PARTON DISTRIBUTIONS FOR BSM SEARCHES AT THE LHC

Maria Ubiali
(University of Cambridge)
In collaboration with:
* J. I. Latorre (Barcelona)
* R. D. Ball, L. Del Debbio, P. Mermaid (Edinburgh)
* Z. Dim, S. Forte (Milano), A. Guffanti (Torino)
* V. Bertone, N. Hartland, J. Rojo, L. Rottoli (Oxford)
Outline

- Motivation
- What (NN)PDFs are
- Impact of LHC data on PDFs
- Needs for BSM searches
- Conclusion
LHC physics at Run II

- Hadron colliders regarded as discovery machines, while lepton colliders seen as precision machines for characterisation.
- **LHC:** change of paradigm, getting close to precision physics at pp collider, thanks to theoretical and experimental progress.
- 20 years of exciting LHC physics in front of us and perturbative QCD could be the key for new discoveries.

---

LHC / HL-LHC Plan

[Diagram showing LHC and HL-LHC plan with key periods and luminosity milestones.]
LHC physics at Run II

Is the discovered scalar truly the SM Higgs?

→ Still substantial uncertainties

→ Need accuracy for indirect detection of new particles
LHC physics at Run II

Is the discovered scalar truly the SM Higgs?
- Still substantial uncertainties
- Need accuracy for indirect detection of new particles

Are there new particles within the reach of LHC Run-II?
- Need robust search strategies not to miss any signal
- Need solid predictions for SM background to establish significance and characterise it
Why PDFs?

Motivation
1) PDFs are ubiquitous

Proton Proton X-section = Parton Distribution Functions \times Partonic X-section

Lepton pair production
1) PDFs are ubiquitous

Proton Proton X-section = Parton Distribution Functions \times Partonic X-section

Higgs production

short distance

long distance
1) PDFs are ubiquitous

Proton Proton X-section = Parton Distribution Functions ⊗ Partonic X-section

New physics particle production and decay
2) The role of PDF uncertainty

PDF uncertainties are a limiting factor in the accuracy of theoretical predictions, both within SM and beyond.

<table>
<thead>
<tr>
<th>Higgs Production Channel</th>
<th>% theo. uncertainty</th>
<th>xsec @13 TeV</th>
</tr>
</thead>
<tbody>
<tr>
<td>ggF</td>
<td><img src="image" alt="ggF" /></td>
<td>43.9 pb</td>
</tr>
<tr>
<td>VBF</td>
<td><img src="image" alt="VBF" /></td>
<td>3.75 pb</td>
</tr>
<tr>
<td>WH</td>
<td><img src="image" alt="WH" /></td>
<td>1.38 pb</td>
</tr>
<tr>
<td>ZH</td>
<td><img src="image" alt="ZH" /></td>
<td>0.87 pb</td>
</tr>
<tr>
<td>ttH</td>
<td><img src="image" alt="ttH" /></td>
<td>0.51 pb</td>
</tr>
<tr>
<td>bbH</td>
<td><img src="image" alt="bbH" /></td>
<td>0.51 pb</td>
</tr>
</tbody>
</table>

Pre-YR4 numbers from HXSWG Wiki for mH = 125 GeV
2) The role of PDF uncertainty

ggF @ NNNLO

D. Wackeroth’s talk at KITP

Anastasiou et al.
PRL 114(2015) 212001

gluino pair production

Mw determination

Beenakker et al.
EPJC76 (2016)2, 53

\[ K_{NLO+NLL}(pp \rightarrow gg + X) \]
\[ \sqrt{s} = 13 \text{ TeV} \]

\[
\begin{array}{c|c|c|c|c}
\Delta M_W [\text{MeV}] & \text{present} & \text{CDF} & \text{D0} & \text{combined} \\
\hline
\mathcal{L} [\text{fb}] & 7.6 & 10 & 10 & 20 \\
PDF & 10 & 5 & 5 & 5 \\
QED rad. & 4 & 4 & 3 & 3 \\
\rho_T(W) model & 2 & 2 & 2 & 2 \\
other systematics & 9 & 4 & 11 & 4 \\
W statistics & 9 & 6 & 8 & 5 \\
Total & 16 & 10 & 15 & 9 \\
\end{array}
\]
3) The choice of PDFs matters

- A reliable understanding of PDF uncertainties plays a crucial role in precision physics.

- How do we interpret the difference predictions using different PDF sets?

- Shall we just pick a set out of the PDFs “supermarket” shelf or take the envelope of ALL predictions?
3) The choice of PDFs matters

- What does PDF uncertainty include? How reliable is it?
- How do we interpret the difference predictions using different PDF sets?
- Shall we just pick a set out of the PDFs “supermarket” shelf or take the envelope of ALL predictions?

\[ \delta \sigma_{PDF} = 2\% \]

\[ \delta \sigma_{PDF} = 5\% \]
3) The choice of PDFs matters

- What does PDF uncertainty include? How reliable is it?

- How do we interpret the difference predictions using different PDF sets?

- Shall we just pick a set out of the PDFs “supermarket” shelf or take the envelope of ALL predictions?
What PDFs are
Collinear Factorisation Theorem

\[
\frac{d\sigma_{pp \to ab}}{dX} = \sum_{i,j=1}^{N_f} f_i(x_1, \mu_F) f_j(x_2, \mu_F) \frac{d\sigma_{ij \to ab}}{dX}(x_1 x_2 S_{\text{had}}, \alpha_s(\mu_R), \mu_F) + \mathcal{O}\left(\frac{\Lambda_{\text{QCD}}^{2n}}{S_{\text{had}}^n}\right)
\]

\[
\mu^2 \frac{\partial f(x, \mu^2)}{\partial \mu^2} = \int_z^1 \frac{dz}{z} \frac{\alpha_s}{2\pi} P(z) f\left(\frac{x}{z}, \mu^2\right)
\]

Dokshitzer, Gribov, Lipatov, Altarelli, Parisi
renormalization group equations

Q-dependence: pert. theory

x-dependence: from data

LO - Dokshitzer; Gribov, Lipatov; Altarelli, Parisi, 1977

NLO - Floratos, Ross, Sachrajda; Floratos, Lacaze, Kounnas, Gonzalez-Arroyo, Lopez, Yndurain; Curci, Furmanski Petronzio, 1981

NNLO - Moch, Vermaseren, Vogt, 2004
The PDF extraction process

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

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

**LHAPDF interface**

http://lhapdf.hepforge.org
A steady progress

- **< 2002**: sets without uncertainty
- **2003-2004**: first MRST, CTEQ, Alekhin sets with uncertainties
- **2004-now**: huge progress made in statistical and theoretical understand, new players

PDG “Structure Functions” 2013

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

Experimental Data

MC generation

NN parametrization

Fi, i=1,...,Ndata
NMC, BCDMS, SLAC, HERA, CHORUS...

Fi(1), Fi(2), Fi(N-1), Fi(N)

TRAINING

q0net(1), q0net(2), q0net(N-1), q0net(N)

EVOLUTION

REPRESENTATION OF PROBABILITY DENSITY

\[
\left< F_i^{\text{net}} \right>_{\text{rep}} = \frac{1}{N_{\text{rep}}} \sum_{k=1}^{N_{\text{rep}}} F_i^{\text{net}(k)}
\]
The NNPDF approach

The N(eural)N(etwork)PDFs:

- Monte Carlo techniques: sampling the probability measure in PDF functional space
- Neural Networks: all independent PDFs are associated to an unbiased and flexible parametrization: $O(300)$ parameters versus $O(20)$ in polynomial parametrization

✓ Precise error estimate not driven by theoretical prejudice
✓ Statistical interpretation of uncertainty bands
Advantages

- No need to add new parameters when new data are included
- Reliable estimate of theoretical uncertainties not driven by parametrisation bias
- Possibility to include data via re-weighting: no need to refit
Past frontiers

- First NNPDF set - only DIS data (2008)
- Determination of the proton strangeness: solved NuTeV anomaly (2009)
- First NNPDF global set (2010)
- Heavy quark mass effects (2011)
  Determination of $\alpha_s$ from PDF fit
- First PDF set with threshold resummation (2014)
- First PDF set with methodology validated with closure test
- First PDF set with fitted photon PDF (2013)
- First PDF set with LHC data
- Reweighting PDFs (2012)
- First PDF set with fitted charm (2016)
  First PDF set with fitted charm
Experimental data
The data (before LHC)

NC \[ F_1^{\gamma,Z} = \sum_i e_i^2 (q_i + \bar{q}_i) \]
CC \[ F_1^{W^+} = \bar{u} + d + s + c \]
CC \[ -F_3^{W^+}/2 = \bar{u} - d - s + c \]
\[ F_2 = 2x F_1 \]

HERA DIS data
- Backbone of any PDF fit
- Structure functions known up to order $\alpha_s^3$
- Constrain $q, q\bar{q}$ at $10^{-4}$
- Constrain $g$ at small and moderate $x$
The data (before LHC)

Deuteron data:
- disentangle isospin triplet and singlet contributions
- Constrain strange and anti-strange at moderate $x > 10^{-2}$

Fixed Target DIS data
The data (before LHC)

- Constrain light quark and antiquark separation
- Up and down separation
- Ubar and Dbar separation

DY and EW vector boson data
The data (before LHC)

- Constrain quarks and gluons at large $x$
- So far cross section known only at NLO + threshold approximation
PDF uncertainties are a crucial input at the LHC, often being the limiting factor in the accuracy of theoretical predictions, both SM and BSM.

Exploit the power of precise LHC data to reduce PDF uncertainties and discriminate among PDF sets.
The LHC data

- Inclusive jets and dijets
  - (medium/large $x$)
- Isolated photon and $\gamma$+jets
  - (medium/large $x$)
- Top pair production (large $x$)
- High $p_T V$(+jets) distribution
  - (small/medium $x$)

- High $p_T W$(+jets) ratios
  - (medium/large $x$)
- $W$ and $Z$ production
  - (medium $x$)
- Low and high mass Drell-Yan
  - (small and large $x$)
- $W_c$ (strangeness at medium $x$)

- Low and high mass Drell-Yan
- WW production
## Effect of LHC data on PDFs

<table>
<thead>
<tr>
<th>Experiment</th>
<th>PDF Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATLAS jets 2.76 TeV and 7 TeV</td>
<td>gluon large x</td>
</tr>
<tr>
<td>ATLAS high-mass DY at 7 TeV</td>
<td>q/q~ separation</td>
</tr>
<tr>
<td>ATLAS W pT data at 7 TeV</td>
<td>g and q at moderate x</td>
</tr>
<tr>
<td>CMS (Y,M) double diff distributions 7 TeV</td>
<td>flavour separation</td>
</tr>
<tr>
<td>CMS jets at 7 TeV</td>
<td>gluon large x</td>
</tr>
<tr>
<td>CMS muon charge asymmetry at 7 TeV</td>
<td>quark separation</td>
</tr>
<tr>
<td>CMS W+c at 7 TeV</td>
<td>strangeness</td>
</tr>
<tr>
<td>LHCb Z rapidity distribution at 7 TeV</td>
<td>small/large x quarks</td>
</tr>
<tr>
<td>ATLAS+CMS tt total xsec at 7/8 TeV</td>
<td>gluon large x</td>
</tr>
</tbody>
</table>

NNPDF3.0
## Effect of LHC data on PDFs

### ATLAS jets 2.76 TeV and 7 TeV + 2011 data 7 TeV
- gluon large x

### ATLAS high-mass DY at 7 TeV + low mass
- q/q~ separation

### ATLAS W pT data at 7 TeV + ATLAS & CMS double diff Z pT
- g and q at moderate x

### CMS (Y,M) double diff distributions 7 TeV + 8 TeV
- flavour separation

### CMS jets at 7 TeV + 2.76 and 8 TeV jet data
- gluon large x

### CMS muon charge asymmetry at 7 TeV + 8 TeV
- quark separation

### CMS W+c at 7 TeV
- strangeness

### LHCb Z rapidity distribution at 7 TeV + 8 TeV (legacy data)
- small/large x quarks

### ATLAS+CMS tt total xsec at 7/8 TeV + differ. distributions
- gluon large x

### D0 legacy W asymmetry data
- q/q~ separation
Effect of LHC data on PDFs

Data give increasingly stronger constraints in known and less-known kinematic regions => PDF experimental uncertainties reduced

In precision region are we keeping up with theory settings in PDF fits?

Large x still affected by huge uncertainties

Ball et al.
JHEP 1504 (2015) 040
NNPDF3.0 / CT14 / MMHT

- common value of $\alpha_S(M_Z) = 0.118$
- comparable GM-VFN schemes for inclusion of HQ masses
- global sets: inclusion of $O(4000)$ experimental data
- extensive benchmarking

J. Butterworth et al
State of the art

ABM 12
- fitted $\alpha_s(M_Z)=0.1132$
- Fixed-Flavour-Number scheme

HERAPDF2.0
- HERA-only data
Needs for BSM searches
Large-x gluon/quarks

- Large-x g/q uncertainty can be reduced thanks to inclusion of LHC data
- NNLO calculation now available for some key processes for PDF determination
- Great progress also in tools to interface NLO codes to PDF fitting code

- NNLO top pair production
  Czakon, Fiedler, Mitov [PRL 116(2016) 082003]
  Czakon, Mitov [JHEP 1301(2015)]

- W/Z+j and W/Z transverse momentum distributions
  Gehrmann-De Ridder et al [1605.04295]
  Boughezal, Liu, Petriello [1602.08140]
  Boughezal, Liu, Petriello [1602.06965]
  Boughezal et al [PRL 116(2016) 152001 & 062002]
  Gehrmann-De Ridder et al [1507.02850]

- Inclusive jet cross section
  Currie et al [JHEP 1401 (2014) 110 ]
  Gehrmann-De Ridder et al [PRL 110 (2016) 162003]
Top data

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

Total cross section →

Differential cross section ↓

Inclusion of top pair production data (total cross section and differential distributions) competitive to jets data and cleaner from non-perturbative effects

Czakon et al [JHEP 1307 (2013) 167]
Beneke et al [JHEP 1207 (2012) 194]

Courtesy of J. Rojo
Czakon, Hartland, Mitov, Nocera and Rojo, in preparation
Experimental precision < 1% up to \( p_T \sim 200 \) GeV

Expect a great impact on the quark-gluon luminosity

To fit the data NNLO corrections are needed, discrepancies in non-normalised distributions

ATLAS collaboration [1512.02192]
Jets data

- Plenty of data from LHC
- NNLO corrections only partially known (gg channel)
- Several PDF groups make different choices: CT14 includes all jet data in NNLO fit assuming overall C-factor small, MMHT14 and ABM12 do not include LHC jet data at NNLO, NNPDF3.0 include some jet data based on goodness of threshold approximation
- These choices affect precision of the gluon, full NNLO calculation is very much needed
EW corrections and photon PDF

- EW corrections become relevant at the current precision level as are sizeable at large invariant mass.
- Full inclusion of EW corrections requires initial $\gamma$ PDF, which induces large uncertainty.

Bertone et al [JHEP 1511 (2015) 194]
The photon PDF

- **NNPDF23QED** provides $\gamma$ PDF and its uncertainty at (N)NLO QCD + LO QED, by reweighting photon PDF
  Ball et al [Nucl.Phys. B877 (2013)]

- **CT14QED** set based on two-parameter ansatz from model of photon radiate from valence quarks (extension to MRST2004QED model)
  Schmidt et al [1509.02905]

\[
\begin{align*}
  f_{\gamma/p}(x, Q_0) &= \frac{\alpha}{2\pi} \left( A_u e_u^2 \bar{P}_{\gamma q} \circ u^0(x) + A_d e_d^2 \bar{P}_{\gamma q} \circ d^0(x) \right) \\
  f_{\gamma/n}(x, Q_0) &= \frac{\alpha}{2\pi} \left( A_u e_u^2 \bar{P}_{\gamma q} \circ d^0(x) + A_d e_d^2 \bar{P}_{\gamma q} \circ u^0(x) \right)
\end{align*}
\]

- $\gamma$ PDF poorly determined by DIS data. Need hadron collider processes where $\gamma$ contributes at LO (on-shell W,Z production and low/high mass DY)

- NNPDF plan: fit photon along with other PDFs (thanks to upgrade of APFEL - simultaneous diagonalization of QCD and QED evolution matrices - and APFELgrid - now includes photon-induced processes)
Large-x and resummation

- Multi-scale processes: $\log(Q_i/Q_j) = L$ arise, which may spoil perturbative expansion
- If $(\alpha_s \cdot L) \sim O(1)$ fixed order perturbative QCD is no longer justified
- Resummation effectively rearranges perturbative series

Various kinds of logs:

- $L = \log(1-x)$ threshold (soft-gluon) resummation
- $L = \log(1/x)$ high-energy (small-x) resummation
- $L = \log(pT/M)$ transverse momentum resummation

Ball et al, JHEP09(2015)091 in progress
Threshold-resummed PDFs will be suppressed as compared to fixed-order PDFs
Mostly due to enhancement of NLO+NLL xsecs used in the fit of DIS structure functions and DY distributions
This suppression partially or totally compensates enhancements in partonic cross sections
Phenomenologically relevant for new physics processes [Beenakker et al. EPJC76 (2016)2, 53]
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 that its understanding (theory or experiment) is incomplete.
- Set of conservative partons based on measure of consistency are crucial to systematically study inclusion of new data.

NNPDF collaboration, JHEP04(2015)040
Beyond LHC - 100 TeV

NNPDF2.3 NNLO $N_F = 6$

NNPDF2.3 NNLO, $x = 10^{-2}$

$NF=6$ vs $NF=5$
Conclusions

- Parton Distribution Functions essential ingredient for LHC phenomenology
- Accurate PDFs are required for precision SM measurements, Higgs characterisation and New Physics
- NNPDF approach provides parton distributions based on a robust, unbiased methodology, the most updated theoretical information and most relevant hard scattering data including LHC data
- Frontiers for PDFs and BSM searches:
  - Bring down uncertainty of large-x gluon and quarks via inclusion of new data
  - Bring the pQCD loop revolution & resummations into the PDF world
  - How not to include effects that go beyond DGLAP/SM formalism into PDF fits?
- At 100 TeV collider big large-x uncertainties, top-quark, larger photon contribution, larger impact of large-x and small-x resummation
- Choice of heavy flavour schemes is also crucial: 4FS versus 5FS versus 6FS
- A challenging and exciting road ahead!

THANK YOU!