PARTON DISTRIBUTIONS FUNCTIONS

STEFANO FORTE
UNIVERSITÀ DI MILANO & INFN

UNIVERSITÀ DEGLI STUDI DI MILANO

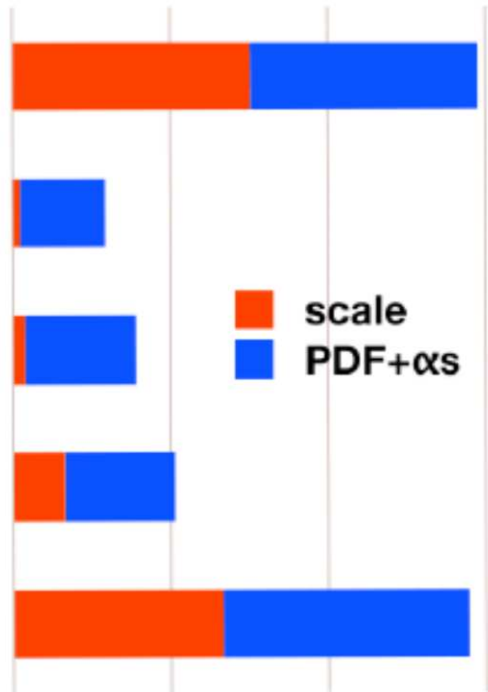
GRK 1694

FREUDENSTADT, SEPT. 26-28, 2016

THE IMPACT OF PDFS: HIGGS DISCOVERY

HIGGS PRODUCTION

		σ (8 TeV)	uncertainty
NNLL QCD +NLO EW	gg→H	19.5 pb	14.7%
	VBF	1.56 pb	2.9%
NNLO QCD +NLO EW	WH	0.70 pb	3.9%
	ZH	0.39 pb	5.1%
NLO QCD	ttH	0.13 pb	14.4%



(J. Campbell, HCP2012)

PDF UNCERTAINTY EITHER DOMINANT, OR VERY LARGE, OR BOTH

. . . AND NOT ONLY FOR THE HIGGS!

(W MASS DETERMINATION, NEW PHYSICS SEARCHES FOR HEAVY STATES,. . .)

SUMMARY

LECTURE I: THE BASICS

- FACTORIZATION
  - RENORMALIZATION AND FACTORIZATION IN QCD
  - ELECTROPRODUCTION AND HADROPRODUCTION
  - EVOLUTION EQUATIONS AND SUM RULES
- FROM DATA TO PDFs
  - DATA FROM HERA TO THE LHC
  - DISENTANGLING QUARK FLAVORS
  - DETERMINING THE GLUON

LECTURE II: PDF DETERMINATION

- STATISTICS AND METHODOLOGY
  - HESSIAN VS MONTE CARLO APPROACH
  - HESSIAN UNCERTAINTY ESTIMATES AND TOLERANCE
  - PARAMETRIZATION BIAS
  - GENERAL PARAMETRIZATIONS AND CROSS-VALIDATION
  - COMPRESSION METHODS AND MONTE CARLO ↔ HESSIAN CONVERSION
  - NON-GAUSSIAN BEHAVIOUR
  - CLOSURE TESTING

LECTURE III: THE STATE OF THE ART

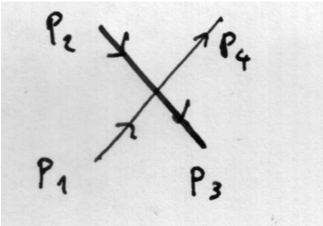
- THEORETICAL ISSUES
  - PERTURBATIVE STABILITY AND HIGHER ORDER CORRECTIONS
  - RESUMMATION
  - HEAVY QUARKS: RESUMMATION AND MATCHING
- PDFs NOW
  - GLOBAL AND NONGLOBAL PDF DETERMINATIONS
  - RECENT PROGRESS: METHODOLOGY AND LHC DATA
  - THE PDF4LHC COMBINED PDF SETS

FACTORIZATION

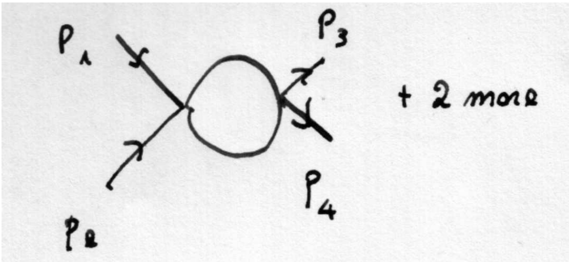
RENORMALIZATION

A QUICK REMINDER

$\mathcal{L} = -\frac{g}{24}\phi^4 \quad \phi\phi \to \phi\phi$ ELASTIC SCATTERING OF MASSIVE SCALAR FIELDS



$$\frac{d\sigma}{d\cos\theta} = \frac{g^2}{128\pi}\frac{1}{s}; \qquad s = (p_1 + p_2)^2$$



$$\frac{d\sigma}{d\cos\theta} = \frac{g^2}{128\pi}\frac{1}{s}F(s,t):\ \text{DIVERGES!};$$

$$t = (p_1 - p_3)^2,\ u = (p_1 - p_4)^2$$

$$F(s,t) = \lim_{\Lambda\to\infty} 1 + \frac{g}{32\pi}\left(3 + \int_0^1 \ln\frac{M^2(s)}{\Lambda^2} + s \to t + s \to u\right); \quad M^2(s) = m^2 - x(1-x)s$$

RENORMALIZATION: EXPRESS A PHYSICAL OBSERVABLE IN TERMS OF OTHER PHYSICAL OBSERVABLES:

WHAT IS THE CHARGE g? DEFINE $g_{\rm phys}$ FROM $\quad \left.\frac{d\sigma}{d\cos\theta}\right|_{s=4m^2} = \frac{g^2_{\rm phys}}{128\pi}\frac{1}{m^2}$:

$$\frac{d\sigma}{d\cos\theta} = \frac{g^2_{\rm phys}}{128\pi}\frac{1}{s}F(s,t); \quad F(s,t) = 1 + \frac{g_{\rm phys}}{32\pi}\left(\int_0^1 \ln\frac{M^2(s)}{M^2(4m^2)} + s \to t + s \to u\right)$$

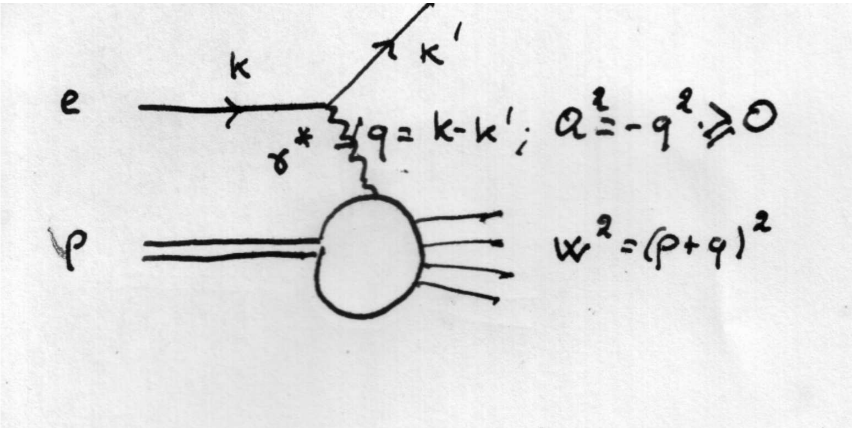
UV SINGULARITY IS UNIVERSAL $\Rightarrow$ REABSORBED IN DEF. OF THE COUPLING

FACTORIZATION

IN PERTURBATIVE QCD

DEEP-INELASTIC LEPTON-HADRON SCATTERING

PROBE THE PROTON WITH A SHORT-WAVELENGTH PHOTON:



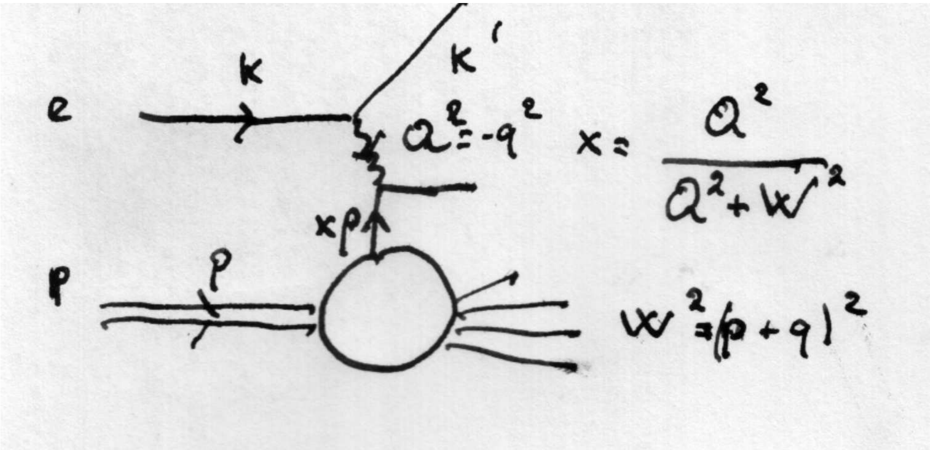
$e = P \to e + X$: for fixed energy, cross section depends on two variables $(\cos\theta,\, W^2)$

FACTORIZATION
IN PERTURBATIVE QCD

DEEP-INELASTIC LEPTON-HADRON SCATTERING

PROBE THE PROTON WITH A SHORT-WAVELENGTH PHOTON:

QCD IS ASYMPTOTICALLY FREE, USE PERTURBATION THEORY:



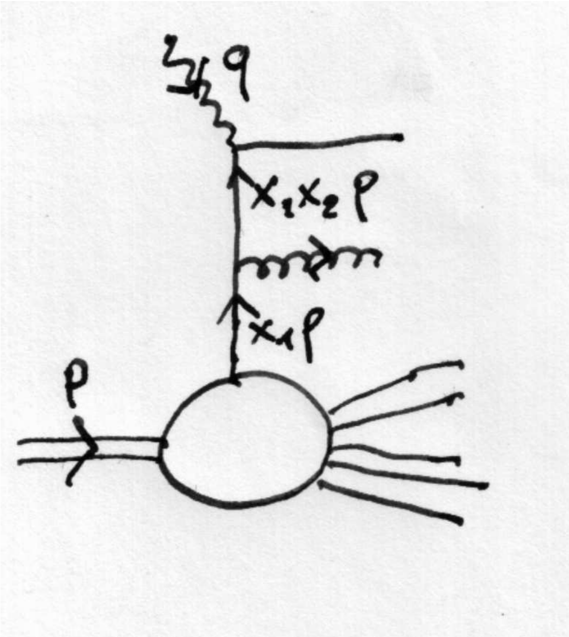
$$\frac{d\sigma}{d\cos\theta dW^2} = \sum_i e_i^2 \frac{d\hat{\sigma}}{d\cos\theta dW^2} q_i(x(\theta, W^2)) + O\left(\frac{M^2}{Q^2}\right)$$

- AT FIRST PERTURBATIVE ORDER, INCOHERENT SUM OF CONTRIBUTIONS FROM CHARGED CONSTITUENTS (QUARKS), PROPORTIONAL TO THEIR CHARGE
- MOMENTUM OF THE CONSTITUENT PROPORTIONAL TO PROTON MOMENTUM $\hat{p} = xp$
- "MOMENTUM FRACTION" x ENTIRELY FIXED BY KINEMATICS, BY Q^2 & W^2 (I.E. $\cos\theta$ & W^2)
- THE CROSS-SECTION IS PROPORTIONAL, UP TO KINEMATIC FACTORS, TO THE PROBABILITY $q_i(x)$ OF THE PHOTON STRIKING A QUARK OF THE i-TH FLAVOR OR ANTIFLAVOR WITH MOMENTUM $\hat{p} = xp$

PARTON MODEL; PARTON= STRUCK CONSTITUENT:
$q_i(x) \Rightarrow$ PARTON DISTRIBUTION (PDF)

FACTORIZATION
IN PERTURBATIVE QCD
DEEP-INELASTIC LEPTON-HADRON SCATTERING

WHAT HAPPENS AT HIGHER ORDERS?
COLLINEAR SINGULARITIES!

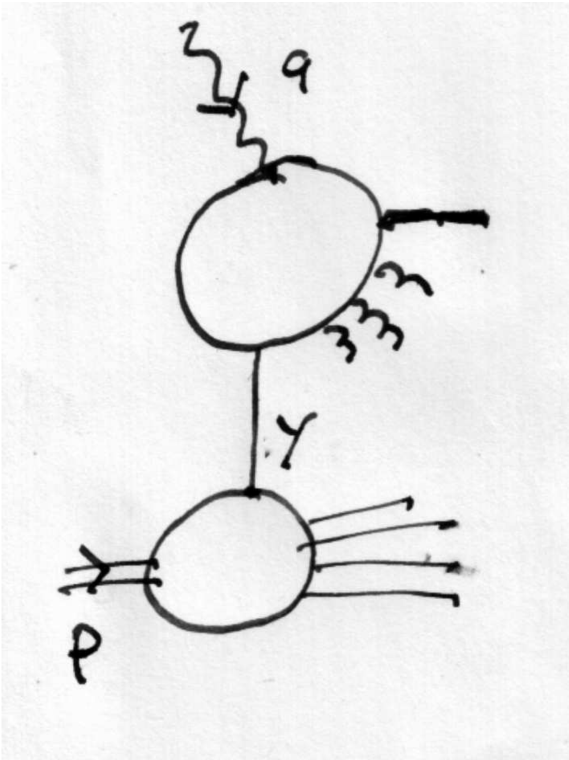


$$\frac{d\sigma}{d\cos\theta\, dW^2} = \sum_i e_i^2 \int \frac{d\hat\sigma}{d\cos\theta\, dW^2} P(x_1 x_2) q_i(x_1) dx_2 \ln\frac{Q^2}{M^2} + O\left(\frac{M^2}{Q^2}\right)$$
$$= \sum_i e_i^2 \frac{d\hat\sigma}{d\cos\theta\, dW^2}(x) q_i^{\rm phys}(x, Q^2) + O\left(\frac{M^2}{Q^2}\right)$$

- HIGHER ORDER CORRECTIONS ARE SINGULAR;
  SINGULARITY REGULATED BY THE PROTON SCALE M
- DEFINE A PHYSICAL PDF WITH SINGULARITY FACTORED IN IT $\Rightarrow$ SCALE DEPENDENT

FACTORIZATION
IN PERTURBATIVE QCD
DEEP-INELASTIC LEPTON-HADRON SCATTERING

WHAT HAPPENS AT HIGHER ORDERS?
COLLINEAR SINGULARITIES!

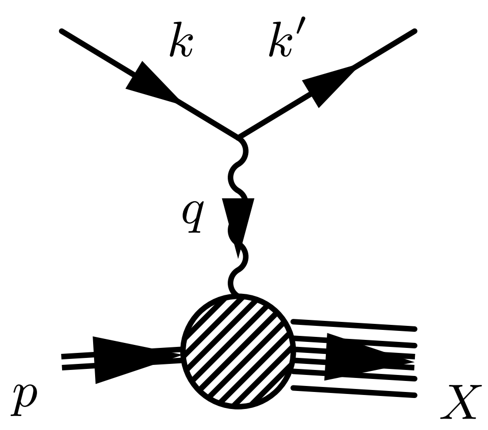


$$\frac{d\sigma}{d\cos\theta\, dW^2} = \sum_i e_i^2 \int \frac{d\hat\sigma}{d\cos\theta\, dW^2} P(x_1 x_2) q_i(x_1) dx_2 \ln\frac{Q^2}{M^2} + O\left(\frac{M^2}{Q^2}\right)$$
$$= \sum_i e_i^2 \int \frac{d\hat\sigma}{d\cos\theta\, dW^2} P(x) q_i^{\rm phys}\left(\frac{y}{x}, Q_0^2\right) \frac{dy}{y} \ln\frac{Q^2}{Q_0^2} + O\left(\frac{M^2}{Q^2}\right)$$

- EXPRESS THE PROCESS IN TERMS OF THE PROCESS AT ANOTHER SCALE
- UP TO POWER CORRECTIONS, PROCESS FACTORIZES (NO INTERFERENCE): PARTONIC PROCESS $\otimes$ SCALE DEPENDENT PDF
- SCALE DEPENDENCE CAN BE SEEN AS A BRANCHING DRIVEN BY THE SPLITTING FUNCTION KERNEL $P(x)$

DEEP-INELASTIC SCATTERING

THE STRUCTURE FUNCTIONS



Lepton fractional energy loss: $y = \frac{p\cdot q}{p\cdot k}$;

gauge boson virtuality: $q^2 = -Q^2$

Bjorken x: $x = \frac{Q^2}{2p\cdot q}$

lepton-nucleon CM energy: $s = \frac{Q^2}{xy}$;

virtual boson-nucleon CM energy $W^2 = Q^2\,\frac{1-x}{x}$;

$$\frac{d^2\sigma^{\lambda_p\lambda_\ell}(x,y,Q^2)}{dxdy} = \frac{G_F^2}{2\pi(1+Q^2/m_W^2)^2}\,\frac{Q^2}{xy}\left\{\left[-\lambda_\ell\, y\left(1-\frac{y}{2}\right) xF_3(x,Q^2) + (1-y)F_2(x,Q^2)\right.\right.$$

$$\left.\left. +y^2xF_1(x,Q^2)\right] - 2\lambda_p\left[-\lambda_\ell\, y(2-y)xg_1(x,Q^2) - (1-y)g_4(x,Q^2) - y^2xg_5(x,Q^2)\right]\right\}$$

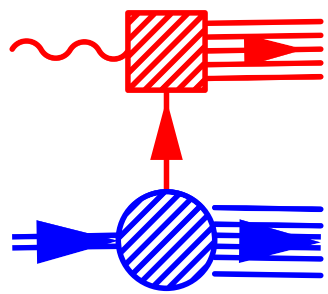
$\lambda_l \rightarrow$ lepton helicity
$\lambda_p \rightarrow$ proton helicity

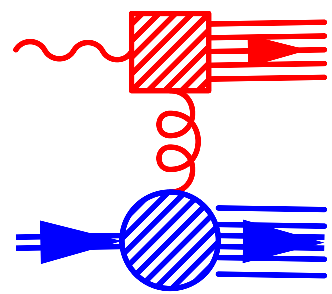
	PARITY CONS.	PARITY VIOL.
UNPOL.	$F_1,\ F_2$	F_3
POL.	g_1	$g_4,\ g_5$

FACTORIZATION:

STRUCTURE FUNCTIONS AND PDFs

STRUCTURE FUNCTION=HARD COEFF. (PARTONIC STRUCTURE FUNCTION) $\otimes$PARTON DISTN.

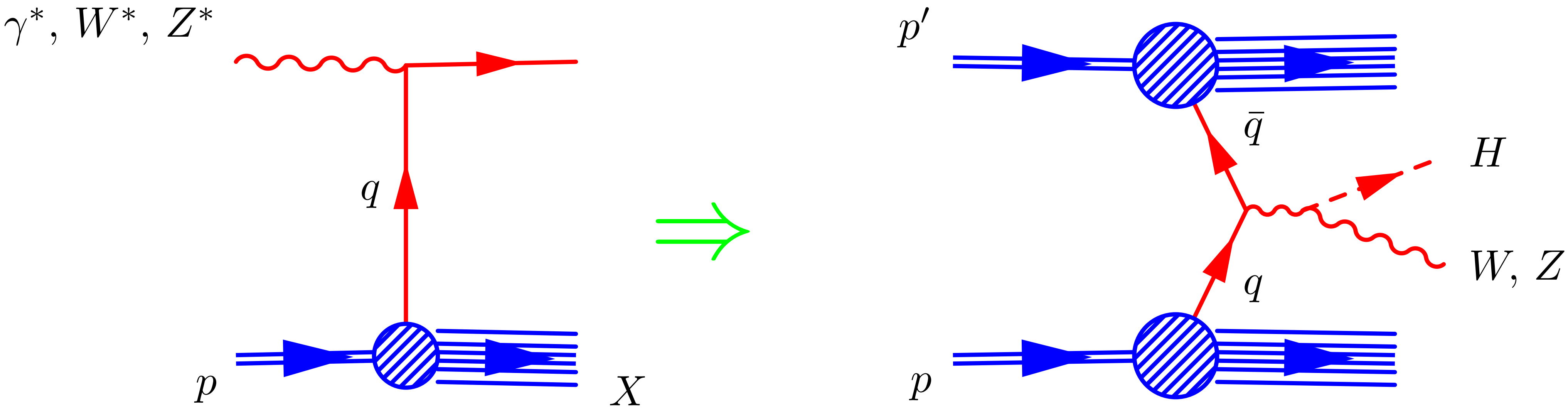




$$F_2(x,Q^2) = x\sum_i \int_1^1 \frac{dy}{y} C_i\left(\alpha_s(Q^2), \frac{x}{y}\right)\left[q_i(y,Q^2) + \bar{q}_i(y,Q^2)\right] + C_g\left(\alpha_s(Q^2), \frac{x}{y}\right) g(y,Q^2)$$

q_i quark, $\bar{q}_i$ antiquark, g gluon

FACTORIZATION
FROM DIS TO HADRONIC PROCESSES



- ONE PARTON PER HADRON: $\hat{p}_1 = x_a p_1$; $\hat{p}_2 = x_2 p_2$
- COLLINEAR EMISSION FROM PARTON LEGS
  $\Rightarrow$ UNIVERSAL (PROCESS-INDEPENDENT) REDEFINITION OF PDFs
- SUPPRESSION OF INTERFERENCE $\Rightarrow$ FACTORIZATION

HADRONIC PROCESSES

THE PARTON LUMINOSITY

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

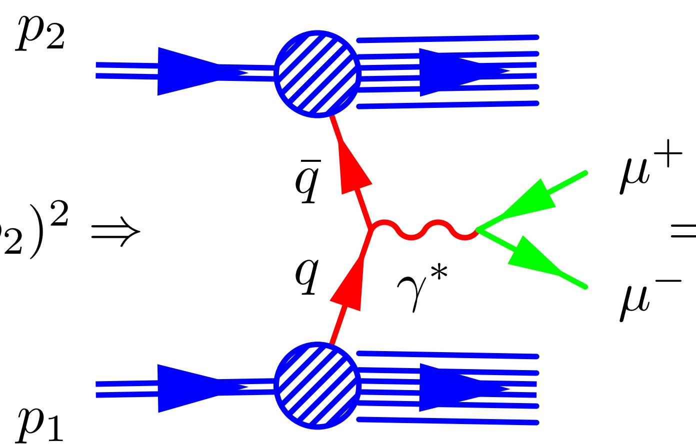
$$\sigma_X(s, M^2) = \sigma_0 \sum_{a,b} \int_\tau^1 \frac{dx}{x} \mathcal{L}_{ab}\left(\frac{\tau}{x}\right) C\left(x, \alpha_s(M_H^2)\right)$$

- PARTON LUMINOSITY $\mathcal{L}_{ab}(\tau) = \int_\tau^1 \frac{dx}{x} f_{a/h_1}(x) f_{b/h_2}(\tau/x)$

- COEFFICIENT FUNCTION $\hat\sigma_{q_a q_b \to X}\left(x_1 x_2 s, M_X^2\right) = \sigma_0 C\left(\frac{M_X^2}{x_1 x_2 s}, \alpha_s(M_H^2)\right)$

EXAMPLE: THE DRELL-YAN PROCESS
AT LEADING ORDER

$s = (p_1 + p_2)^2 \Rightarrow$



$\Rightarrow M_X^2$

- Hadronic c.m. energy: $s = (p_1 + p_2)^2$
- Momentum fractions $x_{1,\,2} = \sqrt{\frac{\hat s}{s}} \exp \pm y$;

  Lead. Ord. $\hat s = M^2$
- Partonic c.m. energy: $\hat s = x_1 x_2 s$
- Invariant mass of final state X (dilepton, Higgs,...):

  $M_W^2 \Rightarrow$ scale of process

- Scaling variable $\tau = \frac{M_X^2}{s}$

$$\Rightarrow M^2 \frac{d\sigma}{dM^2} = \sigma_0 \mathcal{L}(\tau); \quad \sigma_0 = \frac{4}{9}\pi\alpha\frac{1}{s};$$

HADRONIC FACTORIZATION
BEYOND TOTAL CROSS-SECTIONS

WITH MORE DIFFERENTIAL KINEMATICS, MUST IMPOSE EXTRA CONSTRAINTS $\Rightarrow$ MORE INFORMATION

EXAMPLE: THE DRELL-YAN RAPIDITY DISTRIBUTION

$$\frac{d\sigma_X(s,M_X^2)}{dY\,dM^2} =$$
$$\sum_{a,b}\int_{x_{\min}}^1 dx_1\,dx_2\,dy\,f_{a/h_1}(x_1)f_{b/h_2}(x_2)\frac{d\hat\sigma_{q_a q_b\to X}}{dy\,dM_X^2}\left(x_1x_2s,M_X^2\right)\delta\left(Y-\frac{1}{2}\ln\frac{x_1}{x_2}-y\right)$$

LEADING ORDER: $\frac{d\hat\sigma_{q_a q_b\to X}}{dy\,dM_X^2}=\sigma_0\delta(y)\delta(1-\frac{\tau}{x_1x_2})$

$$\Rightarrow \frac{d\sigma_X(s,M_X^2)}{dY\,dM^2}=\sigma_0\sum_{ab}f_a(x_1)fb(x_2);\quad x_i=\tau e^{\pm Y}$$

- RAPIDITY: LONGITUDINAL BOOST OF FINAL STATE WR TO CM OF HADRONIC COLLISION
- AT LEADING ORDER, INVARIANT MASS DETERMINES $\tau=x_1x_2$, HADRONIC RAPIDITY FIXES BOTH x_1, x_2

FACTORIZATION HOLDS FOR A WIDE CLASS OF SUFFICIENTLY INCLUSIVE OBSERVABLES (INCLUSIVE JETS, HIGGS AND GAUGE BOSON PRODUCTION CHANNELS, ETC.)
FAILS FOR EXCLUSIVE OBVSERVABLES (E.G. ELASTIC SCATTERING)

THE SCALE DEPENDENCE OF PDFs
EVOLUTION EQUATIONS

- DEFINE MELLIN MOMENTS OF PARTON DISTRIBUTIONS
  $f(N,Q^2) \equiv \int_0^1 dx\, x^{N-1} f_2(x,Q^2)$
  NOTE LARGE/SMALL $x \Leftrightarrow$ LARGE/SMALL N

- DEFINE LOGARITHMIC SCALE $t = \ln \frac{Q^2}{\Lambda^2}$:
  EVOLUTION GIVEN BY ORDINARY DIFFERENTIAL EQUATIONS (NO CONVOLUTION)

- ANOMALOUS DIMENSIONS RELATED TO DGLAP SPLITTING FUNCTIONS
  $\gamma(N,\alpha_s(t)) \equiv \int_0^1 dx\, x^{N-1} P(x,\alpha_s(t))$

$$\frac{d}{dt}\Delta q_{NS}(N,Q^2) = \frac{\alpha_s(t)}{2\pi}\gamma_{qq}^{NS}(N,\alpha_s(t))\Delta q_{NS}(N,Q^2),$$

$$\frac{d}{dt}\begin{pmatrix}\Delta\Sigma(N,Q^2)\\ \Delta g(N,Q^2)\end{pmatrix} = \frac{\alpha_s(t)}{2\pi}\begin{pmatrix}\gamma_{qq}^{S}(N,\alpha_s(t)) & 2n_f\gamma_{qg}^{S}(N,\alpha_s(t))\\ \gamma_{gq}^{S}(N,\alpha_s(t)) & \gamma_{gg}^{S}(N,\alpha_s(t))\end{pmatrix} \otimes \begin{pmatrix}\Delta\Sigma(N,Q^2)\\ \Delta g(N,Q^2)\end{pmatrix},$$

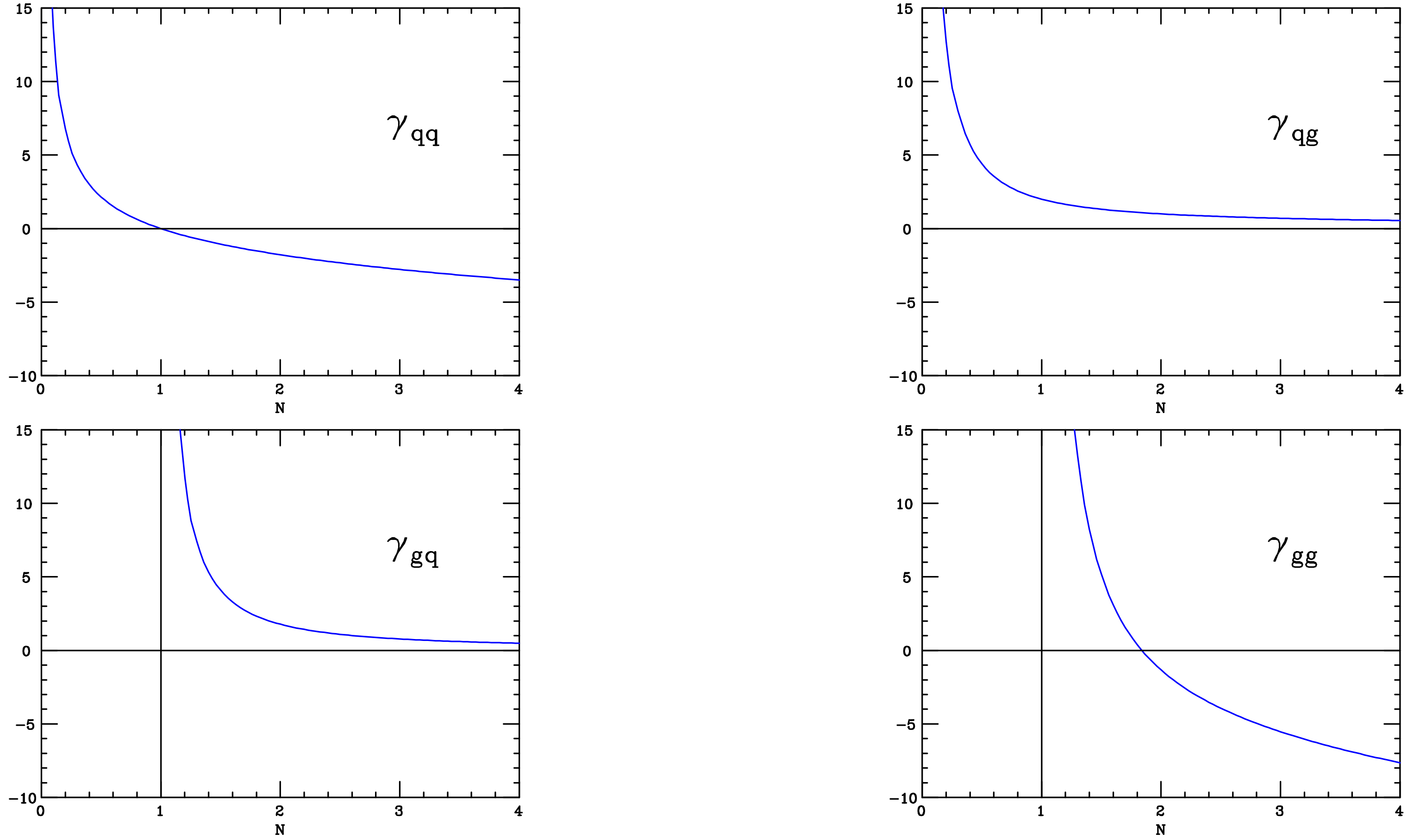
- EVOLUTION OF SINGLET $\Sigma(x,Q^2) = \sum_{i=1}^{n_f}\left(q_i(x,Q^2) + \bar{q}_i(x,Q^2)\right)$ COUPLED TO GLUON

- ALL "NONSINGLET" QUARK COMBINATIONS $q^{NS}(x,Q^2) = q_i(x,Q^2) - q_j(x,Q^2)$ EVOLVE INDEPENDENTLY

- ANOMALOUS DIMENSIONS COMPUTED IN PERTURBATION THEORY:
  $\gamma_i(N,\alpha_s(t)) = \gamma_i^{(0)}(N) + \alpha_s(t)\gamma_i^{(1)}(N) + \ldots$

PERTURBATIVE EVOLUTION

THE LEADING ORDER ANOMALOUS DIMENSIONS



QUALITATIVE FEATURES

- AS Q^2 INCREASES, PDFS DECREASE AT LARGE x & INCREASE AT SMALL x DUE TO RADIATION
- GLUON SECTOR SINGULAR AT $N = 1 \Rightarrow$ GLUON GROWS MORE AT SMALL x
- $\gamma_{qq}(1) = 0 \Rightarrow$ NUMBER OF QUARKS CONSERVED

SUM RULES

CONSTRUCT CONSERVED QUANTUM NUMBERS CARRIED BY PARTON DISTRIBUTIONS:

- BARYON NUMBER $\int_0^1 dx\,(u^p - \bar{u}^p) = 2 = 2\int_0^1 dx\,\left(d^p - \bar{d}^p\right)$
- MOMENTUM $\int_0^1 dx\,x \left[\sum_{i=1}^{N_f} \left(q^i(x) + \bar{q}_i(x)\right) + g(x)\right] = 1$

CANNOT DEPEND ON SCALE

- BARYON NUMBER $\gamma_{qq}(1) - \gamma_{q\bar{q}}(1) = 0$; AT LO $\gamma_{q\bar{q}}(1) = 0$ SO $\gamma_{qq}(1) = 0$
- MOMENTUM $\gamma_{qq}(2) + \gamma_{qg}(2) = 0$, $\gamma_{gq}(2) + \gamma_{gg}(2) = 0$

CAN EXTRACT FROM PHYSICAL OBSERVABLES: BARYON NUMBER

- GROSS–LLEWELLYN-SMITH SUM RULE $\frac{1}{2}\int_0^1 dx\,\left(F_3^{\nu p}(x,Q^2) + F_3^{\nu n}(x,Q^2)\right) =$
  $= C_{\mathrm{GLS}}(Q^2)\int_0^1 dx\,\left[u(x,Q^2) - \overline{u}(x,Q^2) + d(x,Q^2) - \overline{d}(x,Q^2)\right]$

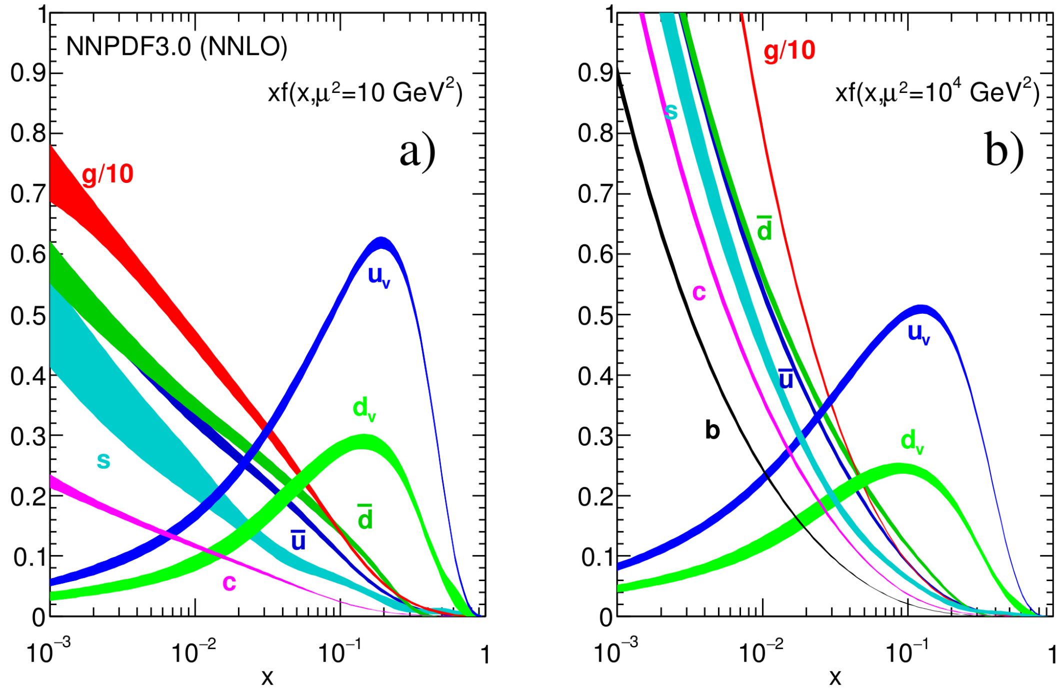
FACTORIZATION SUMMARY

- PHYSICAL OBSERVABLES INVOLVING THE STRONG INTERACTION ARE COMPUTABLE IN PERTURBATIVE QCD WHEN THEY INVOLVE A LARGE SCALE (ASYMPTOTIC FREEDOM)
- INITIAL-STATE (OR FINAL-STATE) HADRONS CAN BE TREATED THANKS TO FACTORIZATION;
  - COLLINEAR SINGULARITIES CAN BE REABSORBED INTO PDFS
  - THEY LEAD TO ENHANCEMENT OF FACTORIZABLE CONTRIBUTIONS (POWER SUPPRESSION OF INTERFERENCE)
  - OBSERVABLE "HADRONIC" CROSS-SECTIONS ARE A CONVOLUTION OF A "PARTONIC" PROCESS WITH INCOMING QUARKS AND GLUONS, TIMES PDFS
  - PDFS ARE UNIVERSAL, I.E. PROCESS-INDEPENDENT
- PREDICTIVITY AT A HADRON COLLIDER INVOLVES THE PERTURBATIVE COMPUTATION OF THE PARTONIC PROCESSES, AND A DETERMINATION OF PDFS
- PDFS ARE

  A WAY OF EXPRESSING A PROCESS IN TERMS OF OTHER PHYSICAL PROCESSES

FROM DATA TO PDFs

THE PDFs

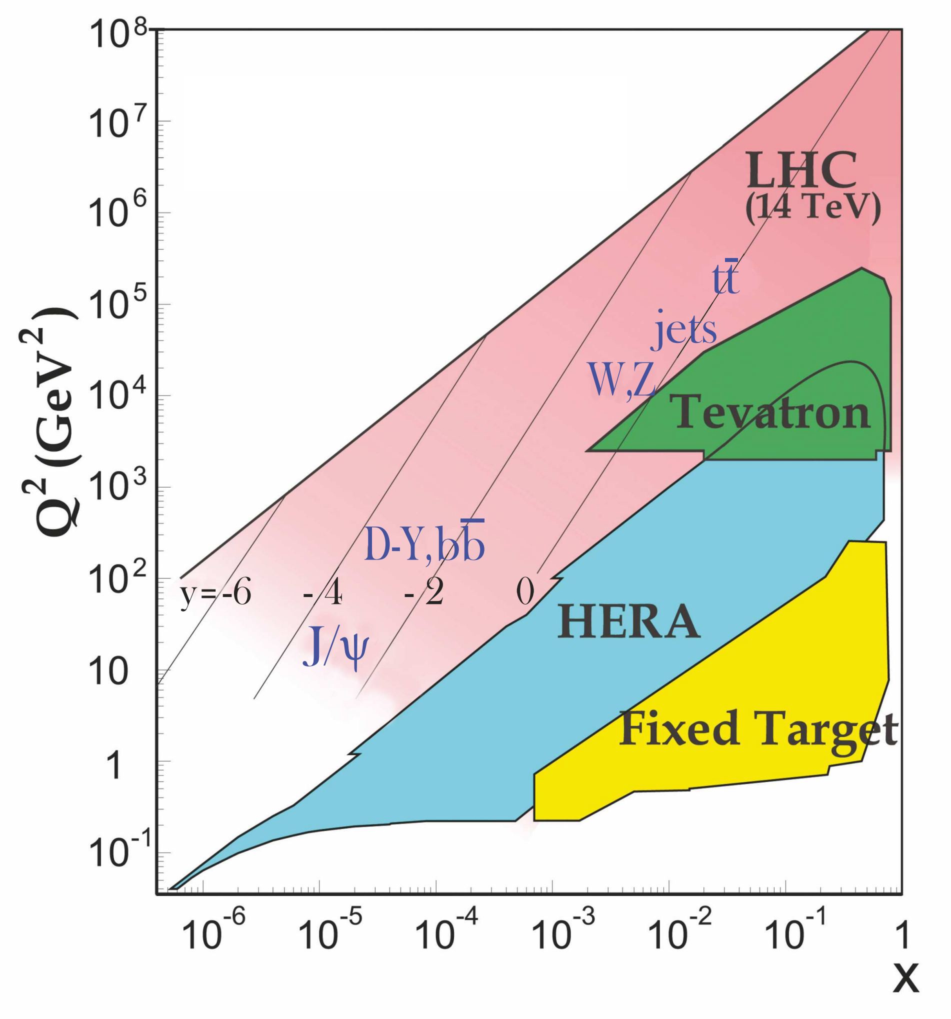


(PDG 2016)

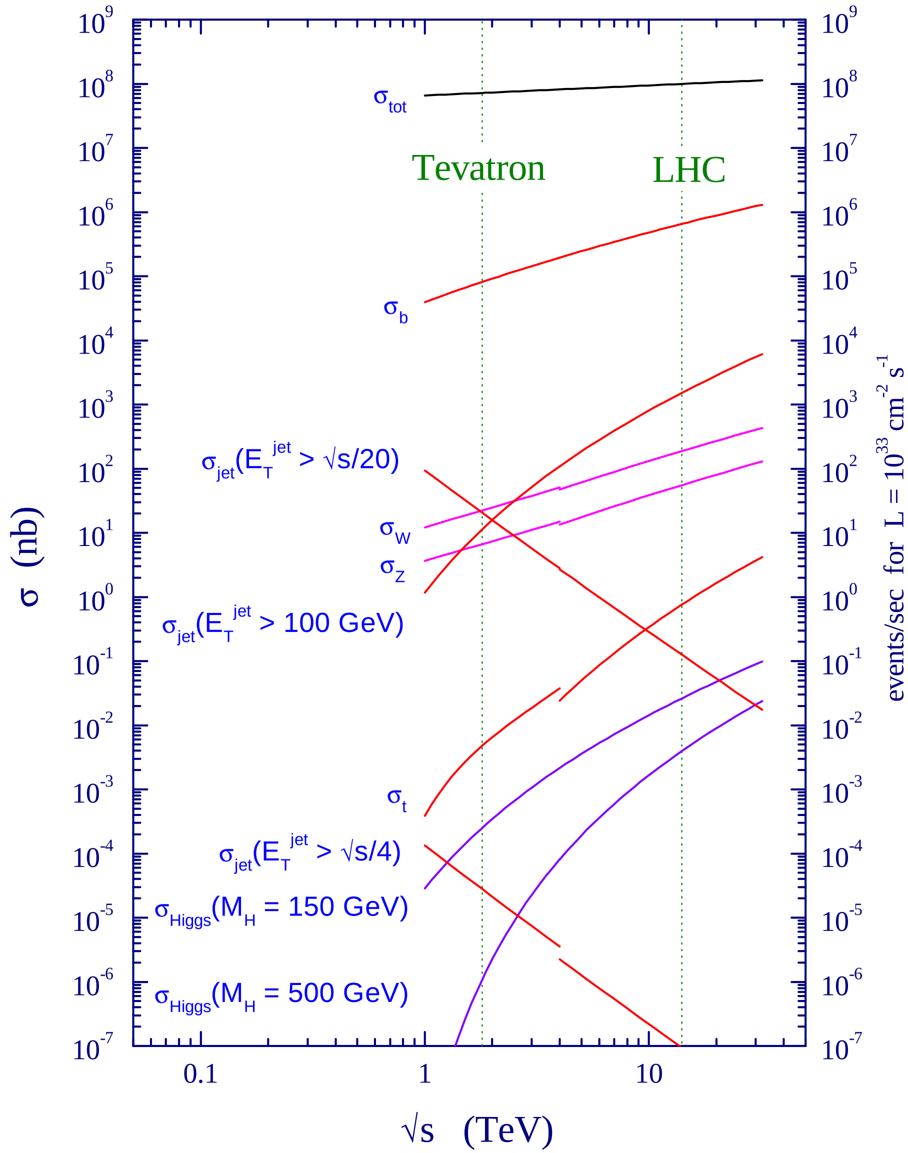
- THE MOMENTUM PROBABILITY DENSITY $xf_i(x)$ IS SHOWN AT TWO DIFFERENT SCALES (LEFT $\Rightarrow$ LOW SCALE; RIGHT $\Rightarrow$ HIGH SCALE)
- GIVEN PDFs VS x AT ONE SCALE Q_0^2 $\Rightarrow$ DETERMINED FOR ALL SCALES BY EVOLUTION EQUATIONS
- AS $x \geq 1$ KINEMATIC CONSTRAINT $f_i(x) = 0$
- "VALENCE" UP AND DOWN: PEAKED AT $x \sim 0.3$; EXPECT $f_x(x) \underset{x\to 1}{\sim} (1-x)_i^\beta$
- "SEA" ANTIQUARK AND GLUON GROW AT SMALL x
- "SINGLET" AND GLUON MIX $\Rightarrow$ ALL PDFs LOOK THE SAME AS $x \to 0$

BEFORE AND AFTER THE LHC

KINEMATIC PLANE



HADRONIC CROSS-SECTIONS



- Q^2 : INVARIANT MASS OF FINAL STATE $\Rightarrow$ WIDENING OF AVAILABLE PROCESSES
- AS ENERGY GROWS, DROP OF CROSS-SECTION MAY BE OFFSET BY GROWTH OF SMALL x PDFs

DISENTANGLING PDFs

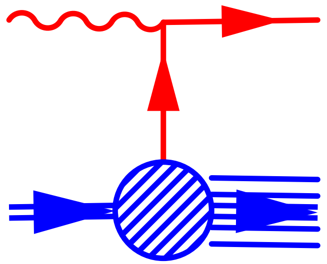
A SCIENTIFIC ART

- DEEP-INELASTIC SCATTERING DATA ON PROTON ABUNDANT AND PRECISE
- CC F_1 AND F_3 IN PRINCIPLE PROVIDE FOUR COMBINATIONS, AND NC F_1 TWO MORE
  $\Rightarrow$ ALL LIGHT FLAVORS
- HERA DATA ONLY DETERMINE FOUR COMBINATIONS OF PDFs:
  FIXED COMBINATION OF F_1 F_3, SO NC AND $\pm$CC WITH $e^{\pm}$, PLUS SEPARATE NC γ AND Z FROM SCALE DEPENDENCE
- $W^{\pm}$ AND Z PRODUCTION (INCLUDING DOUBLE DIFFERENTIAL: MASS AND RAPIDITY) PROVIDE A LARGE AMOUNT OF INFORMATION
- WHEN PRODUCING ELECTROWEAK FINAL STATES, THE GLUON CAN ONLY BE ACCESSED FROM SCALE DEPENDENCE OR HIGHER ORDERS
  …EXCEPT IN HIGGS PRODUCTION!
- JET PRODUCTION GIVES A DIRECT HANDLE ON THE GLUON

Leading Parton Content

(up to $O[\alpha_s]$ corrections)

DEEP-INELASTIC SCATTERING



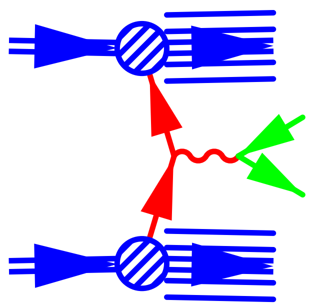
NC	$F_1^{\,\gamma} = \sum_i e_i^2\,(q_i + \bar{q}_i)$
NC	$F_1^{\,Z,\,\mathrm{int.}} = \sum_i B_i\,(q_i + \bar{q}_i)$
NC	$F_3^{\,Z,\,\mathrm{int.}} = \sum_i D_i\,(q_i + \bar{q}_i)$
CC	$F_1^{W^+} = \bar{u} + d + s + \bar{c}$
CC	$-F_3^{W^+}/2 = \bar{u} - d - s + \bar{c}$

ℓ	e	V	A
u,c,t	+2/3	$(+1/2 - 4/3\sin^2\theta_W)$	+1/2
d,s,b	-1/3	$(-1/2 + 2/3\sin^2\theta_W)$	-1/2
ν	0	+1/2	+1/2
e,μ,τ	-1	$(-1/2 + 2\sin^2\theta_W)$	-1/2

$B_q(Q^2) = -2e_q V_\ell V_q P_Z + (V_\ell^2 + A_\ell^2)(V_q^2 + A_q^2)P_Z^2$; $D_q(Q^2) = -2e_q A_\ell A_q P_Z + 4V_\ell A_\ell V_q A_q P_Z^2$; $P_Z = Q^2/(Q^2 + M_Z^2)$

$$W^+ \to W^- \;\Rightarrow\; u \leftrightarrow d,\; c \leftrightarrow s;\; p \to n \;\Rightarrow\; u \leftrightarrow d$$

Drell-Yan



$$L^{ij}\,(x_1, x_2) \equiv q_i(x_1, M^2)\bar{q}_j(x_2, M^2)$$

γ $\quad \frac{d\sigma}{dM^2 dy}(M^2, y) = \frac{4\pi\alpha^2}{9M^2 s}\sum_i e_i^2 L^{ii}(x_1, x_2)$

W $\quad \frac{d\sigma}{dy} = \frac{\pi G_F M_V^2\sqrt{2}}{3s}\sum_{i,j}|V_{ij}^{\mathrm{CKM}}|L^{ij}(x_1, x_2)$

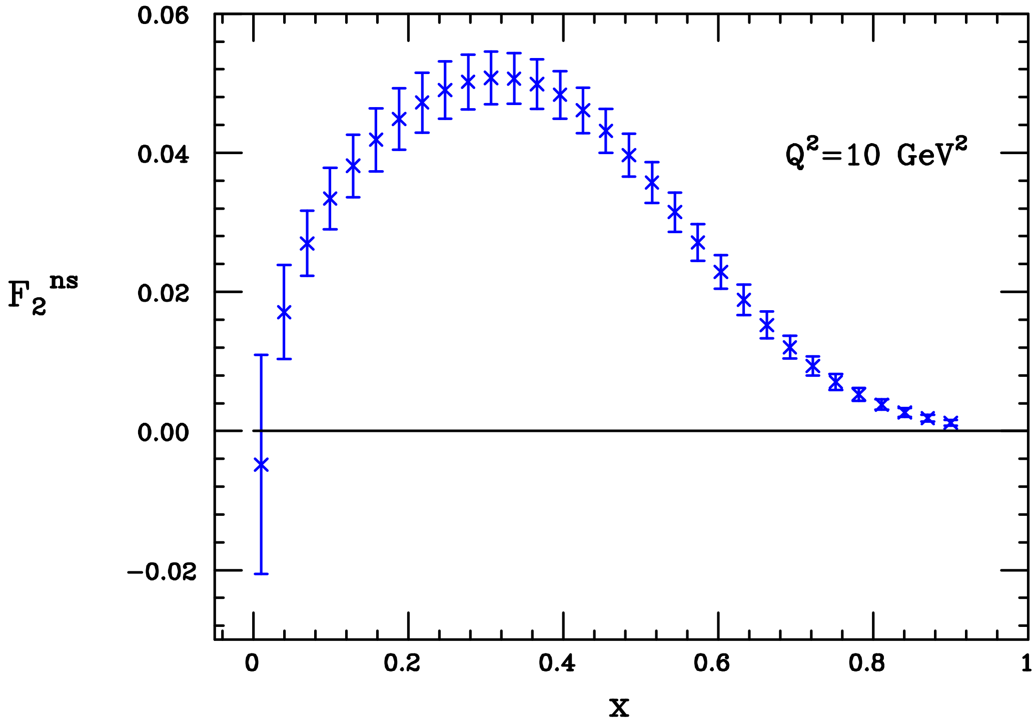
Z $\quad \frac{d\sigma}{dy} = \frac{\pi G_F M_V^2\sqrt{2}}{3s}\sum_i\left(V_i^2 + A_i^2\right)L^{ij}(x_1, x_2)$

$V_{ij}^{\mathrm{CKM}} \to$ CKM MATRIX ($i = u,\ c\,t,\ j = d,\ s\,b$), $V_{ij}^{\mathrm{CKM}} = 1 + O(\lambda)$; $\lambda = \sin\theta_C \approx 0.22$

EXPLOITING ISOSPIN:
THE ISOTRIPLET STRUCTURE FUNCTION

$$u^p(x,Q^2) = d^n(x,Q^2); \quad d^p(x,Q^2) = u^n(x,Q^2)$$

$$F_2^p(x,Q2) - F_2^d(x,Q^2) = \frac{1}{3}\left[(u^p+\bar{u}^p) - \left(d^p+\bar{d}^p\right)\right]\left[1+O(\alpha_s)\right]$$



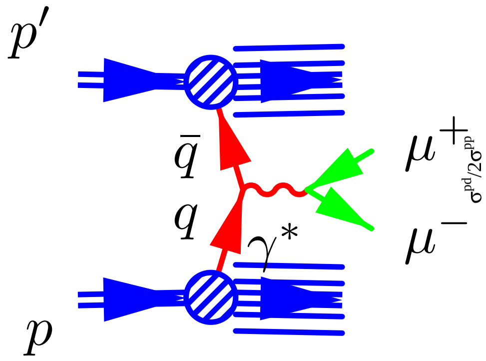
(NNPDF, 2005)

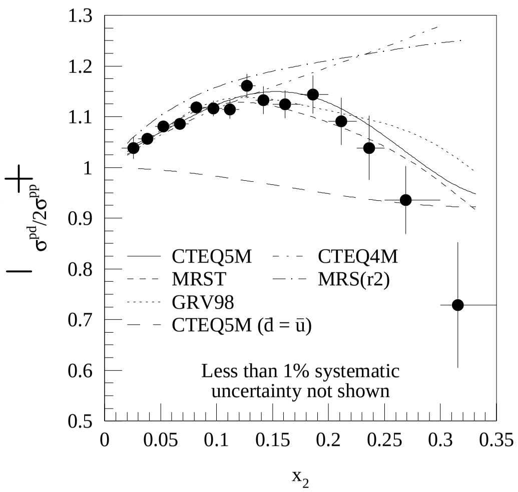
…BUT IN ORDER TO ACCESS THE NEUTRON ONE MUST ASSUME $F_2^d = \frac{1}{2}\left(F_2^p + F_2^n\right)$

EXPLOITING CHARGE CONJUGATION AND ISOSPIN:
QUARKS AND ANTIQUARKS AT A $p\bar{p}$ COLLIDER (TEVATRON)

BY CHARGE CONJUGATION $\bar{q}_{\bar{P}} = q_p$

DRELL-YAN p/d ASYMMETRY (ISOSPIN)

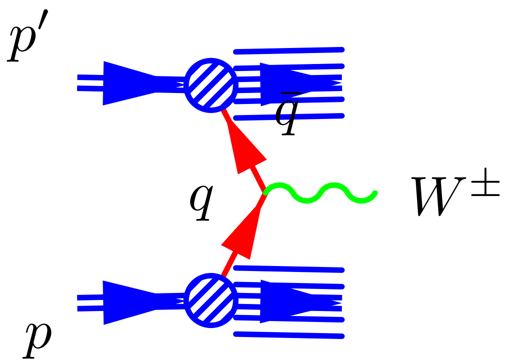


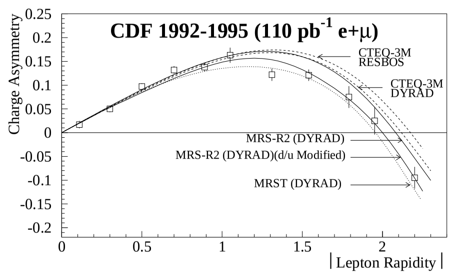


$$\frac{\sigma^{pn}}{\sigma^{pp}} \sim \left.\frac{\frac{4}{9} u^p \bar{d}^p + \frac{1}{9} d^p \bar{u}^p}{\frac{4}{9} u^p \bar{u}^p + \frac{1}{9} d^p \bar{d}^p}\right|_{\text{large } x} \approx \frac{\bar{d}}{\bar{u}}$$

E866 (2001)

$W^\pm$ ASYMMETRY (C-CONJUGATION)





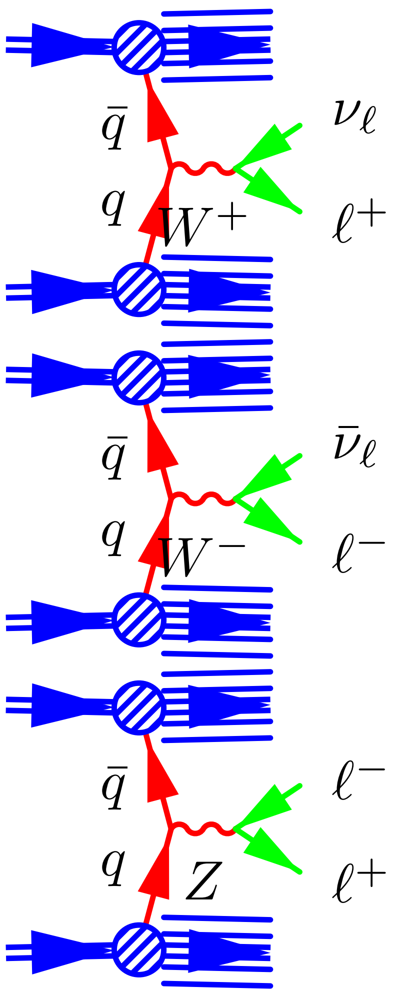
$$\frac{\sigma^{p\bar{p}}_{W+}}{\sigma^{p\bar{p}}_{W-}} = \frac{u^p(x_1)d^p(x_2) + \bar{d}^p(x_1)\bar{u}^p(x_2)}{d^p(x_1)u^p(x_2) + \bar{u}^p(x_1)\bar{d}^p(x_2)} \sim \frac{u^p d^p}{d^p u^p}$$

if x_1, x_2 in valence region,
neglecting HQ & Cabibbo suppr.

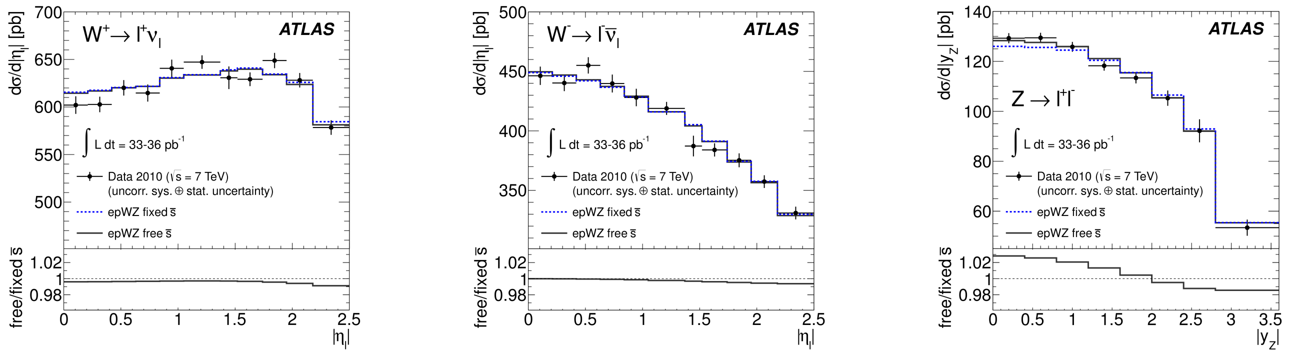
CDF (1998)

EXPLOITING KINEMATIC COVERAGE AND FINAL STATES: LIGHT FLAVORS AND STRANGENESS AT THE LHC

$W^{\pm}$ AND Z PRODUCTION



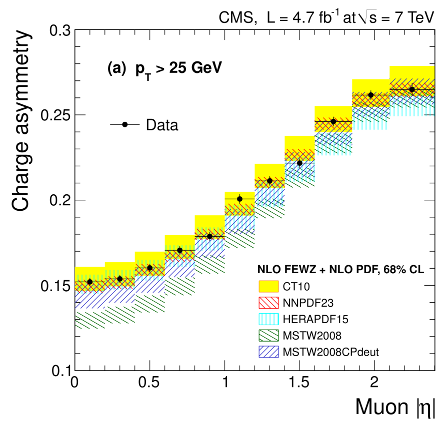
W AND Z CROSS SECTIONS



ATLAS (2012)

$\sigma^{p\bar{p}}_{W^+} = u\bar{d} + c\bar{s}$; $\sigma^{p\bar{p}}_{Z} = u\bar{u} + d\bar{d} + s\bar{s}$: STRANGENESS DETERMINED BY COMPARISON

W MUON ASYMMETRY



CMS (2013)

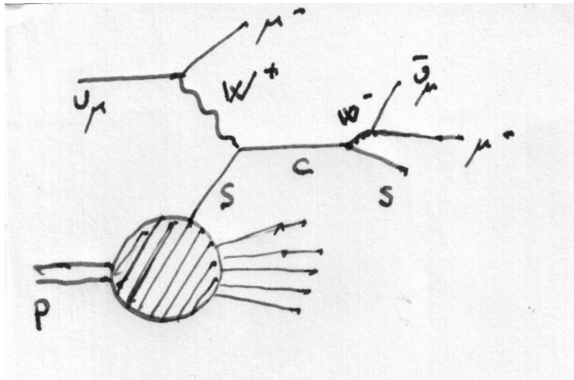
$$\frac{\sigma^{p\bar{p}}_{W^+}}{\sigma^{p\bar{p}}_{W^-}} = \frac{u(x_1)\bar{d}(x_2)+\bar{d}(x_1)u(x_2)}{d(x_1)\bar{u}(x_2)+\bar{u}(x_1)d(x_2)}$$

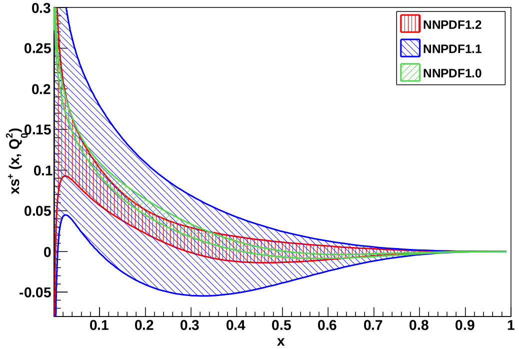
"VALENCE" $x \Rightarrow$ NEGLECT STRANGENESS

$\Rightarrow$ DETERMINE $\bar{u} - \bar{d}$

TAGGING FINAL STATES
STRANGENESS IN DIS AND AT THE LHC

DIMUONS
TOTAL STRANGENESS



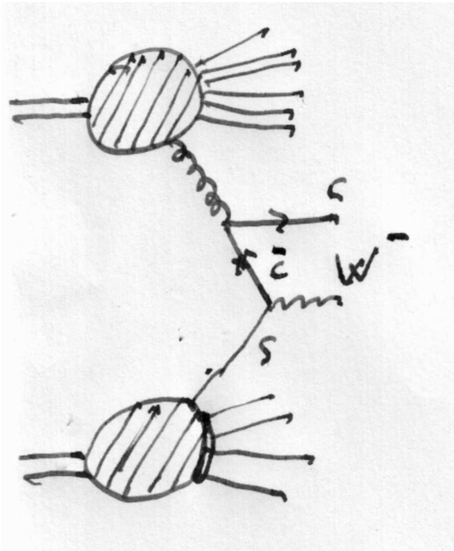


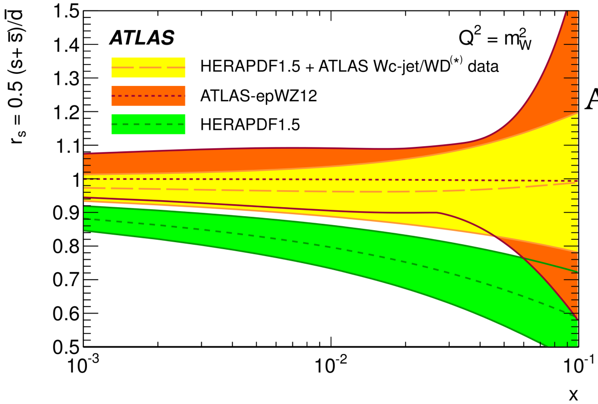
AT LO $F_{2,\,c}^{\nu} = 2xc$;
$F_{2,\,c}^{\bar{\nu}} = 2x\bar{c}$
UP TO CABIBBO SUPPR.
ASSUMPTION VS. INCLUSIVE DIS
VS. INCLUSIVE+ DIMUONS

NNPDF (2009)

W+ CHARM

STRANGE TO LIGH FRACTION





AT LO σ PROPORTIONAL TO THE STRANGE-GLUON LUMI, UP TO CABIBBO SUPPR.
ASSUMPTION VS. DIS+ TOTAL WZ VS. DIS+ TOTAL WZ + W+CHARM

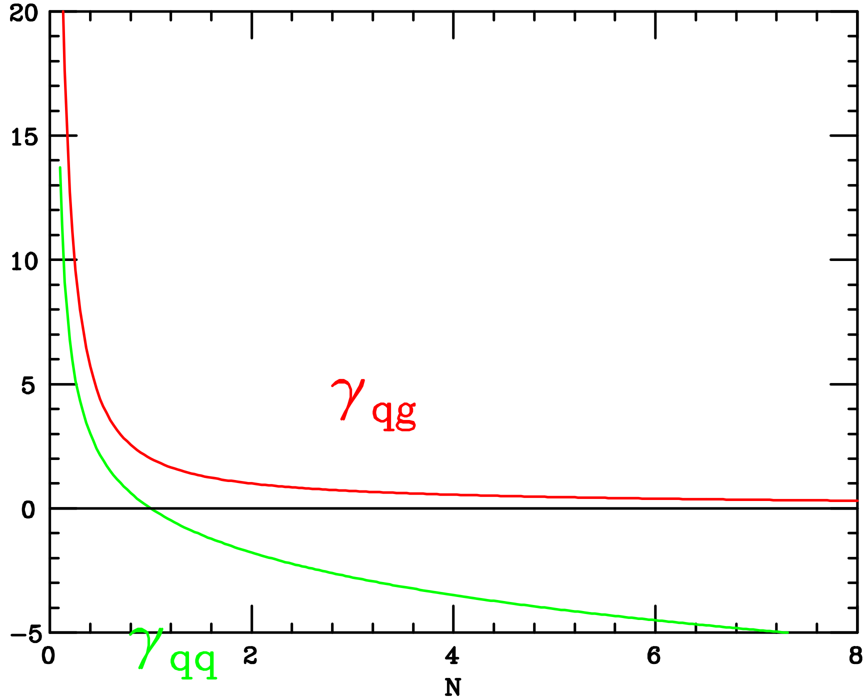
ATLAS (2013)

THE GLUON

SCALE DEPENDENCE OF FLAVOR SINGLET STRUCTURE FUNCTIONS

$$\frac{d}{dt}F_2^s(N,Q^2) = \frac{\alpha_s(Q^2)}{2\pi}\left[\gamma_{qq}(N)F_2^s + 2\,n_f\,\gamma_{qg}(N)g(N,Q^2)\right] + O(\alpha_s^2)$$

ANOMALOUS DIMENSIONS

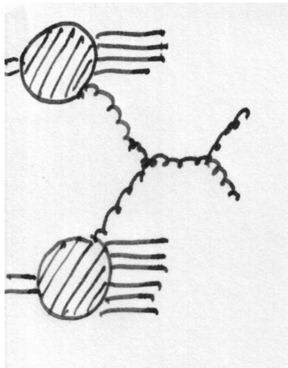


LARGE x GLUON DIFFICULT TO DETERMINE FROM DEEP-INELASTIC SCATTERING

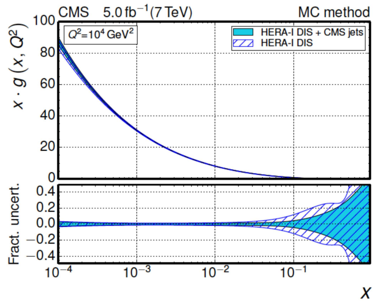
THE GLUON IN HADRONIC COLLISIONS

THE GLUON ONLY INTERACTS THROUGH QCD

JETS



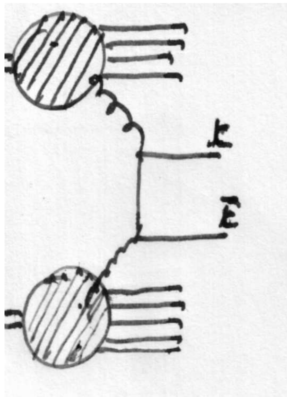
GLUON



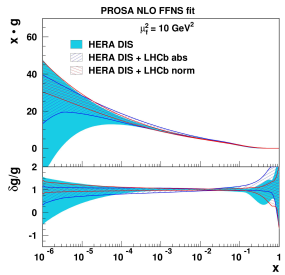
CMS (2014)

- ONE-JET INCLUSIVE USED TO CONSTRAIN THE LARGE x GLUON SINCE TEVATRON
- WIDE KINEMATIC REGION AT LHC

TOP



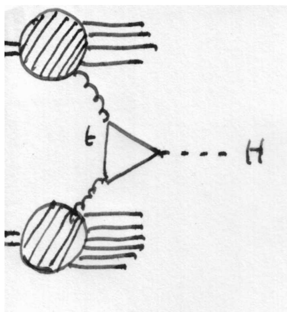
GLUON



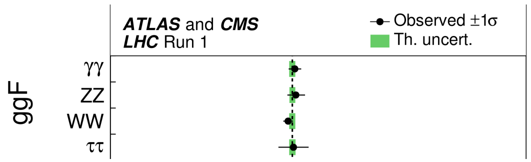
prosa (LHCb data) 2015)

- WIDE RAPIDITY RANGE: CAN ACCESS WIDE x REGION

HIGGS



TOTAL CROSS-SECTION



ATLAS+CMS (2016)

- NOT YET A STANDARD CANDLE
- EXPERIMENTAL ACCURACY ALREADY COMPETITIVE

PDF DETERMINATION SUMMARY

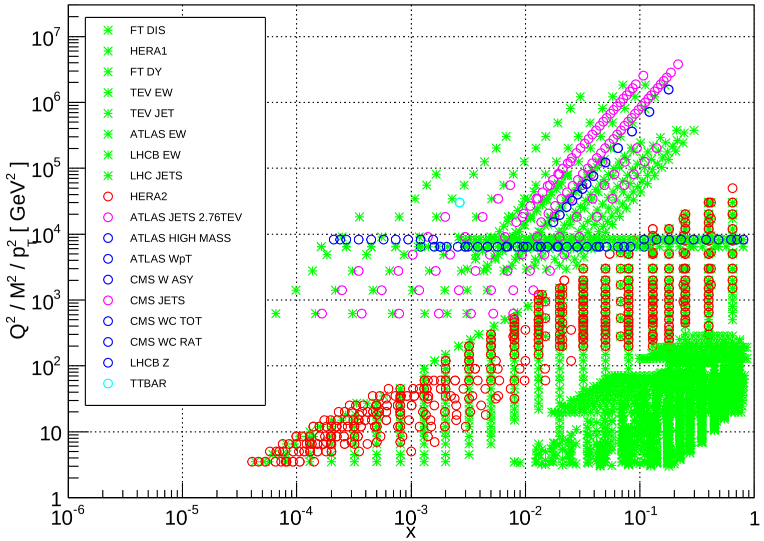
- DEEP-INELASTIC SCATTERING PROVIDES THE BULK OF INFORMATION ON PDFs:
  - HERA COLLIDER $e^{\pm}p$ CC+NC DATA PROVIDE FOUR INDEPENDENT COMBINATIONS IN WIDE KINEMATIC REGION $\Rightarrow$ LIGHT QUARKS AND ANTIQUARKS
  - FIXED-TARGET μp & μd GIVES DIRECT HANDLE ON UP-DOWN SEPARATION, ESPECIALLY AT LARGER x
  - HERA+FT GLUON FROM SCALE DEPENDENCE ("SCALING VIOLATIONS")
  - NEUTRINO (ESPECIALLY DIMUON) $\Rightarrow$ STRANGENESS
- DRELL-YAN γ^* ON FIXED p AND d TARGET $\Rightarrow$ UP-DOWN SEPARATION AT LARGE x
- W AND Z PRODUCTION AT THE TEVATRON $\Rightarrow$ ANTIUP/ANTIDOWN
- LHC W, Z HIGH AND LOW MASS
  - FULL FLAVOR SEPARATION IN WIDE KINEMATIC REGION
  - STRANGENESS BOTH FROM TOTAL CROSS-SECTION AND TAGGED $W + c$ FINAL STATE
  - GLUON FROM Z TRANSVERSE MOMENTUM DISTRIBUTION
- GLUON FROM COLLIDER PROCESSES:
  - LARGE x FROM TEVATRON JETS
  - SMALL x FROM LHC JETS
  - MEDIUM x FROM LHC TOP

METHODOLOGY & STATISTICS

PARTON FITS

DATA $\rightarrow$ PARTON DISTRIBUTIONS

NNPDF3.0 NLO dataset



ISSUES AND TASKS:

- FROM PHYSICAL OBSERVABLES TO PDFS: SOLVE EVOLUTION EQUATIONS, CONVOLUTE WITH PARTON-LEVEL CROSS-SECTIONS
- DISENTANGLING PDFS: CHOOSE A BASIS OF PDFS ($2N_f$ QUARKS + 1 GLUON) & A SET OF SUITABLE PHYSICAL PROCESSES TO DETERMINE THEM ALL
- PROBABILITY IN THE SPACE OF FUNCTIONS: CHOOSE A STATISTICAL APPROACH (HESSIAN, MONTE CARLO, ...)
- UNCERTAINTY ON FUNCTIONS: CHOOSE A FUNCTIONAL FORM

THE HESSIAN APPROACH

- CHOOSE A FIXED FUNCTIONAL FORM
  - SINCE 1973, PHYSICALLY MOTIVATED ANSATZ $f_i(x, Q_0^2) = x^\alpha (1-x)^\beta g_i(x)$; $g_i(x)$ POLYNOMIAL IN x OR $\sqrt{x}$
  - MMHT 2015:
    * BASIS FUNCTIONS g; $u_v = u - \bar{u}$; $d_v = d - \bar{d}$; $S = 2(\bar{u} + \bar{d}) + s + \bar{s}$; $s_+ = s + \bar{s}$; $\Delta = \bar{d} - \bar{u}$; $s_- = s - \bar{s}$.
    * FOR ALL BUT Δ s_-, $g \Rightarrow x f_i(x, Q_0^2) = A x^\alpha (1-x)^\beta \left(1 + \sum_{i=1}^4 a_i T_i(y(x))\right)$; T_i CHEBYSHEV POLYNOMIALS, $y = 1 - 2\sqrt{x} \leftrightarrow$ MUST MAP $x = [0, 1]$ INTO $y = [-1, 1]$; $T_i(-1) = T_i(1) = 1$
    * GLUON $x g(x, Q_0^2) = A x^\alpha (1-x)^\beta \left(1 + \sum_{i=1}^2 a_i T_i(y(x))\right) + A' x T \alpha' (1-x)^{\beta'}$
    * SEA ASYMMETRY $x\Delta(x, Q_0^2) = A x^\alpha (1-x)^\beta (1 + \gamma x + \epsilon x^2)$
    * STRANGENESS ASYMMETRY $x\Delta(x, Q_0^2) = A x^\alpha (1-x)^\beta (1 - x/x_0)$
    * 41 PARAMETERS, 4 FIXED BY SUM RULES
    * 12 PARMS FIXED AT BEST FIT, REMAINING 25 USED FOR (HESSIAN) COVARIANCE MATRIX
- EVOLVE TO DESIRED SCALE & COMPUTE PHYSICAL OBSERVABLES
- DETERMINE BEST-FIT VALUES OF PARAMETERS
- DETERMINE ERROR BY PROPAGATION OF ERROR ON PARMS. ('HESSIAN METHOD'); PARM. SCANS ALSO POSSIBLE ('LAGR. MULTIPLIER METHOD')

HESSIAN ERROR ESTIMATES

GENERAL FEATURES

OBSERVABLE X DEPENDING ON PARAMETERS $\vec{z}$: (LINEAR ERROR PROPAGATION)

$X(\vec{z}) \approx X_0 + z_i \partial_i X(\vec{z})$ ASSUMING MOST LIKELY VALUE AT $\vec{z} = 0$

VARIANCE: $\sigma_X^2 = \sigma_{ij} \partial_i X \partial_j X,$

$\sigma_{ij} \Rightarrow$ COVARIANCE MATRIX IN PARAMETER SPACE

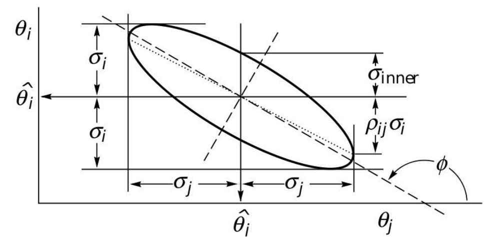
MAXIMUM LIKELIHOOD: COVARIANCE $\Leftrightarrow$ HESSIAN $\sigma_{ij} = \partial_i \partial_j \chi^2$ EVALUATED AT MIN. OF χ^2

DIAGONALIZATION: CHOOSE z_i AS EIGENVECTORS OF σ_{ij} WITH UNIT EIGENVALUES

$\sigma_X^2 = |\vec{\nabla} X|^2$ (LENGTH OF GRADIENT)

SOME INTERESTING CONSEQUENCES

- THE ONE-σ CONTOUR IN PARAMETER SPACE IS ELLIPSE $\chi^2 = \chi^2_{\min} + 1$
- THE TOTAL UNCERTAINTY IS THE SUM IN QUADRATURE OF UNCERTAINTIES DUE TO EACH PARAMETER (LENGTH OF VECTOR) EVEN WHEN NOT DIAGONALIZING (Lai et al, CTEQ 2010)



- ANY ROTATION (ORTHOGONAL TRANSF.) IN THE SPACE OF PARMS PRESERVES THE GRADIENT $\rightarrow$ CAN DIAGONALIZE A CHOSEN OBSERVABLE WITHOUT SPOILING RESULT (Pumplin 2009)

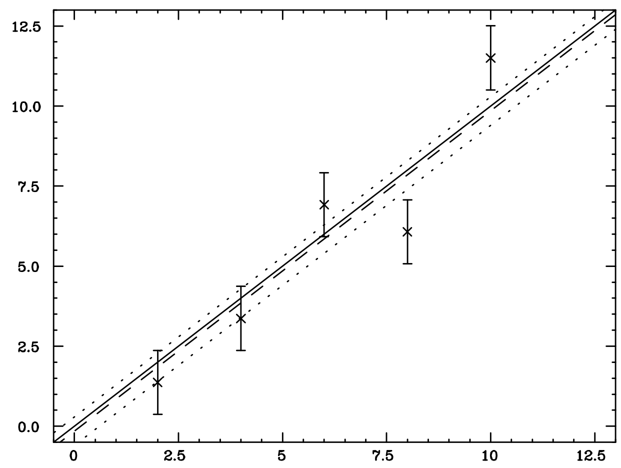
THE HESSIAN UNCERTAINTY

"PARADOX"

- THE STANDARD DEVIATION OF χ^2 FOR $N_{\rm dat}$ DATA $\sigma_{\chi^2} = \sqrt{2N_{\rm dat}}$
  HYPOTESIS-TESTING RANGE: COMPARE $\Delta\chi^2 = \chi^2 - \langle\chi^2\rangle$ TO $\sigma^2_{\chi^2}$.
  IF TOO LARGE, SOMETHING WRONG WITH THEORY (OR DATA)
- BUT THE ONE-σ RANGE FOR A PARM. OF THE THEORY IS THE CURVE $\chi^2 - \chi^2_{\rm min} = 1$
  PARAMETER-FITTING RANGE: UNIT DEVIATION FROM THE PARAMETRIC MINIMUM $\chi^2_{\rm min}$

WHY?

- CONSIDER DEVIATIONS Δ_i FROM LINEAR FIT $y = x + k$; DETERMINE INTERCEPT k AS FREE PARAMETER
- IF STANDARD DEVIATION FOR EACH Δ_i IS σ_Δ, THEN AVERAGE SQUARE DEVIATION IN UNITS OF σ_Δ FOR $N_{\rm dat}$ DATA: $\sigma_{\chi^2} = N_{\rm dat}$
- BEST-FIT INCTERCEPT: $k = \langle\Delta_i\rangle$
- UNCERTAINTY ON IT: $\sigma_k = \frac{\sigma_\Delta}{N_{\rm dat}}$
- IF $\Delta k = \sigma_k$, THEN $\Delta\chi^2 = 1$

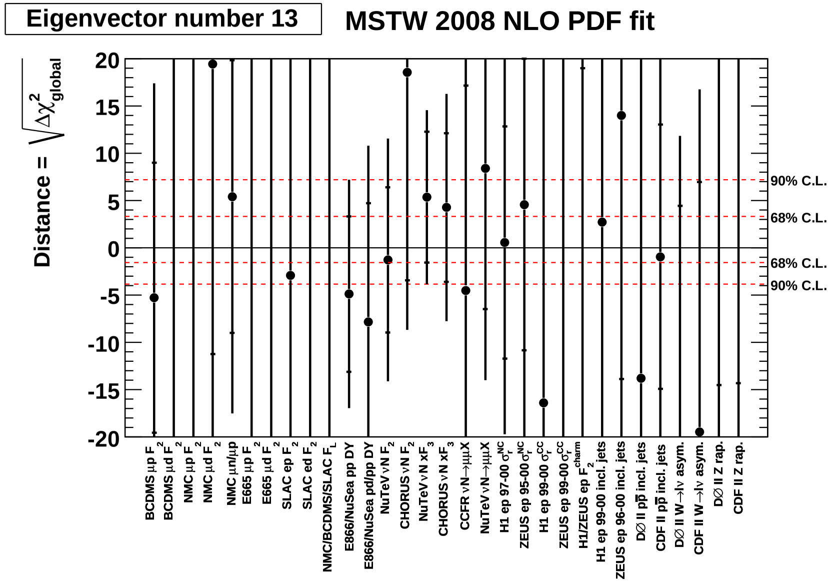


TOLERANCE

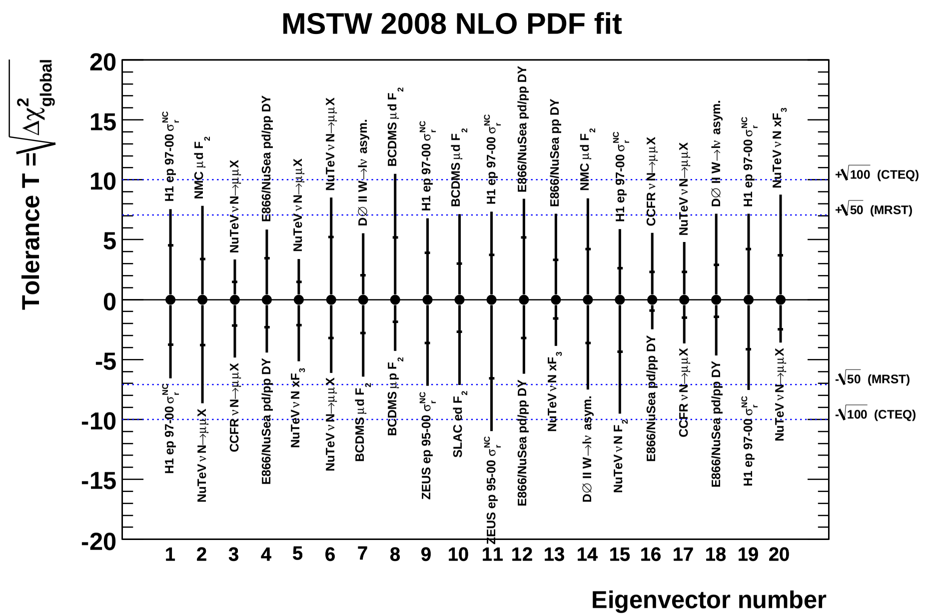
- IN GLOBAL HESSIAN FITS, UNCERTAINTITES OBTAINED BY $\Delta\chi^2 = 1$ UNREALISTICALLY SMALL

- UNCERTAINTIES TUNED TO DISTRIBUTION OF DEVIATIONS FROM BEST-FITS FOR INDIVIDUAL EXPERIMENTS

MSTW TOLERANCE PLOT FOR 13TH EIGENVEC.



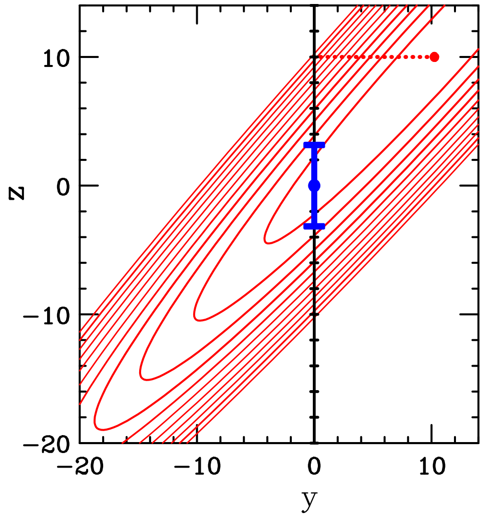
GLOBAL MSTW TOLERANCE



- (MSTW/MMHT) FOR EACH EIGENVECTOR IN PARAMETER SPACE DETERMINE CONFIDENCE LIMIT FOR THE DISTRIBUTION OF BEST-FITS OF EACH EXPERIMENT

- RESCALE $\Delta\chi^2 = T$ INTERVAL SUCH THAT CORRECT CONFIDENCE INTERVALS ARE REPRODUCED

PARAMETRIZATION BIAS?

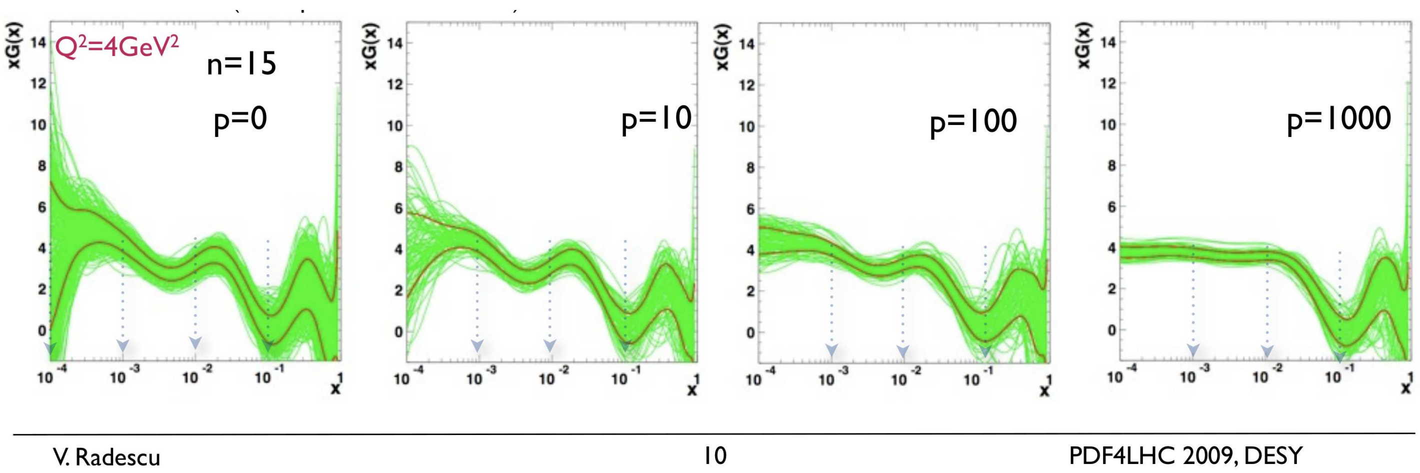


(Pumplin, 2009)

- IF PARM. NOT GENERAL ENOUGH, GLOBAL MIN. IS NOT TRUE MIN.
- ONE-σ VARIATION ABOUT FAKE MIN CORRESP. TO LARGE χ^2 VARIATION
- IN DATA REGION, UNCERTAINTY CAN BE TUNED TO DATA, BUT WHAT ABOUT INTERPORATION/ESTRAPOLATION?

UNBIASED BASES

- OLD IDEA (Parisi, Sourlas, 1978):EXPAND PDFs OVER BASIS OF ORTHOGONAL POLYNOMIALS, WITHOUT ANY FURTHER ASSUMPTION
- DIFFICULT TO AVOID SPURIOUS FLUCTUATIONS
- MUST IMPOSE LENGTH PENALTY TO STABILIZE THE FIT (ANOTHER BIAS?)



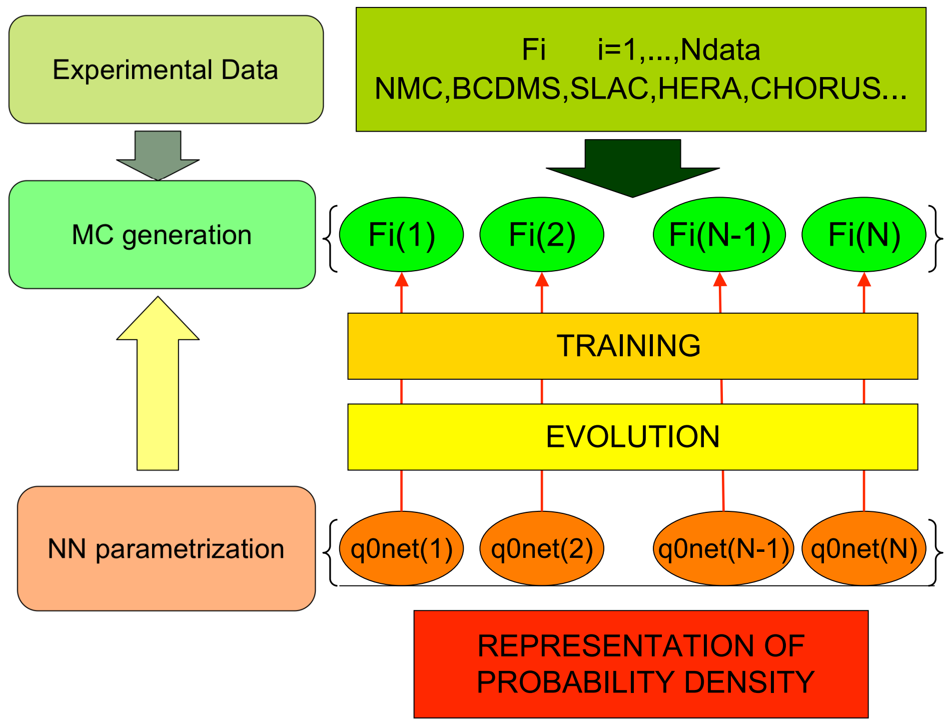
(Glazov, Radescu, 2009)

THE MONTE CARLO METHOD

BASIC IDEA: MONTE CARLO SAMPLING OF THE PROBABILITY MEASURE IN THE (FUNCTION) SPACE OF PDFs

- GENERATE A SET OF MONTE CARLO REPLICAS $\sigma^{(k)}$ OF THE ORIGINAL DATASET $\sigma^{(\mathrm{data})}$
  $\Rightarrow$ REPRESENTATION OF $\mathcal{P}[\sigma]$ AT DISCRETE SET OF POINTS IN DATA SPACE
- FIT A PDF REPLICA TO A DATA REPLICA
  $\Rightarrow$ EACH PDF REPLICA $f_i^{(k)}$ IS A BEST-FIT PDF SET FOR GIVEN DATA REPLICA
- THE SET OF NEURAL NETS IS A REPRESENTATION OF THE PROBABILITY DENSITY:

$$\langle f_i \rangle = \frac{1}{N_{rep}} \sum_{k=1}^{N_{rep}} f_i^{(k)}$$



MONTE CARLO ERROR ESTIMATES

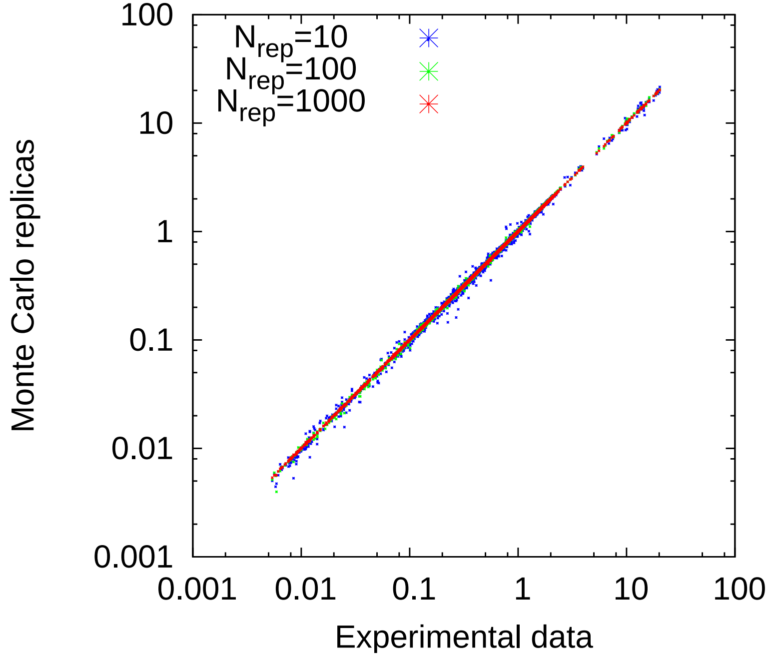
EXACT ERROR PROPAGATION

OBSERVABLE X DEPENDS ON PARAMETERS $\vec{z}$
VARIANCE: $\sigma_X^2 = \langle X^2 \rangle - \langle X \rangle^2$
AVERAGES: $\langle X \rangle = \int d^d z X(\vec{z}) P(\vec{z})$, WITH
$P(\vec{z}) \Rightarrow$ PROBABILITY DISTRIBUTION OF PARAMETER VALUES

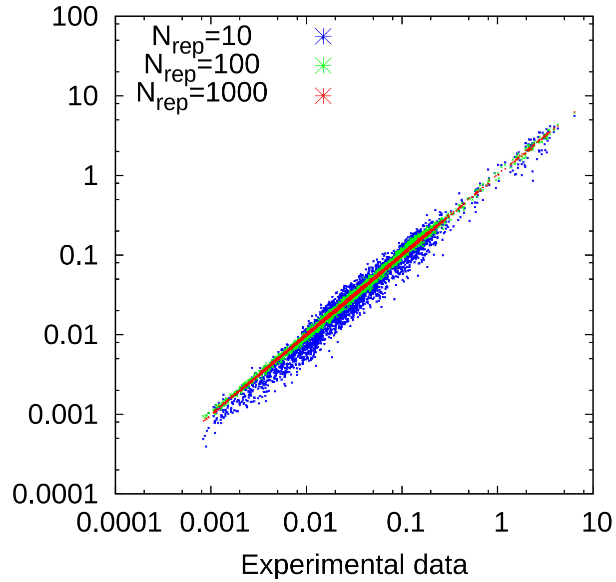
IMPORTANCE SAMPLING

- SPACE OF FUNCTIONS HUGE
  5 BINS FOR 10 PTS$\times$ 7 FCTNS $\rightarrow 5^{70} \sim 10^{49}$ BINS
- BUT EACH OBSERVABLE DEPENDS ONLY ON ONE PARAMETER, & OBSERVABLES CORRELATED
  $\Rightarrow$ DATA TELL US WHICH BINS ARE POPULATED

replica averages
vs. central values



replica standard dev.
vs. uncertainties

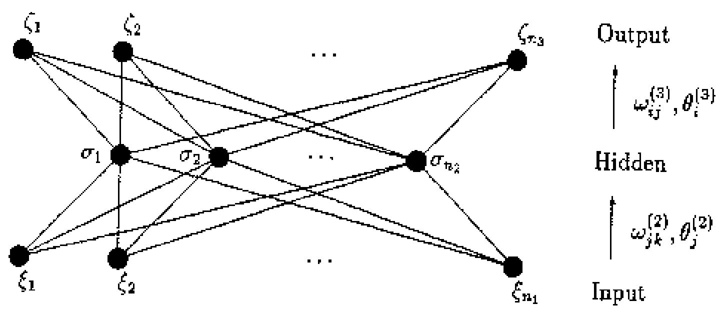


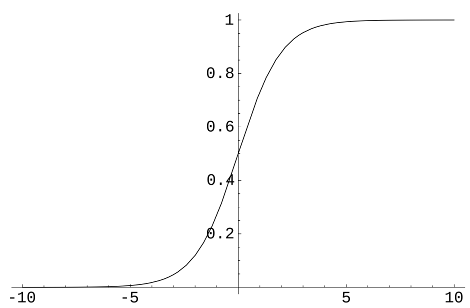
10 REPLICAS ENOUGH FOR CENTRAL VALS, 100 FOR UNCERTAINTIES, 1000 FOR CORRELNS

FLEXIBLE PARAMETRIZATION

- EACH PDF REPLICA FITTED TO A DATA REPLICA
  $\Rightarrow$ NEED BEST-FIT, BUT NOT COVARIANCE MATRIX IN PARAMETER SPACE
- CAN USE VERY LARGE PARAMETRIZATION

NEURAL NETWORKS





MULTILAYER FEED-FORWARD NETWORKS

- Each neuron receives input from neurons in preceding layer and feeds output to neurons in subsequent layer
- Activation determined by weights and thresholds
  $\xi_i = g\left(\sum_j \omega_{ij}\xi_j - \theta_i\right)$
- Sigmoid activation function
  $g(x) = \frac{1}{1+e^{-\beta x}}$

EXAMPLE: A 1-2-1 NN

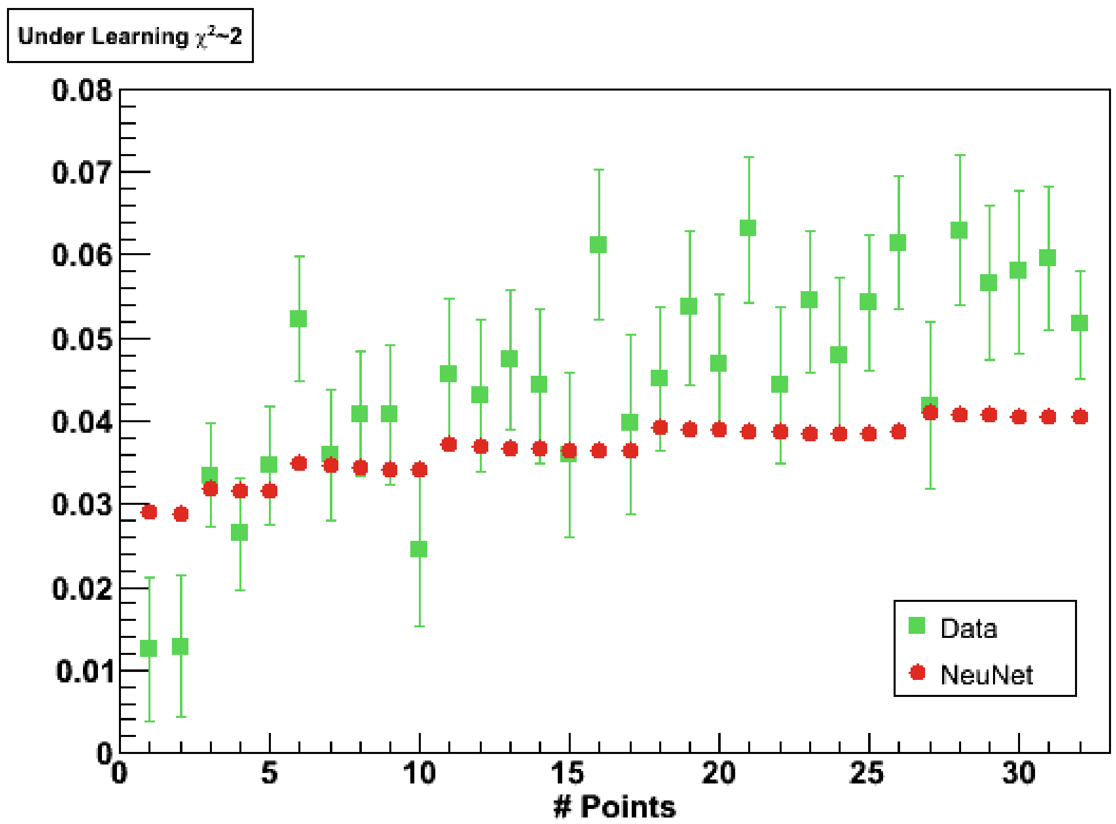
$$f(x) = \frac{1}{1+e^{\theta_1^{(3)} - \frac{\omega_{11}^{(2)}}{1+e^{\theta_1^{(2)} - x\omega_{11}^{(1)}}} - \frac{\omega_{12}^{(2)}}{1+e^{\theta_2^{(2)} - x\omega_{21}^{(1)}}}}}$$

THANKS TO NONLINEAR BEHAVIOUR,
ANY FUNCTION CAN BE REPRESENTED BY A SUFFICIENTLY BIG NEURAL NETWORK

LEARNING

- ONE CAN CHOOSE A HIGHLY REDUNDANT PARAMETRIZATION
  EXAMPLE: NNPDF: $2-5-3-1$ NN FOR EACH PDF: $37 \times 7 = 259$ PARAMETERS
- MINIMIZATION ("LEARNING") CAN BE PERFORMED USING GENETIC ALGORITHMS
- COMPLEXITY INCREASES AS THE FITTING PROCEEDS
- $\Rightarrow$ THE BEST FIT IS NOT THE ABSOLUTE MINIMUM:
  MUST LOOK FOR OPTIMAL LEARNING POINT

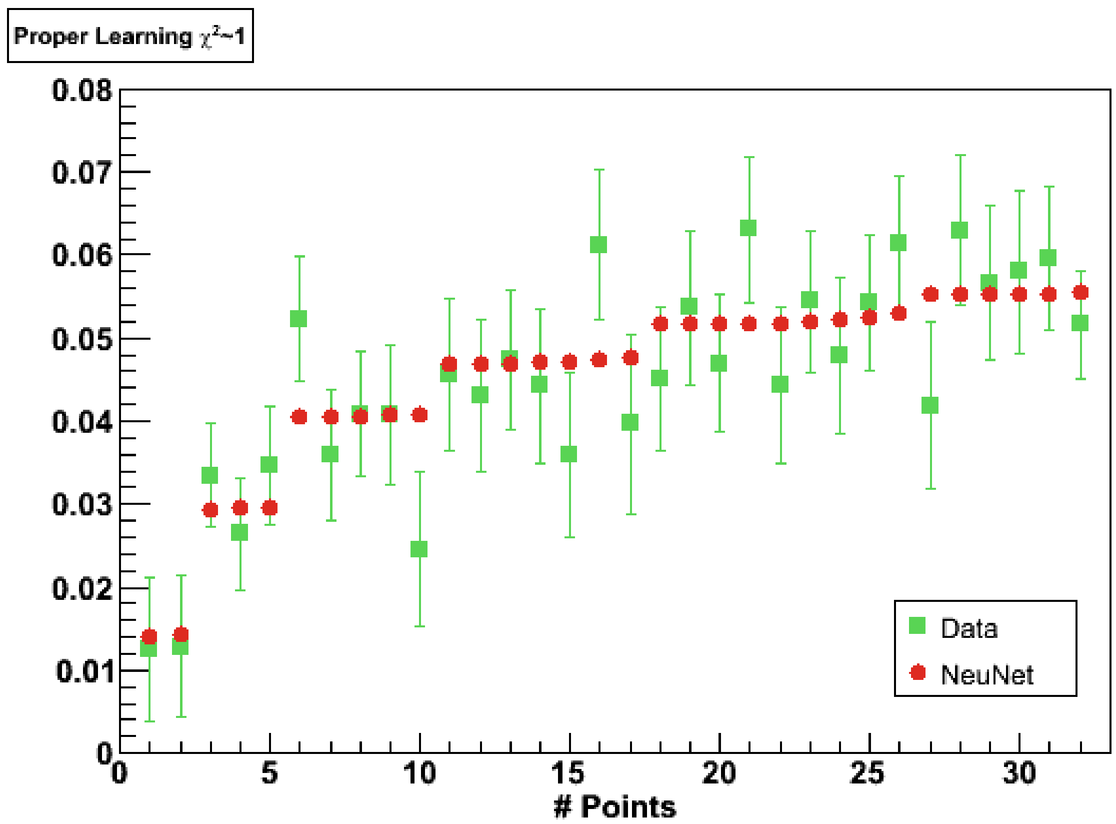
UNDERLEARNING



LEARNING

- ONE CAN CHOOSE A HIGHLY REDUNDANT PARAMETRIZATION
  EXAMPLE: NNPDF: $2-5-3-1$ NN FOR EACH PDF: $37\times 7 = 259$ PARAMETERS
- MINIMIZATION ("LEARNING") CAN BE PERFORMED USING GENETIC ALGORITHMS
- COMPLEXITY INCREASES AS THE FITTING PROCEEDS
- $\Rightarrow$ THE BEST FIT IS NOT THE ABSOLUTE MINIMUM:
  MUST LOOK FOR OPTIMAL LEARNING POINT

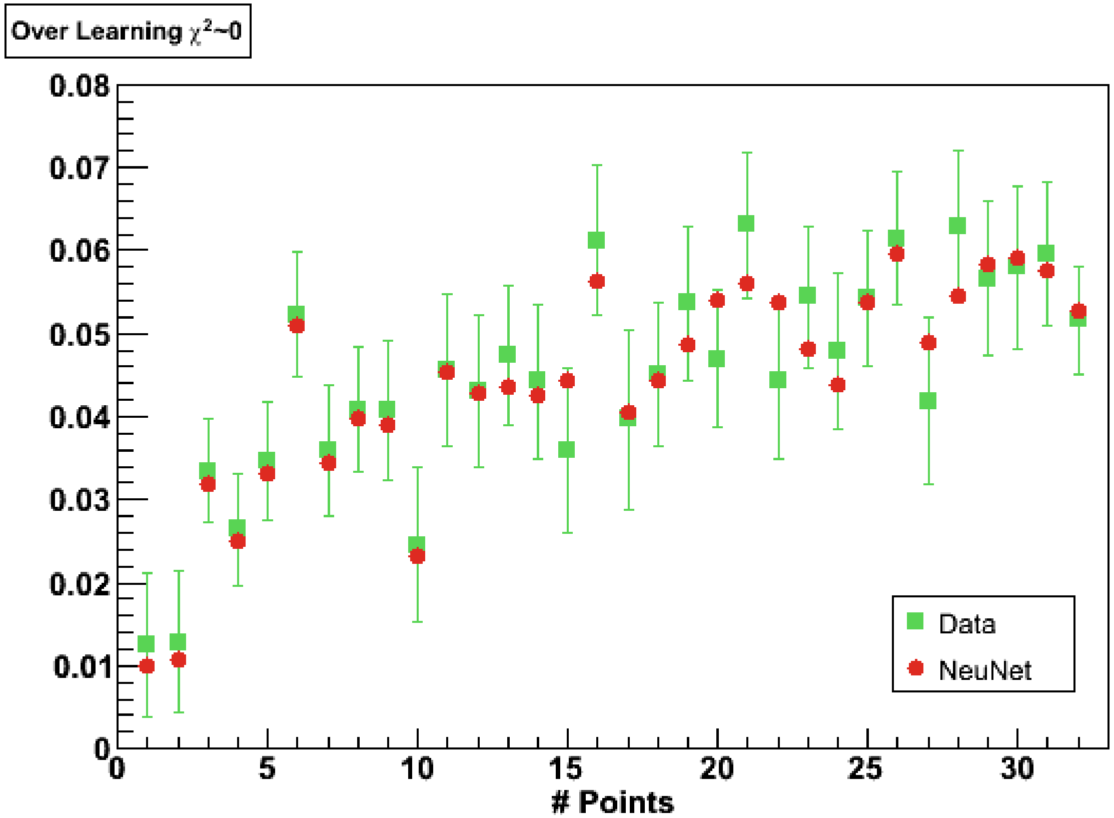
PROPER LEARNING



LEARNING

- ONE CAN CHOOSE A HIGHLY REDUNDANT PARAMETRIZATION
  EXAMPLE: NNPDF: $2-5-3-1$ NN FOR EACH PDF: $37 \times 7 = 259$ PARAMETERS
- MINIMIZATION ("LEARNING") CAN BE PERFORMED USING GENETIC ALGORITHMS
- COMPLEXITY INCREASES AS THE FITTING PROCEEDS
- $\Rightarrow$ THE BEST FIT IS NOT THE ABSOLUTE MINIMUM:
  MUST LOOK FOR OPTIMAL LEARNING POINT

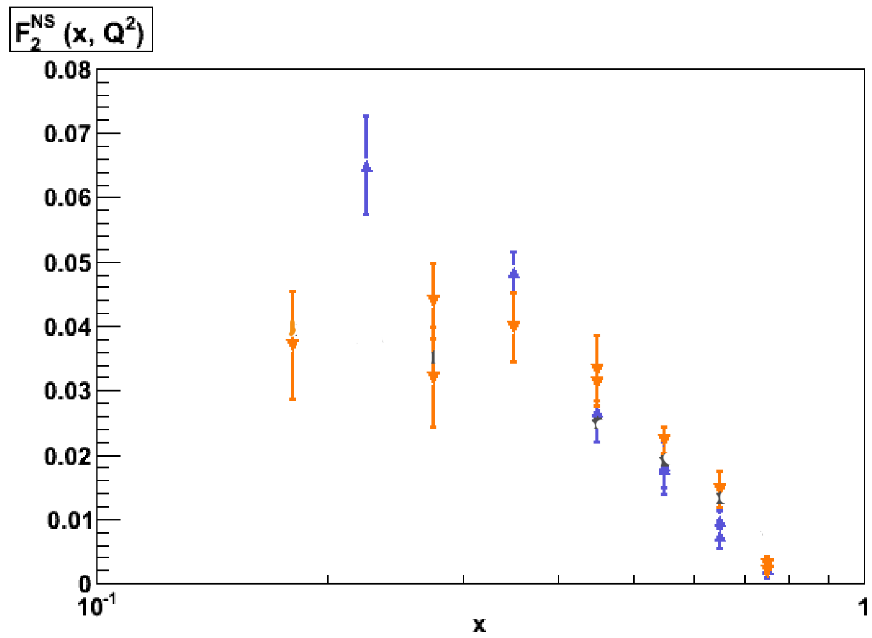
OVERLEARNING



CROSS-VALIDATION

GENETIC MINIMIZATION:
AT EACH GENERATION, χ^2 EITHER UNCHANGED OR DECREASING

- DIVIDE THE DATA IN TWO SETS: TRAINING AND VALIDATION
- MINIMIZE THE χ^2 OF THE DATA IN THE TRAINING SET
- AT EACH ITERATION, COMPUTE THE χ^2 FOR THE DATA IN THE VALIDATION SET (NOT USED FOR FITTING)
- WHEN THE VALIDATION χ^2 STOPS DECREASING, STOP THE FIT

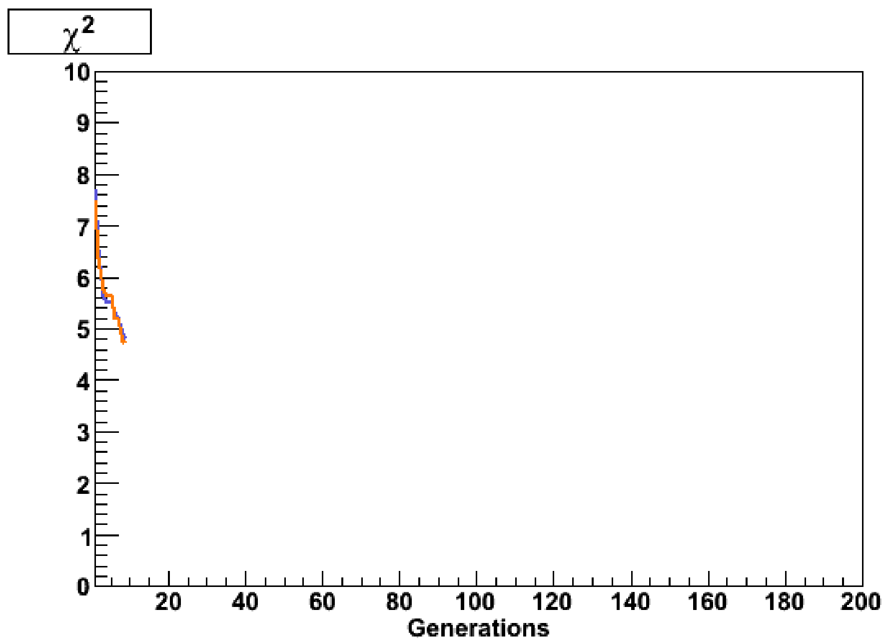


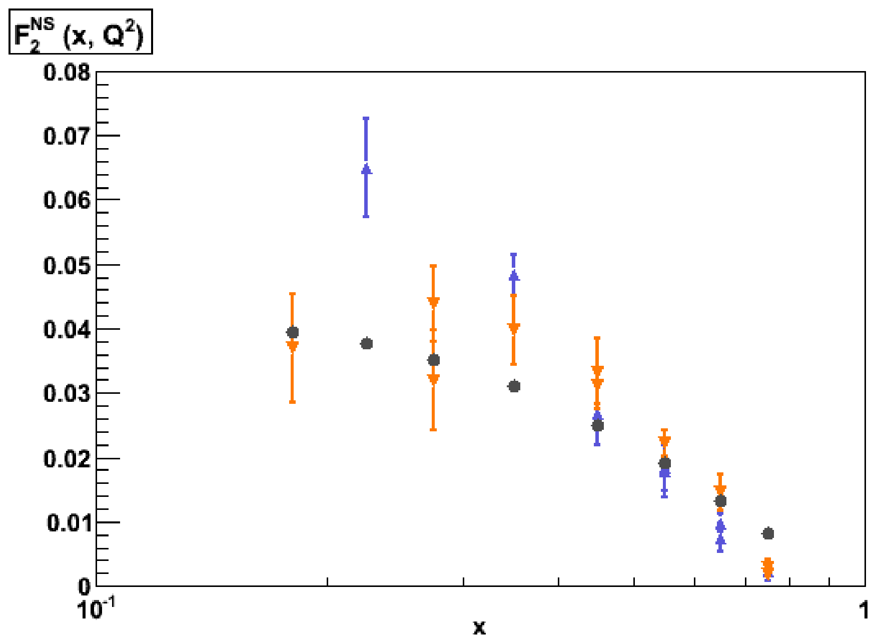
CROSS-VALIDATION

GENETIC MINIMIZATION:
AT EACH GENERATION, χ^2 EITHER UNCHANGED OR DECREASING

- DIVIDE THE DATA IN TWO SETS: TRAINING AND VALIDATION
- MINIMIZE THE χ^2 OF THE DATA IN THE TRAINING SET
- AT EACH ITERATION, COMPUTE THE χ^2 FOR THE DATA IN THE VALIDATION SET (NOT USED FOR FITTING)
- WHEN THE VALIDATION χ^2 STOPS DECREASING, STOP THE FIT

GO!



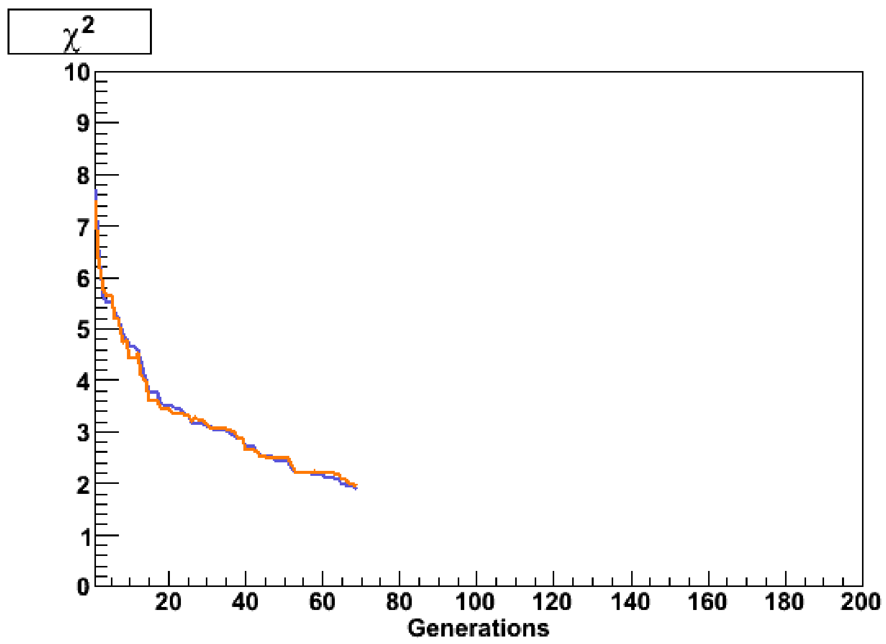


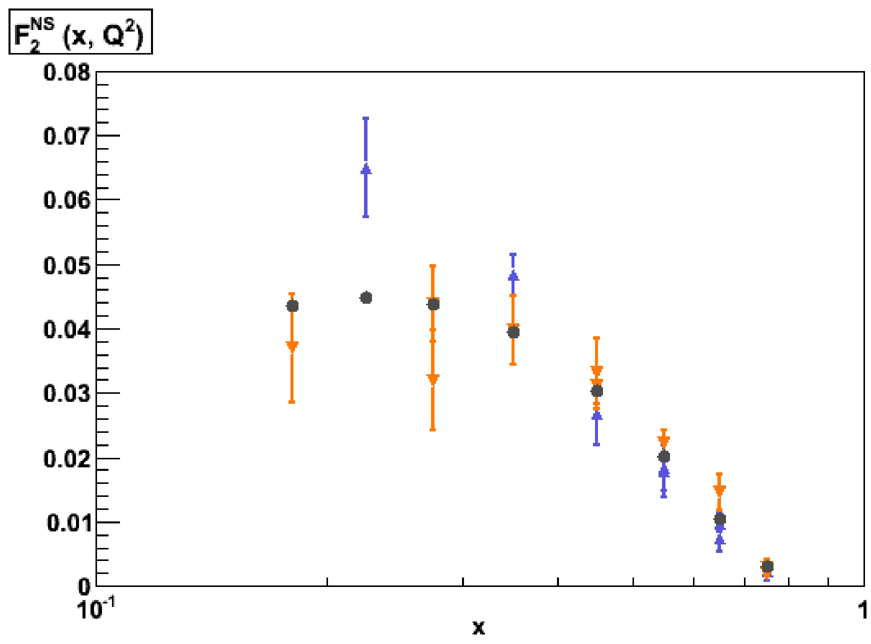
CROSS-VALIDATION

GENETIC MINIMIZATION:
AT EACH GENERATION, χ^2 EITHER UNCHANGED OR DECREASING

- DIVIDE THE DATA IN TWO SETS: TRAINING AND VALIDATION
- MINIMIZE THE χ^2 OF THE DATA IN THE TRAINING SET
- AT EACH ITERATION, COMPUTE THE χ^2 FOR THE DATA IN THE VALIDATION SET (NOT USED FOR FITTING)
- WHEN THE VALIDATION χ^2 STOPS DECREASING, STOP THE FIT

STOP!



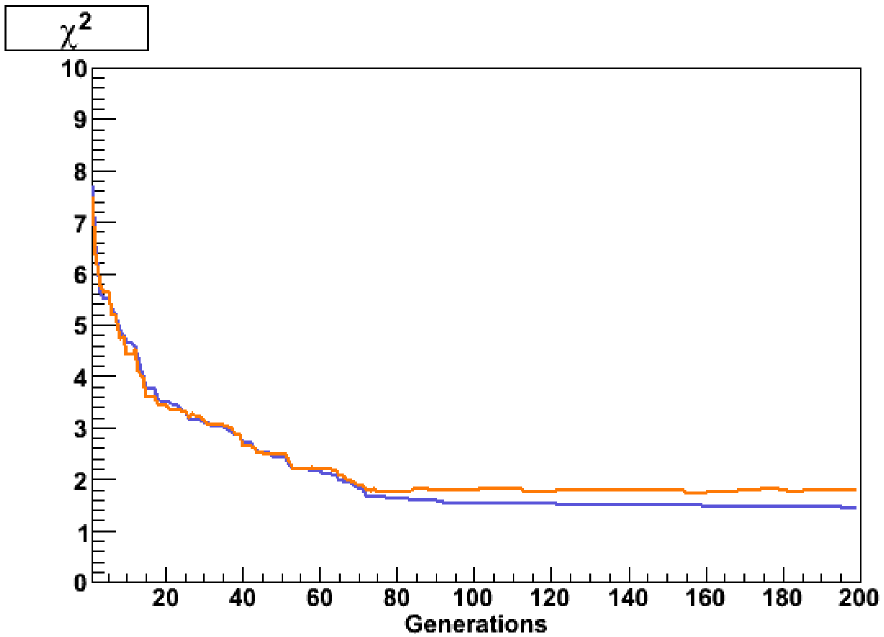


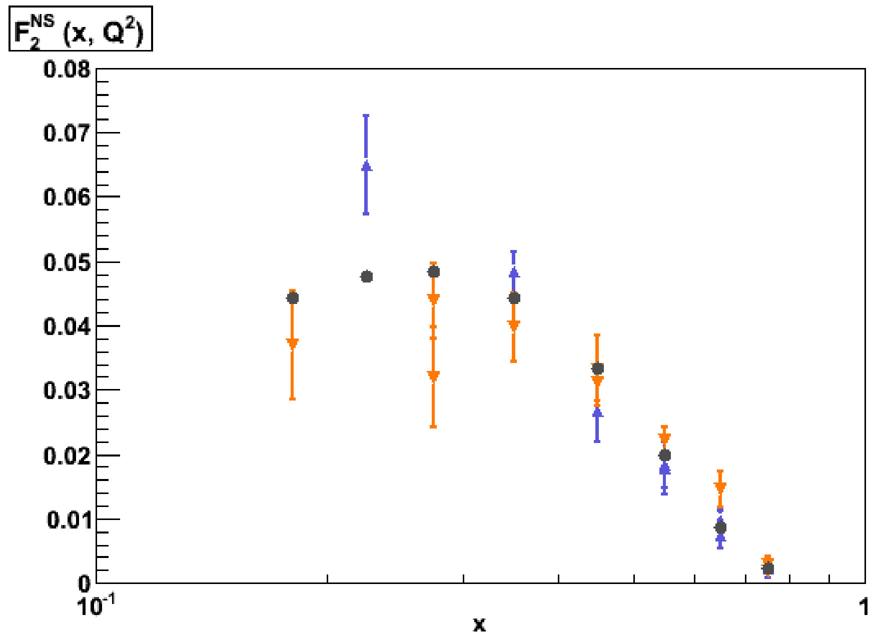
CROSS-VALIDATION

MINIMIZE BY GENETIC ALGORITHM:
AT EACH GENERATION, χ^2 EITHER UNCHANGED OR DECREASING

- DIVIDE THE DATA IN TWO SETS: TRAINING AND VALIDATION
- MINIMIZE THE χ^2 OF THE DATA IN THE TRAINING SET
- AT EACH ITERATION, COMPUTE THE χ^2 FOR THE DATA IN THE VALIDATION SET (NOT USED FOR FITTING)
- WHEN THE VALIDATION χ^2 STOPS DECREASING, STOP THE FIT

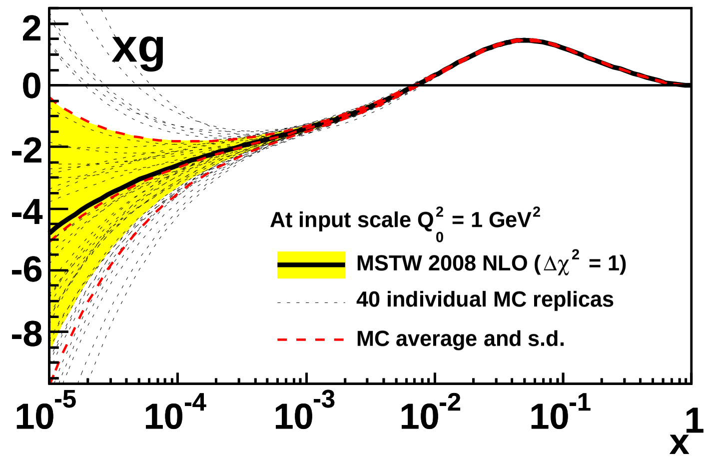
TOO LATE!



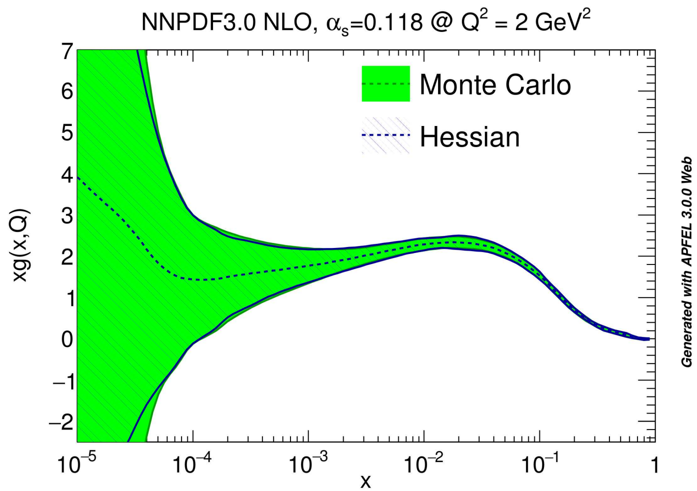


MC ⇔ HESSIAN

- TO CONVERT HESSIAN INTO MONTECARLO GENERATE MULTIGAUSSIAN REPLICAS IN PARAMETER SPACE
- ACCURATE WHEN NUMBER OF REPLICAS SIMILAR TO THAT WHICH REPRODUCES DATA



(Thorne, Watt, 2012)

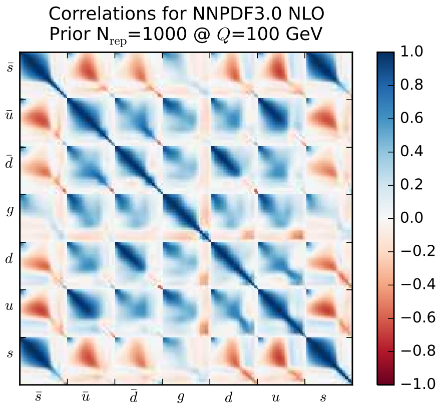


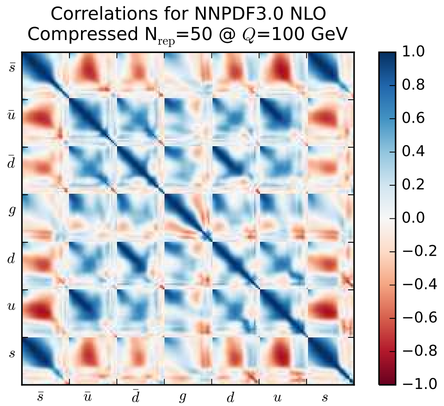
(Carrazza, SF, Kassabov, Rojo, 2015)

- TO CONVERT MONTE CARLO INTO HESSIAN, SAMPLE THE REPLICAS $f_i(x)$ AT A DISCRETE SET OF POINTS & CONSTRUCT THE ENSUING COVARIANCE MATRIX
- EIGENVECTORS OF THE COVARIANCE MATRIX AS A BASIS IN THE VECTOR SPACE SPANNED BY THE REPLICAS BY SINGULAR-VALUE DECOMPOSITION
- NUMBER OF DOMINANT EIGENVECTORS SIMILAR TO NUMBER OF REPLICAS ⇒ ACCURATE REPRESENTATION

COMPRESSION

MONTECARLO





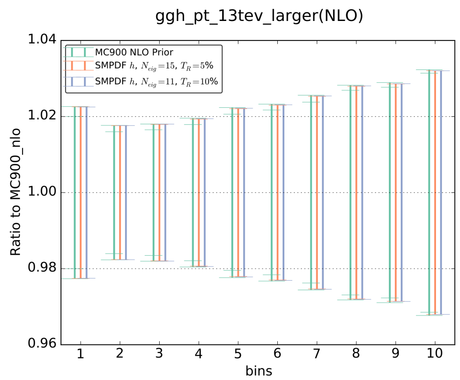
(Carrazza, Latorre, Kassabov, Rojo, 2015)

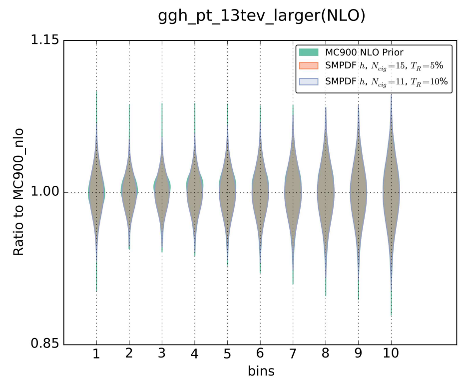
- CONSTRUCT A VERY LARGE REPLICA SAMPLE
- SELECT (BY GENETIC ALGORITHM) A SUBSET OF REPLICAS WHOSE STATISTICAL FEATURES ARE AS CLOSE AS POSSIBLE TO THOSE OF THE PRIOR
- $\Rightarrow$ FOR ALL PDFs ON A GRID OF POINTS// MINIMIZE DIFFERENCE OF: FIRST FOUR MOMENTS, CORRELATIONS; OUTPUT OF KOLMOGOROV-SMIRNOV TEST (NUMBER OF REPLICAS BETWEEN MEAN AND σ, 2σ, INFINITY)
- 50 COMPRESSED REPLICA REPRODUCE 1000 REPLICA SET TO PRECENT ACCURACY

CAN REPRODUCE NONGAUSSIAN FEATURES WITH REASONABLY SMALL REPLICA SAMPLE

HESSIAN

- SELECT SUBSET OF THE COVARIANCE MATRIX CORRELATED TO A GIVEN SET OF PROCESSES
- PERFORM SVD ON THE REDUCED COVARIANCE MATRIX, SELECT DOMINANT EIGENVECTOR, PROJECT OUT ORTHOGONAL SUBSPACE
- ITERATE UNTIL DESIRED ACCURACY REACHED
- CAN ADD PROCESSES TO GIVEN SET; CAN COMBINE DIFFERENT OPTIMIZED SETS
- 15 EIGENVECTORS DESCRIBE ALL HIGGS MODES + JETS + W, Z PRODUCTION





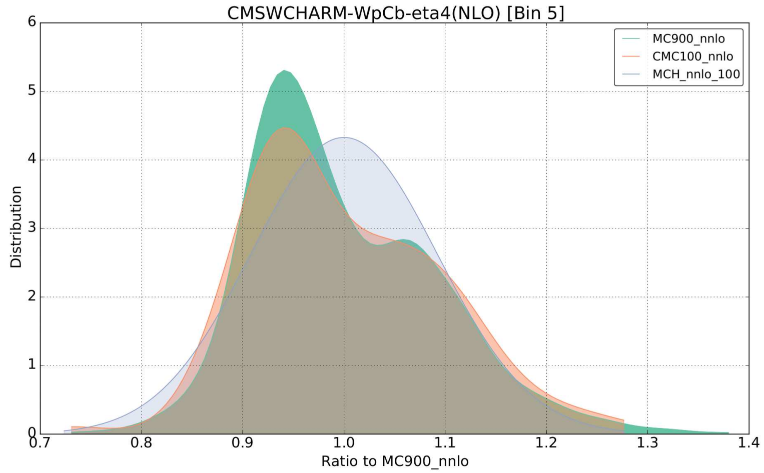
(Carrazza, SF, Kassabov, Rojo, 2016)

VERY SMALL NUMBER OF EVECS; CAN COMBINE WITH NUISANCE PARMS

NONGAUSSIAN BEHAVIOUR

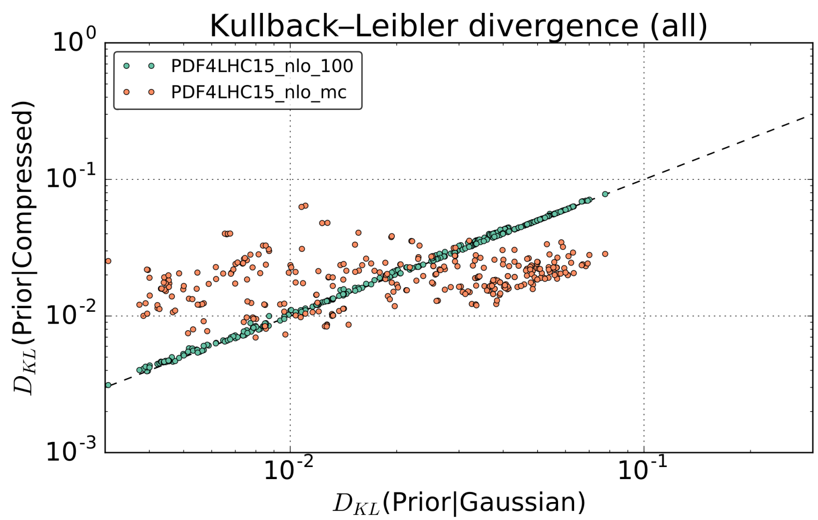
MONTE CARLO COMPARED TO HESSIAN

CMS $W + c$ production



- DEVIATION FROM GAUSSIANITY E.G. AT LARGE x DUES TO LARGE UNCERTAINTY + POSITIVITY BOUNS $\Rightarrow$ RELEVANT FOR SEARCHES
- CANNOT BE REPRODUCED IN HESSIAN FRAMEWORK
- WELL REPRODUCED BY COMPRESSED MC

- DEFINE KULLBACK-LEIBLER DIVERGENCE $D_{\mathrm{KL}} = \int_{-\infty}^{\infty} P(x) \frac{\ln P(x)}{\ln Q(x)}\, dx$ BETWEEN A PRIOR P AND ITS REPRESENTATION Q
- D_{KL} BETWEEN PRIOR AND HESSIAN DEPENDS ON DEGREE OF GAUSSIANITY
- D_{KL} BETWEEN PRIOR AND COMPRESSED MC DOES NOT



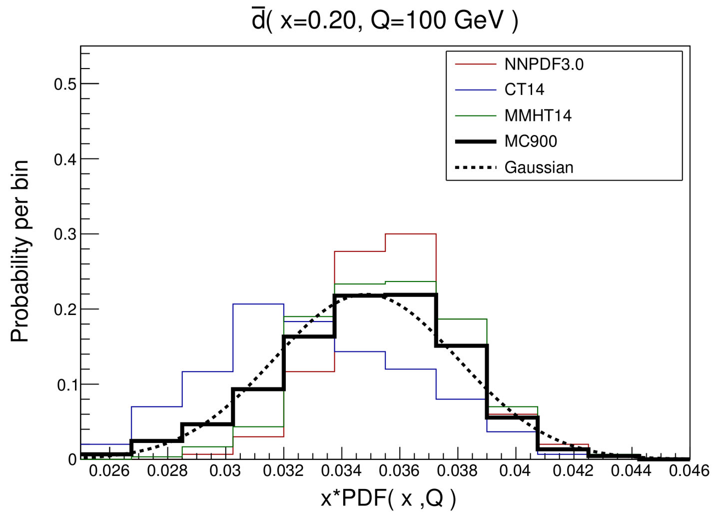
CAN GAUGE WHEN MC IS MORE ADVANTAGEOUS THAN HESSIAN!

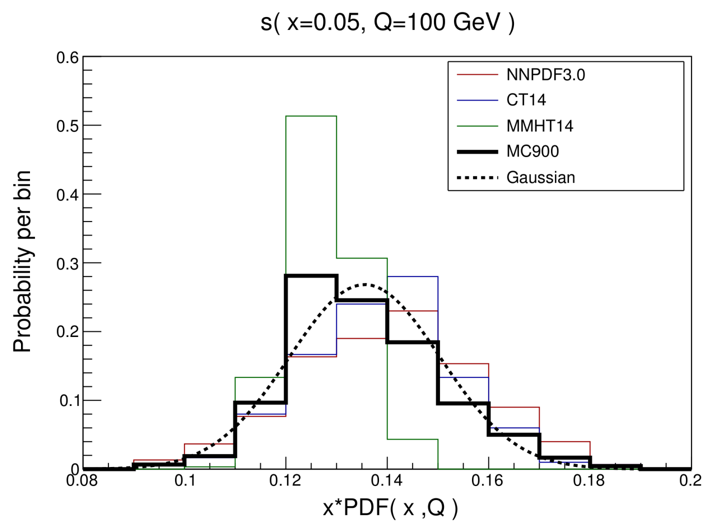
MONTE CARLO COMBINATION

(Watt, S.F., 2010-2013)

- MAY COMBINE DIFFERENT PDF SETS,
  AFTER MC CONVERSION OF HESSIAN SETS
- COMBINE MONTE CARLO REPLICAS INTO SINGLE SET
- USEFUL FOR CONSERVATIVE UNCERTAINTY ESTIMATE
- COMBINED SET APPROXIMATELY GAUSSIAN

COMBINED PDF4LHC SETS FOR ANTIDOWN & STRANGE





THE ULTIMATE CHECK OF PDF DETERMINATION

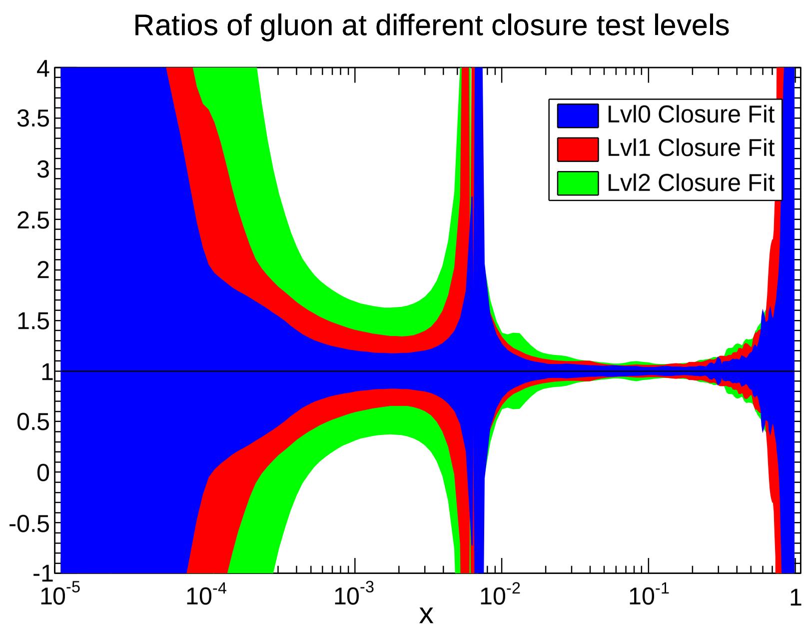
CLOSURE TESTS

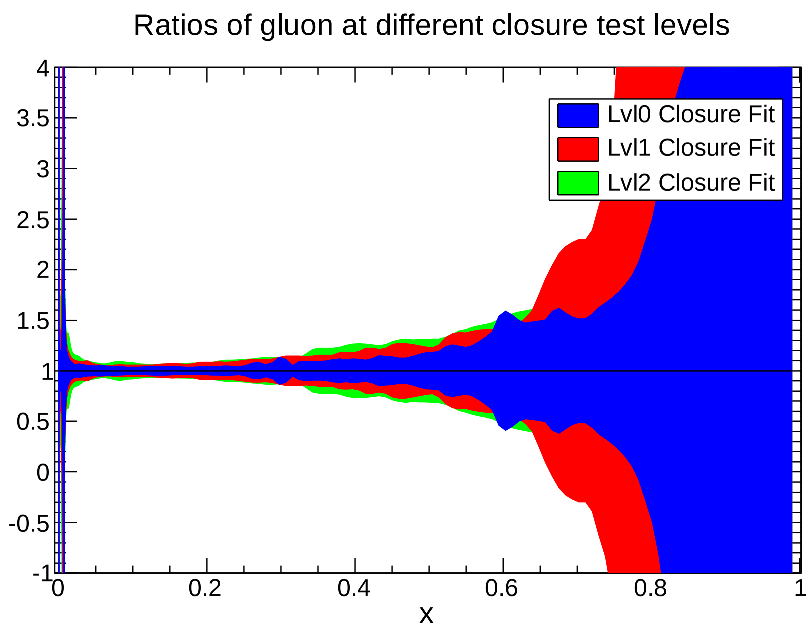
- ASSUME PDFS KNOWN: GENERATE FAKE EXPERIMENTAL DATA
- CAN DECIDE DATA UNCERTAINTY (ZERO, OR AS IN REAL DATA, OR . . .)
- FIT PDFS TO FAKE DATA
- CHECK WHETHER FIT REPRODUCES UNDERLYING "TRUTH":
  - CHECK WHETHER TRUE VALUE GAUSSIANLY DISTRIBUTED ABOUT FIT
  - CHECK WHETHER UNCERTAINTIES FAITHFUL
  - TRACE DIFFERENT SOURCES OF UNCERTAINTY

TRACING SOURCES OF UNCERTAINTY

- LEVEL 0: FAKE DATA GENERATED WITH NO UNCERTAINTY
  $\rightarrow$ INTERPOLATION AND EXTRAPOLATION UNCERTAINTY
- LEVEL 1-2: FAKE DATA GENERATED WITH SAME UNCERTAINTY AS REAL DATA (INCLUDING CORRELATIONS)
- LEVEL 1: NO PSEUDODATA REPLICAS:
  $\Rightarrow$ REPLICAS FITTED TO SAME DATA OVER AND OVER AGAIN
  $\rightarrow$ FUNCTIONAL UNCERTAINTY DUE TO INFINITY OF EQUIVALENT MINIMA
- LEVEL 2: STANDARD NNPDF METHODOLOGY
  $\Rightarrow$ REPLICAS FITTED TO PSEUDODATA REPLICAS
  $\rightarrow$ DATA UNCERTAINTY
- THREE SOURCES OF UNCERTAINTY COMPARABLE IN DATA REGION

THE GLUON: LEVEL 0, LEVEL 1 AND LEVEL 2



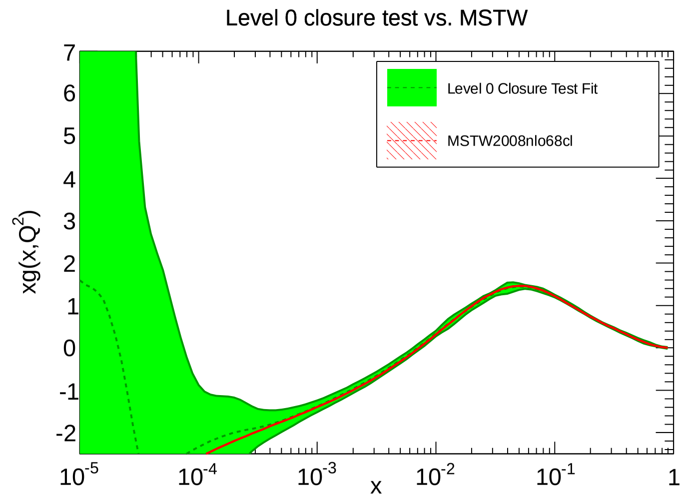


FITTING EFFICIENCY

LEVEL 0

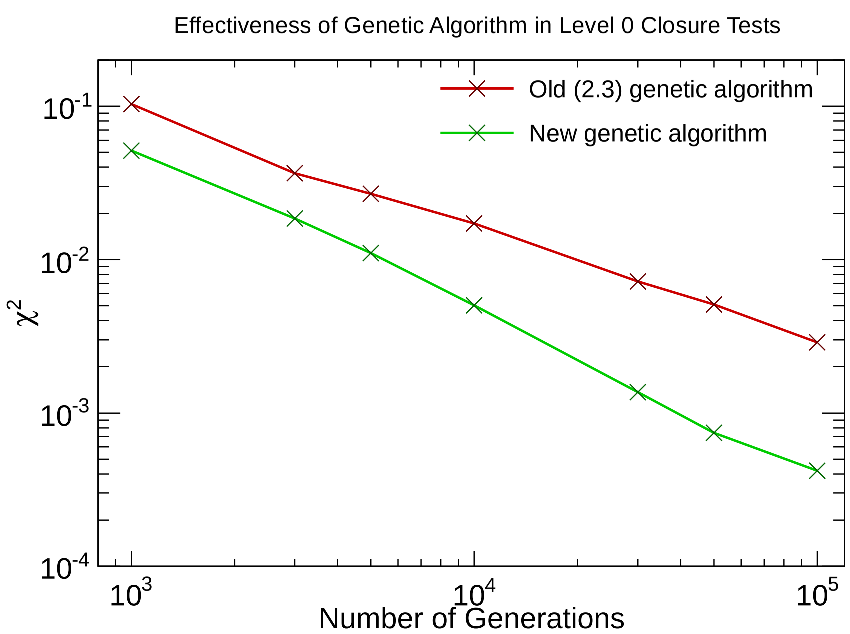
- ASSUME VANISHING EXPERIMENTAL UNCERTAINTY
- MUST BE ABLE TO GET $\chi^2 = 0$
- UNCERTAINTY AT DATA POINTS TENDS TO ZERO (NOT NECESSARILY ON PDF!)
  DEFINE $\phi \equiv \sqrt{\langle \chi^2_{rep} \rangle - \chi^2}$,
  EQUALS FIT UNCERTAINTY/DATA UNCERTAINTY; CHECK $\phi \to 0$
- CAN STUDY EFFICIENCY OF MINIMIZATION

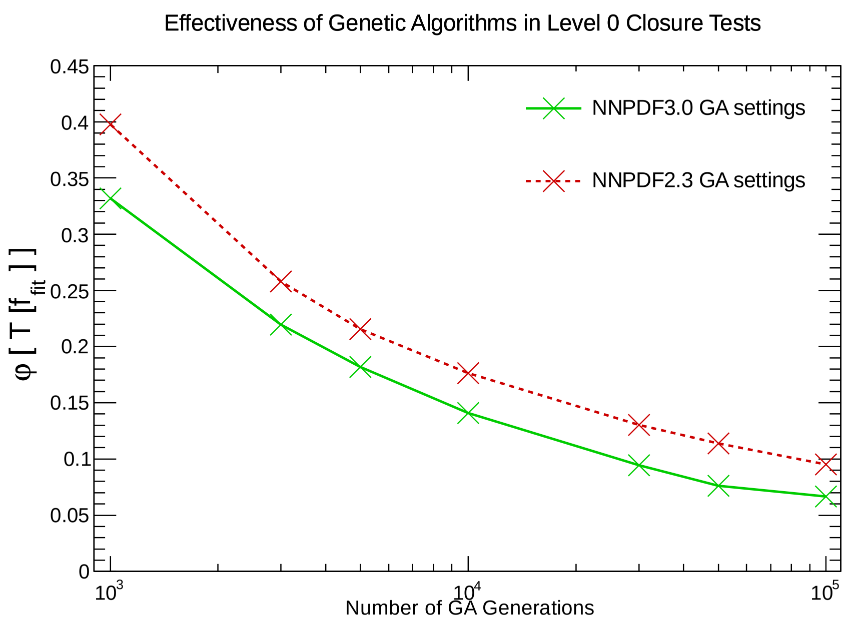
THE GLUON



FRACTIONAL UNCERTAINTY VS TRAINING LENGTH

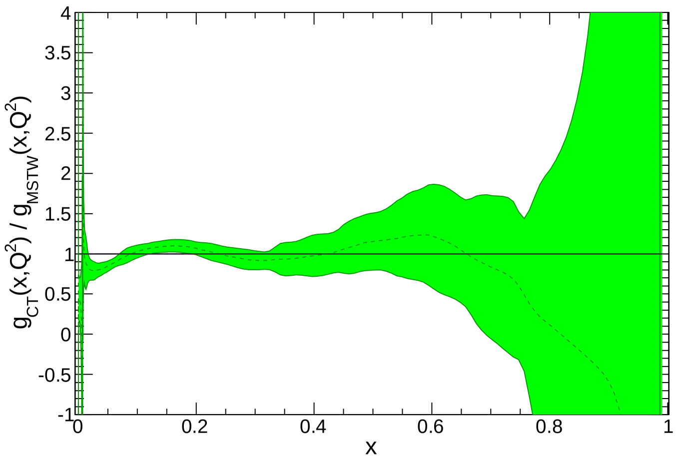
χ^2 VS TRAINING LENGTH





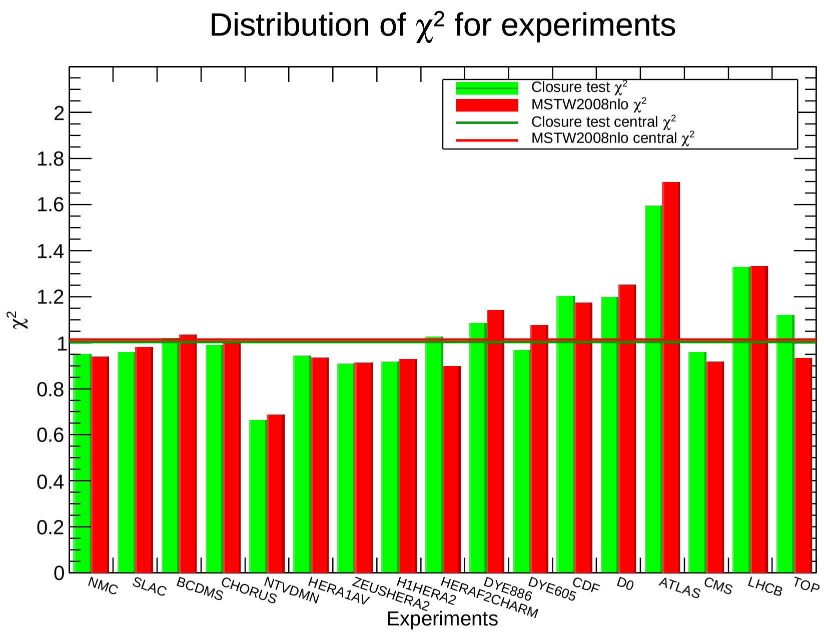
TESTING THE PDF DETERMINATION
LEVEL 2: THE WORKS

THE GLUON: FITTED/"TRUE"

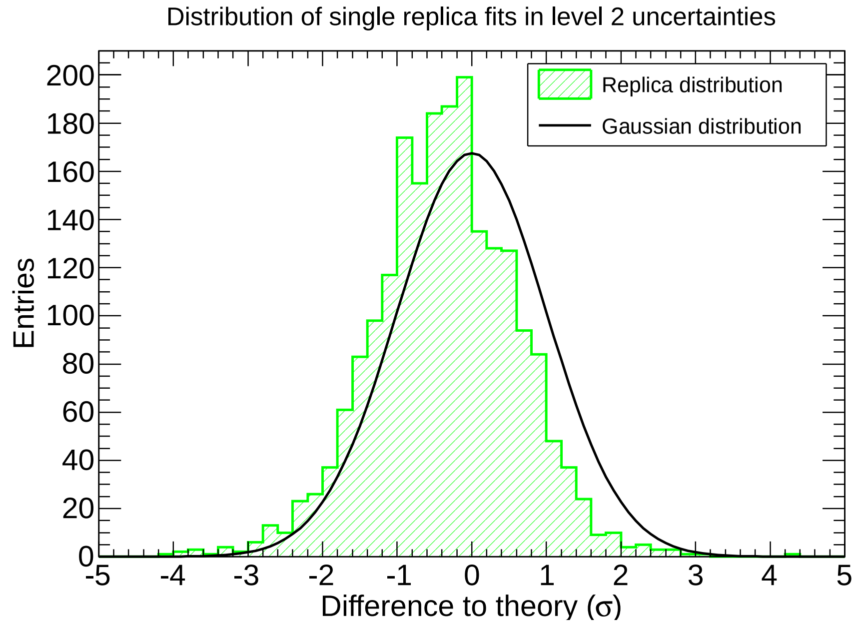


- CENTRAL VALUES:
  COMPARE FITTED VS. "TRUE" χ^2
  BOTH FOR INDIVIDUAL EXPERIMENTS
  & TOTAL DATASET
  FOR TOTAL $\Delta\chi^2 = 0.001 \pm 0.003$
- UNCERTAINTIES: DISTRIBUTION OF DEVIATIONS BETWEEN FITTED AND "TRUE" PDFS
  SAMPLED AT 20 POINTS BETWEEN 10^{-5} AND 1
  FIND 0.699% FOR ONE-SIGMA,
  0.948% FOR TWO-SIGMA C.L.

LEVEL-2 FITTED χ^2 VS "TRUE"



NORM. DISTRIBUTION OF DEVIATIONS

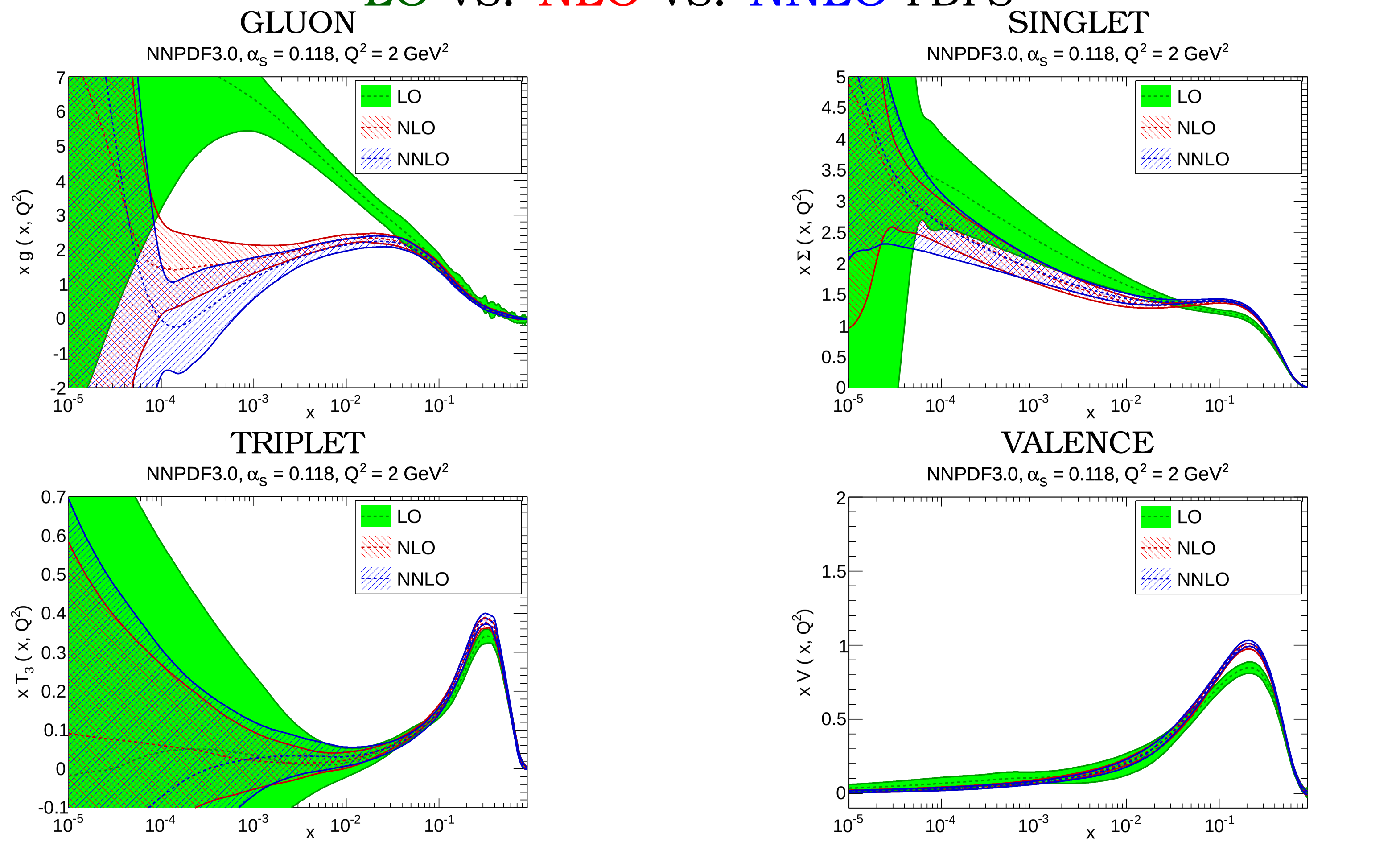


METHODOLOGY SUMMARY

- PDF determination: Hessian method
  - simple linear error propagation
  - tolerance required for realistic uncertainties
  - parametrization bias possible
- PDF determination: Monte Carlo method
  - two-step procedure: data Monte Carlo ⇒ PDF Monte Carlo
  - very general parametrization allowed
  - need optimal fit determination method (cross-validation)
- PDF representation: Hessian vs Monte Carlo
  - conversion possible either way
  - compression methods available either way
  - Monte Carlo very flexible, Hessian very efficient
- PDF validation: closure test
  - performed in the Monte Carlo approach
  - interpolation & functional uncertainties significant

THEORY: ISSUES & PROGRESS

PERTURBATIVE STABILITY I
LO VS. NLO VS. NNLO PDFS



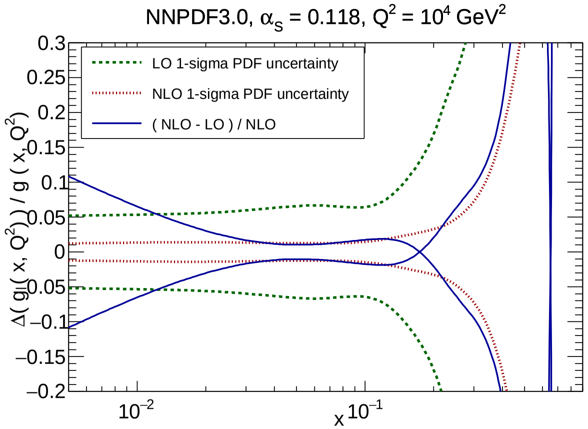
- PERTURBATIVE ACCURACY OF PREDICTION LIMITED BY PERTURBATIVE ACCURACY OF PDF
- $\alpha_s(M_z) \sim 0.1$, $\alpha_s(M_p) \sim 1/2$; $\alpha_s(Q_1^2) = \alpha_s(Q_2^2)(1 + O(\alpha_s^2))$
  $\Rightarrow$ LO: QUALITATIVE; NLO: QUANTITATIVE; NNLO: PRECISION

PERTURBATIVE STABILITY II

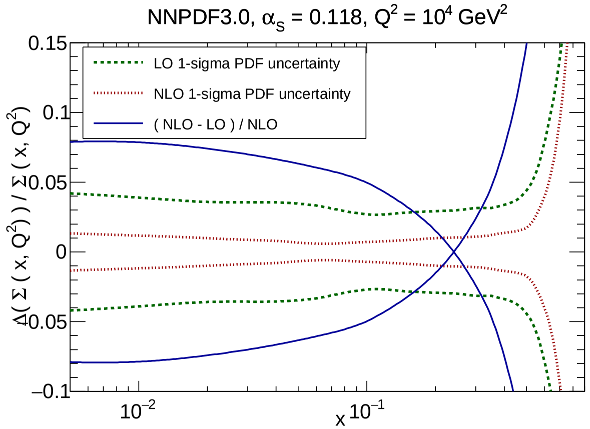
THEORY UNCERTAINTIES VS PDF UNCERTAINTIES

LO VS NLO VS LO-NLO SHIFT

GLUON

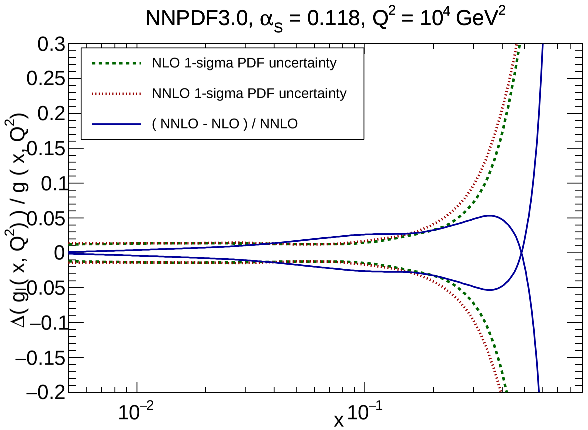


SINGLET

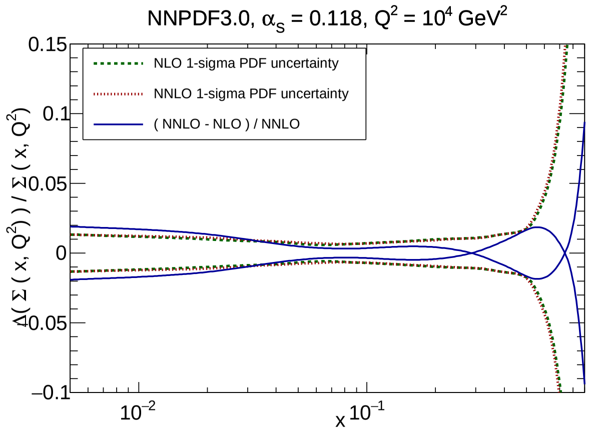


NLO VS NNLO VS NLO-NNLO SHIFT

GLUON



SINGLET



- "DATA" PDF UNCERTAINTY INDEP. OF PERTURBATIVE ORDER
- TH. UNCERTAINTY (MHOU) VS DATA UNCERTAINTY $\Rightarrow$ LO: DOMINANT; NLO, COMPARABLE; NNLO: SUBDDOMINANT

THEORETICAL UNCERTAINTIES ON PDFs:

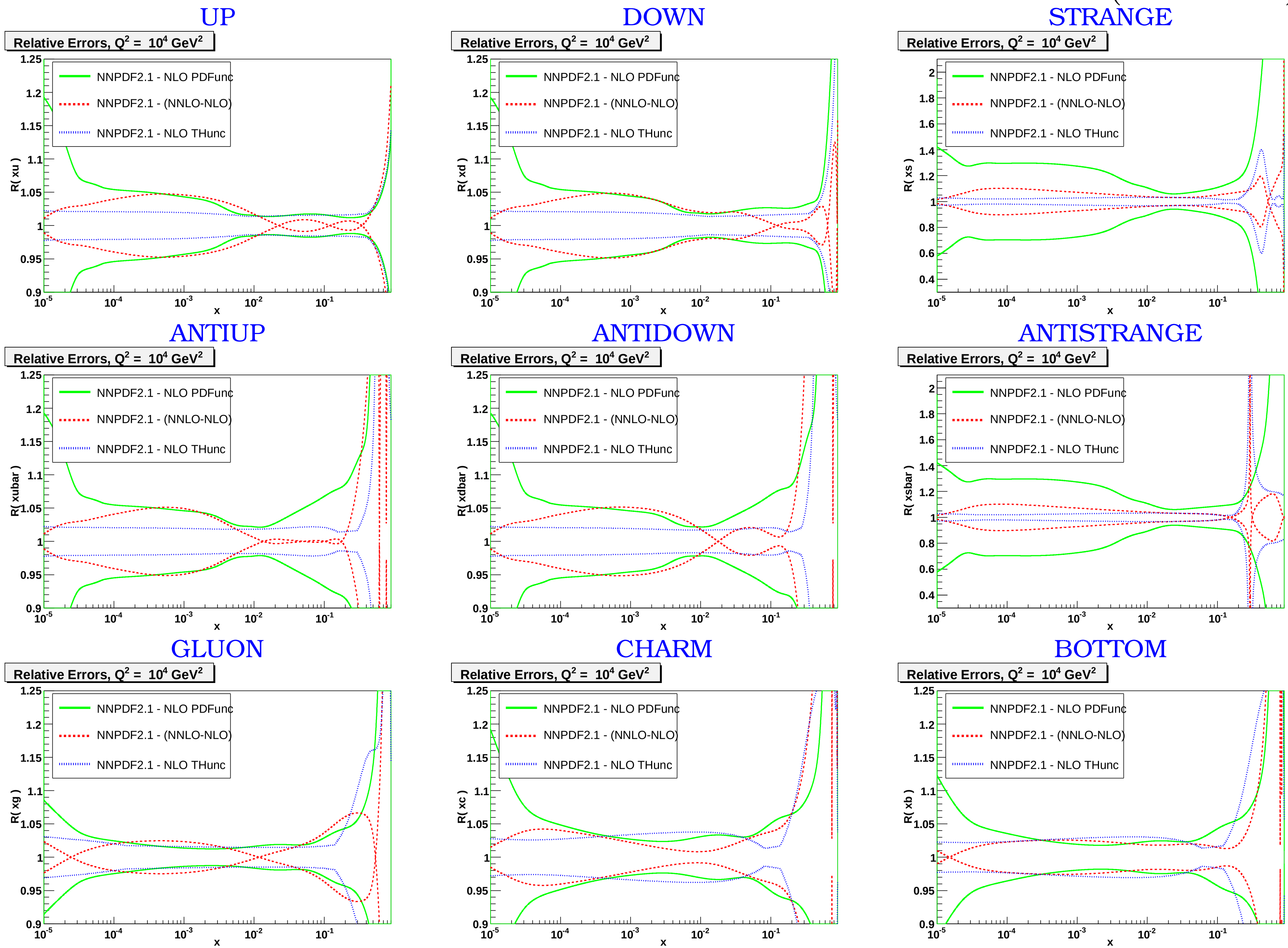
- PDFs ARE DETERMINED BY COMPARING TO DATA THEORY AT SOME FINITE ORDER
- AFFECTED BY THEORETICAL UNCERTAINTY JUST LIKE HARD CROSS-SECTIONS
- NOT INCLUDED IN CURRENT "PDF UNCERTAINTY"
(ACCOUNTS ONLY DATA & METHODOLOGY)

CAN WE ESTIMATE THEM?

- SCALE VARIATION DIFFICULT:
CORRELATED BETWEEN PROCESSES? HOW DOES IT CORRELATE WITH PROCESSES IN WHICH PDFs ARE USED?
- AT NLO: WE KNOW THE SHIFT TO NNLO
- AT NNLO: LOOK AT THE BEHAVIOUR OF THE PERTURBATIVE EXPANSION
(CACCIARI-HOUDEAU)

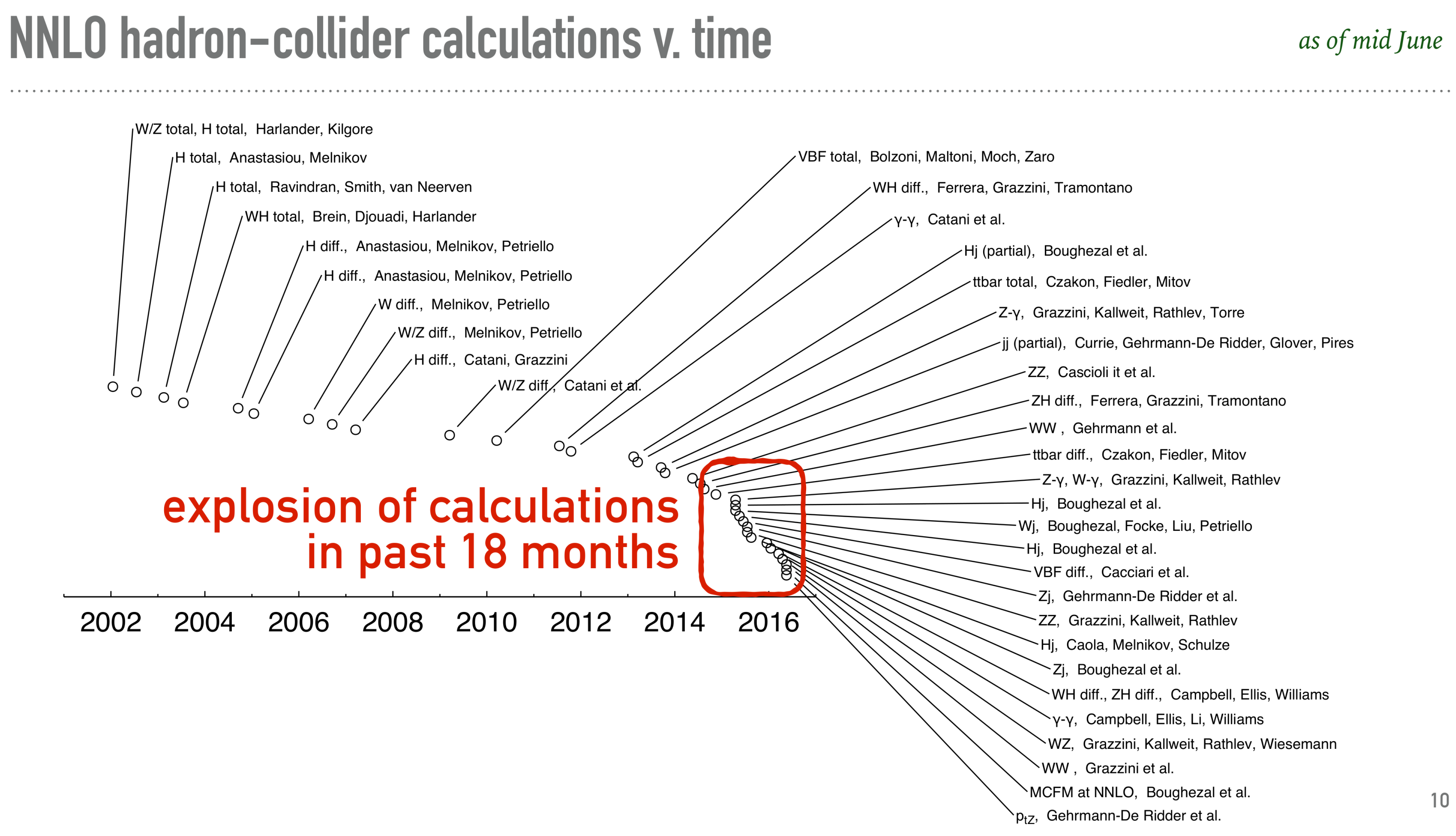
THEORETICAL UNCERTAINTIES

NLO PDF UNC. VS NLO-NNLO SHIFT VS NLO CACCIARI-HOUDEAU (NNPDF2.1)



CACCIARI-HOUDEAU PROMISING?

HIGHER ORDERS: THEORETICAL PROGRESS

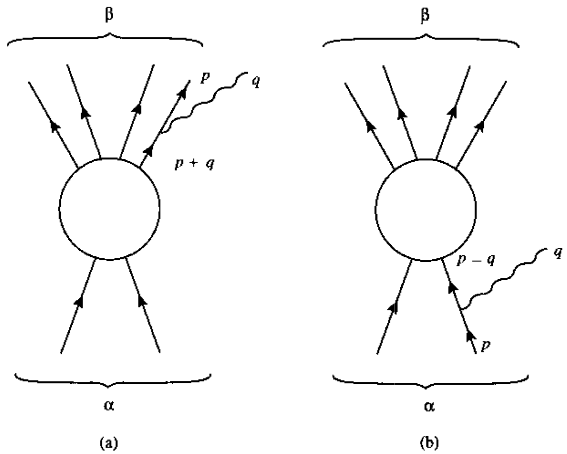


(G. Salam, 2016)

- IMPRESSIVE ACCELERATION OF PERTURBATIVE COMPUTATIONS SINCE INCEPTION OF LHC
- $\Rightarrow$ HANDLE ON THEORETICAL UNCERTAINTIES

BEYOND FIXED ORDER: RESUMMATION

EIKONAL EMISSION



EMISSION OF A SOFT ($q^\mu \to 0$) GAUGE PARTICLE FROM EXTERNAL LINE

$$\sigma(\alpha \to \beta) \to \sigma(\alpha \to \beta) \frac{ep^\mu}{p \cdot q - i\epsilon}$$

(Bloch, Nordiseck, 1937; Yenni, Frautschi, Suura, 1955; Weinberg, 1964)

- SOFT EMISSION $\Rightarrow$ EIKONAL FACTOR
- CROSS SECTION FOR SINGLE (DOUBLE. . .) EMISSION
  INFRARED DIVERGENT; DIVERGENCE CANCELLED BY VIRTUAL CORRECTIONS
- 1,2,. . . ,N EMISSIONS EXPONENTIATE $\Rightarrow \Gamma \sim \exp - \left[\alpha \ln^2\left(1 - \frac{M_\beta^2}{s}\right)\right]$

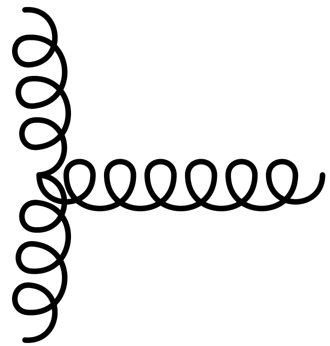
  (Sudakov, 1956)
- AFTER CANCELLATION, LEFTOVER SOFT LOGS:
  $\sigma(\alpha \to \beta) \to \sigma(\alpha \to \beta) \ln^2\left(1 - \frac{M_\beta^2}{s}\right)$

RESUMMATION:
EXPONENTIATION

- EXPONENTIATION OF LEFTOVER LOGS $\Rightarrow$
  THRESHOLD RESUMMATION OF $\alpha_s \ln^2(1-x)$, $x = \frac{M^2}{s}$
- LOGS COME IN PAIRS: SOFT+COLLINEAR $\to \ln p_t$ WHEN INTEGRAL OVER p_t NOT PERFORMED$\Rightarrow$
  TRANSVERSE MOMENTUM RESUMMATION OF $\alpha_s \ln^2 \frac{q_T^2}{M^2}$
- IN GLUON CHANNEL SYMMETRY OF THE TRIPLE GLUON VERTEX $\to$ LARGE LOGS ALSO WHEN THE EXCHANGED GLUON IS SOFT: NO COLLINEAR CONTRIBUTION, SINGLE LOGS $\to \ln \frac{s}{Q^2} \Rightarrow$
  HIGH ENERGY RESUMMATION OF $\alpha_s \ln \frac{1}{x}$

GLUON RADIATION

$$\sigma(\tau, M^2) = \int_y^1 \frac{dy}{y} P\left(\frac{x}{y}\right) \int_{\mu^2}^{(s-M^2)^2/s} \frac{dk_t^2}{k_t^2} \hat{\sigma}(y, M^2)$$



THE GLUON SPLITTING FUNCTION:

$$P_{gg}(x) = 2C_A \left[\frac{x}{(1-x)_+} + \frac{1-x}{x} + x(1-x)\right] + \beta_0 \delta(1-x)$$

LOGARITHMICALLY ENHANCED TERMS

- INFRARED LOGS: $\int_\tau^1 dy \frac{1}{1-y}_+ \sim \ln(1-\tau)$
- UV LOGS: $\int_\tau^1 dy \frac{1}{y} \sim \ln(\tau)$
- COLLINEAR LOGS: $\int_{\mu^2}^{(s-M^2)^2/s} \frac{dk_t^2}{k_t^2} \sim \ln\left[\frac{Q^2}{\mu^2} \frac{(1-\tau)^2}{\tau}\right] = \ln \frac{Q^2}{\mu^2} + \ln(1-\tau)^2 + \ln \tau$

FACTORIZATION REMINDER

THE FACTORIZED CROSS SECTION

$$\sigma(\tau, M^2) = \tau \sum_{ij} \int_\tau^1 \frac{dz}{z}\, \mathcal{L}_{ij}\left(\frac{\tau}{z}, \mu_{\rm F}^2\right) \frac{1}{z}\hat\sigma_{ij}\left(z, M^2, \alpha_s(\mu_{\rm R}^2), \frac{M^2}{\mu_{\rm F}^2}, \frac{M^2}{\mu_{\rm R}^2}\right) \qquad \tau = \frac{M^2}{s}$$

PARTON LUMINOSITIES

$$\mathcal{L}_{ij}(z, \mu^2) = \int_z^1 \frac{dx}{x}\, f_i\left(\frac{z}{x}, \mu^2\right) f_j(x, \mu^2)$$

COEFFICIENT FUNCTIONS

$$\hat\sigma_{ij}\left(z, M^2, \alpha_s(\mu_{\rm R}^2), \frac{M^2}{\mu_{\rm F}^2}, \frac{M^2}{\mu_{\rm R}^2}\right) = z\,\sigma_0\left(M^2, \alpha_s(\mu_{\rm R}^2)\right)\, C_{ij}\left(z, \alpha_s(\mu_{\rm R}^2), \frac{M^2}{\mu_{\rm F}^2}, \frac{M^2}{\mu_{\rm R}^2}\right)$$

$$C_{ij}(z, \alpha_s) = \delta(1-z)\delta_{ig}\delta_{jg} + \alpha_s C_{ij}^{(1)}(z) + \alpha_s^2 C_{ij}^{(2)}(z) + \alpha_s^3 C_{ij}^{(3)}(z) + \mathcal{O}(\alpha_s^4)$$

MELLIN-SPACE FACTORIZATION

$$\sigma(N, M^2) = \sigma_0\left(M^2, \alpha_s\right) \mathcal{L}(N) C(N, \alpha_s),$$

$$\sigma(N, M^2) \equiv \int_0^1 d\tau\, \tau^{N-2} \sigma(\tau, M^2);\ \mathcal{L}(N) \equiv \int_0^1 dz\, z^{N-1} \mathcal{L}(z)\ C(N, \alpha_s) \equiv \int_0^1 dz\, z^{N-1} C(z, \alpha_s)$$

THE STRUCTURE OF RESUMMED EXPRESSIONS
THRESHOLD RESUMMATION

LOG COUNTING

$$C_{\rm res}(N,\alpha_s) = g_0(\alpha_s)\exp\left[\frac{1}{\alpha_s}g_1(\alpha_s\ln N) + g_2(\alpha_s\ln N) + \alpha_s g_3(\alpha_s\ln N) + \dots\right];$$

$g_0(\alpha_s) = 1 + \alpha_s g_{0,1} + \alpha_s^2 g_{0,2} + \mathcal{O}(\alpha_s^3)$; $\quad g_1(\lambda) = \sum_{k=2}^{\infty} g_{1,k}\lambda^k$, $g_i(\lambda) = \sum_{k=1}^{\infty} g_{i,k}\lambda^k$ FOR $i \geq 2$

LOG APPROX.	XSECT ACCURACY	EXP. ACCURACY: $\alpha_s^n L^k$	g_0 ACCURACY: α_s^i
LL	$k = 2n$	$k = n+1$	0
NLL	$2n-2 \leq k \leq 2n$	$k = n$	1
NNLL	$2n-4 \leq k \leq 2n$	$k = n-1$	2

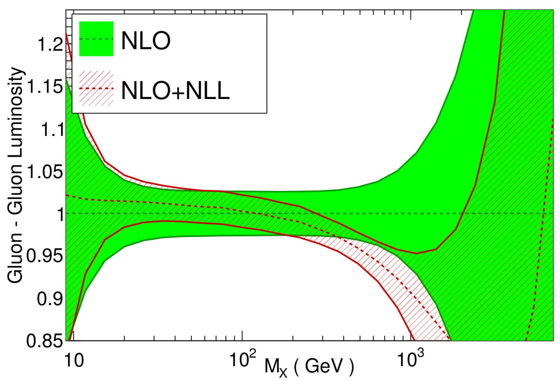
THE RESUMMED EXPONENT

$$S\left(M^2, \frac{M^2}{N^2}\right) = \int_{M^2}^{M^2/N^2} \frac{d\mu^2}{\mu^2}\bar\gamma\left(\alpha_s(\mu^2), \frac{M^2}{N^2\mu^2}\right)$$
$$= \int_{M^2}^{M^2/N^2} \frac{d\mu^2}{\mu^2}\left[-A(\alpha_s(\mu^2))\ln\left(\frac{M^2/N^2}{\mu^2}\right) + B[\alpha_s(\mu^2)]\right].$$

- A, B ARE POWER SERIES IN α_s
- A IS UNIVERSAL, COEFFICIENT OF $\ln N$ IN (DIAGONAL QUARK OR GLUON) ANOMALOUS DIMENSION AT EACH ORDER
- B CONTAINS PROCESS-DEPENDENT TERMS, STARTS AT NLL

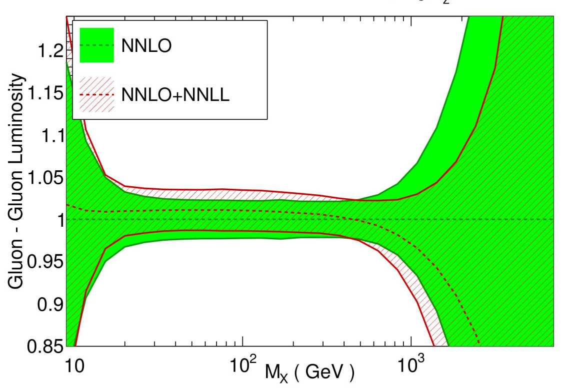
RESUMMED PDFs

GLUON: NLO vs NLL

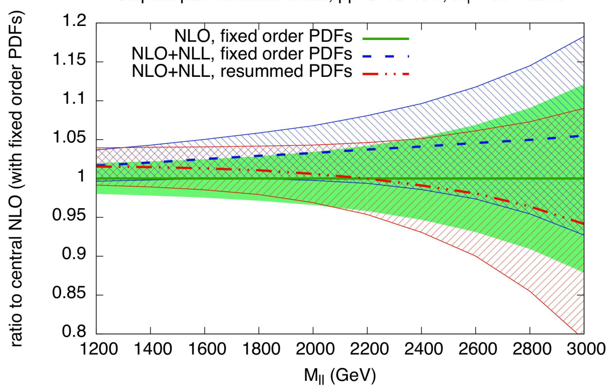


- SO FAR NO RESUMMED PDF SETS AVAILABLE
- PRELIMINARY STUDY: IF THRESHOLD RESUMMATION INCLUDED IN FIT (DIS, DY, TOP DATA), EFFECTS NOT NEGLIGIGLE AT NLLO, LARGE x, MORE MODERATE AT NNLO
- EFFECT ON PDFs COMPARABLE TO EFFECT ON MATRIX ELEMENT, ANTICORRELATED TO IT
- RELEVANT FOR NEW PHYSICS SEARCHES

GLUON: NNLO vs NNLL

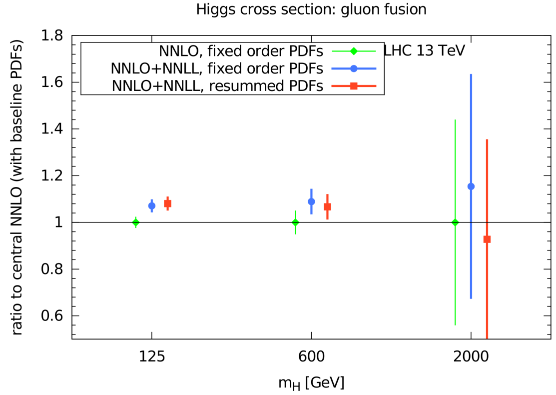


SLEPTON PAIR PRODUCTION



Bonvini et al., 2015

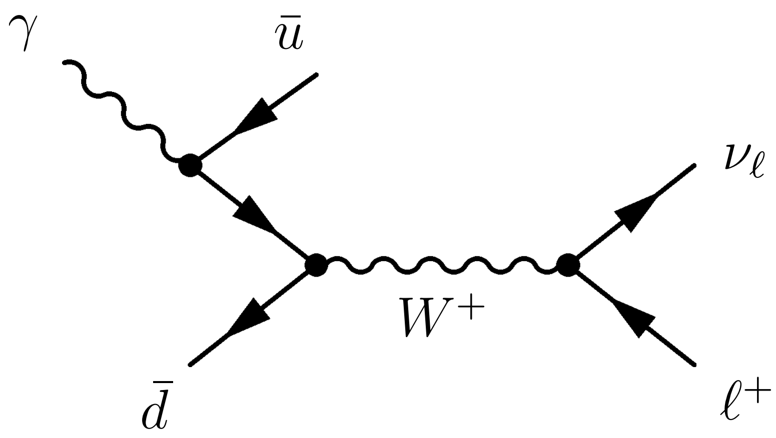
HIGGS IN GLUON FUSION VS m_H

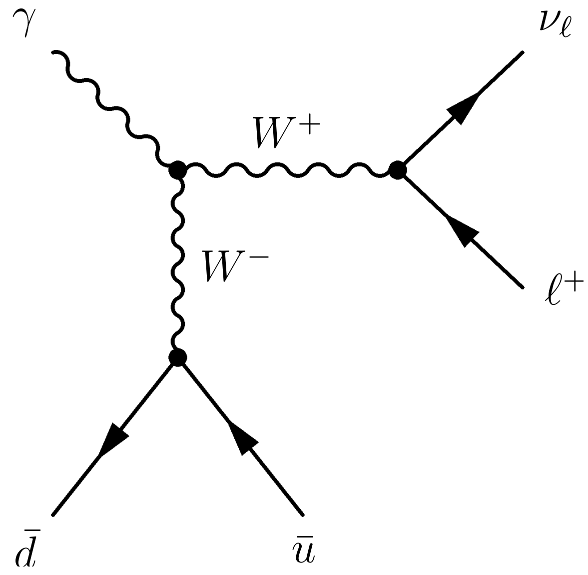


THE PHOTON PDF

- THE SAME FACTORIZATION ARGUMENTS APPLY TO QCD AND QED
- THE PROTON ALSO HAS A PHOTON CONTENT $\Rightarrow$ PHOTON PDF
- PHOTONS RADIATED BY QUARKS;
  $\alpha_s \to \alpha$;
- QED INDUCED CONTRIBUTIONS TO HADRON COLLIDER PROCESSES
  $\alpha(M_z) \sim \alpha_s(M_z)/10 \Rightarrow$ NLO QED CORRECTIONS$\sim$ NNLO QCD CORRECTIONS

EXAMPLE: PHOTON-INDUCED DRELL-YAN





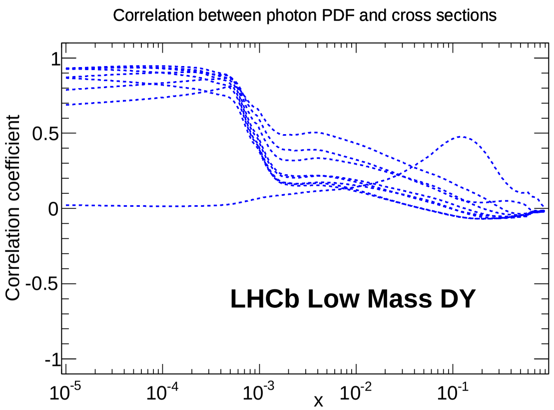
THE PHOTON PDF FROM DATA

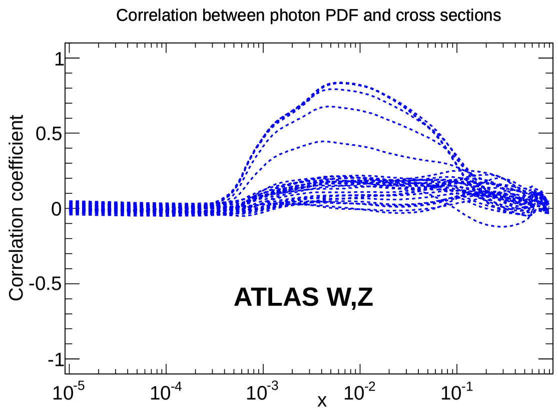
NNPDF2.3QED/NNPDF3.0QED DATASET

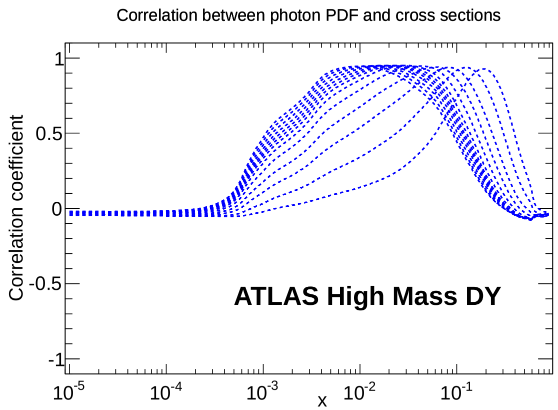
Dataset	Observable	$N_{\rm dat}$	$[\eta_{\rm min},\eta_{\rm max}]$	$[M_{\rm ll}^{\rm min},M_{\rm ll}^{\rm max}]$
LHCb γ^*/Z Low Mass	$d\sigma(Z)/dM_{ll}$	9	[2,4.5]	[5,120] GeV
ATLAS W,Z	$d\sigma(W^\pm,Z)/d\eta$	30	[-2.5,2.5]	[60,120] GeV
ATLAS γ^*/Z High Mass	$d\sigma(Z)/dM_{ll}$	13	[-2.5,2.5]	[116,1500] GeV

IMPACT

CORRELATION BETWEEN DATA AND γ PDF



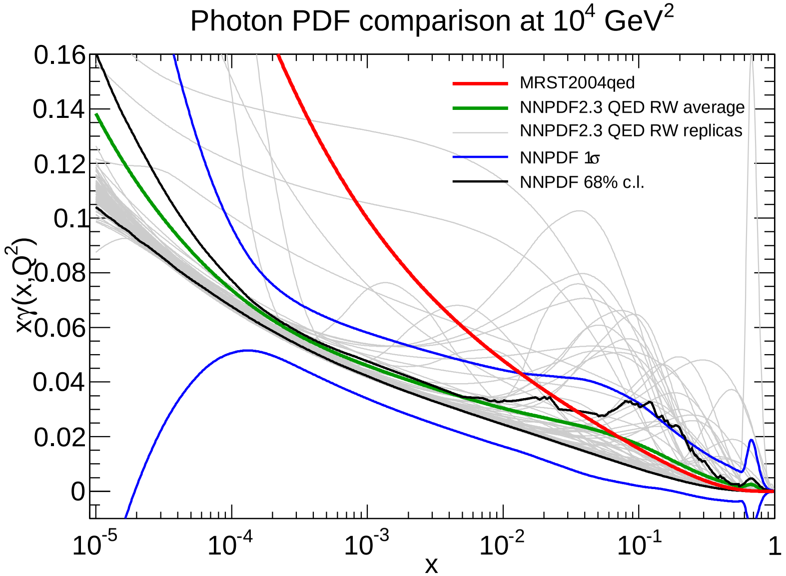


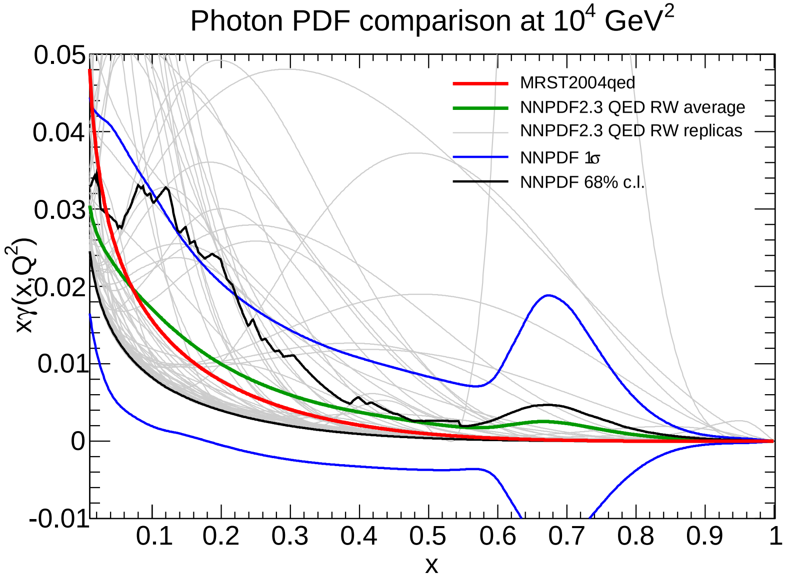


THE PHOTON PDF

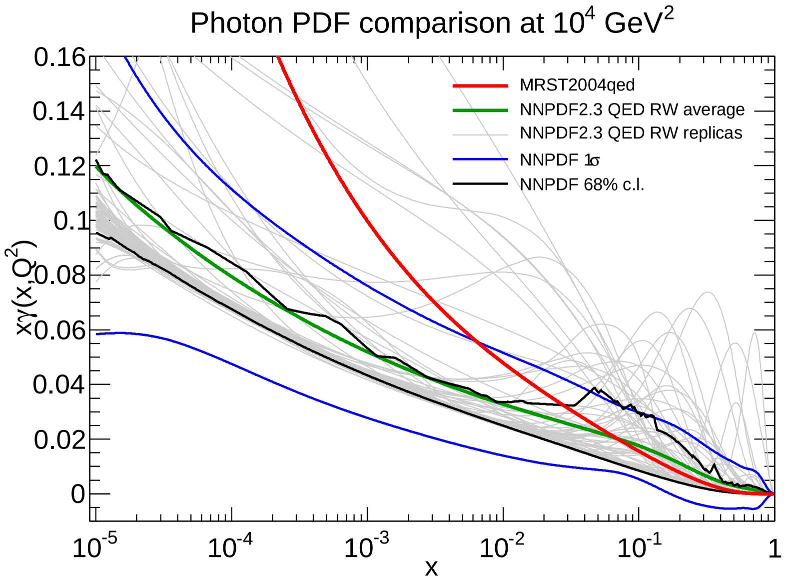
NNPDF2.3QED-NNPDF3.0QED

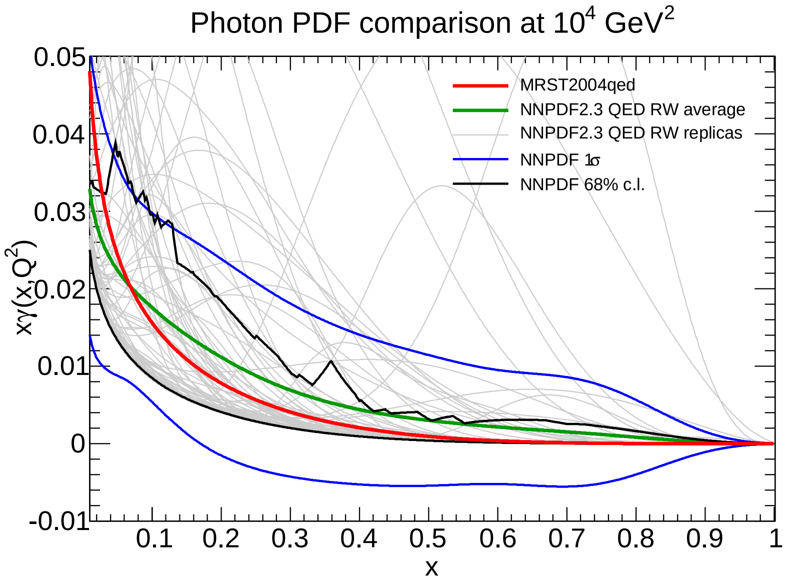
NLO RESULTS





NNLO RESULTS



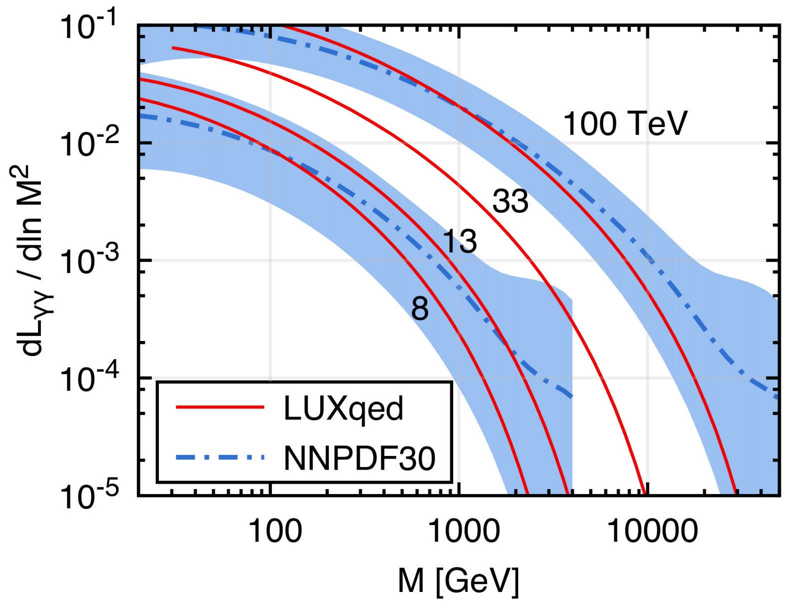


THE PHOTON PDF BREAKTHROUGH

(Manohar, Nason, Salam, Zanderighi, 2016)

- QED IS PERTURBATIVE DOWN TO LOW SCALES ⇒ THE PHOTON PDF MUST BE COMPUTABLE IF THE INPUT QUARK SUBSTRUCTURE IS KNOWN
- WRITE THE CROSS-SECTION FOR A CHOSEN PROCESS: SUSY PRODUCTION IN EP COLLISION (Drees, Zeppenfeld, 1989)
- COMPUTE IT DIRECTLY, OR USING THE PHOTON PDF
- ⇒ PDF EXPRESSED IN TERMS OF THE STRUCTURE FUNCTION INTEGRATED OVER ALL SCALES, INCLUDING ELASTIC FORM FACTORS

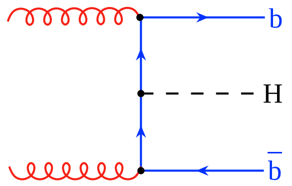
$$x f_{\gamma/p}(x,\mu^2) = \frac{1}{2\pi\alpha(\mu^2)}\int_x^1 \frac{dz}{z}\left\{\int_{\frac{x^2 m_p^2}{1-z}}^{\frac{\mu^2}{1-z}} \frac{dQ^2}{Q^2}\alpha^2(Q^2)\left[\left(z p_{\gamma q}(z) + \frac{2x^2 m_p^2}{Q^2}\right)F_2(x/z,Q^2) - z^2 F_L\left(\tfrac{x}{z},Q^2\right)\right] - \alpha^2(\mu^2) z^2 F_2\left(\tfrac{x}{z},\mu^2\right)\right\},$$

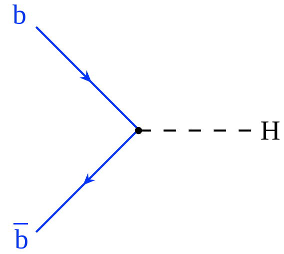


HEAVY QUARKS

EXAMPLE: HIGGS IN BOTTOM QUARK FUSION

MASSIVE (DECOUPLING) SCHEME vs. MASSLESS SCHEME





Massive (decoupling) scheme:

- THE b QUARK IS MASSIVE
- THE b QUARK DECOUPLES FROM QCD EVOLUTION AND THE RUNNING OF α_s: $n_f = 4$
- THE DEPENDENCE ON m_b IS FULLY RETAINED, INCLUDING $O\left(\frac{m_b^2}{m_H^2}\right)$ TERMS
- THERE ARE NO COLLINEAR SINGULARITIES, $\ln \frac{m_b^2}{m_H^2}$ ARE INCLUDED UP TO FINITE ORDER
- b INITIATED PROCESSES START BEYOND LEADING ORDER: $Hb\bar{b}$ STARTS AT $O(\alpha_s^2)$.

Massless scheme:

- b QUARK MASS EFFECTS NEGLECTED
- THE b QUARK IS JUST ANOTHER MASSLESS PARTON, $n_f = 5$
- COLLINEAR LOGS $\lim_{m_b \to 0} \ln \frac{m_b^2}{m_H^2}$ ARE FACTORIZED AND RESUMMED TO ALL ORERS THROUGH QCD EVOLUTION
- b INITIATED PROCESSES START AT LEADING ORDER

ACCURACY REQUIRES MATCHING (ACOT, FONLL, SCET-BASED)

THE FONLL METHOD

(Cacciari, Greco, Nason, 1998)

BASIC IDEA: COMBINE N^iLL MASSLESS RESUMMED & N^jLO MASSIVE FIXED-ORDER (UNRESUMMED) $\Rightarrow$ EXPAND OUT THE RESUMMED RESULT AND REPLACE THE FIRST j ORDERS WITH THEIR MASSIVE COUNTERPARTS

$$F(x,Q^2) = F^{(3)}(x,Q^2) + F^{(4)}(x,Q^2) - F^{(3,0)}(x,Q^2)$$

$$F^{(3)}(x,Q^2) = x\int_x^1 \frac{dy}{y} \sum_{i=g,q,\bar{q}} C_i^{(3)}\left(\frac{x}{y}, \frac{Q^2}{m_h^2}, \alpha_s^{(3)}(Q^2)\right) f_i^{(3)}(y,Q^2)$$

$$F^{(4)}(x,Q^2) = x\int_x^1 \frac{dy}{y} \sum_{i=g,q,\bar{q},h,\bar{h}} C_i^{(4)}\left(\frac{x}{y}, \alpha_s^{(4)}(Q^2)\right) f_i^{(4)}(y,Q^2)$$

ADVANTAGES

- RELIES ON STANDARD FACTORIZATION & DECOUPLING
- THE RESUMMED AND UNRESUMMED ORDERS CAN BE CHOSEN FREELY & INDEPENDENTLY

COMPLICATIONS

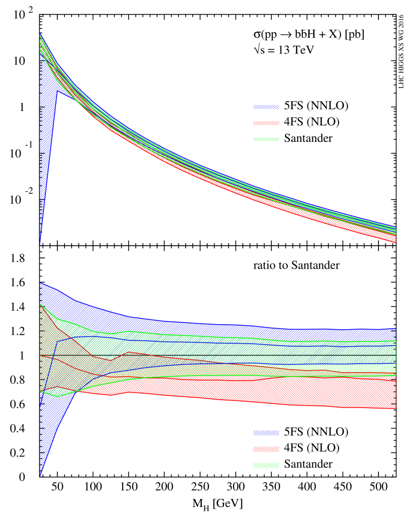
- RESUMMED & FIXED-ORDER CALCULATION ARE PERFORMED IN DIFFERENT RENORMALIZATION & FACTORIZATION SCHEMES: 3F (MASSIVE, DECOUPLING) VS. 4F (MASSLESS)
- MUST MATCH α_s & PDFS

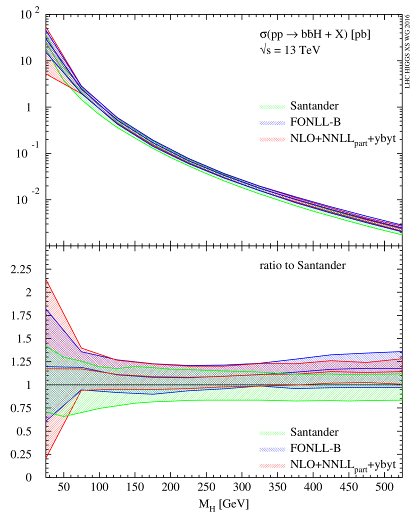
SOLUTION

RE-EXPRESS 3F-SCHEME PDFS & α_s IN TERMS OF THE 4F-SCHEME ONES

IMPACT
PROCESSES

$b\bar{b}H$
4FS, F5S AND MATCHED

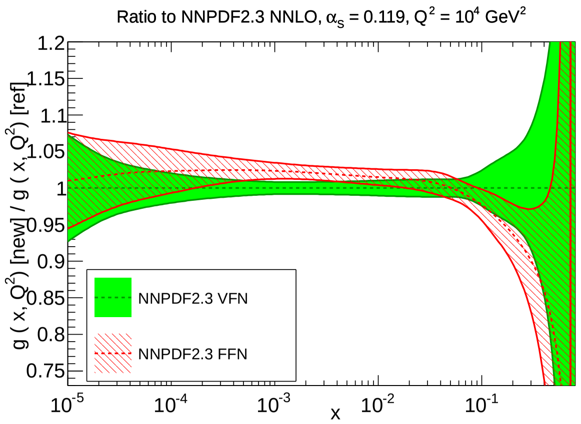


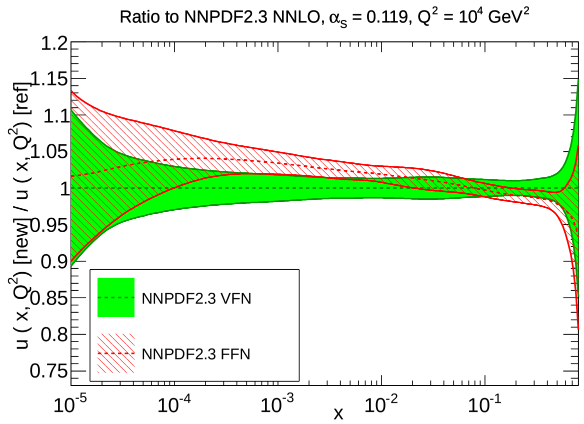


- FOR HIGH-SCALE PROCESSES EFFECT OF LOG RESUMMATION IS SIZABLE
- NAIVE MATCHING OF MASSIVE+MASSLESS COMPUTATIONS (“SANTANDER”) FAILS
- DIFFERENT MATCHING METHODOLOGIES AGREE WELL
  (FONLL: sf, Napoletano, Ubiali; SCET: Bonvini, Tackmann, 2016) HXSWG

PDFS
GLUON & UP

- IN FFN, SMALL x PDFs MUST GROW TO COMPENSATE FOR MISSING LARGE LOGS
- PDFs IN FFN SCHEME DIFFER SIGNIFICANTLY $\Rightarrow$ ACCURACY LOSS





THEORY SUMMARY

- PDF ACCURACY STARTS BEING LIMITED BY THEORY ACCURACY: NNLO NEEDED FOR PERCENT ACCURACY
- PDF THEORY UNCERTAINTIES NOT YET INCLUDED, NEEDED FOR PERCENT ACCURACY
- MANY LHC PROCESSES NOW KNOWN AT NNLO, FEW INCLUDED IN PDF DETERMINATION:
  CURRENTLY ONLY DRELL-YAN RAPIDITY DISTRIBUTION
- RESUMMATION NOT INCLUDED IN PDF DETERMINATION
- MATCHED FIXED-ORDER & RESUMMATION CURRENTLY INCLUDED ONLY FOR HEAVY QUARKS
- MATCHED COMPUTATIONS $\Rightarrow$ REALISTIC DESCRIPTION OF MEASURABLE QUANTITIES

PDFs NOW

CONTEMPORARY PDF TIMELINE

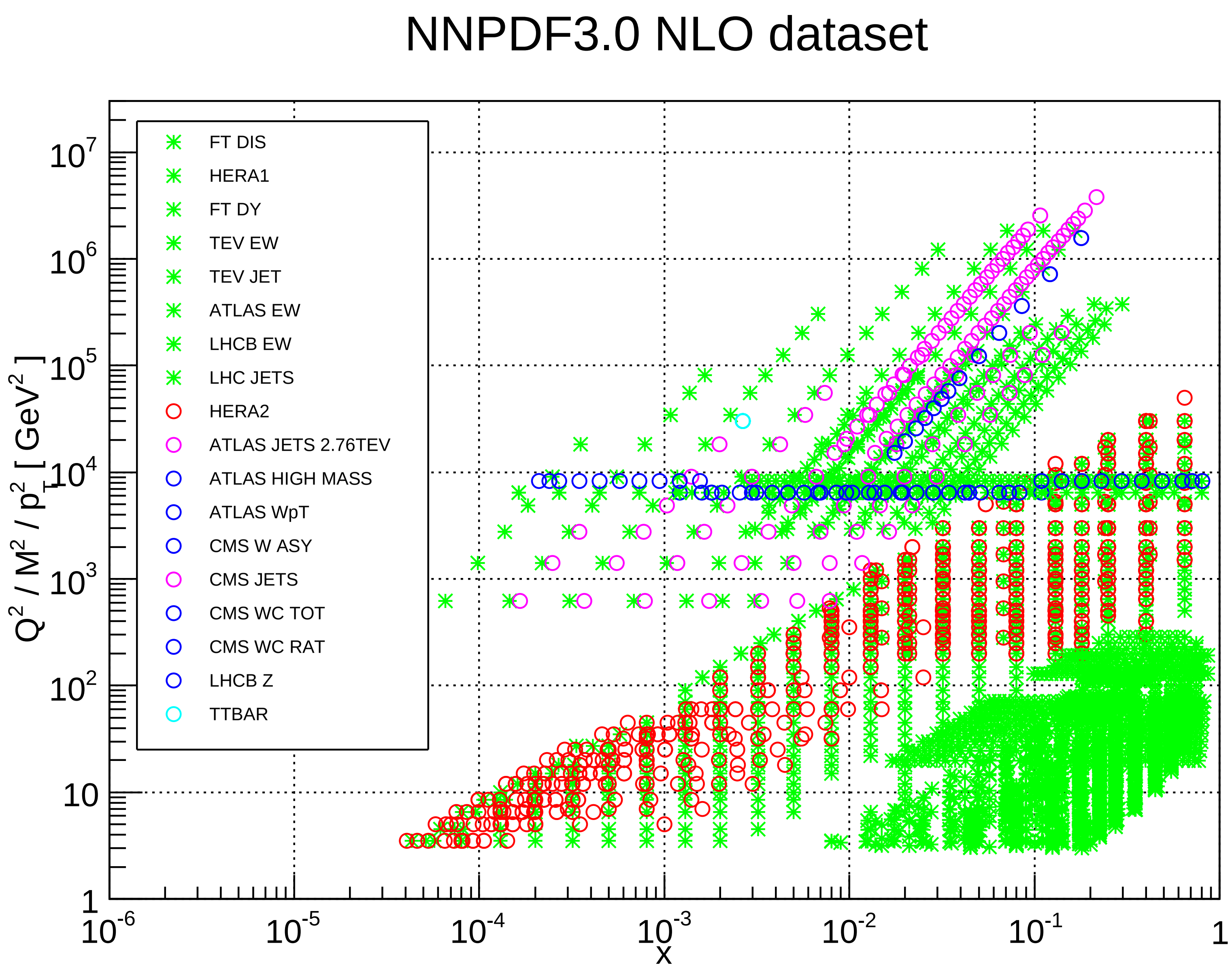
	2008		2009		2010		2011	2012		2013		2014		2015
SET MONTH	CT6.6 (02)	NN1.0 (08)	MSTW (01)	ABKM09 (08)	NN2.0 (02)	CT10(N) (07)	NN2.1(NN) (07)	ABM11 (02)	NN2.3 (07)	CT10(NN) (02)	ABM12 (10)	NN3.0 (10)	MMHT (12)	CT14 (06)
F. T. DIS	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
ZEUS+H1-HI	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
COMB. HI	✗	✗	✗	✗	✔	✗	✔	✗	✔	✗	✔	✔	✗	✗
ZEUS+H1-HII	✗	✗	✗	✗	✗	✗	some	✗	✗	some	✗	✔	✗	✗
HERA JETS	✗	✗	✔	✗	✗	✗	✗	✗	✗	✗	✗	✗	✔	✗
F. T. DY	✔	✗	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
TEV. W+Z	✔	✗	✔	✗	✔	✔	✔	✗	✔	✔		✔	✔	✔
TEV. JETS	✔	✗	✔	✗	✔	✔	✗	✔	✔	✔	✗	✔	✔	✔
LHC W+Z	✗	✗	✗	✗	✗	✗	✗	✗	✔	✗	some	✔	✔	✔
LHC JETS	✗	✗	✗	✗	✗	✗	✗	✗	✔	✗	✗	✔	✔	✔
TOP	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✔	✔	✗	✗
W+C	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✔	✗	✗
W p_T	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✔	✗	✗

- INCREASINGLY WIDE DATASET USED FOR PDF DETERMINATION
- HERAPDF: ONLY HERA STRUCTURE FUNCTION DATA $\Rightarrow$ EXTREME CONSISTENCY
- MANY THEORETICAL AND METHODOLOGICAL IMPROVEMENTS:
  - MSTW, ABKM: ALL NNLO; NNPDF NNLO SINCE 07/11 (2.1), CT SINCE 02/13 (CT10)
  - MSTW, CT ALL MATCHED HEAVY QUARK SCHEMES; NNPDF GM-VFN SINCE 01/11 (2.1)

GLOBAL FITS: THE DATASET IN DETAIL

	NNPDF3.0	MMHT14	CT14
SLAC P,D DIS	✔	✔	✗
BCDMS P,D DIS	✔	✔	✔
NMC P,D DIS	✔	✔	✔
E665 P,D DIS	✗	✔	✗
CDHSW NU-DIS	✗	✗	✔
CCFR NU-DIS	✗	✔	✔
CHORUS NU-DIS	✔	✔	✗
CCFR DIMUON	✗	✔	✔
NuTeV DIMUON	✔	✔	✔
HERA I NC,CC	✔	✔	✔
HERA I CHARM	✔	✔	✔
H1,ZEUS JETS	✗	✔	✗
H1 HERA II	✔	✗	✗
ZEUS HERA II	✔	✗	✗
E605 & E866 FT DY	✔	✔	✔
CDF & D0 W ASYM	✗	✔	✔
CDF & D0 Z RAP	✔	✔	✔
CDF RUN-II JETS	✔	✔	✔
D0 RUN-II JETS	✗	✔	✔
D0 RUN-II W ASYM	✗	✗	✔
ATLAS HIGH-MASS DY	✔	✔	✔
CMS 2D DY	✔	✔	✗
ATLAS W,Z RAP	✔	✔	✔
ATLAS W p_T	✔	✗	✗
CMS W ASY	✔	✔	✔
CMS W +C	✔	✗	✗
LHCB W,Z RAP	✔	✔	✔
ATLAS JETS	✔	✔	✔
CMS JETS	✔	✔	✔
TTBAR TOT XSEC	✔	✔	✗
TOTAL NLO	4276	2996	3248
TOTAL NNLO	4078	2663	3045

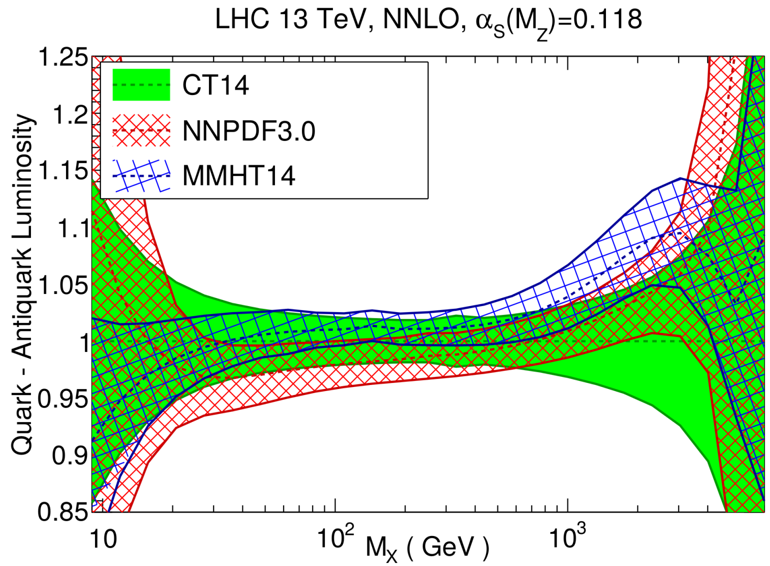
THE NNPDF3.0 DATASET



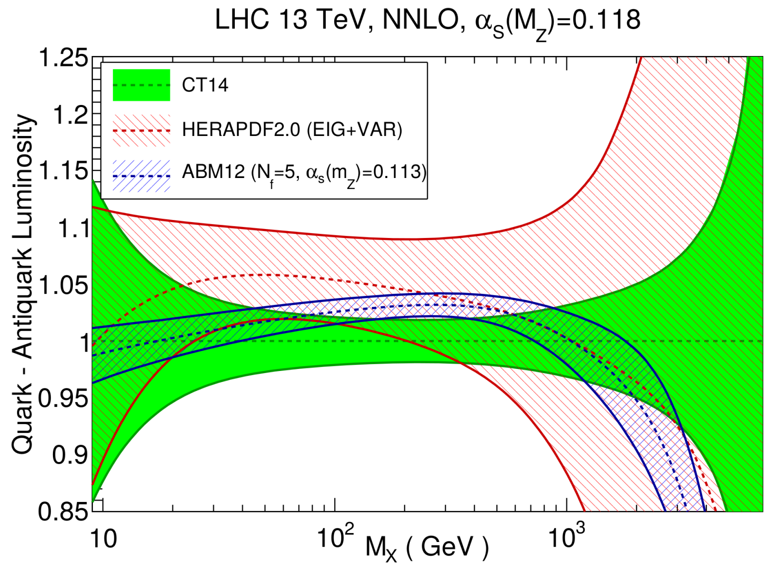
PARTON LUMINOSITIES

QUARK-ANTIQUARK

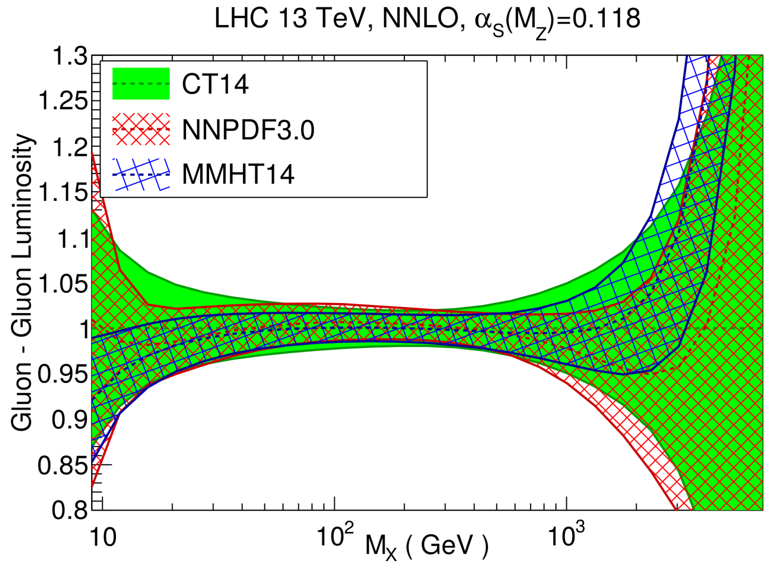
GLOBAL

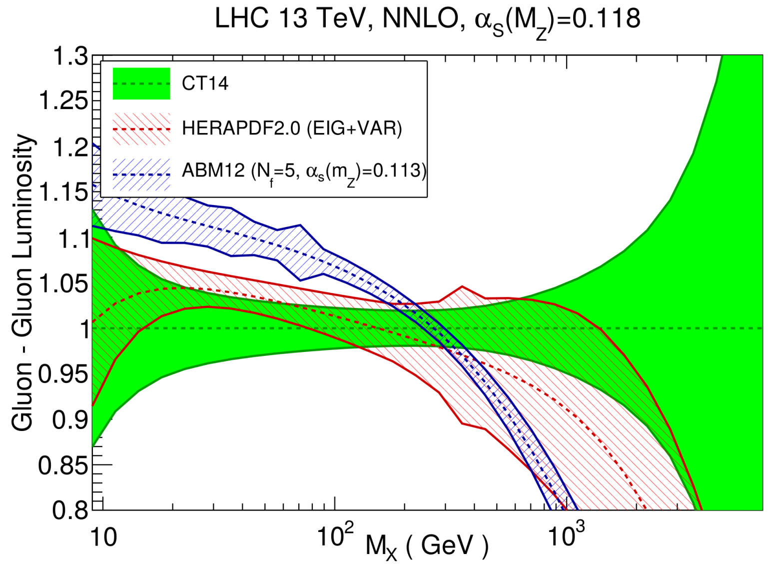


REDUCED



GLUON-GLUON



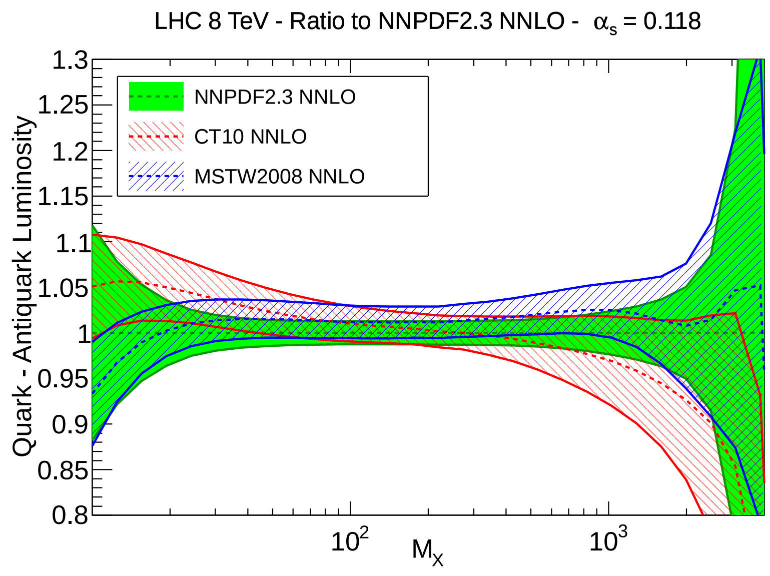


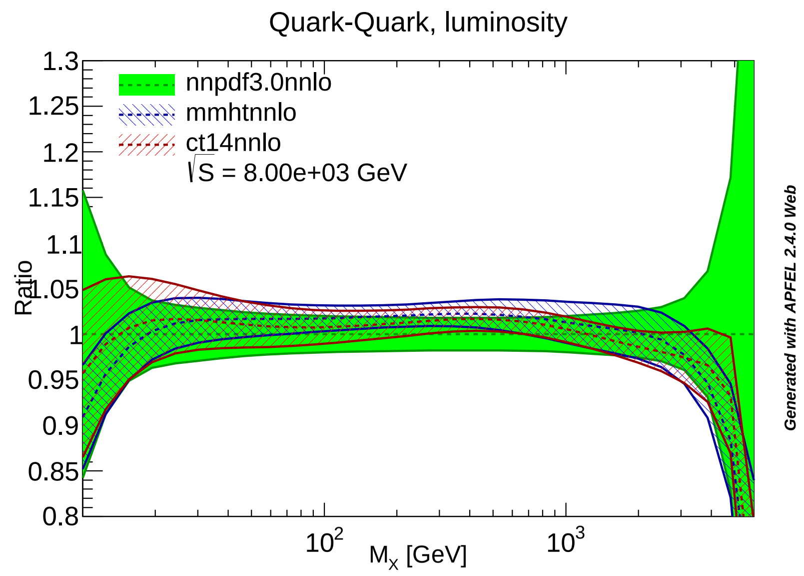
- GLOBAL FITS AGREE WELL
- FITS BASED ON REDUCED DATASET HAVE EITHER LARGE UNCERTAINTIES OR SHOW SIZABLE DEVIATIONS

PARTON LUMINOSITIES

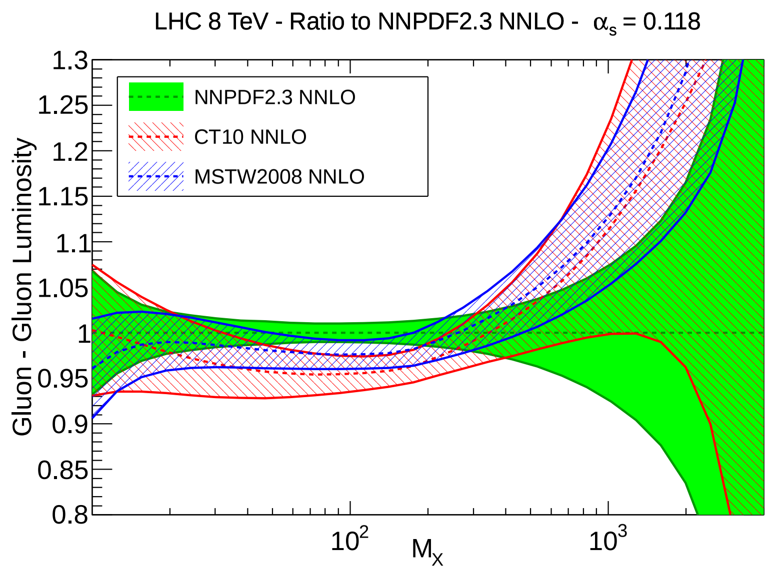
QUARK-QUARK

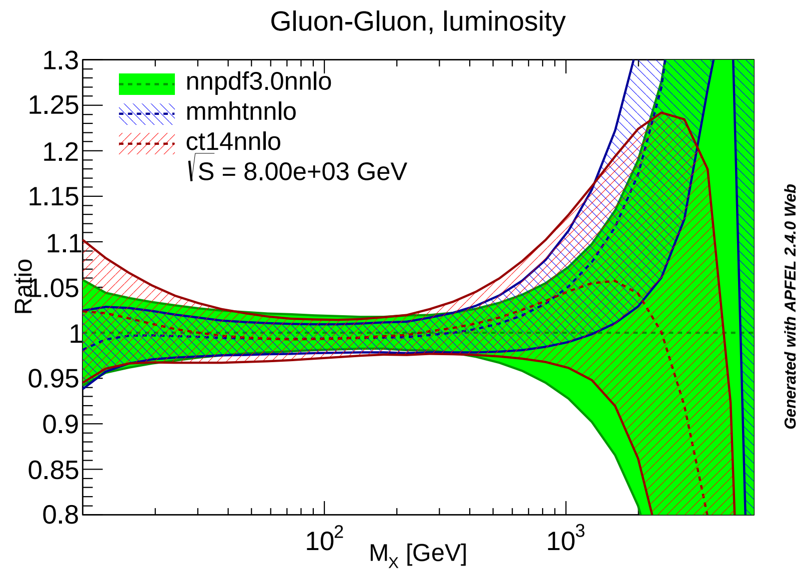
2012 | **2015**





GLUON-GLUON





- LONGSTANDING DISCREPANCY BETWEEN GLUON LUMINOSITIES IS GONE $\Rightarrow$ IMPACT ON HIGGS
- UNCERTAINTIES BLOW UP FOR LIGHT ($\lesssim$ 10 GeV) OR HEAVY ($\gtrsim$ 1 TeV) FINAL STATES $\Rightarrow$ IMPACT ON SEARCHES

PROGRESS

- Q: WHY ARE PDF UNCERTAINTIES ON GLOBAL FITS OF SIMLAR SIZE?
  - SIMILAR DATASETS
  - BUT DIFFERENT PROCEDURES
- A: UNCERTAINTY TUNED TO DATA THROOUGH TOLERANCE (MMHT & CT) OR CLOSURE TESTING (NNPDF)
- Q: WHAT HAS DRIVEN THE IMPROVED AGREEMENT OF GLOBAL FITS
  - SIMILAR DATASETS
  - BUT DIFFERENT PROCEDURES
- A: DATA+METHODOLOGY

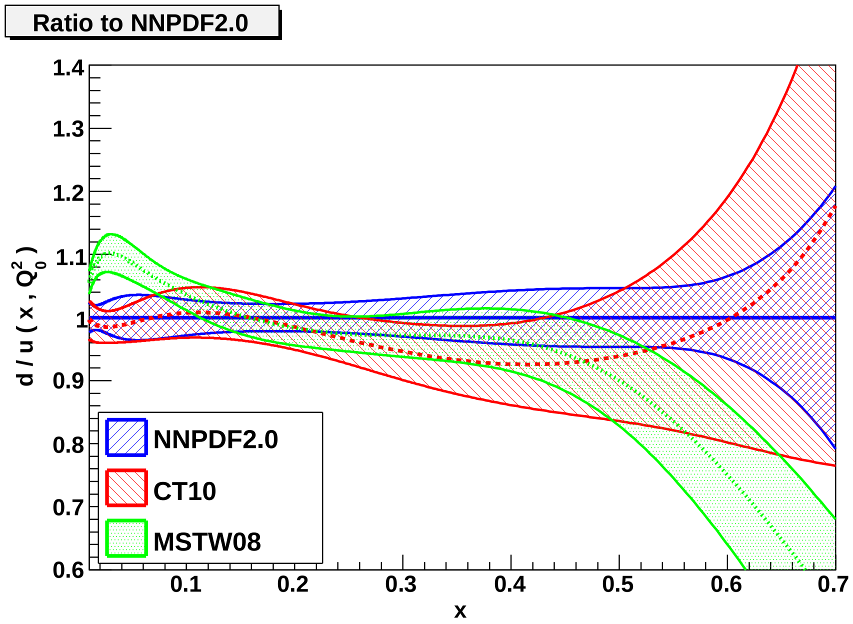
METHODOLOGY

	NNPDF3.0	MMHT14	CT14
No. of fitted PDFs	7	7	6
Parametrization	neural nets	$x^a(1-x)^b\times$ Chebyschev	$x^a(1-x)^b\times$Bernstein
Free parameters	259	37	30-35
Uncertainties	Replicas	Hessian	Hessian
Tolerance	None	Dynamical	Dynamical
Closure test	✔	✗	✗
Reweighting	replicas	eigenvectors	eigenvectors

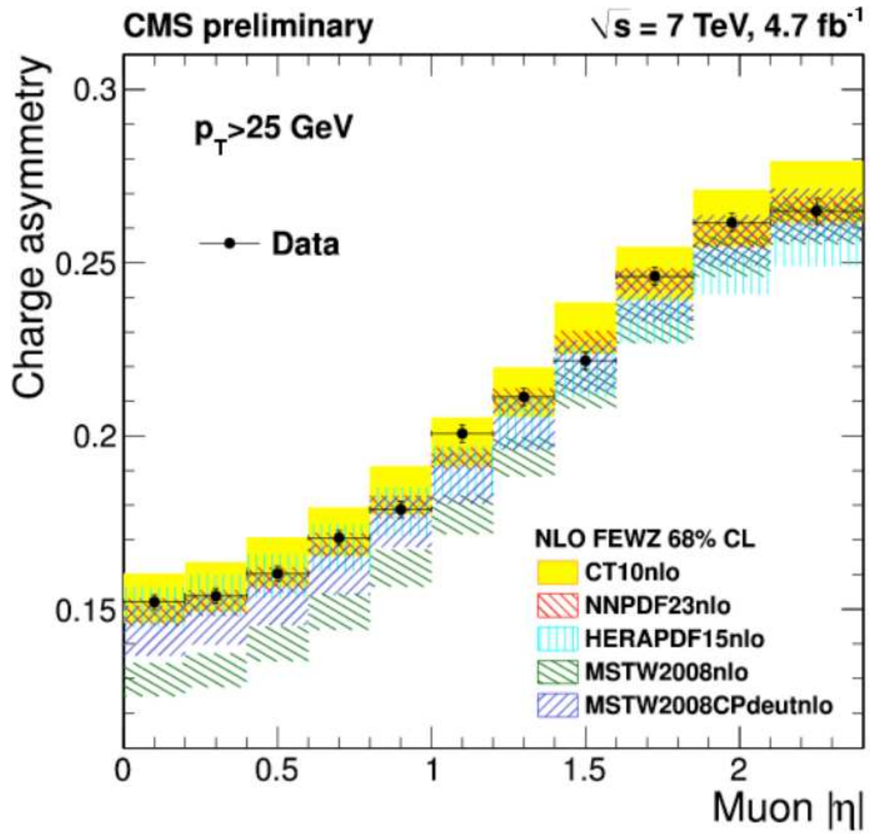
- MMHT, CT10 LARGER # OF PARMS., ORTHOGONAL POLYNOMIALS
- NNPDF CLOSURE TEST

EXAMPLE OF DATA-DRIVEN PROGRESS
MSTW/MMHT: THE d/u RATIO

THE d/u RATIO



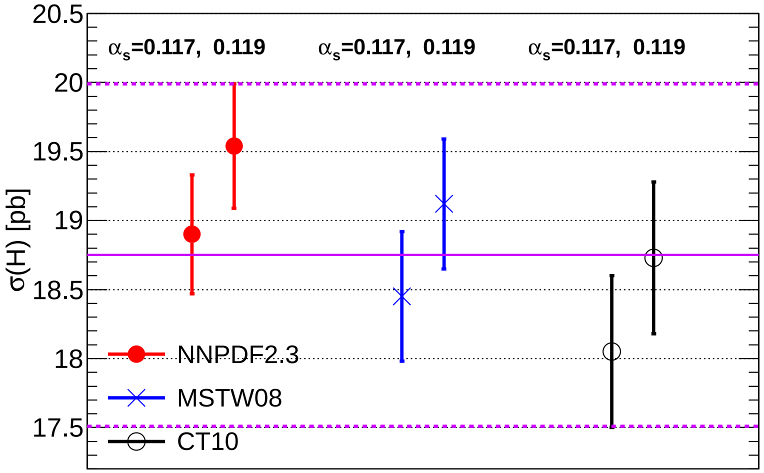
THE CMS W ASYMMETRY



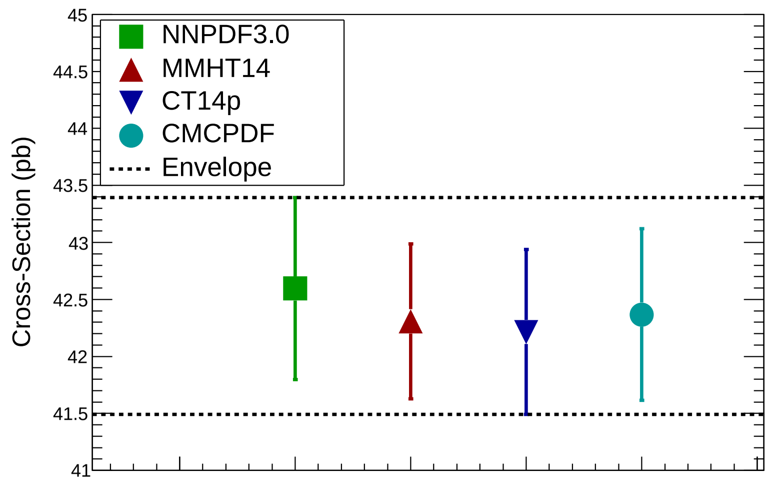
- LONG-STANDING DISCREPANCY IN THE d/u RATIO BETWEEN MSTW AND OTHER GLOBAL FITS
- RESOLVED BY W ASYMMETRY DATA
- EXPLAINED BY INSUFFICIENTLY FLEXIBLE PDF PARAMETRIZATION $\Rightarrow$ FIXED IN MSTW08DEUT/MMHT

THE IMPACT ON HIGGS IN GLUON FUSION

2012



2015



- PDF4LHC PRESCRIPTION 2012: ENVELOPE, PDF UNCERTAINTY $\sim 6\%$
- PDF4LHC PRESCRIPTION 2015: STATISTICAL COMBINATION, PDF UNCERTAINTY $\sim 2\%$

THE NEW PDF4LHC PRESCRIPTION

- PERFORM MONTE CARLO COMBINATION OF UNDERLYING PDF SETS
- SETS ENTERING THE COMBINATION MUST SATISFY COMMON REQUIREMENTS
- DELIVER A SINGLE COMBINED PDF SET THROUGH SUITABLE TOOLS

MONTE CARLO VS. HESSIAN DELIVERY

- MONTE CARLO: A SET OF PDF REPLICAS IS DELIVERED; QUANTITIES COMPUTED FOR EACH REPLICA: CENTRAL VALUE IS THE MEAN, UNCERTAINTY IS STANDARD DEVIATION
- HESSIAN: A CENTRAL SET AND ERROR SETS ARE DELIVERED; CENTRAL SET PROVIDES CENTRAL PREDICTION, UNCERTAINTY IS THE SUM IN QUADRATURE OF ERROR DEVIATIONS

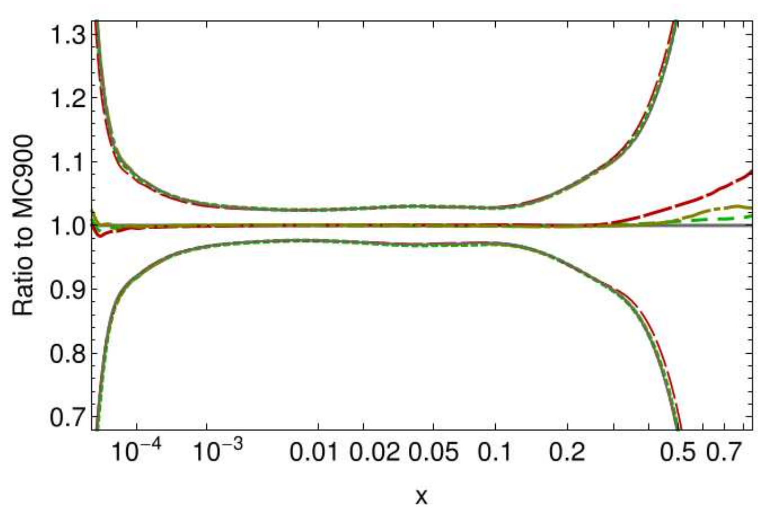
TREATMENT OF α_s

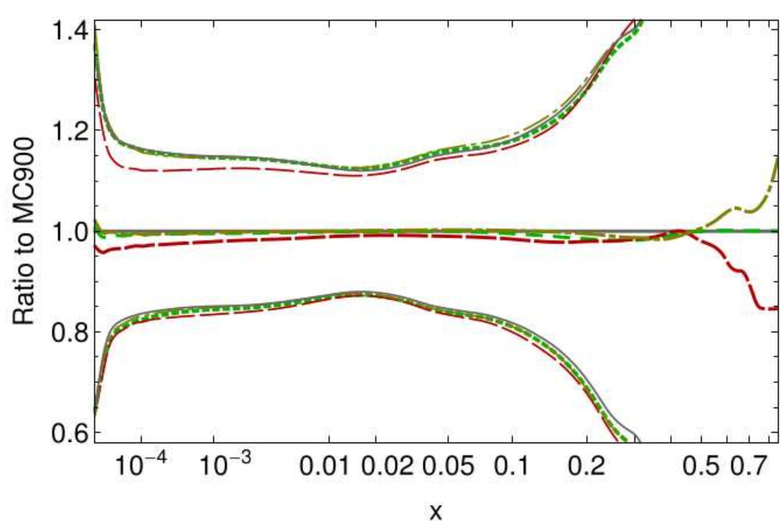
- PDFS ARE DELIVERED FOR EACH VALUE OF α_s
- PDF AND α_s UNCERTAINTIES TO BE KEPT SEPARATE

CURRENT COMBINED SET

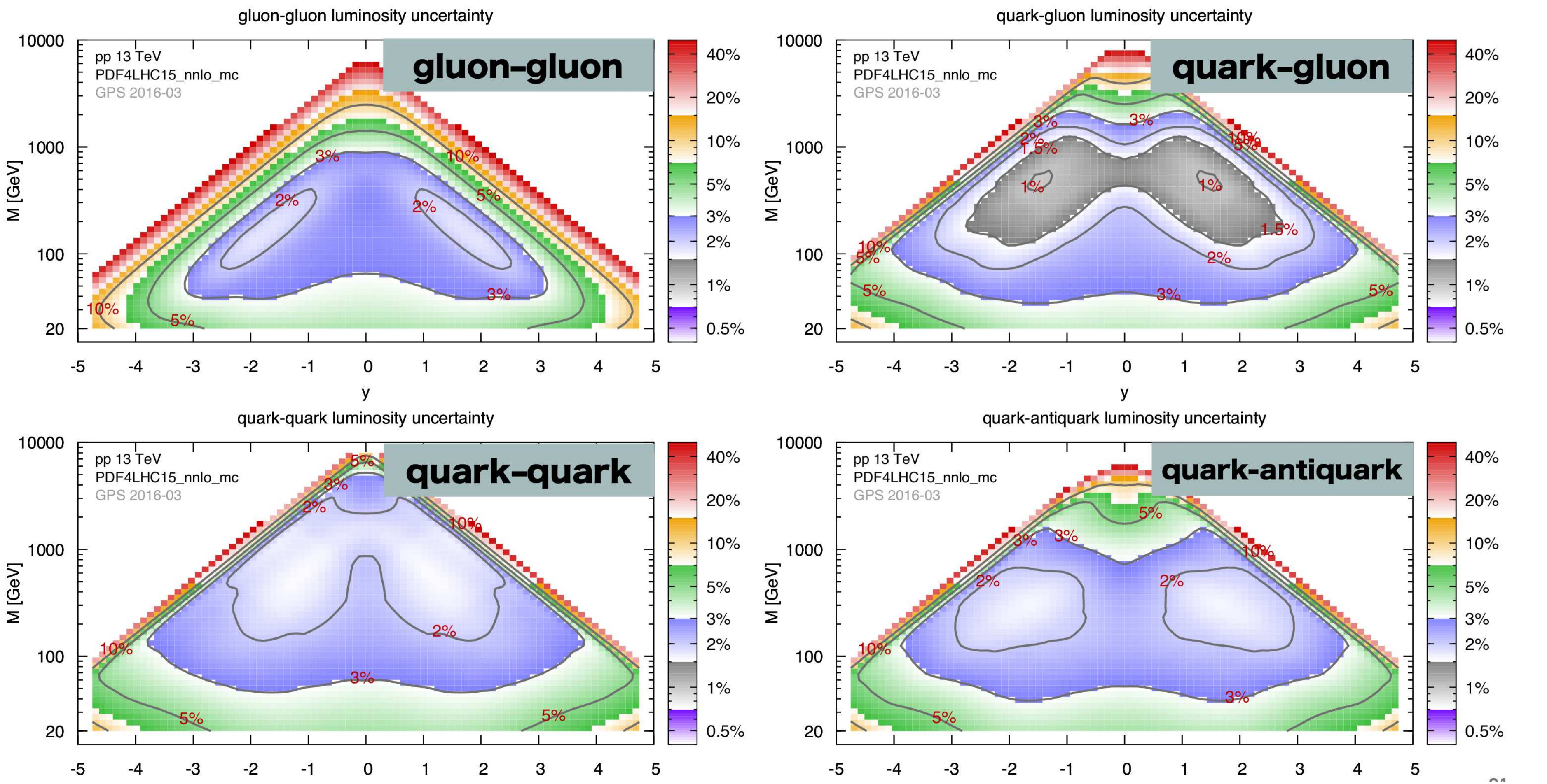
- INCLUDES CT14, MMHT, NNPDF3.0
- 900 REPLICAS (300 FOR EACH SET) ENSURE PRECENTAGE ACCURACY ON ALL QUANTITIES

300, 900, 1800 REPLICAS
(RATIO TO 900)





LUMINOSITY UNCERTAINTIES
THE PDF4LHC15 SET



G.P. Salam, 2016

OUTLOOK

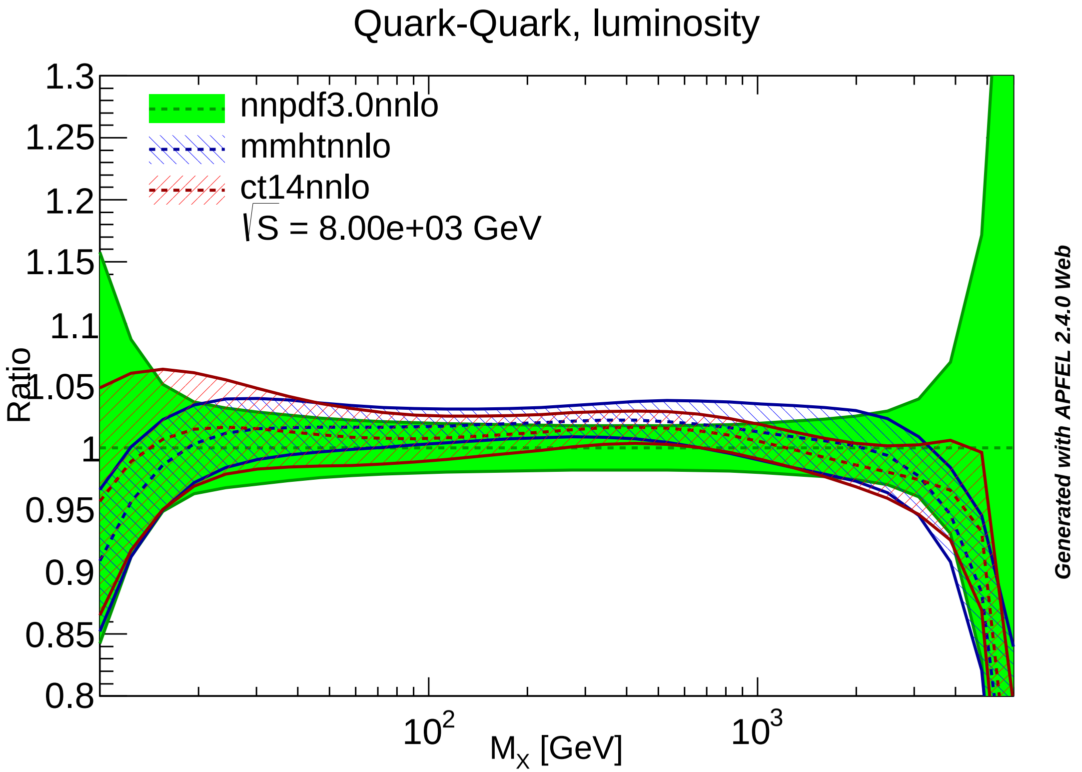
- “PDF uncertainties” as propagated from data and methodology are by and large reliable, of order of 5% in a wide kinematic range
- Lack of experimental information can be a limitation in regions relevant for searches but will be remedied by LHC data
- Theoretical uncertainties must be handled
- LHC data are likely to challenge the proper handling of theoretical uncertainties

PDFs ARE NOT PLUMBING!

EXTRAS

HEAVY QUARK MASSES

REMEMBER THE QUARK LUMINOSITY

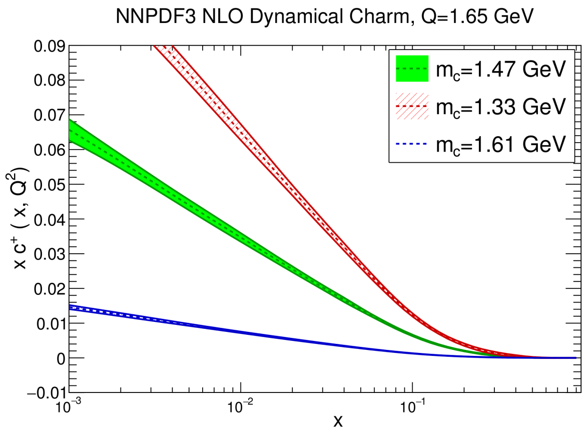


- PDFS DEPEND ON HEAVY QUARK MASSES
- NNPDF3.0 $m_c = 1.275$ Gev; MMHT $m_c = 1.4$ Gev; CT14 $m_c = 1.3$ Gev (POLE) [PDG: 1.47 ± 0.03]
- INDICATIONS THAT DIFFERENCE MAY BE TO m_c VALUE

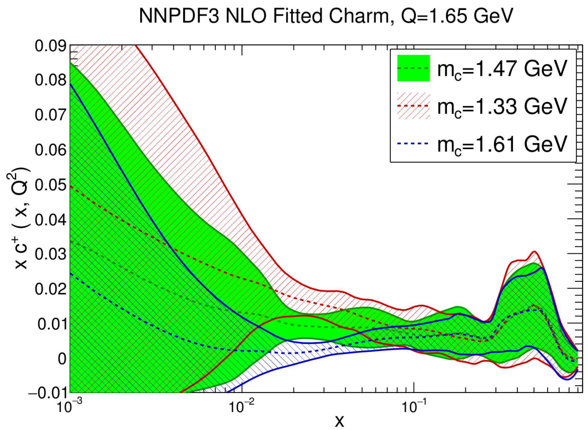
THE CHARM PDF:
STABILITY

LOW SCALE

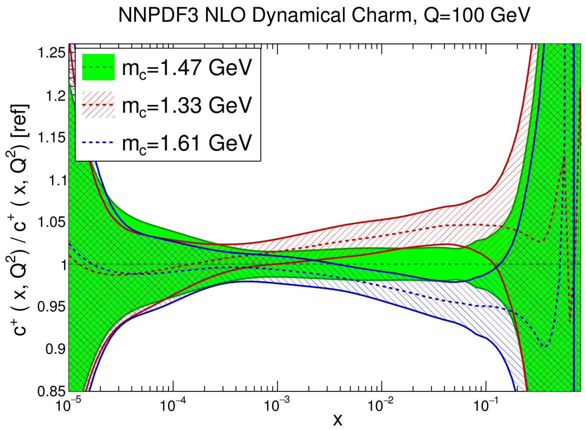
DYNAMICAL

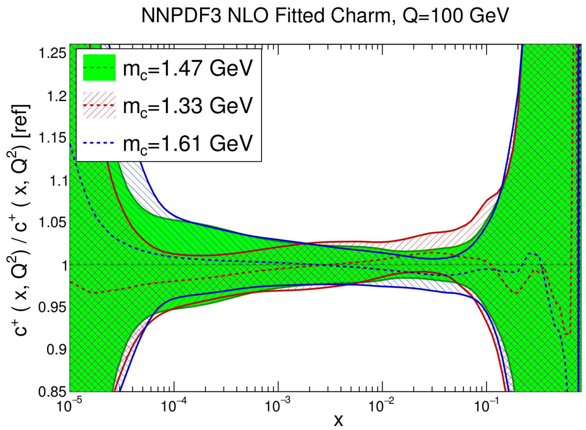


FITTED



HIGH SCALE



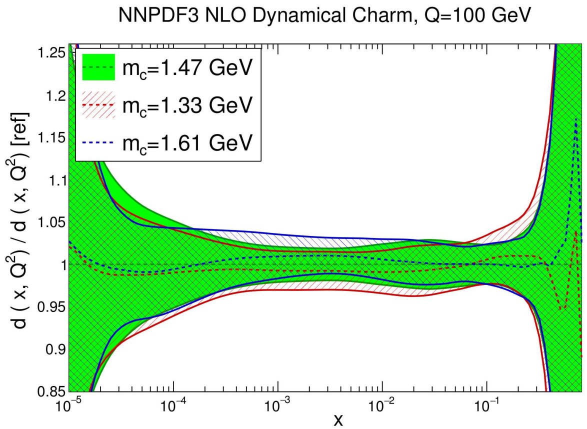


- DYNAMICAL: DEPENDS SIGNIFICANTLY ON THE MASS WHICH SETS THE PHYSICAL THRESHOLD; DEPENDENCE SEEN BOTH AT LOW AND HIGH SCALE;
- FITTED: EXTREMELY STABLE AT ALL SCALES STRUCTURE APPEARS AT LARGE x

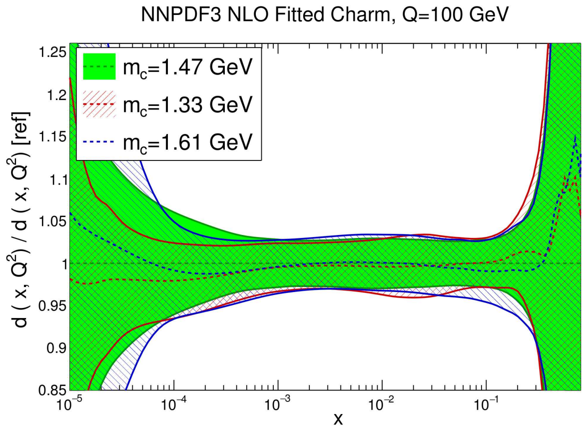
STABILITY: THE LIGHT QUARKS

DOWN

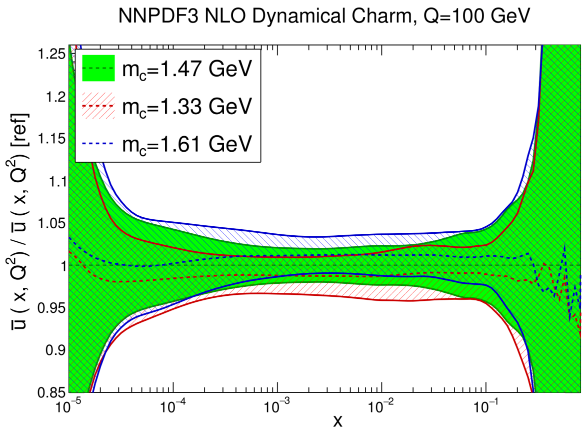
DYNAMICAL

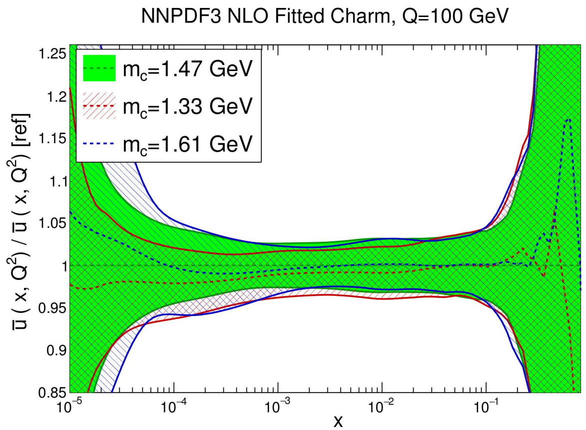


FITTED



ANTIUP





- DYNAMICAL CHARM: LIGHT QUARKS DEPEND (WEAKLY) ON THE MASS WHICH SETS THE PHYSICAL THRESHOLD FOR CHARM, BOTH AT LOW AND HIGH SCALE;
- FITTED CHARM: LIGHT QUARKS BECOME INDEPENDENT OF CHARM MASS AT ALL SCALES
- GLUON LARGELY INSENSITIVE TO CHARM MASS IN ALL CASES