The CT14 MC replicas

Tie-Jiun Hou
In collaboration with CTEQ-TEA

Southern Methodist University

April 13, DIS16, DESY, Hamburg

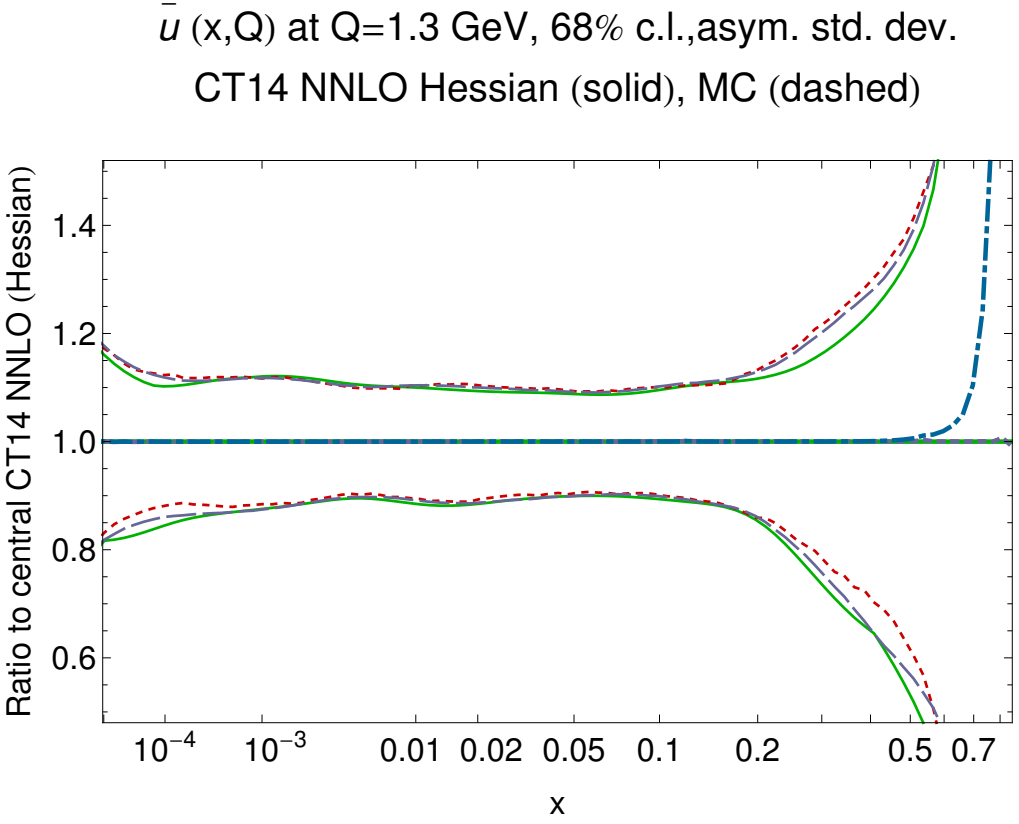
CTEQ-TEA group

- CTEQ – Tung et al. (TEA)
  in memory of Prof. Wu-Ki Tung, who established CTEQ Collaboration in early 90's.
- Current members:
  Sayipjamal Dulat (Xinjiang U.)
  Tie-Jiun Hou, Pavel Nadolsky (Southern Methodist U.)
  Jun Gao(Argonne Nat. lab.)
  Marco Guzzi(U. of Manchester)
  Joey Huston, Jon Pumplin, Carl Schmidt,
  Dan Stump, C.-P. Yuan(Michigan State U.)

CT14 NNLO error PDFs

- CT10 includes only pre-LHC data
- CT14 is the first CT analysis including LHC Run 1 data
- CT14 also includes the new Tevatron D0 Run 2 data on W-electron charge asymmetry
- CT14 uses a more flexible parametrization in the non-perturbative PDFs.
- We have published its results at NNLO, NLO and LO.

Monte-Carlo replicas for CT14 asymmetric errors

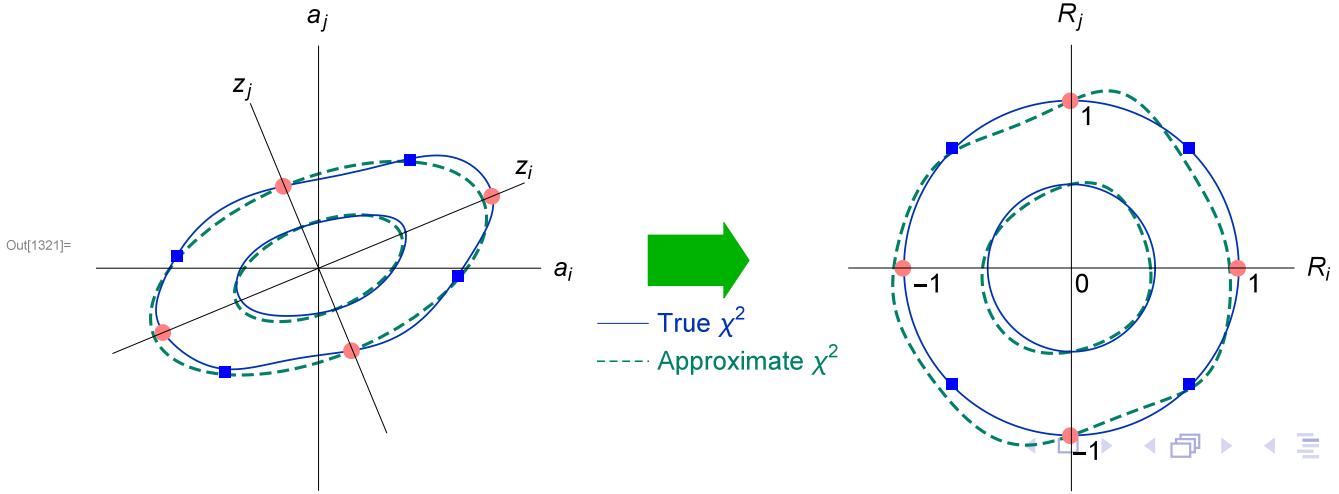


Green:Hessian 68% c.l. errors
Blue:Asymmetric MC replicas

Generation of MC replica from CT14 Hessian eigenvector sets

MC replicas for $f_a(x, Q) \equiv f$....

- are constructed from the best-fit (central) PDF values f_0 and 68% c.l. extreme displacements $f_{\pm i}$ along eigenvector directions $\vec{u}_i$, $i = 1, \ldots, 28$ in parameter space near χ^2 minimum
- retain exact information about boundaries of 68%/90% probability regions; approximate probability everywhere using Gaussian approximation
- approximate asymmetric Hessian errors using modified standard deviations

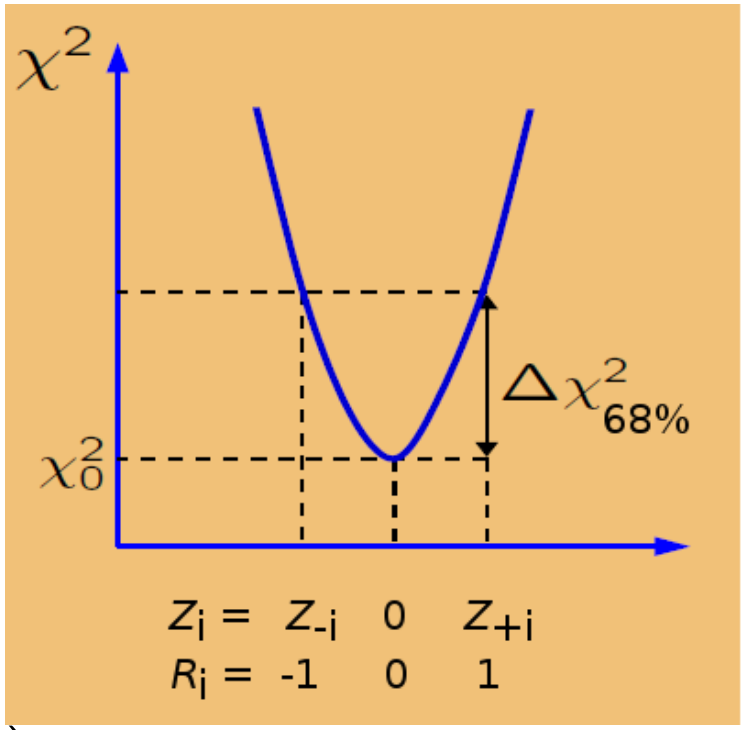


Sources of asymmetry of PDF errors for QCD predictions

$\chi^2 \Rightarrow$ PDFs $f_a(x, Q) \Rightarrow$ Cross section X

1. The asymmetry of χ^2 is usually mild near the minimum; can approximate

$$\chi^2 \approx \chi_0^2 + \sum_{i=1}^{D} R_i^2,$$



Where R_i (rescaled z_i) obeys the standard normal distribution:

$$Probability(\{R\}) \sim e^{-\sum_{i=1}^{D} R_i^2/2}$$

$$f_{\pm i}(\{R\}) = f(0, 0, ..., R_i = \pm 1, ..., 0)$$

$$f_{\pm i, \pm j}(\{R\}) = f(0, 0, ..., R_i = \pm\frac{1}{\sqrt{2}}, ..., R_j = \pm\frac{1}{\sqrt{2}}, 0)$$

Additional χ^2 contribution from dynamic penalty (Tier-2) for CT PDFs.

Sources of asymmetry of PDF errors for QCD predictions

$\chi^2 \Rightarrow$ PDFs $f_a(x, Q) \Rightarrow$ Cross section X 2. PDFs and cross sections are generally asymmetric function of R_i

$$X(\{R\}) = X(\{0\}) + \sum_{i=1}^{D} \frac{\partial X}{\partial R_i} R_i + \frac{1}{2} \sum_{i=1}^{D} \frac{\partial^2 X}{\partial R_i \partial R_j} R_i R_j + \ldots$$

Evaluate partial derivatives by finite differences

$$\frac{\partial X}{\partial R_i} \approx (X_{+i} - X_{-i})/2 \qquad \text{need 2D eigenvectors sets}$$

$$\frac{\partial^2 X}{\partial R_i^2} \approx X_{+i} - X_{-i} - 2X_0 \qquad \text{need 2D eigenvectors sets}$$

$$\frac{\partial^2 X}{\partial R_i \partial R_j} \approx (X_{+i,+j} + X_{-i,-j} - X_{+i,-j} - X_{-i,+j})/2 \qquad \text{need } 2D(D-1) \text{ New evsets}$$

Symmetric PDF errors

Keep only linear terms

$$X(\{R\}) = X(\{0\}) + \sum_{i=1}^{D} \frac{X_{+i} - X_{-i}}{2} R_i$$

1. The Hessian method produces a symmetric master formula (Stump, Pumplin, Tung, et al., 1999):

$$\delta_{68}^{H} X = |\Delta X| = \frac{1}{2}\sqrt{\sum_i (X_{+i} - X_{-i})^2}$$

2. The MC generation produces N_{rep} symmetric replicas

$$X^{(k)} = X(\{0\}) + \sum_{i=1}^{D} \frac{X_{+i} - X_{-i}}{2} R_i^{(k)}, \quad k = 1, ..., N_{rep}$$

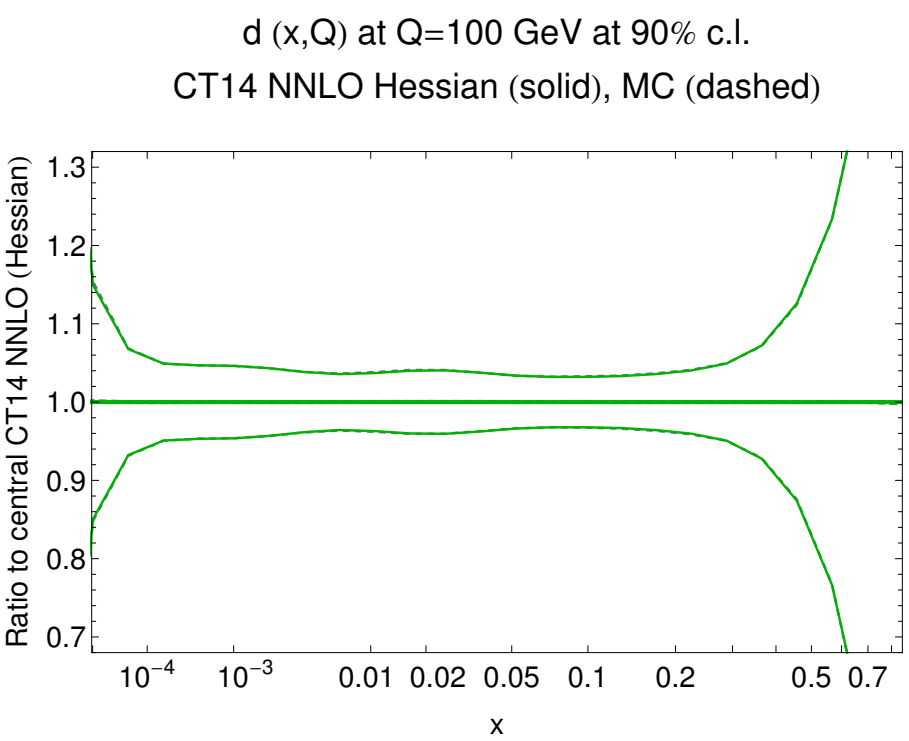
$R_i^{(k)}$ are normally distributed. We choose $N_{rep} = 1000$.

Hessian and MC symmetric errors for PDFs (X=f)...

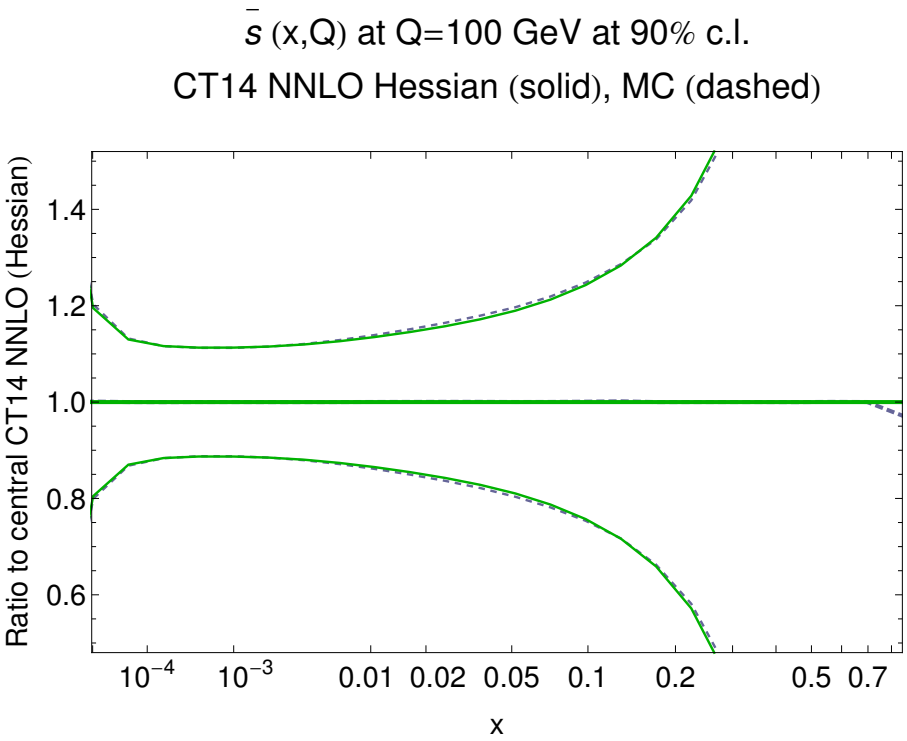
agree well. The MC mean can deviate when the PDFs vanish

The MC error is estimated by the standard deviation of X,

$$\delta_{68}^{MC} X = \sqrt{\langle (X - \langle X \rangle)^2 \rangle}$$



Symmetric uncer., **preliminary**



Symmetric uncer., **preliminary**

CT14 asymmetric PDF errors

Include the diagonal second derivatives

$$X(\{R\}) = X(\{0\}) + \sum_{i=1}^{D} \frac{X_{+i} - X_{-i}}{2} R_i + \frac{1}{2} \sum_{i,j=1}^{D} (X_{+i} + X_{-i} - 2X_0) R_i^2$$

1.The Hessian method produces asymmetric master formulas
(Nadolsky, Sullivan, 2001)

$$\delta_{68}^{H,>} X = \sqrt{\sum_i \left(max[X_{+i} - X_0, X_{-i} - X_0, 0]\right)^2}$$

$$\delta_{68}^{H,<} X = \sqrt{\sum_i \left(max[X_0 - X_{+i}, X_0 - X_{-i}, 0]\right)^2}$$

CT14 asymmetric PDF errors

2.The MC generation produces N_{rep} asymmetric replicas

$$X^{(k)} = X(\{0\}) + \delta X^{(k)} - \langle \delta X \rangle$$

$$\delta X^{(k)} = \sum_{i=1}^{D} \frac{X_{+i} - X_{-i}}{2} R_i^{(k)} + \frac{1}{2} \sum_{i,j=1}^{D} (X_{+i} + X_{-i} - 2X_0)(R_i^{(k)})^2$$

With this definition,$\langle X \rangle = X(\{0\})$, does not fluctuate about $X(\{0\})$
The MC errors can be estimated by **asymmetric** standard deviations,

$$\delta_{68}^{MC,>} X = \sqrt{\langle (X - \langle X \rangle)^2 \rangle_{X > \langle X \rangle}}$$

$$\delta_{68}^{MC,<} X = \sqrt{\langle (X - \langle X \rangle)^2 \rangle_{X < \langle X \rangle}}$$

Alternatively,$\delta_{68}^{MC,\gtrless} X$ can be estimated by 68% central probability intervals for ordered X_i values (more cumbersome and noisy than the std. deviations)

Comparison with Watt-Thorne algorithm

CT14 algorithm:

$$X^{(k)} = X(\{0\}) + \sum_{i=1}^{D} \frac{X_{+i} - X_{-i}}{2} R_i^{(k)} + \frac{1}{2} \sum_{i,j=1}^{D} (X_{+i} + X_{-i} - 2X_0)(R_i^{(k)})^2 - \langle \delta X \rangle$$

$$\delta_{68}^{MC,\gtrless} X = \sqrt{\langle (X - \langle X \rangle)^2 \rangle_{X \gtrless \langle X \rangle}}$$

Recommended

Asymmetric algorithm in Watt, Thorne (arXiv:1205.4024)

$$X^{(k)} = X(\{0\}) + \sum_{i=1}^{D} \frac{\partial X}{\partial R_i} R_i^{(k)}$$

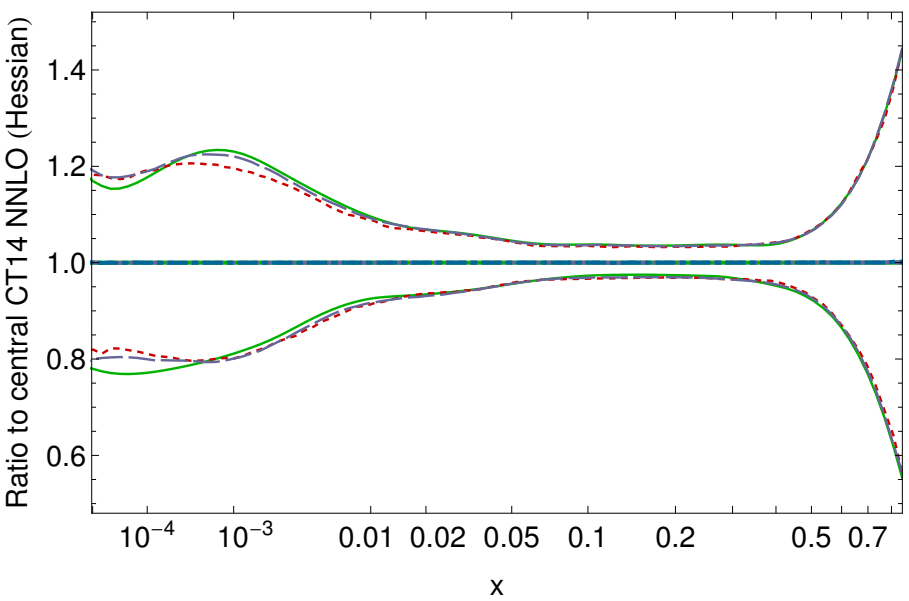
$$\frac{\partial X}{\partial R_i} = \begin{cases} X_i - X_0, & R_i^{(k)} > 0 \\ X_0 - X_{-i}, & R_i^{(k)} < 0 \end{cases}$$

Different from the CT14 algorithm if $R_i^{(k)} \neq 0, \pm 1$

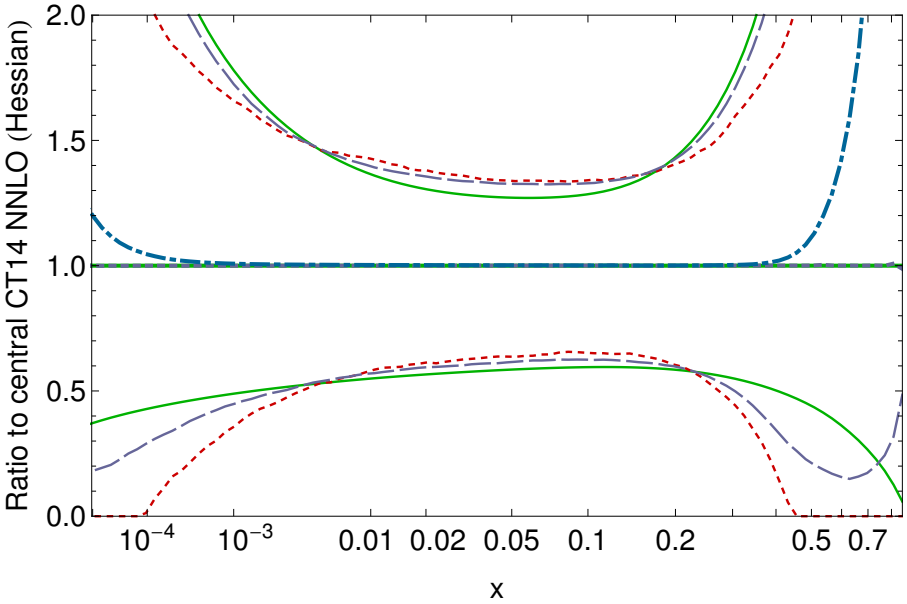
We find that separate averaging of positive and negative displacements is essential for recovering the asymmetry of $\delta_{68}^{H,\lessgtr} X$ in CT14.

Asymmetric standard deviations for PDFs ($X = f$)

Asymmetric uncer., **preliminary**



Asymmetric uncer., **preliminary**

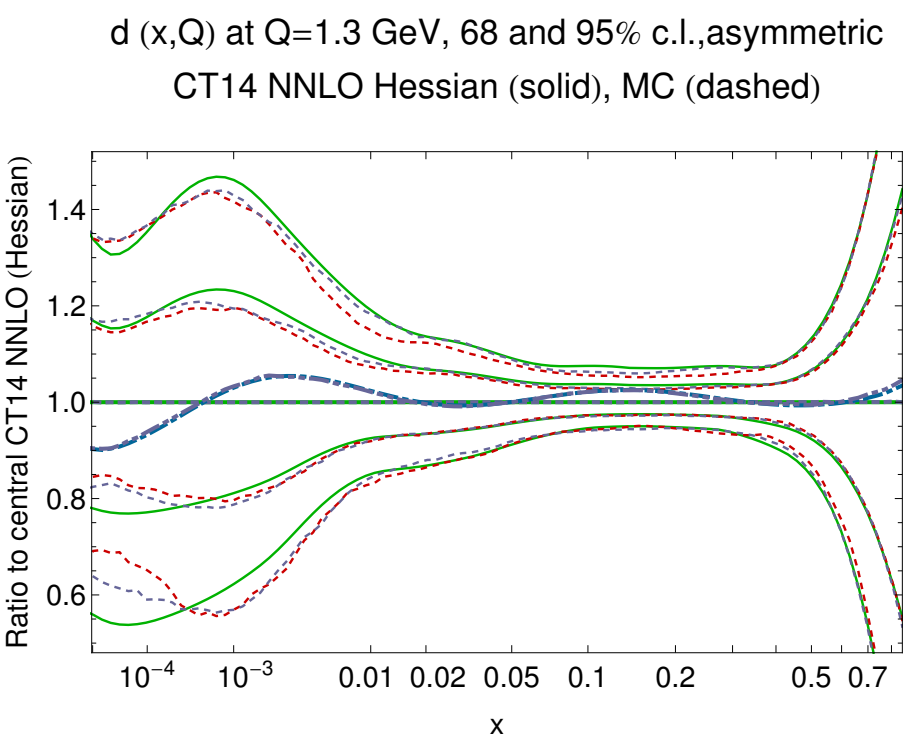


Green: Hessian std. deviation
Red: Symmetric MC std. dev.
Thin blue: Asymmetric MC std. dev.
Thick blue: Asymmetric MC median

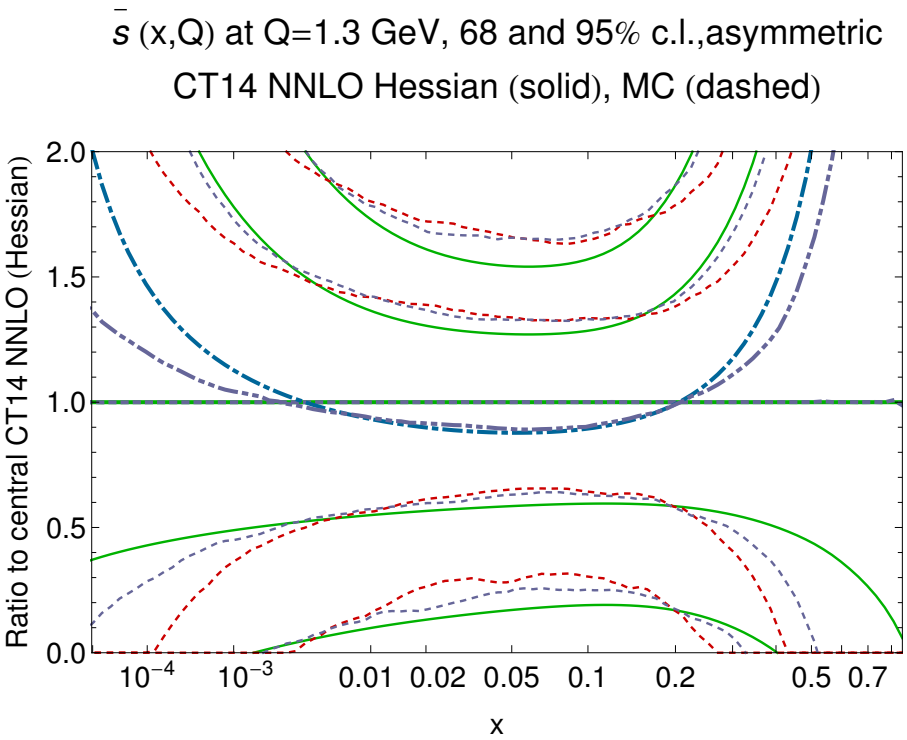
Good agreement between green and light blue, smooth behavior

Asymmetric central probability intervals

Asymmetric uncer., **preliminary**



Asymmetric uncer., **preliminary**

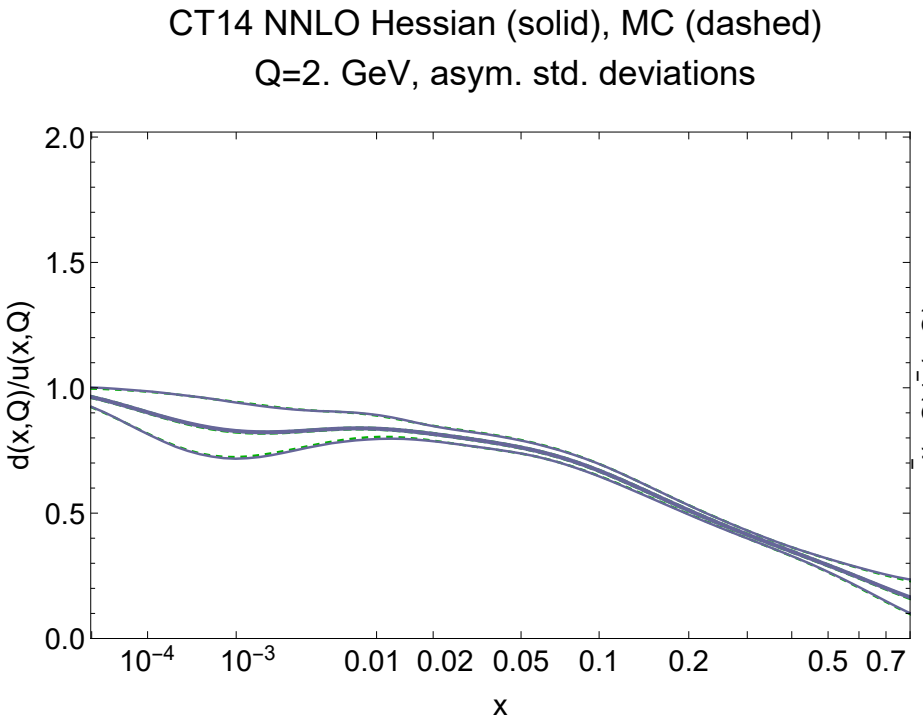


Green: Hessian probability intervals
Red: Symmetric MC generation
Thin blue: Asymmetric MC generation, Watt-Thorne formula
Thick blue: Asymmetric MC median

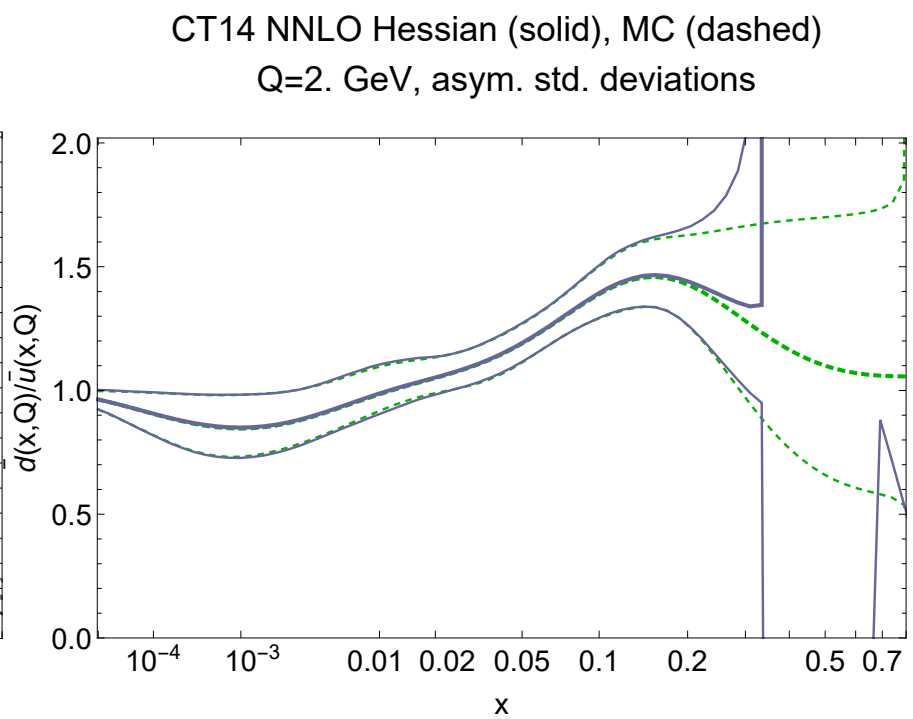
Probability intervals are more sensitive to behavior of individual replica

Asymmetric central probability intervals

Asymmetric uncer., **preliminary**



Asymmetric uncer., **preliminary**

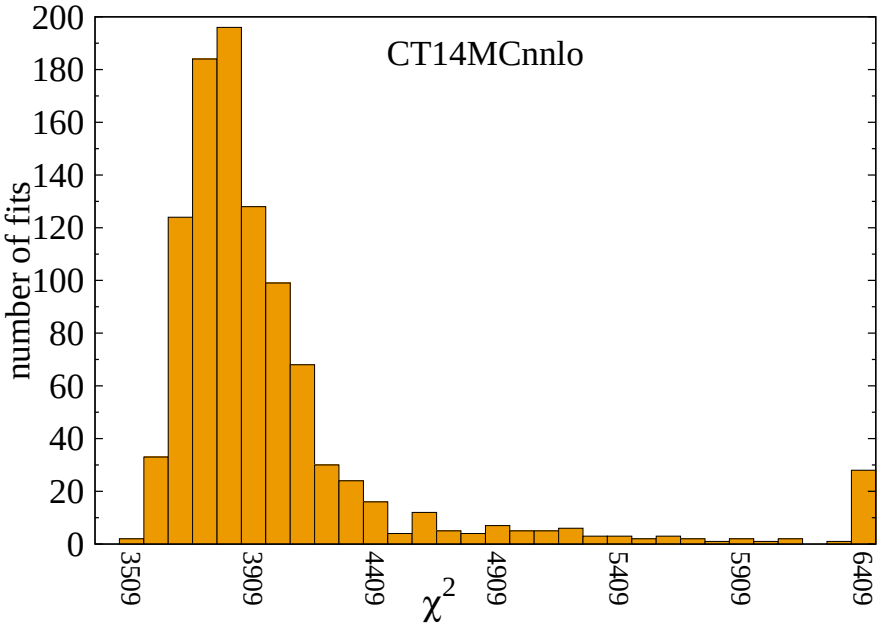


Green: Hessian probability intervals
Thin blue: Asymmetric MC generation, Watt-Thorne formula

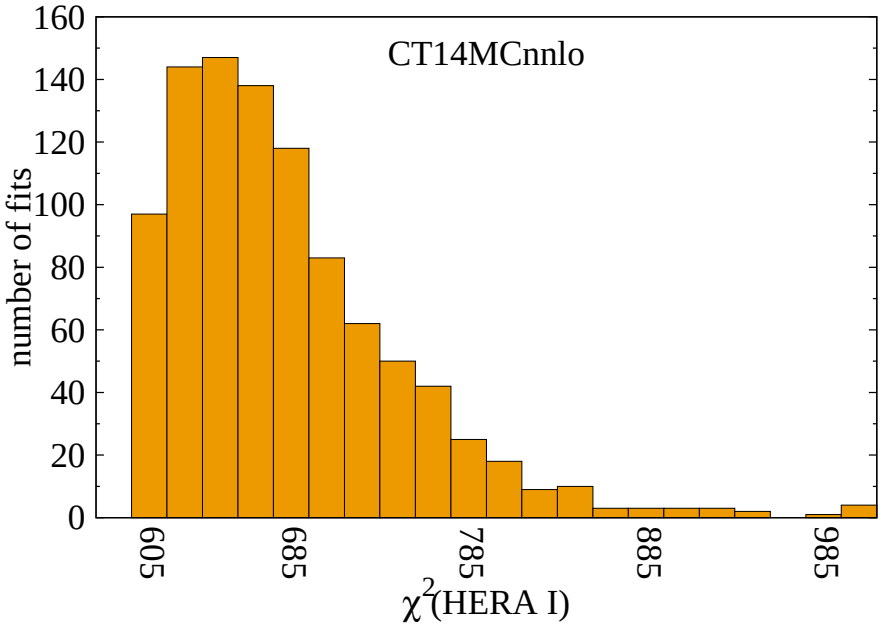
Large deviation happen on the numerical accuracy sensitive region.

Large χ^2 in replicas

preliminary



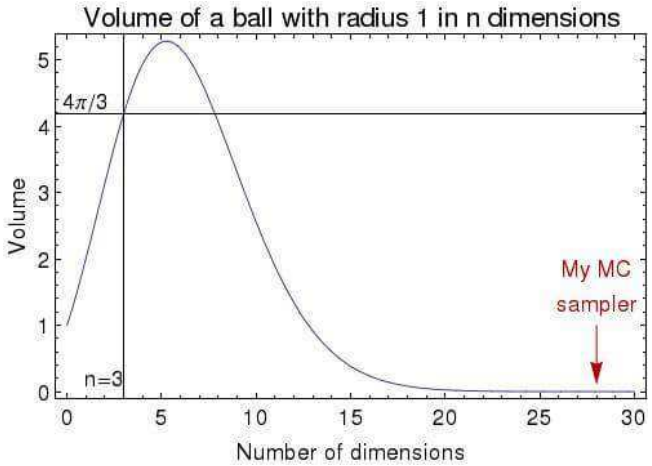
preliminary



Typical CT14MC replicas sets have large χ^2.
Here, we show χ^2 distributions for 1000 replicas, with about 3000 data points (579 for HERA-I) included in the CT14 fit.

Large χ^2 in replicas

Imagine we construct an N -dimensional vector $\vec{R}$ whose coordinates are given by random parameters R_i sampled from a standard normal distribution.

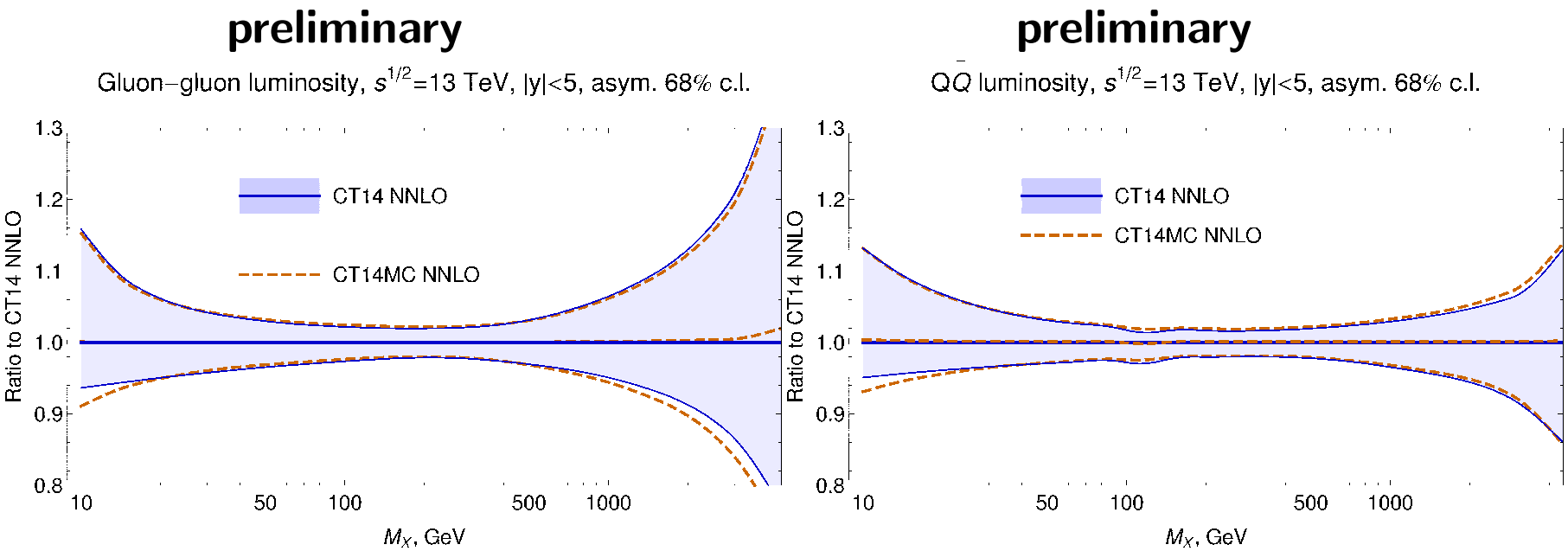


$$\langle \vec{R} \rangle = \frac{\int_{-\infty}^{\infty} dR_1 \int_{-\infty}^{\infty} dR_2 \ldots \int_{-\infty}^{\infty} dR_N \sqrt{\sum_{i=1}^{N} R_i^2}\, e^{-\frac{\sum_{i=1}^{N} R_i^2}{2}}}{\int_{-\infty}^{\infty} dR_1 \int_{-\infty}^{\infty} dR_2 \ldots \int_{-\infty}^{\infty} dR_N\, e^{-\frac{\sum_{i=1}^{N} R_i^2}{2}}}$$

$$= \frac{\int_0^{\infty} dR\, R^N e^{-\frac{R^2}{2}} \int d\Omega_N}{(2\pi)^{N/2}} = \frac{\sqrt{2}\Gamma(\frac{N+1}{2})}{\Gamma(\frac{N}{2})}$$

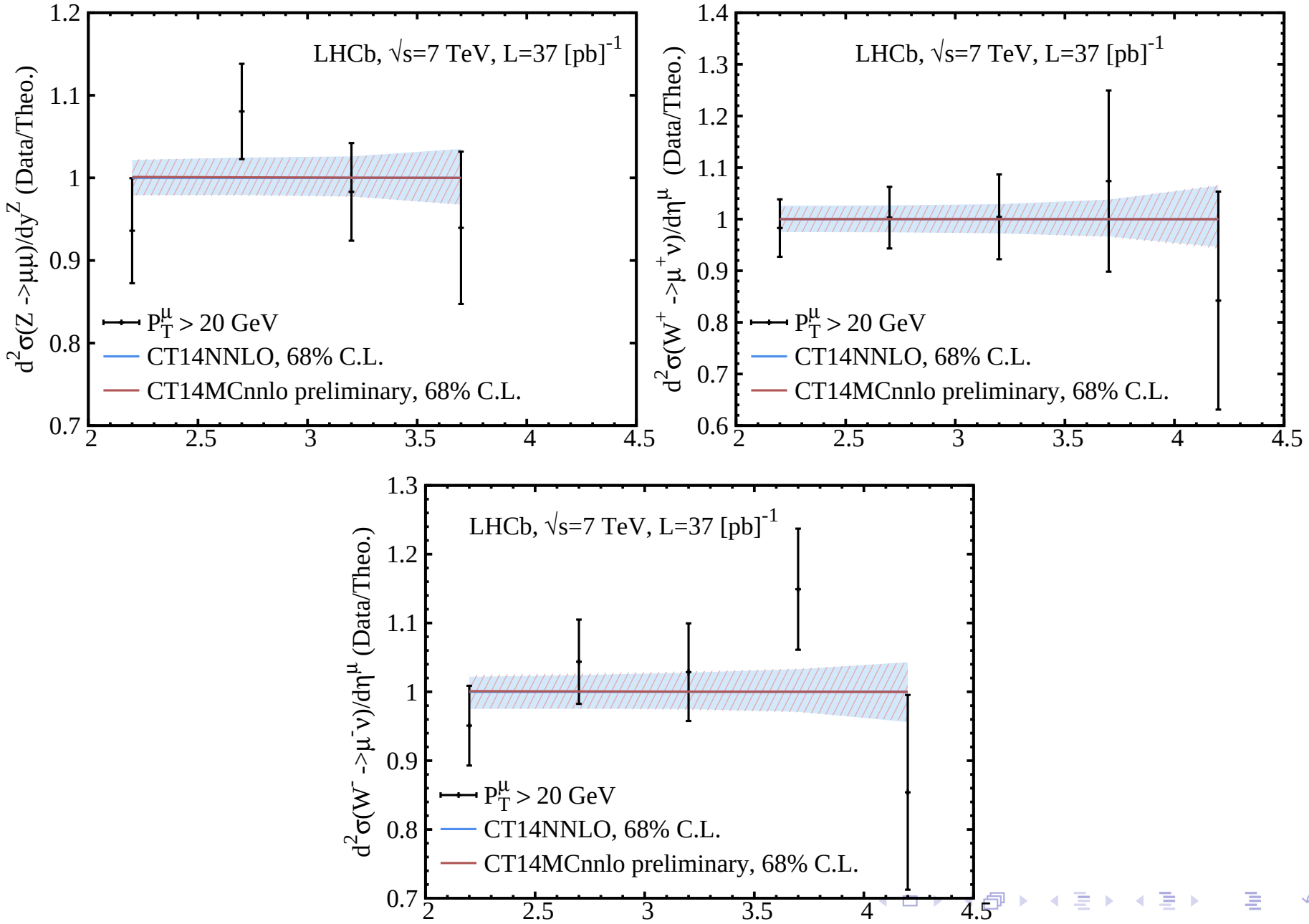
For $N = 28$, this formula gives $\langle \vec{R} \rangle \approx 5.24$, that is, a typical displacement vector of a CT14 replica is more than five standard units in length.

Agreement between CT14 and CT14MC

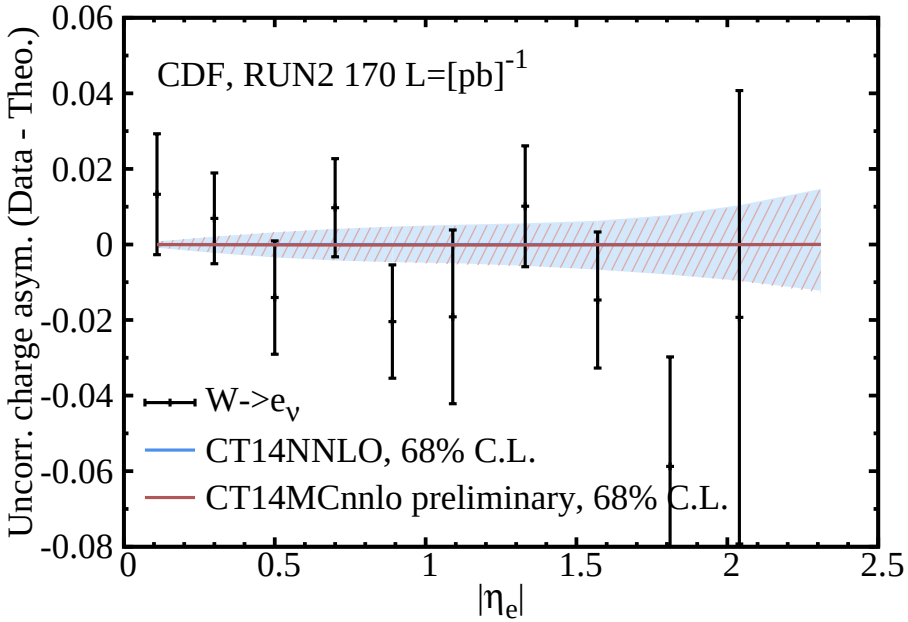


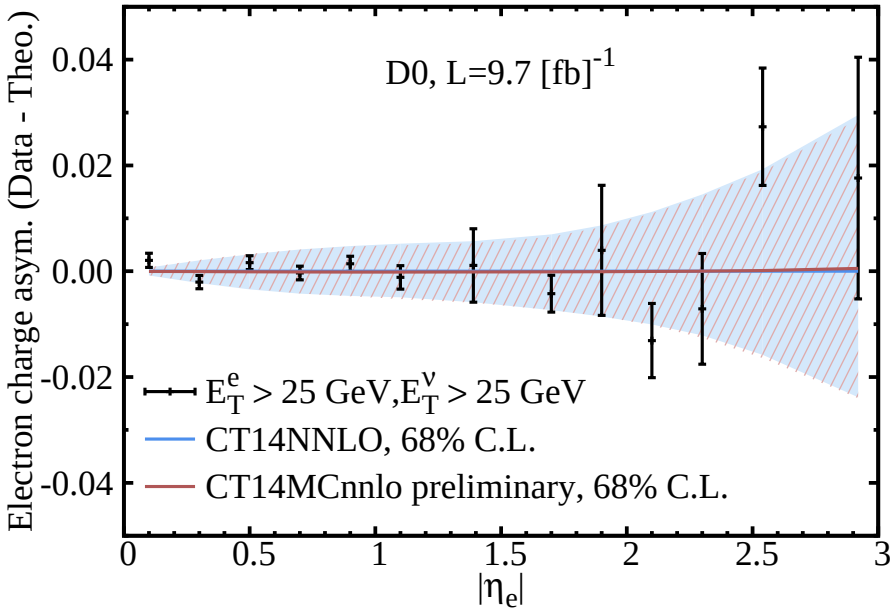
Good agreement in the typical region for gluon-gluon and q-$\bar{q}$ luminosity. The deviation at large and small M_x come from numerical fluctuation.

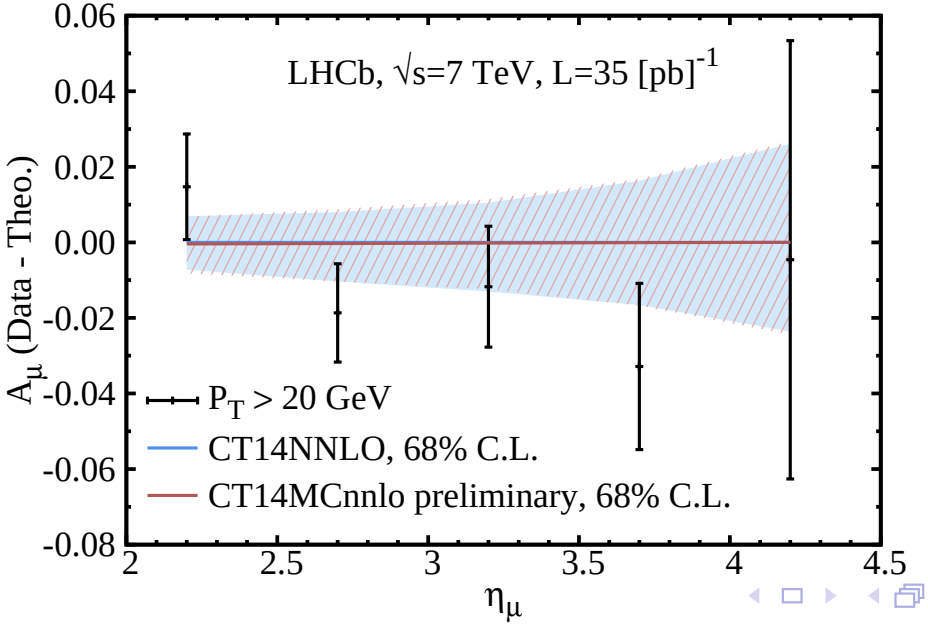
Agreement between CT14 and CT14MC



Agreement between CT14 and CT14MC







Summary

In this work, the asymmetric CT14MC replica set with 1000 replicas is generated base on the CT14 Hessian eigen vector sets. Learning from the Hessian asymmetric master formula, the diagonal second derivative contribution is included.
Uncertainties between CT14 Hessian eigenvector set and CT14MC replica set are in good agreement. Deviation mainly come from the numerical fluctuation and the dynamic penalty of the CT14 Hessian eigenvector set.

Hessian method

spartyness and Tier-2 penalty

The χ^2/N_{pt} is not sufficient to discribe the goodness-of-fit for individual experiment. The chi-square probability for $\chi^2/N_{pt} \geq 11.0/10$ is 0.358; the probability for $\chi^2/N_{pt} \geq 110.0/100$ is 0.232. And the probability for $\chi^2/N \geq 1100.0/1000$ is only 0.015. For this reason, we define the effective Gaussian variable, the "*spartyness*", S_n,

$$C(\chi^2, N) = \int_0^{\chi^2} P(\xi, N) d\xi = \int_{-\infty}^{S_n} \frac{e^{-x^2/2} dx}{\sqrt{2\pi}}.$$

In practice, we use the approximation for S_n

$$S_n \approx L(\chi^2, N_{pt}) = \frac{(18N_{pt})^{3/2}}{18N_{pt}+1} \left\{ \frac{6}{6 - \ln(\chi^2/N_{pt})} - \frac{9N_{pt}}{9N_{pt}-1} \right\}$$

Ideally, the variable S_n has an approximately Gaussian distribution with mean 0 and standard deviation 1. Additional Tier-2 penalty for each experiment to preventing poor fit for individual experiment. (S.Dulat et.al. Phys.Rev. D89 (2014) no.11, 113002)