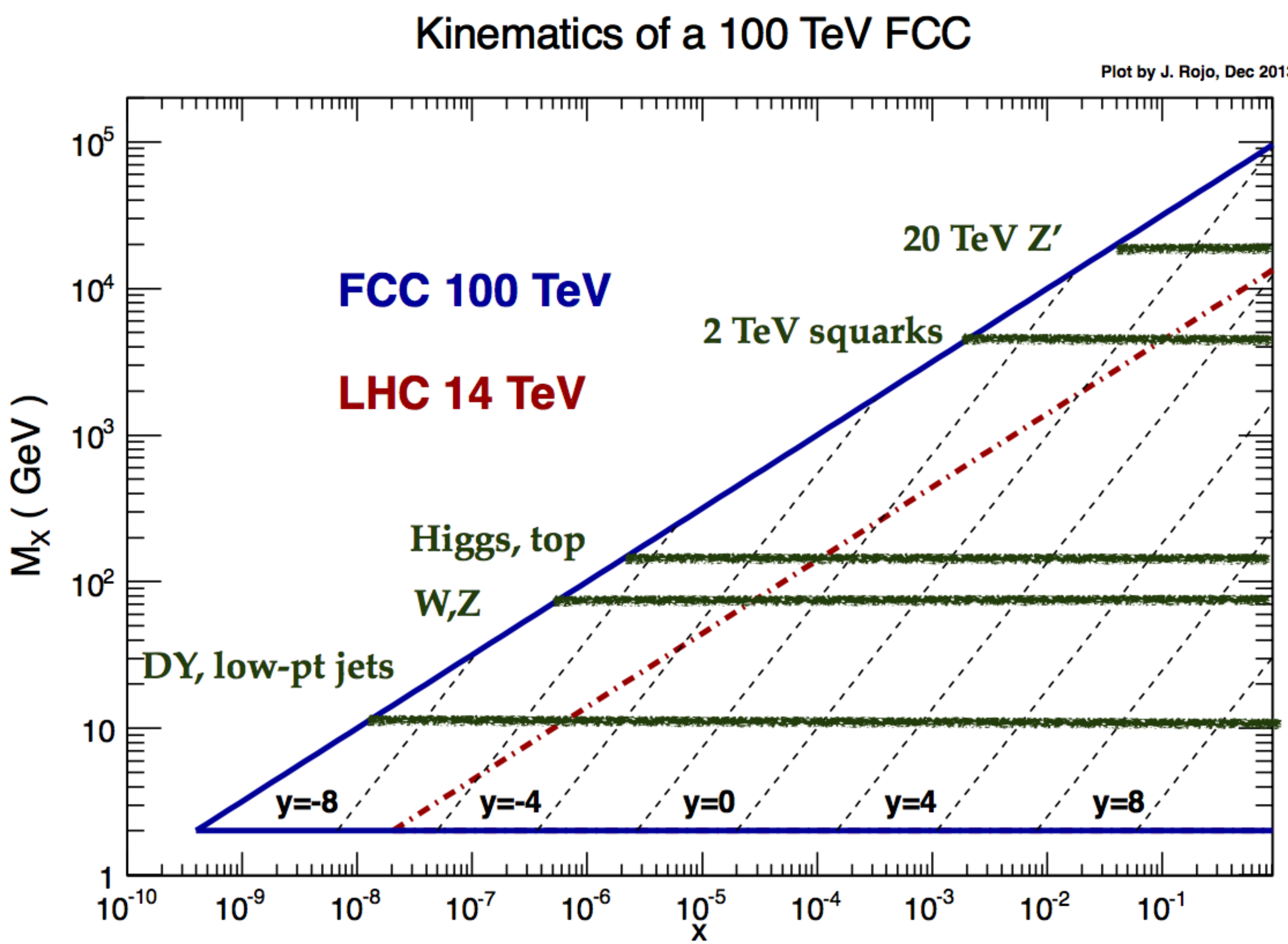
Best description of **small-*x* HERA data** only possible with **BFKL effects!**

J Rojo, BNL talk

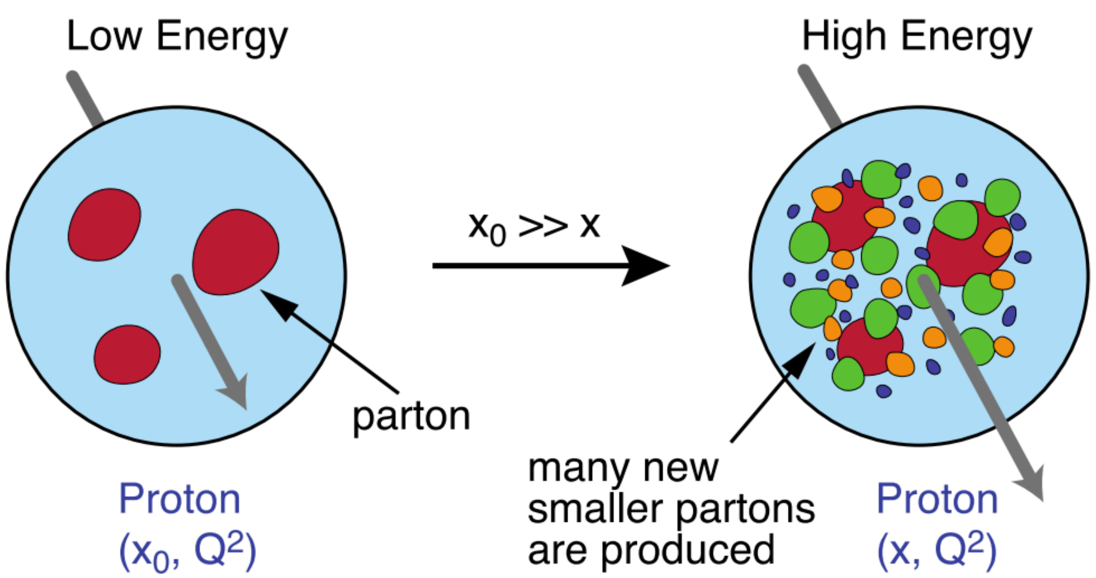
Ball et al, 1710.05935

(i) Small-x resummation

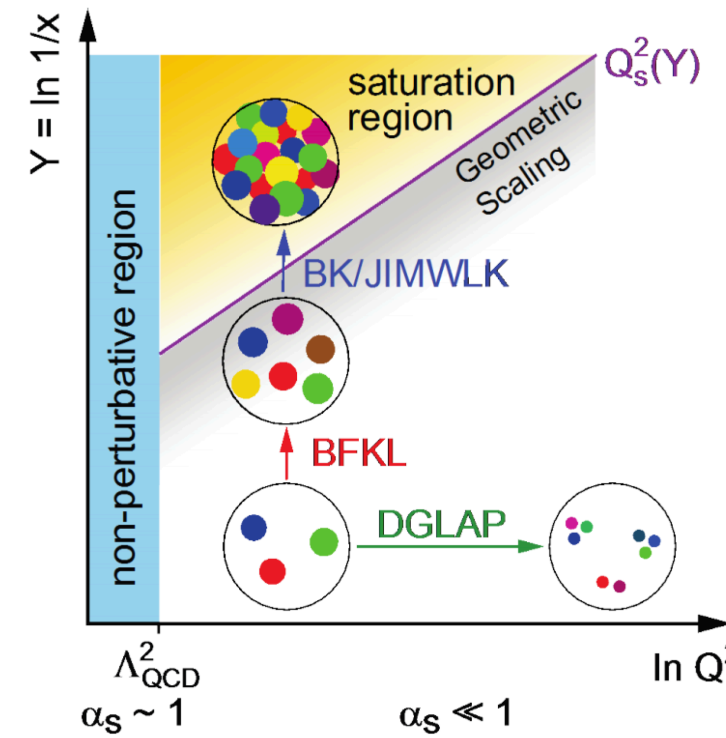
- Will this be enough when we will reach even smaller values of x?



(ii) Non-linear evolution and saturation

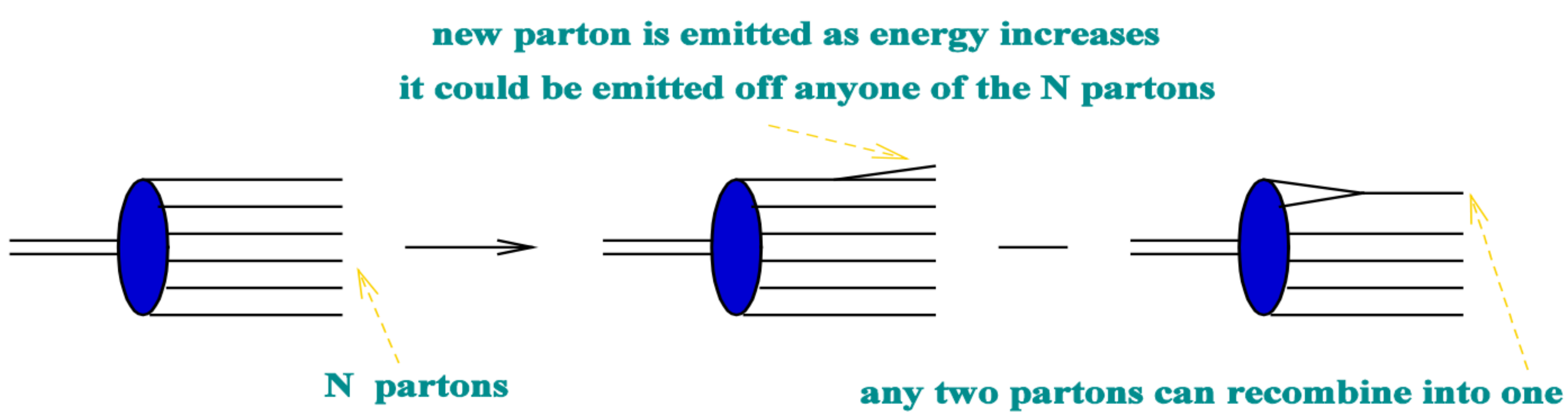


"Color Glass Condensate"



$Q_s^2 \sim A^{1/3} s^\lambda$

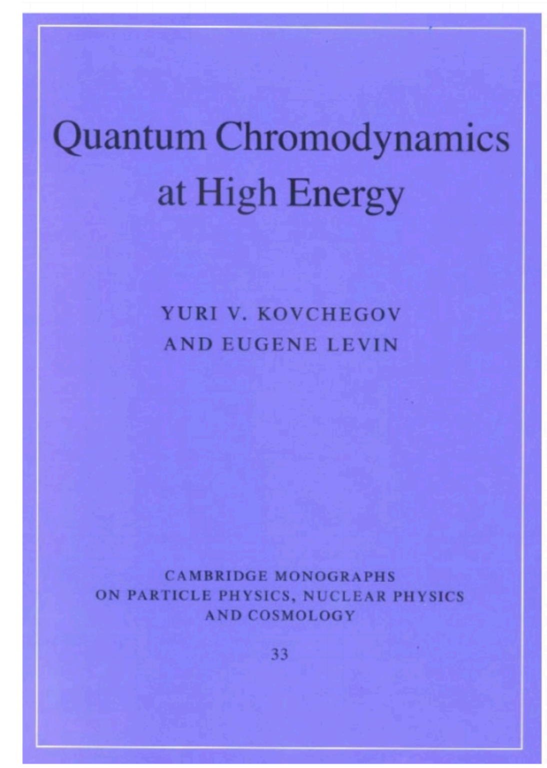
Important for ion-ion collisions



$$\frac{\partial}{\partial Y} N(x,k_T^2) = \alpha_s \, K_{BFKL} \otimes N(x,k_T^2) - \alpha_s \, [N(x,k_T^2)]^2$$

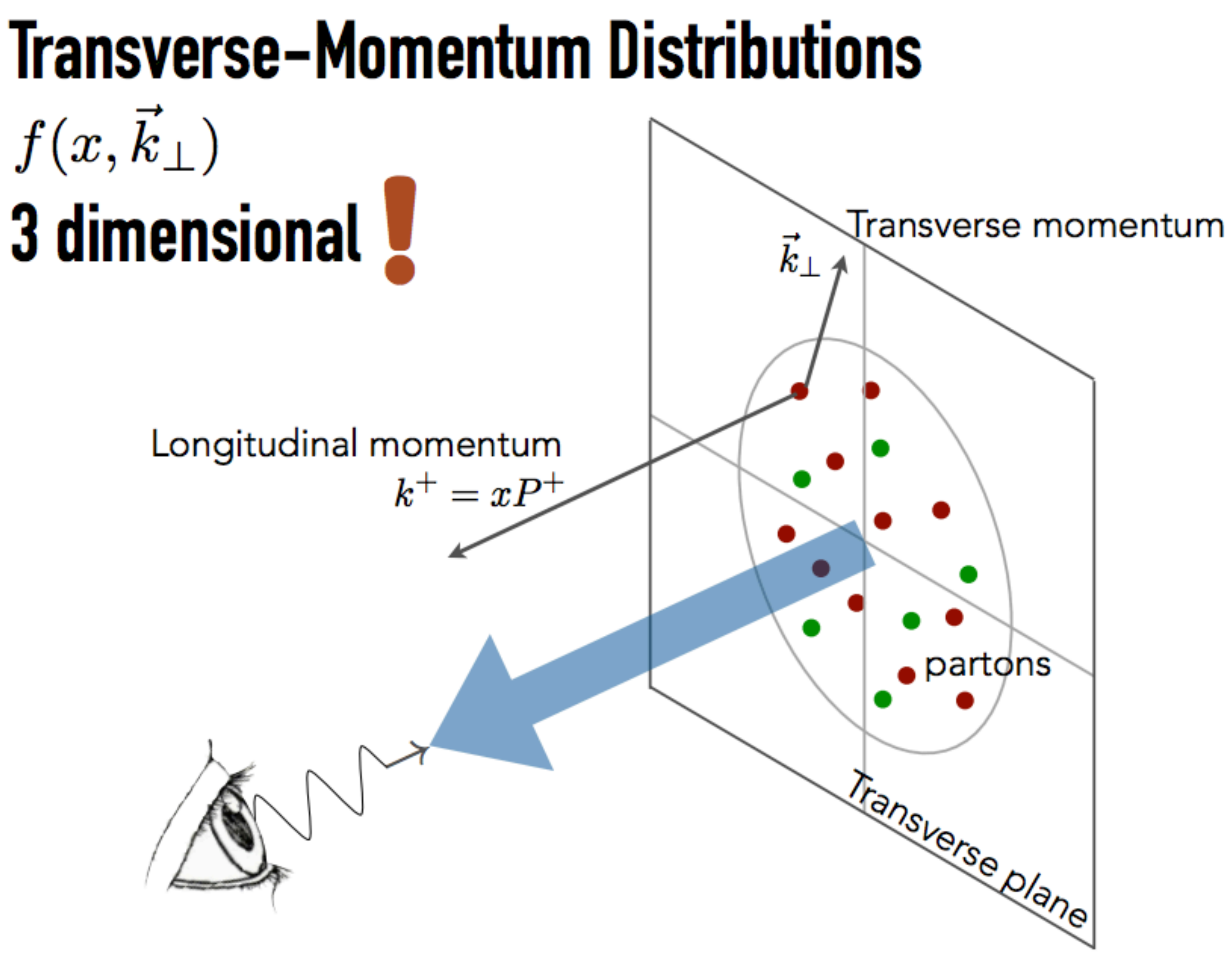
Number of parton pairs ~ N^2

I. Balitsky ' 96 (effective Lagrangian)
Yu. K. ' 99 (large N_c QCD)
JIMWLK '98-'01 (beyond large-N_c)



Beyond DGLAP: TMDs

- Much less mature field (universality and factorisation not well established but lots of interesting developments)



A. Bacchetta, talk at DIS2017

Covariance matrix

- Theory is perturbative expansion to some order : $t_p = \sum_{m=0}^{p} c_m$
- Standard case: $P(d|t_p) \propto \exp\Big(-\frac{1}{2} \underbrace{(d-t_p)^T \mathrm{cov}_{\mathrm{exp}}^{-1} (d-t_p)}_{\chi^2_{\mathrm{exp}}} \Big)$
- Bayes' theorem: $P(t_p|d) = \dfrac{P(d|t_p)P(t_p)}{P(d)} \propto P(d|t_p)P(t_p)$
- Assume Gaussian theory prior:

$$P(t_p) = \prod_{m=0}^{p} P(c_m) \quad \text{where} \quad P(c_m) \propto \exp\Big(-\frac{1}{2} \underbrace{c_m^T \mathrm{cov}_{\mathrm{th},m}^{-1} c_m}_{\chi^2_{\mathrm{th}}} \Big)$$

- Assume MHOUs due to $\mathcal{O}(\alpha^{p+1})$ terms only → marginalise these terms:

$$P(t_p|d) \propto \int dc_{p+1} P(d|c_{p+1}) P(t_{p+1})$$

$$\propto \exp\Big(-\frac{1}{2} \underbrace{(d-t_p)^T (\mathrm{cov}_{\mathrm{exp}} + \mathrm{cov}_{\mathrm{th}})^{-1} (d-t_p)}_{\chi^2_{\mathrm{tot}}} \Big)$$

- Include higher order terms by induction

Covariance matrix

$$\chi^2 = \sum_{m,n=1}^{N} (d_m - t_m)(\mathrm{cov}_{\mathrm{exp}} + \mathrm{cov}_{\mathrm{th}})^{-1}_{mn}(d_n - t_n)$$

➡ How to build correlations between different points?

$$(\mathrm{cov}_{\mathrm{th}})_{mn} = \langle (t_p(\mu_R, \mu_F) - t_p(\mu_R^0, \mu_F^0))_m (t_p(\mu_R, \mu_F) - t_p(\mu_R^0, \mu_F^0))_n \rangle$$

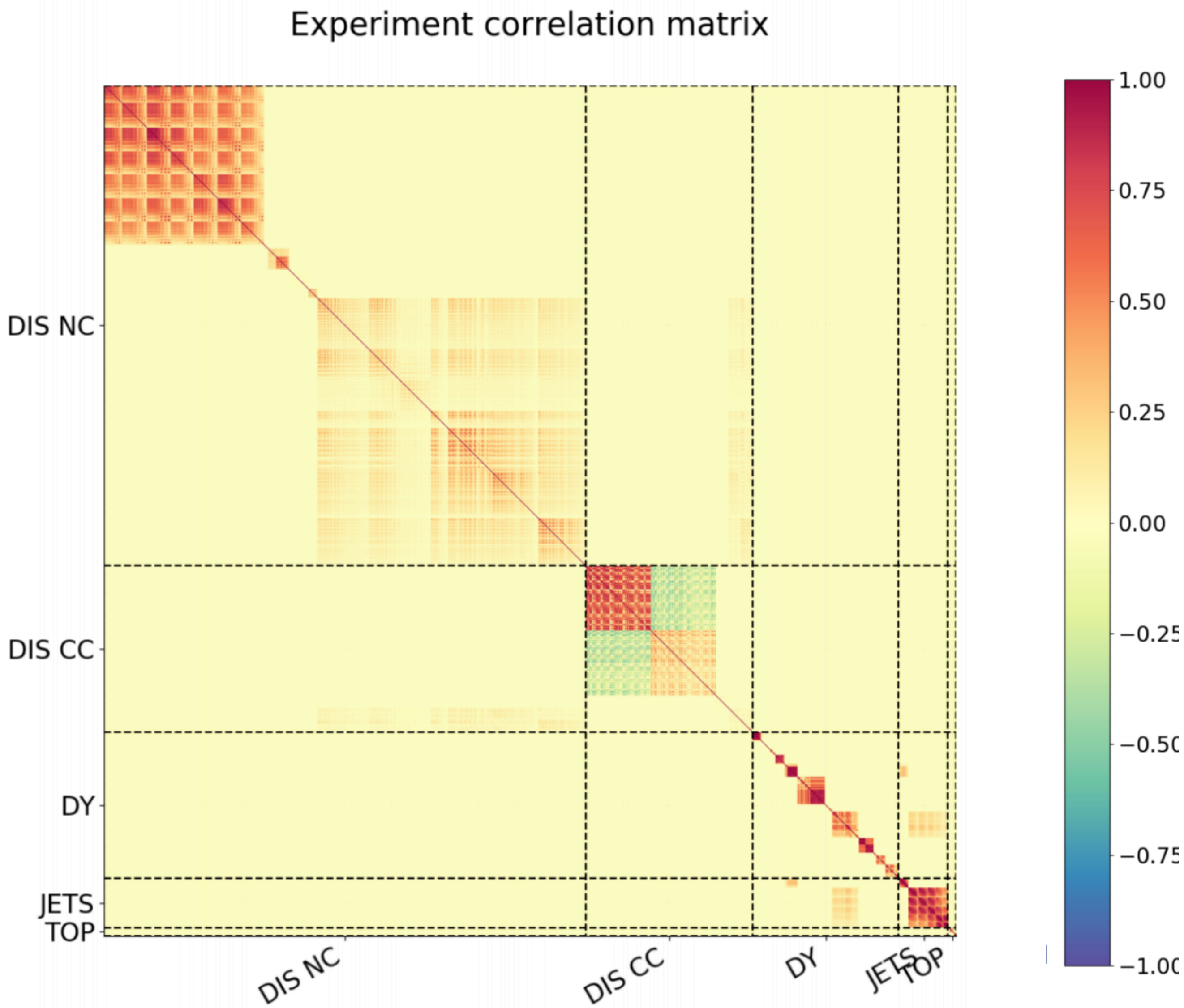
- $\boldsymbol{\mu}_F$ variations correlated across all processes by PDF evolution
- $\boldsymbol{\mu}_R$ variation correlated by process (hard cross section)

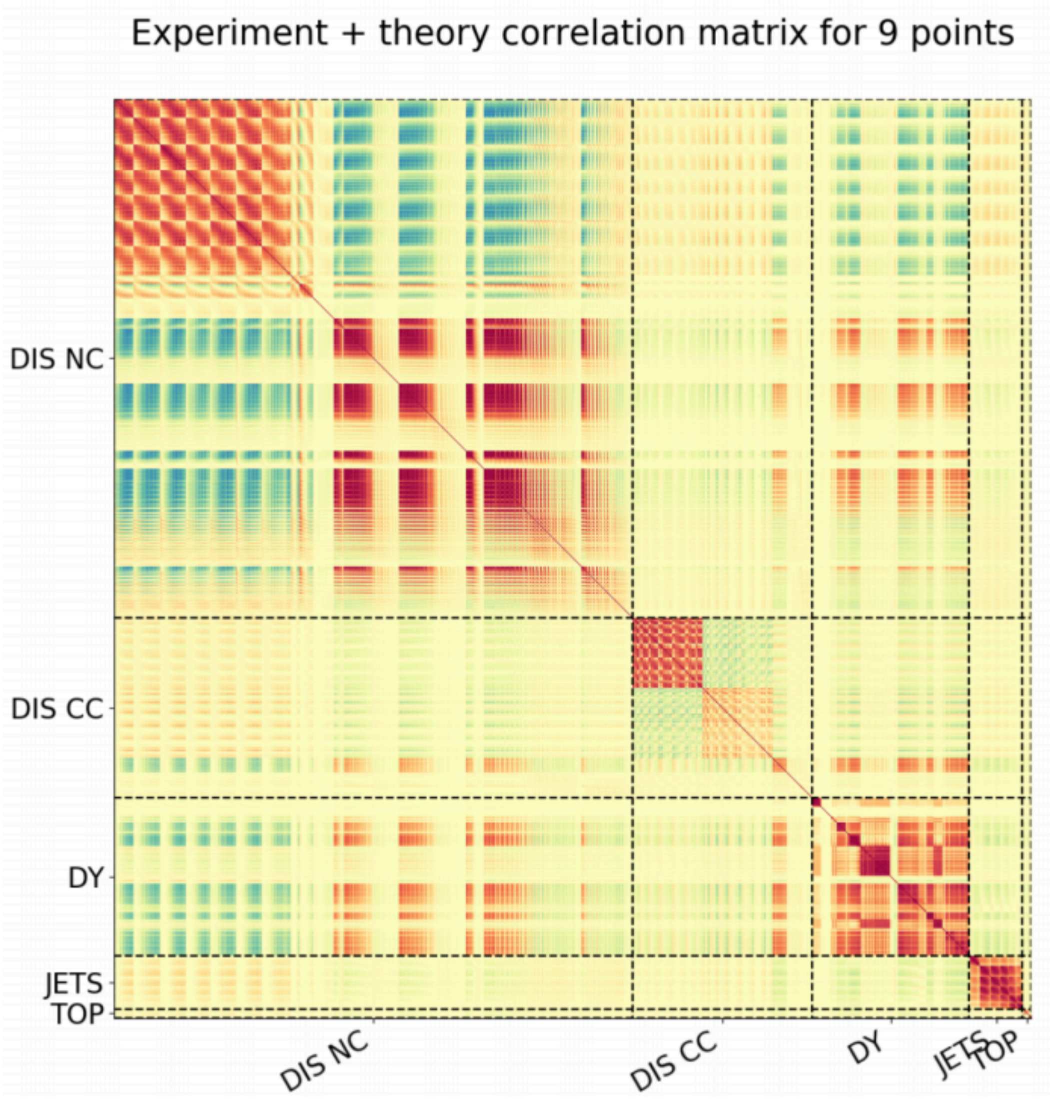
➡ Several recipes possible (3-points prescriptions, 7-points…)

➡ Details of correlations are also important

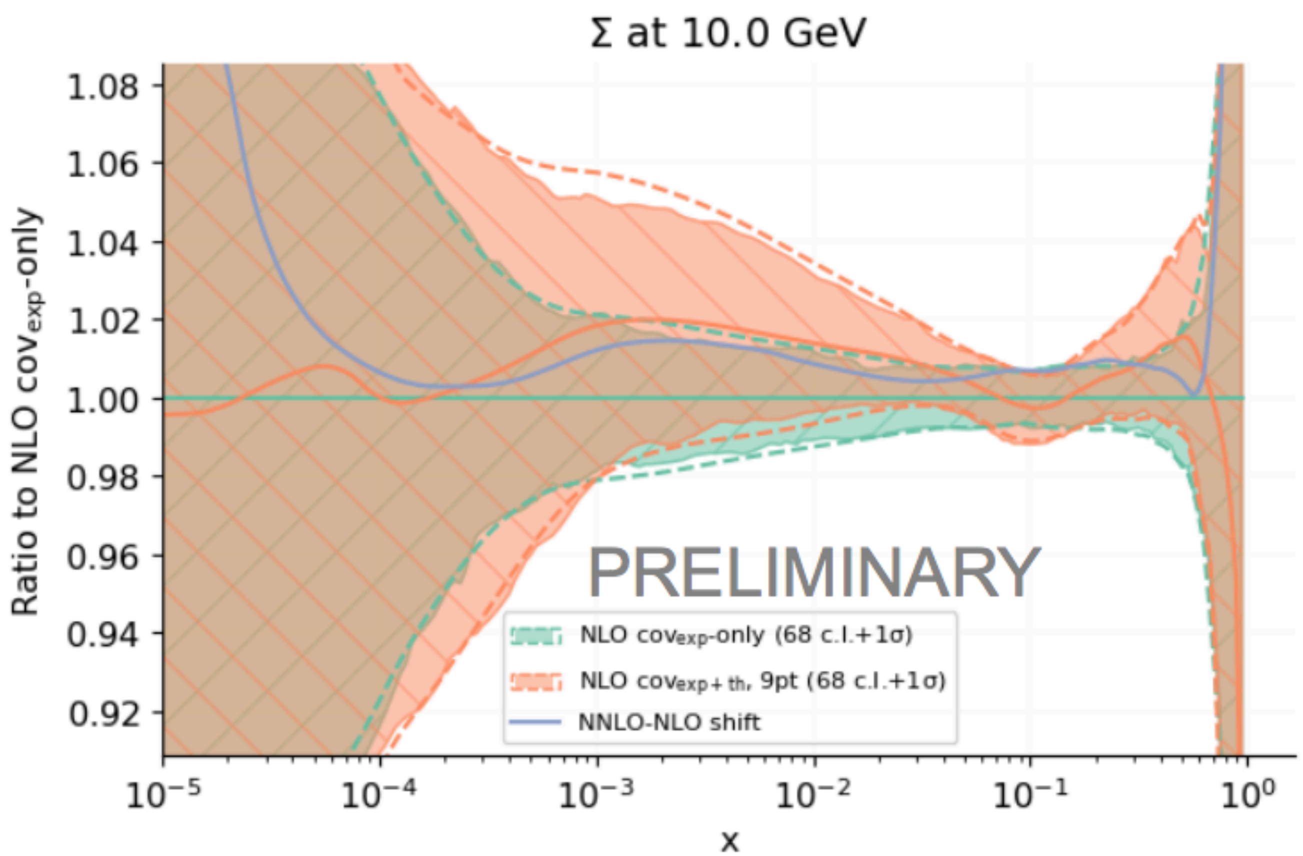
➡ A lot to be investigated

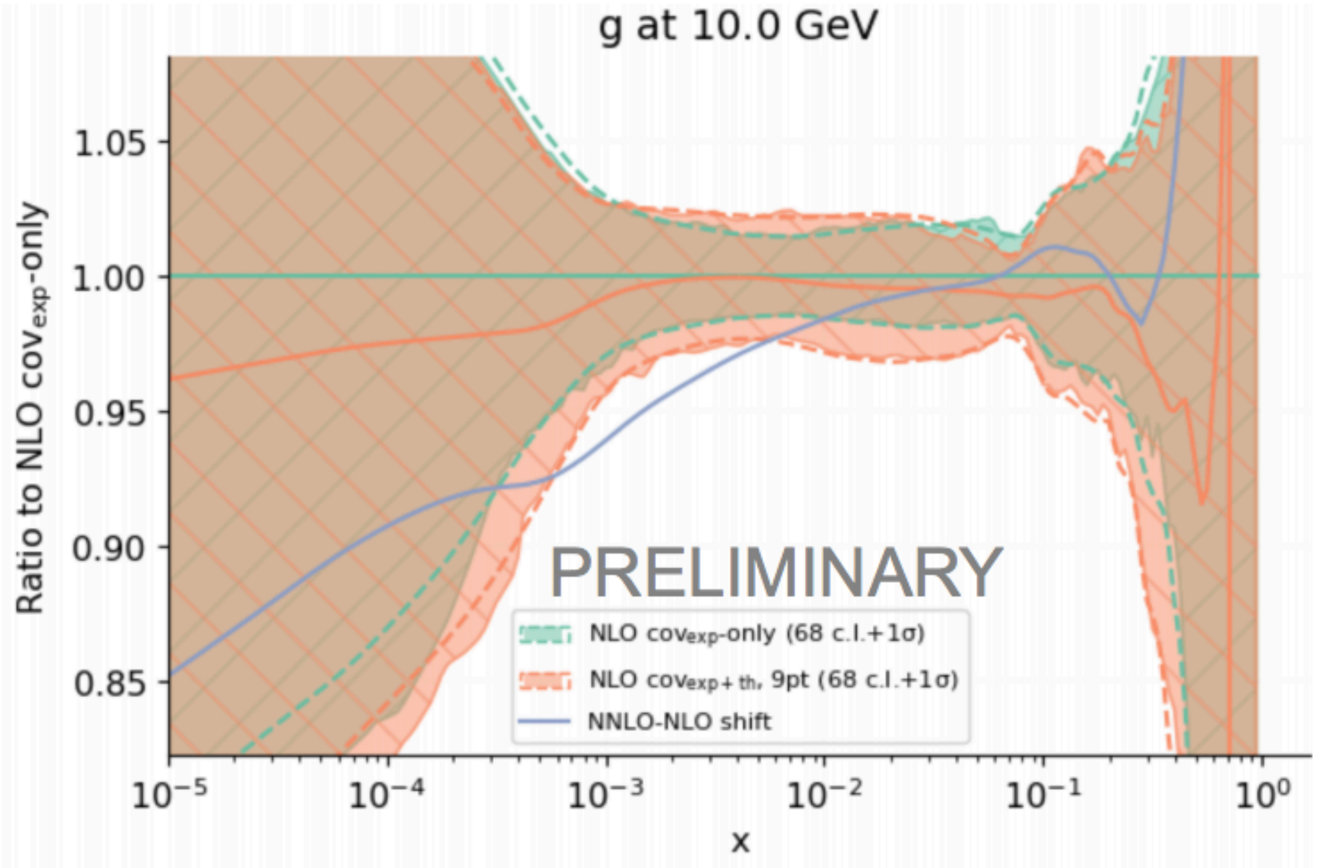
Covariance matrix





More reliable uncertainties?





Beyond fixed order

- Multi-scale processes: log(Qi/Qj) = L arise, which may spoil perturbative expansion
- If (α_S * L) ~ O(1) fixed order perturbative QCD is no longer justified
- Resummation effectively rearranges perturbative series

fixed order		all order (L = some large logarithm)	
$\frac{\sigma}{\sigma_0} = 1$	LO	$\ln\frac{\sigma}{\sigma_0} = \alpha^n L^{n+1}$	LL
$+ c_1\alpha$	NLO	$+ \alpha^n L^n$	NLL
$+ c_2\alpha^2$	NNLO	$+ \alpha^n L^{n-1}$	NNLL
$+ \ldots$		$+ \ldots$	

- Various kinds of logs:

L = log (1-x) threshold (soft-gluon) resummation ⟵ **Ball et al, JHEP09(2015)091**

L = log (1/x) high-energy (small-x) resummation ⟵ **BFKL**

L = log (pT/M) transverse momentum resummation

Threshold resummation

- Threshold resummation: initial energy just enough to produce final state with mass M, so emissions forced to be soft and logs at each order in PT are enhanced

$$x = \frac{M^2}{\hat{s}} \qquad \mathrm{NLO}: \; M^2 = z\hat{s} \qquad \left[\frac{\log^k(1-z)}{(1-z)}\right]_+$$

- Transform factorised cross section into Mellin space

$$\sigma(x,Q^2) = x\sum_{a,b}\int_x^1 \frac{dz}{z}\mathcal{L}_{ab}\left(\frac{x}{z},\mu_\mathrm{F}^2\right)\frac{1}{z}\hat{\sigma}_{ab}\left(z,Q^2,\alpha_s(\mu_\mathrm{R}^2),\frac{Q^2}{\mu_\mathrm{F}^2},\frac{Q^2}{\mu_\mathrm{R}^2}\right)$$

$$\sigma(N,Q^2) = \int_0^1 dx\, x^{N-2}\sigma(x,Q^2) = \sum_{a,b}\mathcal{L}_{ab}(N,Q^2)\hat{\sigma}_{ab}\left(N,Q^2,\alpha_s\right)$$

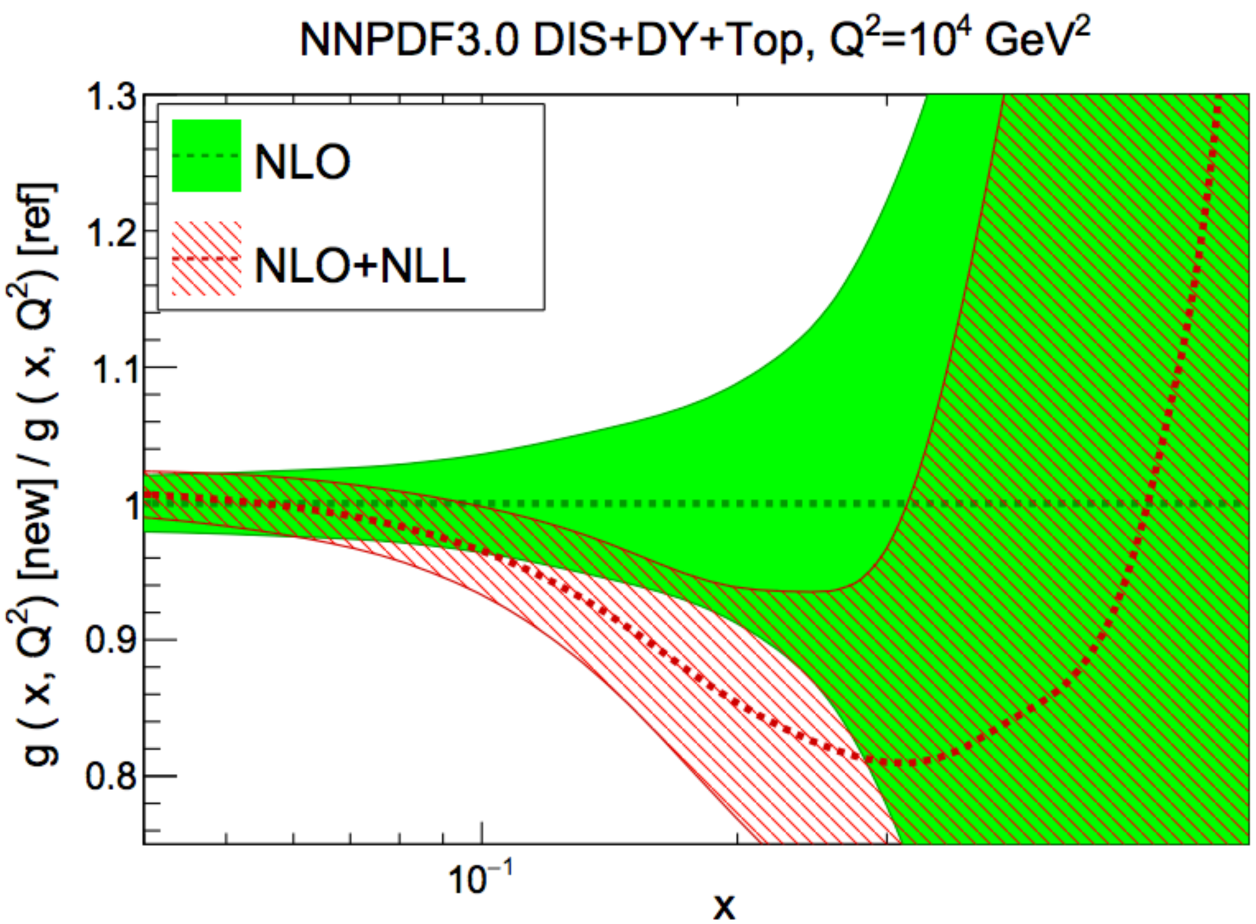
- In the MSbar scheme PDF evolution does not contain large-x logs and the effect of resummation can be included in resummed coefficient functions

$$\hat{\sigma}_{ab}^{(\mathrm{res})}(N,Q^2,\alpha_s) = \sigma_{ab}^{(\mathrm{born})}(N,Q^2,\alpha_s)\,C_{ab}^{(\mathrm{res})}(N,\alpha_s)$$

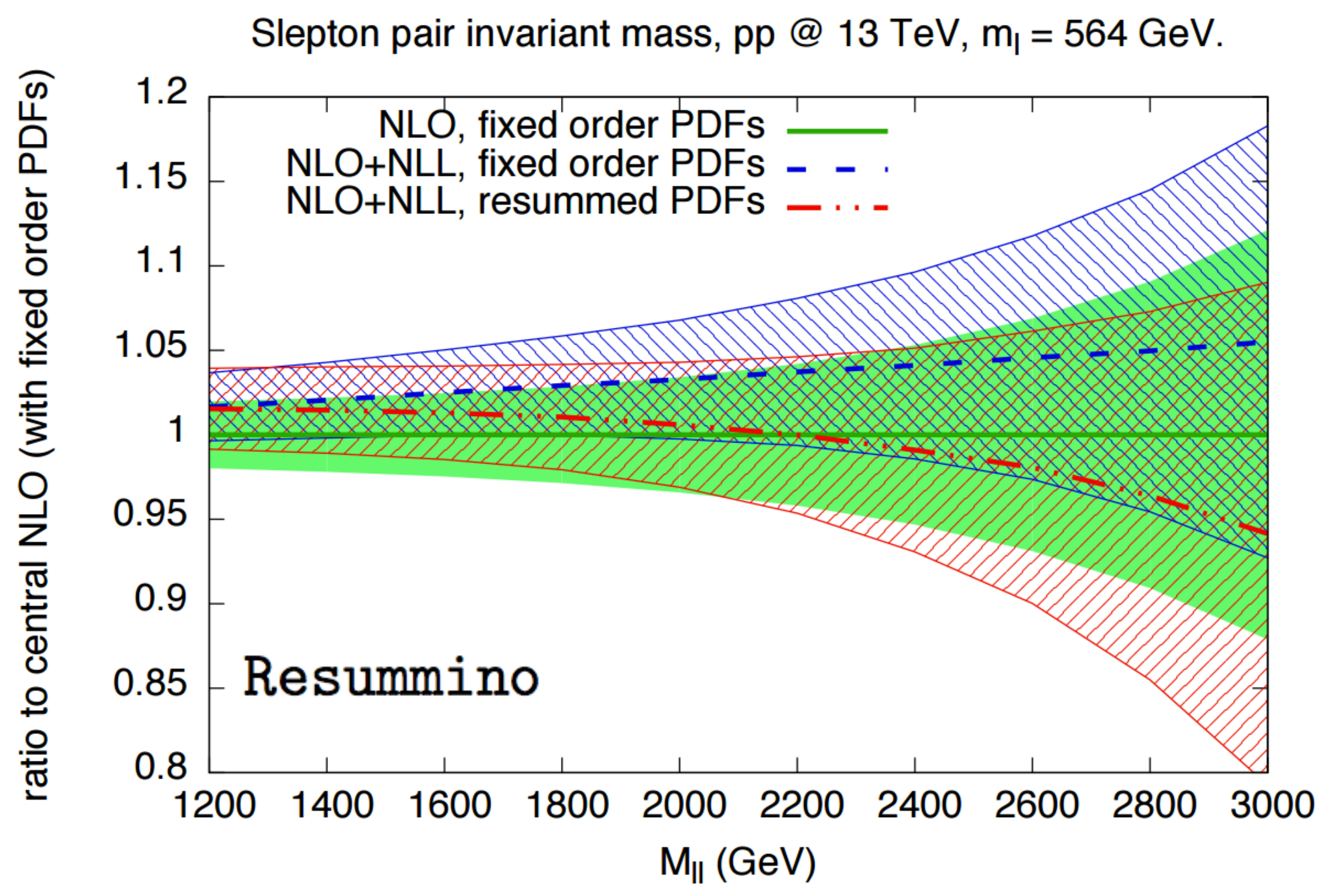
$$C^{(N\text{-soft})}(N,\alpha_s) = g_0(\alpha_s)\exp\mathcal{S}(\ln N,\alpha_s),$$

$$\mathcal{S}(\ln N,\alpha_s) = \left[\frac{1}{\alpha_s}g_1(\alpha_s\ln N) + g_2(\alpha_s\ln N) + \alpha_s g_3(\alpha_s\ln N) + \ldots\right]$$

Threshold resummation



Bonvini et al, JHEP 1509 (2015) 191



- Threshold-resummed PDFs will be suppressed as compared to fixed-order PDFs
- Mostly due to enhancement of NLO+NLL xsecs used in the fit of DIS structure functions and DY distributions
- This suppression partially or totally compensates enhancements in partonic cross sections
- Phenomenologically relevant for new physics processes [Beenakker et al. EPJC76 (2016)2, 53]