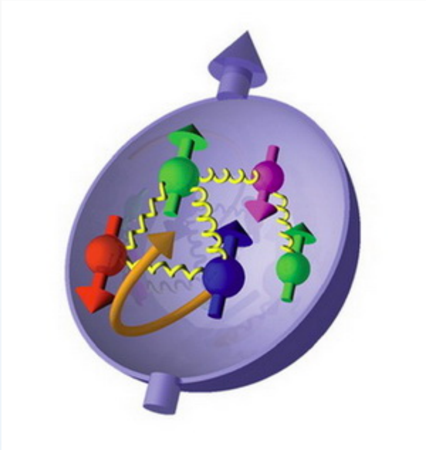
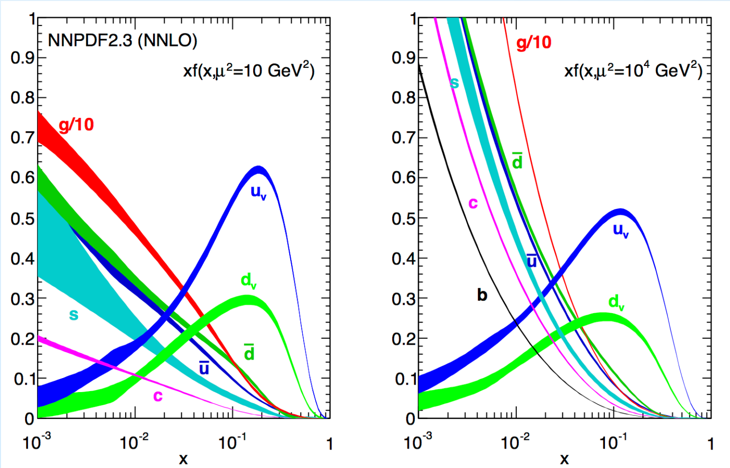
The structure of the proton and precision LHC phenomenology

Juan Rojo
VU Amsterdam & Theory group, Nikhef

High Energy Physics & Cosmology Seminar
SUNY Buffalo, 30/11/2016

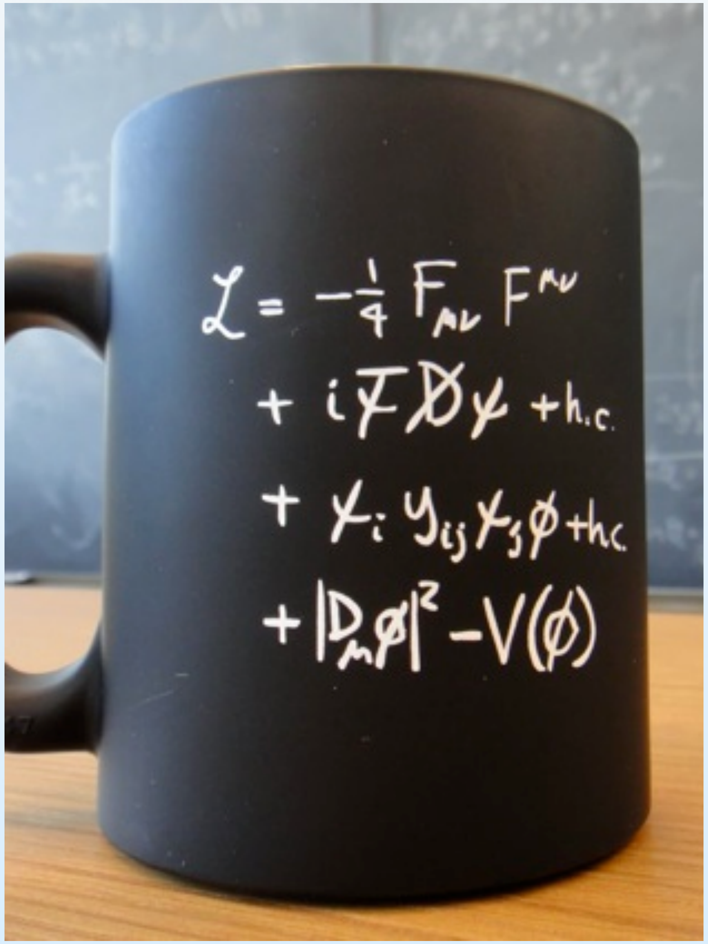


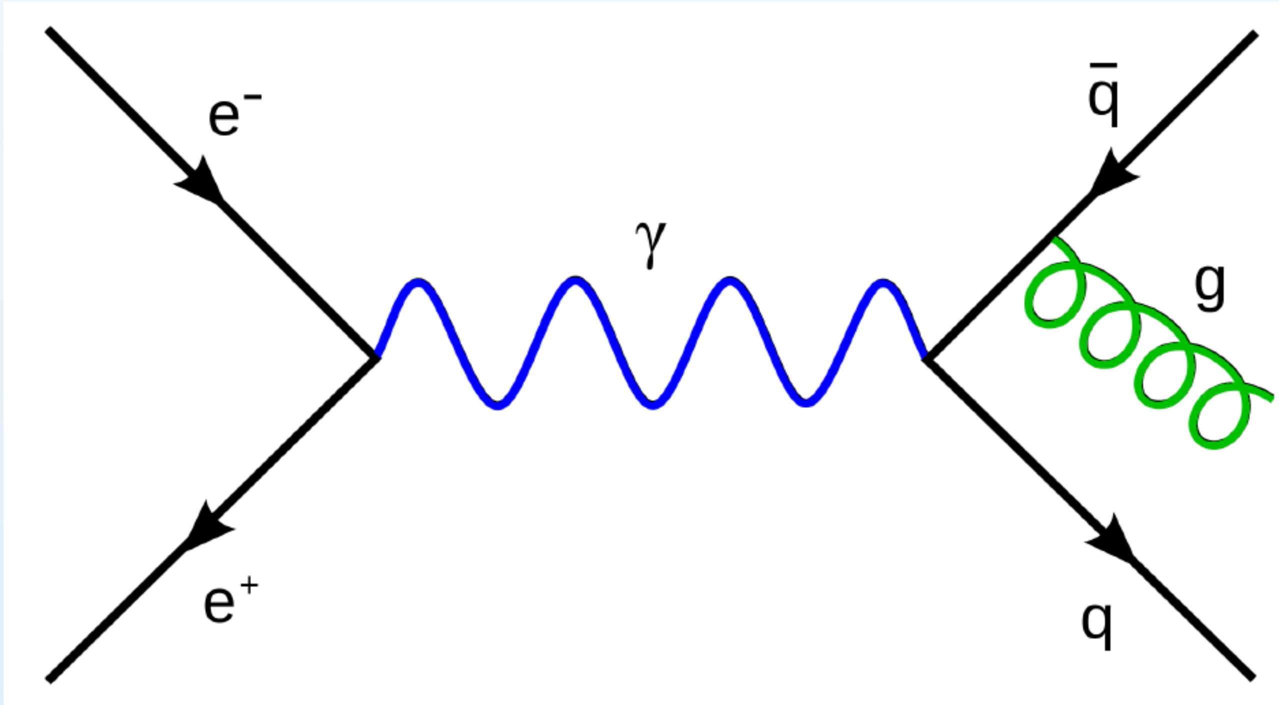
The inner life of protons : Parton Distribution Functions



Lepton vs Hadron Colliders

In high-energy **lepton colliders**, such as the **Large Electron-Positron Collider** (LEP) at CERN, the collisions involve **elementary particles** without substructure

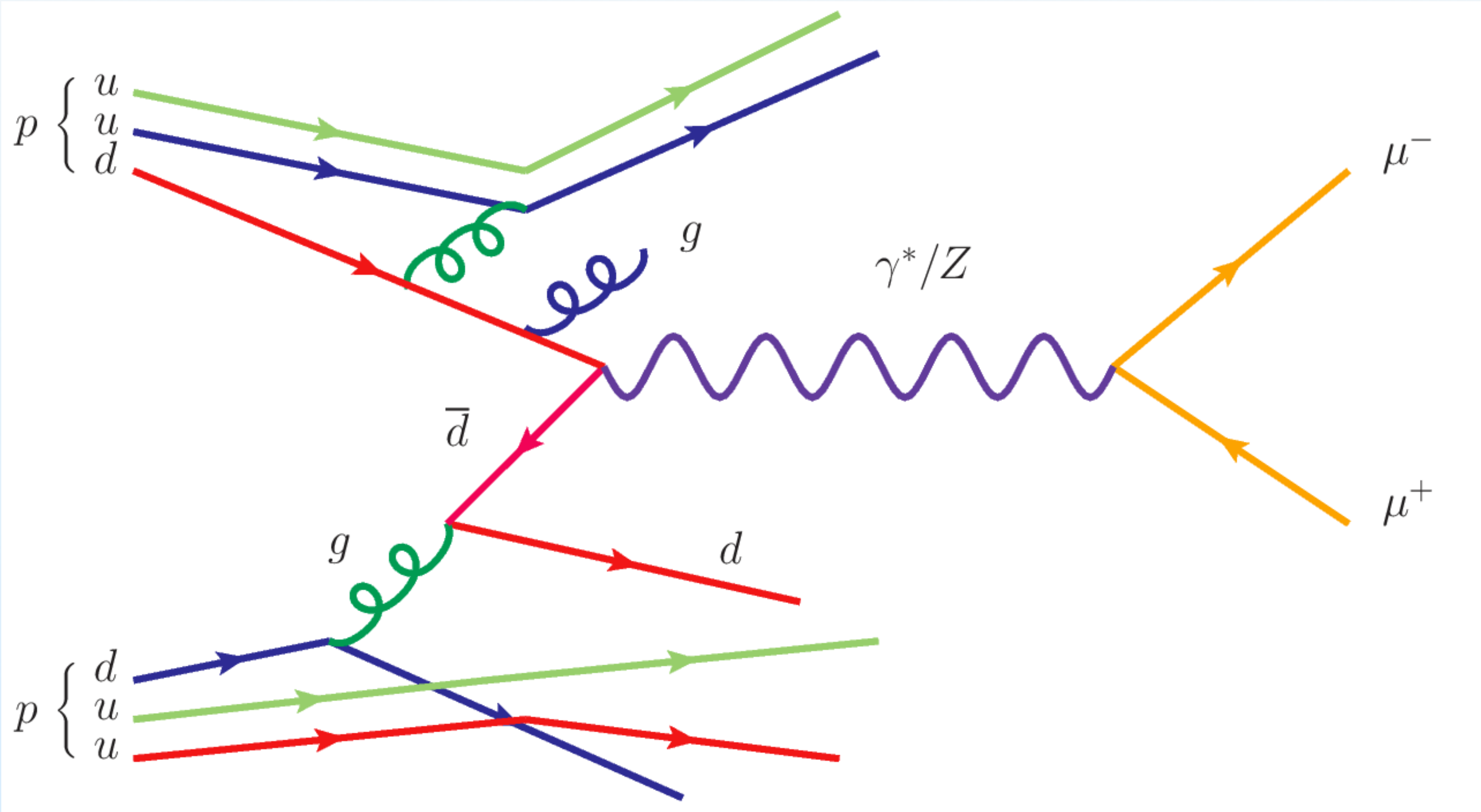




Cross-sections in lepton colliders can be computed in perturbation theory using the **Feynman rules of the Standard Model Lagrangian**

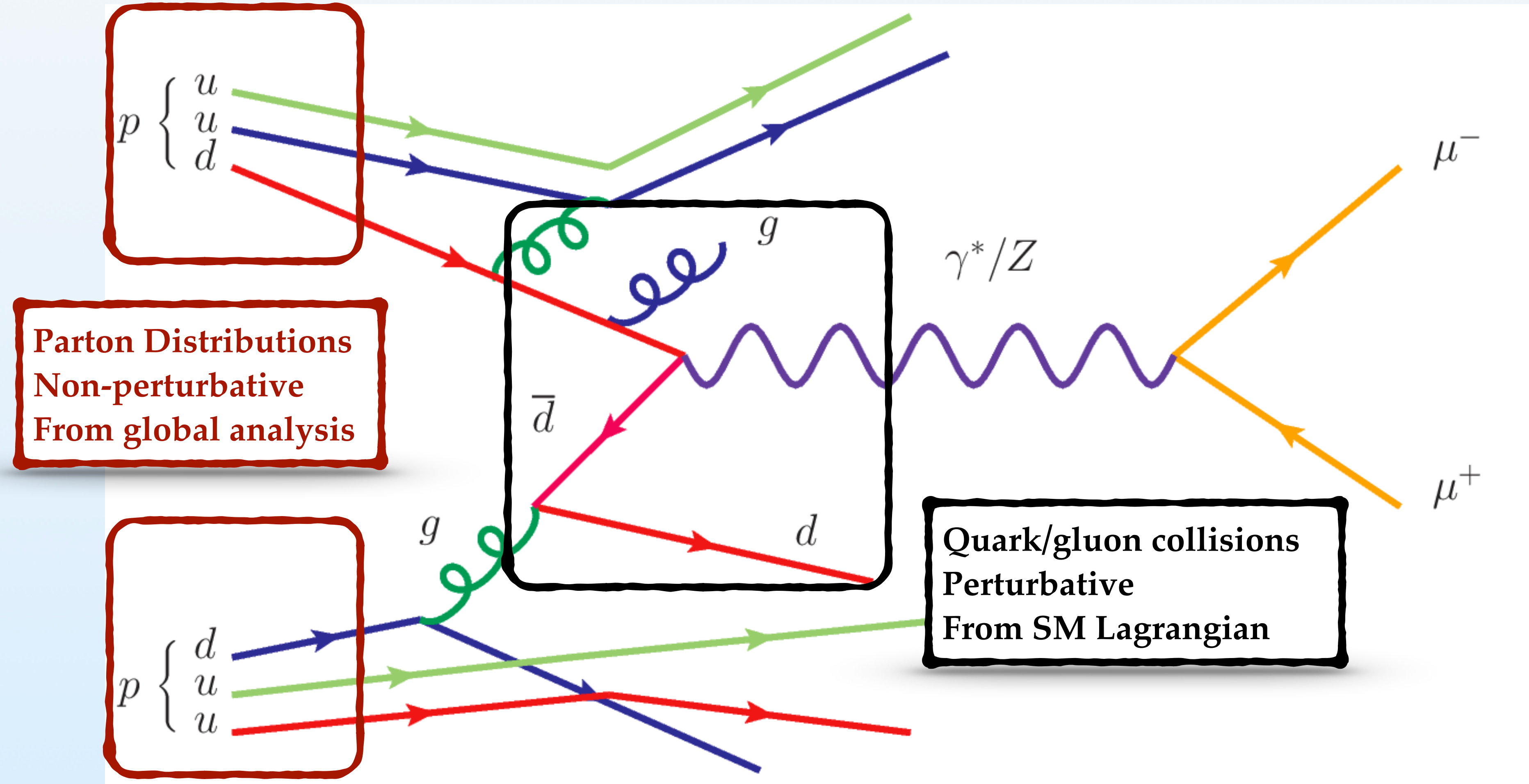
Lepton vs Hadron Colliders

In high-energy **hadron colliders**, such as the LHC, the collisions involve **composite particles** (protons) with internal structure (quarks and gluons)



Lepton vs Hadron Colliders

In high-energy **hadron colliders**, such as the LHC, the collisions involve **composite particles** (protons) with internal structure (quarks and gluons)



Calculations of **cross-sections** in hadron collisions require the combination of **perturbative, quark/gluon-initiated processes**, and **non-perturbative, parton distributions**, information

Parton Distributions

The distribution of energy that **quarks and gluons carry inside the proton** is quantified by the **Parton Distribution Functions (PDFs)**

$$g(x, Q)$$

Q: **Energy of the quark/gluon collision**
Inverse of the resolution length

g(x,Q): **Probability** of **finding a gluon inside a proton**, carrying a fraction ***x*** of the proton momentum, when probed with energy ***Q***

x: Fraction of the proton's momentum

Parton Distributions

The distribution of energy that **quarks and gluons carry inside the proton** is quantified by the **Parton Distribution Functions (PDFs)**

$$g(x,Q)$$

***Q*: Energy of the quark/gluon collision**
Inverse of the resolution length

g(x,Q): **Probability** of **finding a gluon inside a proton**, carrying a fraction ***x*** of the proton momentum, when probed with energy ***Q***

x: Fraction of the proton's momentum

PDFs are determined by **non-perturbative QCD dynamics,** cannot be computed from first principles, and need to be **extracted from experimental data** with a **global analysis**

Parton Distributions

The distribution of energy that **quarks and gluons carry inside the proton** is quantified by the **Parton Distribution Functions (PDFs)**

$$g(x, Q)$$

Q: **Energy of the quark/gluon collision**
Inverse of the resolution length

g(x,Q): **Probability** of **finding a gluon inside a proton**, carrying a fraction ***x*** of the proton momentum, when probed with energy ***Q***

x: **Fraction of the proton's momentum**

PDFs are determined by **non-perturbative QCD dynamics,** cannot be computed from first principles, and need to be **extracted from experimental data** with a **global analysis**

- **Energy conservation** $$\int_0^1 dx \left(g(x,Q) + \sum_q q(x,Q) \right) = 1$$

- **Dependence with quark/gluon collision energy Q** determined in perturbation theory

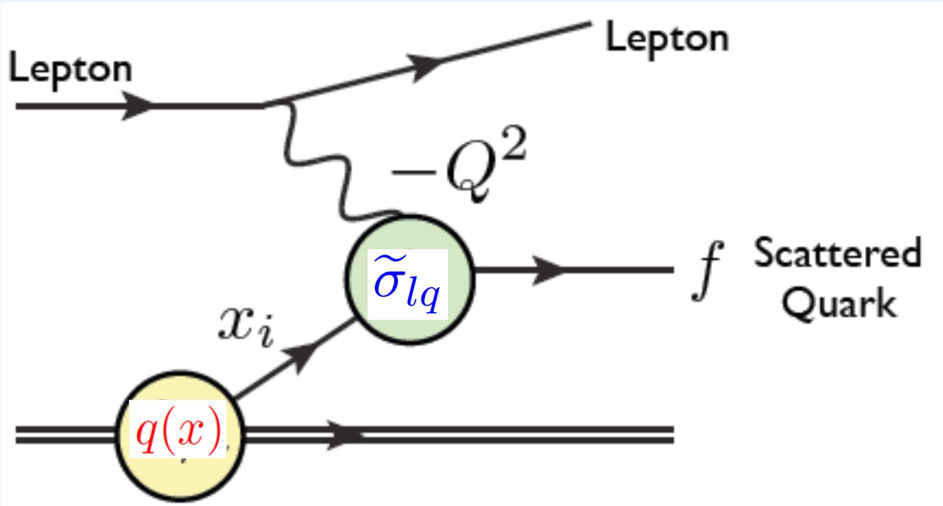
$$\frac{\partial g(x,Q)}{\partial \ln Q} = P_g\left(\alpha_s\right) \otimes g(x,Q) + P_q\left(\alpha_s\right) \otimes q(x,Q)$$

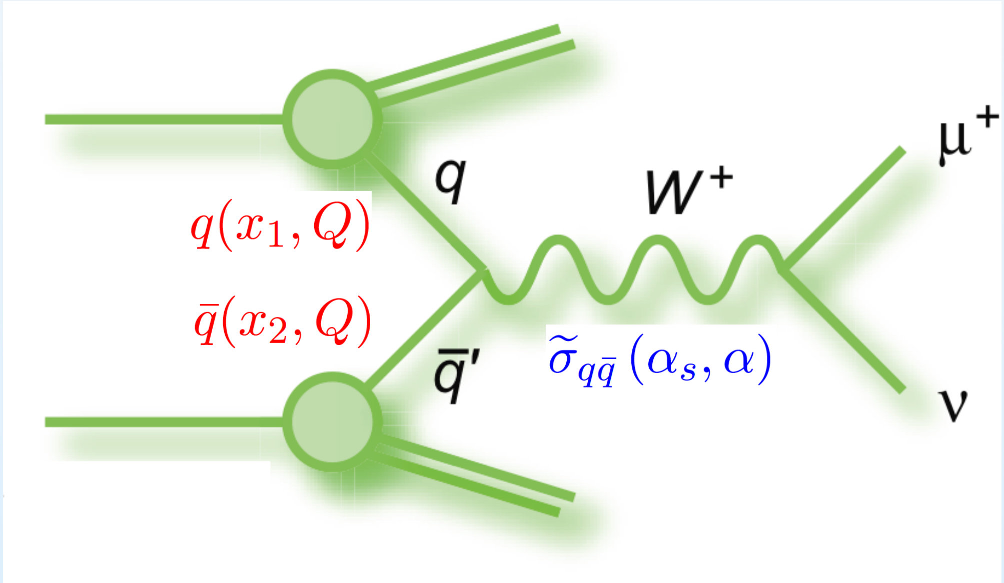
The QCD Factorization Theorem

The **QCD factorization theorem** guarantees **PDF universality: extract them from a subset of** process and use them to provide pure predictions for new processes

$$\sigma_{lp} \simeq \widetilde{\sigma}_{lq}\left(\alpha_s, \alpha\right) \otimes q(x, Q)$$

$$\sigma_{pp} \simeq \widetilde{\sigma}_{q\bar{q}}\left(\alpha_s, \alpha\right) \otimes q(x_1, Q) \otimes \bar{q}(x_2, Q)$$



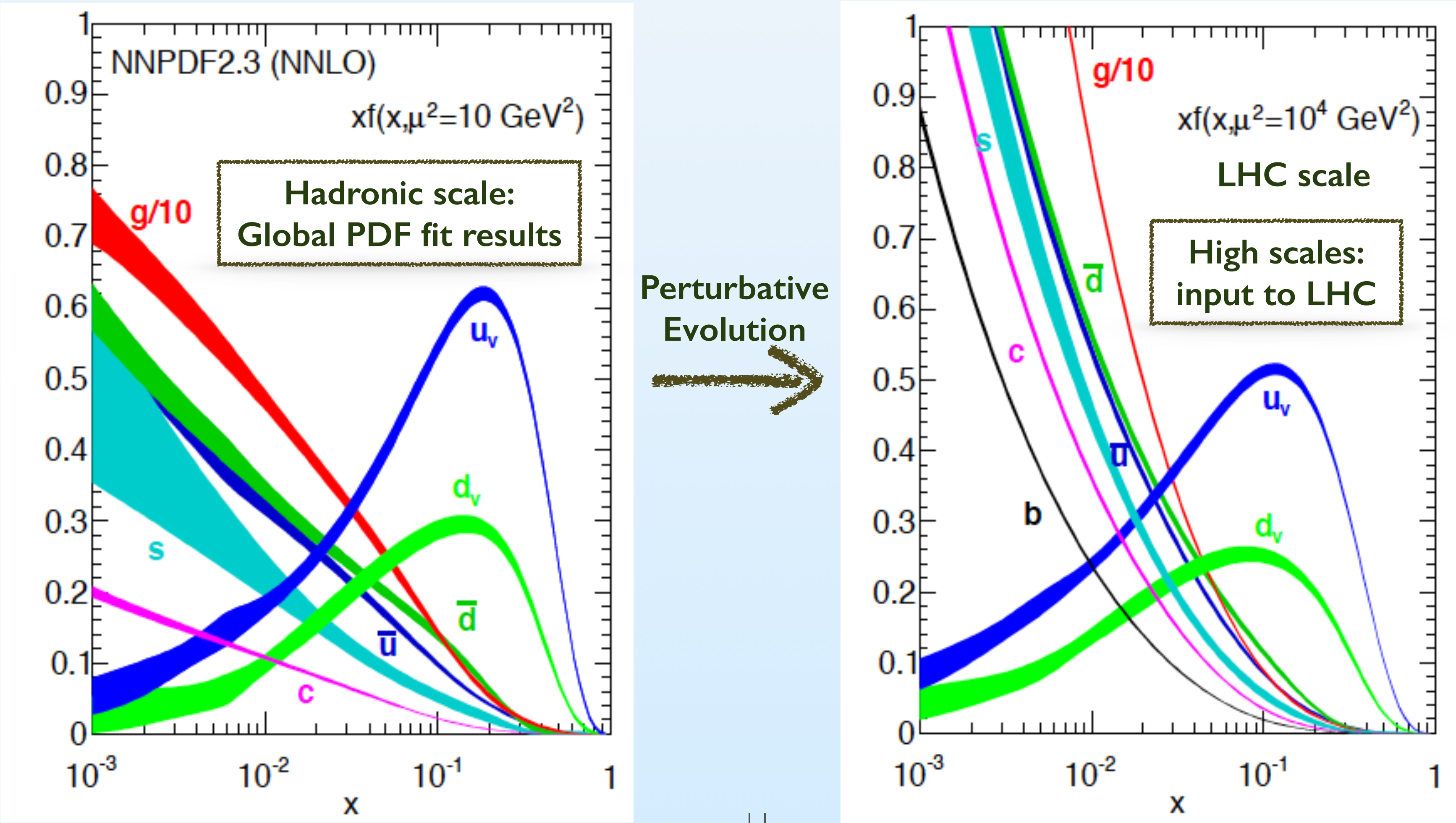


Determine PDFs in lepton-proton collisions ….

And use them to compute cross-sections in proton-proton collisions at the LHC

The global PDF analysis

- Combine **state-of-the-art theory calculations,** the **constraints from PDF-sensitive measurements** from different processes and colliders, and a **statistically robust fitting methodology**
- **Extract Parton Distributions** at hadronic scales of **a few GeV,** where non-perturbative QCD sets in
- Use **perturbative evolution** to compute PDFs at high scales as **input to LHC predictions**



The NNPDF approach

A **novel approach to PDF determination**, improving the limitations of the traditional PDF fitting methods with the use of **advanced statistical techniques** such as **machine learning** and **multivariate analysis**

Non-perturbative PDF parametrization

- **Traditional approach**: based on **restrictive functional forms** leading to strong theoretical bias
- **NNPDF solution**: use **Artificial Neural Networks** as universal unbiased interpolants

PDF uncertainties and propagation to LHC calculations

- **Traditional approach**: limited to Gaussian/linear approximation
- **NNPDF solution**: based on the **Monte Carlo replica method** to create a probability distribution in the space of PDFs. Specially critical in **extrapolation regions** (**i.e.** high-x) for New Physics searches

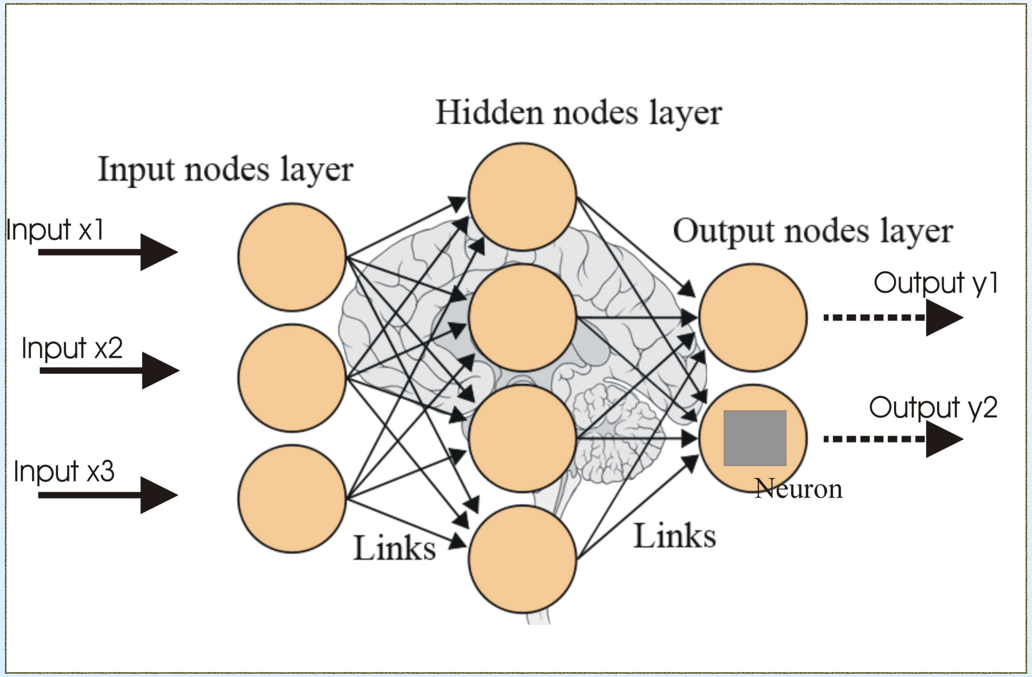
Fitting technique

- **Traditional approach**: deterministic minimization of χ^2, **flat directions** problem
- **NNPDF solution**: **Genetic Algorithms** to explore efficiently the vast parameter space, with **cross-validation** to avoid fitting stat fluctuations

Artificial Neural Networks

Inspired by **biological brain models**, **Artificial Neural Networks** are **mathematical algorithms** widely used in a wide range of applications, from **high energy physics** to **targeted marketing** and **finance forecasting**

From Biological to Artificial Neural Networks



Artificial neural networks aim to excel in the same domains as their biological counterparts: **pattern recognition, forecasting, classification**, ... where our **evolution-driven biology** outperforms traditional algorithms

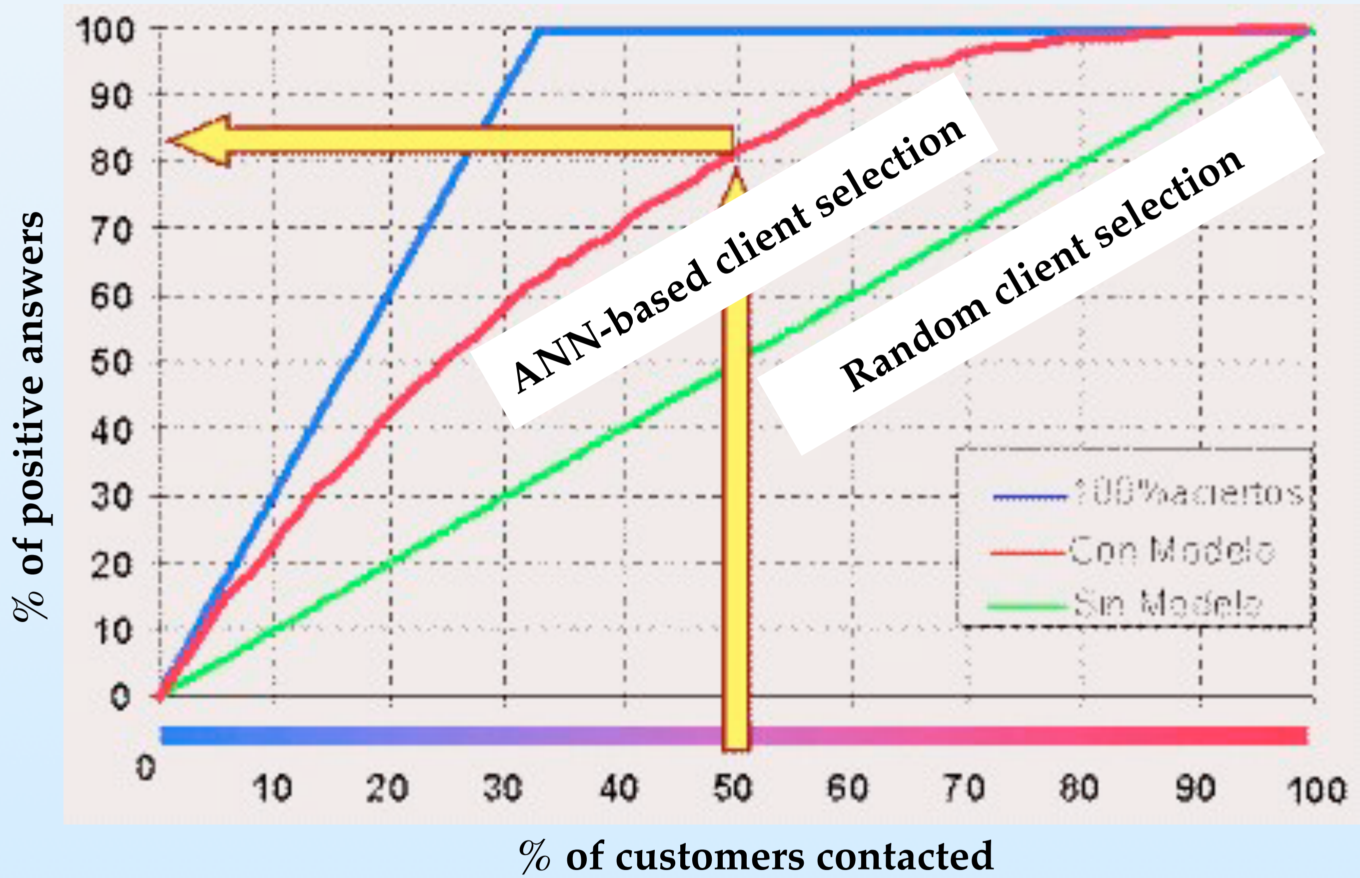
Artificial Neural Networks

Application in **Marketing**

A bank wants to offer a new credit card to their clients. Two possible strategies:

- **Contact all customers**: slow and costly
- Contact **5%** of the customers, **train a ANN with their input** (gender, income, loans) and **their output** (yes/no) and use the information to **contact only clients likely to accept the product**

Cost-effective method to improve marketing performance



ANN in high-energy physics

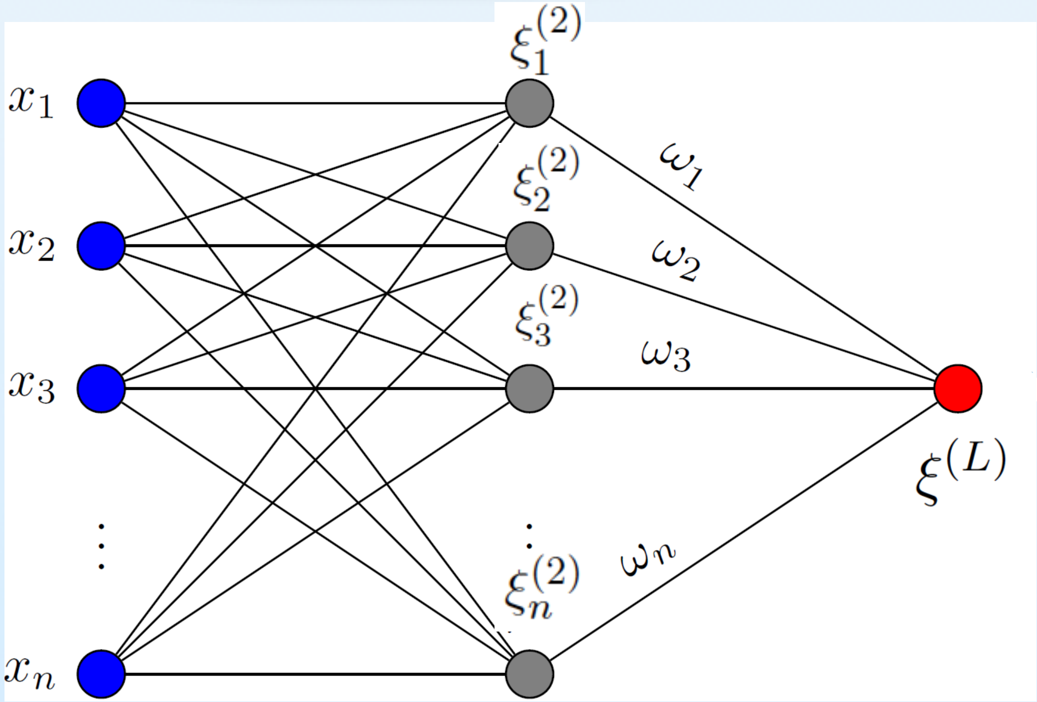
- ANNs are routinely exploited in **high-energy physics**, in most cases as **classifiers** to separate between interesting and more mundane events
- **ANNs also** provide **universal unbiased interpolants** to parametrize the non-perturbative dynamics that determines the **size and shape of the PDFs** from experimental data

Traditional approach

$$g(x,Q_0) = A_g(1-x)^{a_g}x^{-b_g}\left(1+c_g\sqrt{s}+d_g x+\ldots\right)$$

NNPDF approach

$$g(x,Q_0) = A_g \mathrm{ANN}_g(x)$$



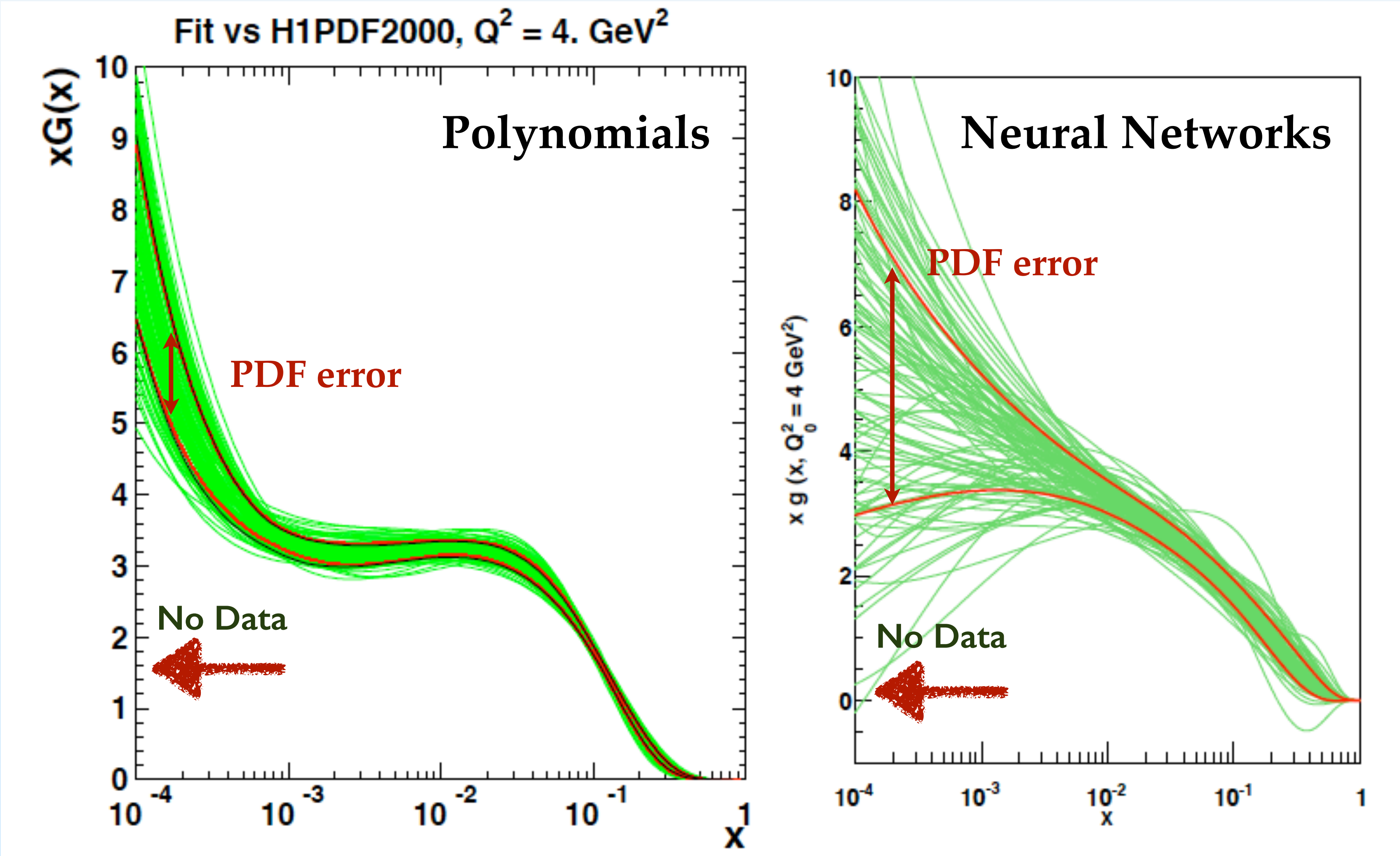
$$\mathrm{ANN}_g(x) = \xi^{(L)} = \mathcal{F}\left[\xi^{(1)},\{\omega_{ij}^{(l)}\},\{\theta_i^{(l)}\}\right]$$

$$\xi_i^{(l)} = g\left(\sum_{j=1}^{n_{l-1}}\omega_{ij}^{(l-1)}\xi_j^{(l-1)} - \theta_i^{(l)}\right)$$

- ANNs eliminate **theory bias** introduced in PDF fits from choice of *ad-hoc* functional forms
- NNPDF fits used **O(400) free parameters**, to be compared with **O(10-20) in traditional fits**. Results stable if **O(4000) parameters used!**
- **Faithful extrapolation**: PDF uncertainties **blow up** in regions with scarce experimental data

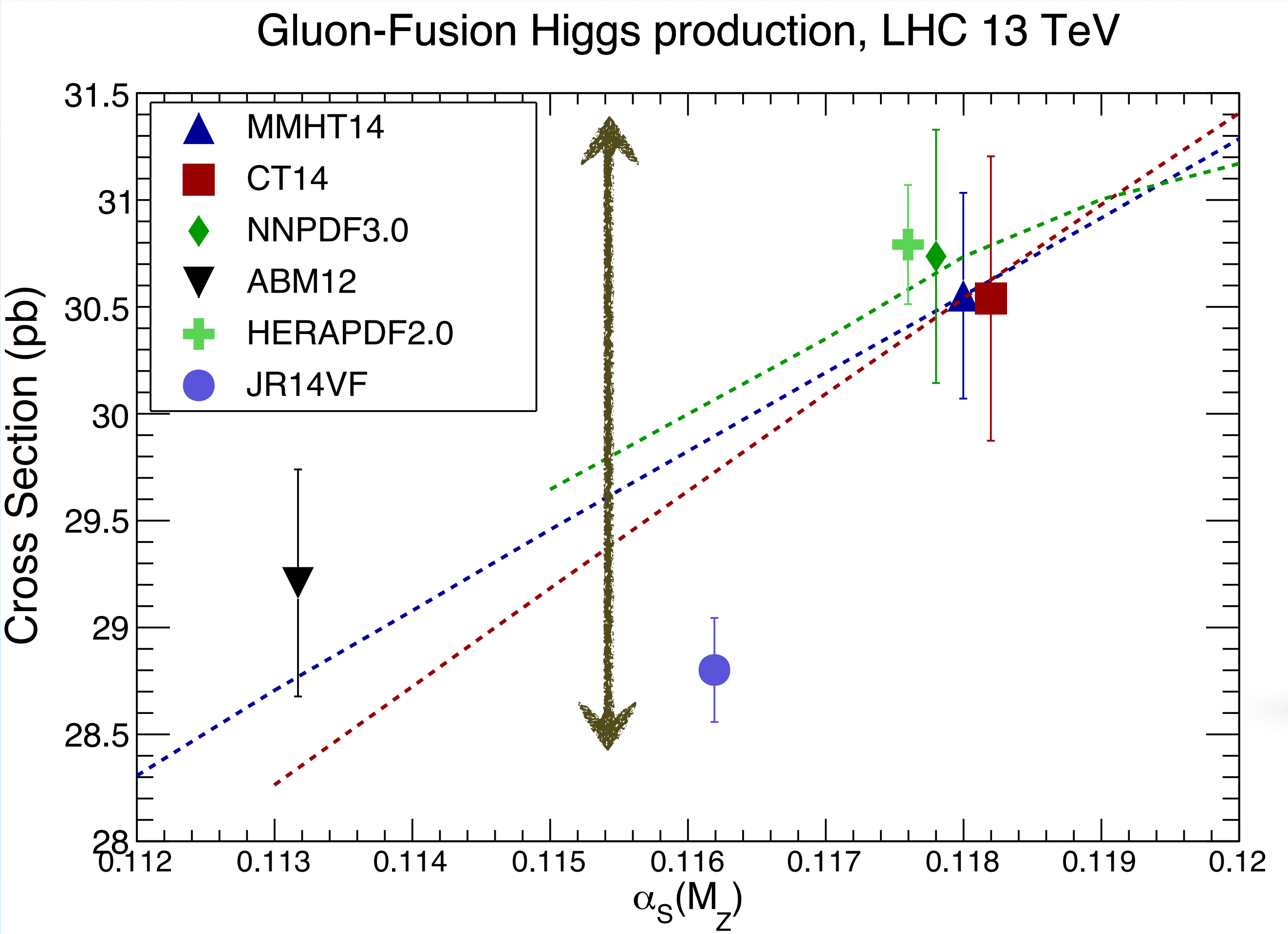
Artificial Neural Networks vs Polynomials

- Compare a **benchmark PDF analysis** where **the same dataset** is fitted with **Artificial Neural Networks** and with **standard polynomials,** other settings identical)
- **ANNs** avoid biasing the PDFs, **faithful extrapolation at small-x** (very few data, thus error blow up)

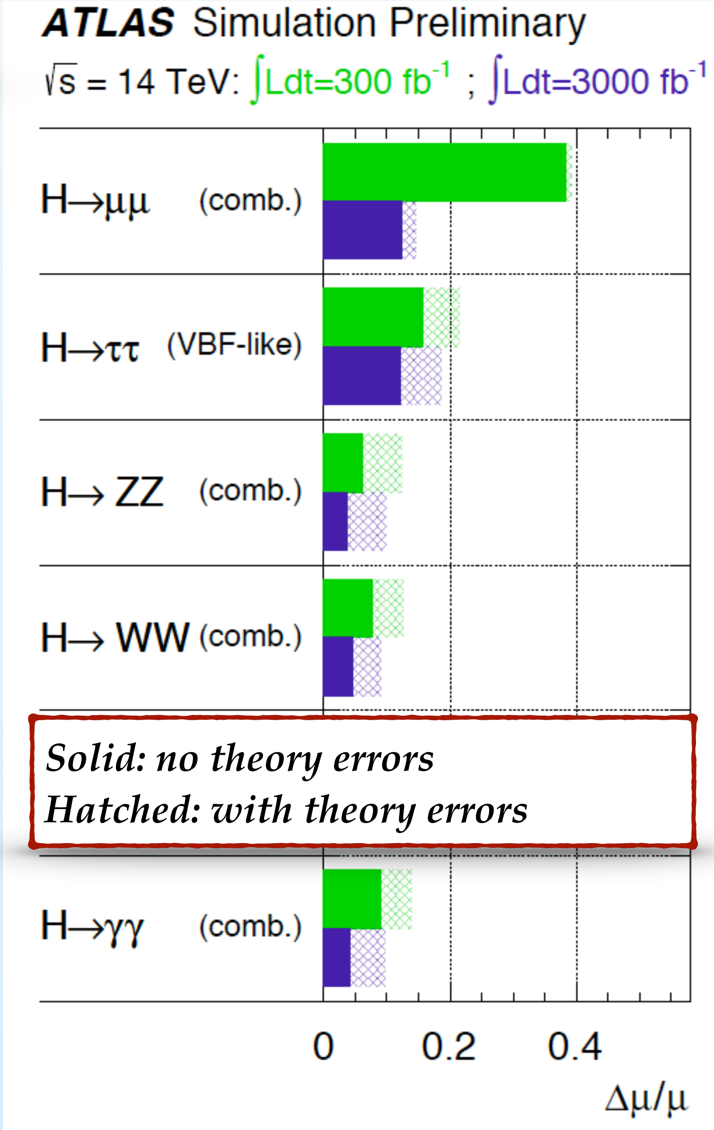


Fingerprinting the Higgs sector

Uncertainties from Parton Distributions *are one of the limiting factors of theory predictions of Higgs production,* ***degrading the exploration of the Higgs sector***



Higgs Cross-Section Working Group YR4 16

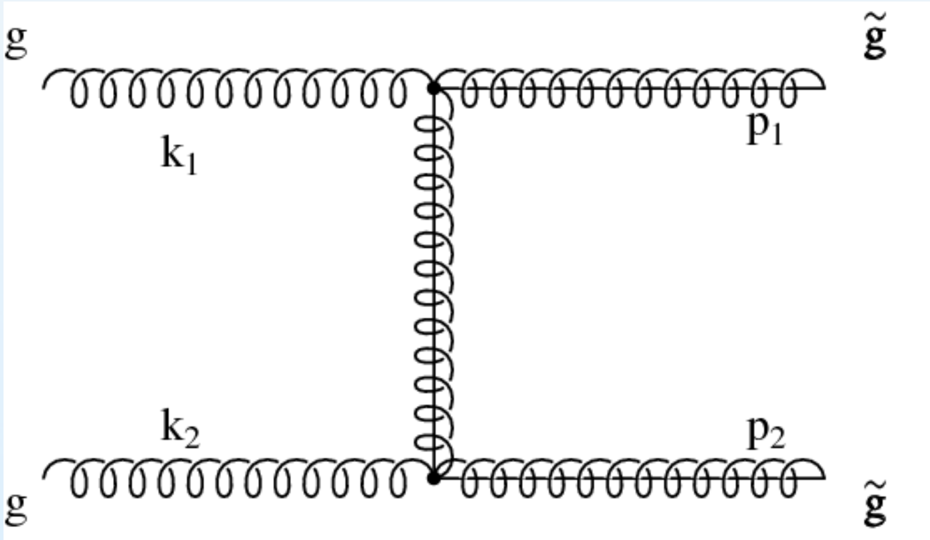


Solid: no theory errors
Hatched: with theory errors

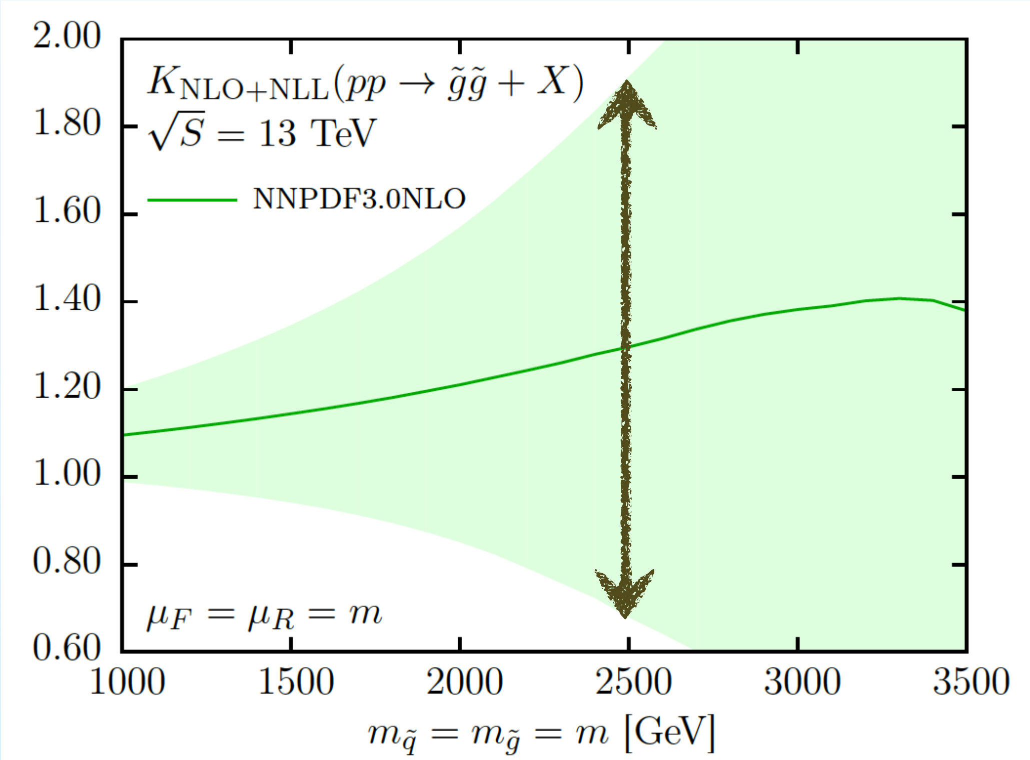
Characterising Heavy New Physics

PDF uncertainties in the production of **New Physics heavy resonances** can be al large as **100%**

The reason is that massive particle production probes **large-x PDFs**, which are poorly known due to very limited direct experimental constraints



Gluino pair production at the LHC

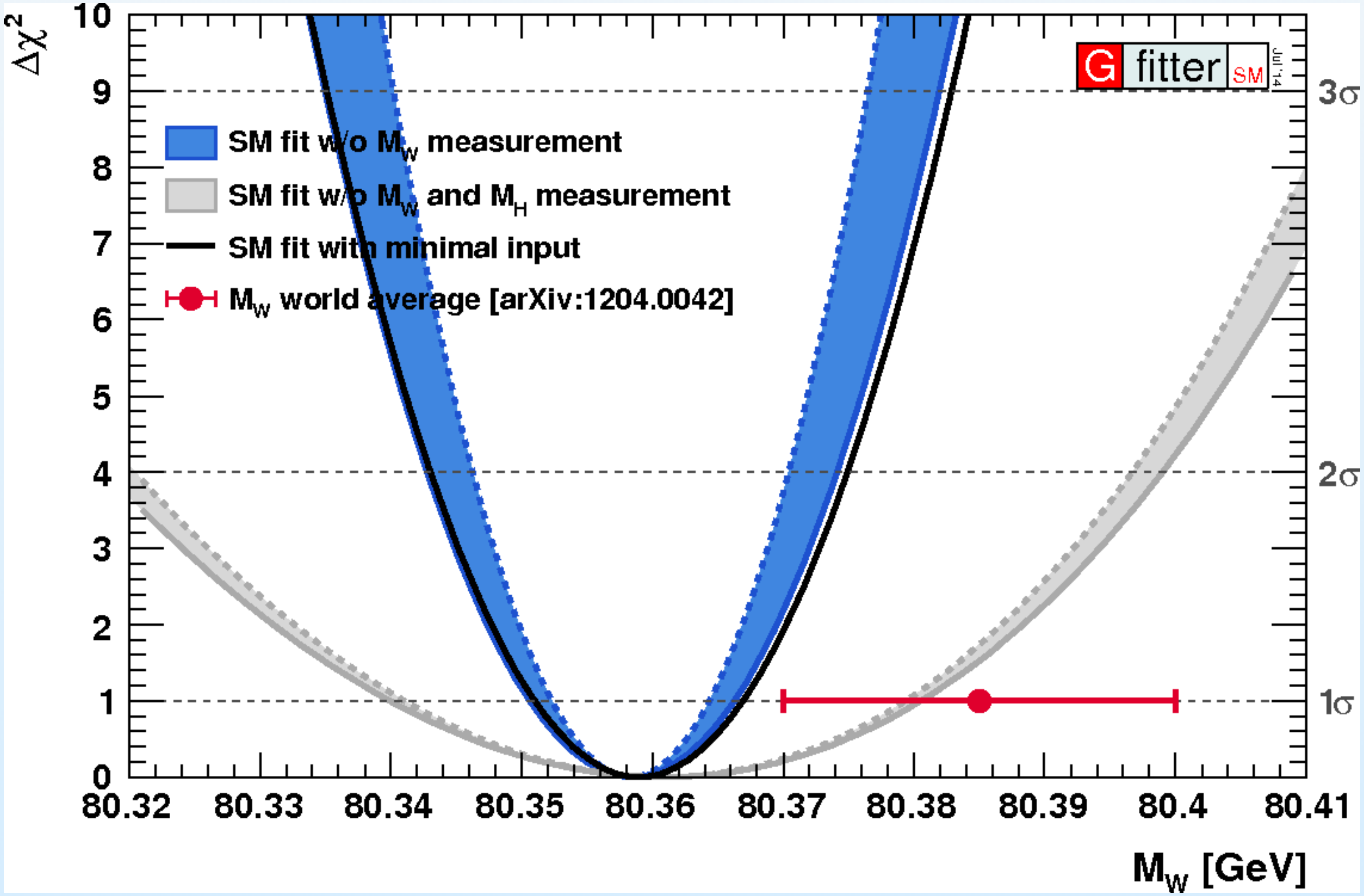


Beenakker, Borchensky, Kramer, Kulesza, Laenen, Marzani, Rojo 15

*Unless we **improve PDF uncertainties**, even if we discover New Physics, it will be extremely **difficult to characterise the underlying dynamics***

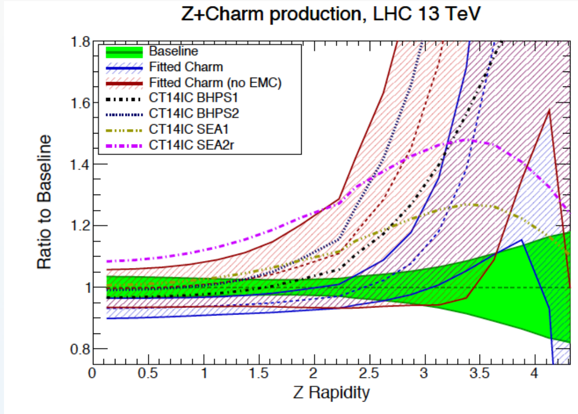
PDFs and precision determination of SM parameters

PDFs dominant systematic for precision measurements, like W boson mass, that provide stringent consistency stress-tests of the Standard Model

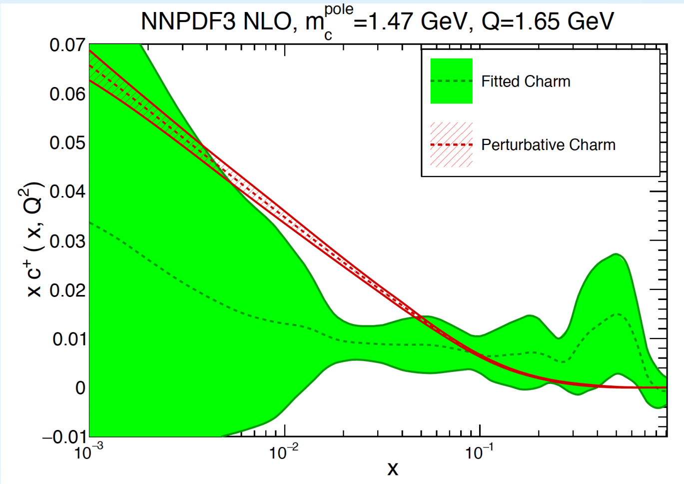


GFitter collaboration 16

- Reducing TH systematics could lead to **indirect BSM discovery** from precision measurements
- Precision in M_W will **improve by a factor 3** in 10 years
- If SM confirmed, **ruled out a broad class of BSM scenarios**
- First measurements from ATLAS and CMS to be presented soon!



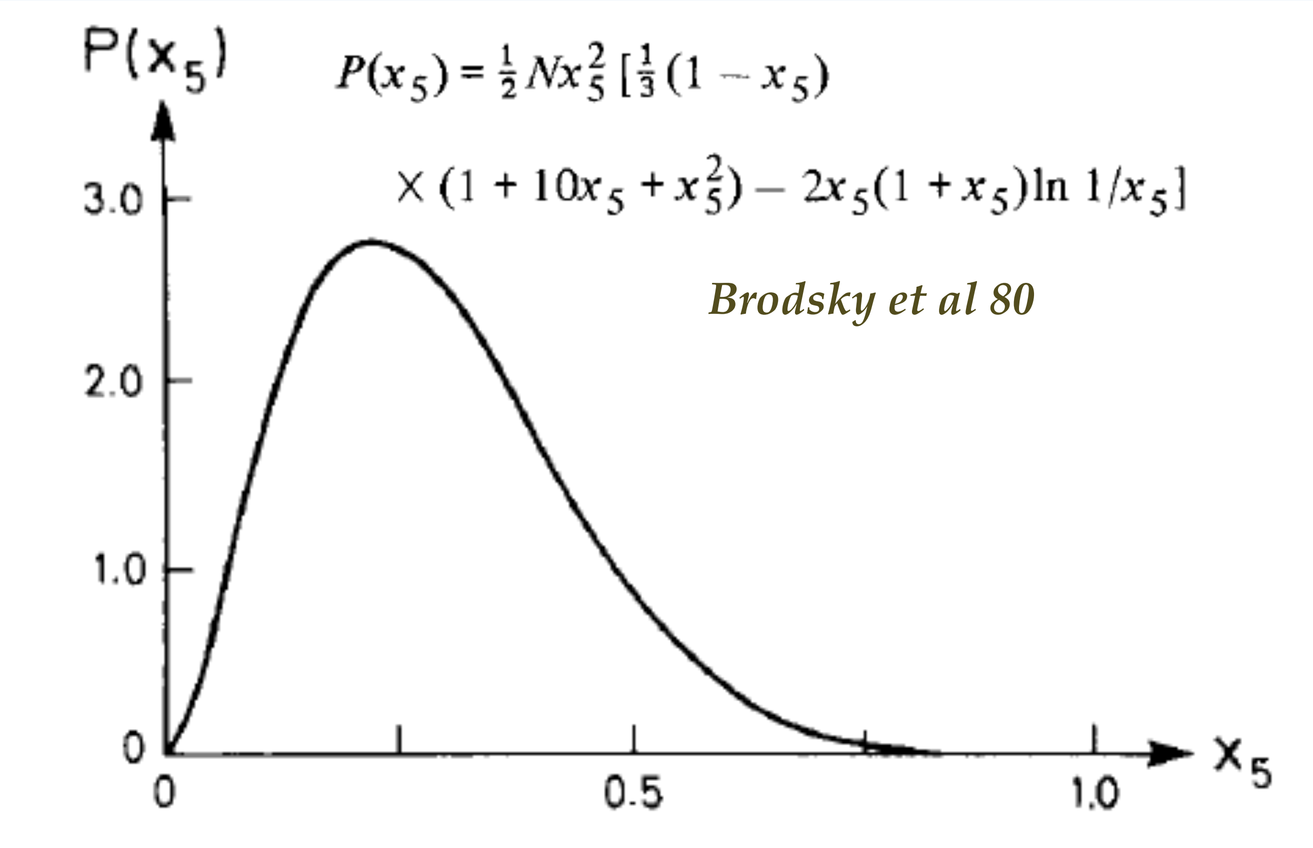
A first unbiased determination of the charm content of the proton



NNPDF Collaboration 16

The charm content of the proton

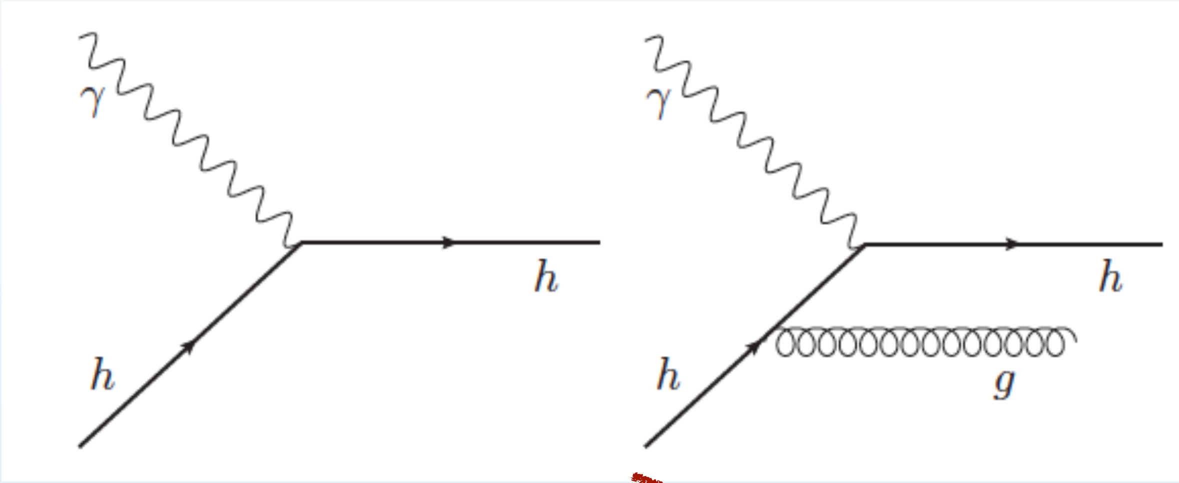
- The motivation **to fit a charm PDF** in a global analysis is **two-fold:**
  - ☑ Stabilise the dependence of **LHC calculations** with respect to **value of the charm mass**
  - ☑ Quantify the **non-perturbative charm component in the proton** and compare with models



A 30-years old conundrum of QCD!

Global QCD analysis with fitted charm

General-mass VFN scheme **heavy quark deep-inelastic structure functions** are modified to account for **massive charm-initiated diagrams**



$$F(x,Q^2) = F^{\mathrm{FLNR}}(x,Q^2) + \Delta F(x,Q^2)$$

Forte, Laenen, Nason, Rojo 10

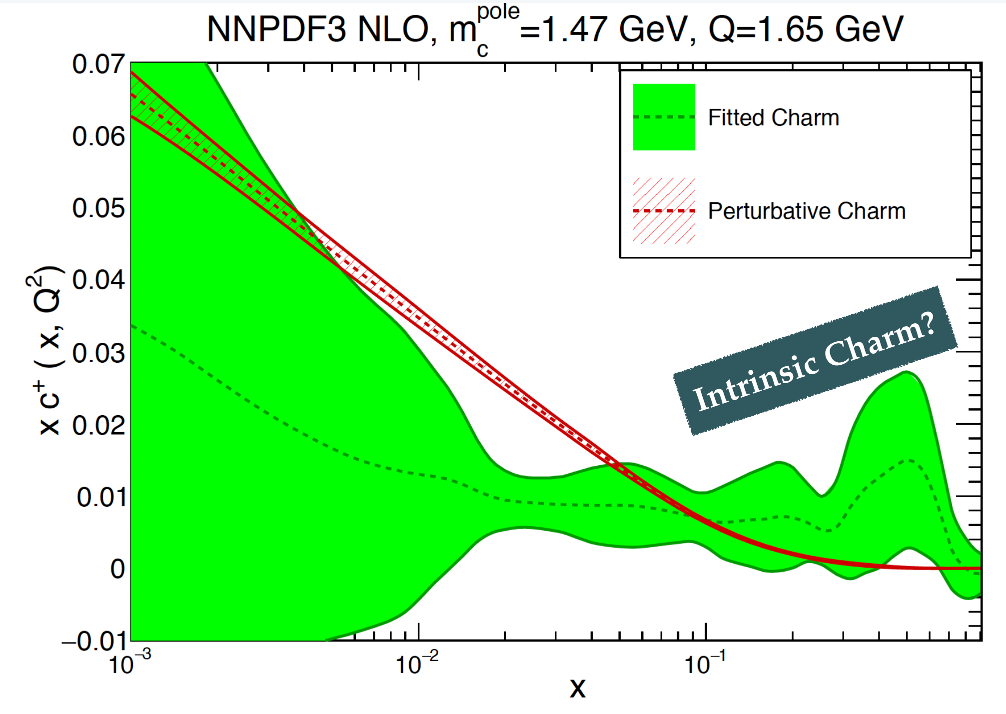
Ball, Bertone, Bonvini, Forte, Groth-Merrild, Rojo and Rottoli 15

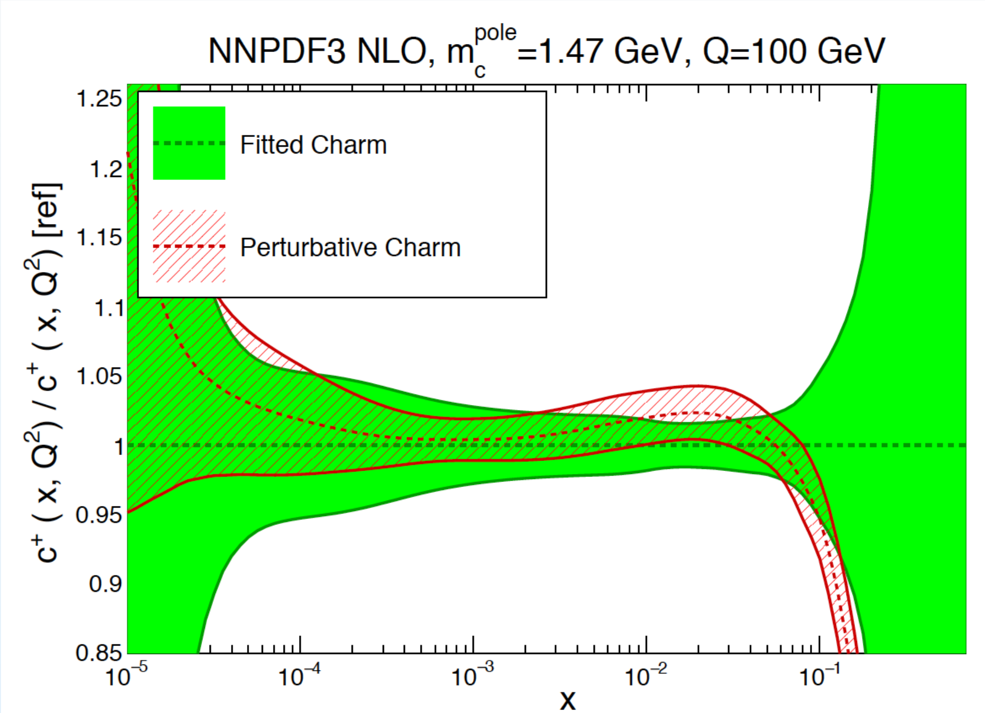
Parametrize the charm PDF $c(x,Q_0)$ on an equal footing to light quarks: **ANN with 37 params**

Fit quality improvement when **charm is fitted** as compared to the traditional **dynamical charm:** χ^2/N_{dat} from **1.18 to 1.15**

NNPDF Collaboration 16

Global QCD analysis with fitted charm





At low scales, hints of a **non-perturbative charm component**, though PDF errors still large

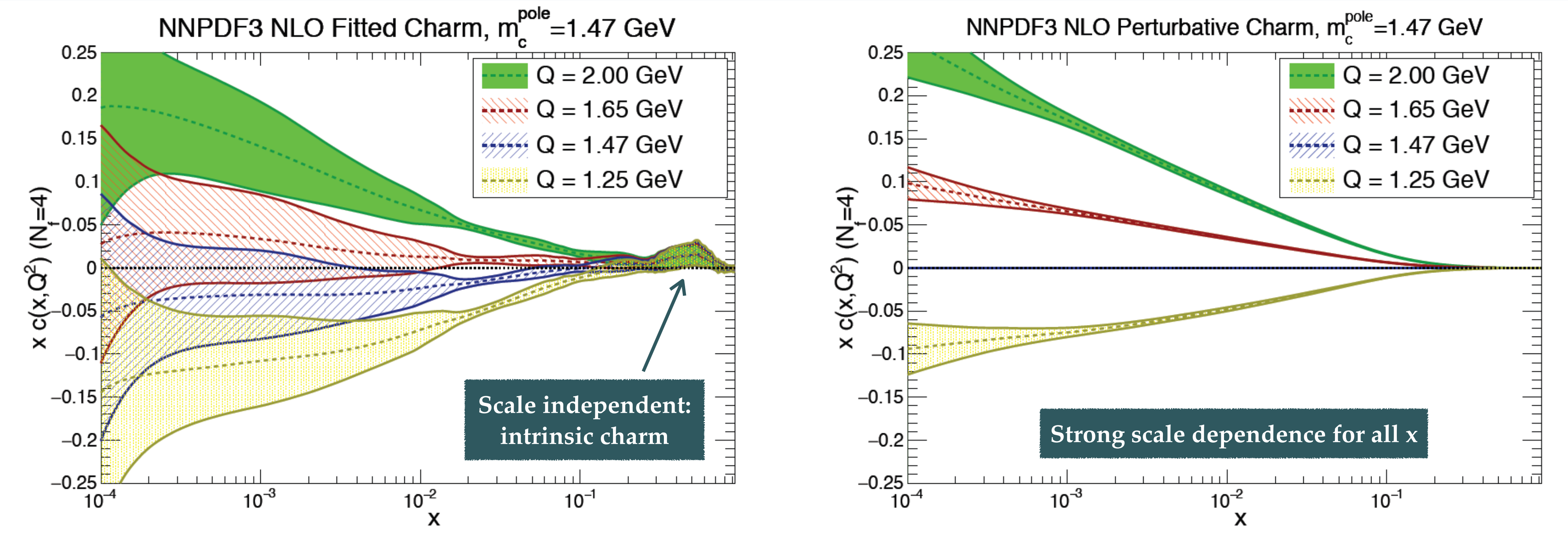
At LHC scales, **fitted and perturbative charm in good agreement** for x < 0.08

Charm can account up to **1% of the total proton momentum** at low scales

Charm momentum fraction

PDF set	$C(Q = 1.65\ \mathrm{GeV})$
NNPDF3 perturbative charm	$(0.239 \pm 0.003)\%$
NNPDF3 fitted charm	$(0.7 \pm 0.3)\%$
NNPDF3 fitted charm (no EMC)	$(1.6 \pm 1.2)\%$

Disentangling perturbative and intrinsic charm



The fitted charm PDF is, by construction, the **sum of the intrinsic and perturbative components**

These can be disentangled from the **scale dependence** of the fitted charm PDF

Intrinsic charm is scale independent, while perturbative charm has strong scale dependence

The perturbative component vanishes for $Q \simeq$ **m_{charm}** $\simeq$ **1.5 GeV** as expected since for **Q** $\lesssim$ **m_{charm}** the perturbative description of the charm PDF breaks down

Fitted and perturbative charm very similar for **x** $\lesssim$ **0.1**

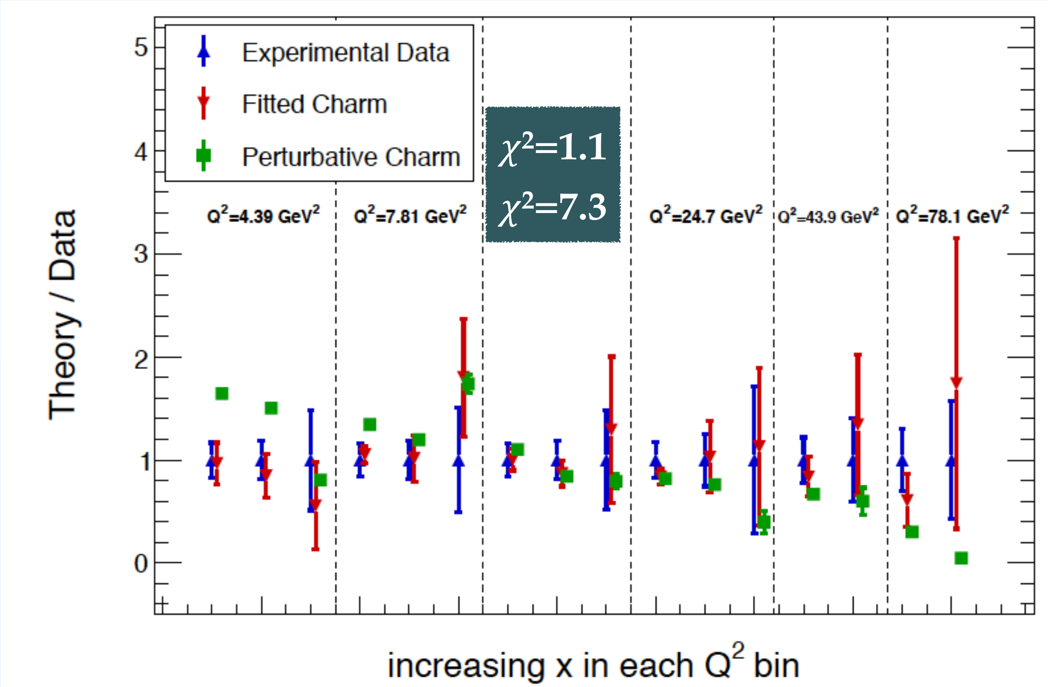
The EMC charm mystery solved?

The **EMC measurements of charm production** in the early 80s have since then been advocated as *smoking gun* for intrinsic charm, but **including them in a global PDF fit** had been impossible

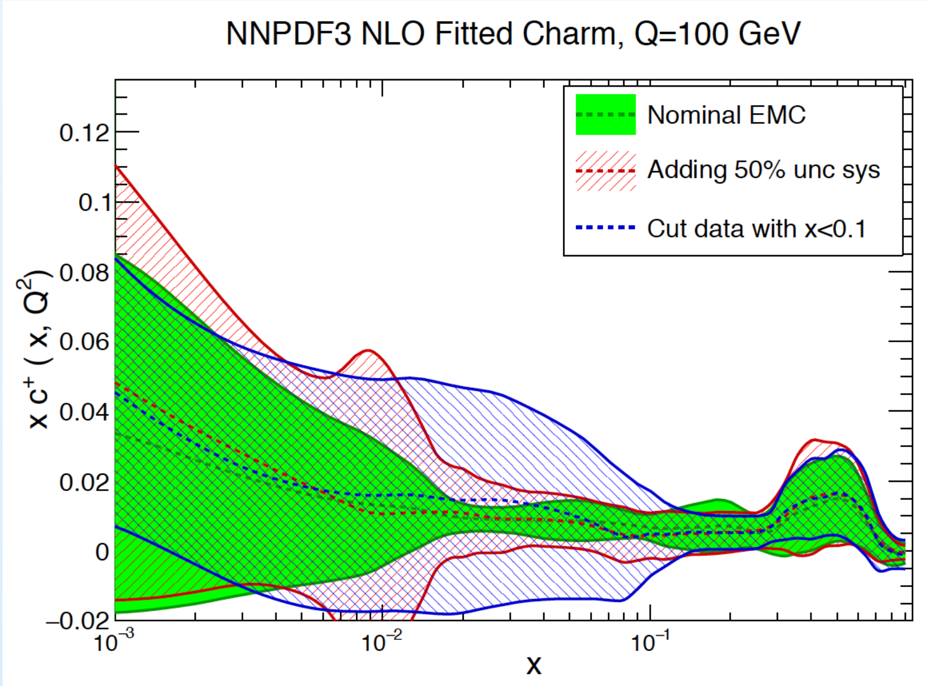
This NNPDF analysis with fitted charm includes EMC data, leading to an **excellent fit quality**

Our results stable upon **variations of the EMC data** (such as cutting data for $x < 0.1$)

Theory/data ratio for EMC charm production



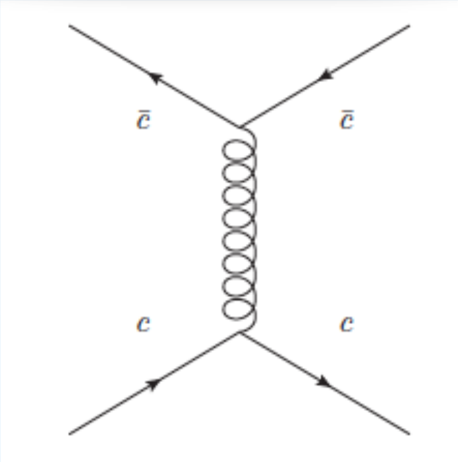
Fitted charm stability

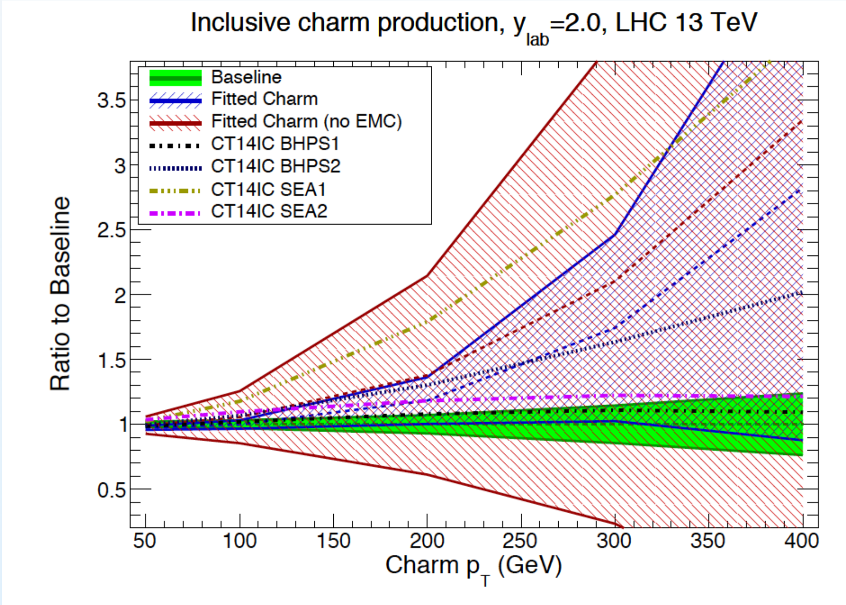


intrinsic charm: implications for the LHC

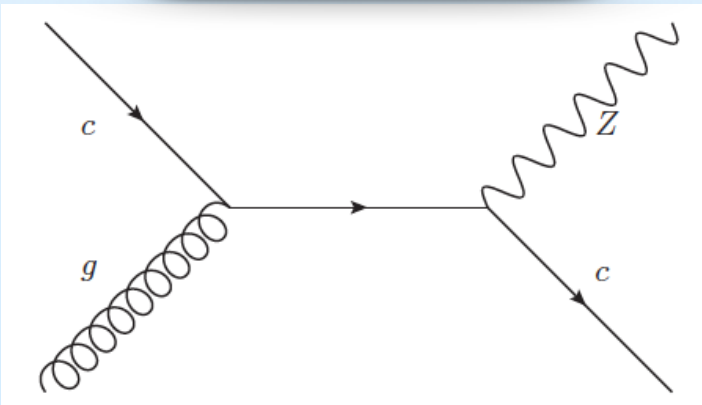
- A number of LHC processes are sensitive to the **charm content of the proton**
- To probe **large-x charm** we need either **large p_T** or **forward rapidities** production
- Within the **reach of the LHC** at Run II+III

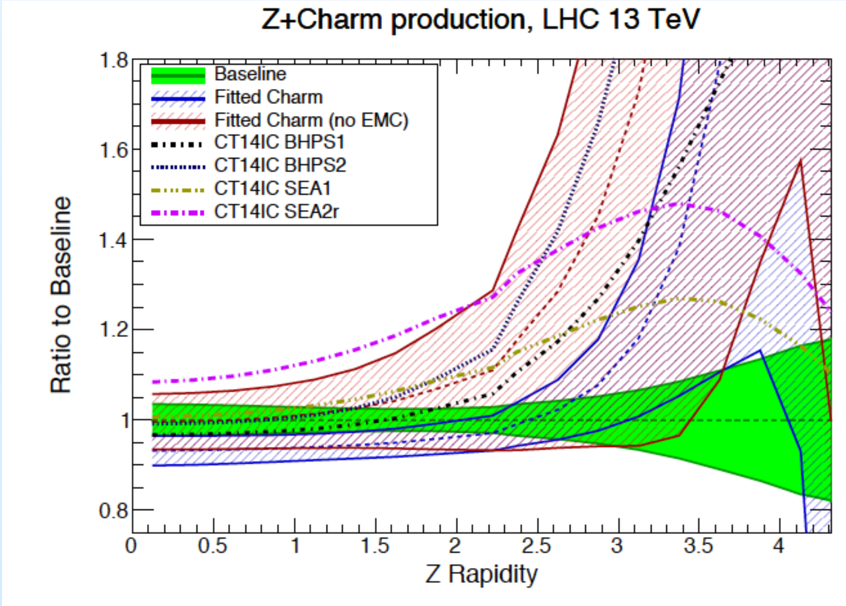
D meson production

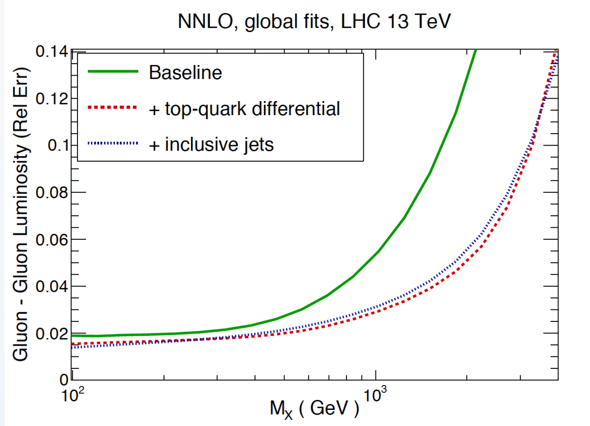




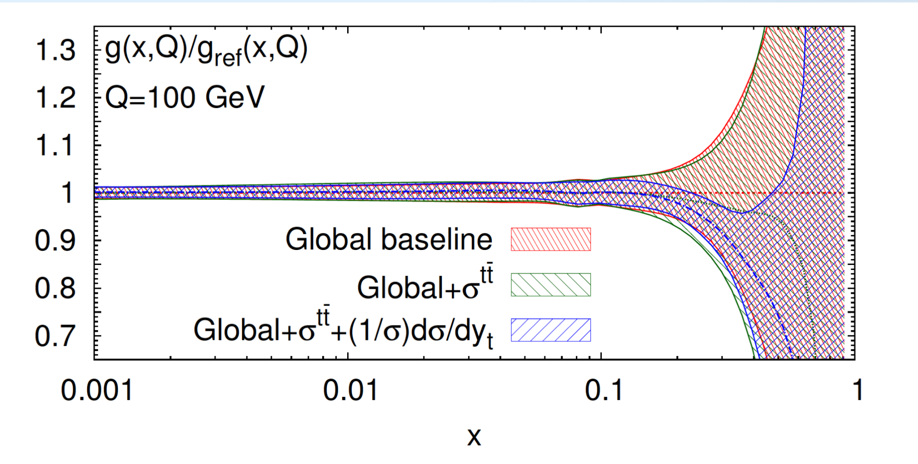
Z+charm production







Pinning down the large-x gluon with NNLO top differential distributions



Czakon, Hartland, Mitov, Nocera, Rojo 16

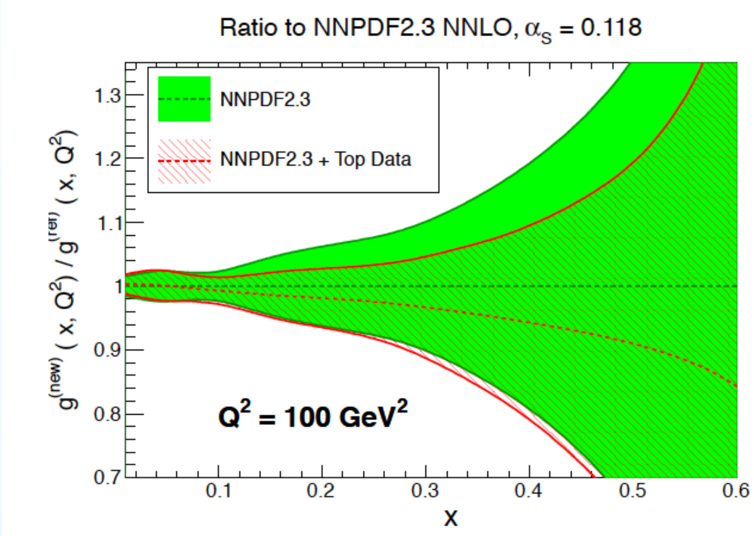
Why top-pair differential data in PDF fits?

- **Total cross-sections already used** in some PDF fits, using **differential measurements** will only increase constraining power

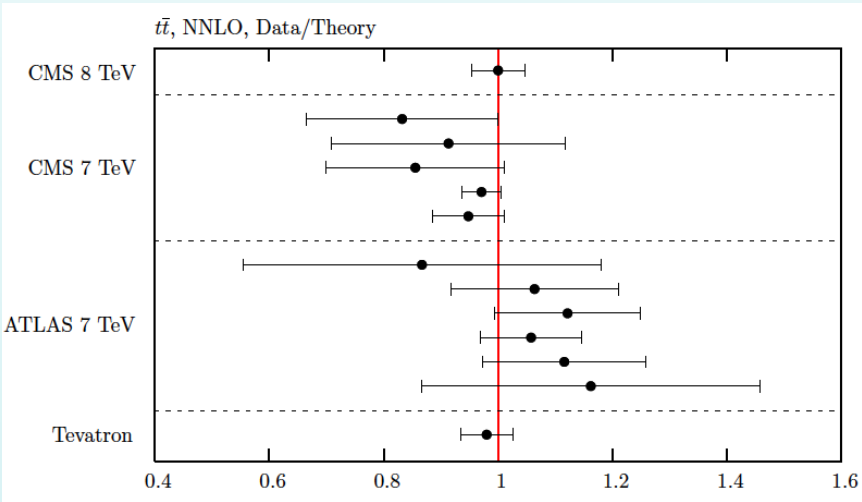
- **NNLO calculation for stable top quarks** available, together with **scale optimization,** allows for the use of state-of-the art theory for top data in global PDF fits

- Available p**recise data from ATLAS and CMS at 8 TeV** with full breakdown of statistical and systematic uncertainties

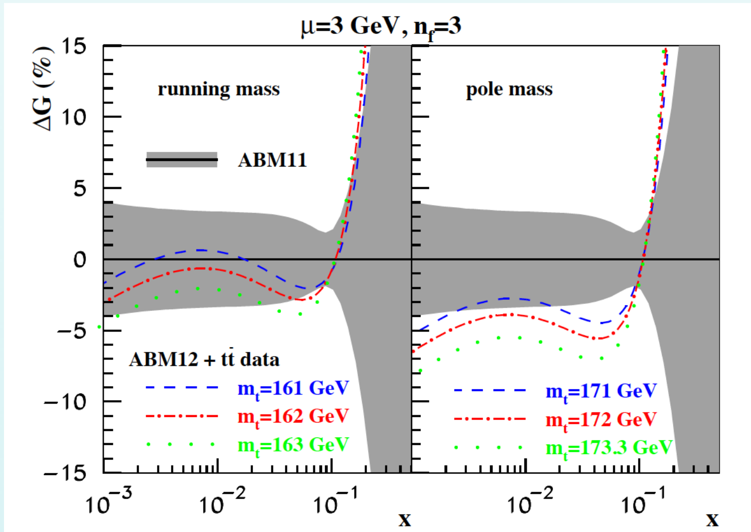
- Study **interplay with inclusive jet production** measurements (NNLO calculation very recently completed)



Czakon, Mangano, Mitov, Rojo 13



MMHT14



ABM12

PDF fit settings

NNPDF Collaboration 14-16

- Based on the **NNPDF3 global analysis framework**
- Baseline fit based on the **same dataset as in NNPDF3.0**, except for:
  - ☑ Separate H1 and ZEUS data from HERA-II replaced by **final HERA combination**
  - ☑ **Jet data excluded** (since NNLO for inclusive jets only very recently available) *Currie, Glover, Pires 16*
- **NNLO theory,** with $\alpha_S(m_Z)$ =0.118, $\mathbf{m_{top}}$**=173.3 GeV**, m_{charm} =1.51 GeV, m_{bottom}=4.92 GeV
- Optimized choice of **factorization and renormalization scales**

Czakon,Heines, Mitov 16

For y_t, y_{tt}, m_{tt}

$$\mu_R = \mu_F = \mu = H_T/4\,, \qquad H_T \equiv \sqrt{m_t^2 + \left(p_T^t\right)^2} + \sqrt{m_t^2 + \left(p_T^{\bar{t}}\right)^2}$$

For p^t_T

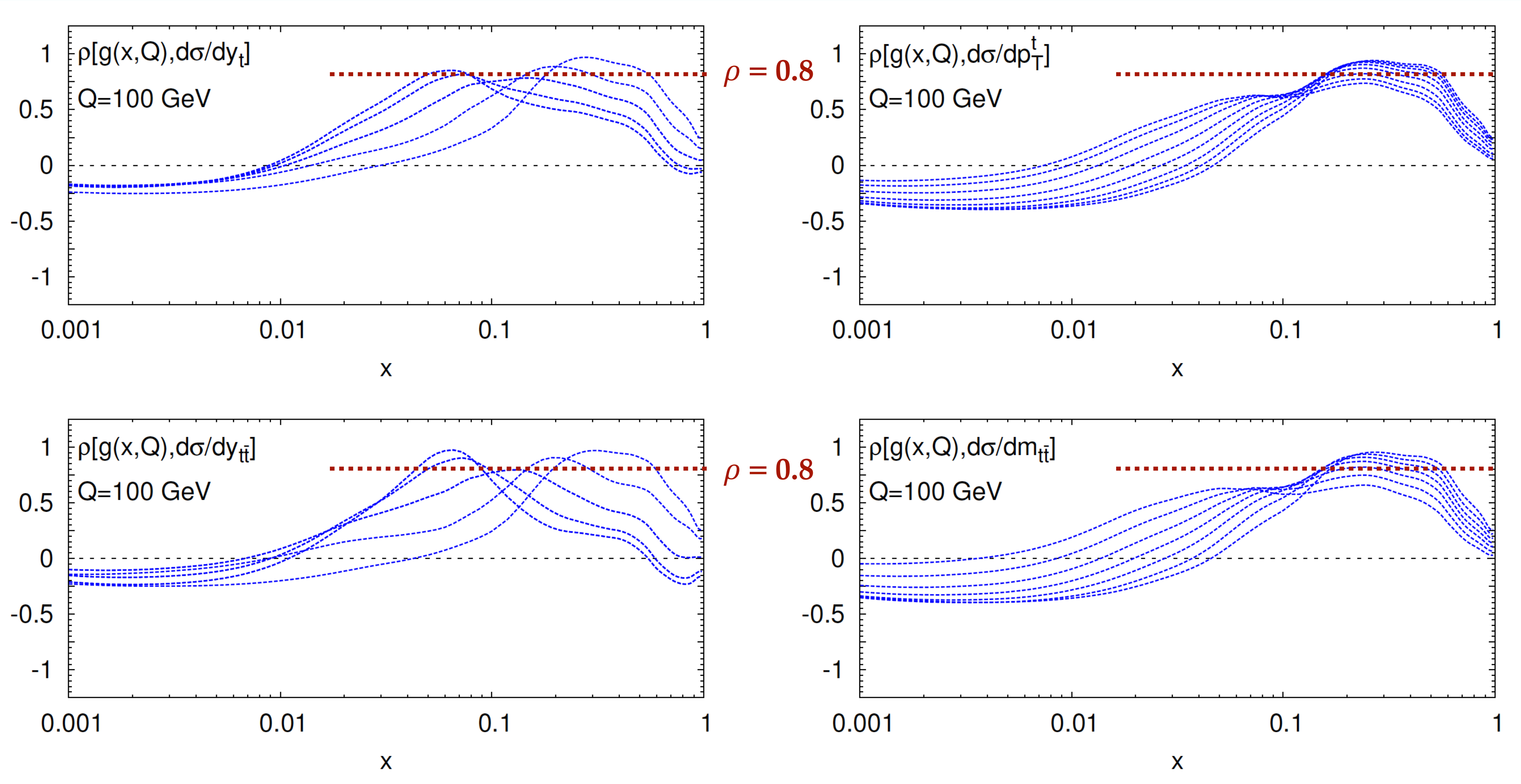
$$\mu'_R = \mu'_F = \mu' = \sqrt{m_t^2 + \left(p_T^t\right)^2}\Big/2$$

- Fast NLO calculations based on **Sherpa/MCgrid**, supplemented by bin-by-bin **NNLO/NLO K-factors**
- All available sources of **statistical and systematic correlated uncertainties** accounted for

Sensitivity to the large-x gluon

- The **correlation coefficient** between the gluon PDF and each of the bins of the four kinematic distributions determines the region of Bjorken-x where **available data has sensitivity**

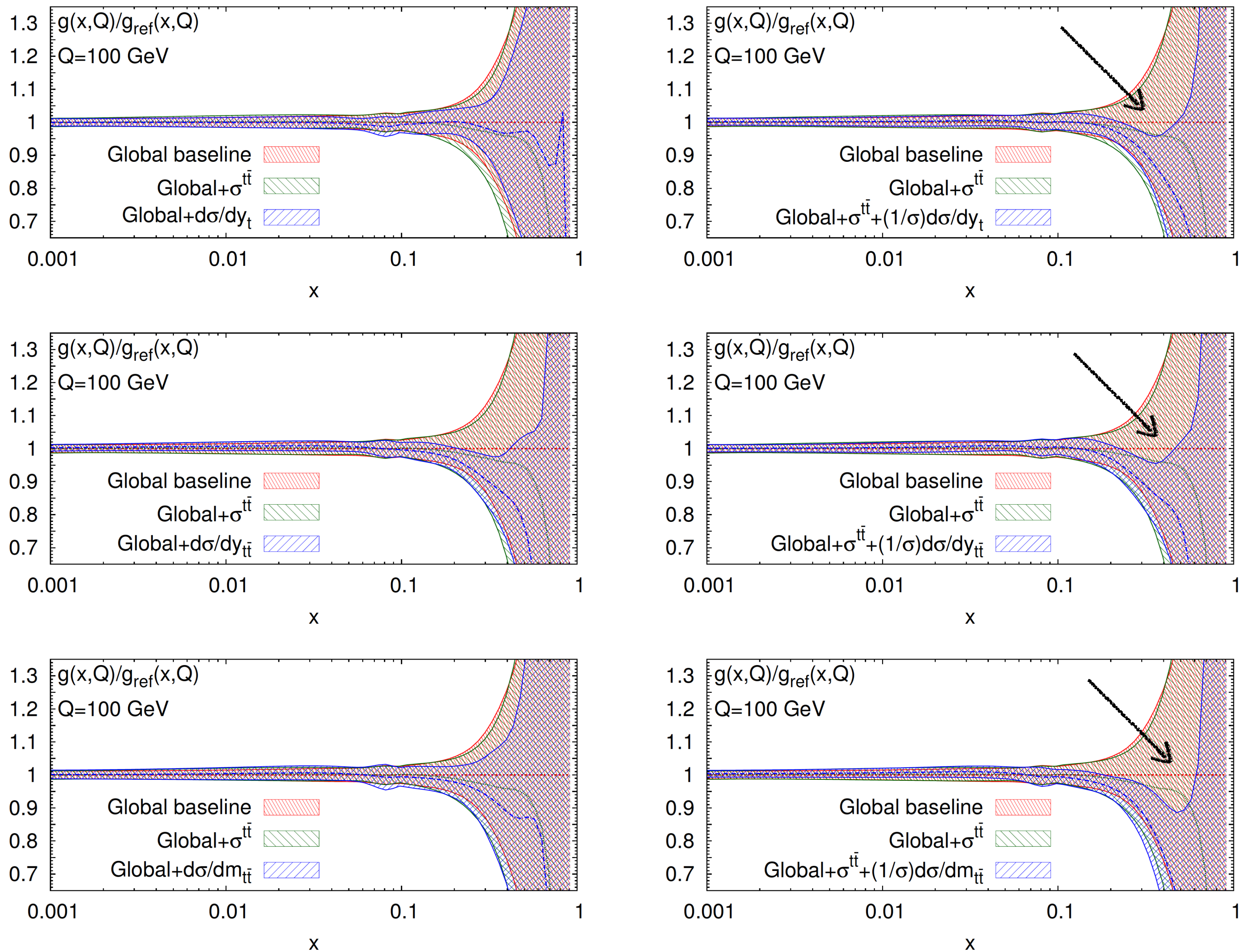
- Large values, $\rho \geq$ **0.8**, of the correlation for **0.07 < x < 0.6,** representing a significant improved over the **coverage in x of inclusive cross-section measurements**



Impact on the large-x gluon

Absolute

Normalized



Which top-quark differential data to fit?

- Fit **normalized distributions**, which exhibit somewhat enhanced constraining power

- Among these, select the distributions with which lead to **larger reduction of PDF uncertainties**

- Require **good agreement between data and theory, $\chi^2/N_{dat} \sim 1$,** to avoid distorting the fit in the case of (still not understood) tensions between ATLAS and CMS

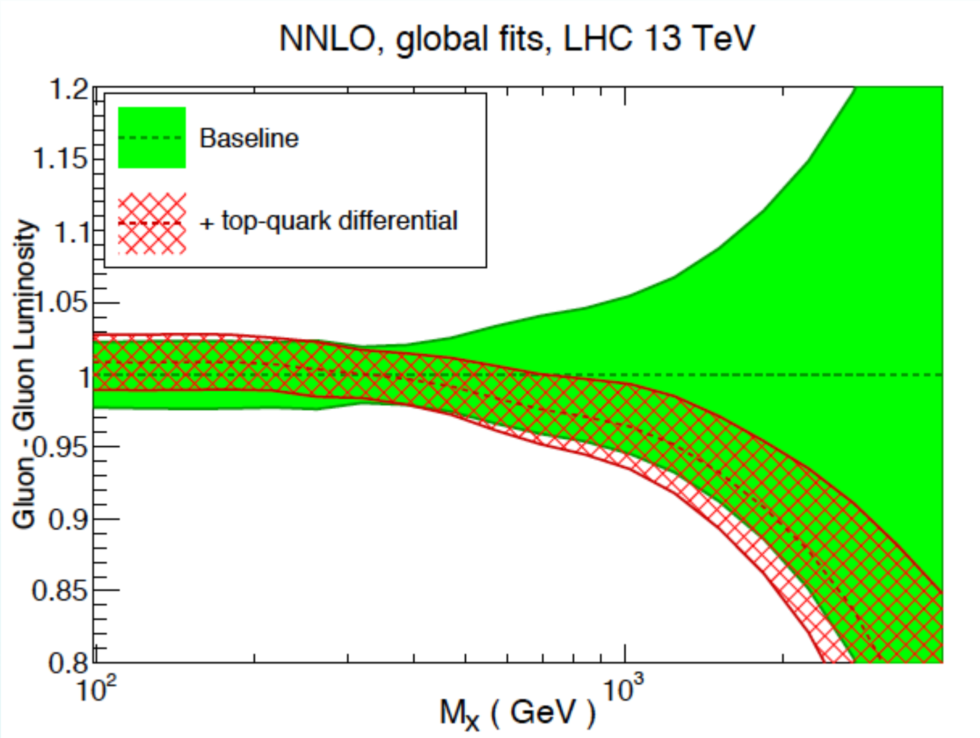
- Include **one kinematic distribution from ATLAS** and **a different one from CMS**, to achieve better kinematical coverage on the gluon PDF

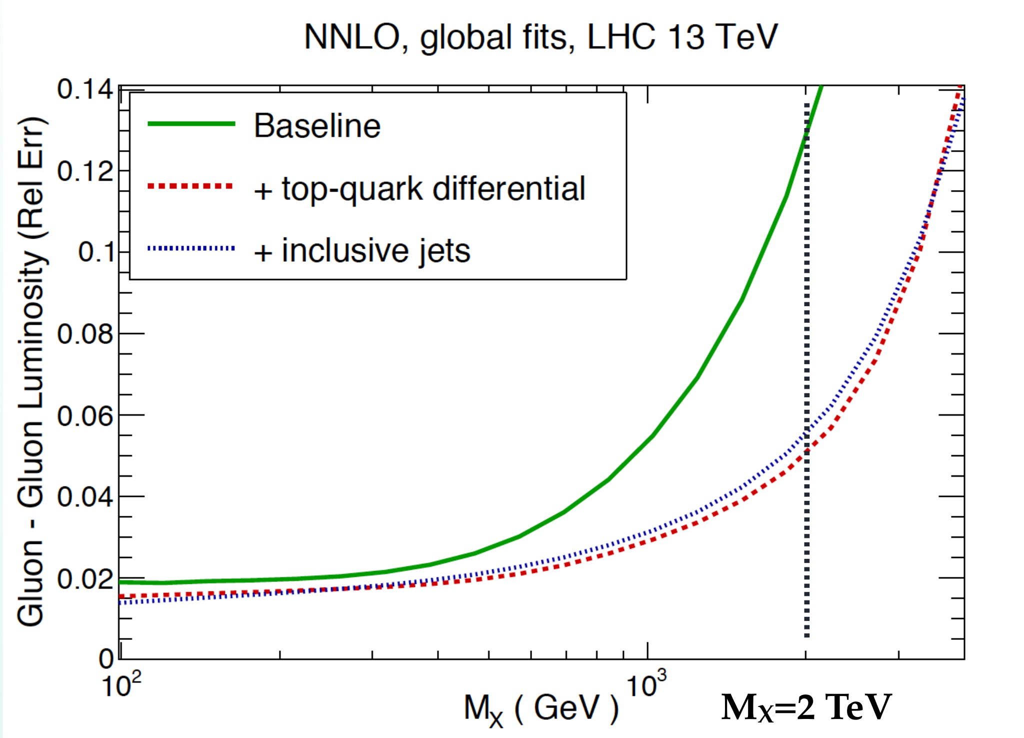
Our recommendation for the 8 TeV lepton+jets differential distributions:

- the normalized top-quark rapidity distribution $(1/\sigma)d\sigma/dy_t$ from ATLAS;
- the normalized top-quark pair rapidity distribution $(1/\sigma)d\sigma/dy_{t\bar{t}}$ from CMS;
- and the total inclusive cross-section $\sigma_{t\bar{t}}$ from ATLAS and CMS at $\sqrt{s} = 8$ TeV.

Other possible choices, within the above guidelines, would lead to consistent results, since the **pull on the large-x gluon is similar for all the ATLAS and CMS distributions** (even when $\chi^2/N_{dat} \gg 1$)

Impact on the large-x gluon





- Significant reduction of PDF uncertainties in the **gluon-gluon luminosity at high invariant masses**
- For instance, for **M_X=2 TeV**, the PDF uncertainties decrease **from 13% to 5%**
- Remarkably, the **constraints from top differential data** in the global fit are comparable to those from **inclusive jets**, despite coming from much fewer data points: **N_{dat} =17 for top vs N_{dat} =470 for jets**

Impact on differential distributions

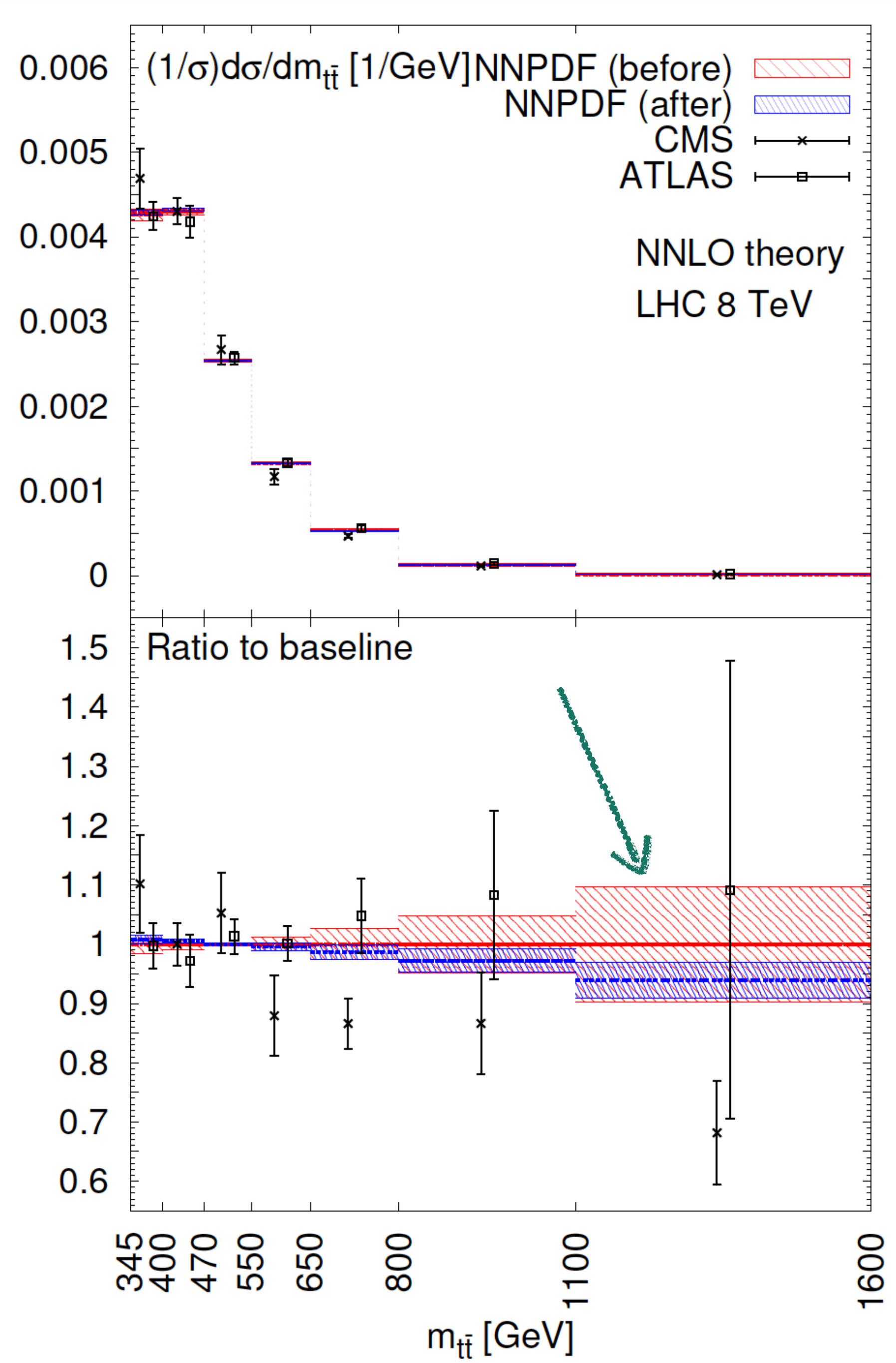
- PDF uncertainties reduced by more than a factor two for **$m_{tt} \gtrsim$ 500 GeV**

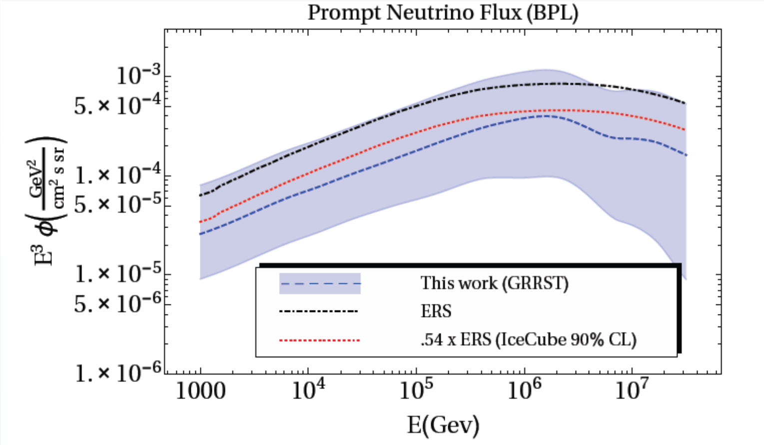
- Similar improvements for **gluon-driven processes,** either SM or BSM, at high masses

- Self-consistent program to **use top data to provide improved theory predictions**

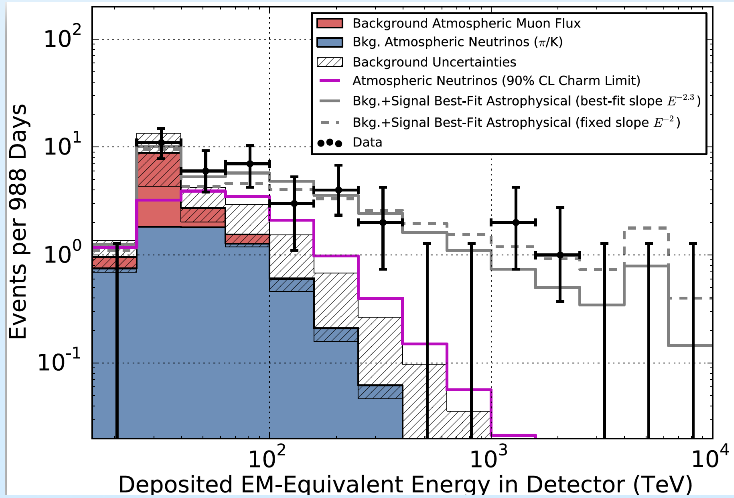
- Our **choice of fitted distributions** (y_t and y_{tt}) reduce the **risk of *BSM contamination*** (kinematical suppression of heavy resonances) which can show up instead in the tails of the m_{tt} and p^t_T distributions (whose PDF uncertainties are now greatly reduced)

Improved sensitivity to BSM dynamics with top-quark final states





From hadron colliders to neutrino telescopes



Gauld, Rojo, Rottoli, Talbert 15

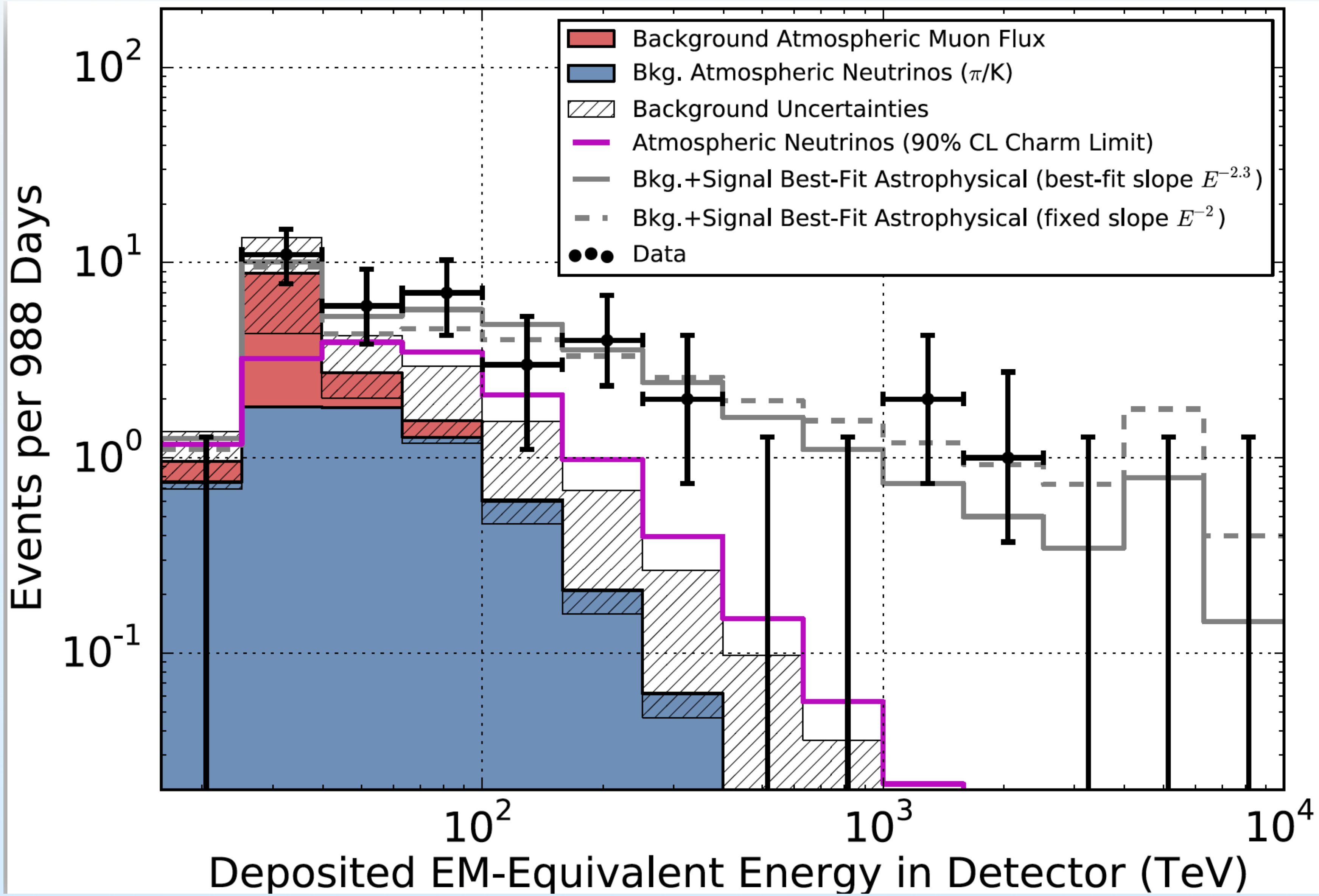
Gauld, Rojo, Rottoli, Sarkar, Talbert 15

Gauld, Rojo 16

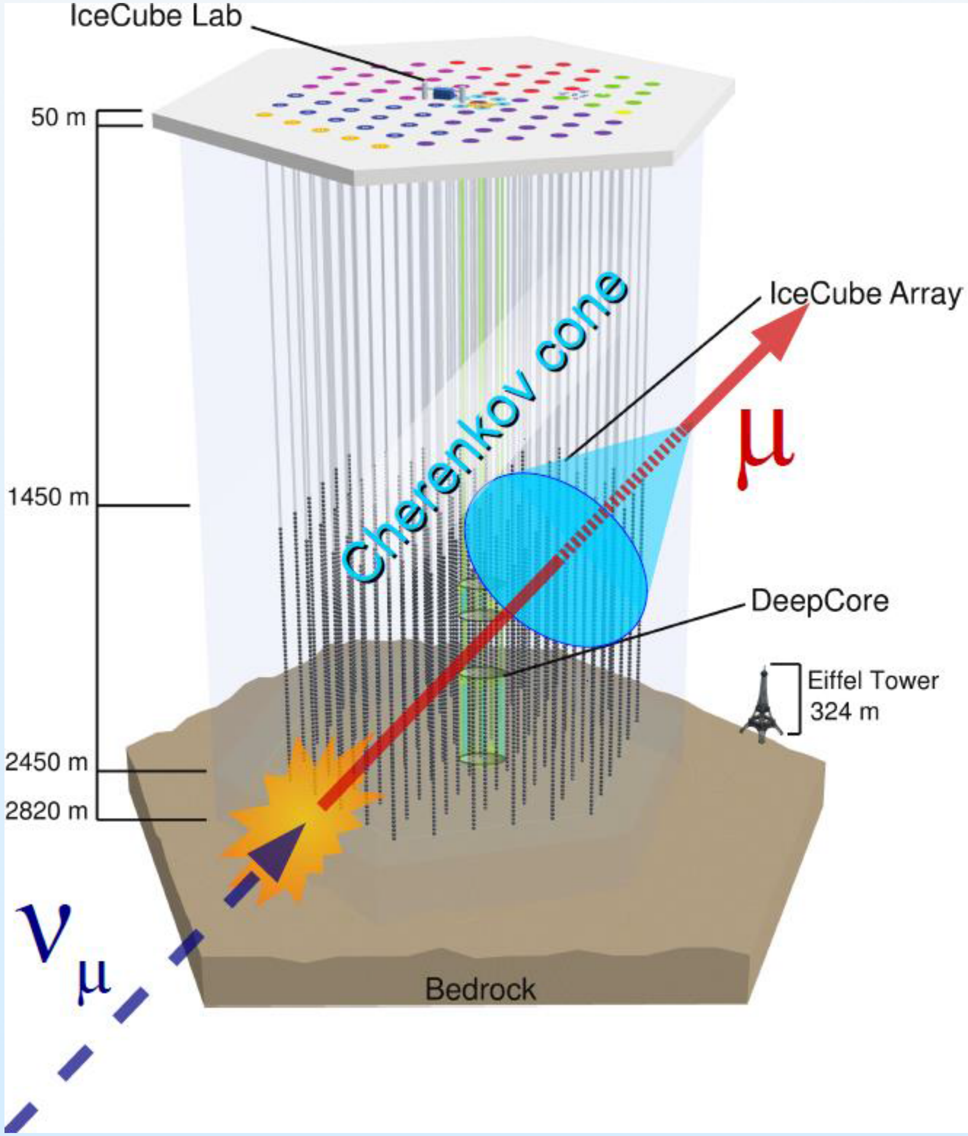
From the LHC to Neutrino Telescopes

Observation of Ultra-High Energy (UHE) neutrino events at IceCube heralds start of **Neutrino Astronomy**

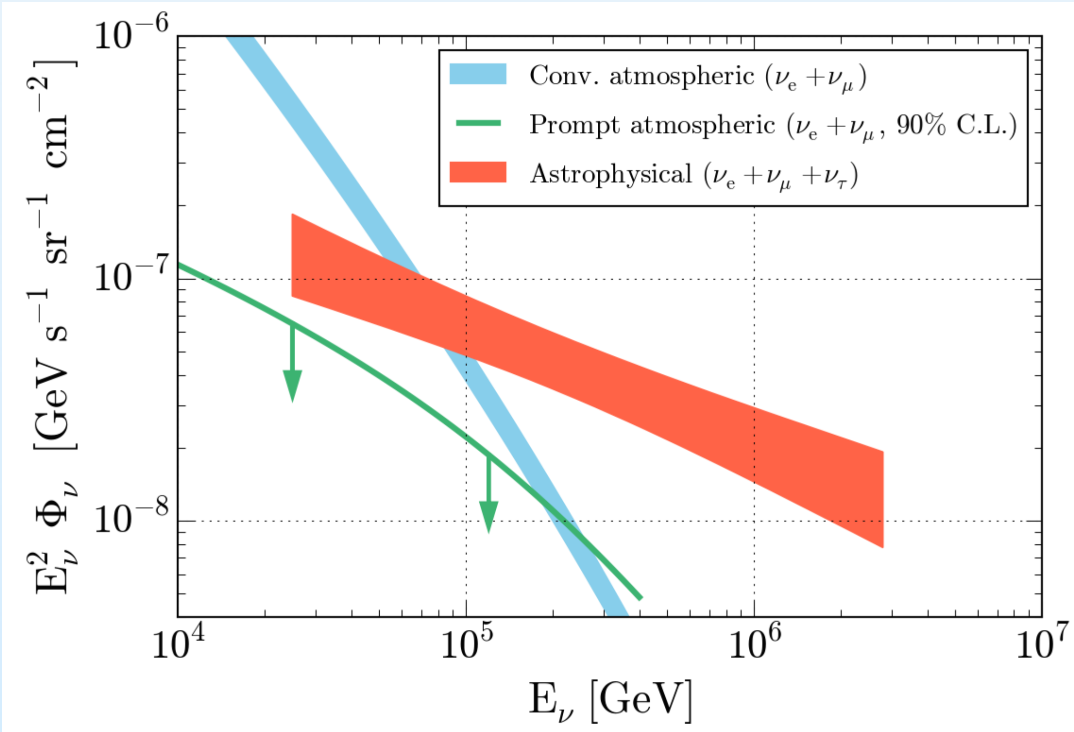
New window to the Universe, but interpretation of IceCube data requires **control over backgrounds**



From the LHC to Neutrino Telescopes



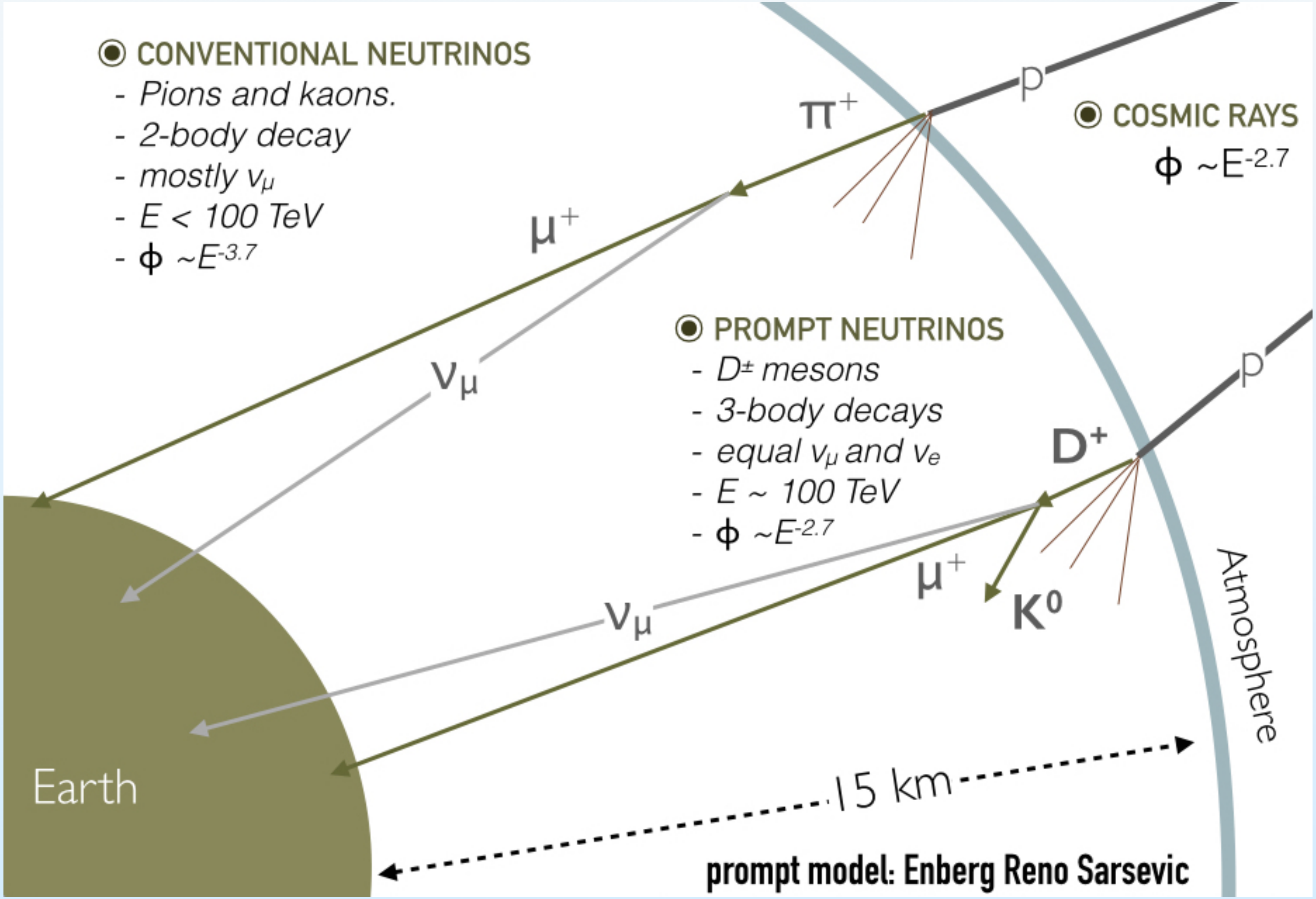
- The main **background for astrophysical neutrinos at IceCube** is the flux of neutrinos from the **decays of charm mesons** in cosmic ray collisions in atmosphere
- Theoretically, this **prompt neutrino flux** is affected by large uncertainties: very small-x PDFs, very low scales - can pQCD be applied?
- Use **LHC data itself** to pin down this prompt flux



From the LHC to Neutrino Telescopes

Observation of Ultra-High Energy (UHE) neutrino events at IceCube heralds start of **Neutrino Astronomy**

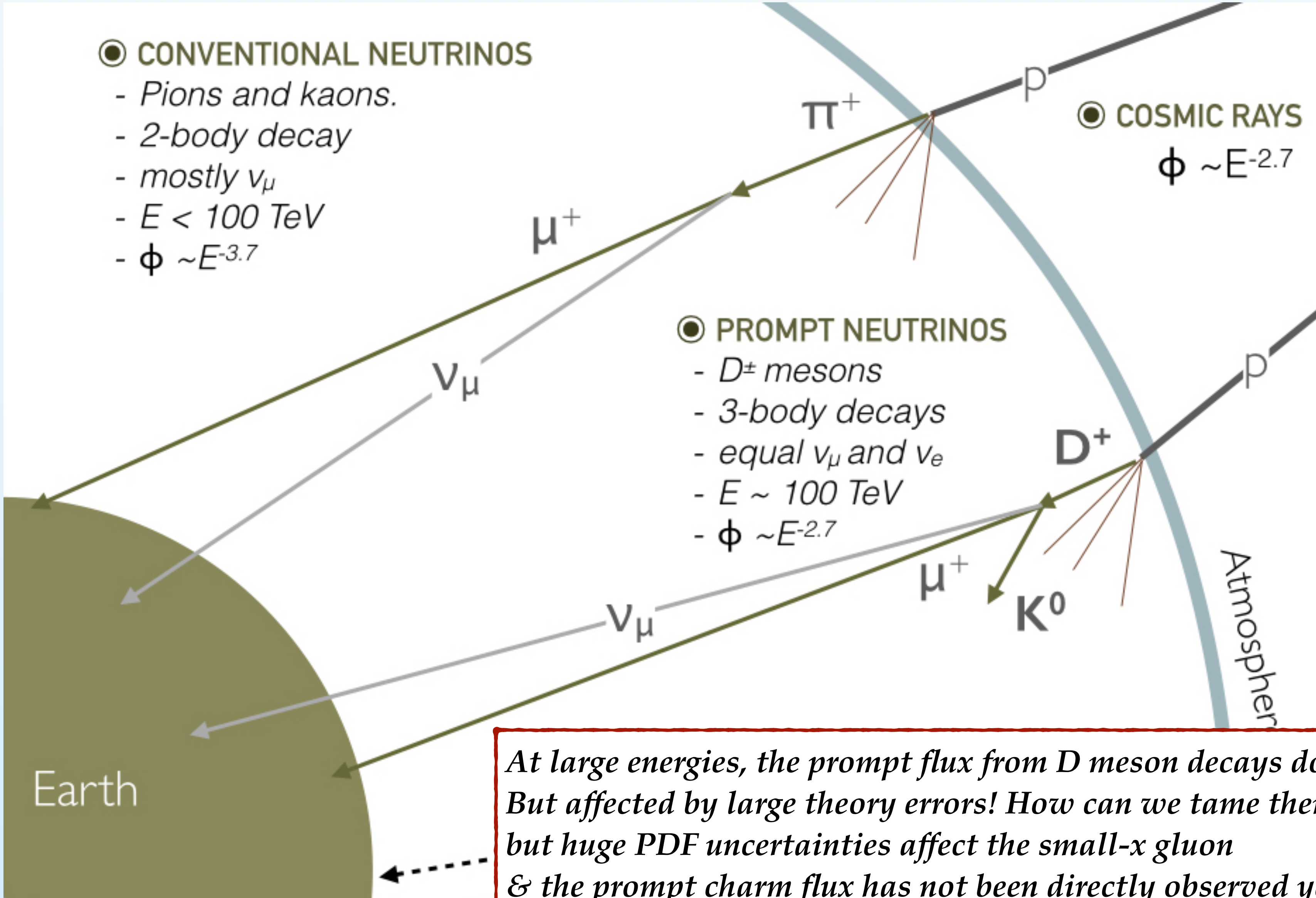
New window to the Universe, but interpretation of IceCube data requires **control over backgrounds**



From the LHC to Neutrino Telescopes

Observation of Ultra-High Energy (UHE) neutrino events at IceCube heralds start of **Neutrino Astronomy**

New window to the Universe, but interpretation of IceCube data requires **control over backgrounds**



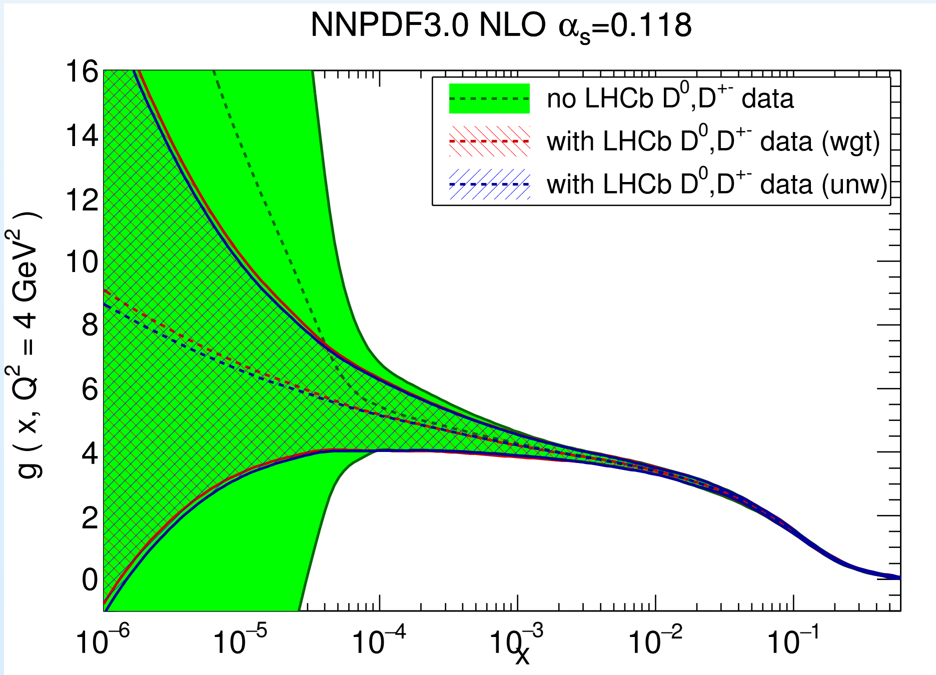
At large energies, the prompt flux from D meson decays dominates
But affected by large theory errors! How can we tame them?
but huge PDF uncertainties affect the small-x gluon
& the prompt charm flux has not been directly observed yet …

From the LHC to Neutrino Telescopes

Use **collider data** to provide state-of-the-art **predictions for backgrounds at Neutrino Telescopes**

The **LHCb forward charm production data** at 7 TeV cover the **same kinematical region** as prompt neutrino production in high-energy cosmic rays

- ☑ Include **7 TeV LHCb forward charm production data** in the global fit
- ☑ Validate **perturbative QCD calculations** on collider data, and **constrain the small-x gluon**
- ☑ Compute optimised predictions for **prompt neutrino fluxes at high energies**



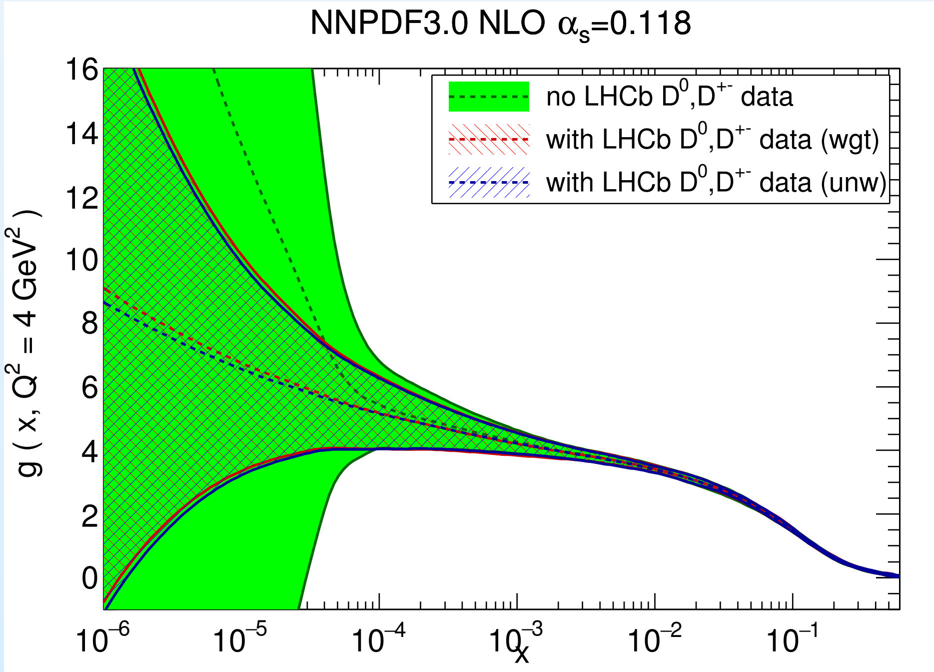
Gauld, Rojo, Rottoli, Talbert 15

From the LHC to Neutrino Telescopes

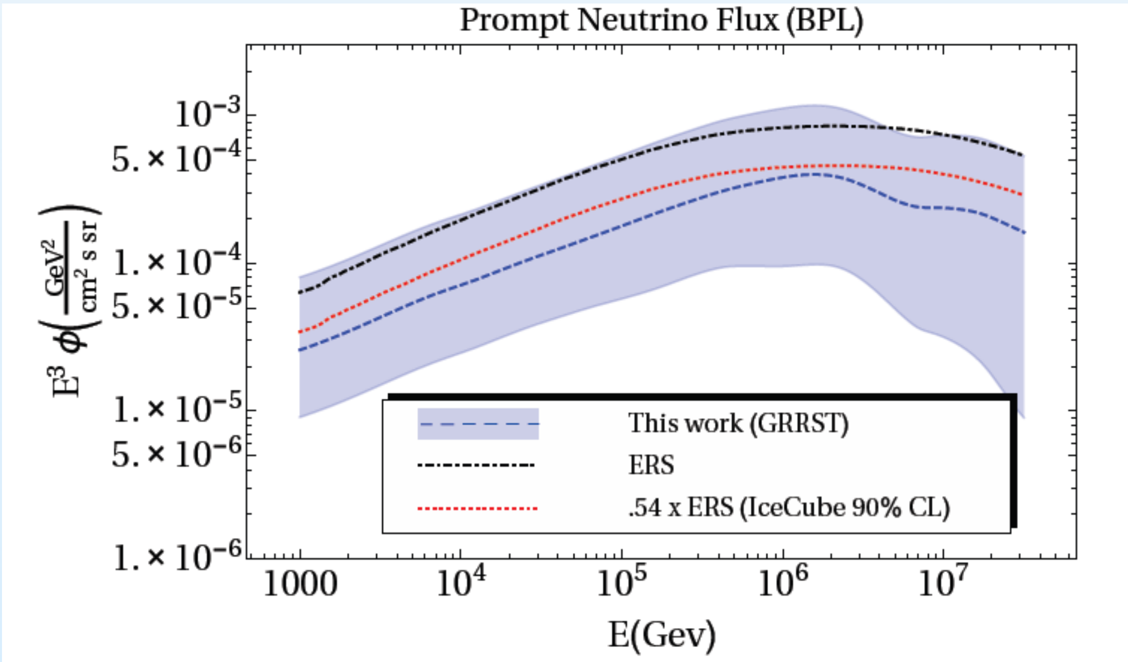
Use **collider data** to provide state-of-the-art **predictions for backgrounds at Neutrino Telescopes**

The **LHCb forward charm production data** at 7 TeV cover the **same kinematical region** as prompt neutrino production in high-energy cosmic rays

- ☑ Include **7 TeV LHCb forward charm production data** in the global fit
- ☑ Validate **perturbative QCD calculations** on collider data, and **constrain the small-x gluon**
- ☑ Compute optimised predictions for **prompt neutrino fluxes at high energies**



Gauld, Rojo, Rottoli, Talbert 15



Gauld, Rojo, Rottoli, Sarkar, Talbert 15

We predict that detection of the prompt neutrino flux should be within IceCube reach

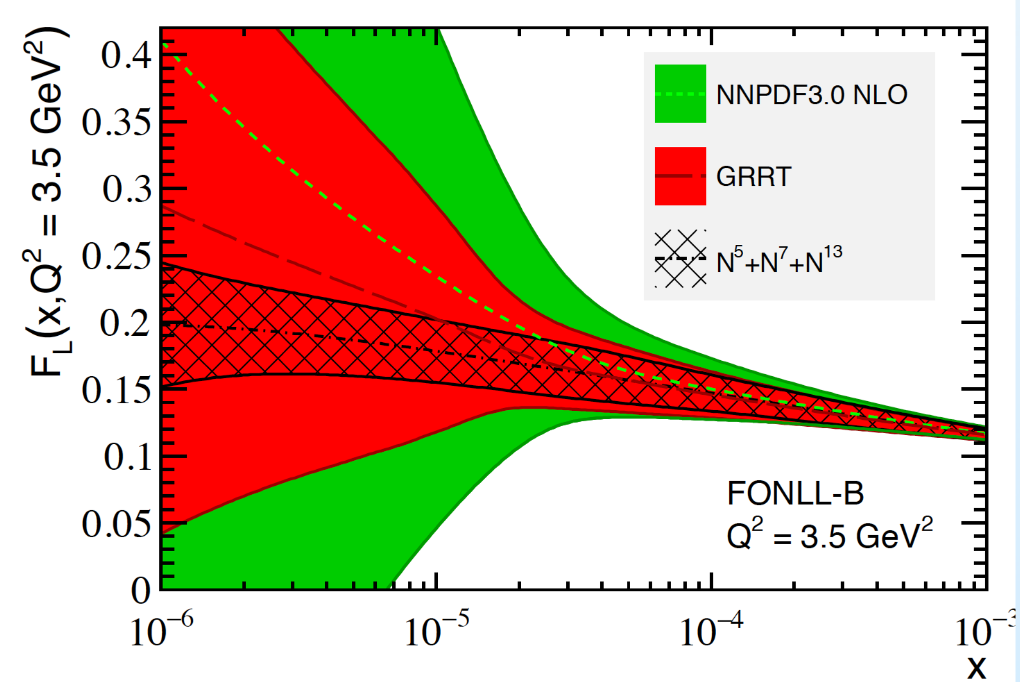
Precision determination of the small-x gluon

- Recently LHCb presented charm production data at **5 and 13 TeV**
- Different options to fit data, either normalised distributions or cross-section ratios (13 TeV / 7 TeV and 13 TeV / 5 TeV)
- As compared to NNPDF3.0, we achieve up to a **factor 10 reduction on the small-x gluon** uncertainties down to **x ≲ 10^{-6}**
- High-precision predictions of UHE neutrino-nucleus cross section for energies up to **$E_\nu \simeq 10^6$ PeV**

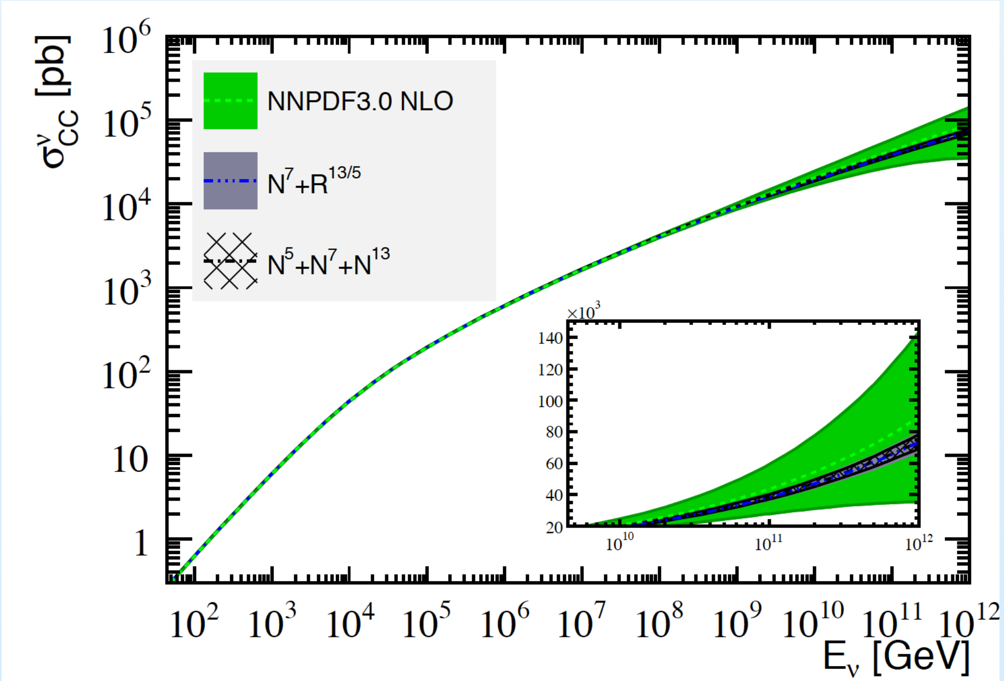
$$N_X^{ij} = \frac{d^2\sigma(\mathrm{X\ TeV})}{dy_i^D d(p_T^D)_j} \Big/ \frac{d^2\sigma(\mathrm{X\ TeV})}{dy_{\mathrm{ref}}^D d(p_T^D)_j}$$

$$R_{13/X}^{ij} = \frac{d^2\sigma(\mathrm{13\ TeV})}{dy_i^D d(p_T^D)_j} \Big/ \frac{d^2\sigma(\mathrm{X\ TeV})}{dy_{\mathrm{i}}^D d(p_T^D)_j}$$

F_L at future high-energy lepton-proton colliders



Ultra-high energy neutrino-nucleus cross-section



Gauld, Rojo 16

Summary and outlook

- **Parton Distributions** are an essential ingredient for **LHC phenomenology**

- **At the LHC,** accurate PDFs are required for

  - ☑ **Precision SM measurements** and determination of **fundamental parameters** like **W mass**

  - ☑ **The characterisation of the Higgs sector**

  - ☑ **Searches for New Physics** beyond the Standard model

  - ☑ **Monte Carlo event generators** and higher-order **perturbative calculations**

- The NNPDF approach provides state-of-the art Parton Distributions based on a **robust, unbiased methodology,** the most updated **theoretical information** and all the relevant experimental information **including LHC data**

- The recent years have seen a **revolution** in PDF fits: PDFs with **LHC data**, PDFs with **QED corrections**, PDFs with **all-order resummations**, PDFs tailored for **neutrino telescopes,** and much more

Beyond unpolarized (NN)PDFs

Unpolarized proton structure: *NNPDF3.0*
LHC phenomenology
Higgs boson characterisation
New Physics searches

Polarized proton structure: *NNPDFpol1.1*
The proton spin puzzle
The polarization of gluons

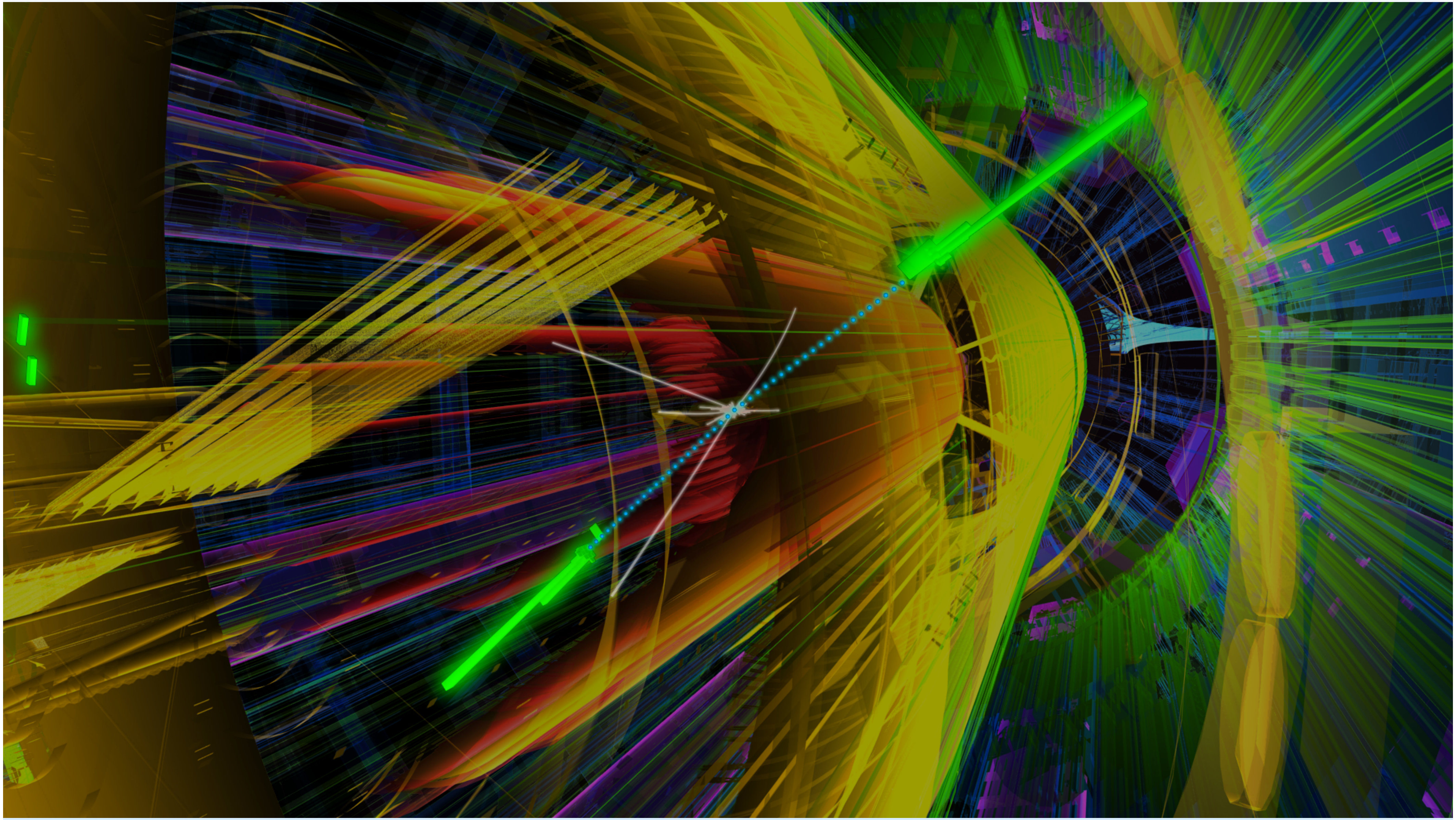
Fragmentation Functions: *NNFF1.0*
inclusive hadron production
in DIS, pp, p+Pb, Pb+Pb collisions
in preparation

Nucleon Structure

Hadron Fragmentation

Nuclear parton distributions: *nNNPDF1.0*
from nuclear DIS and LHC p+Pb data
QCD in the high density regime
Quark-Gluon Plasma characterisation
in preparation

Fascinating times ahead at the high-energy frontier!



we should stay tuned for news from the LHC!

Fascinating times ahead at the high-energy frontier!

we should stay tuned for news from the LHC!