TBA

Alberto Guffanti

Niels Bohr International Academy & Discovery Center
Niels Bohr Institute - Copenhagen

Heavy Ions: Experiment Confront Theory

Niels Bohr Institute (Copenhagen)

November 7-9, 2011

NNPDF for the LHC

and a couple of comments on Nuclear (NN)PDFs

Alberto Guffanti

Niels Bohr International Academy & Discovery Center
Niels Bohr Institute - Copenhagen

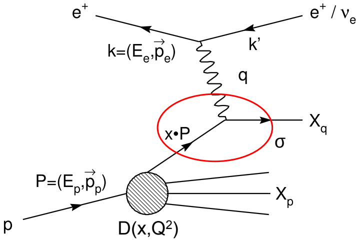
Heavy Ions: Experiment Confront Theory

Niels Bohr Institute (Copenhagen)

November 7-9, 2011

What are Parton Distribution Functions?

- Consider a process with one hadron in the initial state



- According to the **Factorization Theorem** we can write the cross section as

$$d\sigma = \sum_a \int_0^1 \frac{d\xi}{\xi} D_a(\xi,\mu^2) d\hat{\sigma}_a\left(\frac{x}{\xi},\frac{\hat{s}}{\mu^2},\alpha_s(\mu^2)\right) + \mathcal{O}\left(\frac{1}{Q^p}\right)$$

What are Parton Distribution Functions?

- The **absolute value** of PDFs at a given x and Q^2 **cannot be computed** in QCD Perturbation Theory
  (Lattice? In principle yes, but ...)

- ... but the **scale dependence** is governed by **DGLAP** evolution equations

$$\frac{\partial}{\ln Q^2} q^{NS}(\xi, Q^2) = P^{NS}(\xi, \alpha_s) \otimes q^{NS}(\xi, Q^2)$$

$$\frac{\partial}{\ln Q^2} \begin{pmatrix} \Sigma \\ g \end{pmatrix} (\xi, Q^2) = \begin{pmatrix} P_{qq} & P_{qg} \\ P_{gq} & P_{gg} \end{pmatrix} (\xi, \alpha_s) \otimes \begin{pmatrix} \Sigma \\ g \end{pmatrix} (\xi, Q^2)$$

- ... and the **splitting functions** P can be computed in PT and are known up to **NNLO**

[**LO** - Dokshitzer; Gribov, Lipatov; Altarelli, Parisi; 1977]
[**NLO** - Floratos, Ross, Sachrajda; Gonzalez-Arroyo, Lopez, Yndurain; Curci, Furmanski, Petronzio, 1981]
[**NNLO** - Moch, Vermaseren, Vogt; 2004]

Problem

Faithful estimation of errors on PDFs

- Single quantity: **1-σ error**
- Multiple quantities: **1-σ contours**
- Function: need an **"error band" in the space of functions**
  (*i.e.* the probability density $\mathcal{P}[f]$ in the space of functions $f(x)$)

Expectation values are **Functional integrals**

$$\langle \mathcal{F}[f(x)] \rangle = \int \mathcal{D}f \mathcal{F}[f(x)] \mathcal{P}[f(x)]$$

Problem

Faithful estimation of errors on PDFs

- Single quantity: **1-σ error**
- Multiple quantities: **1-σ contours**
- Function: need an **"error band" in the space of functions**
  (*i.e.* the probability density $\mathcal{P}[f]$ in the space of functions $f(x)$)

Expectation values are **Functional integrals**

$$\langle \mathcal{F}[f(x)] \rangle = \int \mathcal{D}f \mathcal{F}[f(x)] \mathcal{P}[f(x)]$$

Determine a function from a finite set of data points

Solution

Standard Approach

- Introduce a simple functional form with enough free parameters

$$q(x, Q^2) = x^{\alpha}(1-x)^{\beta} P(x; \lambda_1, ..., \lambda_n).$$

- Fit parameters minimizing χ^2.

Solution

Standard Approach

- Introduce a simple functional form with enough free parameters

$$q(x, Q^2) = x^{\alpha}(1-x)^{\beta} P(x; \lambda_1, ..., \lambda_n).$$

- Fit parameters minimizing χ^2.

Open problems:

- **Error propagation** from data to parameters and from parameters to observables is **not trivial**.

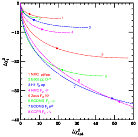
- **Theoretical bias** due to the chosen **parametrization** is difficult to assess.

Shortcomings of the Standard approach

What is the meaning of a one-σ uncertainty?

- Standard $\Delta\chi^2 = 1$ criterion is **too restrictive** to account for large discrepancies among experiments in a **global fit**.

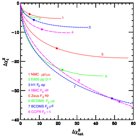
  [Collins & Pumplin, 2001]



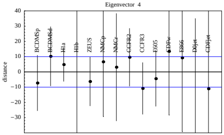
Shortcomings of the Standard approach

What is the meaning of a one-σ uncertainty?

- Standard $\Delta\chi^2 = 1$ criterion is **too restrictive** to account for large discrepancies among experiments in a **global fit**.
  [Collins & Pumplin, 2001]



- Introduce a **TOLERANCE** criterion, i.e. take the envelope of uncertainties of experiments to determine the $\Delta\chi^2$ to use for the global fit (CTEQ).
  [Tung et al., 2006]

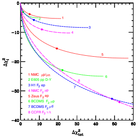


Shortcomings of the Standard approach

What is the meaning of a one-σ uncertainty?

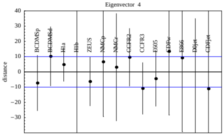
- Standard $\Delta\chi^2 = 1$ criterion is **too restrictive** to account for large discrepancies among experiments in a **global fit**.

  [Collins & Pumplin, 2001]



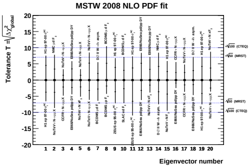
- Introduce a **TOLERANCE** criterion, i.e. take the envelope of uncertainties of experiments to determine the $\Delta\chi^2$ to use for the global fit (CTEQ).

  [Tung et al., 2006]



- Make it **DYNAMICAL**, i.e. determine $\Delta\chi^2$ separately for each hessian eigenvector (MSTW).

  [Thorne et. al, 2008]

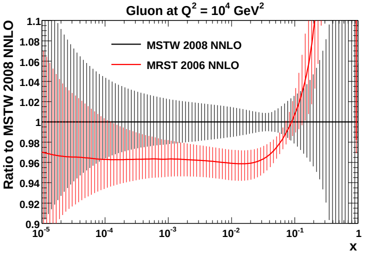


Shortcomings of the standard approach

What determines PDF uncertainties?

- **Uncertainties** in standard fits often **increase when adding data** (i.e. when adding information) even if they are compatible with the old data.
- **Reason**: need change the parametriztion in order to accomodate the new data.

Smaller high-x gluon (and slightly smaller α_S) results in larger small-x gluon – now shown at NNLO.



Larger small-x uncertainty due to extrat free parameter.

[R. Thorne, PDF4LHC]

THE NNPDF METHODOLOGY

[R. D. Ball, V. Bertone, F. Cerutti, L. Del Debbio, S. Forte, J. I. Latorre, A. Piccione, J. Rojo, M. Ubiali and AG]

NNPDF Methodology

Main Ingredients

- **Monte Carlo** determination of errors
  - No need to rely on linear propagation of errors
  - Possibility to test for the impact of non gaussianly distributed errors
  - Possibility to test for non-gaussian behaviour in fitted PDFs ($1 - \sigma$ vs. 68% CL)
- **Neural Networks**
  - Provide an **unbiased** parametrization
- **Stopping based on Cross-Validation**
  - Ensures proper fitting avoiding overlearning

NNPDF Methodology

... in a Nutshell

- Generate N_{rep} **Monte-Carlo replicas** of the experimental data (sampling of the probability density in the space of data)
- Fit a set of Parton Distribution Functions on each replica (sampling of the probability density in the space of PDFs)
- **Expectation values** for observables are **Monte Carlo integrals**

$$\langle \mathcal{F}[f_i(x,Q^2)] \rangle = \frac{1}{N_{rep}} \sum_{k=1}^{N_{rep}} \mathcal{F}\left(f_i^{(net)(k)}(x,Q^2) \right)$$

... the same is true for errors, correlations, etc.

NNPDF Methodology

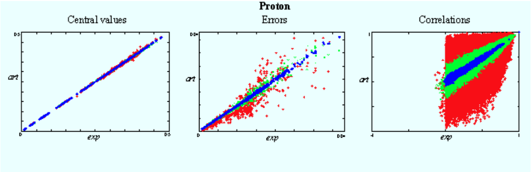
Monte Carlo replicas generation

- **Generate** artificial data according to distribution

$$O_i^{(art)\,(k)} = (1 + r_N^{(k)}\,\sigma_N)\left[O_i^{(exp)} + \sum_{p=1}^{N_{sys}} r_p^{(k)}\,\sigma_{i,p} + r_{i,s}^{(k)}\,\sigma_s^i\right]$$

  where r_i are univariate (gaussianly distributed) random numbers

- **Validate** Monte Carlo replicas against experimental data (statistical estimators, faithful representation of errors, convergence rate increasing N_{rep})



- $\mathcal{O}(1000)$ replicas needed to reproduce correlations to percent accuracy

Neural Networks

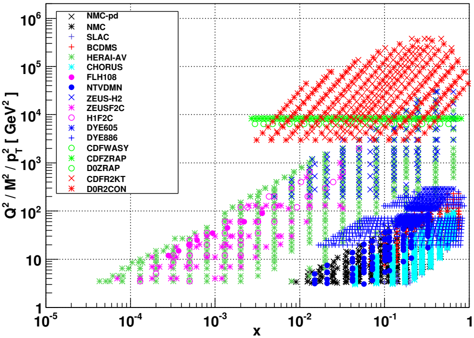
... a suitable basis of functions

- We use **Neural Networks** as **functions** to represent **PDFs at the starting scale**
  - They provide a parametrization which is **redundant** and **robust** against variations
- We employ **Multilayer Feed-Forward** Neural Networks trained using a **Genetic Algorithm**
  - Very efficient for complex parameter spaces
  - Reduces the probability of being trapped in local minima
- Stopping of the fit for each replica is determined by **cross-validation** method
  - Ensures proper proper fitting avoiding overlearning

NNPDF2.1

A family of global fits ...

NNPDF2.1 dataset



3338 data points (NLO fit)

(3330 - LO and 3445 - NNLO)

[R. D. Ball et. al, arXiv:1101.1300] - **NLO**
[R. D. Ball et. al, arXiv:1107.2652] - **LO/NNLO**

OBS	***Data set***
Deep Inelastic Scattering	
F_2^d / F_2^p	NMC-pd
F_2^p	NMC, SLAC, BCDMS
F_2^d	SLAC, BCDMS
$\sigma_{NC}^{\pm}$	HERA-I, ZEUS (HERA-II)
$\sigma_{CC}^{\pm}$	HERA-I, ZEUS (HERA-II)
F_L	H1
$\sigma_{\nu}, \sigma_{\bar{\nu}}$	CHORUS
dimuon prod.	NuTeV
F_2^c	ZEUS, H1
Drell-Yan & Vector Boson prod.	
$d\sigma^{\mathrm{DY}} / dM^2 dy$	E605
$d\sigma^{\mathrm{DY}} / dM^2 dx_F$	E866
W asymm.	CDF
Z rap. distr.	D0/CDF
Inclusive jet prod.	
Incl. $\sigma^{(\mathrm{jet})}$	CDF (k_T) - Run II
Incl. $\sigma^{(\mathrm{jet})}$	D0 (cone) - Run II

NNPDF2.1

... based on the NNPDF Methodology

Parton Distributions Combination		NN architechture
Singlet ($\Sigma(x)$)	$\Longrightarrow$	2-5-3-1 (37 pars)
Gluon ($g(x)$)	$\Longrightarrow$	2-5-3-1 (37 pars)
Total valence ($V(x) \equiv u_V(x) + d_V(x)$)	$\Longrightarrow$	2-5-3-1 (37 pars)
Non-singlet triplet ($T_3(x)$)	$\Longrightarrow$	2-5-3-1 (37 pars)
Sea asymmetry ($\Delta_S(x) \equiv \bar{d}(x) - \bar{u}(x)$)	$\Longrightarrow$	2-5-3-1 (37 pars)
Total Strangeness ($s^+(x) \equiv (s(x) + \bar{s}(x))/2$)	$\Longrightarrow$	2-5-3-1 (37 pars)
Strange valence ($s^-(x) \equiv (s(x) - \bar{s}(x))/2$)	$\Longrightarrow$	2-5-3-1 (37 pars)

259 parameters
Standard fits have $\sim$ 25 parameters in total

No change in the parametrization since NNPDF1.2 ... despite substantial **enlargement of the dataset**.

NNPDF2.1

... including Heavy Flavour contributions (FONLL)

- We adopt the **FONLL** General Mass-Variable Flavour Number Scheme

  [M. Cacciari, M. Greco and P. Nason, (1998)]

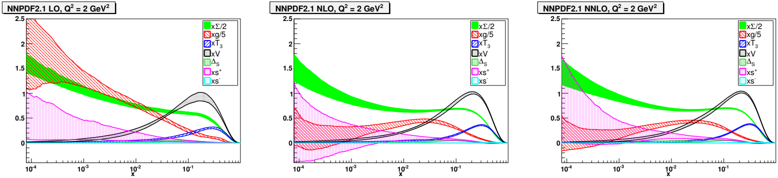
  [S. Forte, P. Nason E. Laenen and J. Rojo, (2010)]

- FONLL gives a prescription to **combine FFN** (Massive) and **ZM-VFN** (Massless) computations, at any given order, **avoiding double counting**.
- With available computations three implementations of FONLL are possibile:
  - **FONLL-A**: $\mathcal{O}(\alpha_s)$ Massless + $\mathcal{O}(\alpha_s)$ Massive - (NLO fit)
  - **FONLL-B**: $\mathcal{O}(\alpha_s)$ Massless + $\mathcal{O}(\alpha_s^2)$ Massive
  - **FONLL-C**: $\mathcal{O}(\alpha_s^2)$ Massless + $\mathcal{O}(\alpha_s^2)$ Massive - (NNLO fit)
- **Fixed Flavour Number Scheme** (3-, 4-, 5-) fits **available**.

NNPDF2.1

... a (PDF) family portrait

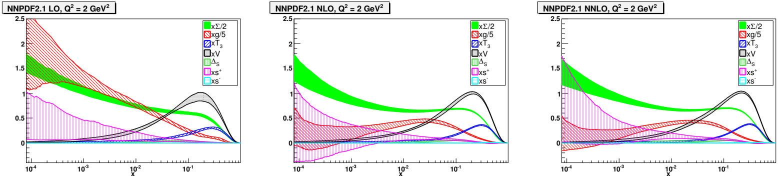
- At the starting scale (2 GeV^2) ...



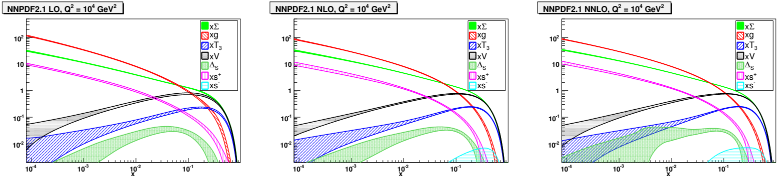
NNPDF2.1

... a (PDF) family portrait

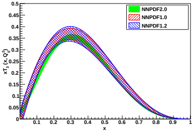
- At the starting scale (2 GeV^2) ...



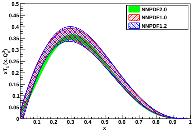
- ... and at the typical EW scale (100 GeV^2)



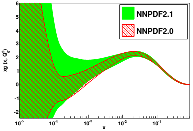
- **Reduction of uncertainties** with respect to older NNPDF sets due to **inclusion of new data**



- **Reduction of uncertainties** with respect to older NNPDF sets due to **inclusion of new data**



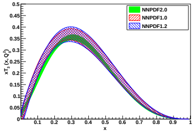
- When **uncertainties increase** we know it is **not a parametrzation effect**



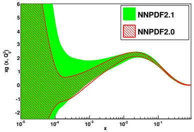
NNPDF2.1

Partons - A couple of upshots

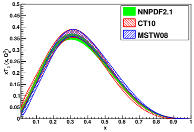
- **Reduction of uncertainties** with respect to older NNPDF sets due to **inclusion of new data**



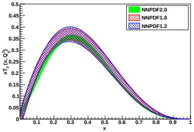
- When **uncertainties increase** we know it is **not a parametrzation effect**



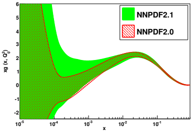
- **Uncertainties** on PDFs have **size comparable** to those obtained by other groups when there are significant **contraints from data** ...



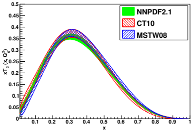
- **Reduction of uncertainties** with respect to older NNPDF sets due to **inclusion of new data**



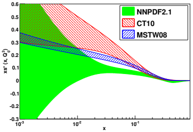
- When **uncertainties increase** we know it is **not a parametrzation effect**



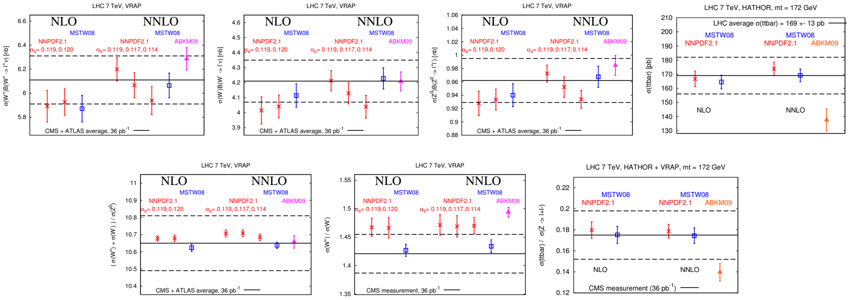
- **Uncertainties** on PDFs have **size comparable** to those obtained by other groups when there are significant **contraints from data** ...



- ... but **no functional bias** in kinematic regions where there are little or **no experimental constraints**.



- Predictions for **LHC Standard Candles** compared to **LHC data**



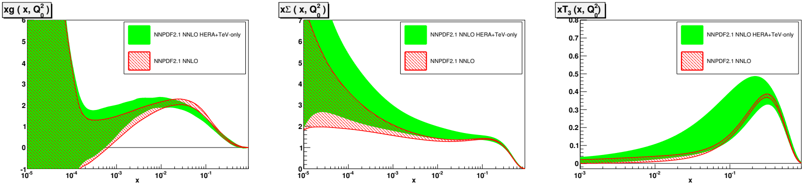
- LHC data will soon be precise enough to distinguish between different predictions.

NNPDF2.1: fits to reduced datasets

Collider-only fit

- The (proton) fit we would love to have ...
  - Only **high energy data**: minimize effects of higher-twist contribiutions
  - Only **proton data**: no assumptions based on models for nuclear corrections

- Based on **HERA and Tevatron** (inclusive jets and W/Z prduction) data



- **LHC** (and HERA-II combined) **data are crucial** in order to improve collider-only fit

LHC4PDFs

... the data we would love to have from the LHC

- Medium- and large-x **gluon**
  - Prompt photons
  - Inclusive Jets
  - t-quark distributions ($p_{\perp}$, y) (?)

- **Light flavour separation** at medium- & small-x
  - Low-mass Drell-Yan
  - High-mass W prduction
  - Z rapidity distribution
  - W(+jets) asymmetry

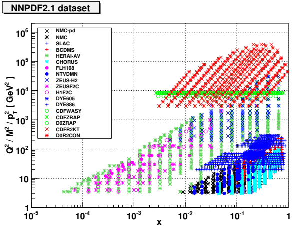
- **Strangeness & Heavy Flavours**
  - $W + c$
  - $Z + c$, $\gamma + c$
  - $Z + b$

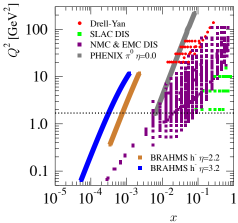
A QUICK LOOK IN THE NUCLEAR TERRITORY

Nuclear NNPDFs

The case for having them

- Nuclear PDFs are determined from a much more restricted dataset than proton ones





- In the standard approach to PDF extractions (used by EPS, HKN and nDS sets) uncontrolled theoretical assumption on the functional form of PDFs can be a source of large bias

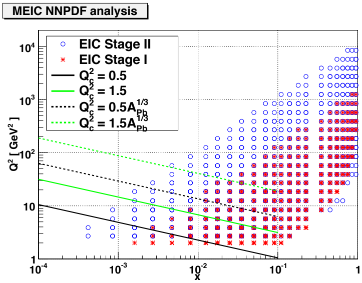
- The NNPDF methodology is ideally suited to provide a reliable estimation of the uncertanties in a situation where experimental constraints are very loose

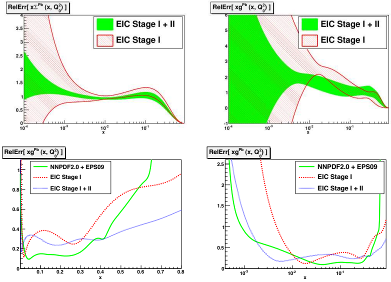
Nuclear NNPDFs

EIC White Book - the first NNPDF steps in the nuclear territory

[A. Accardi, V. Guzey and J. Rojo, arXiv:1106.3839]

- One of the phyiscs goal of the proposed Electron Ion Collider (EIC) is the determination of nuclear PDFs
- The NNPDF methodology has been used to extract the Singlet and gluon PDFs from a fit to simulated EIC *e*-Pb pseudodata





Conclusions & Outlook

- A set of PDF with a **reliable** estimation of **uncertainties** is crucial in order to exploit the full physics potential of the LHC experiments (both for *pp* and *AA*).
- The **NNPDF2.1** family of PDF sets fulfills the requirement of an ideal parton densities set for precision (*pp*) phenomenology at the LHC
- The NNPDF Methodology is ideally suited to tackle problems that affect standard global (nuclear) fits (loose constraints from data, parametrization bias, etc.)
- ... moreover, plenty of new data will come from the LHC probing different nPDF combinations and providing more stringent constraints