PARTON DISTRIBUTIONS FROM HIGH PRECISION DATA

MARIA UBIALI
UNIVERSITY OF CAMBRIDGE

HIGH ENERGY THEORY AND GENDER
CERN, 26-28 SEPTEMBER 2018
A NEW PRECISION ERA IN PARTICLE PHYSICS

- LHC at the forefront of the exploration of the high energy regime
- Precise predictions are the key not to miss this unique opportunity!

Apollinari et al., CERN Yellow Report 4 (2017)

- High-Luminosity LHC
- Higher mass reach
- Huge increase in integrated luminosity
A NEW PRECISION ERA IN PARTICLE PHYSICS

- LHC at the forefront of the exploration of the high energy regime
- Precise predictions are the key not to miss this unique opportunity!

Direct detection: precision required to characterise NP

**ATLAS, SM summary plots**
A NEW PRECISION ERA IN PARTICLE PHYSICS

➡️ LHC at the forefront of the exploration of the high energy regime
➡️ Precise predictions are the key not to miss this unique opportunity!

Indirect detection: precision essential to spot deviations

ATLAS, SM summary plots
THE PRECISION CHALLENGE

Theoretical predictions must catch up with experimental precision!

ATLAS, SM summary plots
THE PRECISION INGREDIENTS

- Hard scattering of partons (Perturbative QCD+EW)
- Parton Distribution Functions
- Parton Showering and Hadronization
- Multiple Parton Interaction, Underlying Events

SHERPA artist
OUTLINE OF MY TALK

- INTRODUCTION
  - WHAT PDFS ARE
  - WHY THEY CRUCIAL AT HADRON COLLIDERS

- HOW PDFS BECAME PRECISION PHYSICS
  - EXPERIMENTAL DATA
  - THEORY PROGRESS
  - ROBUST STATISTICS

- NEW FRONTIERS
  - THEORETICAL UNCERTAINTIES IN PDFS
  - PDFS AND NEW PHYSICS

- Hard scattering of partons (Perturbative QCD+EW)
- Parton Distribution Functions
- Parton Showering and Hadronization
- Multiple Parton Interaction, Underlying Events
PART I

INTRODUCTION
Parton Distribution Functions (non-perturbative)

Hard Scattering (Perturbative QCD)

\[
\frac{d\sigma_{pp\rightarrow ab}}{dX} = \sum_{i,j} \int \frac{dz_1}{z_1} \frac{dz_2}{z_2} f_i(z_1, \mu_F) f_j(z_2, \mu_F) \frac{d\hat{\sigma}^{ij\rightarrow ab}}{dX}(z_1 z_2 S, \alpha_s(\mu_R), \mu_F) + \mathcal{O}\left(\frac{\Lambda}{S}\right)
\]

- Universal: PDFs do not depend on hard scattering process
- Dependence on scale $\mu$ predicted by pQCD (DGLAP evolution equations)
WHAT PDFS ARE

\[ f_i(x, \mu) \]

Data

Perturbative QCD

Hadronic scale: global fit of PDFs

High scale: input to the LHC
WHY THEY ARE CRUCIAL

Yellow Report 3 (2013)

PDF uncertainties limiting factor in the accuracy of theoretical predictions
PDF uncertainties down to 1-5% but crucial for Higgs physics

Why They Are Crucial

Yellow Report 4 (2016)

M. Ubiali - High Energy Theory and Gender - PDFs from High Precision Data

Higgs Production Channel

<table>
<thead>
<tr>
<th>Channel</th>
<th>% Theo. Uncertainty</th>
<th>σ@13 TeV</th>
</tr>
</thead>
<tbody>
<tr>
<td>ggF (N3LO)</td>
<td>6.5</td>
<td>48.5 pb</td>
</tr>
<tr>
<td>VBF (N2LO)</td>
<td>3.7</td>
<td>3.78 pb</td>
</tr>
<tr>
<td>WH (N2LO)</td>
<td>1.4</td>
<td>1.37 pb</td>
</tr>
<tr>
<td>ZH (N2LO)</td>
<td>0.9</td>
<td>0.88 pb</td>
</tr>
<tr>
<td>ttH (N1LO)</td>
<td>0.5</td>
<td>0.51 pb</td>
</tr>
</tbody>
</table>

PDF uncertainties down to 1-5% but crucial for Higgs physics.
**WHY THEY ARE CRUCIAL**

**M\(_w\) determination**

\[ m_W = 80370 \pm 19 \text{ MeV} \]

**Gluino production**

\[ K_{NLO+NLL}(pp \to \tilde{g}\tilde{g} + X) \]

\[ \sqrt{S} = 13 \text{ TeV} \]

\[ m_{\tilde{q}} = m_{\tilde{g}} = m \text{ [GeV]} \]

**Fundamental parameters of SM**

**New physics searches/characterisation**
PART II
HOW PDFS BECAME PRECISION PHYSICS
30 years of steady progress in PDF community have produced a huge impact on understanding of proton structure and precision physics.
A COMPLEX MACHINERY

EXPERIMENTAL DATA
- FIXET TARGET EXPERIMENTS, TEVATRON, HERA, LHC...
- DIS, DRELL-YAN, JETS, TOP, DIRECT PHOTON...

\[ \sigma = f_1(\otimes f_2) \otimes \hat{\sigma} \]
A COMPLEX MACHINERY

EXPERIMENTAL DATA
- Fixet Target Experiments, Tevatron, HERA, LHC...
- DIS, Drell-Yan, Jets, Top, Direct Photon...

THEORY
- Fixed Order Calculation
- Fast Interfaces
- Heavy Quark Schemes
- EW Corrections...

\[ \sigma = f_1(\otimes f_2) \otimes \hat{\sigma} \]
A COMPLEX MACHINERY

EXPERIMENTAL DATA
- FIXET TARGET EXPERIMENTS, TEVATRON, HERA, LHC ...
- DIS, DRELL-YAN, JETS, TOP, DIRECT PHOTON ...

THEORY
- FIXED ORDER CALCULATION
- FAST INTERFACES
- HEAVY QUARK MASSES
- EW CORRECTIONS ...

STATISTICAL FRAMEWORK
- PDF PARAMETRISATION
- PDF UNCERTAINTIES AND PROPAGATION
- MINIMISATION ROBUSTNESS

$$\sigma = f_1(\otimes f_2) \otimes \hat{\sigma}$$
A COMPLEX MACHINERY

EXPERIMENTAL DATA
- FIXET TARGET EXPERIMENTS, TEVATRON, HERA, LHC...
- DIS, DRELL-YAN, JETS, TOP, DIRECT PHOTON...

THEORY
- FIXED ORDER CALCULATION
- FAST INTERFACES
- HEAVY QUARK MASSES
- EW CORRECTIONS...

STATISTICAL FRAMEWORK
- PDF PARAMETRISATION
- PDF UNCERTAINTIES AND PROPAGATION
- MINIMISATION ROBUSTNESS
EXPERIMENTAL DATA

Ball et al, EPJC 77 (2017)
EXPERIMENTAL DATA

- Increased reach at large and low-x and new constraints
- Amazing precision of experimental data
EXPERIMENTAL DATA

- Increased reach at large and low-x and new constraints
- Amazing precision of experimental data
Large-x gluon constrained by three independent processes
Consistent picture and analysis robustness

EXPERIMENTAL DATA & INSIGHT ON THE GLUON

NNLO, Q = 100 GeV

$\delta g(x, Q^2)/g(x, Q^2)_{[\text{ref}]}$

- $\text{NNPDF3.1}$
- $\text{NNPDF3.0}$

$g(x, Q^2)/g(x, Q^2)_{[\text{ref}]}$

- Global
- no jets
- no top
- no Z $p_T$

Robust under dataset variations

$g$ $\rightarrow$ Jet $\rightarrow$ $g$

$g$ $\rightarrow$ $q$

$g$ $\rightarrow$ $t$ $\rightarrow$ $tb$

$g$ $\rightarrow$ jet $\rightarrow$ $Z$
**THEORY: THE NNLO FRONTIER**

- NNLO calculations are essential to reduce theoretical uncertainties in PDF analyses

- Stunning progress made on calculation of some key process for PDF determination

- **NNLO top pair production**
  - Czakon et al [PRL 110(2013)]
  - Czakon et al [JPCP (2014)]
  - Czakon et al [JHEP 1301(2015)]

- **W/Z+j and W/Z transverse momentum distributions**
  - Gehrmann-De Ridder et al [JHEP 07 (2016)]
  - Gehrmann-De Ridder et al [JHEP 11 (2016)]
  - Boughezal et al [PRL 16 (2016)]
  - Boughezal et al [PRD 14 (2016)]

- **Inclusive jet and di-jets**
  - Currie et al [PRL 118 (2017)]
  - Currie et al [PRL 119 (2017)]
  - Gehrmann-De Ridder et al [PRL 110 (2016)]

- **Inclusive DIS jets**
  - Currie et al [JHEP 17 (2017)]

- **Direct photon**
  - Campbell et al [PRL 118 (2017)]

- **Single top**
  - Bruchersfeier et al [PRB 736 (2014)]
THEORY: THE NNLO FRONTIER

- NNLO calculations are essential to reduce theoretical uncertainties in PDF analyses

- Stunning progress made on calculation of some key process for PDF determination


- Breakthrough in determining photon PDF (Manohar et al 2016)

- NNLO top pair production
  - Czakon et al [PRL 110(2013)]
  - Czakon et al [JPCP (2014)]
  - Czakon et al [JHEP 1301(2015)]

- W/Z+j and W/Z transverse momentum distributions
  - Gehrmann-De Ridder et al [JHEP 07 (2016)]
  - Gehrmann-De Ridder et al [JHEP 11 (2016)]
  - Boughezal et al [PRL 16 (2016)]
  - Boughezal et al [PRD 14 (2016)]

- Inclusive jet and di-jets
  - Currie et al [PRL 118 (2017) ]
  - Currie et al [PRL 119 (2017) ]
  - Gehrmann-De Ridder et al [PRL 110 (2016)]

- Inclusive DIS jets
  - Currie et al [JHEP 17 (2017) ]

- Direct photon
  - Campbell et al [PRL 118 (2017) ]

- Single top
  - Bruchersfeier et al [PRB 736 (2014) ]
STATS: ERROR PROPAGATION

\[ \langle O[\{f\}] \rangle = \int [Df] O[\{f\}] P[\{f\}] \]

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

Propagation of experimental uncertainty

\[ \langle O[\{f\}] \rangle = \int da_1 da_2 \cdots da_n O[\vec{a}] P[\vec{a}] \]

- Hessian approach: Project into a n-dimensional space of parameters and use linear approximation around minimum \( \chi^2 \)

Parametrisation

- Introduce a simple functional form with enough free parameters
- Typically about 20-40 free parameters for 7 independent functions

\[ f_i(x, Q_0) = a_0 x^{a_1} (1 - x)^{a_2} P(x, a_3, a_4, \ldots) \]
PDF uncertainties tuned to data (tolerance $\Delta \chi^2 > 1$ - many studies/improvements)

Fixed parametrisation was forced to be more flexible by new data $\Rightarrow$ less biased parametrisation form (a posteriori data-driven progress)

$$xg = A_g x^{\delta_g (1 - x)} \eta_g (1 + \epsilon_g \sqrt{x} + \gamma_g x) + A_{g'} x^{\delta_{g'} (1 - x)} \eta_{g'}$$
PDF uncertainties tuned to data (tolerance $\Delta \chi^2 > 1$ - many studies/improvements)

Fixed parametrisation was forced to be more flexible by new data => less biased parametrisation form (a posteriori data-driven progress)
Monte Carlo techniques: sampling the probability measure in PDF functional space from data

\[ \langle \mathcal{O}[\{f\}] \rangle = \frac{1}{N} \sum_{k=1}^{N} \mathcal{O}[f_k] \]

Neural Networks: all independent PDFs are associated to an unbiased and flexible parametrization: \( O(300) \) parameters versus \( O(30) \) in polynomial parametrization
Great progress in cross-talk between Hessian and MC approaches to propagate experimental data error into PDF error bands.

- Hessian $\Rightarrow$ Monte Carlo by generate multi-gaussian replicas in parameter space.
- Monte Carlo $\Rightarrow$ Hessian by sampling replicas $f(x)$ at discrete sets of points and contract covariance matrix.

**Watt, Thorne, JHEP 08 (2012)**

Pretty good agreement for gluon PDF
Some differences in quark PDFs, especially down and strangeness
How to improve?
- Closure test to test methodology (!)
- Data
- Theory (K-factors and MC integration error…)

PART III

NEW FRONTIERS
PDF UNCERTAINTIES

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
- Changes in theory?
- Updated parametrization
- Differences in fitting methodology/minimisation?

Theory boundaries
Data region
Extrapolation region

Tolerance/Statistical estimators
Closure Test
THEORY UNCERTAINTIES

- PDF fits performed at given perturbative order
- PDF uncertainties only reflect lack of information from data
- How to estimate MHOU in PDF fits?
- Idea: build theory covariance matrix
- Add it to experimental one [NNPDF collaboration in progress]

Ball et al, EPJC 77 (2017)
NEW PHYSICS AND PDFS

- Q: As more data at higher energy will be released, how can we make sure that we will not absorb new physics in the PDFs?

- Inconsistencies between data that enter a global PDF analysis can distort statistical interpretation of PDF uncertainties.

- Inconsistency of any individual dataset with the bulk of global fit may suggest its understanding is incomplete.

- Conservative partons & studies on impact of EFT operators on PDF fits way forward to systematically include new data.
CONCLUSIONS AND OUTLOOK

[...] Global QCD Analysis of available hard processes critically tests the validity of the PQCD framework, allows the determination of the non-perturbative parton distribution functions, thereby provides the necessary input to calculate and predict most Standard Model and New Physics processes for future, higher, energy interactions. After two decades of steady progress in this venture, has global QCD analysis of parton distributions reached the End of the Road (as some have proclaimed); or, will the physics challenges of the next generation of colliders usher in the Dawn of a New Era, with fresh ideas and more powerful methodology (as some have promised)? That, is the question.

Wu-Ki Tung - CERN-TH colloquium 2000
CONCLUSIONS AND OUTLOOK

[... ] Global QCD Analysis of available hard processes critically tests the validity of the PQCD framework, allows the determination of the non-perturbative parton distribution functions, thereby provides the necessary input to calculate and predict most Standard Model and New Physics processes for future, higher, energy interactions.

After two decades of steady progress in this venture, has global QCD analysis of parton distributions reached the End of the Road (as some have proclaimed); or, will the physics challenges of the next generation of colliders usher in the Dawn of a New Era, with fresh ideas and more powerful methodology (as some have promised)? That, is the question.

Wu-Ki Tung - CERN-TH colloquium 2000
EXTRA MATERIAL
ROBUST STATISTICS: A POWERFUL TEST

Try harder!

New Fitting Methodology

Define Underlying Physical Law
ie input PDFs from MSTW08, CT10, NNPDF2.3...

Generate random pseudo-data for the NNPDF3.0 dataset from info of experimental uncertainties and correlations

Perform (NN)PDF fit

Validate resulting PDF set:
- Reproduce input PDFs
- Both central values and uncertainties
- Expected values of $\chi^2$ are determined by pseudo-data
- PDF reweighting equal to refitting (Bayesian inference)

Fail?

Now you can fit real exp data!

Closure Test successful!

OK!
THEORY UNCERTAINTIES

- Theory is perturbative expansion to some order: 
  \[ t_p = \sum_{m=0}^{p} c_m \]

- Standard case: 
  \[ P(d|t_p) \propto \exp \left( -\frac{1}{2} (d - t_p)^T \text{cov}_{\exp}^{-1} (d - t_p) \right) \]

- Bayes’ theorem: 
  \[ P(t_p|d) = \frac{P(d|t_p)P(t_p)}{P(d)} \propto P(d|t_p)P(t_p) \]

- Assume Gaussian theory prior: 
  \[ P(t_p) = \prod_{m=0}^{p} P(c_m) \quad \text{where} \quad P(c_m) \propto \exp \left( -\frac{1}{2} c_m^T \text{cov}_{\text{th},m}^{-1} c_m \right) \]

- Assume MHOUs due to \( \mathcal{O}(a^{p+1}) \) terms only \( \rightarrow \) marginalise these terms: 
  \[ P(t_p|d) \propto \int dc_{p+1} P(d|c_{p+1})P(t_{p+1}) \]
  \[ \propto \exp \left( -\frac{1}{2} (d - t_p)^T (\text{cov}_{\exp} + \text{cov}_{\text{th}})^{-1} (d - t_p) \right) \]

- Include higher order terms by induction