COLLINEAR STRUCTURE

LECTURE 2

Stefano Forte
Università di Milano & INFN

UNIVERSITÀ DEGLI STUDI DI MILANO
DIPARTIMENTO DI FISICA

INFN
Istituto Nazionale di Fisica Nucleare

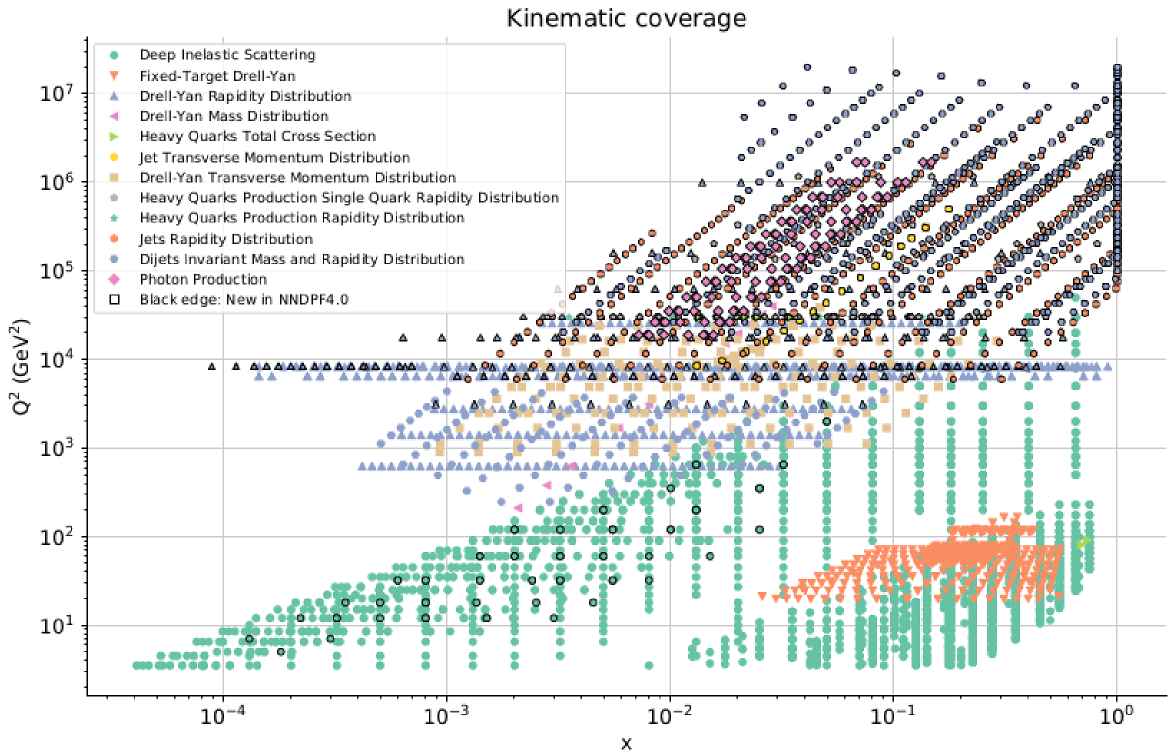
CFNS EIC school

Stony Brook, June 7 2024

PDF DETERMINATION

DATA → PARTON DISTRIBUTIONS

Experimental data in NNPDF4.0



More than 4000 datapoints!

New processes:

- direct photon
- single top
- dijets
- W+jet
- DIS jet

ISSUES AND TASKS:

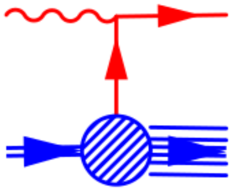
- FROM PHYSICAL OBSERVABLES TO PDFs: SOLVE EVOLUTION EQUATIONS, CONVOLUTE WITH PARTON-LEVEL CROSS-SECTIONS
- DISENTANGLING PDFs: CHOOSE A BASIS OF PDFs ($2N_f$ QUARKS + 1 GLUON) & A SET OF SUITABLE PHYSICAL PROCESSES TO DETERMINE THEM ALL
- PROBABILITY IN THE SPACE OF FUNCTIONS: CHOOSE A STATISTICAL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

DISENTANGLING PDFs

- DIS STRUCTURE FUNCTIONS CC F_1 AND F_3 $\Rightarrow$ FOUR COMBINATIONS, AND NC F_1 TWO MORE $\Rightarrow$ ALL LIGHT FLAVORS
- $W^{\pm}$ AND Z PRODUCTION (INCLUDING DOUBLE DIFFERENTIAL: MASS AND RAPIDITY) $\Rightarrow$ INDEPENDENT COMBINATIONS
- TAGGED CHARM $\Rightarrow$ HEAVY FLAVORS IN INITAL STATE
- DIS+DY $\Rightarrow$ GLUON FROM SCALE DEPENDENCE OR HIGHER ORDERS
  p_T $\Rightarrow$ GLUON AT LO
- JETS $\Rightarrow$ GLUON AT LO

FLAVOR SEPARATION (DIS & DY)
LEADING ORDER PARTON CONTENT

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 $F_1^{W^+} = \bar{u} + d + s + \bar{c}$

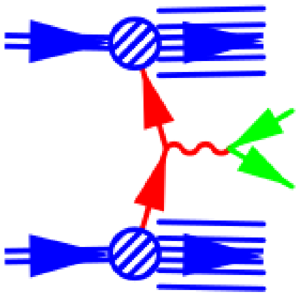
CC $-F_3^{W^+}/2 = \bar{u} - d - s + \bar{c}$

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

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

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

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

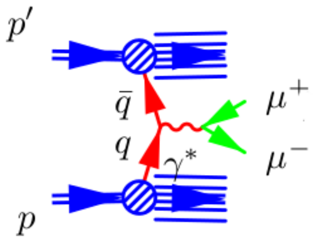
ISOSPIN

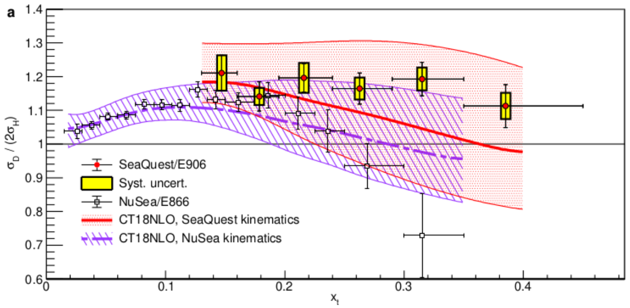
PROTON VS. NEUTRON $u^p = d^n$; $d^p = u^n$; $\bar{u}^p = \bar{d}^n$; $\bar{d}^p = \bar{u}^n$; $s^p = s^n$; $\bar{s}^p = \bar{s}^n$; $\bar{c}^p = \bar{c}^n$

FIXED-TARGET DRELL-YAN (TEVATRON)
QUARK ANTIQUARK SEPARATION

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

DRELL-YAN p/d ASYMMETRY

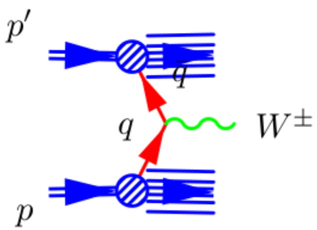


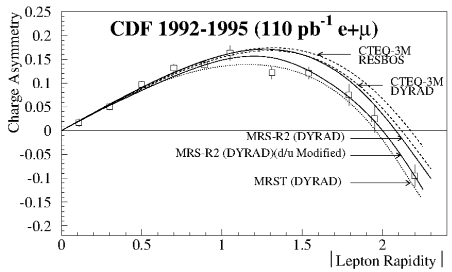


$$\frac{\sigma^{pn}}{\sigma^{pp}} \sim \left.\frac{\frac{4}{9}\, u^p\, \bar{d}^p + \frac{1}{9}\, d^p\, \bar{u}^p}{\frac{4}{9}\, u^p\, \bar{u}^p + \frac{1}{9}\, d^p\, \bar{d}^p}\right|_{\text{large } x} \approx \frac{\bar{d}}{\bar{u}}$$

SeaQuest (2023)

$W^\pm$ ASYMMETRY





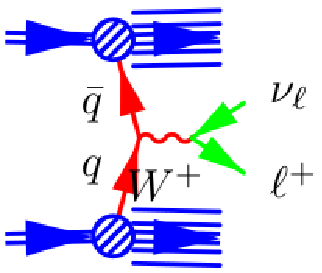
$$\frac{\sigma^{p\bar{p}}_{W+}}{\sigma^{p\bar{p}}_{W-}} = \frac{u^p(x_1)d^p(x_2)+\bar{d}^p(x_1)\bar{u}^p(x_2)}{d^p(x_1)u^p(x_2)+\bar{u}^p(x_1)\bar{d}^p(x_2)} \sim \frac{u^p\, d^p}{d^p\, u^p}$$

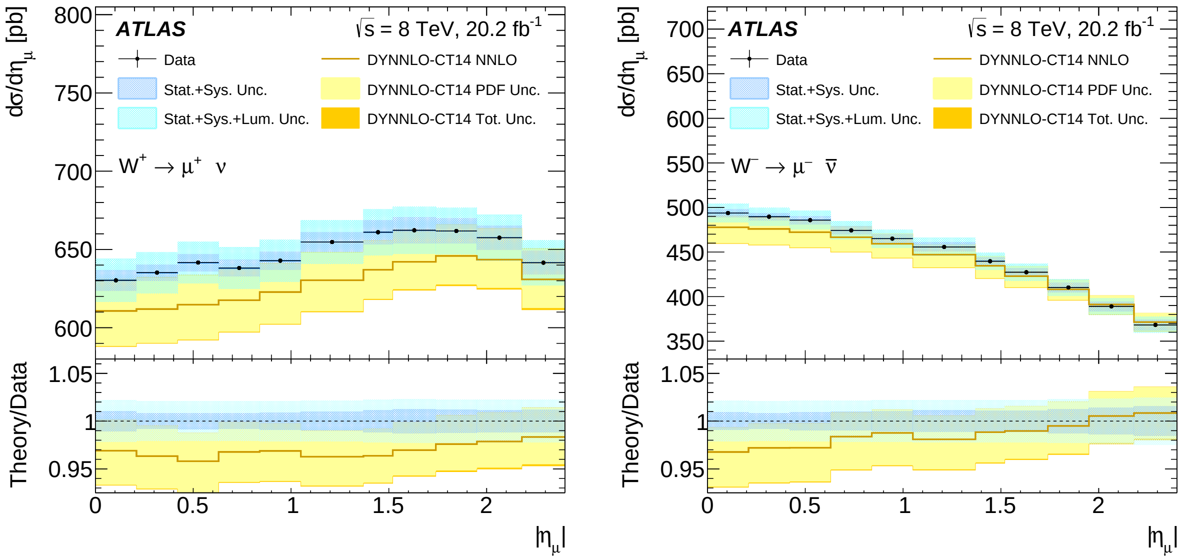
if x_1, x_2 in valence region,
neglecting HQ & Cabibbo suppr.

CDF (1998)

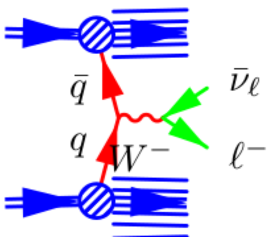
COLLIDER DRELL-YAN (LHC)

$W^\pm$ AND Z PRODUCTION



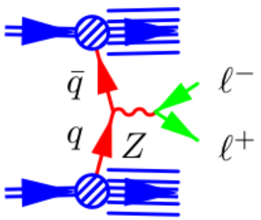


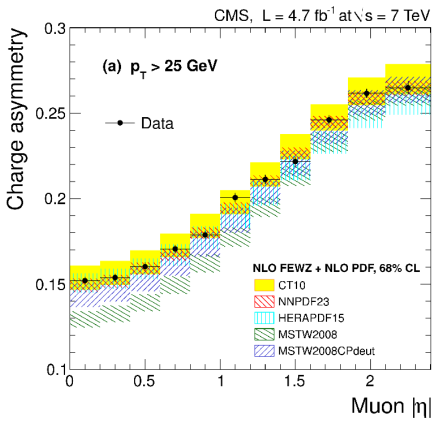
ATLAS (2019)



$$\sigma^{p\bar{p}}_{W+} = u\bar{d} + c\bar{s}; \quad \sigma^{p\bar{p}}_{Z} = u\bar{u} + d\bar{d} + s\bar{s} \Rightarrow \text{STRANGENESS}$$

W MUON ASYMMETRY





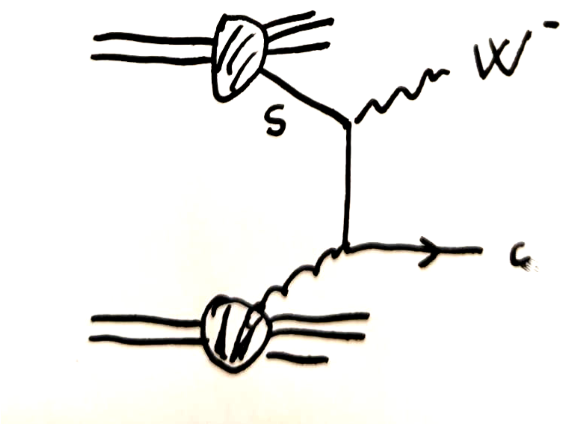
CMS (2013)

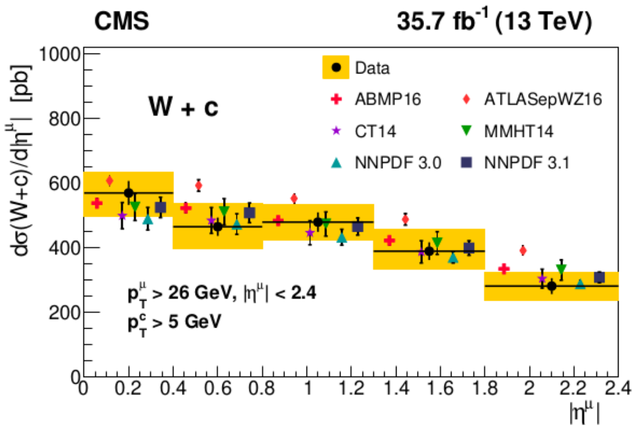
$$\frac{\sigma^{p\bar{p}}_{W+}}{\sigma^{p\bar{p}}_{W-}} = \frac{u(x_1)\bar{d}(x_2) + \bar{d}(x_1)u(x_2)}{d(x_1)\bar{u}(x_2) + \bar{u}(x_1)d(x_2)}$$

VALENCE x (NEGLIGIBLE STRANGE) $\Rightarrow \bar{u} - \bar{d}$ DETERMINED

W, Z + TAGGED JET: HEAVIER FLAVORS

$W + c$

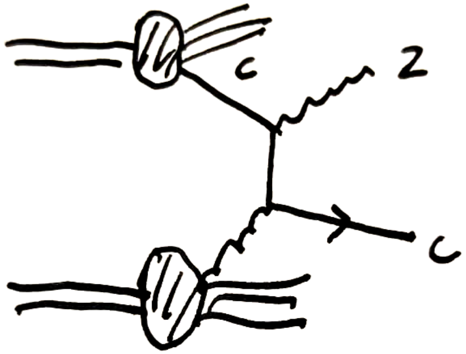


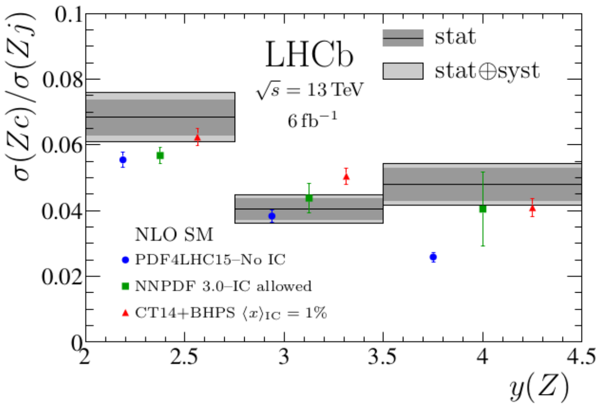


CMS (2018)

- CHARM TAG $\Rightarrow$ STRANGENESS
- $W^\pm$ $\Rightarrow$ STRANGE-ANTISTRANGE SEPARATION

$Z + c$





LHCb (2022)

- CHARM TAG $\Rightarrow$ CHARM
- LARGE RAPIDITY $Rightarrow\ x_1 \gg x_2$ $\Rightarrow$ ACCESS SMALL & LARGE x
- INTRINSIC CHARM!

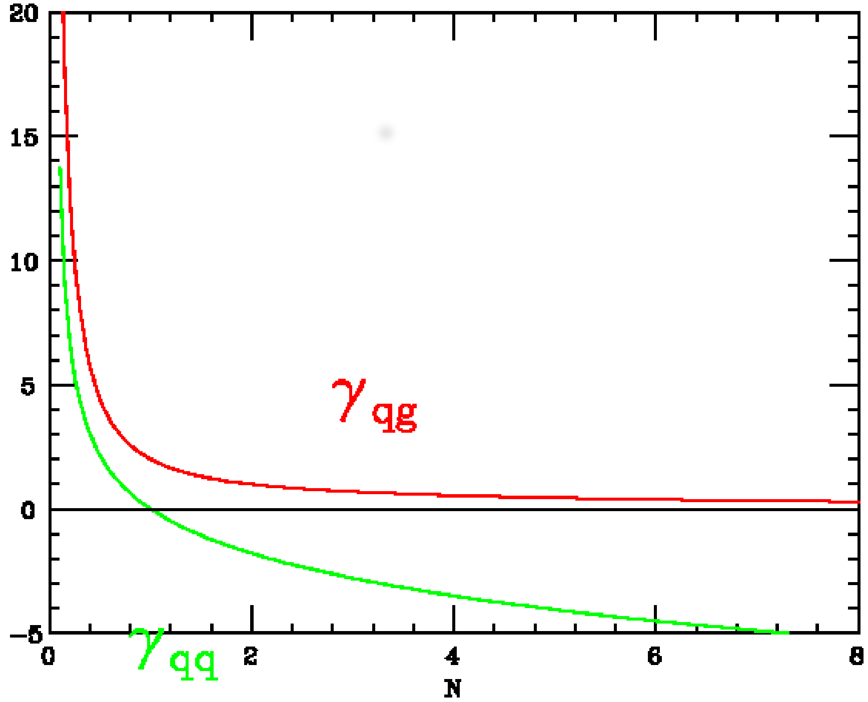
GLUON FROM SCALING VIOLATIONS

THE GLUON HAS ONLY QCD INTERACTIONS!

SCALE DEPENDENCE OF SINGLET STRUCTURE FUNCTION

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

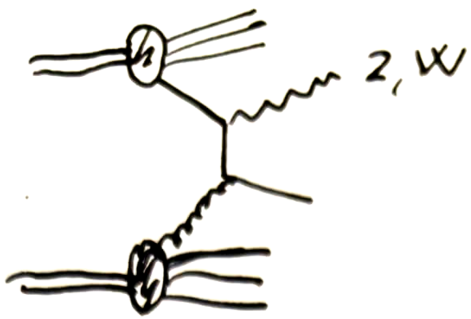
ANOMALOUS DIMENSIONS

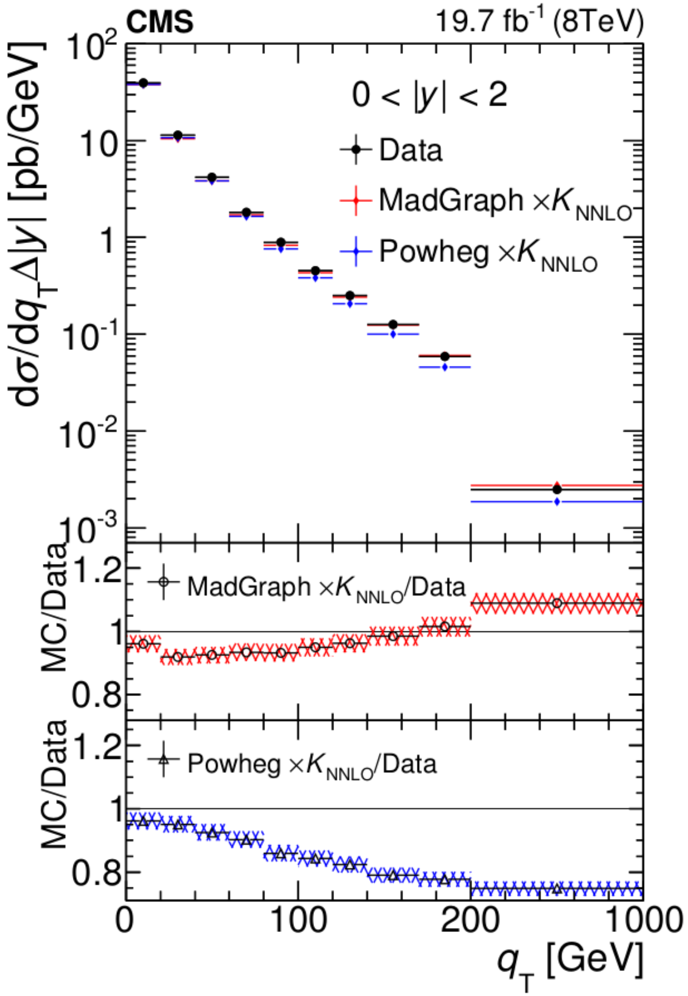


$\Rightarrow$ GLUON AT SMALL x ONLY

GLUON FROM EW P_T DISTRIBUTIONS

Z P_t





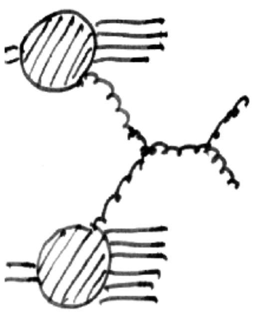
CMS (2015)

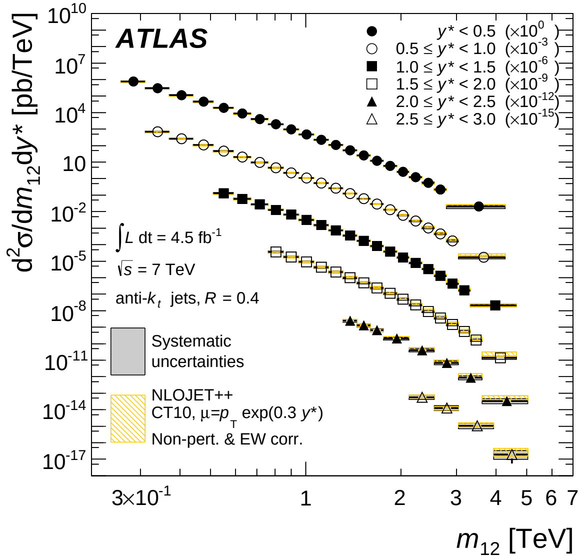
- NEED ONE FINAL-STATE PARTON ⇒ INITIAL-STATE QUARK & GLUON ON SAME FOOTING
- WIDE p_T RANGE ⇒ WIDE x & Q^2 RANGE

GLUON: QCD PROCESSES

JETS

GLUON



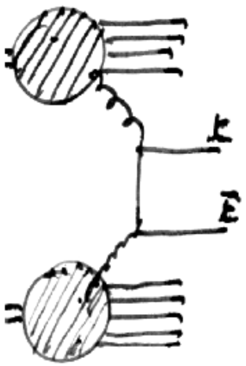


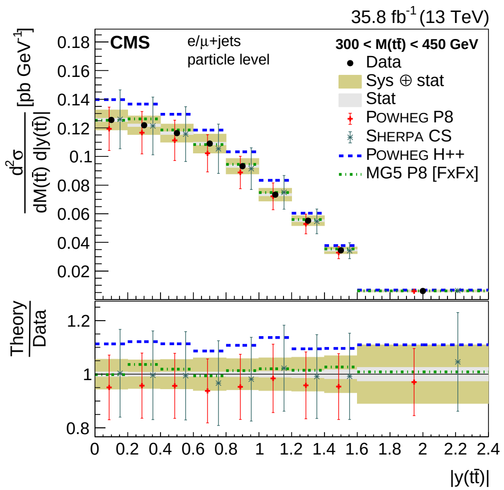
ATLAS (2014)

- ONE-JET/DIJET INCLUSIVE USED TO LARGE x GLUON
- WIDE KINEMATIC REGION AT LHC

TOP

GLUON





CMS (2018)

- WIDE RAPIDITY RANGE $\Rightarrow$ MEDIUM AND LARGE x REGION

PDF DETERMINATION SUMMARY

- DEEP-INELASTIC SCATTERING $\Rightarrow$ CLEAN AND ABUNDANT INFORMATION ON PDFS:
  - HERA $e^{\pm}p$ CC+NC DATA $\Rightarrow$ FOUR INDEPENDENT COMBINATIONS, WIDE KINEMATIC REGION $\Rightarrow$ LIGHT QUARKS AND ANTIQUARKS
  - FIXED-TARGET μp & μd $\Rightarrow$ DIRECT HANDLE ON UP-DOWN SEPARATION
  - HERA $\Rightarrow$ SMALL x GLUON FROM SCALE DEPENDENCE
  - NEUTRINO (ALSO TAGGED c) $\Rightarrow$ STRANGENESS
- DRELL-YAN γ^* ON FIXED p AND d TARGET $\Rightarrow$ UP-DOWN SEPARATION AT LARGE x
- LHC W, Z HIGH AND LOW MASS
  - ANTIUP/ANTDOWN FROM W ASYMMETRY
  - FULL FLAVOR SEPARATION IN WIDE KINEMATIC REGION
  - STRANGENESS $\Leftarrow$ TOTAL CROSS-SECTION AND TAGGED $W + c$ FINAL STATE
  - CHARM $\Leftarrow$ TAGGED $Z + c$ FINAL STATE
  - GLUON $\Leftarrow$ Z TRANSVERSE MOMENTUM DISTRIBUTION
- GLUON AT LHC:
  - TOP $\Rightarrow$ MEDIUM x, FEW DATAPOINTS, HIGH ACCURACY

DATA UNCERTAINTIES:
COVARIANCE MATRIX APPROACH

PREDICTIONS VS. DATA

$$\chi^2 = \sum_{i,j}^{N_{\rm pt}} (T_i - D_i)(\mathrm{cov}^{-1})_{ij}(T_j - D_j)$$

THE COVARIANCE MATRIX

$$\mathrm{cov}_{ij} = \delta_{ij} s_i^2 + \sum_{\alpha=1}^{N_c} \sigma_{i,\alpha}^{(c)} \sigma_{j,\alpha}^{(c)} + \left(\sum_{\alpha=1}^{N_{\mathcal{L}}} \sigma_{i,\alpha}^{(\mathcal{L})} \sigma_{j,\alpha}^{(\mathcal{L})} \right) D_i D_j$$

- D_i: DATA; T_i: PREDICTION
- s_i: UNCORRELATED STATISTICAL UNCERTAINTY FOR i-TH DATAPOINT
- $\sigma_{i,\alpha}^{(c)}$: α-TH CORRELATED ADDITIVE SYSTEMATICS FOR i-TH DATAPOINT
- $\sigma_{i,\alpha}^{(\mathcal{L})}$: α-TH CORRELATED MULTIPLICATIVE SYSTEMATICS FOR i-TH DATAPOINT

DATA UNCERTAINTIES:
NUISANCE PARAMETER APPROACH
THE PARAMETERS

$$\chi^2(\{a\},\{\lambda\}) = \sum_{k=1}^{N_{\rm pt}} \frac{1}{s_k^2}\left(D_k - T_k - \sum_{\alpha=1}^{N_\lambda}\beta_{k,\alpha}\lambda_\alpha\right)^2 + \sum_{\alpha=1}^{N_\lambda}\lambda_\alpha^2$$

SYSTEMATIC UNCERTAINTIES

$$\beta_{i,\alpha} = \sigma_{i,\alpha}^{(c)} \text{ for } \alpha = 1,\ldots,N_c;\ \beta_{i,\alpha} = \sigma_{j,\alpha}^{(\mathcal{L})} D_i \text{ for } \alpha = N_c+1,\ldots,N_{\mathcal{L}}$$

BEST-FIT VALUES

$$\lambda_{0\alpha} = \sum_{i=1}^{N_{\rm pt}} \frac{D_i - T_i}{s_i} \sum_{\delta=1}^{N_\lambda} \mathcal{A}_{\alpha\delta}^{-1} \frac{\beta_{i,\delta}}{s_i}$$

REDUCED COVARIANCE MATRIX

$$\mathcal{A}_{\alpha\beta} = \delta_{\alpha\beta} + \sum_{k=1}^{N_{\rm pt}} \frac{\beta_{k,\alpha}\beta_{k,\beta}}{s_k^2}$$

CONSTRUCTION OF THE COVARIANCE MATRIX: INVERSE

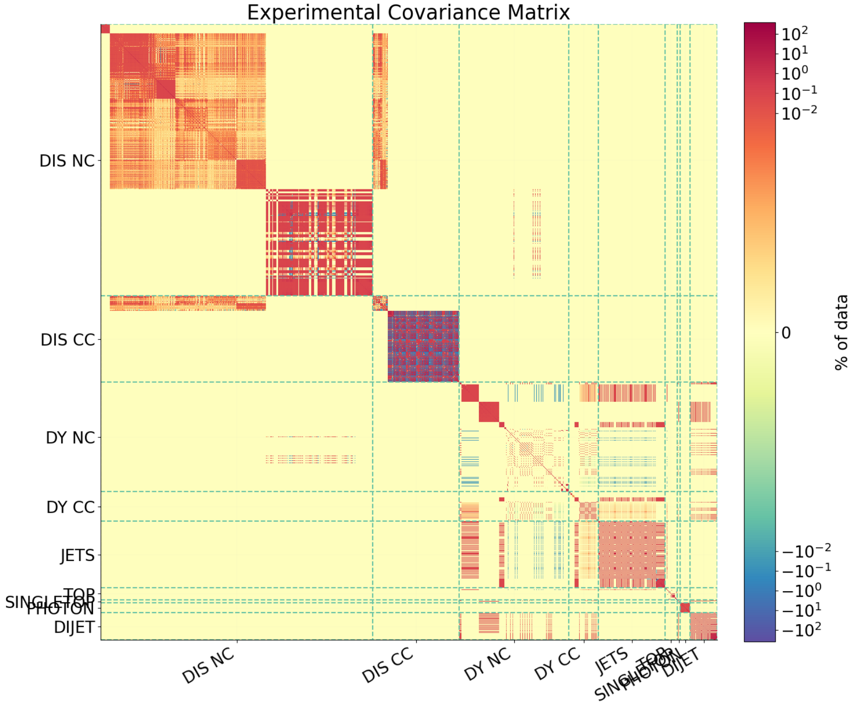
$$({\rm cov})_{ij}^{-1} = \left[\frac{\delta_{ij}}{s_i^2} - \sum_{\alpha,\beta=1}^{N_\lambda} \frac{\beta_{i,\alpha}}{s_i^2}\mathcal{A}_{\alpha\beta}^{-1}\frac{\beta_{j,\beta}}{s_j^2}\right],$$

THE COVARIANCE MATRIX

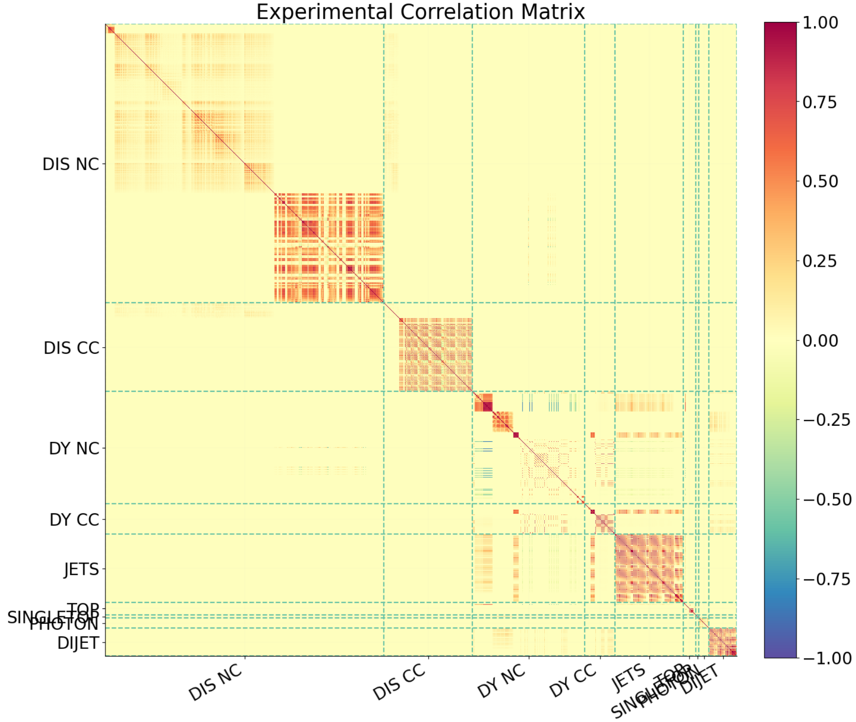
$$({\rm cov})_{ij} \equiv s_i^2\delta_{ij} + \sum_{\alpha=1}^{N_\lambda}\beta_{i,\alpha}\beta_{j,\alpha}$$

A LOOK AT THE EXPERIMENTAL COVARIANCE MATRIX

COVARIANCE



CORRELATION



HESSIAN UNCERTAINTIES

- CHOOSE A FIXED FUNCTIONAL FORM $f(x, Q^2; \vec{p})$, p_i, $i = 1, \ldots, N_{\text{par}}$ PARAMETERS
  - 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_-, $g \Rightarrow xf_i(x, Q_0^2) = Ax^\alpha (1-x)^\beta \left(1 + \sum_{i=1}^4 a_i T_i(y(x))\right)$; 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 $xg(x, Q_0^2) = Ax^\alpha (1-x)^\beta \left(1 + \sum_{i=1}^2 a_i T_i(y(x))\right) + A' x T \alpha' (1-x)^{\beta'}$
    * SEA ASYMMETRY $x\Delta(x, Q_0^2) = Ax^\alpha (1-x)^\beta (1 + \gamma x + \epsilon x^2)$
    * STRANGENESS ASYMMETRY $x\Delta(x, Q_0^2) = Ax^\alpha (1-x)^\beta (1 - x/x_0)$
    * 41 PARAMETERS, 4 FIXED BY SUM RULES
    * 12 PARMS FIXED AT BEST FIT, REMAINING 25 USED FOR COVARIANCE MATRIX $\Rightarrow$ INCREASED TO 30 IN MSHT 2019
- EVOLVE TO DESIRED SCALE & COMPUTE PHYSICAL OBSERVABLES
- MINIMUM OF $\chi^2(\vec{p})$ BEST-FIT VALUES OF PARAMETERS $p_i^{(0)}$
- COVARIANCE MATRIX IN PARM. SPACE $\sigma_{ij} = \partial_i \partial_j \chi^2(\vec{p})$

HESSIAN UNCERTAINTY PROPOAGATION

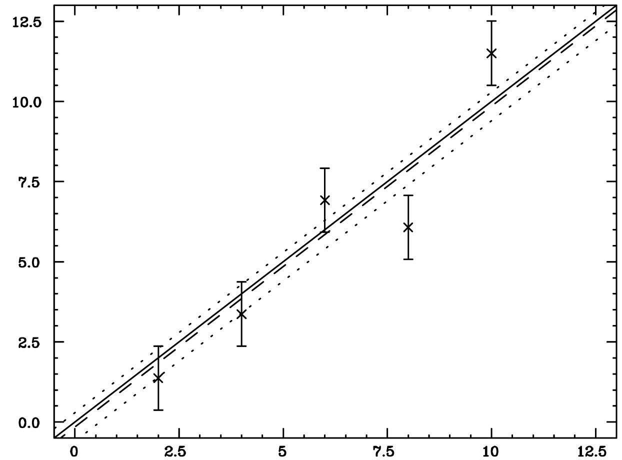
ONE SIGMA PARM. RANGE $\Rightarrow \Delta\chi^2 = 1$ (ERROR PROPAGATION)

"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 INTERCEPT: $k = \langle\Delta_i\rangle$
- UNCERTAINTY ON IT: $\sigma_k = \frac{\sigma_\Delta}{N_{\rm dat}}$
- IF $\Delta k = \sigma_k$, THEN $\Delta\chi^2 = 1$



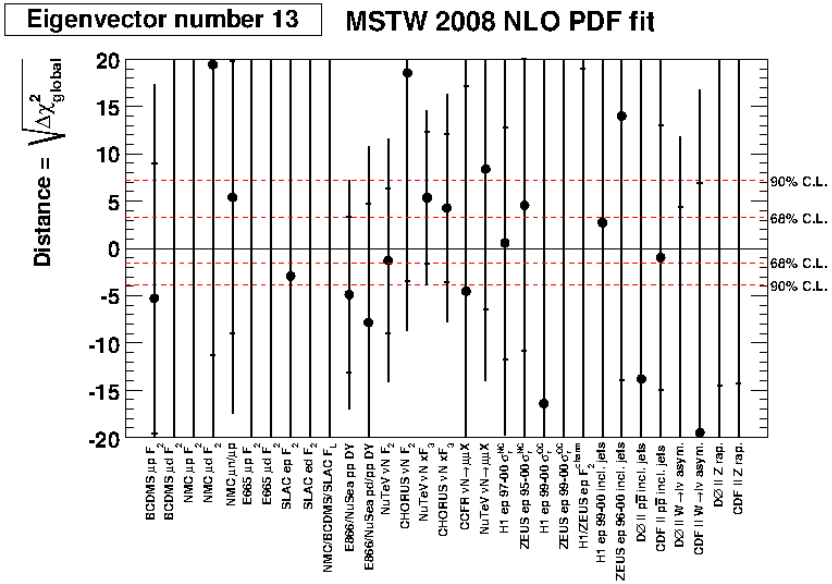
TOLERANCE

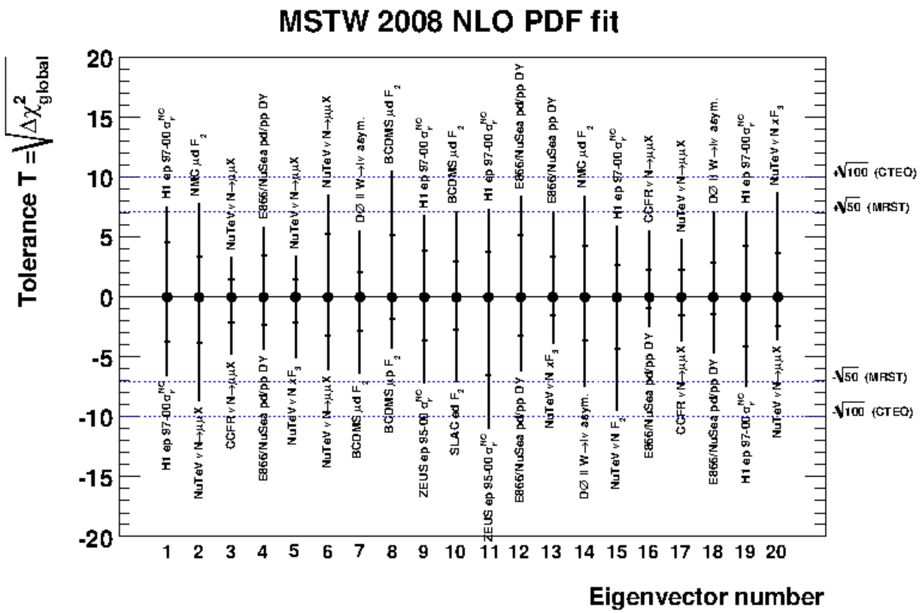
- IN GLOBAL HESSIAN FITS, UNCERTAINTITES OBTAINED BY $\Delta\chi^2 = 1$ UNREALISTICALLY SMALL

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

GLOBAL MSTW TOLERANCE

MSTW TOLERANCE PLOT FOR 13TH EIGENVEC.





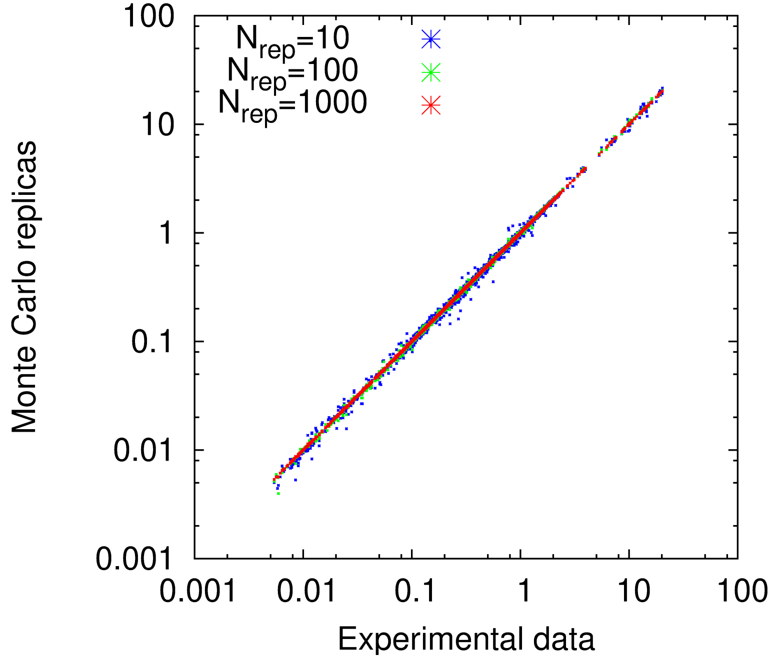
- (MSTW/MMHT) FOR EACH EIGENVECTOR IN PARAMETER SPACE DETERMINE CONFIDENCE LIMIT FOR THE DISTRIBUTION OF BEST-FITS OF EACH EXPERIMENT

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

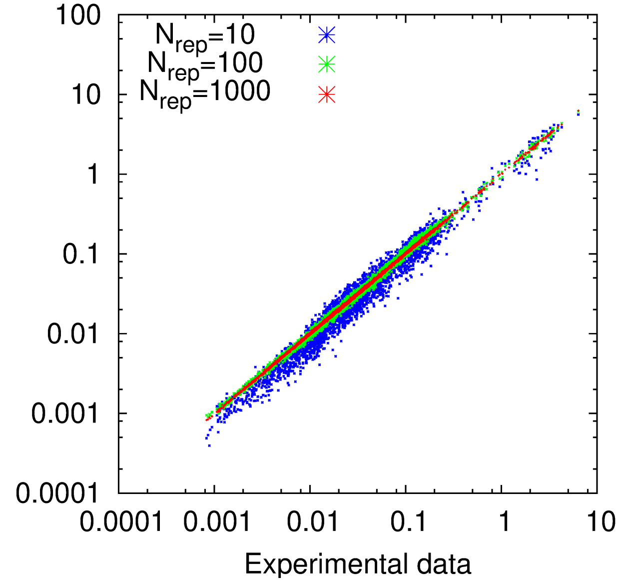
MONTE CARLO UNCERTAINTIES

- DATA+UNCERTAINTIES ⇒ PROBABILITY $P(\vec{z})$ (MULTIGAUSSIAN); z_i, $i = 1, \ldots, N_{\rm dat}$
- MEAN $\langle \vec{z} \rangle = \int d^d z X(\vec{z}) P(\vec{z})$; COVARIANCE $\sigma_{ij} = \langle (z_i - \langle z_i \rangle)(z_j - \langle z_j \rangle) \rangle$
- GENERATE REPLICAS OF ORIGINAL DATA $\vec{z}^{(k)}$, $k = 1, \ldots, N_{\rm rep}$
- MEAN $\langle \vec{z} \rangle = \frac{1}{M_{\rm rep}} \sum_1^{N_{\rm rep}} \vec{z}^{(k)}$

replica averages
vs. central values

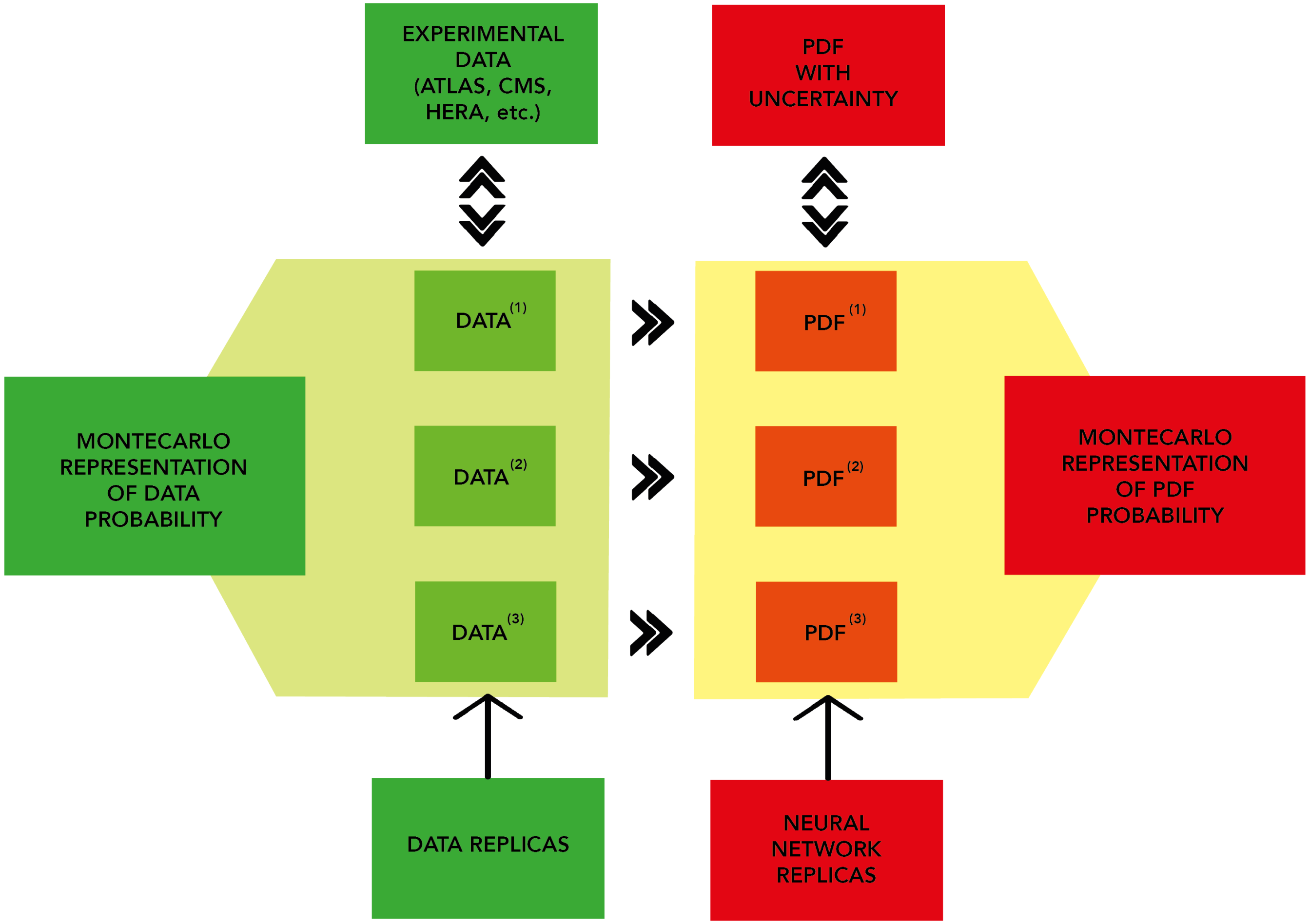


replica standard dev.
vs. uncertainties



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

MONTE CARLO UNCERTAINTY PROPAGATION



- DETERMINE BEST-FIT PDF REPLCA $f^i(x, Q_0^)$ FOR EACH DATA REPLICA
  $\Rightarrow$ DOES NOT HAVE TO BE MIN. OF χ^2
- MC REPRESENTATION OF PROBABILITY DISTRIBUTION IN PDF SPACE

MONTE CARLO UNCERTAINTIES

IMPORTANCE SAMPLING

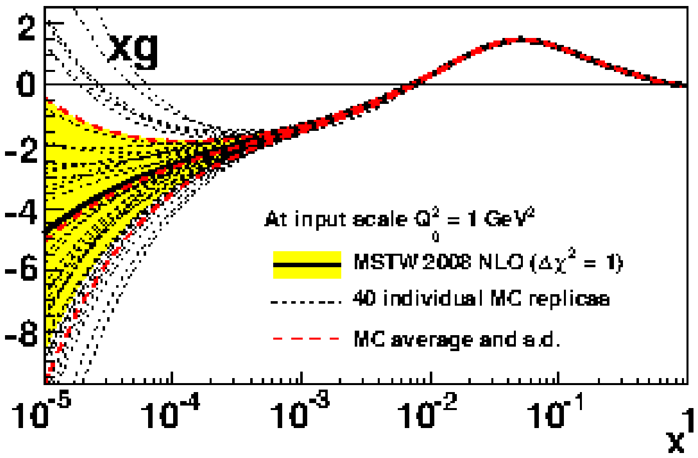
- PROBABILITY DISTRIBUTION SAMPLED DIRECTLY
  $\Rightarrow$ ALL INSTANCES EQUALLY WEIGHTED $\langle f \rangle = \frac{1}{N} \sum_{I=1}^{N} f_i$
- CONTRAST TO A MODEL DEPENDING ON PARAMETERS θ_i WITH KNOWN PROBABILITY $p(\theta_i)$:
- $\langle f \rangle = \frac{1}{N} \sum_{i=1}^{N} f(\theta_i) \, p(\theta_i) = \frac{1}{N} \sum_{i=1}^{N} f\left(\theta_i^p\right)$;
  θ_i^p sampled with probability $p(\theta_i)$
  - IF $p(\theta_i)$ SMALL FOR SOME $\theta_i \Rightarrow$ INEFFICIENCY
  - REDEFINE $\langle f \rangle = \frac{1}{N} \sum_{I=1}^{N} f(\theta_i) \frac{p(\theta_i)}{q(\theta_i)} q(\theta_i) = \frac{1}{N} \sum_{I=1}^{N} f\left(\theta_i^q\right) \frac{p(\theta_i^q)}{q(\theta_i^q)}$;
    θ_i^q sampled with probability $q(\theta_i)$
  - OPTIMIZE CHOICE OF $q(\theta_i)$
- EQUAL WEIGHTING $\Rightarrow$ OPTIMAL CHOICE

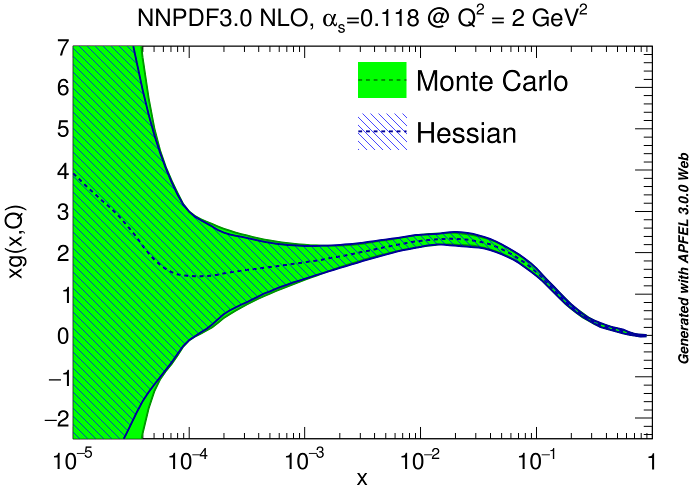
WHY IT IS IMPORTANT

- SPACE OF FUNCTIONS HUGE
  5 BINS FOR 10 PTS$\times$ 7 FCTNS $\rightarrow 5^{70} \sim 10^{49}$ BINS
- BUT OBSERVABLES CORRELATED $\Rightarrow$ DATA TELL US WHICH BINS ARE POPULATED

MC ⇔ HESSIAN

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

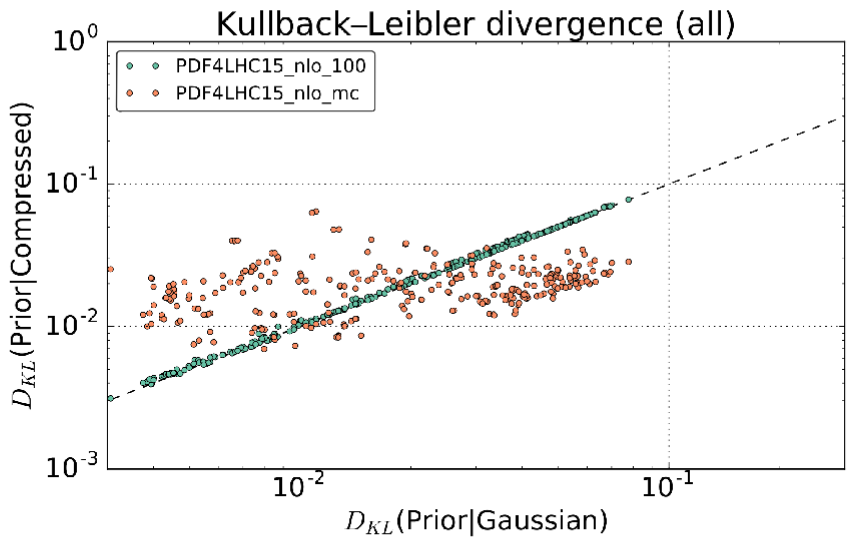




- TO CONVERT MONTE CARLO INTO HESSIAN, SAMPLE REPLICAS $f_i(x)$ AT A DISCRETE SET OF POINTS & CONSTRUCT THE ENSUING COVARIANCE MATRIX
- EIGENVECTORS OF THE COVARIANCE MATRIX ⇒ A BASIS IN VECTOR SPACE SPANNED BY REPLICAS BY SINGULAR-VALUE DECOMPOSITION
- NUMBER OF DOMINANT EIGENVECTORS ∼ TO NUMBER OF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

ARE UNCERTAINTIES GAUSSIAN?

- REPLICA HISTOGRAM i-TH DATAPOINT z_i FROM MC $\Rightarrow$ CONTINUOUS DISTRIBUTION WITH KDE
  - POINT $\Rightarrow$ KERNEL: $P(z) = \frac{1}{N_{\rm rep}} \sum_{k=1}^{N_{\rm rep}} K\left(z - z_i\right);$
  - Gaussian kernel $K(z - z_i) \equiv \frac{1}{h\sqrt{2\pi}} \exp\left(-\frac{(z-z_i)^2}{h}\right)$
  - Silverman bandwidth $h = \sigma_i \left(\frac{4}{3N_{\rm rep}}\right)^{\frac{1}{5}}$ $\Rightarrow$ MINIMIZES DIFFERENCE TO GAUSSIAN
- DEFINE KULLBACK-LEIBLER DIVERGENCE $D_{\rm KL} = \int_{-\infty}^{\infty} P(x) \ln \frac{P(x)}{Q(x)}\, dx$ BETWEEN A PRIOR P AND ITS REPRESENTATION Q
- COMPUTE $D_{\rm KL}$ MC PRIOR VS REPRESENTATION & MC PRIOR VS GAUSSIAN
- REPRESENTATIONS SHOWN: MULTIGAUSSIAN OR MC COMPRESSION (OPTIMAL MC WITH SAME NUMBER OF REPLICAS)



$D_{\rm KL}$ TO GAUSSIAN SMALL!: $D_{KL} \sim$ PERCENTAGE DIFFERENCE

CAN WE TRUST UNCERTAINTIES?
CLOSURE TESTS

- ASSUME UNDERLYING "TRUTH" PDF (SAY A RANDOM PDF REPLICA)
- GENERATE DATA ACCORDING TO STATISTICAL AND CORRELATED SYSTEMATICS (SAY FOR NNPDF4.0 DATASET)
- DETERMINE PDFs & COMPARED TO "TRUTH" BASED ON INDICATORS

THE NATURE OF UNCERTAINTIES

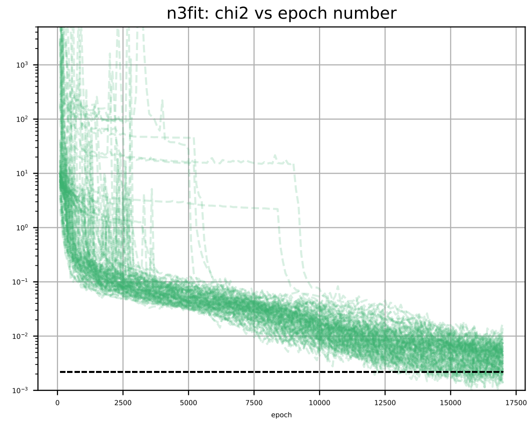
- LEVEL 0:
  - EACH DATAPOINT EQUAL TO THE "TRUTH VALUE"; ZERO UNCERTAINTY
  - FIT → MUST FIND $\chi^2 = 0$ (GET BACK "TRUTH")
  - $\chi^2 \approx 0$ BOTH REPLICA TO REPLICA AND AVERAGE TO TRUTH
  - INTERPOLATION/EXTRAPOLATION UNCERTAINTY
- LEVEL 1:
  - EACH PSEUDO- DATAPOINT IS OBTAINED AS A RANDOM FLUCTUATION WITH GIVEN COVARIANCE MATRIX ABOUT "TRUTH"
⇒ "RUN OF THE UNIVERSE"
  - FIT DATA OVER AND OVER AGAIN
  - $\chi^2 \approx 1$ BOTH REPLICA TO REPLICA AND AVERAGE TO TRUTH
  - FUNCTIONAL UNCERTAINTY
- LEVEL 2:
  - DATA AS IN LEVEL 1
  - GENERATE DATA REPLICAS OF THESE "DATA"
  - FIT PDF REPLICAS TO DATA REPLICAS
  - $\chi^2 \approx 2$ REPLICA TO REPLICA; $\chi^2 \approx 1$ AVERAGE TO TRUTH
  - DATA UNCERTAINTY

UNCERTAINTIES: TYPE AND SIZE

CLOSURE TEST RESULTS (NNPDF4.0)

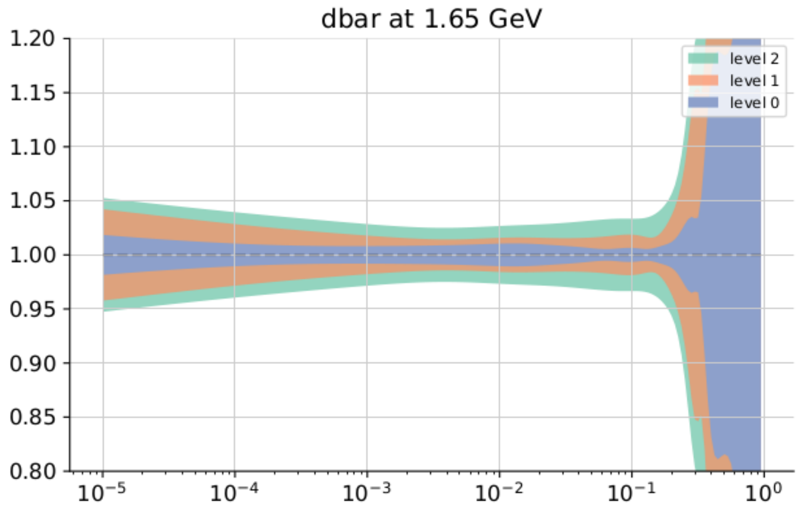
- LEVEL 0 (TRUTH DATA) $\Rightarrow \chi^2 \approx 0$
  UNCERTAINTY NONZERO $\Rightarrow$ INTERPOLATE DISCRETE DATA
- LEVEL 1 (RUNS OF UNIVERSE) $\Rightarrow$ REPLICAS ALL FITTED TO SAME DATA, UNCERTAINTY NONZERO
  $\Rightarrow$ DEGENERACY OF BEST-FITS (FUNCTIONAL FORMS)
- LEVEL 0, 1 AND 2 UNCERTAINTIES COMPARABLE IN SIZE

LEVEL 0 χ^2 VS TRAINING

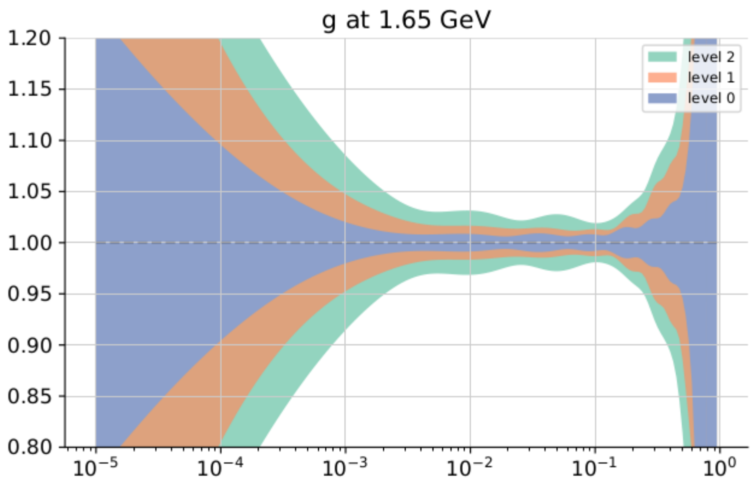


LEVEL 0/1/2 UNCERTAINTIES

ANTIDOWN



GLUON



“PDF” UNCERTAINTIES SUMMARY

- DATA UNCERTAINTIES $\Rightarrow$ MULTIGAUSSIAN
- “PDF” UNCERTAINTIES
  - DATA UNCERTAINTY PROPAGATION + MODEL
  - HESSIAN
    * ABSOLUTE MINIMUM OF χ^2 IN PARAMETER SPACE
    * MULTIGAUSSIAN
  - MONTECARLO
    * IMPORTANCE SAMPLING IN PDF SPACE
    * CAN TEST FOR GAUSSIANITY
  - INTERPOLATION, MODEL, DATA $\Rightarrow$ COMPARABLE SIZE
  - GENERALLY GAUSSIAN

MISSING HIGHER ORDER (THEORY) UNCERTAINTIES

- MAXIMIZE LIKELIHOOD

$$P = N \exp - \left(\frac{d - t}{2\sigma^2_{exp}} \right)$$

d, t ARE REALLY VECTORS AND $1/\sigma^2$ THE INVERSE COVARIANCE MATRIX

- PROBABILITY OF THEORY t GIVEN DATA d; BAYES $\Rightarrow$

$$P(t|d) \propto P(d|t)P(t)$$

- THEORY KNOWN EXACTLY $\Rightarrow$ $P(t) = \delta(t - t^{\mathrm{exact}})$
- THEORY KNOWN PERTURBATIVELY: $t_p \Rightarrow t^{\mathrm{exact}} = t_p + \Delta_p$; $\Delta_p \Leftrightarrow$ MHO
- Δ GAUSSIAN WITH UNCERTAINTY σ_{th};
  INTEGRATE OUT

$$P = N \exp \left[\frac{d - t_p}{2\left(\sigma^2_{exp} + \sigma^2_{th}\right)} \right]$$

- MHOU + EXP COMBINE IN QUADRATURE

MISSING HIGHER ORDER (THEORY) UNCERTAINTIES

- FACTORIZED OBSERVABLE (NONSINGLET STRUCTURE FUNCTION):

$$F_2^{\rm NS}(N,Q^2) = x C_{\rm NS}(\alpha_s(Q^2),N)\exp\left[\int_{Q_0^2}^{Q^2}\frac{d\lambda^2}{\lambda^2}\gamma_{\rm NS}\left(\alpha_s(\mu^2),N\right)\right] f^{\rm NS}(Q_0^2)$$

- SOURCES OF MHOU
  - $\gamma_{\rm NS}^{N^kLO}(\alpha_s,N) = \alpha_s\gamma_{\rm NS}^{(0)}(N) + \alpha_s^2\gamma_{\rm NS}^{(1)}(N) + \alpha_s^{k+1}\cdots + \gamma_{\rm NS}^{(k)}(N)$
  - $C_{\rm NS}^{N^kLO}(\alpha_s(Q^2),N) = 1 + \alpha_s C_{\rm NS}^{(1)}(N) + \cdots + \alpha_s^{k+1} C_{\rm NS}^{(k)}(N)$

SCALE VARIATION

- BASIC IDEA: $\alpha_s(\kappa^2\mu^2) = \alpha_s(\mu^2)[1+O(\alpha_s)]$; AT N^{k}LO DIFFERENCE $\Leftrightarrow$ β-FCTN UP TO β_k
  - $\bar{C}(\alpha_s(\kappa_r^2 Q^2, \kappa_r^2) = C\alpha_s(Q^2)[1+O(\alpha_s)] \Rightarrow$ FIXES $\bar{C}^{(k)}$ IN TERMS OF $C^{(k)}$
  - $\bar{\gamma}(\alpha_s(\kappa_f^2 Q^2, \kappa_f^2) = \gamma\alpha_s(Q^2)[1+O(\alpha_s)] \Rightarrow$ FIXES $\bar{\gamma}^{(k)}$ IN TERMS OF $\gamma^{(k)}$
- $\Delta C = \bar{\gamma}(\alpha_s(\kappa_r^2 Q^2, \kappa_r^2) - \gamma(\alpha_s(Q^2)$ RENORMALIZATION SCALE $\mu_r = \kappa_r Q$ VARN
  SCALE AT WHICH UV DIVS ARE SUBTRACTED
- $\Delta\gamma = \bar{\gamma}(\alpha_s(\kappa_f^2 Q^2, \kappa_f^2) - \gamma(\alpha_s(Q^2)$ FACTORIZATION SCALE $\mu_f = \kappa_f Q$ VARN
  SCALE AT WHICH COLLINEAR DIVS ARE FACTORIZED
  - CHANGE IN $\gamma \Rightarrow$ CHANGE IN PDF $f(Q^2) \Rightarrow$ CAN INCLUDE $\Delta\gamma$ AS Δf
  - FIXED F FACTORIZED AS $C\otimes f \Rightarrow$ CAN INCLUDE Δf AS ΔC

MHOU PRESCRIPTIONS

prediction for datapoint i, scale choice $\mu_r^{(k)}, \mu_f^{(k)}$, default μ_r^0, μ_f^0;

$$\Delta^k(\sigma_i) = \sigma_i[\{\mu^{(k)}\}] - \sigma_i[\{\mu_0\}]$$

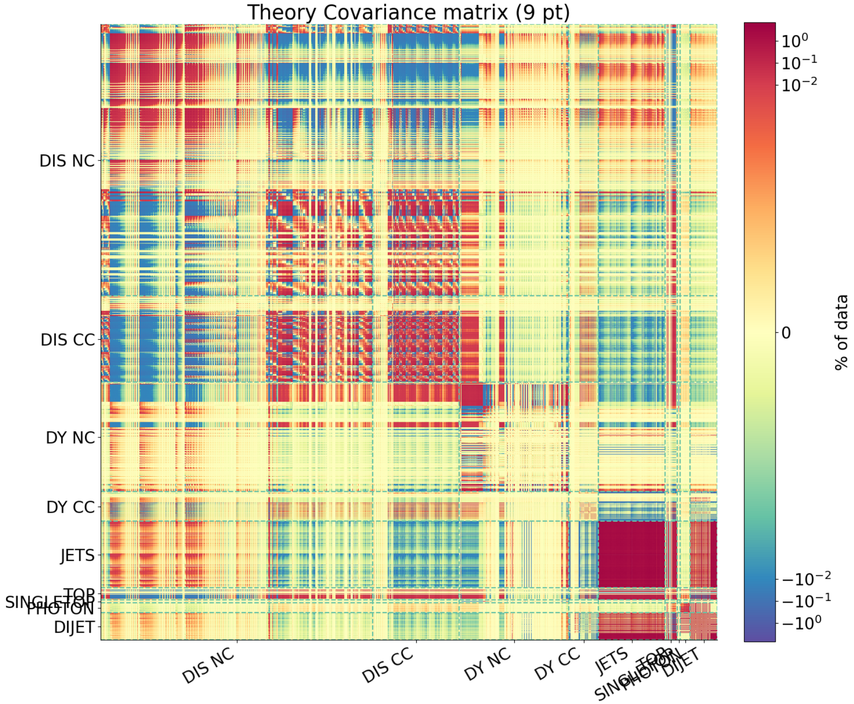
- VARY μ_r, μ_f ABOUT μ_0
- PICK A SET OF POSSIBLE VARIATIONS
  - 3PT $\mu_r = \mu_f$, $\kappa = 2,\ 1/2$
  - 9 PT μ_r, μ_f VARIED INDEP. $\kappa = 2,\ 1/2$
  - 7 PT μ_r, μ_f VARIED INDEP. $\kappa = 2,\ 1/2$, AVOID $mu_r/\mu_f = 4$
- ENVELOPE: TAKE LARGEST AND SMALLEST σ AS UNCERTAINTY BAND
- THEORY COVARIANCE MATRIX:

$$\sigma_{i,j} = \frac{1}{N}\sum_k \Delta^k(\sigma_i)\Delta^k(\sigma_j)$$

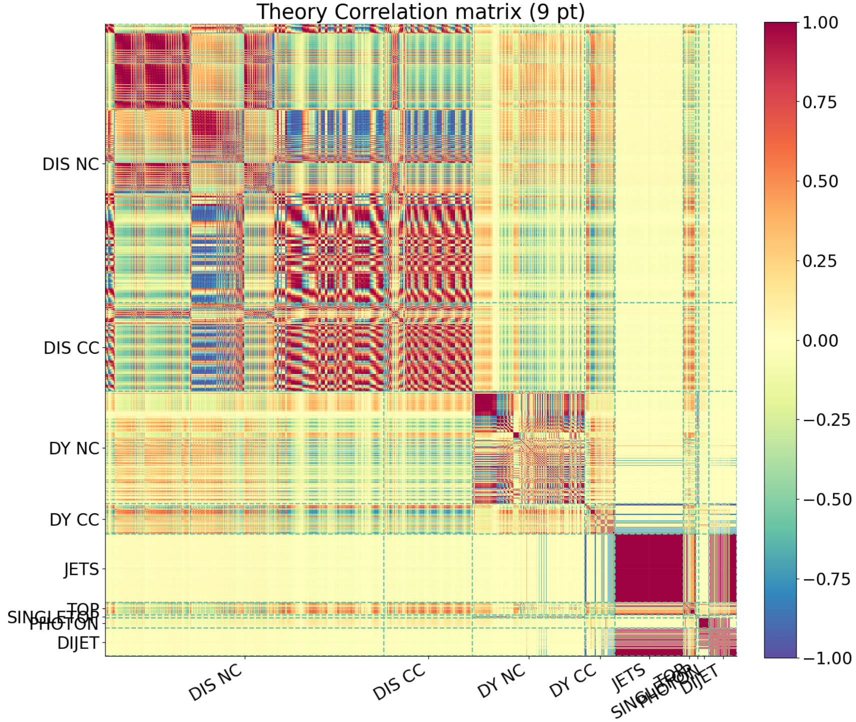
  - SINGLE PROCESS: k RUNS OVER COMMON SET OF SCALE CHOICES
  - MANY PROCESSES:
    - UNCORRELATED RENORMALIZATION: DIFFERENT FOR DIFFERENT HARD PROCESSES
    - CORRELATED FACTORIZATION: MHOU OF PERTURBATIVE EVOLUTION UNIVERSAL

A LOOK AT THE THEORY COVARIANCE MATRIX

COVARIANCE



CORRELATION



HEAVY QUARKS: DECOUPLING

- DECOUPLING SCHEME $\Rightarrow$ HEAVY FLAVOR GRAPHS SUBTRACTED AT ZERO MOMENTUM (Collins, Wilczek, Zee, 1978)
- $N_f = 3$ ACTIVE FLAVORS IN β FUNCTION & EVOLUTION EQUATIONS
- DECOUPLING VS $\overline{\text{MS}}$ $\Leftrightarrow$ DIFFERENT RENORMALIZATION & FACTORIZATION SCHEMES

EXAMPLE: PHOTON SELF-ENERGY

$$\Pi^R(q^2) = \frac{2\alpha}{\pi} \int_0^1 dx x(1-x) \ln \frac{m^2 - x(1-x)q^2}{\mu_r^2}$$

- $\overline{\text{MS}}$ $\ln \frac{m^2 - x(1-x)q^2}{\mu_r^2} = \ln \frac{q^2}{\mu_r^2} + O\left(\frac{m^2}{q^2}\right) \Rightarrow$ RUNNING α
- DECOUPLING: $\ln \frac{m^2 - x(1-x)q^2}{\mu_r^2} = O\left(\frac{q^2}{m^2}\right)$

SOLID $\Rightarrow$ HEAVY; DASHED $\Rightarrow$ LIGHT

M. Buza et al.: Charm

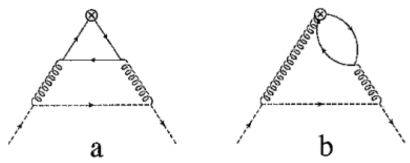


Fig. 2. $O(\alpha_s^2)$ contributions to the purely-singlet OME $A^{\rm PS}_{q'q}$. Here q and q' are represented by the *dashed* and *solid lines* respectively. In the case of $q' = H$ these graphs contribute to the heavy-quark OME $A^{\rm PS}_{Hq}$

MATCHING

- PDFs, α_s IN $N_f = 3$ & $N_f = 4$ RELATED BY MATCHING CONDITIONS
- DETERMINED BY COMPUTING OPERATOR MATRIX ELEMENTS IN EITHER SCHEME AND EQUATING

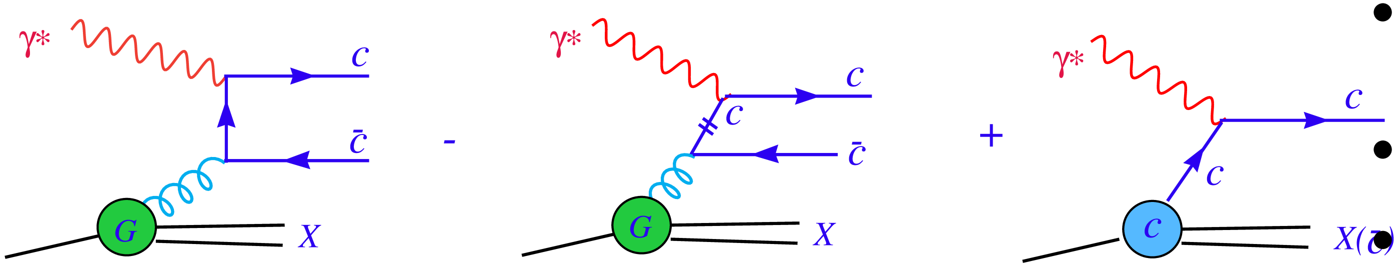
HEAVY QUARKS IN DIS

- $\overline{\text{MS}}$ SCHEME $\Rightarrow$ HQ MASSLESS PARTON $\Rightarrow \ln Q^2/m_h^2$ RESUMMED TO ALL ORDERS BY EVOLUTION EQNS, $O(m^2/Q^2)$ CONTRIBNS NEGLECTED

- DECOUPLING SCHEME $\Rightarrow$ HQ IN HARD XSECT $\Rightarrow O(m^2/Q^2)$ CONTRIBNS INCLUDED, $\ln Q^2/m_h^2$ TREATED AT FIXED ORDER

THE BEST OF TWO WORLDS
MATCHED SCHEMES
ACOT

$m_c \neq 0$, LO charm radiation — $m_c = 0$, LO charm radiation — $m_c = 0$ charm pdf



- USE $\overline{\text{MS}}$ FOR $Q^2 > m_q^2$ WITH FULL MASS DEP. RETAINED
- KEEP ALL FLAVOURS IN RUNNING, DGLAP
- SUBTRACT DOUBLE COUNTING

FONLL

COMBINE N^iLL MASSLESS RESUMMED & N^jLO MASSIVE FIXED-ORDER
$\Rightarrow$ EXPAND RESUMMED RESULT; REPLACE THE FIRST j ORDERS WITH THEIR MASSIVE COUNTERPARTS

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

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

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

ADVANTAGES

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

COMPLICATION

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

SOLUTIONS

- EITHER RE-EXPRESS 3F-SCHEME PDFS & α_s IN TERMS OF THE 4F-SCHEME ONES
- OR HAVE SIMULTANEOUSLY 3F & 4F SCHEME α_s & PDFS

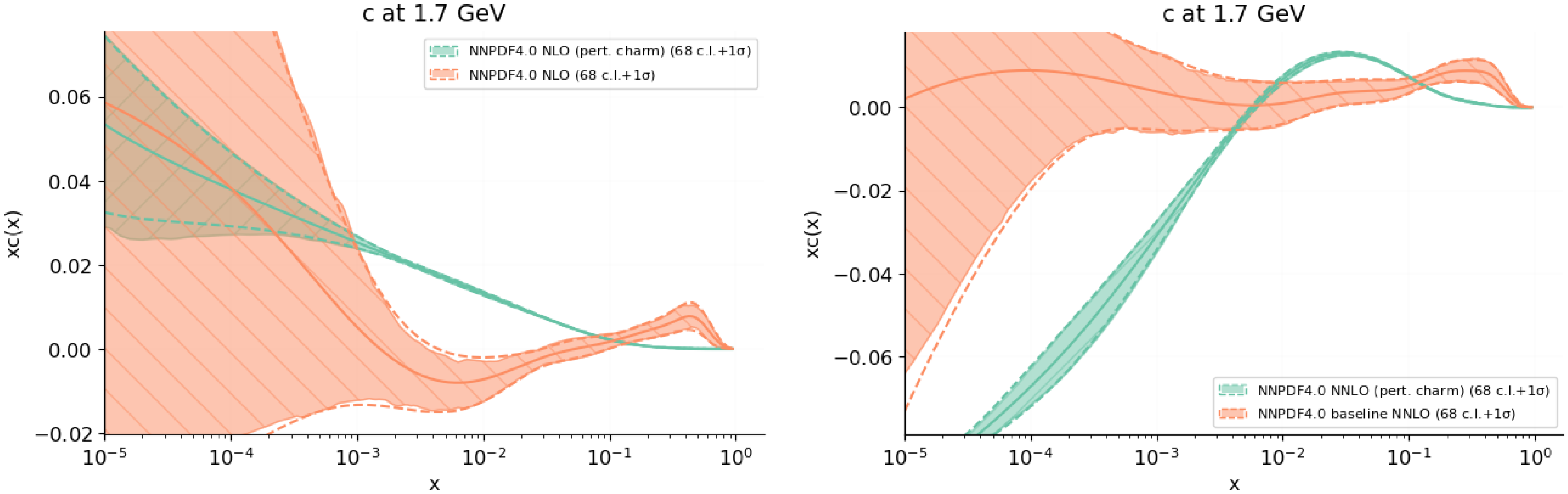
THE CHARM PDF

PERTURBATIVE CHARM

- IN $N_f = 3$ SCHEME CHARM PDF VANISHES
- IN $N_f = 4$ SCHEME, CHARM DETERMINED BY PERTURBATIVE MATCHING
- STARTING AT NNLO (TWO LOOPS) DOES NOT VANISH AT ANY SCALE

PERTURBATIVE CHARM PDF, $n_f = 4$ SCHEME, Q=1.7 GeV

NLO NNLO



INTRINSIC CHARM

- DEFINE CHARM PDF AS OME:

$$\langle p|\bar{c}\gamma^{\mu_1}D^{\mu_2}\dots D^{\mu_n}c|p\rangle = A_c^n p^{\mu_1}\dots p^{\mu_n} - \text{traces}$$

$$A_c^n = \int_0^1 dx\, x^{n-1} c(x)$$

- DECOUPLE CHARM MASS LOGS $\Rightarrow$ CHOOSE $N_f = 3$ SCHEME
- ALLOW NONVANISHING (SCALE-INDEPENDENT) CHARM PDF
- IN $N_f = 4$ SCHEME CHARM PDF DIFFERS FROM THAT FIXED BY MATCHING

NONVANISHING CHARM IN THE $N_F = 3$ (DECOUPLING) SCHEME $\Rightarrow$ INTRINSIC CHARM

THEORY UNCERTAINTIES SUMMARY

- THEORY UNCERTAINTIES $\Rightarrow$ THEORY COVARIANCE MATRIX
- SCALE VARIATION:
  - RENORMALIZATION $\Rightarrow$ MHOU IN PARTONIC CROSS-SECTION
  - FACTORIZATION $\Rightarrow$ MHOU IN ANOMALOUS DIMENSION
- HEAVY QUARKS
  - DECOUPLING SCHEME $\Rightarrow$ QUARK MASS EFFECTS INCLUDED
  - $\overline{\mathrm{MS}}$ SCHEME $\Rightarrow$ COLLINEAR MASS LOGS RESUMMED
  - MATCHING $\Rightarrow$ BOTH INCLUDED
  - HQ PDF DIFFERS FROM RESULT OF MATCHING $\Rightarrow$ INTRINSIC HQ

FOOD FOR THOUGHT

- CAN YOU THINK OF NEW PROCESSES AT EIC FOR PDF DETERMINATION?
  AND CAN YOU THINK OF A SYNERGY BETWEEN EIC & LHC?
- WHAT MIGHT BE THE REASON WHY TOLERANCE IS NEEDED?
  AND CAN YOU THINK HOW TO TEST IT?
- CAN YOU THINK OF ALTERNATIVE WAYS OF ESTIMATING MHOUs?
- IF INTRINSIC HQ PDFs ARE NONZERO, HOW DO YOU EXPECT THEIR SIZE TO SCALE WITH THE HQ MASS?