# Parton Distribution Functions for beginners

#### KAROL KOVAŘÍK

INSTITUT FÜR THEORETISCHE PHYSIK WESTFÄLISCHE WILHELMS-UNIVERSITÄT MÜNSTER

> Westfälische Wilhelms-Universität Münster



### Recap of Lecture 1

Parton Distribution Functions - parameterisation @ initial scale (DGLAP initial condition)

hep-ph/0201195

CTEQ 6

$$x f_k(x, Q_0) = c_0 x^{c_1} (1 - x)^{c_2} e^{c_3 x} (1 + e^{c_4} x)^{c_5} \qquad k = u_v, d_v, g, \bar{u} + \bar{d}$$
  
$$\bar{l}(x, Q_0) / \bar{u}(x, Q_0) = c_0 x^{c_1} (1 - x)^{c_2} + (1 + c_3 x) (1 - x)^{c_4}$$
  
$$s = \bar{s} = 0.2 (\bar{u} + \bar{d}) \qquad c(x, Q_0^2) = b(x, Q_0^2) = 0$$

#### Parton Distribution Functions - Q dependance & evolution (DGLAP)

all contributions together mix different PDF via DGLAP equations

$$\frac{df_q(x,Q^2)}{d\log Q^2} = \frac{\alpha_s(Q^2)}{2\pi} \int_x^1 \frac{d\xi}{\xi} \left[ P_{qq}\left(\frac{x}{\xi}\right) f_q(\xi,Q^2) + P_{qg}\left(\frac{x}{\xi}\right) f_g(\xi,Q^2) \right]$$
$$\frac{df_g(x,Q^2)}{d\log Q^2} = \frac{\alpha_s(Q^2)}{2\pi} \int_x^1 \frac{d\xi}{\xi} \left[ P_{gg}\left(\frac{x}{\xi}\right) f_g(\xi,Q^2) + P_{gq}\left(\frac{x}{\xi}\right) f_q(\xi,Q^2) \right]$$

Westfälische Wilhelms-Universität Münster

#### Parton Distribution Functions

In how do we determine them ? What are the moving parts in a typical PDF fitting-machine ?



Parton Distribution Functions - experimental data & theory

- which data are included in a fit ? Which PDFs do they constrain ?
- kinematic cuts on data
- difference between global fits like CTEQ, MSTW, NNPDF and not so global HERApdf, ABM...

Neutral current DIS (HERA, SLAC, NMC, BCDMS)





Jet data from Tevatron & LHC (D0,CDF, ATLAS, CMS)



🗹 Neutrino DIS & di-muon

(CDHSW, CHORUS, NuTeV)



Westfälische Wilhelms-Universität Münster

**M** DIS data from HERA & fixed target experiments (SLAC,NMC,BCDMS)

If different observables used in the fit (HERA -  $\frac{d\sigma}{dxdy}$ , fixed target -  $F_2(x,Q^2)$ )

$$\frac{d\sigma}{dx \, dQ^2} = \frac{2\pi\alpha^2}{xQ^4} \left[ Y_+ F_2 - y^2 F_L \pm Y_- xF_3 \right]$$

• motivation to still use fixed target is different coverage in x-Q<sup>2</sup> plane & sensitivity to large-x



**M** DIS data from HERA & fixed target experiments (SLAC,NMC,BCDMS)

DIS @ low Q - dominated by photon exchange



NC DIS

$$\begin{split} F_{1,2} &= F_{1,2}^{\gamma\gamma} + \frac{2g_V}{4s_W^2 c_W^2} \frac{Q^2}{Q^2 + M_Z^2} F_{1,2}^{\gamma Z} + \frac{g_V^2 + a_V^2}{16s_W^4 c_W^4} \frac{Q^4}{(Q^2 + M_Z^2)^2} F_{1,2}^{ZZ} \\ F_3 &= \frac{2a_V}{4s_W^2 c_W^2} \frac{Q^2}{Q^2 + M_Z^2} F_3^{\gamma Z} + \frac{2g_V a_V}{16s_W^4 c_W^4} \frac{Q^4}{(Q^2 + M_Z^2)^2} F_3^{ZZ} , \end{split}$$

#### sensitive to quark & anti-quark PDF @ LO

quarks & anti-quarks enter together with different weights depending on the exchange vector boson
 access interference comparing different helicity leptons & electrons/positrons

$$\begin{split} F_2^{\gamma\gamma}(x,Q^2) &= x \sum_q e_q^2 \big[ q(x,Q^2) + \bar{q}(x,Q^2) \big] & \text{photon} \\ F_2^{\gamma Z}(x,Q^2) &= x \sum_i B_i \big[ q_i(x,Q^2) + \bar{q}_i(x,Q^2) \big] & \text{photon-Z interference} \\ x F_3^{\gamma Z}(x,Q^2) &= x \sum_i D_i \big[ q_i(x,Q^2) - \bar{q}_i(x,Q^2) \big] & \text{photon-Z interference} \\ & \underbrace{ = \sum_i D_i \big[ q_i(x,Q^2) - \bar{q}_i(x,Q^2) \big] }_{\text{Westfälische}} & \underbrace{ \text{Westfälische} }_{\text{Wilhelms-Universität}} \end{split}$$

MÜNSTER

Meutrino DIS & di-muon (CDHSW, CHORUS, NuTeV)

 $\circ$  neutrino DIS contributes to  $F_2(x,Q^2)$  and  $F_3(x,Q^2)$ 

Ifferent PDF combinations contribute to flavor separation together with NC DIS

$$N \longrightarrow X$$

$$F_2(x,Q^2) = x \sum_{q} \left[ q(x,Q^2) + \bar{q}(x,Q^2) \right]$$
$$xF_3(x,Q^2) = x \sum_{q} \left[ q(x,Q^2) - \bar{q}(x,Q^2) \right]$$

neutrino DIS data on protons are scarce and hard to come by (WA21/22)

neutrino DIS typically taken on nuclei - need for nuclear corrections

#### Charge current DIS on proton (HERA)

neutrino DIS can be replaced by CC DIS on protons (still experimentally challenging)

$$F_2^{W^{\pm}}(x,Q^2) = x(\bar{u} \pm d \pm s + \bar{c})$$

Westfälische Wilhelms-Universität Münster

#### 🗹 DIS @ NLO

#### sensitive to gluon PDF @ NLO

breaking of Bjorken scaling at small x driven by gluon PDF

$$Q^2 \frac{\mathrm{d}F_2}{\mathrm{d}Q^2} = \frac{\alpha_S}{2\pi} \sum e_i^2 \int_x^1 \frac{\mathrm{d}y}{y} P_{qg}(y) f_g(x/y, Q^2)$$

Iongitudinal structure function  $F_L(x,Q^2)$ 

$$F_L(x,Q^2) = F_2(x,Q^2) - 2xF_1(x,Q^2) = 0$$

#### Callan-Gross relation @ LO

- gluon not subdominant in  $F_L(x,Q^2)$  as in  $F_2(x,Q^2)$
- experimental separation of  $F_2(x,Q^2)$  and  $F_L(x,Q^2)$ requires measurements at different  $\sqrt{s}$  - lower statistics
- using differential cross-section in the fit effective separation of structure functions



Vestfälische Vilhelms-Universität Nünster

**Marcell**-Yan lepton pair production (W,Z production)



• DY dominated by photon exchange away from W & Z resonances  

$$\frac{d\sigma}{dQ^2dy} = \frac{4\pi\alpha^2}{9Q^2s} \sum_i e_i^2 \left[ q_i(x_a, Q^2)\bar{q}_i(x_b, Q^2) + a \leftrightarrow b \right]$$

$$Q^2 = (p_l + p_{\bar{l}})^2 \qquad x_{a,b} = \frac{Q}{\sqrt{s}} \exp(\pm y)$$

DY at the W & Z resonances - different PDF combinations



**M** Drell-Yan lepton pair production (W,Z production)



Interpretended States State

hadronic jet production at leading order proceeds through

 $qq \rightarrow qq \qquad qg \rightarrow qg \qquad gg \rightarrow gg$ 

• qq subprocess dominates high-Et jets but gluon important enough to allow jet data to put constraints on large-x gluon PDF

- combined with low-x constraints on gluon PDF from DIS and with sum rules one has strong constraints on the gluon PDF
- additional direct probes of gluon PDF needed to constrain the gluon PDF at mid-x and large-x for future searches e.g. SUSY @ LHC



#### Using LHC data in Parton Distribution Functions

Motto: "Yesterday's signal is today's background"

#### **M** Top pair production

sensitive to gluon PDF at high-x

- very precise top pair production expected from LHC top-factory
- ratios of top/anti-top cross-sections sensitive also to u/d

#### **M** Direct photons

additional, complementary probe of gluon PDF (same x as gg Higgs production)

39

#### **M**+charm production

sensitivity to strange quark PDF (difficult to extract elsewhere)







Westfälische Wilhelms-Universität Münster

#### • Theory predictions for relevant observables

- Theory predictions can be done at LO, NLO or NNLO (at the moment)
  - at each order PDF has modified meaning need to use LO PDF with LO predictions etc.
    - leading-order
      - hard scattering results simple but with no scale dependance (no cancellation large scale dependance)
      - LO PDF useful for some MC applications where only LO exists
      - data descriptions unsatisfactory

#### next-to-leading-order

- scale cancellation between hard scattering and PDF lesser scale dependance
- hard scattering matrix elements complicated & need to be evaluated many times in a fit
- NLO hard matrix elements together with NLO DGLAP current state-of-the-art

#### next-to-next-to-leading-order

- NNLO splitting functions known but not all relevant hard scattering matrix element known e.g. jets or other additional processes

> Vestfälische Vilhelms-Universität Nünster

#### • Theory predictions for relevant observables

- NLO & NNLO theory predictions complicated functions need a way how to evaluate them quickly & efficiently
- Full cross-section is a convolution of hard matrix elements with PDFs (or two convolutions)

$$\sigma(\mu_r, \mu_f) = \sum_{n,i} c_{n,i}(x_a, x_b, \mu_r, \mu_f) \otimes \left[\alpha_s^n(\mu_r) F_i(x_a, x_b, \mu_f)\right]$$

Old (but effective) method - use K-factors - lose some (N)NLO information about shape
 Alternative - decouple PDFs and strong coupling dependance by putting everything on a (x,Q) grid & pre-compute complicated matrix element once and for all

$$\sigma(\mu) \simeq \sum_{n,i,k,l,m} \tilde{\sigma}_{n,i,k,l,m}(\mu) \, \alpha_s^n(\mu^{(m)}) \, F_i(x_a^{(k)}, x_b^{(l)}, \mu^{(m)})$$

FastNLO hep-ph/0609285

APPLGRID

MS-UNIVERSITÄT

#### Parton Distribution Functions

In how do we determine them ? What are the moving parts in a typical PDF fitting-machine ?



Parton Distribution Functions -  $\chi^2$ -fit & errors

 ${\scriptstyle \diamond}$  Most PDF fitters use  $\chi^2$  - function to measure the goodness of the fit

standard definition

$$\chi^2 = \sum_{i=1}^{N_{\text{dat}}} \left(\frac{D_i - T_i}{\sigma_i}\right)^2$$

definition with correlated errors

$$\chi^{2} = \sum_{i=1}^{N_{\text{dat}}} \sum_{j=1}^{N_{\text{dat}}} (D_{i} - T_{i})(V^{-1})_{ij}(D_{j} - T_{j}) \qquad V_{ij} = \delta_{ij}(\sigma_{i}^{\text{uncorr}})^{2} + \sum_{k=1}^{N_{\text{corr}}} \sigma_{k,i}^{\text{corr}} \sigma_{k,j}^{\text{corr}}$$
covariance matrix

Try to use all possible experimental information available

- statistical errors
- systematic errors (un)correlated
- normalisation uncertainty (might be multiplicative)

/estfälische /ilhelms-Universität 1ünster

#### Parton Distribution Functions - what do we get ?

What are the features we see in a typical PDF result ?

#### I u&d quarks - valence & sea

- valence part causes u&d dominate all other PDF at large-x where u>d
- symmetric sea-quark: q & anti-q comparable at low-x
- at high Q contribution of the sea component increases through gluon radiation (DGLAP)

#### **M** strange quarks

strange quark PDF suppressed at initial scale but enhanced at high-Q



#### Parton Distribution Functions - what do we get ?

What are the features we see in a typical PDF result ?

#### 🗹 gluon

- o dominate at small-x but fall off steeply as x increases
- going to high-Q gluon radiation reduces momenta of partons everything shifts to smaller x
- gluon radiates q-qbar pairs or additional gluons at small-x gluon PDF and sea quark PDF get steeper
- gluon can radiate even heavy quarks at high-Q so charm and bottom PDF are non-zero



Parton Distribution Functions - what are PDF uncertainties

#### **M** error PDF

- uncertainty of experimental data can be interpreted as uncertainty of the underlying PDF parameters
- different approaches how to translate
   experimental uncertainties to PDFs

#### *Solution of the sector of the*

- choice of data sets or observables (include neutrino DIS or not, LHC or not ...)
- choice of kinematic cuts (looser cuts might constrain PDF better but ...)
- parameterisation bias
- pQCD choices (NLO vs NNLO, strong coupling)
- heavy-quark schemes (FFS, ZM-VFNS, VF-VFNS)

higher-twist terms, nuclear corrections etc...

Westfälische Wilhelms-Universität Nünster

#### Parton Distribution Functions - $\chi^2$ -fit & errors

- error PDFs are experimental errors translated to errors of free PDF parameters
- all approaches to determine error PDFs give approx. the same results in regions with data

#### Messian method

the most widely used technique to determine error PDFs



#### Parton Distribution Functions - $\chi^2$ -fit & errors

- error PDFs are experimental errors translated to errors of free PDF parameters
- all approaches to determine error PDFs give approx. the same results in regions with data

#### Messian method

the most widely used technique to determine error PDFs

Expansion of 
$$\chi^2$$
:  

$$\chi^2(a) = \chi_0^2 + \frac{1}{2} \frac{\partial \chi^2}{\partial a_i \partial a_j} (a - a_0)_i (a - a_0)_j + \dots \rightarrow \chi_0^2 + \sum_i^N z_i^2$$
Choice of  $\Delta \chi^2 = \chi^2 - \chi_0^2$ :  
ideal choice  $\Delta \chi^2 = 1$  pragmatic choice  $\Delta \chi^2 \gg 1$   $\Delta \chi^2 \sim 50 - 100$   
error PDFs  
Construct error PDFs for each parameter in 2 directions:  
 $z_i = \pm \sqrt{\Delta \chi^2}$ 
 $X_i^{\pm}(z) = X_i^{\pm}(0, 0, \dots, \pm \sqrt{\Delta \chi^2}, \dots, 0, 0)$   
Calculate PDF uncertainty of cross-section

Westfälische Wilhelms-Universität Münster

 $(\Delta\sigma)^2 \approx \frac{-}{4} \sum_i \left( \sigma(X_i^+) - \sigma(X_i^-) \right)$ 

#### Parton Distribution Functions - $\chi^2$ -fit & errors

#### Messian method - dynamical tolerance criterion

• Ideal case would require using  $\Delta \chi^2 = 1$  for one sigma (68% CL) or  $\Delta \chi^2 = 2.71$  for 90% CL BUT we are fitting data from multiple not necessarily compatible experiments

49

#### dynamical tolerance by MSTW

 Each experiment (N data points) have to be described up to 90%CL for variations along one eigenvector

$$P_N(\chi^2) = \frac{(\chi^2)^{N/2 - 1} e^{-\chi^2/2}}{2^{N/2} \Gamma(N/2)}$$

$$\int_0^{\xi_{90}} \mathrm{d}\chi^2 P_N(\chi^2) = 0.90$$

- ${}^{\diamond}$  For each eigenvector, take  $\Delta\chi^2_n$  where all experiments are described within 90% CL
- Translate  $\Delta \chi^2_n$  to (different) shifts in each parameter



#### Parton Distribution Functions - $\chi^2$ -fit & errors

- error PDFs are experimental errors translated to errors of free PDF parameters
- all approaches to determine error PDFs give approx. the same results in regions with data

#### Monte Carlo method (as used by NNPDF) new technique which allows for more flexible PDF x-shapes

- $^{\circ}$  neural network is used to (over-)parametrize PDFs @ Q\_{0}
- $N_{
  m set}$  artificial replicas of data points generated assuming multi-Gaussian probability distribution
- random separation into training & validation data subsets
- minimize error function (not  $\chi^2$ ) for training set
- stop before overlearning

$$\langle \sigma(X) \rangle = \frac{1}{N_{\text{set}}} \sum_{i=1}^{N_{\text{set}}} \sigma(X^{i})^{\text{error/replica PDFs}}$$

$$\Delta\sigma(X) = \left(\sum_{i=1}^{N_{\text{set}}} \left[\sigma(X^i) - \langle\sigma(X)\rangle\right]^2\right)$$



# Misc topics in PDF

#### Parton Distribution Functions - theory & related issues

theory calculations @ LO, NLO or NNLO include several important constants which can have large impact on PDFs

#### $\mathbf{M}$ strong coupling $\mathbf{X}_{s}$

treated as an external parameter or fitted together with PDFs



#### $\mathbf{V}$ quark masses $m_c, m_b$

#### & heavy quark treatment

quark masses enter the evolution

#### **I** treatment of deuterium

fixed target DIS experiments provide important high-x constraints but done & their treatment influences gluon PDFs not on proton but deuterium





WESTFÄLISCHE WILHELMS-UNIVERSITÄT MÜNSTER

### Misc topics in PDF

#### **Strange quark PDF**

- not many data constrain strange quark PDF the best pre-LHC contraint comes from neutrino DIS and di-muon subset of DIS (CHORUS, NuTeV, NOMAD)
- new measurements of W+c production from ATLAS & CMS provided w constraints
- with no handle on the strange quark assumptions were introduced which tied strange PDF to other sea quarks - now these assumptions can be tested

$$s = \bar{s} = \frac{\kappa}{2} \left( \bar{u} + \bar{d} \right) \qquad \qquad r_s = \frac{1}{2} \left( s + \bar{s} \right) / \bar{d}$$



52

#### arXiv:1404.6469

#### Parton Distribution Functions - summary of existing PDFs

	DATA	Order	HQ	0Ls	Params.	Uncert.
<b>CT14</b>	global	LO,NLO, NNLO	GM-VFNS (s-ACOT)	external	6 indep. PDFs (26 params)	Hessian $(\Delta \chi^2 \sim 100)$
MSTW08	global	LO,NLO, NNLO	GM-VFNS (TR)	fit	7 indep. PDFs (20 params)	Hessian $(\Delta \chi^2 \sim 25)$
NNPDF	global	LO,NLO, NNLO	GM-VFNS (FONLL)	external	7 indep. PDFs (259 params)	Monte Carlo
CJ12	global	LO,NLO	ZM-VFNS	external	5 indep. PDFs (22 params)	Hessian $(\Delta \chi^2 = 100)$
HERApdf	DIS (HERA)	NLO NNLO	GM-VFNS (TR)	external	5 indep. PDFs (14 params)	Hessian $(\Delta \chi^2 = 1)$
<b>ABM11</b>	DIS+DY	NLO NNLO	FFN	fit	6 indep. PDFs (25 params)	Hessian $(\Delta \chi^2 = 1)$

Westfälische Wilhelms-Universität Münster

Parton Distribution Functions in Higgs production

Higgs is pre-dominantly produced through gluon fusion -

gluon PDFs at x=M<sub>H</sub>/ $\sqrt{s} \sim 0.02$  are crucial

sub-leading Higgs production via VBF is sensitive to quark & anti-quark PDFs



Parton Distribution Functions in Higgs production

Higgs is pre-dominantly produced through gluon fusion -

gluon PDFs at x=M<sub>H</sub>/ $\sqrt{s} \sim 0.02$  are crucial

sub-leading Higgs production via VBF is sensitive to quark & anti-quark PDFs



Parton Distribution Functions in top quark pair production

• Top quark pair production is dominated by s-channel diagrams where valence quarks & gluons are important at  $x=2m_t/\sqrt{s} \sim 0.05$ 

S- CHANNEL



NNLO gg luminosity at LHC ( $\sqrt{s} = 8$  TeV) 1.2 <sub>[]</sub> Ratio to MSTW 2008 (68% C.L.) MSTW 2008 1.15 CT10 NNPDF2.3 noLHC 1.1 NNPDF2.3 1.05 0.95 0.9 0.85 2m<sub>top</sub> M<sub>Higgs</sub> 0.8 1000 10 100  $\sqrt{\hat{\mathbf{s}}}$  (GeV) arXiv:1301.6754

T- CHANNEL





56

Parton Distribution Functions in SUSY production

Production of SUSY coloured particles (squarks & gluinos) very sensitive to gluon PDF at very high  $x=2m_X/\sqrt{s} \sim 0.2-0.7$ 

very problematic

57

SQUARK PRODUCTION





GLUINO PRODUCTION





# PDF after LHC

#### Parton Distribution Functions - new dedicated data

new projects with large possible impact on PDFs

#### LHeC

colliding electrons / positrons with LHC protons / nuclei

- unprecedented coverage in x-Q<sup>2</sup> plane
- precise determination of the gluon PDF
- interesting also for Higgs & BSM physics programs
- breakthrough machine for nuclear PDFs

#### EIC

electron ion collider

- high-intensity precision machine with polarized beams
- good coverage in x-Q<sup>2</sup> plane (down to  $x \sim 10^{-4}$ )
- o precise determination of the gluon PDF
- breakthrough machine for nuclear PDFs, saturation, polarized PDFs...

# Point 3.

206,2013



Energy

58



