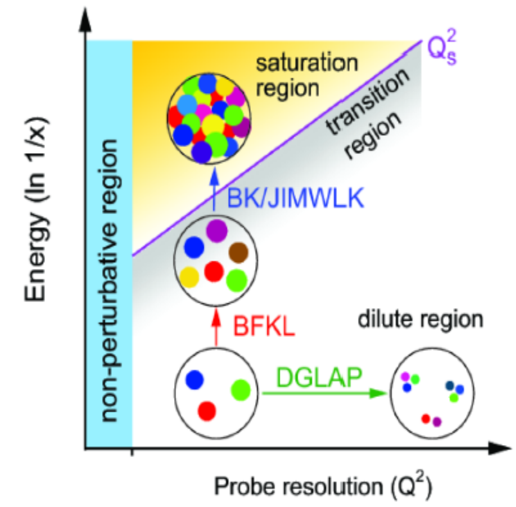
High-performance computing for global fits of parton distribution functions (PDFs)

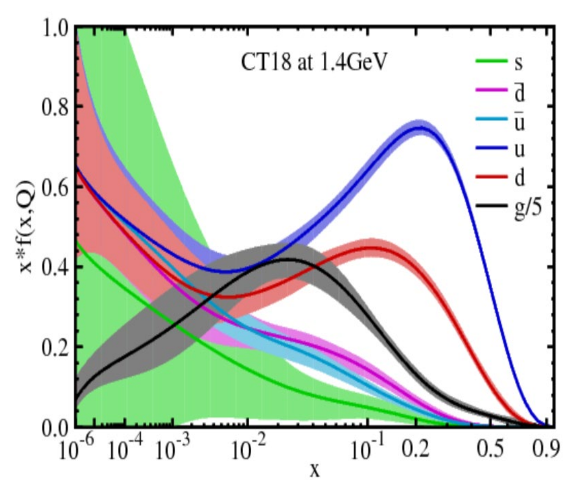


Pavel Nadolsky

Southern Methodist University

Co-convener, EF06 topical group “Hadron structure, forward QCD, and hadron spectroscopy”

CTEQ-TEA (Tung et al.) working group



References

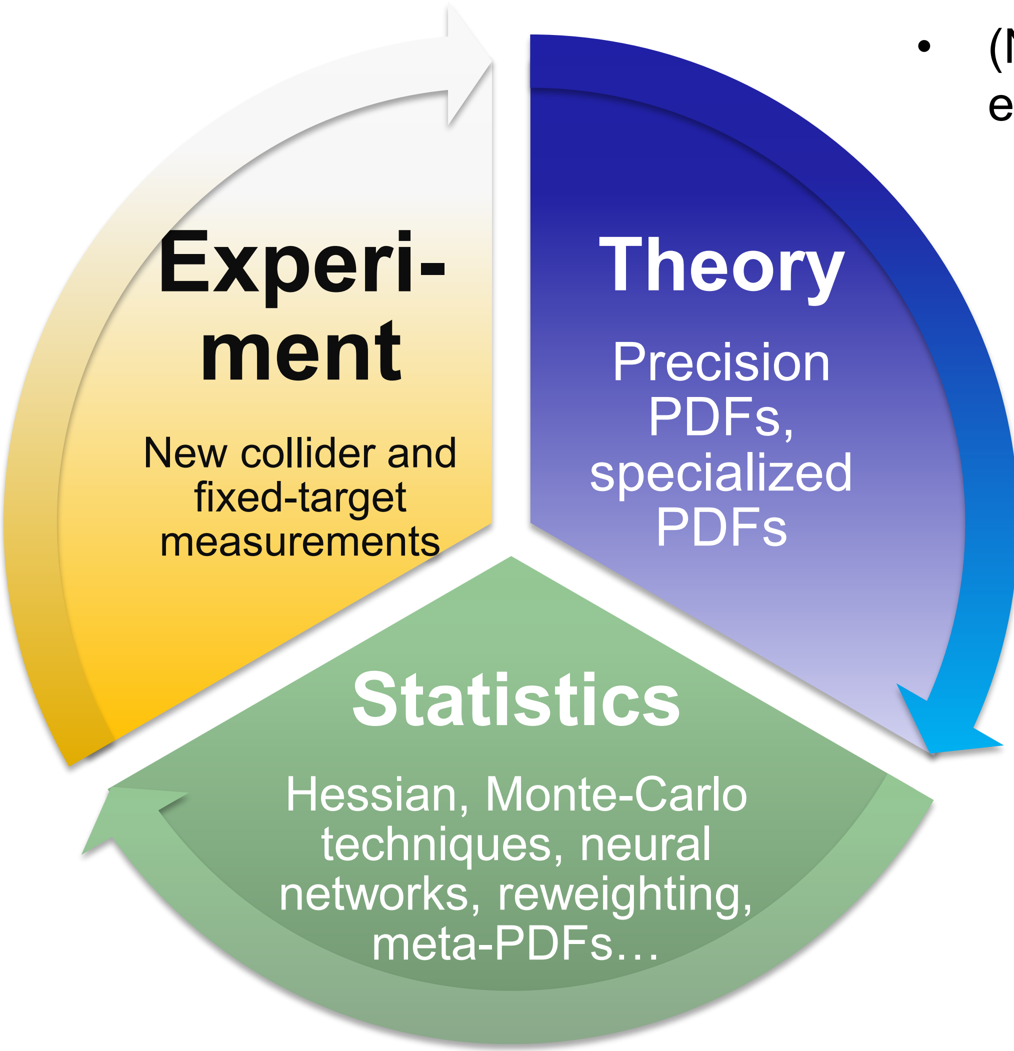
1. CTEQ-TEA group, T.-J. Hou et al., arXiv:1912.10053
2. K. Kovarik, P. Nadolsky, D. Soper, arXiv: 1905.06957

SMU

CTEQ

Thanks to Fred Olness (nCTEQ) for inputs

Frontiers of the PDF analysis



- (N)NNLO QCD computations require equally accurate PDFs
- Significant advances on all aspects of the proton PDF analysis are necessary to meet physics targets of the HL-LHC program
- Exceptional opportunities to learn about the 3-dim. structure of protons, nuclei, pions at new facilities in the HL-LHC era: EIC, LHeC, AMBER, LHCSpin…

PDF fits require speed and accuracy; critically depend on…

…high-performance computing

…data science & machine learning

(N)NNLO QCD, NLO EW computations

Development of fast (N)NNLO interfaces

Benchmarking of multithreaded fitting codes (Fortran, C++, Python,…)

Probability distributions with hundreds of parameters

Shapes of PDFs presented by flexible functions (ABM, CTEQ, HERA, MMHT, …) or CNNs (NNPDF)

Global fits to >40 heterogenous measurements in collider and fixed-target experiments

Minimization/learning with MINUIT, TensorFlow, genetic algorithms…

Algebraic marginalization of labyrinthine experimental systematic uncertainties

PDF4LHC combinations of PDF sets for LHC applications. Dimensionality reduction in METAPDF/compressed PDF methods. The majority of LHC publications use PDF error sets!

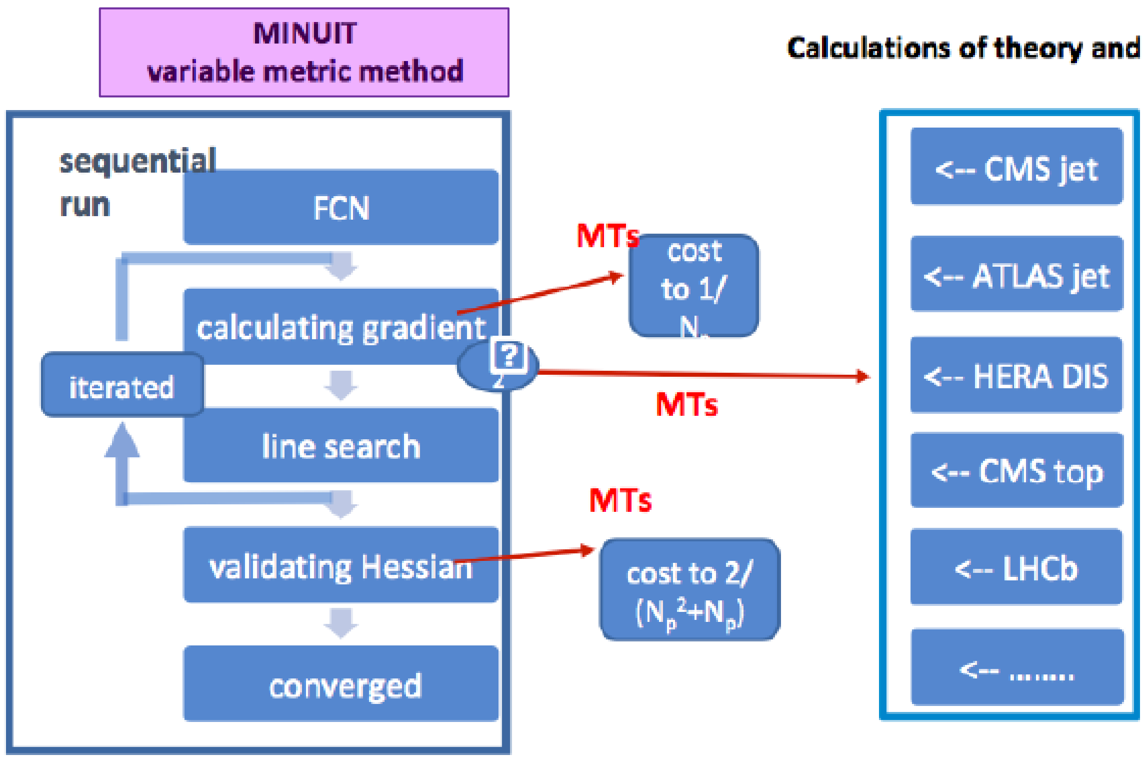
Example: Multi-threaded PDF fitting code in CT18 NNLO analysis (arXiv:1912.10053)

upgrade to a parallelized version of the fitting code, two-layer parallelization: 1. through rearrangement of the minimization algorithm; 2. via redistribution of the data sets

CERNLIB and ROOT 5.36 are still in use

Transiting to new libraries (TensorFlow, NumPy,…)

Benchmarking and quality control?



Layer 1: after all a factor of 4~5 improvement on speed is achieved!

Layer 2: further speed up by a factor of 2, depending on data sets included

External modules written in Fortran 77/95, C++, Python,…

Developed by CTEQ or external groups

Quality control?

Parallelized implementation by Jun Gao

Transition to modern codes requires significant investment and expertise

SMU ManeFrame II HPC cluster

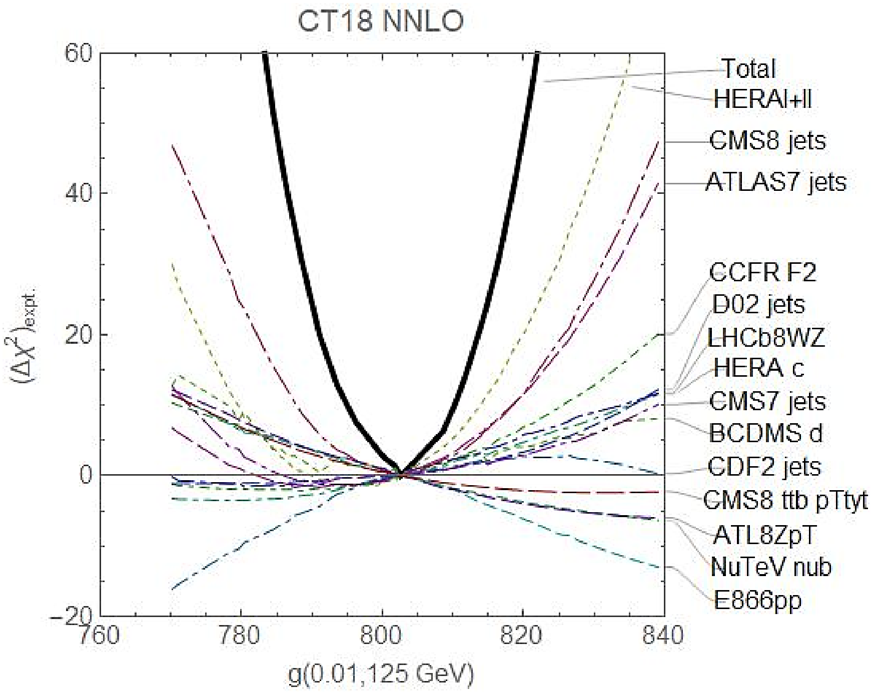
https://www.smu.edu/OIT/Services/HPC

HPC infrastructure at SMU and MSU was essential for completion of the CT18 global PDF analysis

Availability of HPC opens avenues for generating PDFs with alternative assumptions (resummations, EW corrections, alternative selections of experimental data,…) and detailed studies of PDF uncertainties

Computational Ability	870 TFLOPS
Number of Nodes	354
Intel CPU Cores (AVX2)	11,276
Total Accelerator Cores	275,968
Total Memory	120 TB (122,880 GB)
Node Interconnect Bandwidth	100 Gb/s
Scratch Space	2.8 PB (2,867 TB)
Archive Capabilities	Yes
Operating System	CentOS 7

Example, computing requirements, CT18 study



A Lagrange Multiplier scan…

…offers a detailed picture of pulls from experiments on the CT18 gluon PDF in the Higgs production region

…instrumental for reducing PDF uncertainties

Intel Xeon E5-2695 v4 workstations, 18 cores/48 GB RAM per 1 fit

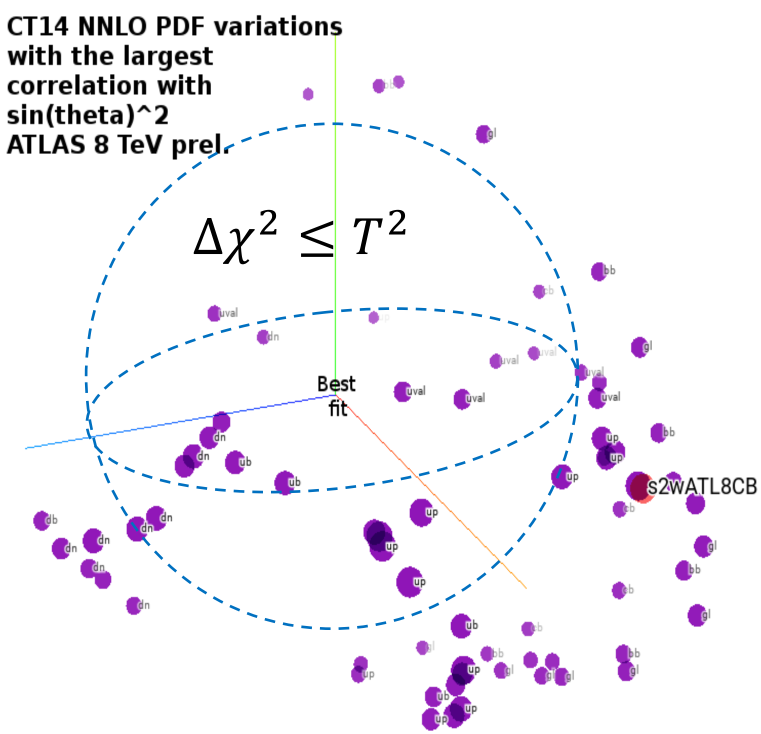
Memory management issues to read large ApplGrid/FastNLO tables

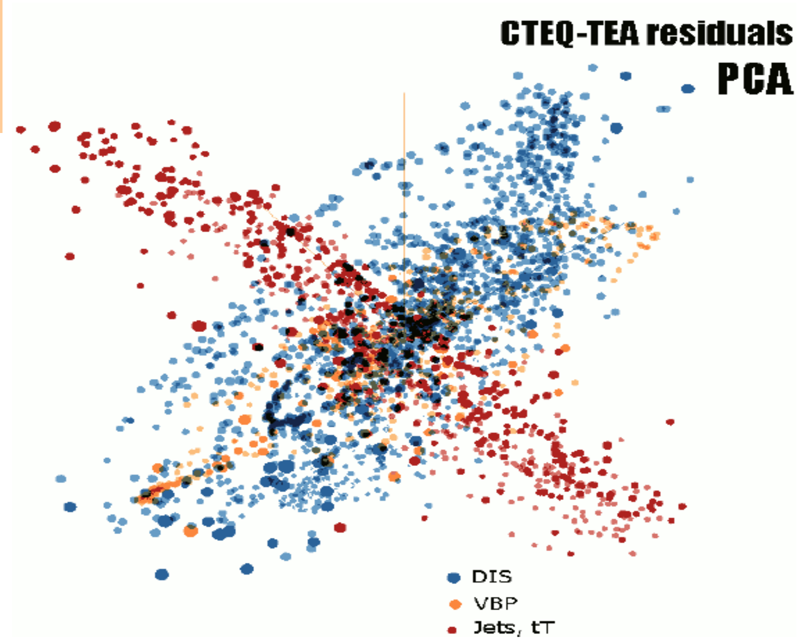
Task	Approximate core-hours
1 candidate NNLO fit	300-430
1 NNLO error set	1300
1 LM scan, for 1 point in x and Q	6500
6000+ fits we performed to study parametric, theoretical, methodological uncertainties	> 6000*300=$1.8 \cdot 10^6$

New tools for the analysis of N-dimensional distributions…

… offer great alternative alternatives to traditional fits. We increasingly use AI/ML for (un)supervised learning, classification, dimensionality reduction in TensorFlow, Mathematica,…

Outcomes of CT14 fits analyzed in TensorFlow Embedding Projector,
B. Wang, T. Hobbs, et al., PRD, **98 (2018) 094030**





Principal Component Analysis (PCA) visualizes the 56-dim. manifold by reducing it to 10 dimensions (à la META PDFs)

Final remarks and requests

Progress toward N3LO accuracy requires numerous advancements beyond the current frameworks for global QCD fits, on the top of implementing N3LO radiative contributions.

Many advancements will take time and computational resources already at NNLO to satisfy the goals of the HL-LHC physics program.

They also require the community culture for enforcing baseline accuracy in all experimental measurements, theoretical computations, and software contributing to the PDF fits.

Future lepton-hadron colliders (EIC, LHeC,…) can greatly improve the accuracy of PDFs in the HL-LHC era.

CPU power for generating higher order ApplGrids and FastNLO tables, replacing K-factor tables in the fits is critical as we push to higher orders.

Docker and Singularity tools are also key as this allow us to distribute uniform environments for the analysis purposes.

The AI/ML techniques increasingly are powerful alternatives to traditional fits.

All these efforts require HPC resources, training, and support.