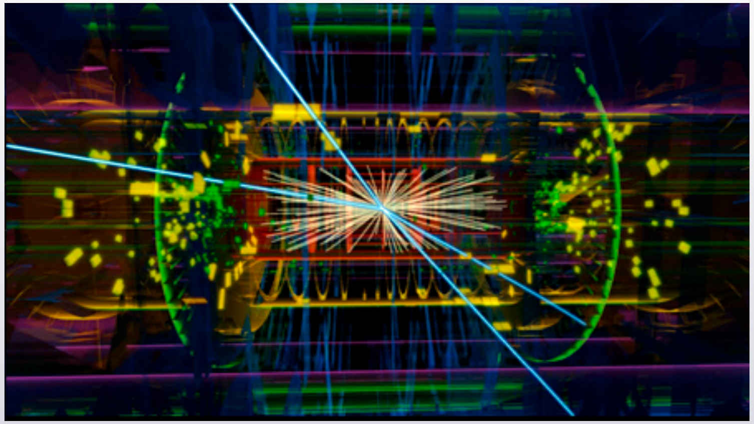
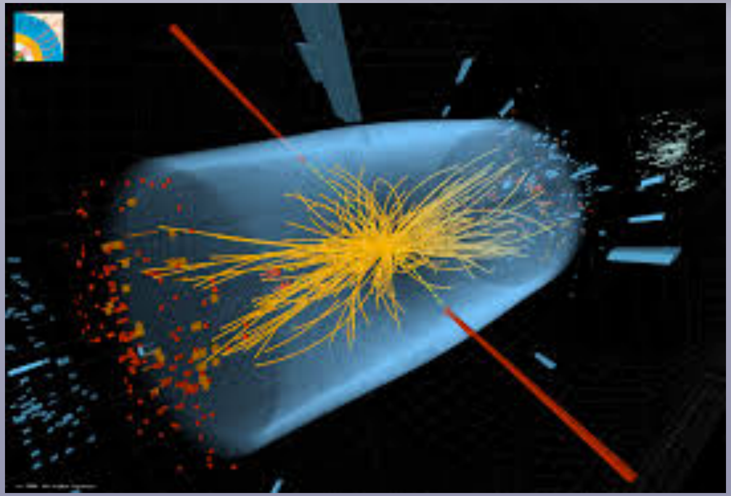
Parton Distributions for the LHC Run II

Juan Rojo
STFC Rutherford Fellow
Rudolf Peierls Center for Theoretical Physics
University of Oxford
http://www.juanrojo.com/
@JuanRojoC

Birmingham Particle Physics Seminar
Birmingham, 07/10/2015



Parton Distributions and LHC phenomenology



Juan Rojo

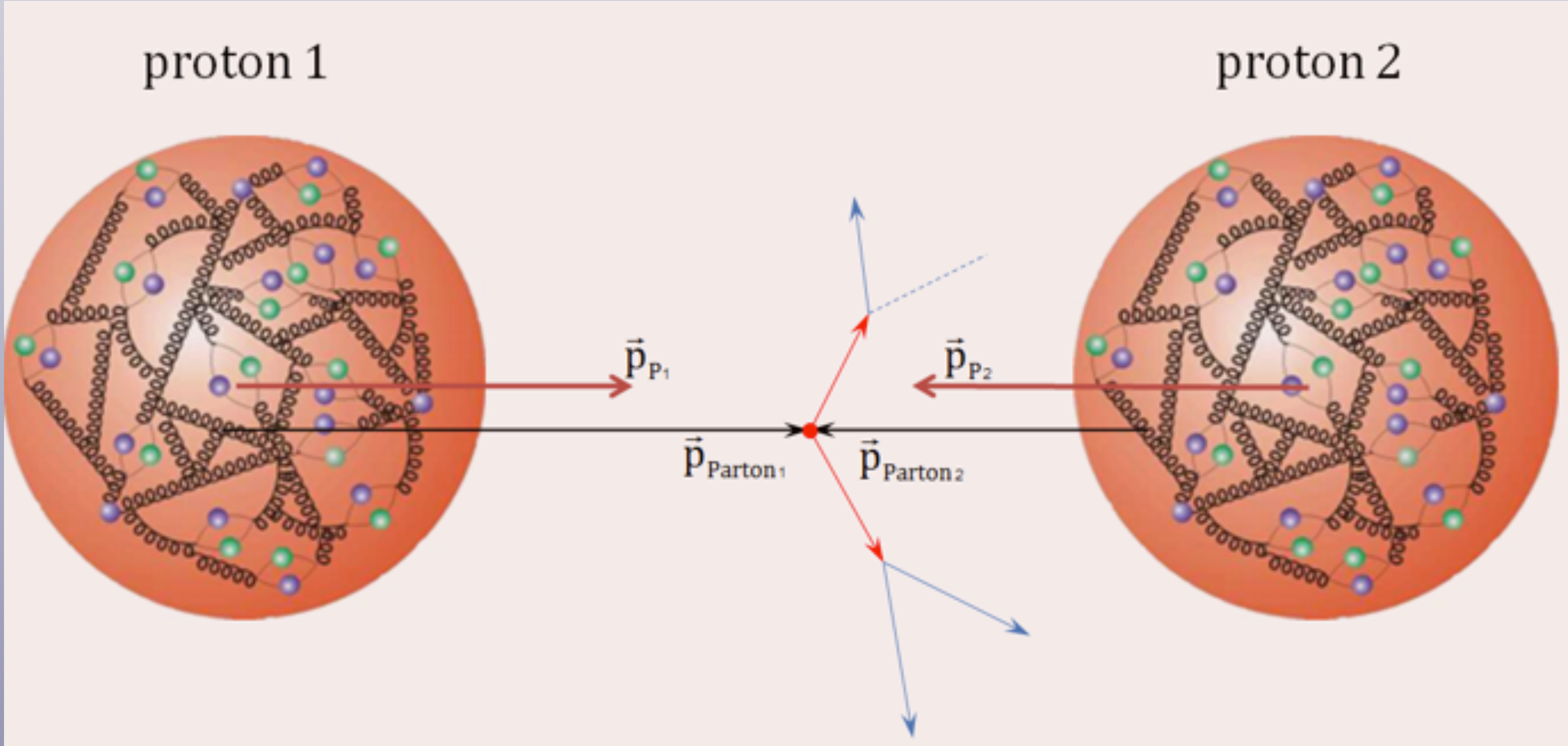
Birmingham, 07/10/2015

The inner life of the protons

☑ The **Large Hadron Collider** collides proton, but these are not fundamental particles: really what the LHC is doing is colliding **quarks and gluons**

☑ The distribution of momentum that the quarks and gluons carry is quantified by the **Parton Distribution Functions (PDFs),** determined by **non-perturbative dynamics:** cannot be computed from first principles and need to be **extracted from experimental data**

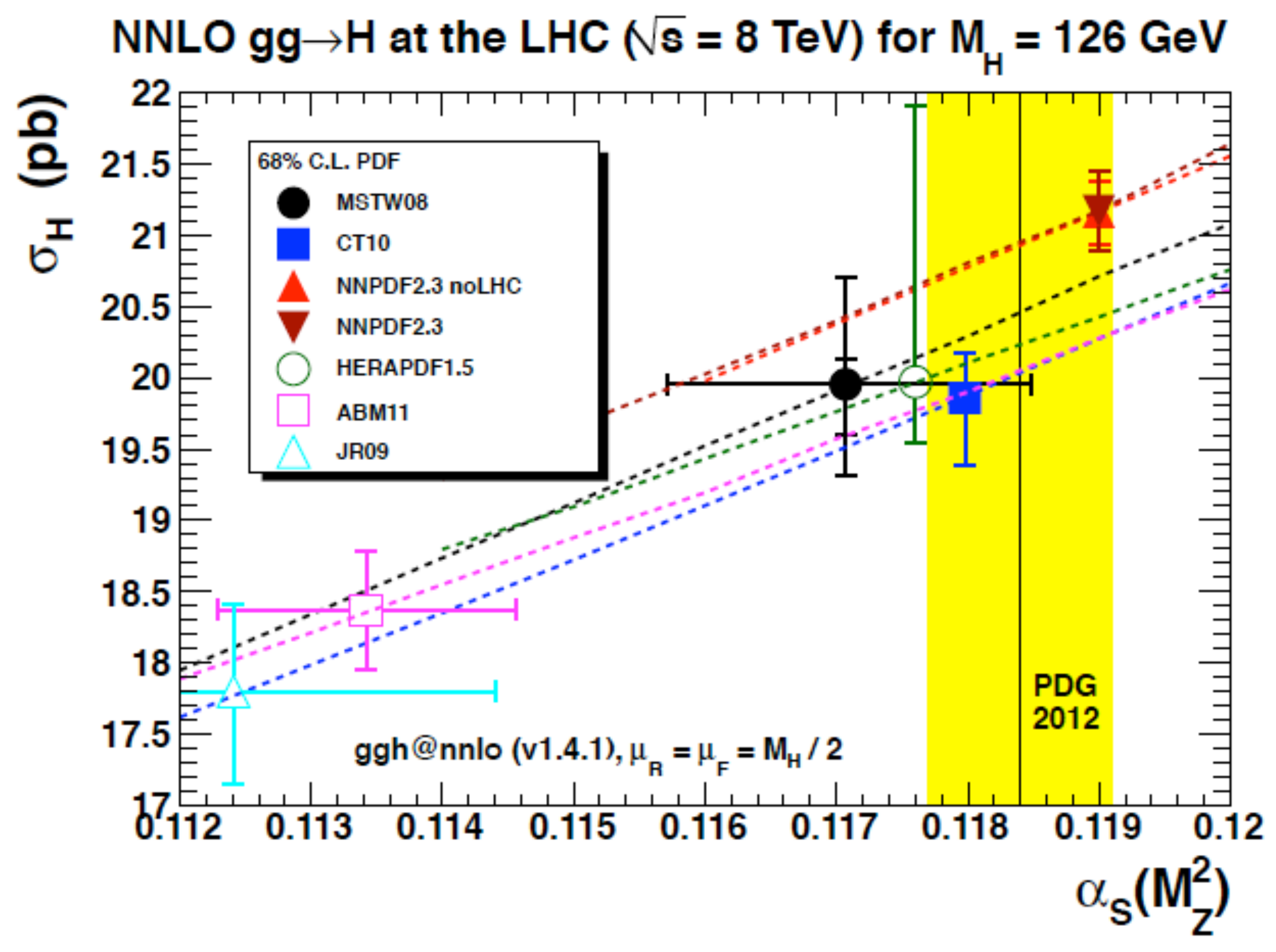
☑ An **accurate determination of PDFs** is of paramount importance to be able to do **precision physics at hadronic colliders as the LHC**

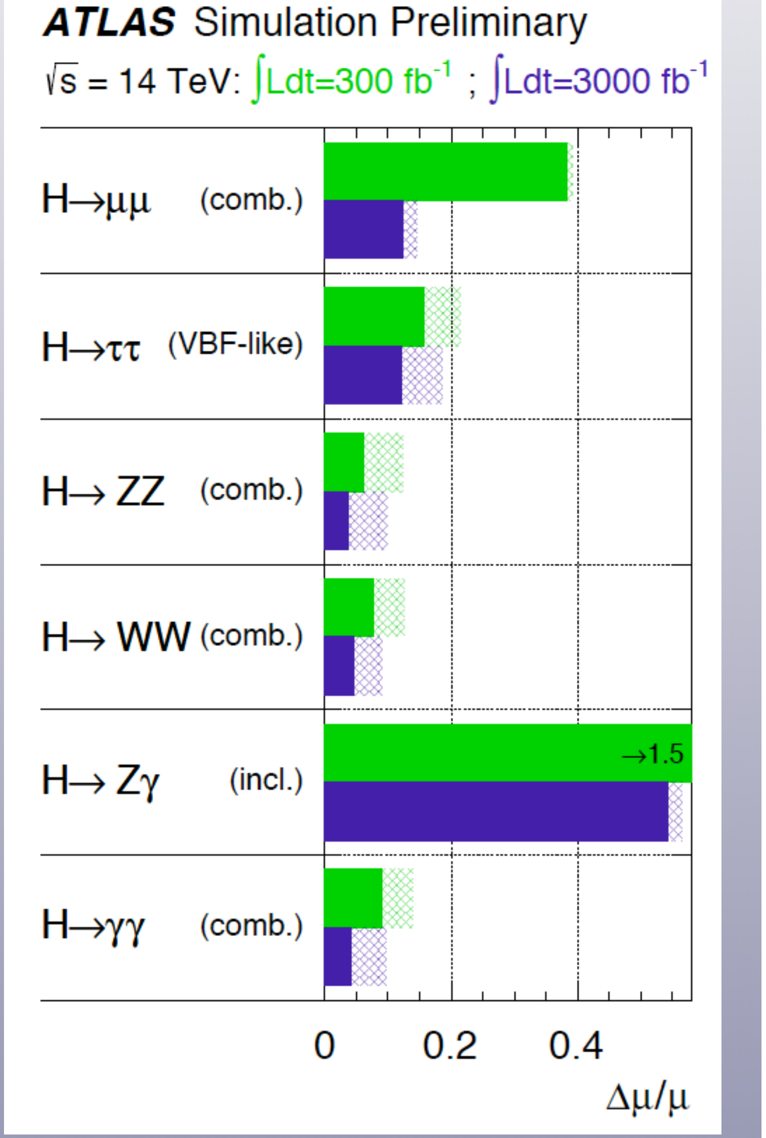


Parton Distributions and LHC phenomenology

1) PDFs fundamental limit for Higgs boson characterization in terms of couplings

Solid: no TH unc
Hatched: with TH unc

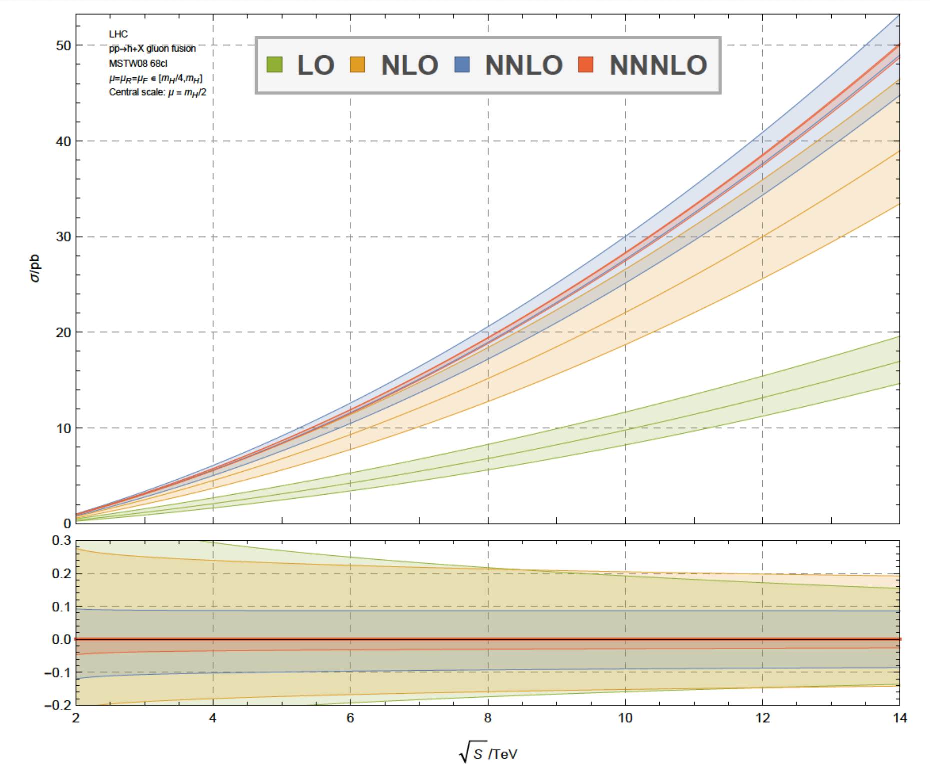


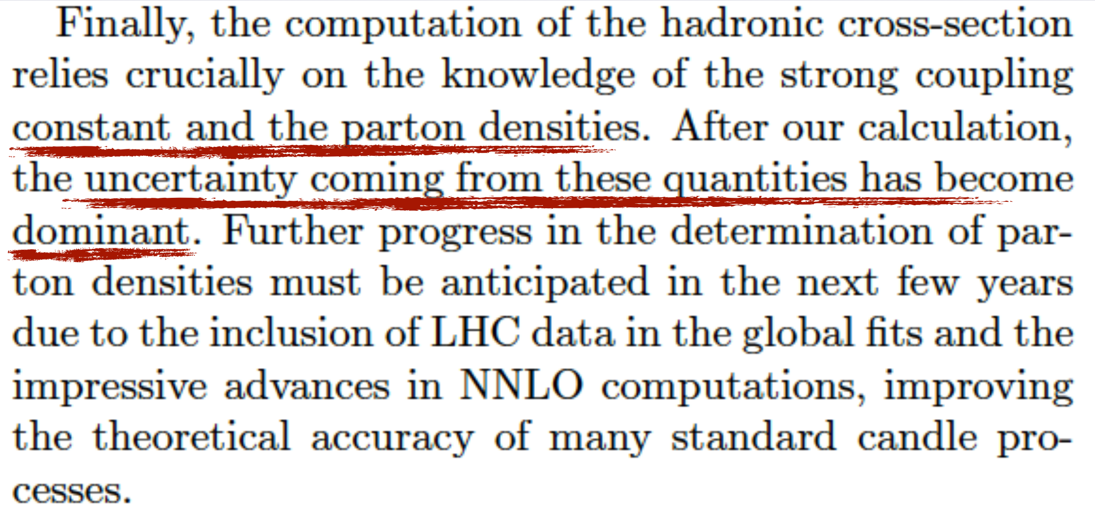


Parton Distributions and LHC phenomenology

☑ Recently **massive development of NNLO higher-order calculations …**

☑ … now we even have the **Higgs gluon fusion xsec at N3LO! Scale uncertainties down to 2%!**




Finally, the computation of the hadronic cross-section relies crucially on the knowledge of the strong coupling constant and the parton densities. After our calculation, the uncertainty coming from these quantities has become dominant. Further progress in the determination of parton densities must be anticipated in the next few years due to the inclusion of LHC data in the global fits and the impressive advances in NNLO computations, improving the theoretical accuracy of many standard candle processes.

Anastasiou et al, arxiv:1503.06056

☑ PDF uncertainties are now **dominant** for a number of crucial LHC processes, and thus it is crucial to match the **accuracy of hard-cross section calculations** with that of the PDFs

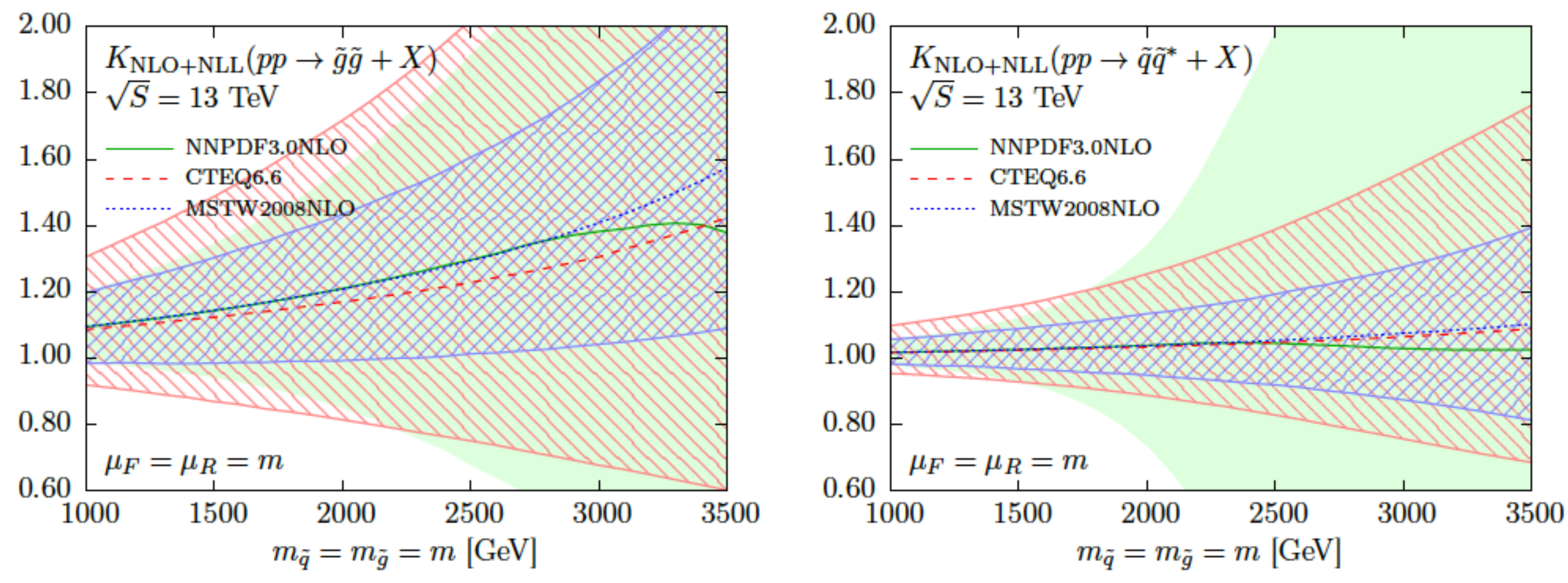
Parton Distributions and LHC phenomenology

2) Very large PDF uncertainties (>100%) for BSM heavy particle production

$K_{\text{NLO+NLL}}$ = (NLO+NLL)/NLO

Gluino Pair Production

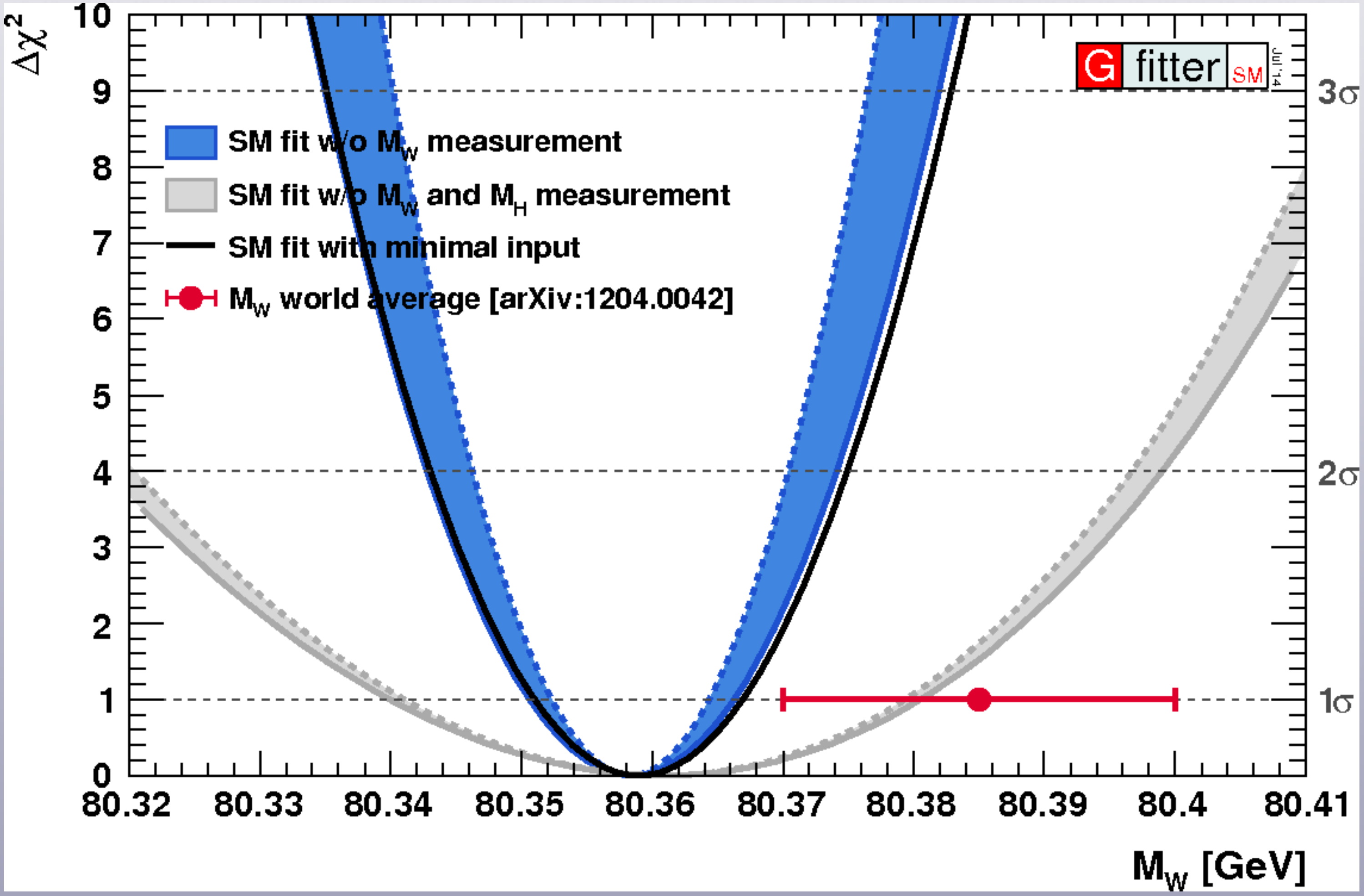
Squark-Antisquark Pair Production



Beenakker, Borschensky, Kramer, Kulesza, Laenen, Marzani, **JR,** arXiv:1510.00375

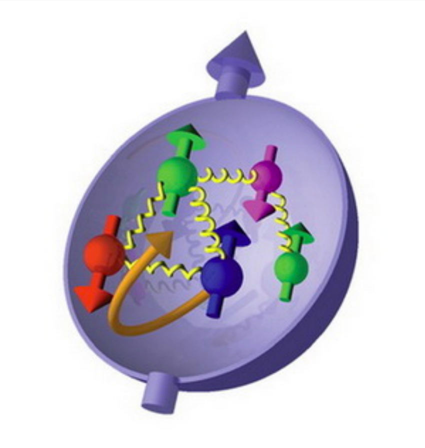
Parton Distributions and LHC phenomenology

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

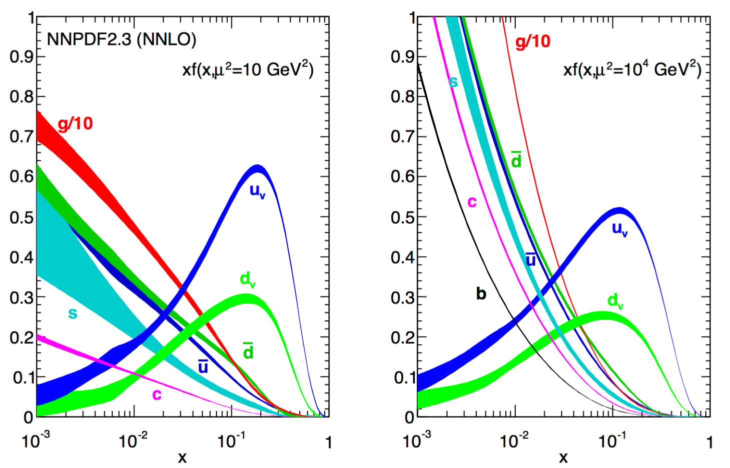


- Reducing TH systematics could lead to **indirect BSM discovery** from precision measurements
- Precision in M_W will **improve by a factor 3** in 10 years
- If SM confirmed, **ruled out a broad class of BSM scenarios**

1.8-sigma tension between direct M_W measurement and global EW fit

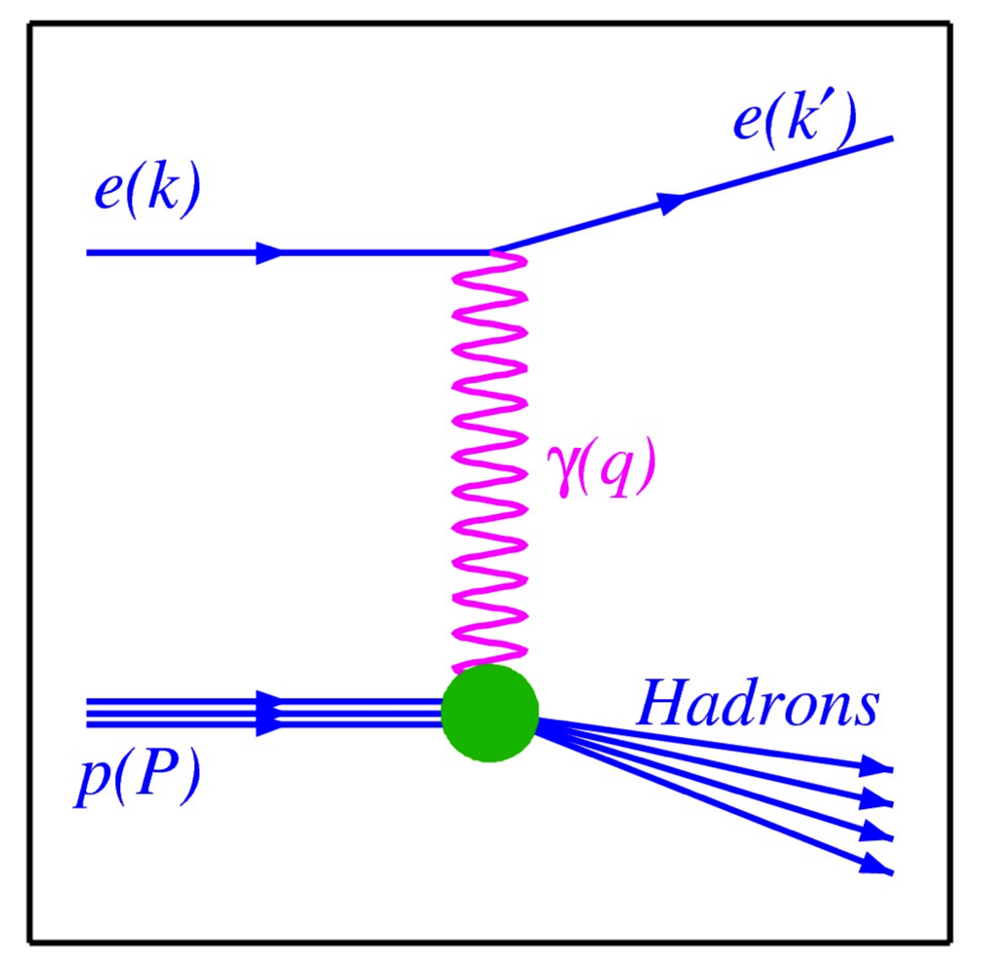


The inner life of protons : Parton Distribution Functions



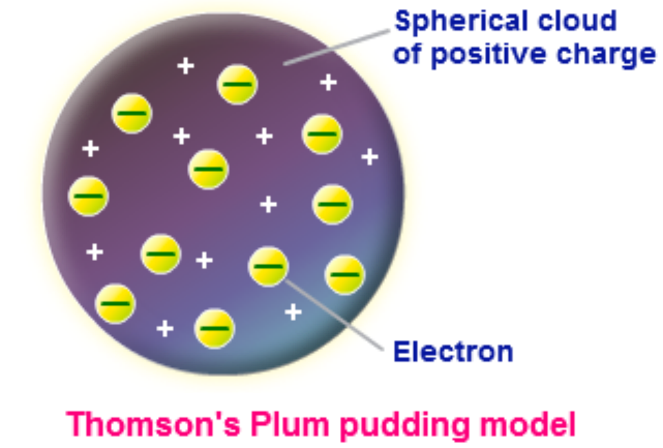
Deep-Inelastic scattering and the discovery of quarks

- Deep-inelastic **lepton-proton scattering**: First evidence for **proton structure**
- Measured scattering cross-section constant as **resolution scale 1/Q decreases.**
- Evidence for **point-like constituents in the proton: the quarks**



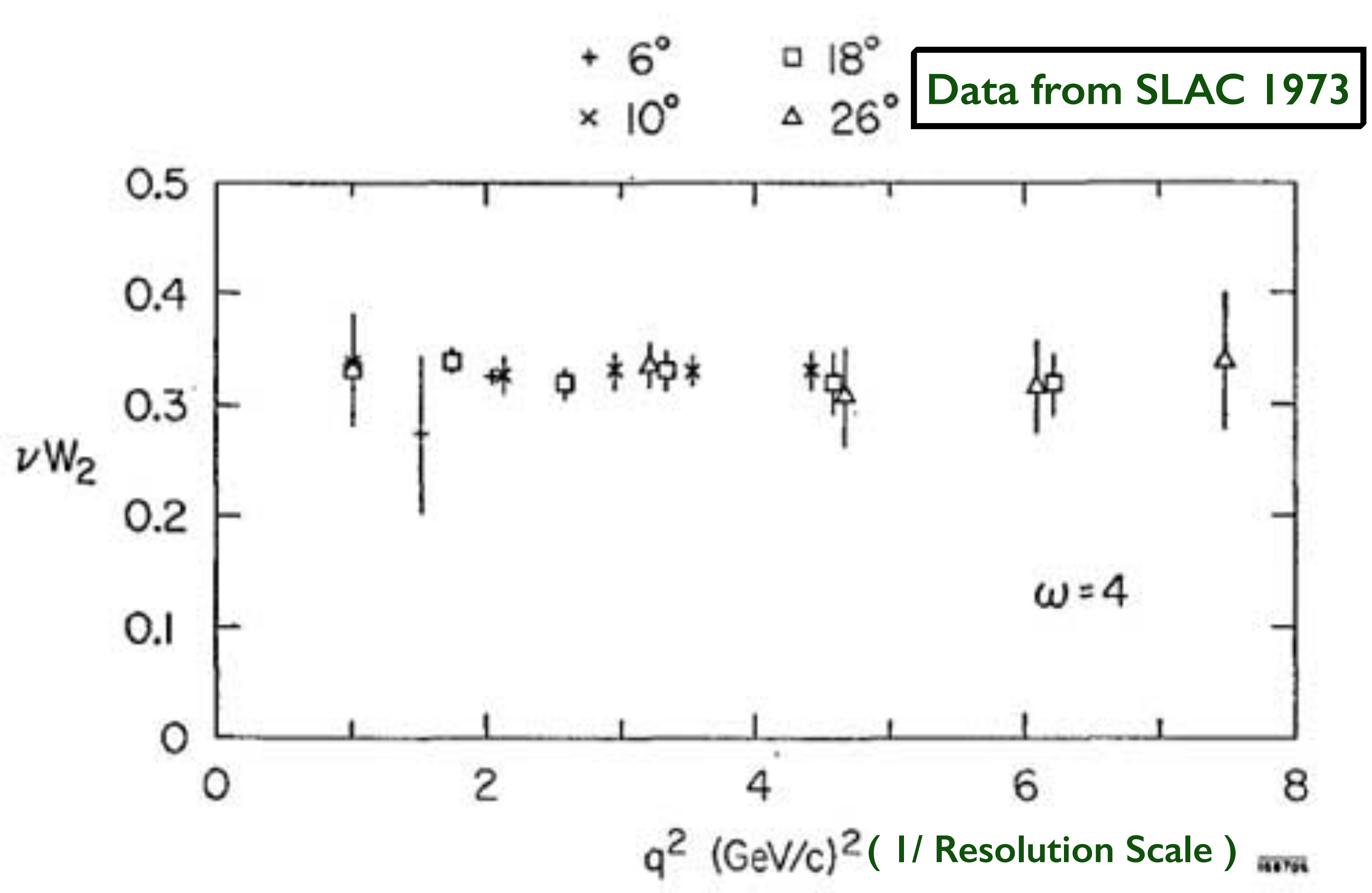
$$x_{\mathrm{Bj}} = \frac{Q^2}{2\,p\cdot q}, \quad Q^2 = -q^2 \quad y = \frac{q\cdot p}{k\cdot p}$$

- If the proton had a different structure, a **form factor F(Q)** would be expected
- Analogous to **Rutherford's discovery of the point-like atomic nucleus**, while expecting Thomson's Plum model



Deep-Inelastic scattering and the discovery of quarks

- Deep-inelastic **lepton-proton scattering**: First evidence for **proton structure**
- Measured scattering cross-section constant as **resolution scale 1/Q decreases.**
- Evidence for **point-like constituents in the proton: the quarks**

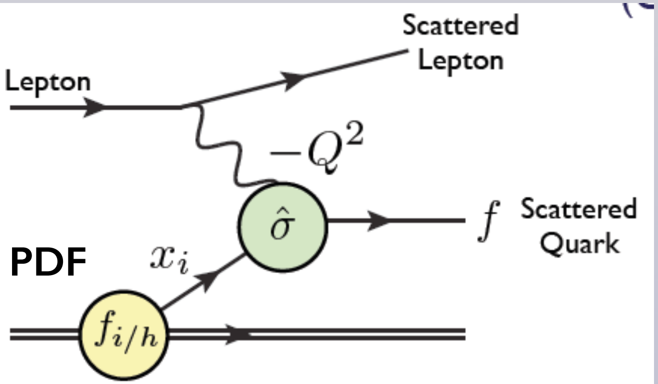


Data from SLAC 1973

(1/ Resolution Scale)

QCD Factorization and PDFs

- **QCD Factorization Theorem:** separate the hadronic cross section into a **perturbative, process dependent partonic cross section** and **non-perturbative, process independent Parton Distributions.** In DIS we have:



$$F_i(x,Q^2) = x\sum_i \int_x^1 \frac{dz}{z} C_i\left(\frac{x}{z},\alpha_s(Q^2)\right) f_i(z,Q^2).$$

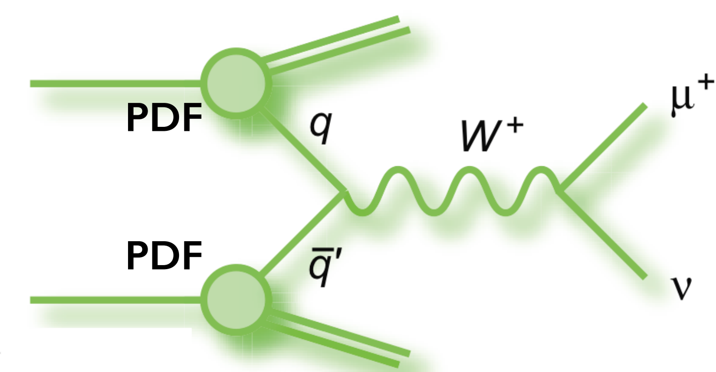
Hadron-level cross section Parton-level cross-section Parton Distribution

- The same **Factorization Theorem** allows to use the same **universal PDFs** to provide **predict ions for proton-proton collisions at the LHC:**

$$\sigma_X(s,M_X^2) \quad = \quad \sum_{a,b}\int_{x_{\min}}^1 dx_1\, dx_2\, f_{a/h_1}(x_1,M_X^2) f_{b/h_2}(x_2,M_X^2)\hat{\sigma}_{ab\to X}\left(x_1x_2s,M_X^2\right)$$

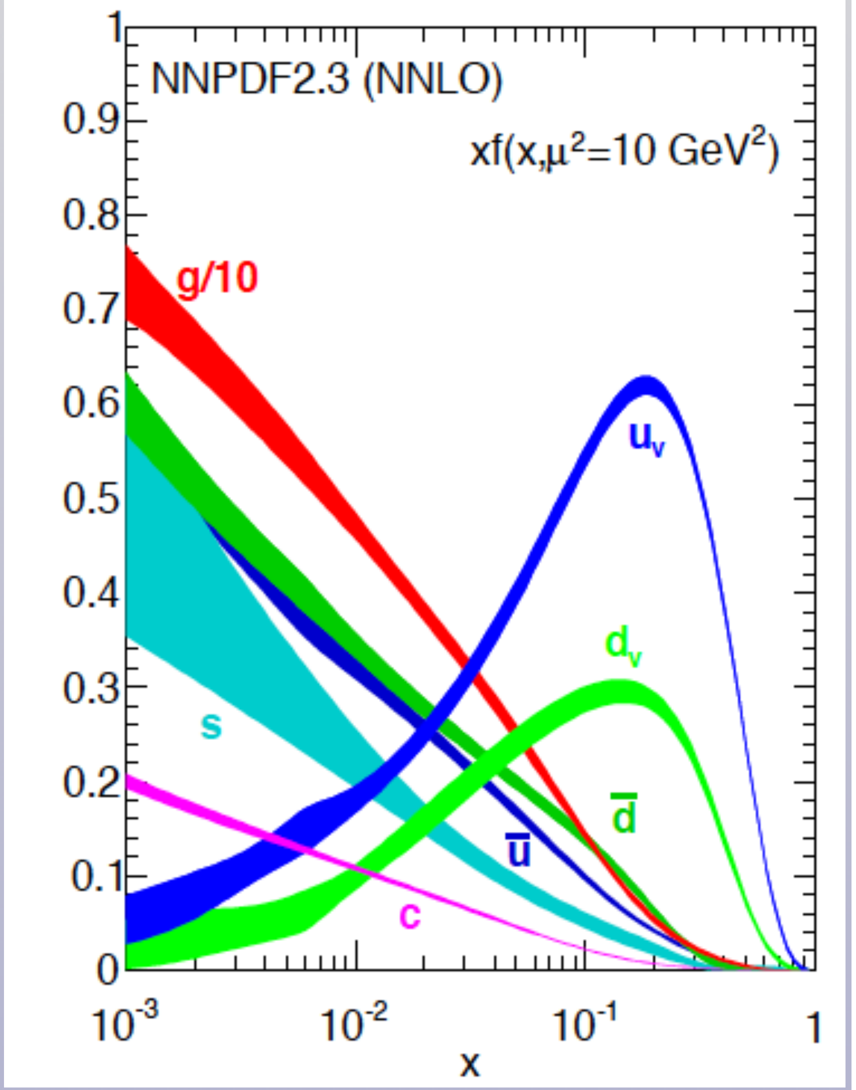
Hadron-level cross section (2) Parton Distributions Parton-level cross-section

- To make sense of **LHC collisions**, we need first of all to **determine the parton distributions of the proton** with good precision!



Parton Distributions

- There is one independent PDF **for each parton in the proton**: **u(x,Q^2), d(x,Q^2), g(x,Q^2)**, …
- A total of **13 PDFs,** but **heavy quark PDFs generated radiatively** from gluon and light quarks
- At Leading Order, PDFs understood as the **probability of finding a parton of a given flavor that carries a fraction x** of the total proton's momentum
- Once QCD corrections included, PDFs become scheme-dependent and have no probabilistic interpretation
- **Shape and normalization** of PDFs are very different for each flavor, reflecting the different underlying **dynamics** that determine each PDF flavor
- QCD imposes **valence** and **momentum** sum rules **valid to all orders** in perturbation theory



PDG Review 2014

Momentum Sum Rule

$$\int_0^1 dx\ x\left[\Sigma(x)+g(x)\right]=1$$

Valence Sum Rules

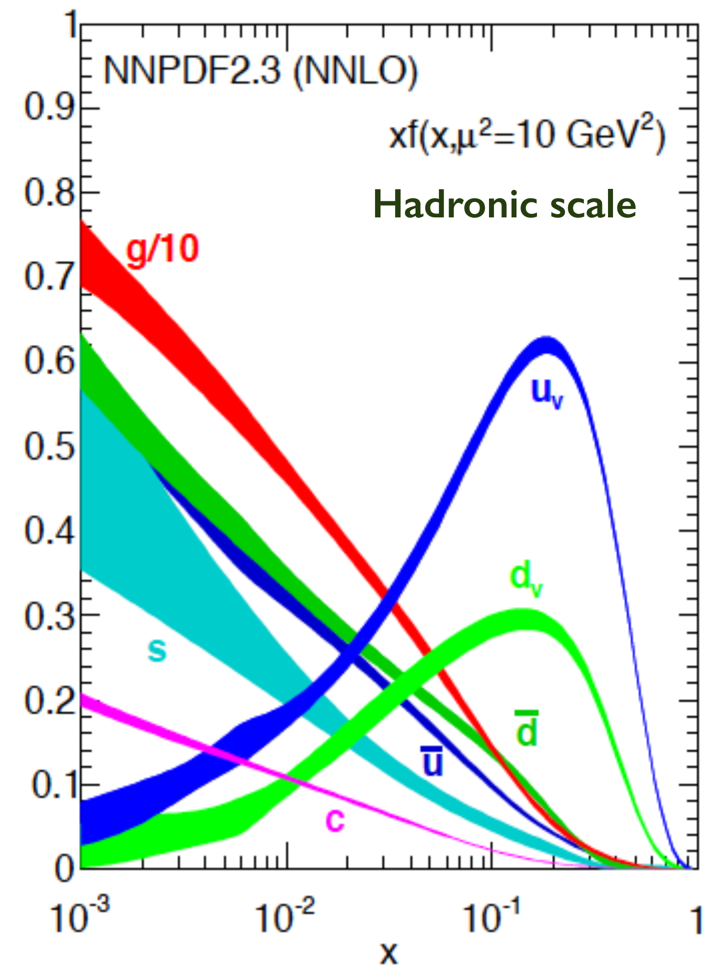
$$\int_0^1 dx\ \left(u(x)-\bar{u}(x)\right)=2\ ,\qquad \int_0^1 dx\ \left(d(x)-\bar{d}(x)\right)=1$$

Perturbative evolution equations

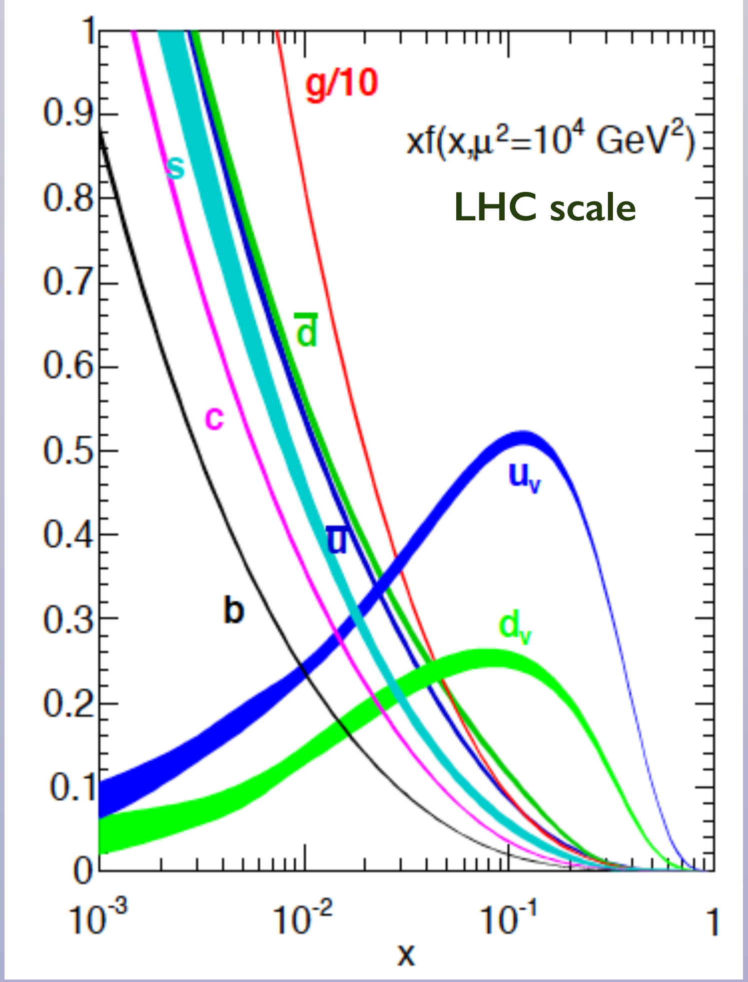
- The dependence of PDFs on **Bjorken-x (momentum fraction) is determined by non-perturbative QCD dynamics**, but that on the **scale Q² (resolution)** is instead known from perturbative QCD: the **DGLAP evolution equations**

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

- Once *x*-dependence **$q(x,Q^2_0)$** extracted from data, **pQCD determines PDFs** at other scales **$q(x,Q^2)$**



Perturbative Evolution



The Neural Network Approach to Parton Distributions

The NNPDF approach

- The **limitations of available PDF sets circa 2005**, and the requirements of **precision physics at the upcoming LHC,** prompted us to develop a completely novel approach to PDF determination

- PDF sets typically based on **restrictive functional forms** leading to strong theoretical bias

$$g(x,Q_0^2) \quad = \quad A_g(1-x)^{m_\Sigma} x^{-n_\Sigma} \left(1 + a_g\sqrt{x} + b_g x + \dots\right)$$

☑ NNPDF solution: use **artificial neural networks** as universal unbiased interpolants

$$g(x,Q_0^2) \quad = \quad A_g(1-x)^{m_\Sigma} x^{-n_\Sigma} \mathrm{NN}_g(x)$$

- PDF sets often rely on the the **Gaussian/linear approximation** for error estimation and propagation

$$F_0 = F(S_0), \qquad \sigma_F = \sqrt{\sum_{i=1}^{N_{\rm par}} \left[F(S_i) - F(S_0)\right]^2}.$$

☑ NNPDF solution: Use the **Monte Carlo method** to create a probability distribution in the space of PDFs

$$F_{I,p}^{({\rm art})(k)} = S_{p,N}^{(k)} F_{I,p}^{({\rm exp})} \left(1 + \sum_{l=1}^{N_c} r_{p,l}^{(k)} \sigma_{p,l} + r_p^{(k)} \sigma_{p,s}\right) \quad , \quad k = 1, \dots, N_{\rm rep}$$

Consistent error propagation to LHC xsecs
no Gaussian assumptions

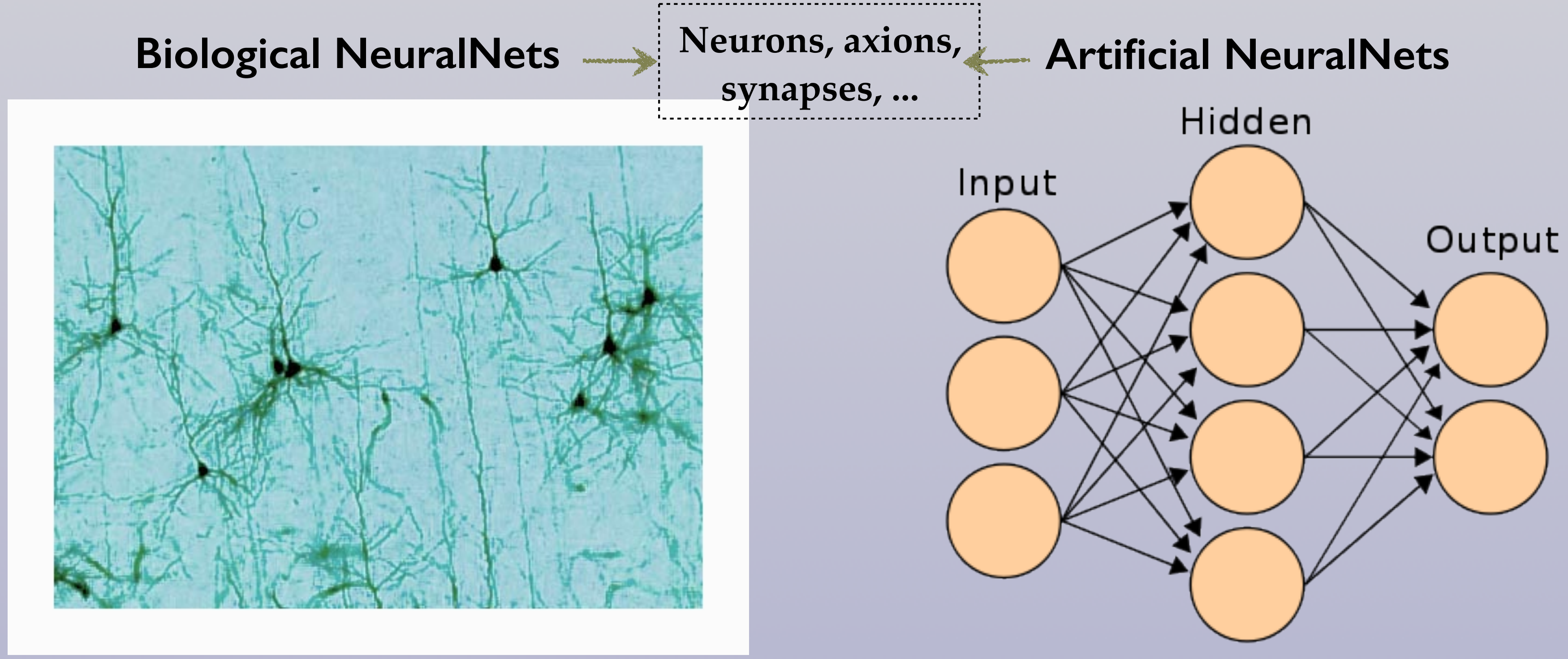
- Traditional PDF analyses based on **deterministic minimisation** of the χ^2 to reach convergence in the fit

☑ NNPDF solution: Use **Genetic Algorithms** to be able to explore efficiently the vast parameter space

$$\chi^2 \; = \; \sum_{i=1}^{N_{\rm dat}} \sum_{i'=1}^{N_{\rm dat}} (D_i - T_i) \left(V^{-1}\right)_{ii'} (D_{i'} - T_{i'}).$$

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**



Biological NeuralNets → Neurons, axions, synapses, ... ← **Artificial NeuralNets**

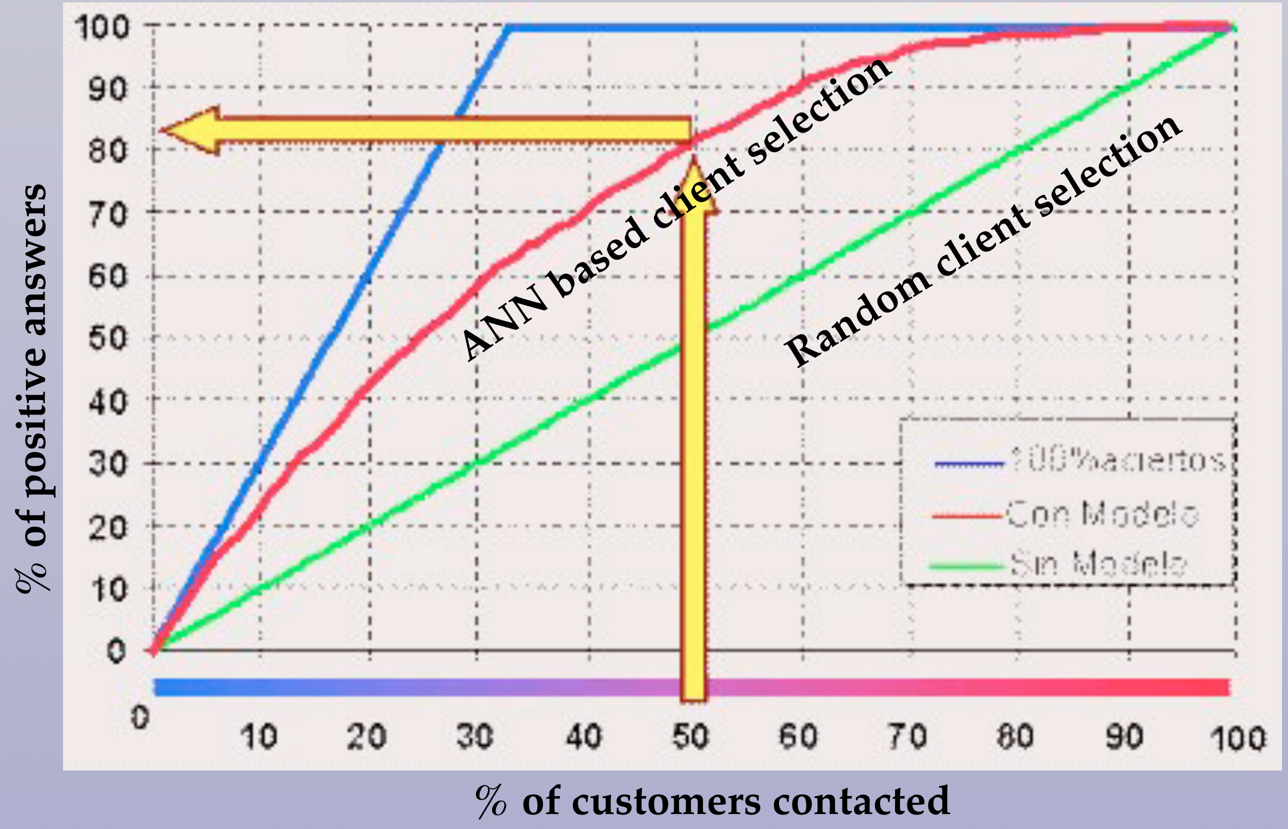
- Artificial neural networks aimed to excel in the same domains as their biological counterparts: **pattern recognition, forecasting, classification**, ... where our **evolution-driven biology** outperforms traditional algorithms

Artificial Neural Networks

Example 1: **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** (sex, income, loans) and **their output** (yes/no) and use the information to contact only clients likely to accept the offer

Cost-effective method to improve marketing performance



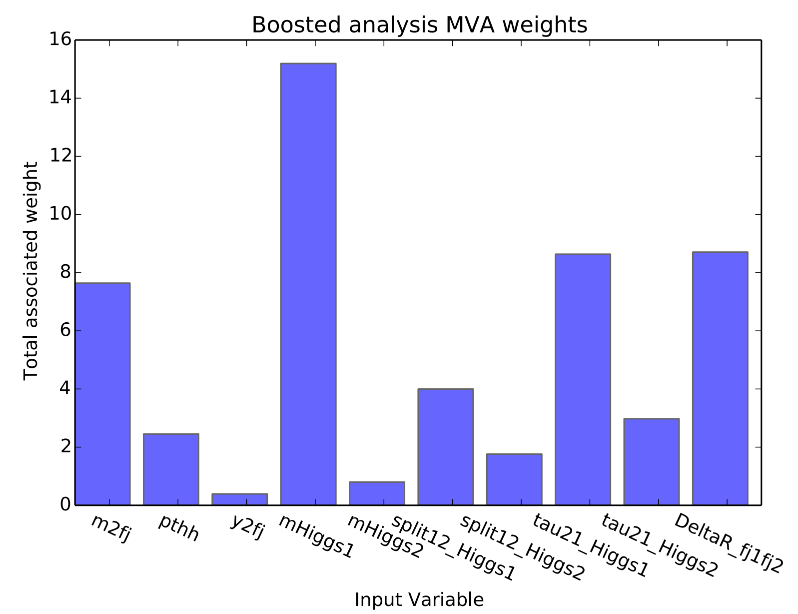
Artificial Neural Networks

Example 2: **Classification.** Discriminate between signal and background events in complicated final states

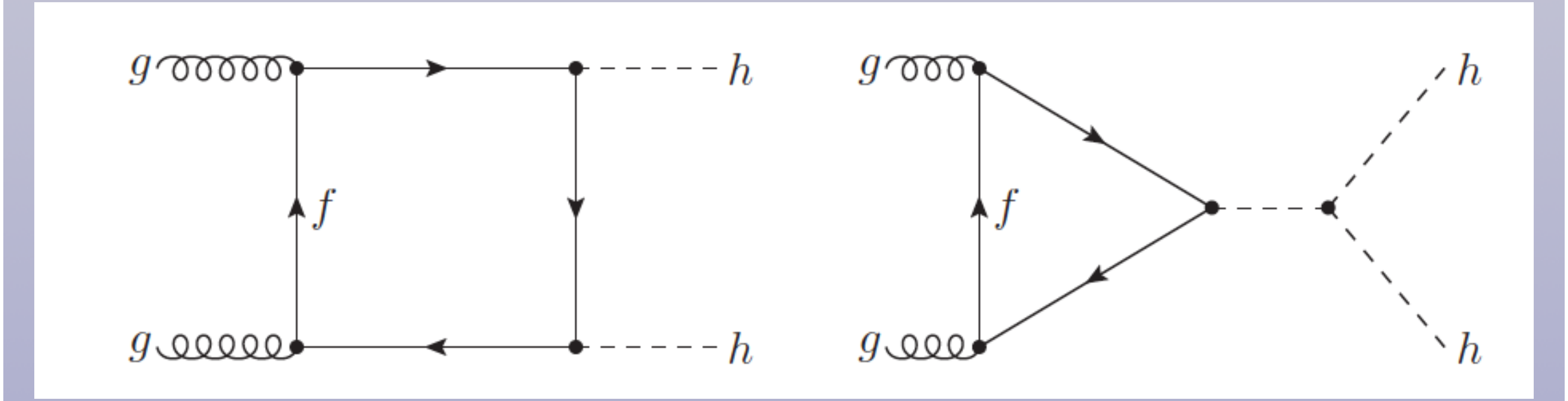
- **Improve S/√B as compared to cut-based analyses**
- **Identify automatically** the kinematical variables with most discrimination power

Redundancy of **NN-based Multivariate Analysis** guarantees the optimisation of signal/background separation

Behr, Bortoletto, Frost, Hartland, Issever, JR, in prep



HH->4b feasibility study



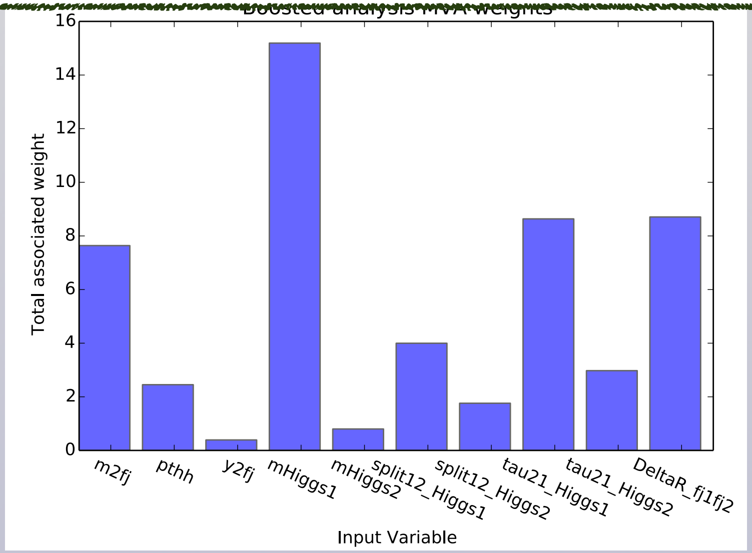
Artificial Neural Networks

Example 2: **Classification.** Discriminate between signal and background events in complicated final states

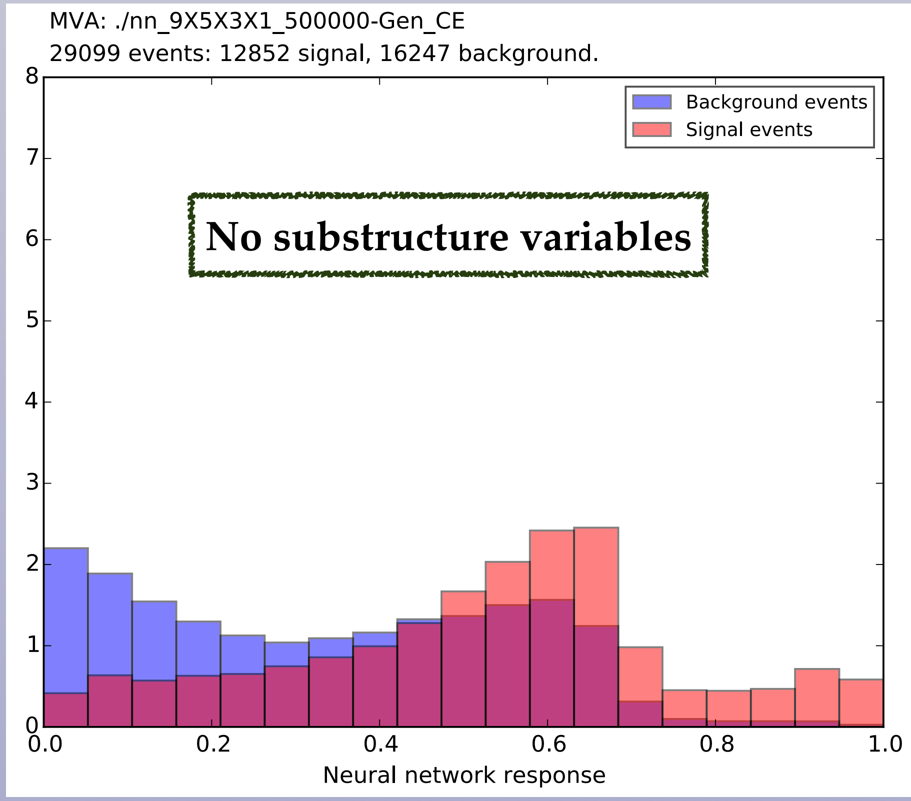
- **Improve S/√B as compared to cut-based analyses**
- **Identify automatically** the kinematical variables with most discrimination power

Redundancy of **NN-based Multivariate Analysis** guarantees the optimisation of signal/background separation

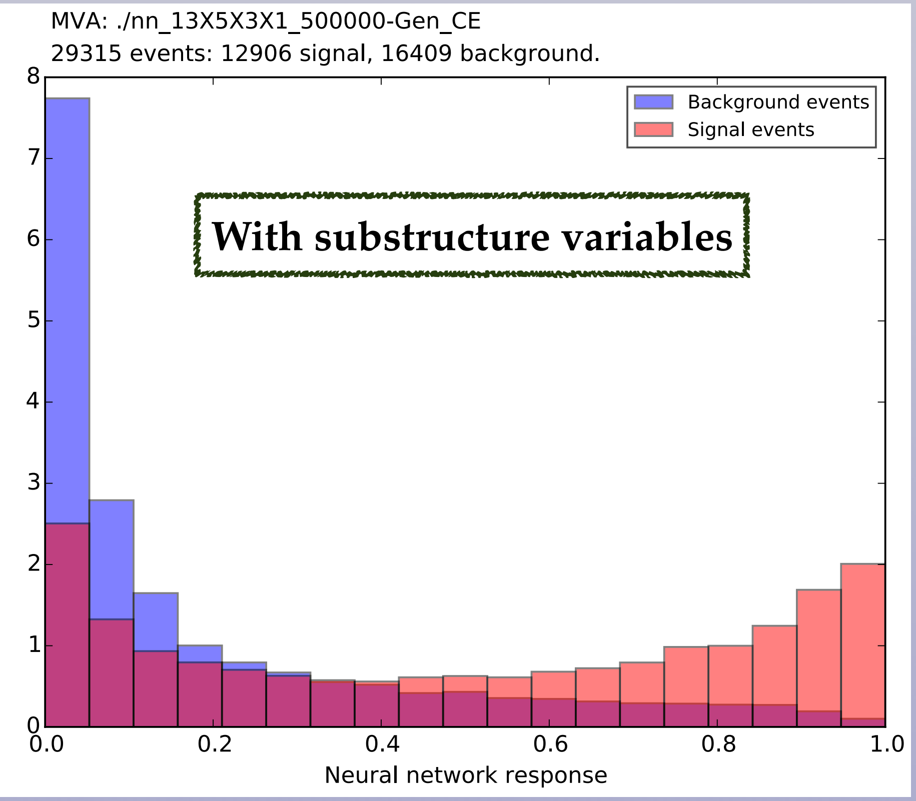
Behr, Bortoletto, Frost, Hartland, Issever, JR, in prep



HH->4b feasibility study



No substructure variables



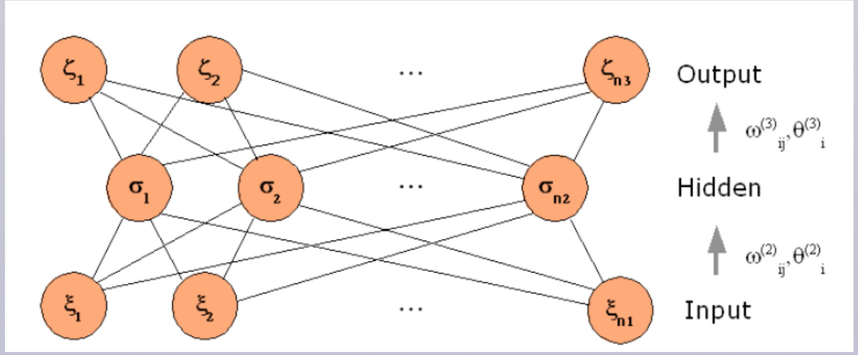
With substructure variables

Artificial Neural Networks

- **Artificial Neural Networks** (ANNs) provide **universal unbiased interpolants** to parametrize PDFs at low input scales

$$\begin{aligned} \Sigma(x,Q_0^2) &= (1-x)^{m_\Sigma} x^{-n_\Sigma} \mathrm{NN}_\Sigma(x) \\ g(x,Q_0^2) &= A_g (1-x)^{m_\Sigma} x^{-n_\Sigma} \mathrm{NN}_g(x) \end{aligned}$$

- The ANN class that we adopt are **feed-forward multilayer neural networks** (perceptrons)



$$\xi_i^{(l)} = g\left(h_i^{(l)}\right) \ , \quad i = 1,\ldots,n_l \ , \quad l = 2,\ldots,L$$

$$h_i^{(l)} = \sum_{j=1}^{n_{l-1}} \omega_{ij}^{(l)} \xi_j^{(l-1)} - \theta_i$$

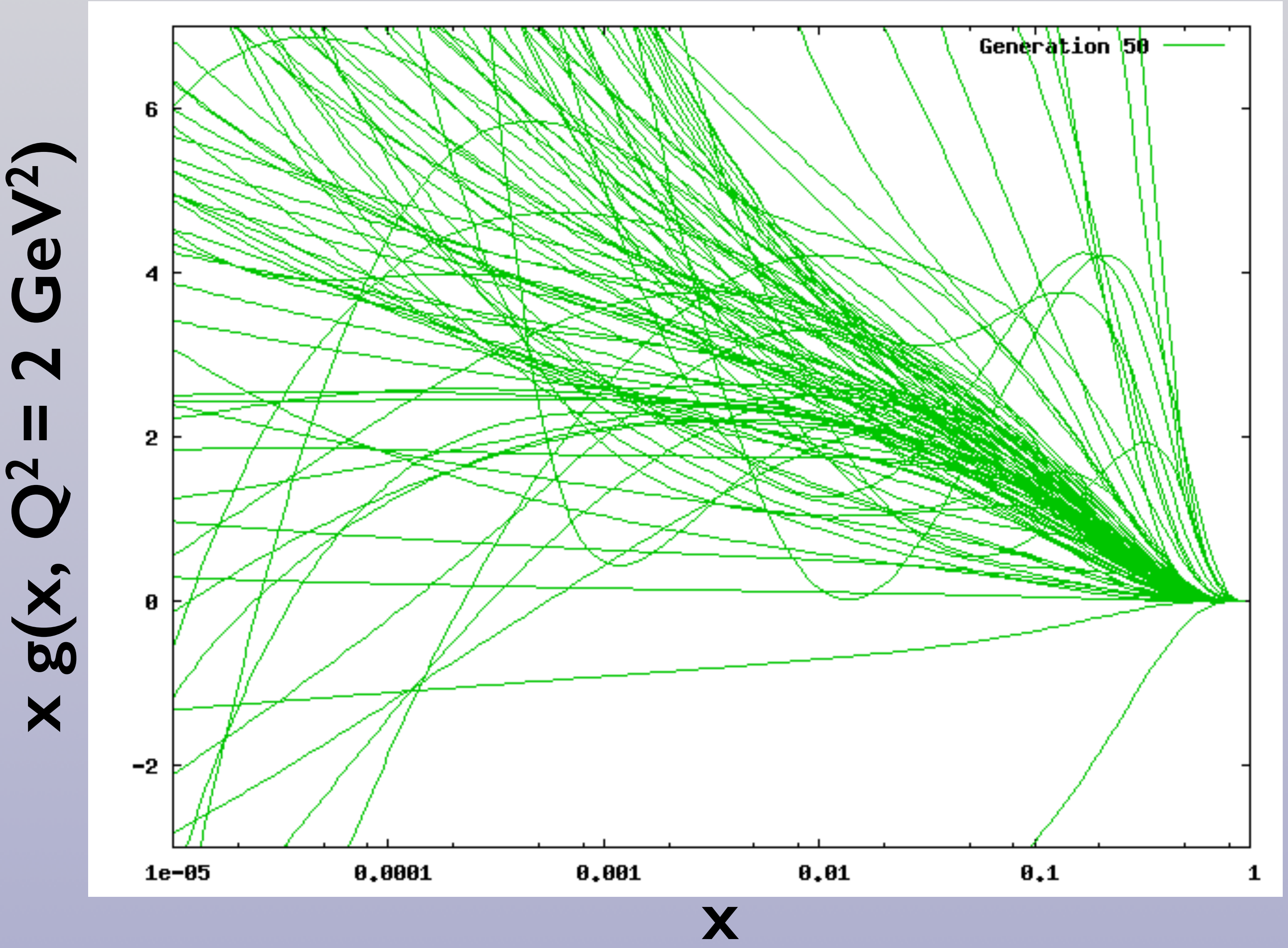
- In traditional PDF determinations, the input *ansatz* is a simple **polynomial**

$$\begin{aligned} \Sigma(x,Q_0^2) &= (1-x)^{m_\Sigma} x^{-n_\Sigma} \left(1 + a_\Sigma \sqrt{x} + b_\Sigma x + \ldots\right) \ , \\ g(x,Q_0^2) &= A_g (1-x)^{m_\Sigma} x^{-n_\Sigma} \left(1 + a_g \sqrt{x} + b_g x + \ldots\right) \end{aligned}$$

- The use of Artificial Neural Networks allows:
  - **No theory bias** introduced in the PDF determination by the choice of *ad-hoc* functional forms
  - The use of very flexible parametrizations for all PDFs - regardless of the dataset used. The NNPDF analysis allow for **O(400) free parameters**, to be compared with **O(10-20) in traditional PDFs**
  - **Faithful extrapolation**: PDF uncertainties **blow up** in regions with scarce experimental data

PDF Replica Neural Network Learning

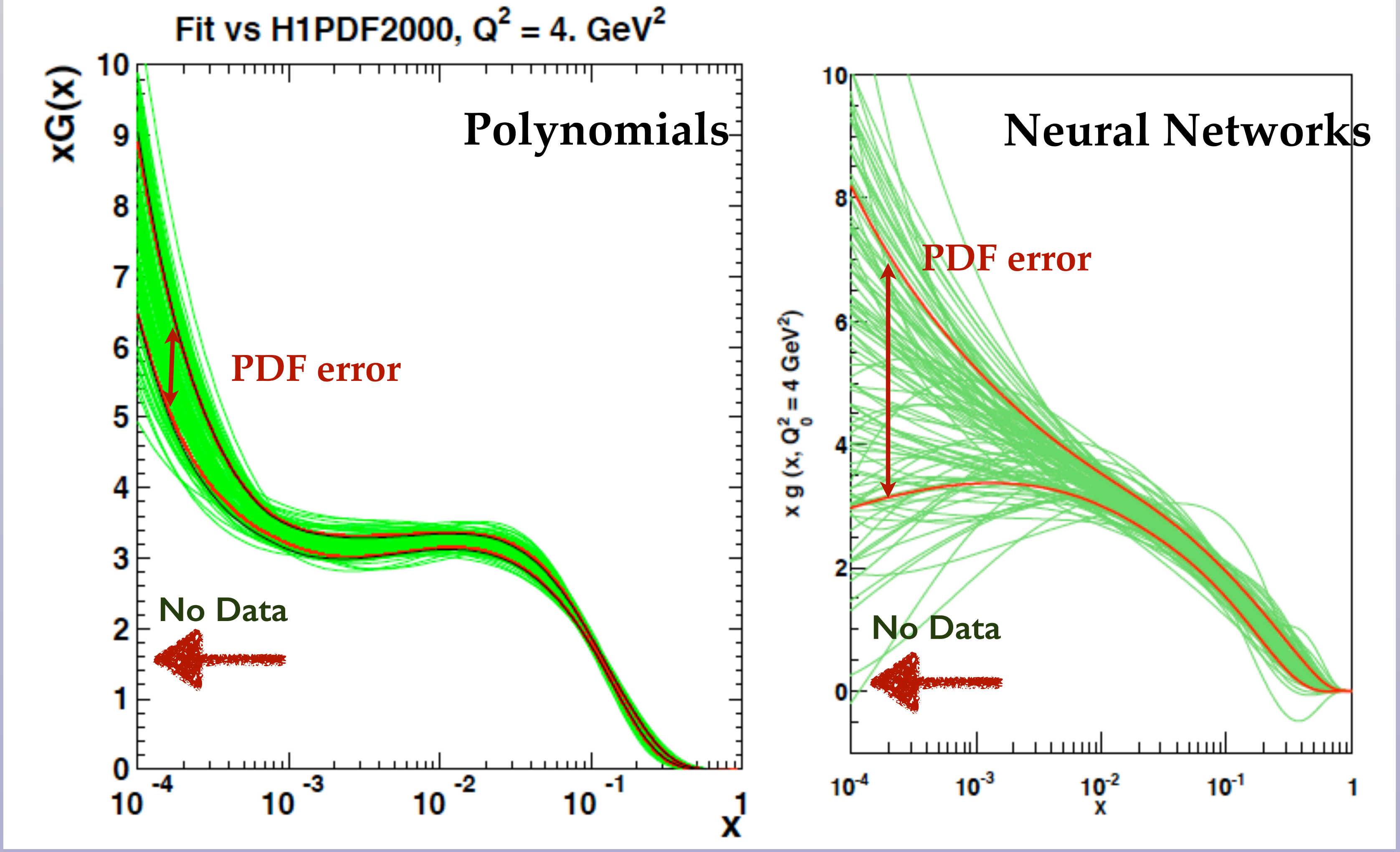
The minimisation of the **data vs theory χ^2** is performed using **Genetic Algorithms**
Each **green curve** corresponds to a **gluon PDF Monte Carlo** replica



Artificial Neural Networks vs. Polynomials

- Compare a **benchmark PDF analysis** where **the same dataset** is fitted with **Artificial Neural Networks** and with **standard polynomials** (everything else identical)

- **ANN** avoid biasing the PDFs, **faithful extrapolation at small-x** (very few data, thus error blow up)

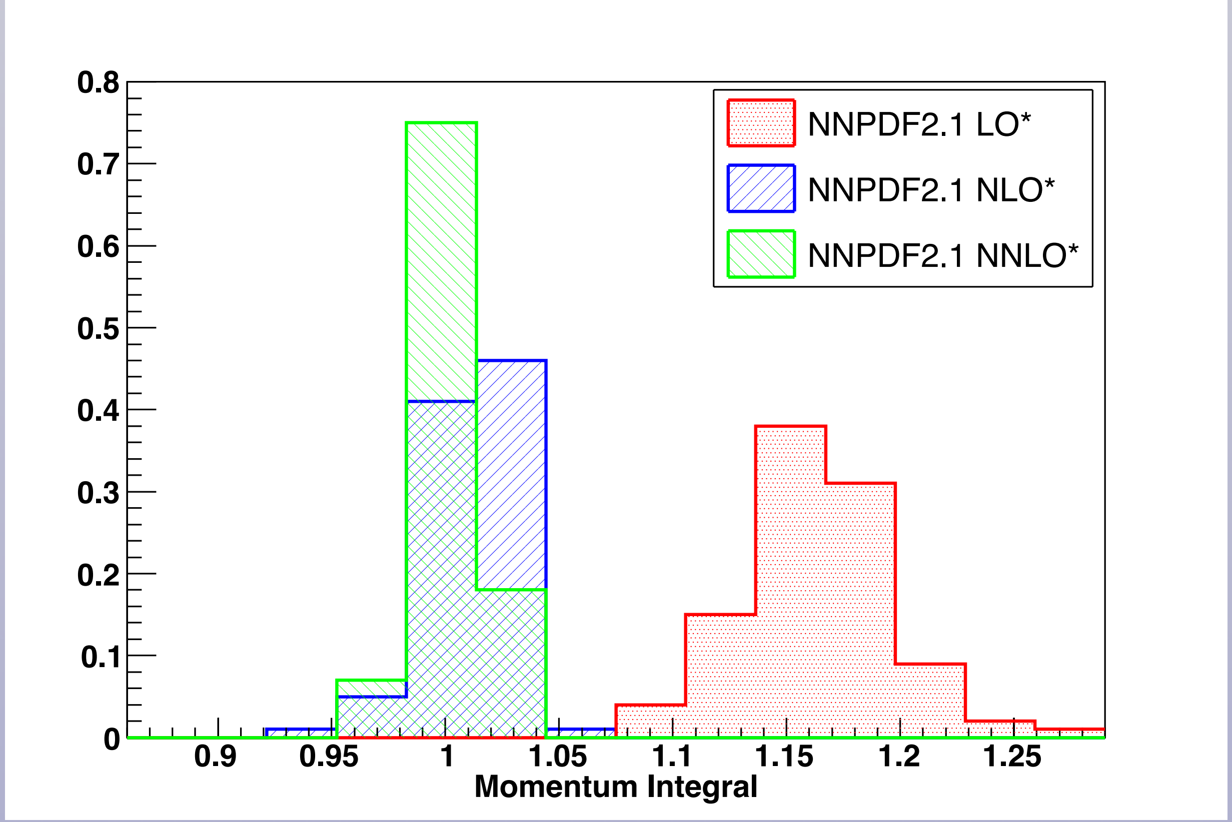


Precision tests of the Factorisation Theorem

- Perturbative QCD requires that the **momentum integral** should be unity to all orders

$$[M]\left(Q^2\right) \equiv \int_0^1 dx \left(xg\left(x,Q^2\right) + x\Sigma\left(x,Q^2\right)\right)$$

- Is it possible to **determine** the value of the momentum integral from the global PDF analysis, rather than **imposing it?** Check in LO*, NLO* and NNLO* fits **without setting M=1**



$$\begin{aligned}[M]_{\rm LO} &= 1.161 \pm 0.032\,,\\ [M]_{\rm NLO} &= 1.011 \pm 0.018\,,\\ [M]_{\rm NNLO} &= 1.002 \pm 0.014\,.\end{aligned}$$

- Experimental data beautifully **confirms the pQCD expectation**
- **Extremely non trivial test** of the global analysis framework and the **factorization hypotheses**
- Very good convergence of the QCD perturbative expansion

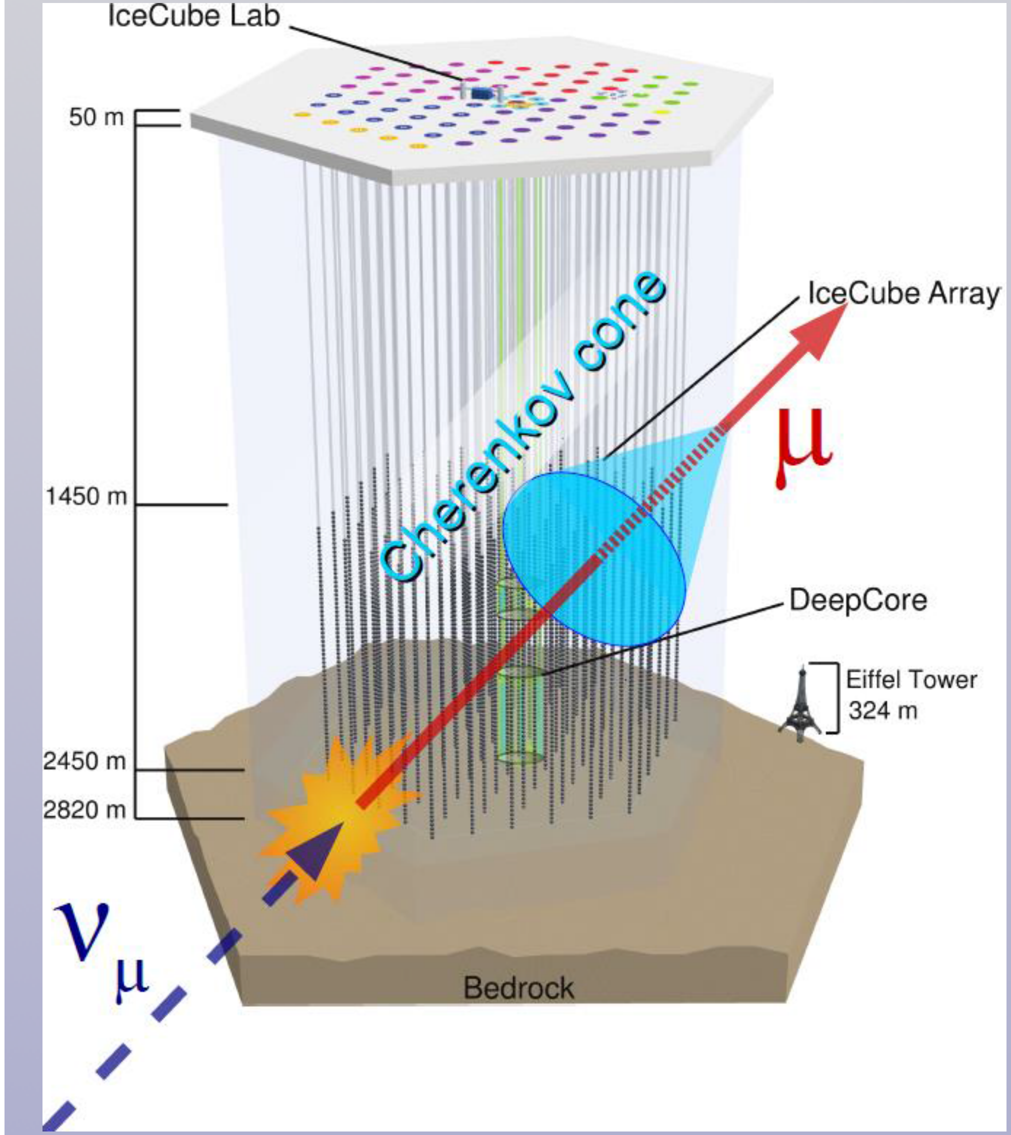
NNPDF, 11

From LHC measurements to neutrino telescopes

Gauld, JR, Rottoli, Talbert, arXiv:1506.08025
Gauld, JR, Rottoli, Sarkar, Talbert, arXiv:1511.aaaaa

From LHC to IceCube

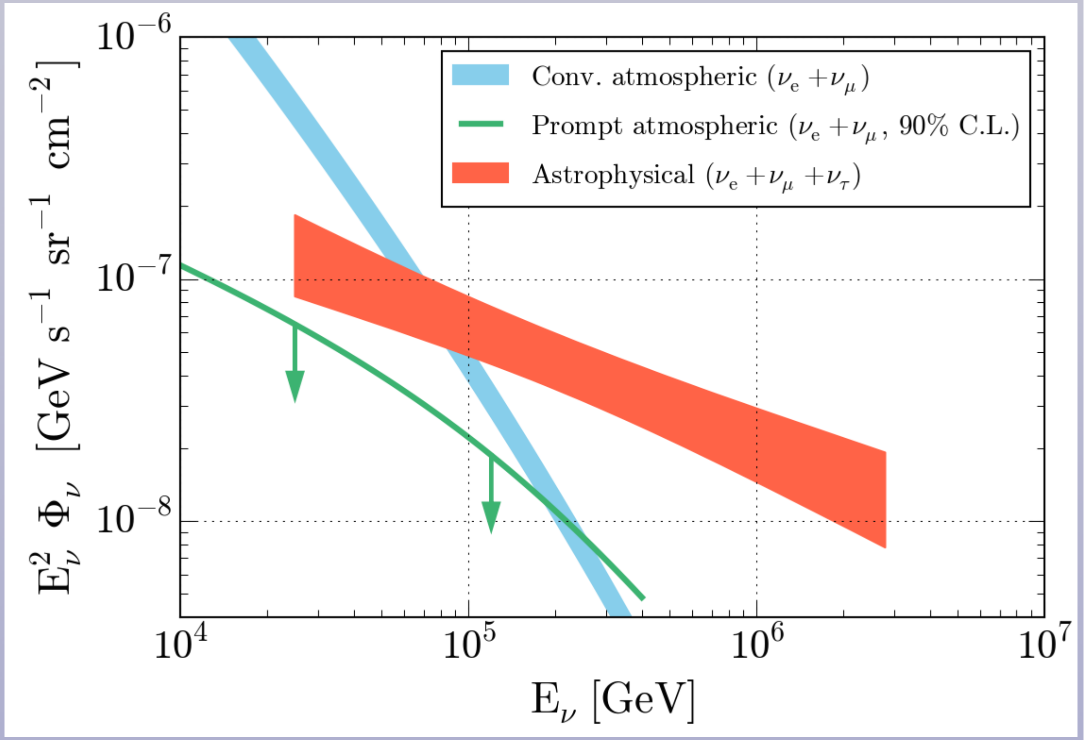
ICECUBE
SOUTH POLE NEUTRINO OBSERVATORY



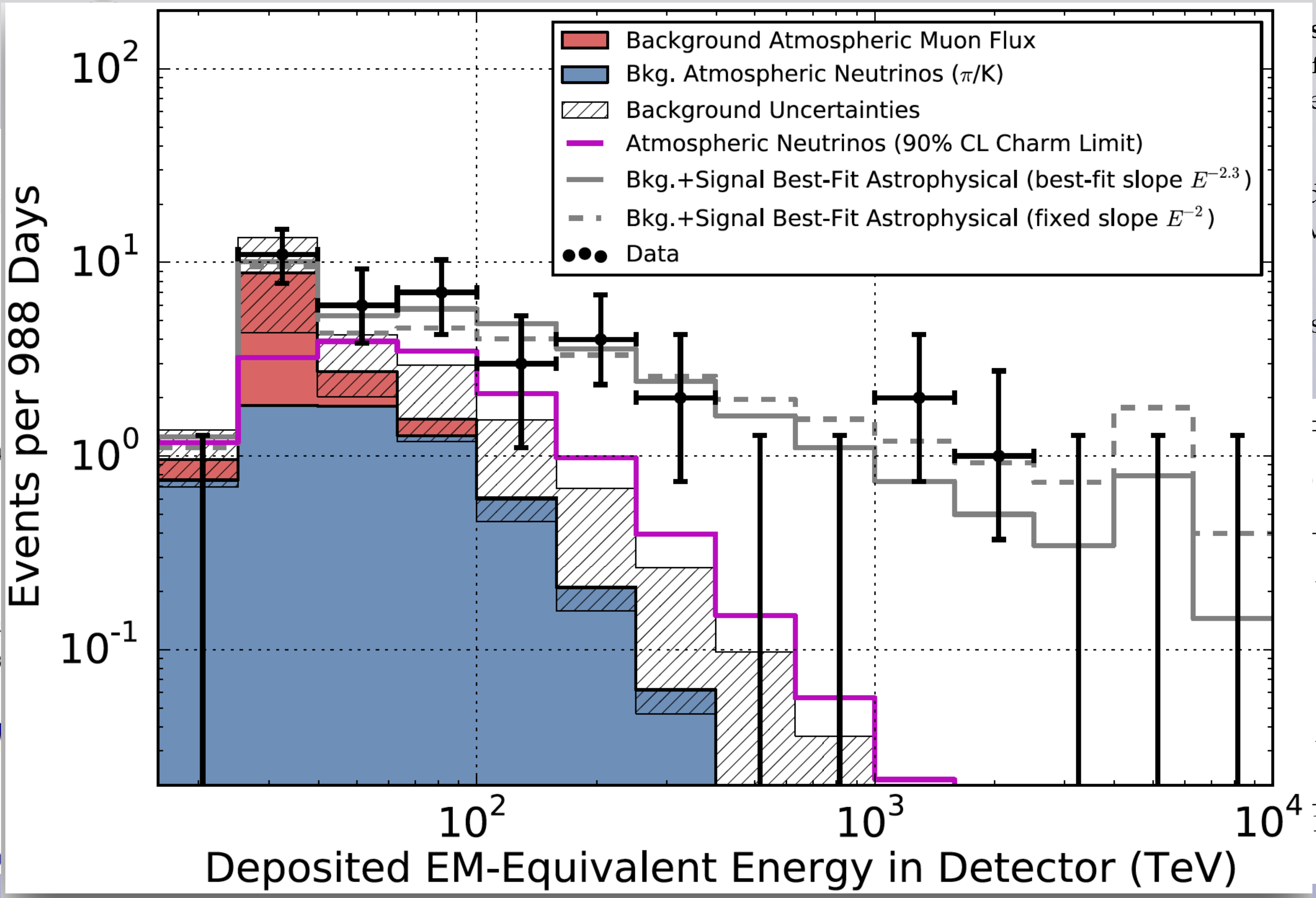
- The main **background for astrophysical neutrinos at IceCube** is the flux of neutrinos from the **decays of charm mesons** in cosmic ray collisions in the atmosphere

- Theoretically, this **prompt neutrino flux** is affected by large uncertainties: very small-x PDFs, very low scales - can pQCD be applied?

- Strategy: use **LHC data itself** to pin down this prompt flux!



From LHC to IceCube



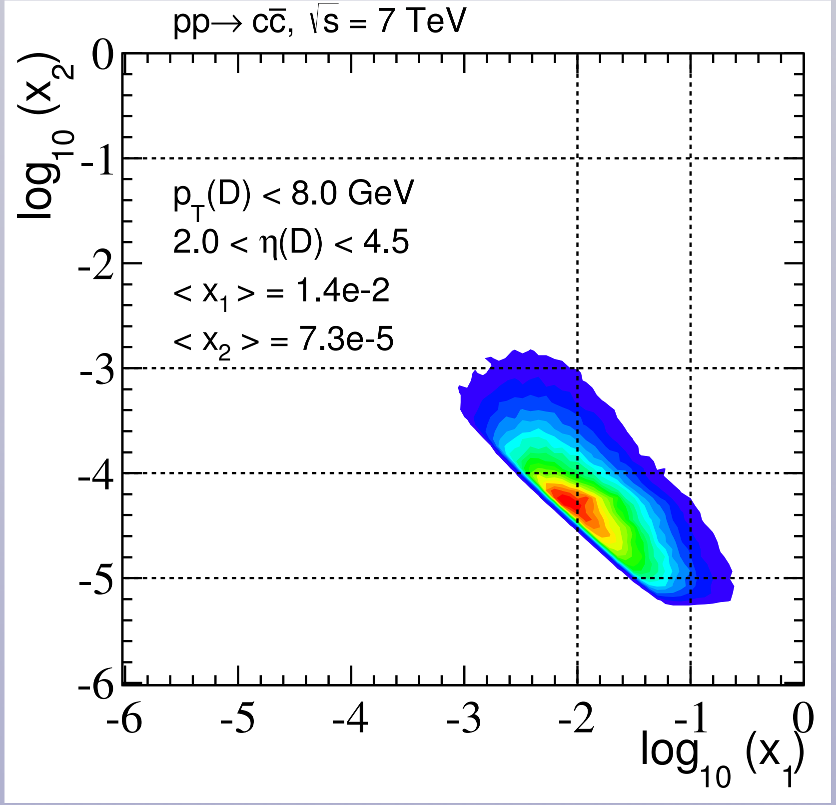
Charm production and the small-x gluon

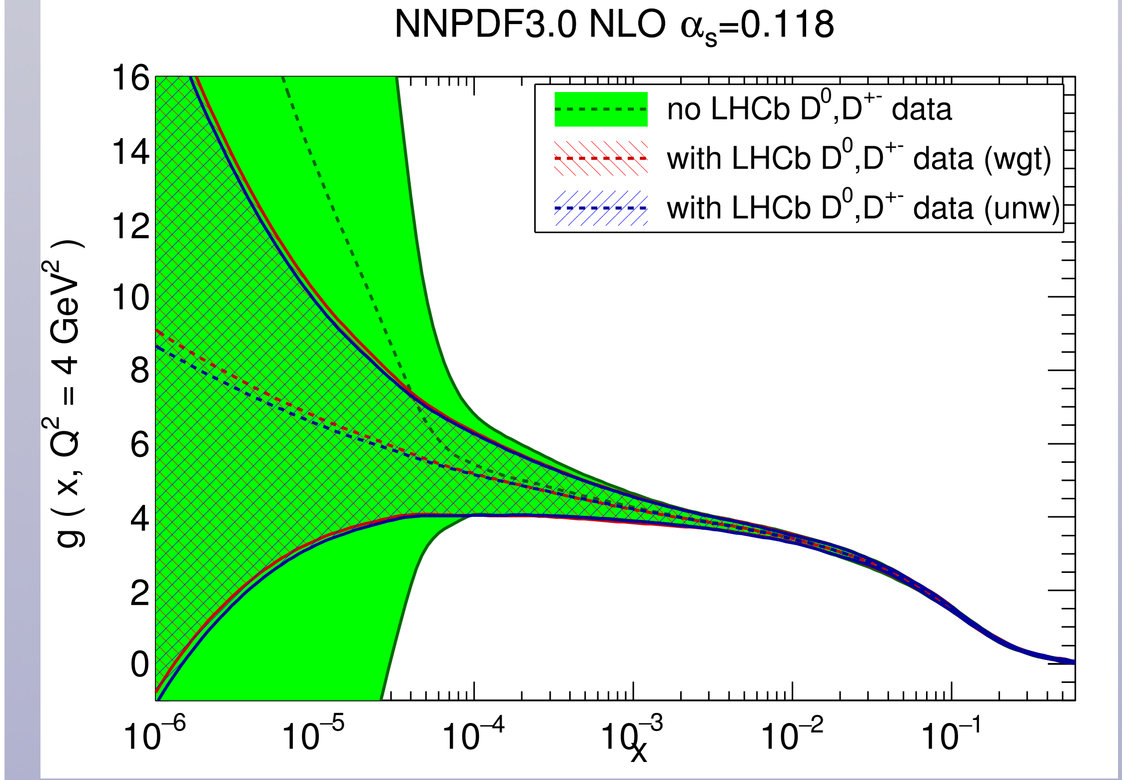
- The production of **charm and bottom mesons in the forward region** is directly sensitive to the **small-x gluon**, where PDF uncertainties are huge from lack of direct constraints

- Using the FONLL calculation and **normalised LHCb 7 TeV D meson data**, we have included these measurements in **NNPDF3.0 NLO** and found a substantial reduction of PDF errors

- Semi-analytical **FONLL** results validated with **POWHEG** and **aMC@NLO** calculations

- Important implications for the calculations of **charm-induced prompt neutrino fluxes at IceCube**

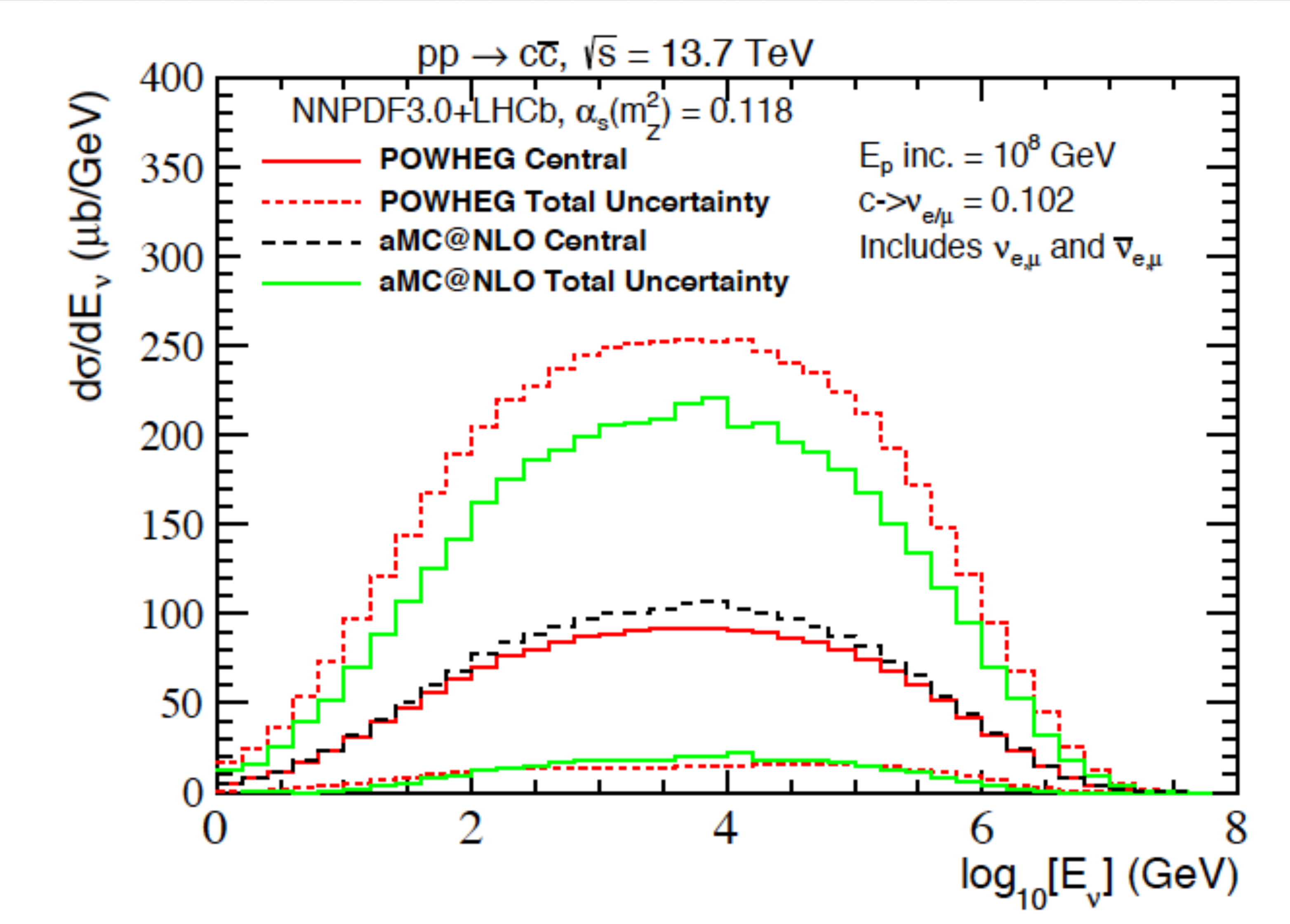




Charm production and the small-x gluon

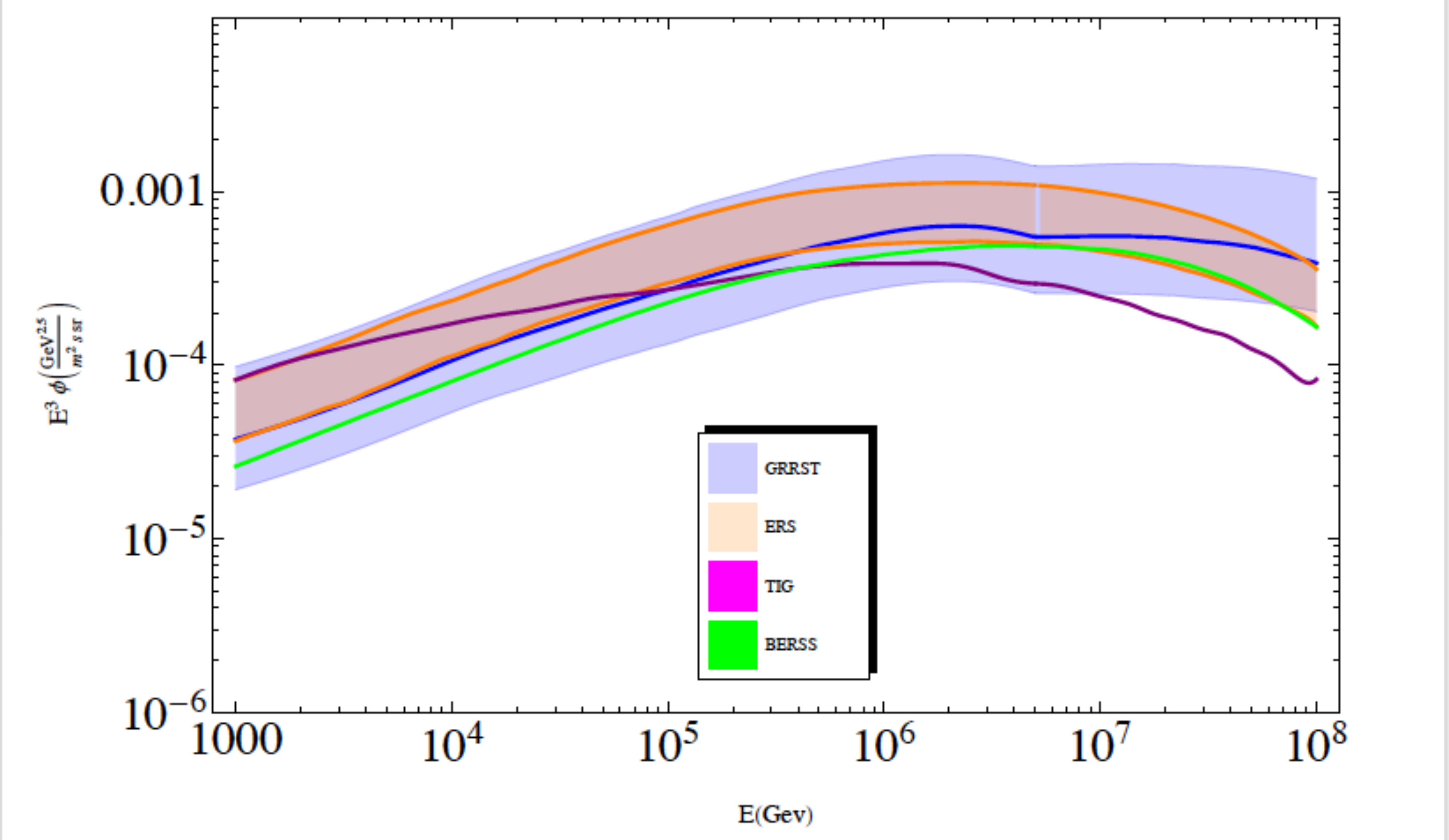
- **Prompt neutrinos in the PeV region** (highest neutrino elevents in IceCube) arise from cosmic rays of energies of 100-1000 PeV, c**orresponding to a center-of-mass energy of 14 TeV**

- **Very good overlap between the LHCb and IceCube kinematics**



Charm production and the small-x gluon

Preliminary results for the **expected prompt lepton fluxes at IceCube** in our framework (GRRST) compared with previous calculations. Final results to be published soon.

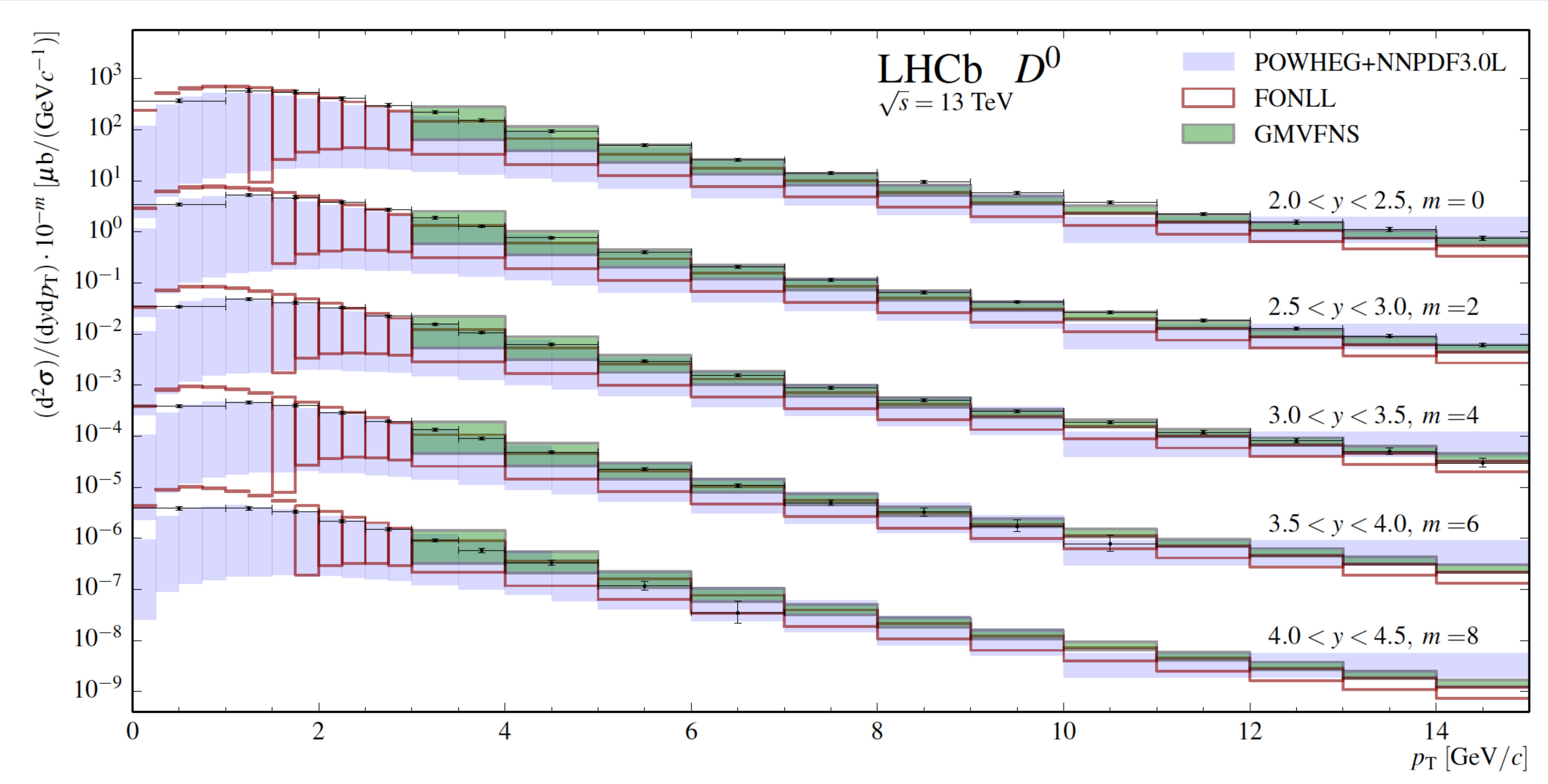


PRELIMINARY-SCALE UNCERTAINTY ONLY****

Comparison with LHCb 13 TeV data

- Using the **improved small-x PDFs with 7 TeV LHCb charm data**, we also provided predictions for 13 TeV, which have been published just today!

- **Good consistency** between our **predictions** (POWHEG+NNPDF3.0L) and the **LHCb 13 TeV** data within theory uncertainties, which further strengthens the robustness of our approach



Updated squark and gluino cross-sections with threshold-improved PDFs

Bonvini, Marzani, JR, Rottoli, Ubiali, Ball, Bertone, Carrazza, Hartland, arXiv:1507.01006
Beenakker, Borschensky, Kramer, Kulesza, Laenen, Marzani, JR, arXiv:1510.00375

Why threshold resummation?

☑ The basic idea of threshold resummation methods is simple. Start from a **factorised cross-section** and transform it to **Mellin (conjugate) space**

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

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

☑ Then, using different techniques, we can computed a **resummed coefficient function** that includes terms or the type $\alpha_s^k \ln^p N$ (corresponding to $\alpha_s^k \ln^r (1-x)$) to **all orders in perturbation theory**

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

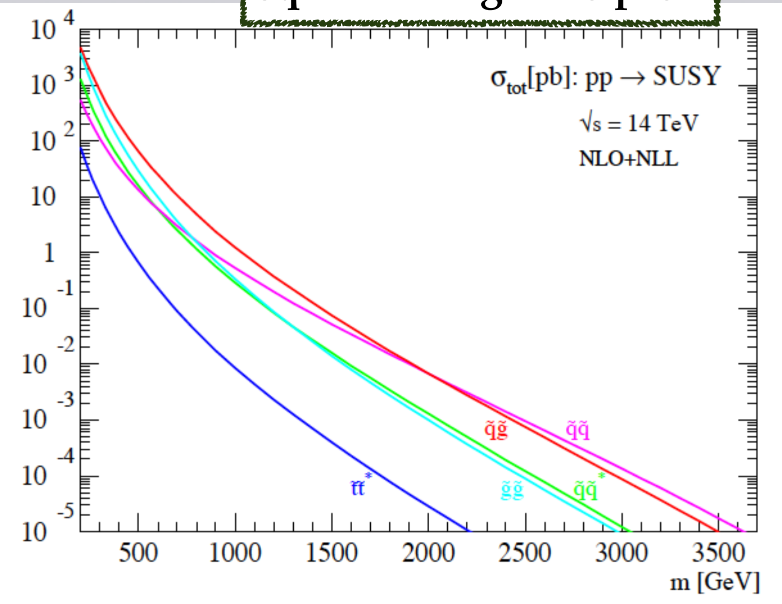
$$\begin{aligned} C^{(N\text{-soft})}(N,\alpha_s) &= g_0(\alpha_s) \exp \mathcal{S}(\ln N,\alpha_s), \\ \mathcal{S}(\ln N,\alpha_s) &= \left[\frac{1}{\alpha_s} g_1(\alpha_s \ln N) + g_2(\alpha_s \ln N) + \alpha_s g_3(\alpha_s \ln N) + \ldots\right] \end{aligned}$$

☑ These terms are numerically large near the **partonic threshold x -> 1 (N -> ∞)**, and thus their resummation **improves the perturbative expansion**, reduces scale uncertainties and allows to **construct approximate higher-order results**

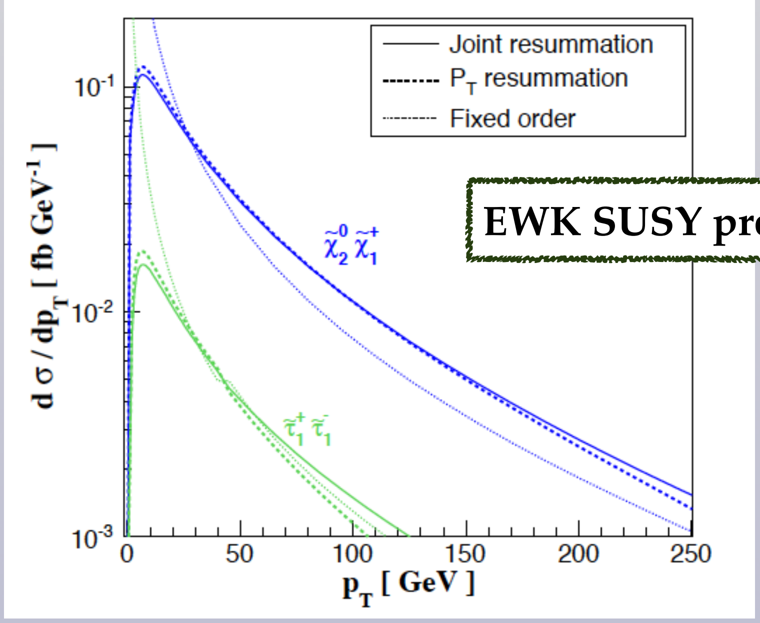
Why threshold resummation?

☑ **Threshold resummation of partonic cross-sections** extensively used in precision LHC pheno

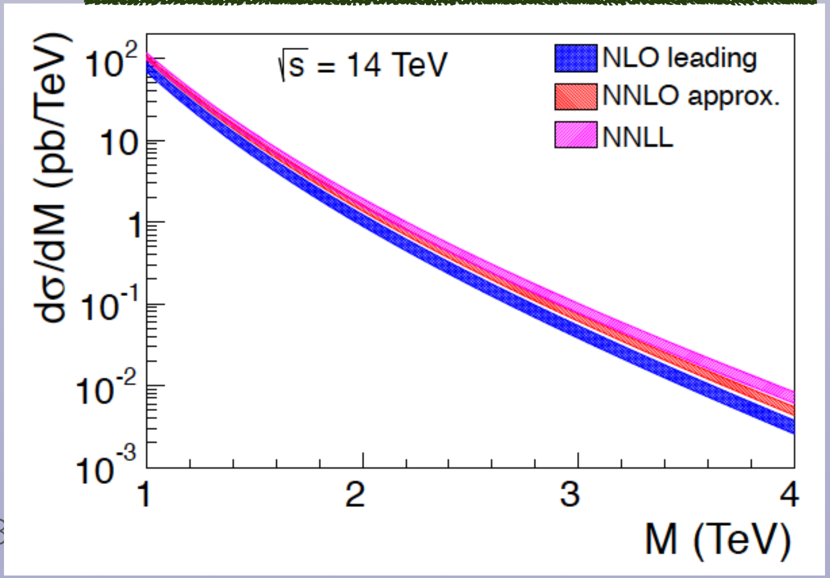
squark and gluino prod



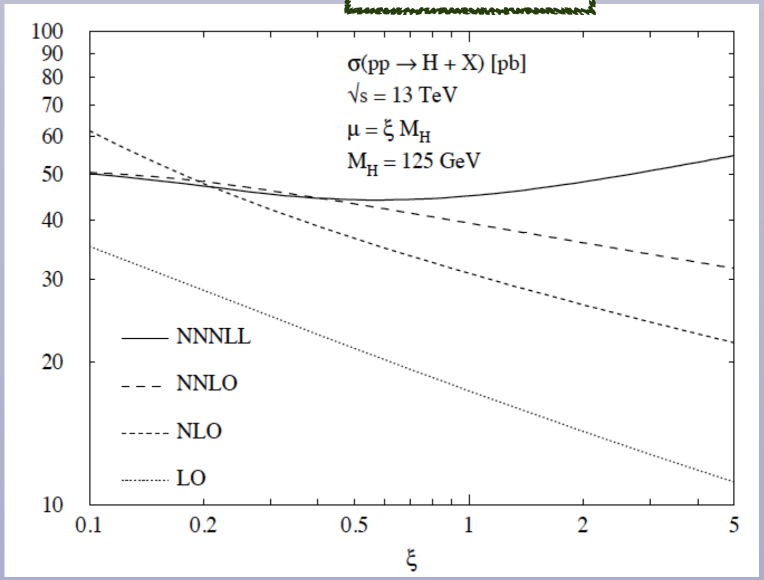
EWK SUSY prod



large mass top quark pair production



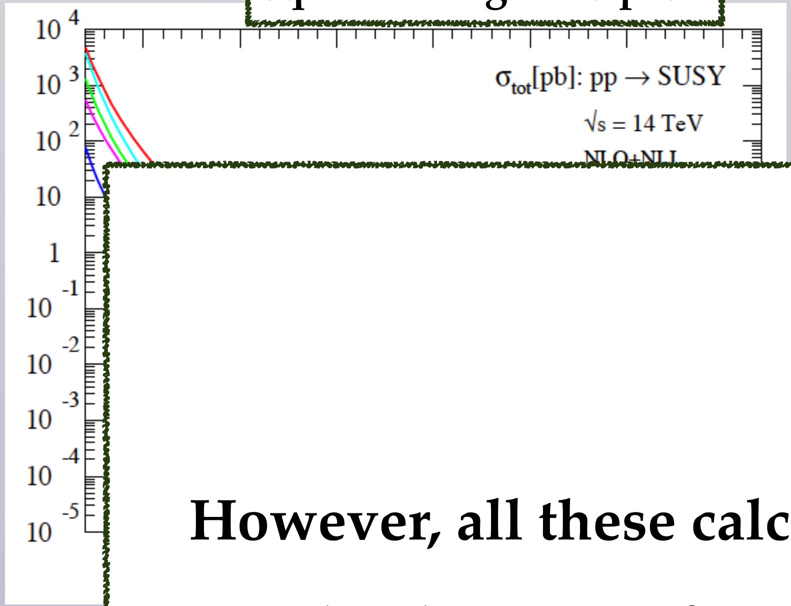
gg -> Higgs

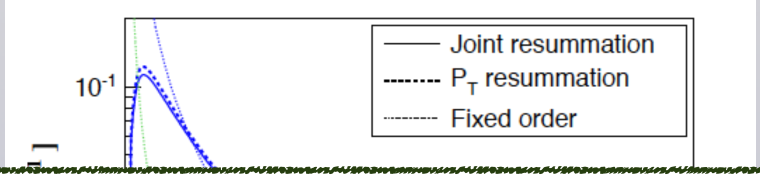


Why threshold resummation?

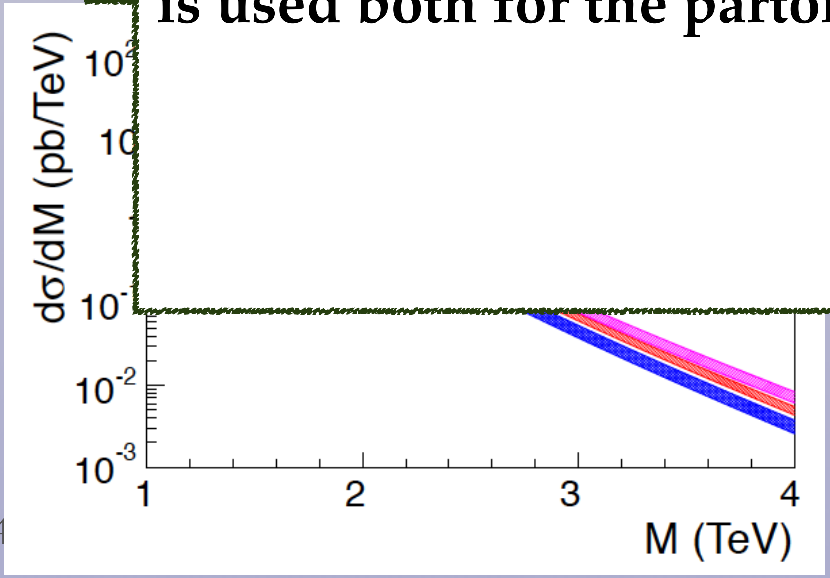
☑ **Threshold resummation of partonic cross-sections** extensively used in precision LHC pheno

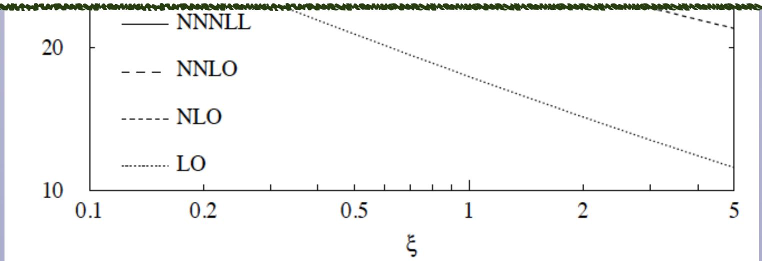
squark and gluino prod





la





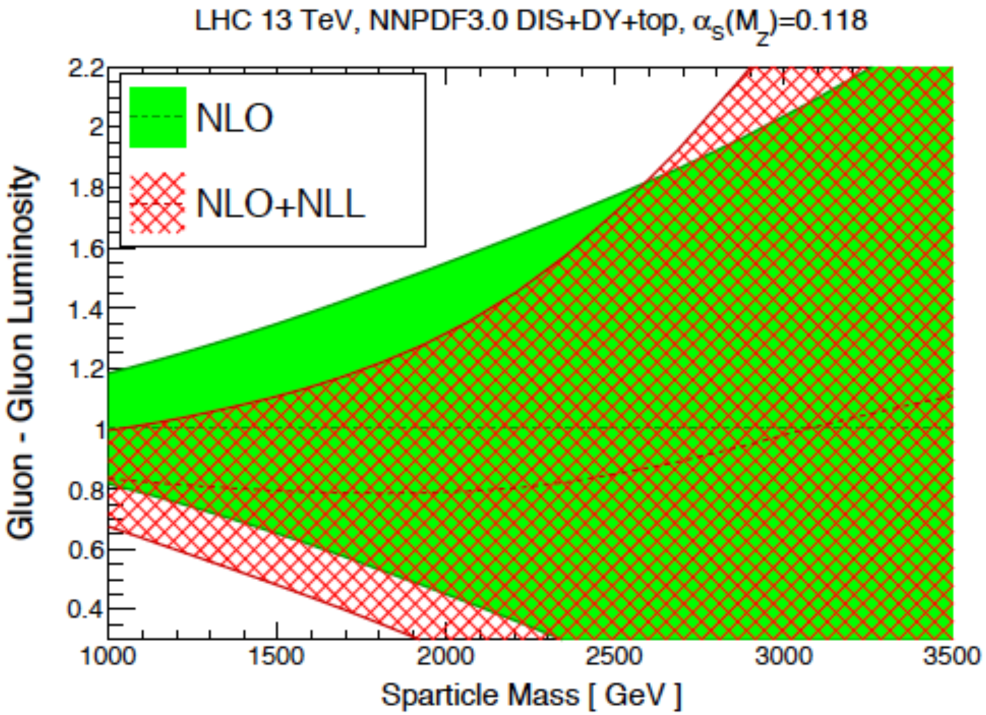
However, all these calculations use fixed-order NLO PDFs!

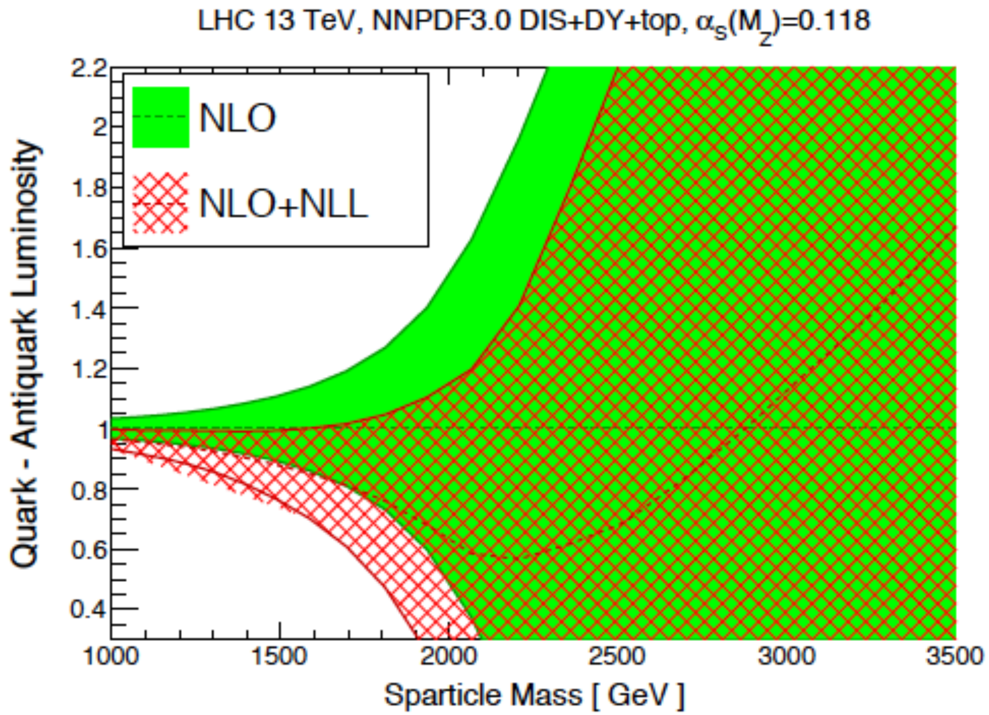
What happens if exactly the same theory (NLO+NLL)

is used both for the partonic matrix elements and for the PDFs?

PDFs with threshold resummation

☑ To determine the relevance of calculations where resummation is included **both** for partonic matrix-elements and the PDFs, we have produced for the first time **threshold-improved PDFs at NLO+NLL and NNLO+NNLL** using a **variant of the NNPDF3.0 fit**



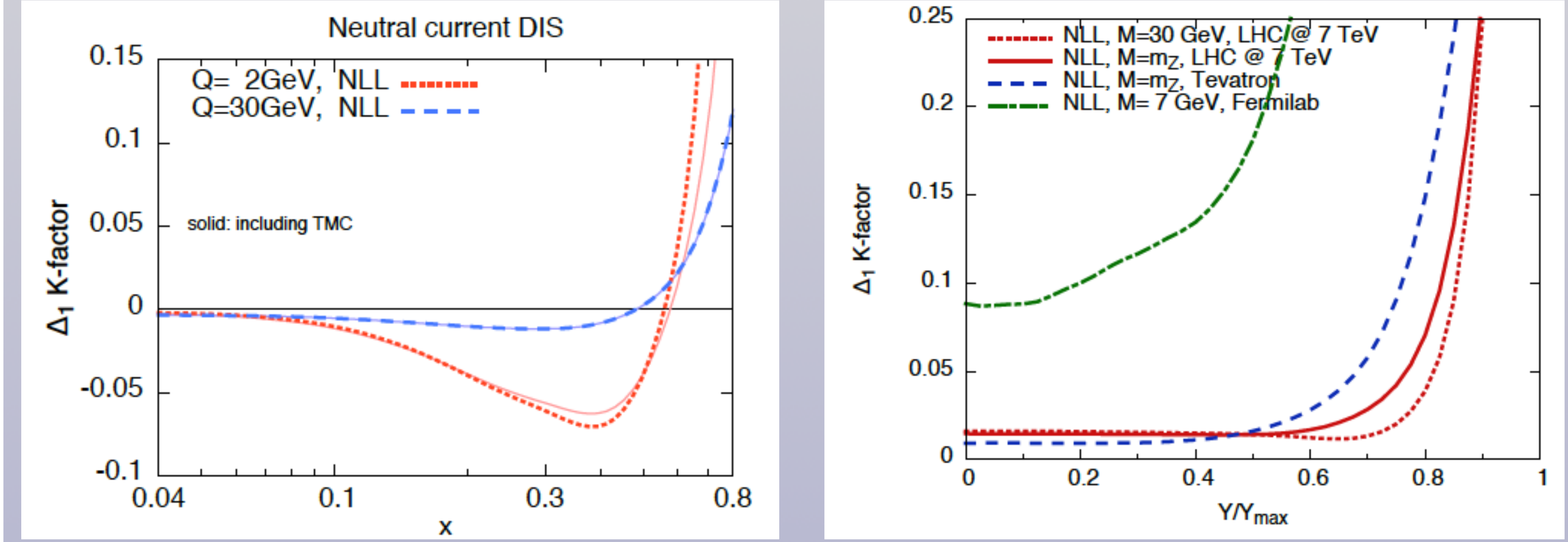


☑**Threshold-improved PDFs** can differ substantially wrt fixed-order PDFs: **up to -20% for gg luminosity and -40% for quark-antiquark luminosity,** in the high-mass region relevant for new BSM heavy particles

PDFs with threshold resummation

☑ The **suppression** observed at **large-x in the resummed PDFs** as compared to the FO ones can be traced back to the **enhancement due to NLO+NLL** used in the fit for DIS structure functions and DY distributions

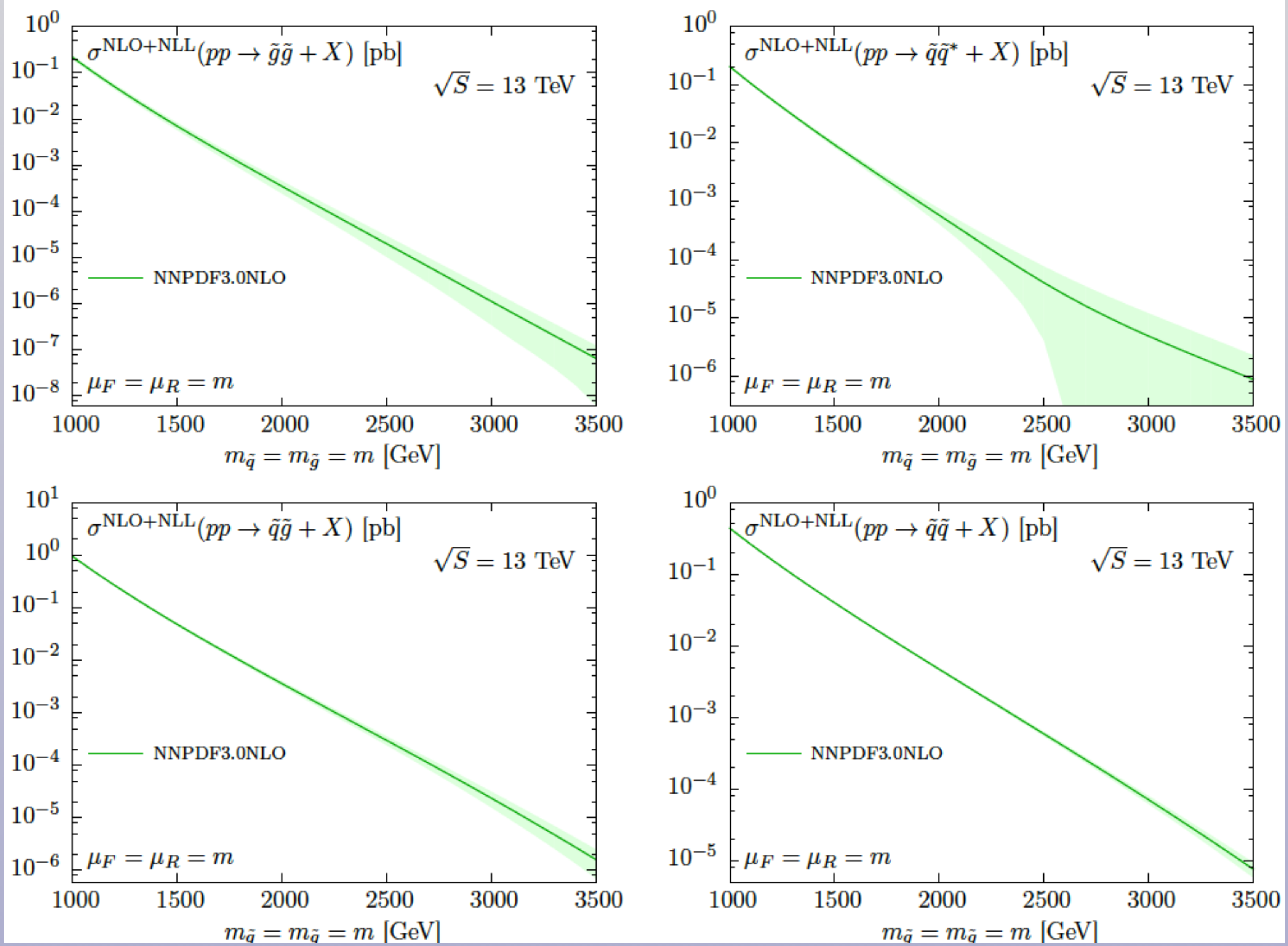
$$\sigma_{\mathrm{N}^j\mathrm{LO}+\mathrm{N}^k\mathrm{LL}} = \sigma_{\mathrm{N}^j\mathrm{LO}} + \sigma_{\mathrm{LO}} \times \Delta_j K_{\mathrm{N}^k\mathrm{LL}}$$



☑ **Phenomenologically most relevant:** this suppression will partially or totally compensate enhancements in partonic cross-sections for new processes (SUSY, Higgs, ttbar differential)

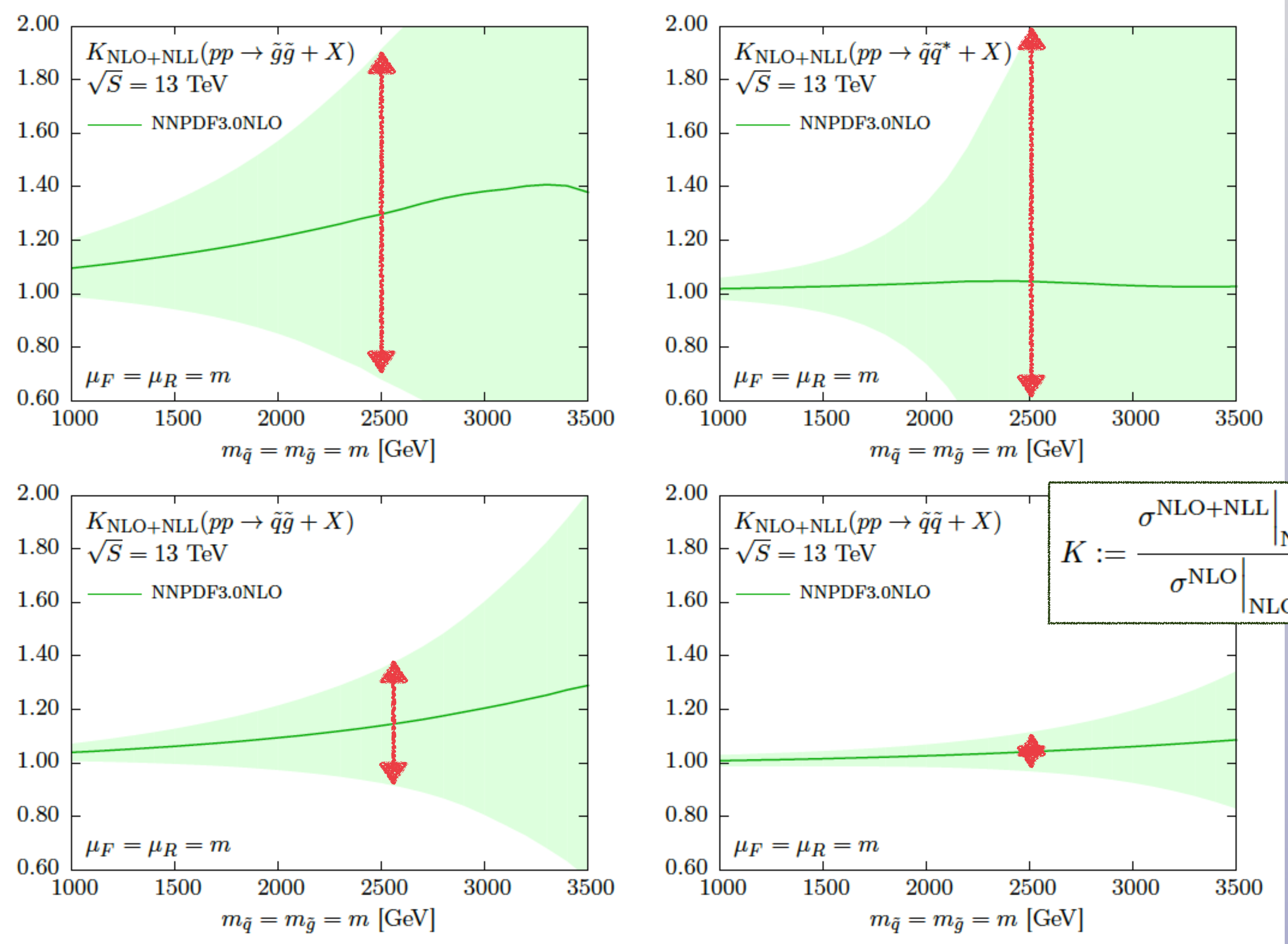
Updated NLO+NLL cross-sections with NNPDF3.0NLO

- ☑ **Previous NLL-fast calculations** at 13 TeV based on the (oldish) CTEQ6.6 and MSTW08 sets
- ☑ **NLL-fast version 3.1** has now been updated to **NLO+NLL cross-sections with NNPDF3.0NLO**



Updated NLO+NLL cross-sections with NNPDF3.0NLO

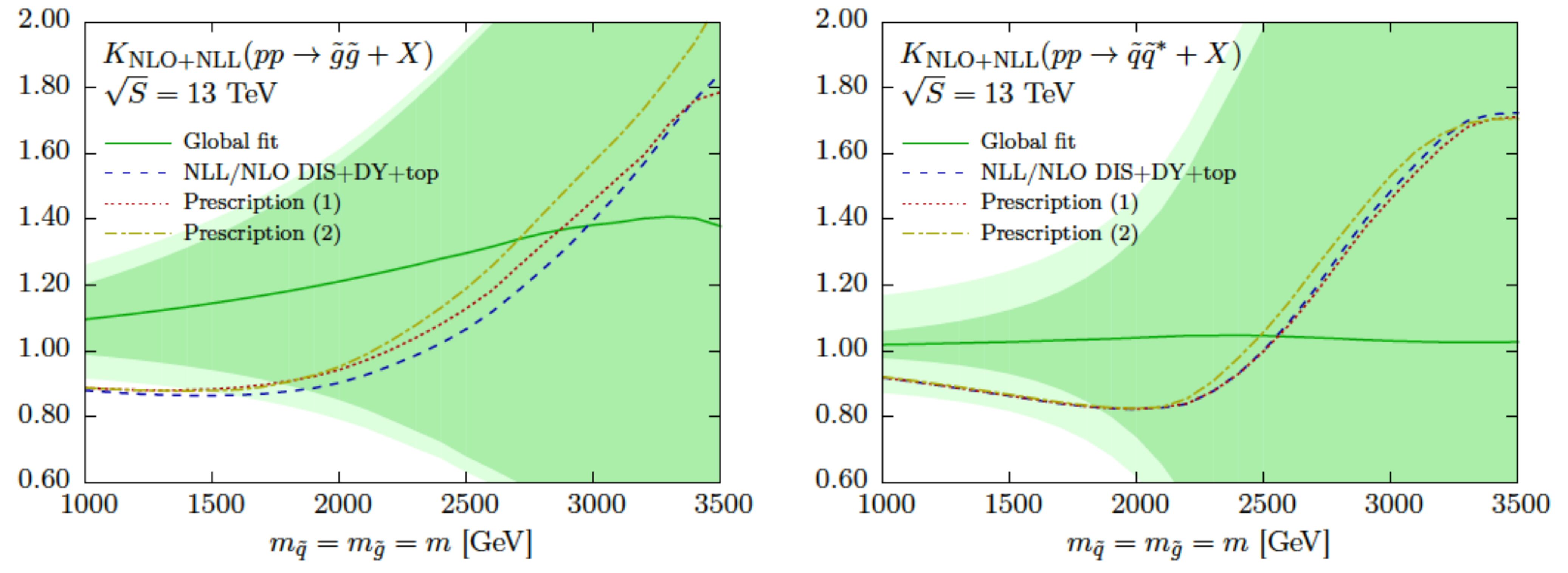
- ☑ **Previous NLL-fast calculations** at 13 TeV based on the (oldish) CTEQ6.6 and MSTW08 sets
- ☑ **NLL-fast version 3.1** has now been updated to **NLO+NLL cross-sections with NNPDF3.0NLO**



$$K := \frac{\sigma^{\text{NLO+NLL}}\Big|_{\text{NLO global}}}{\sigma^{\text{NLO}}\Big|_{\text{NLO global}}}.$$

NLO+NLL SUSY xsecs with threshold-improved PDFs

- ☑ Now include the effect of **NLO+NLL threshold-improved PDF**
- ☑ Substantial shift, **changes qualitatively and quantitatively** the behaviour of NLO+NLL SUSY xsecs
- ☑ Shift within total theory band, so **current exclusion limits unaffected**
- ☑ But will become crucial if we ever need to **characterise SUSY particles from LHC data**, much in the same way as we do for the Higgs boson



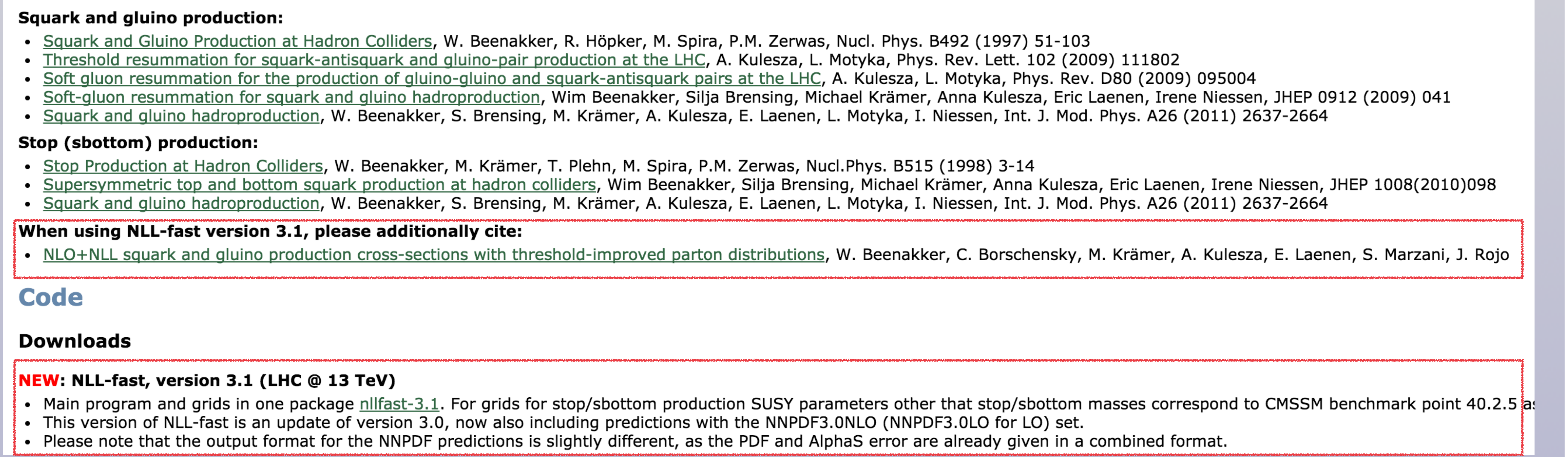
solid band: PDF errors only. lighter band: scales+PDF errors added linearly

NLL-fast grids

☑ The **updated NLO+NLL squark and gluino** production cross-sections at the **LHC 13 TeV** using **NNPDF3.0** can be downloaded from the NLL-fast collaboration webpage

☑ Include a **complete characterisation of theory uncertainties** from PDFs, scales and string coupling

http://pauli.uni-muenster.de/~akule_01/nllwiki/index.php/NLL-fast



Squark and gluino production:

- Squark and Gluino Production at Hadron Colliders, W. Beenakker, R. Höpker, M. Spira, P.M. Zerwas, Nucl. Phys. B492 (1997) 51-103
- Threshold resummation for squark-antisquark and gluino-pair production at the LHC, A. Kulesza, L. Motyka, Phys. Rev. Lett. 102 (2009) 111802
- Soft gluon resummation for the production of gluino-gluino and squark-antisquark pairs at the LHC, A. Kulesza, L. Motyka, Phys. Rev. D80 (2009) 095004
- Soft-gluon resummation for squark and gluino hadroproduction, Wim Beenakker, Silja Brensing, Michael Krämer, Anna Kulesza, Eric Laenen, Irene Niessen, JHEP 0912 (2009) 041
- Squark and gluino hadroproduction, W. Beenakker, S. Brensing, M. Krämer, A. Kulesza, E. Laenen, L. Motyka, I. Niessen, Int. J. Mod. Phys. A26 (2011) 2637-2664

Stop (sbottom) production:

- Stop Production at Hadron Colliders, W. Beenakker, M. Krämer, T. Plehn, M. Spira, P.M. Zerwas, Nucl.Phys. B515 (1998) 3-14
- Supersymmetric top and bottom squark production at hadron colliders, Wim Beenakker, Silja Brensing, Michael Krämer, Anna Kulesza, Eric Laenen, Irene Niessen, JHEP 1008(2010)098
- Squark and gluino hadroproduction, W. Beenakker, S. Brensing, M. Krämer, A. Kulesza, E. Laenen, L. Motyka, I. Niessen, Int. J. Mod. Phys. A26 (2011) 2637-2664

When using NLL-fast version 3.1, please additionally cite:

- NLO+NLL squark and gluino production cross-sections with threshold-improved parton distributions, W. Beenakker, C. Borschensky, M. Krämer, A. Kulesza, E. Laenen, S. Marzani, J. Rojo

Code

Downloads

NEW: NLL-fast, version 3.1 (LHC @ 13 TeV)

- Main program and grids in one package nllfast-3.1. For grids for stop/sbottom production SUSY parameters other that stop/sbottom masses correspond to CMSSM benchmark point 40.2.5 as
- This version of NLL-fast is an update of version 3.0, now also including predictions with the NNPDF3.0NLO (NNPDF3.0LO for LO) set.
- Please note that the output format for the NNPDF predictions is slightly different, as the PDF and AlphaS error are already given in a combined format.

☑ In addition, **cross-sections using the threshold-improved NNPDF3.0** sets is available from the authors upon request.

☑ Important to use **up-to-date theory calculations** for the **interpretation of SUSY searches** at the LHC Run II

Summary

- **Parton Distributions** are an essential ingredient for LHC phenomenology
- Accurate PDFs are required for **precision SM measurements, Higgs characterisation** and **New Physics searches**
- The determination of **fundamental SM parameters** like the **W mass** or **α_S from LHC data** also greatly benefits from improved PDFs
- Improving our understanding of PDFs has important implications in other domains of high-energy physics, such as the **calculations of backgrounds for neutrino telescopes**
- The NNPDF approach provides parton distributions based on a **robust, unbiased methodology,** the most updated **theoretical information** and all the relevant hard scattering data **including LHC data**