EuCAIFCon 2024

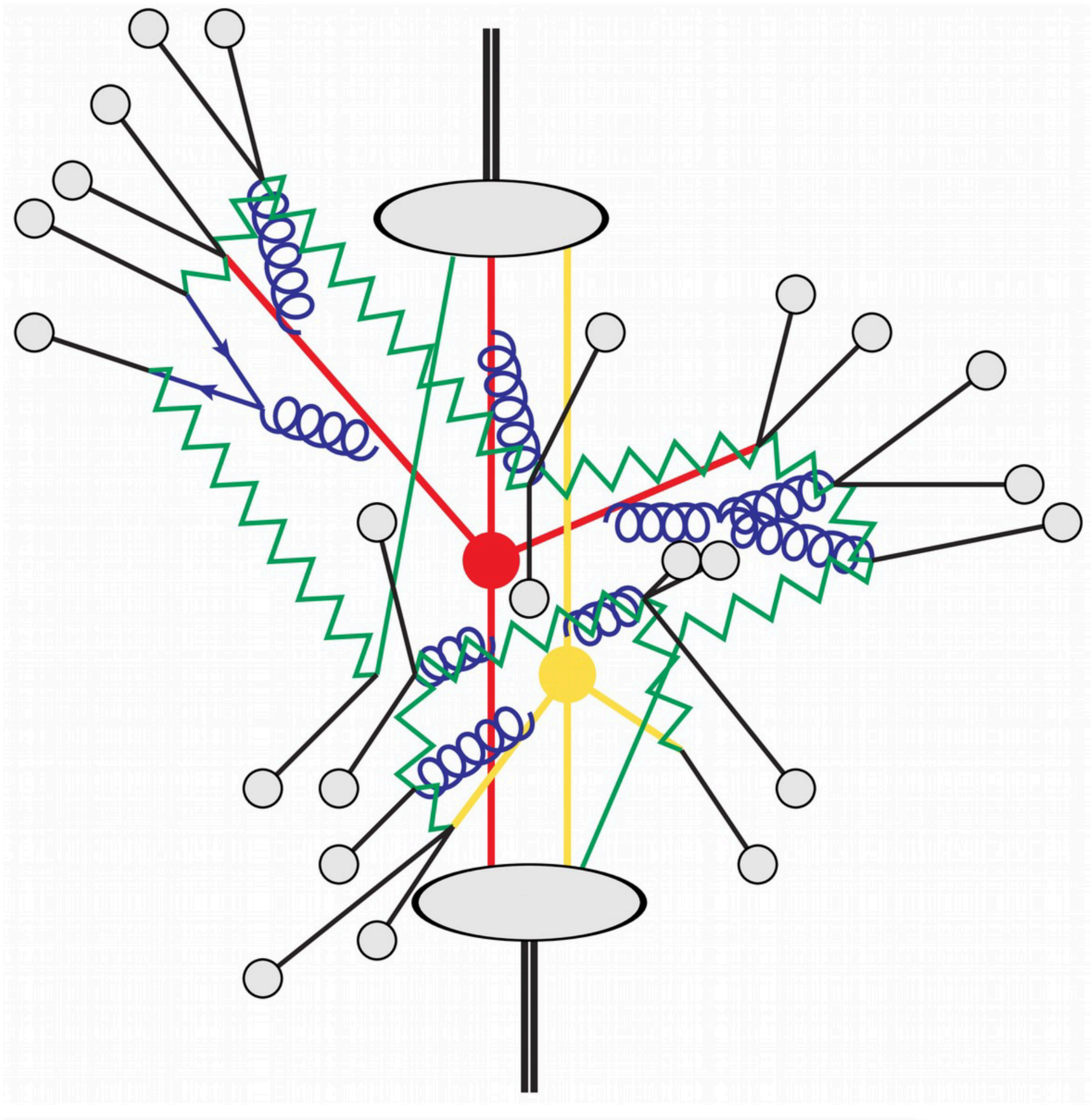
Hyperparameter optimisation of neural networks for proton structure analyses

Tanjona R. Rabemananjara on behalf of
Juan Cruz-Martinez, Aron Jansen, Gijs van Oord, Carlos M. Rocha, Juan Rojo, Roy Stegeman

VU Vrije Universiteit Amsterdam
Nikhef
NNPDF
netherlands eScience center

OUTLINE OF THE TALK:

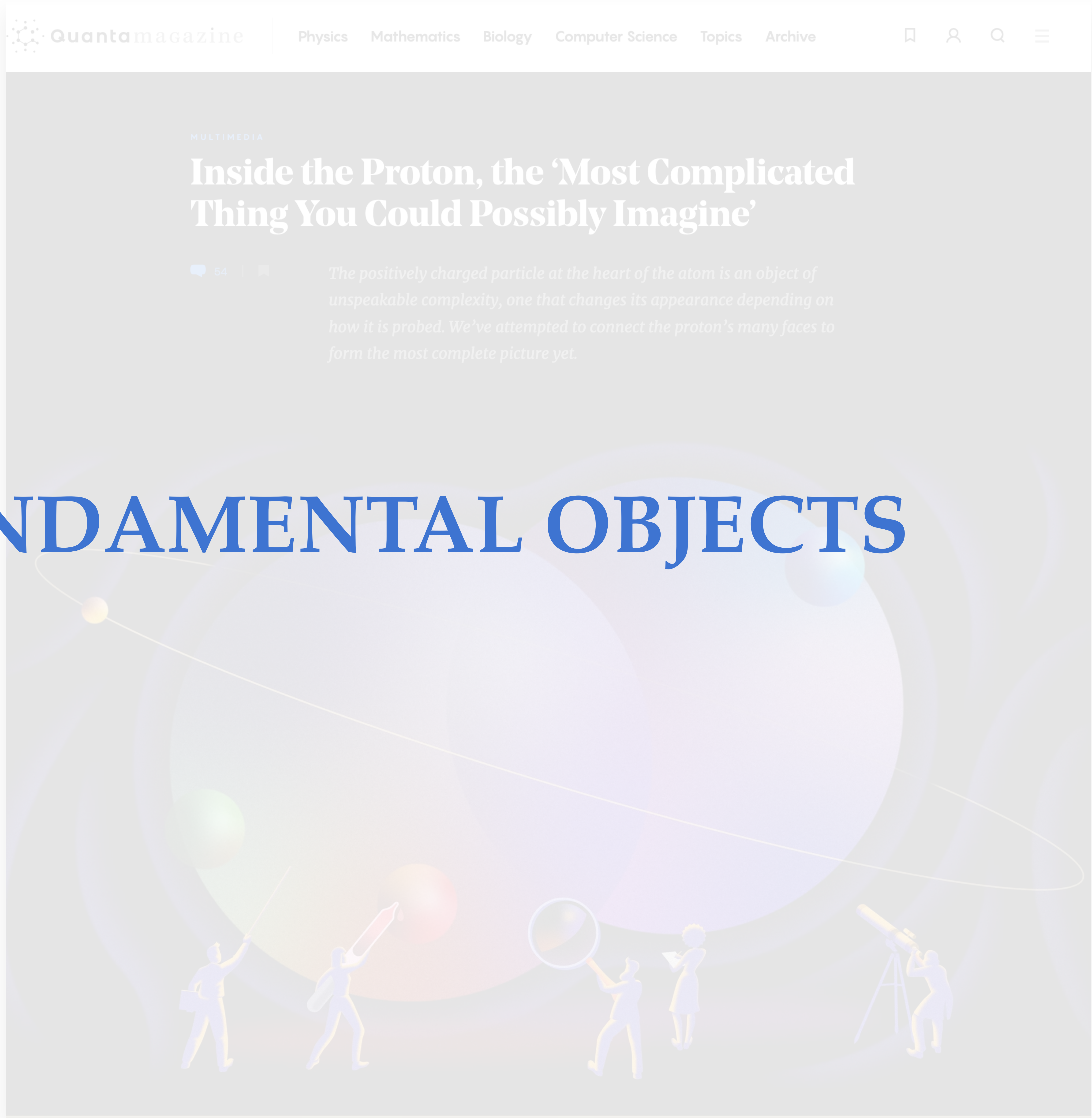
1. Introduction and Motivations
2. The NNPDF Proton PDF determination
3. Hyperparameter Optimisation
4. Uncertainty Validations
5. Conclusions & Outlook



Why study the structure of the Protons?

We want answers to fundamental Questions:

- **Imaging of the Protons:** How are quarks and gluons distributed both in space and in momentum? How do nuclear properties emerge from their interactions?
- **Proton Spin mystery:** How are spins of the sea quarks and gluons distributed inside the protons? How much of the proton spin come from the Orbital motion?
- **Gluon saturation:** Does Gluon density in nuclei exhibit Saturation at high-energy? How does a dense nuclear environment affect the quarks and gluons, their correlations and their interactions?
- **What is the origin of Mass? Can new Physics hide inside the protons? etc.**

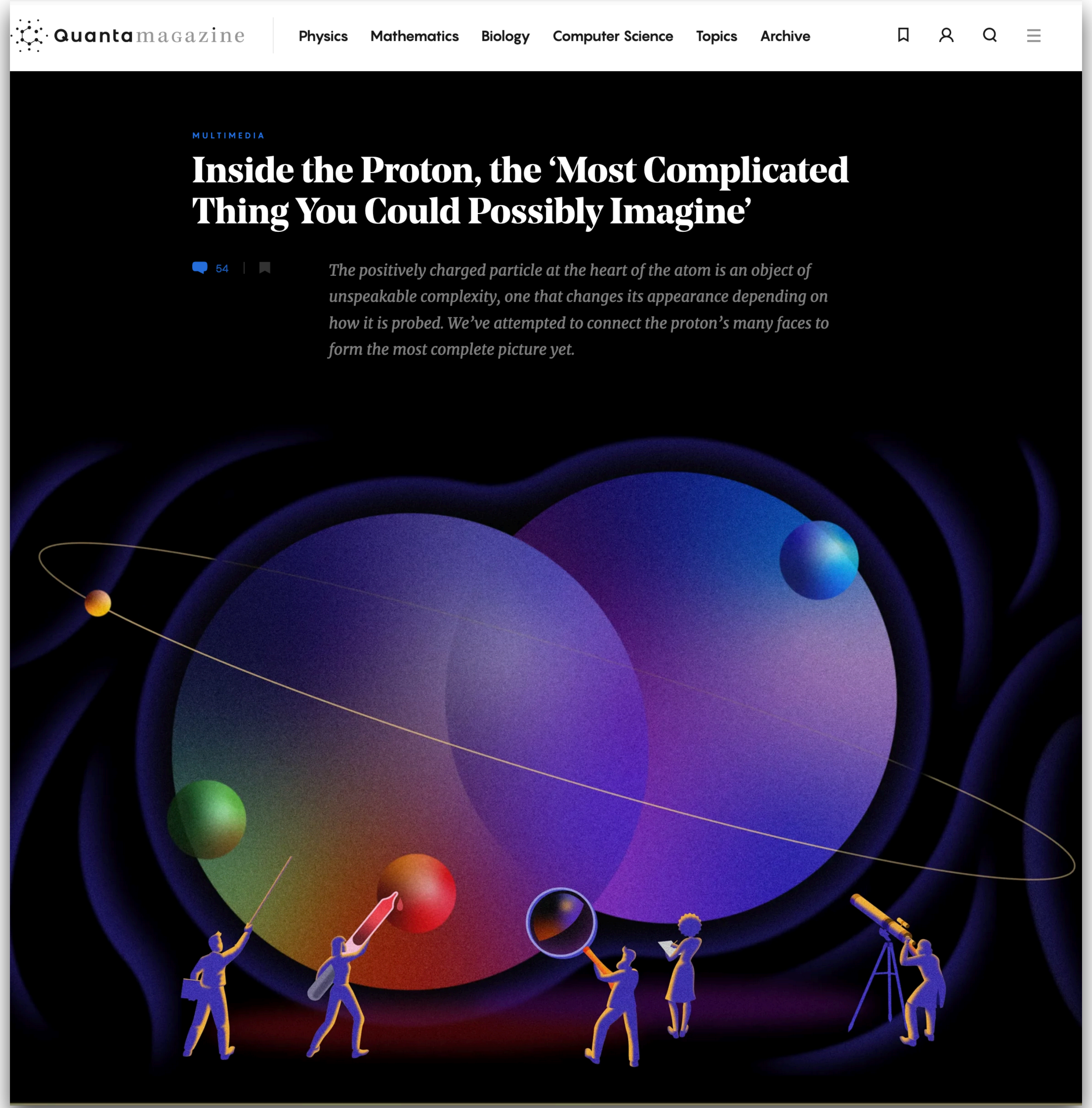


PROTONS ARE NOT FUNDAMENTAL OBJECTS

Why study the structure of the Protons?

We want answers to fundamental Questions:

- **Imaging of the Protons:** How are quarks and gluons distributed both in space and in momentum? How do nuclear properties emerge from their interactions?
- **Proton Spin mystery:** How are spins of the sea quarks and gluons distributed inside the protons? How much of the proton spins comes from the Orbital motion?
- **Gluon saturation:** Does Gluon density in nuclei exhibit Saturation at high-energy? How does a dense nuclear environment affect the quarks and gluons, their correlations and their interactions?
- **What is the origin of Mass? Can new Physics hide inside the protons? etc.**

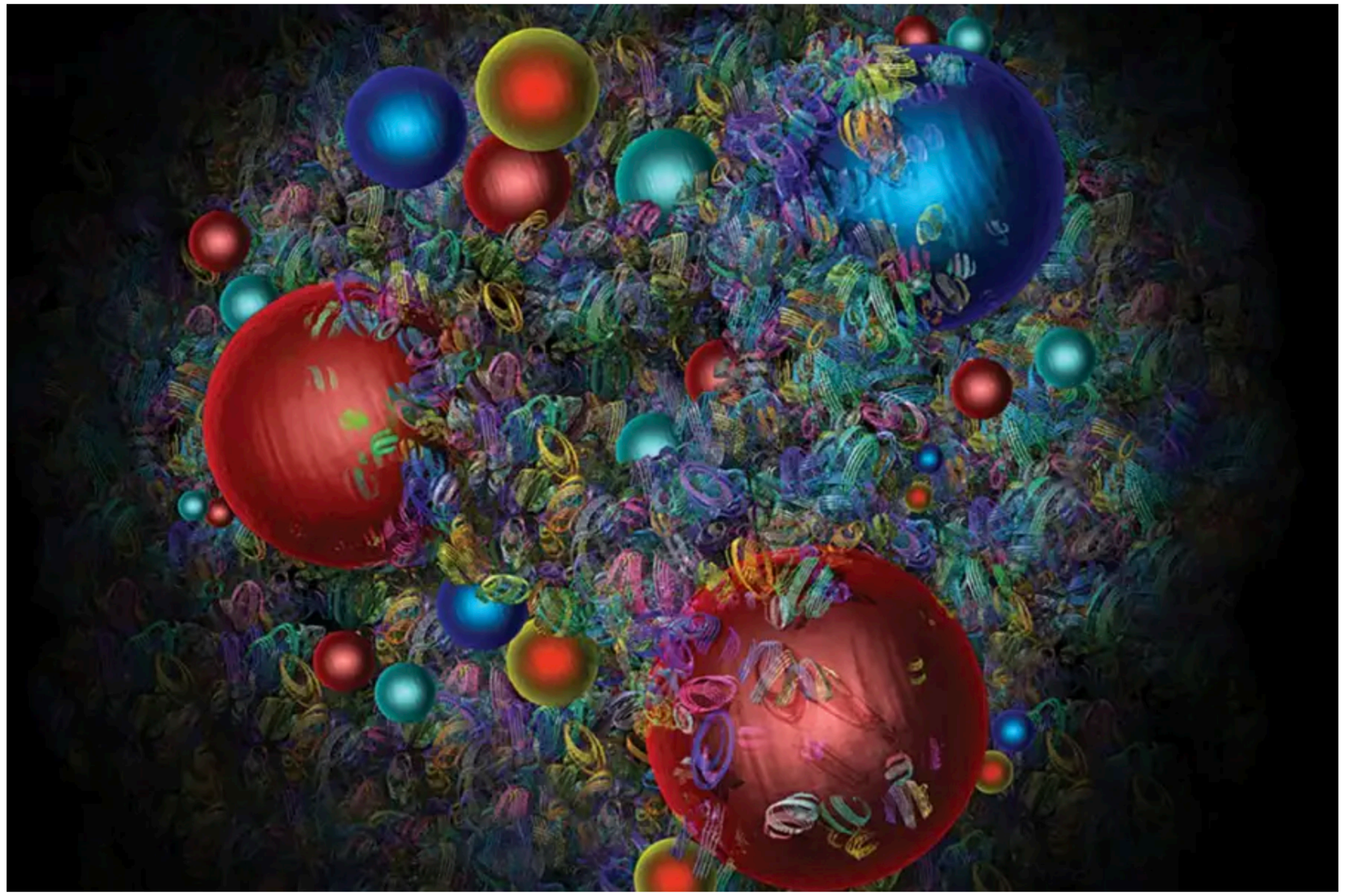


Discoveries using Protons as a Gateway

NewScientist Sign in | Enter search keywords

Physics

Physicists surprised to discover the proton contains a charm quark



See Juan Rojo's talk on "AI-driven discovery of charm quarks in the proton" this Thursday in WG5.3

SCIENTIFIC AMERICAN | SUBSCRIBE

THE SCIENCES

Proton Spin Mystery Gains a New Clue

Physicists long assumed a proton's spin came from its three constituent quarks. New measurements suggest particles called gluons make a significant contribution

Quanta magazine | Physics | Mathematics | Biology | Computer Science | Topics | Archive

QUANTUM PHYSICS

Decades-Long Quest Reveals Details of the Proton's Inner Antimatter

symmetry | dimensions of particle physics | topics | follow +

A joint Fermilab/SLAC publication

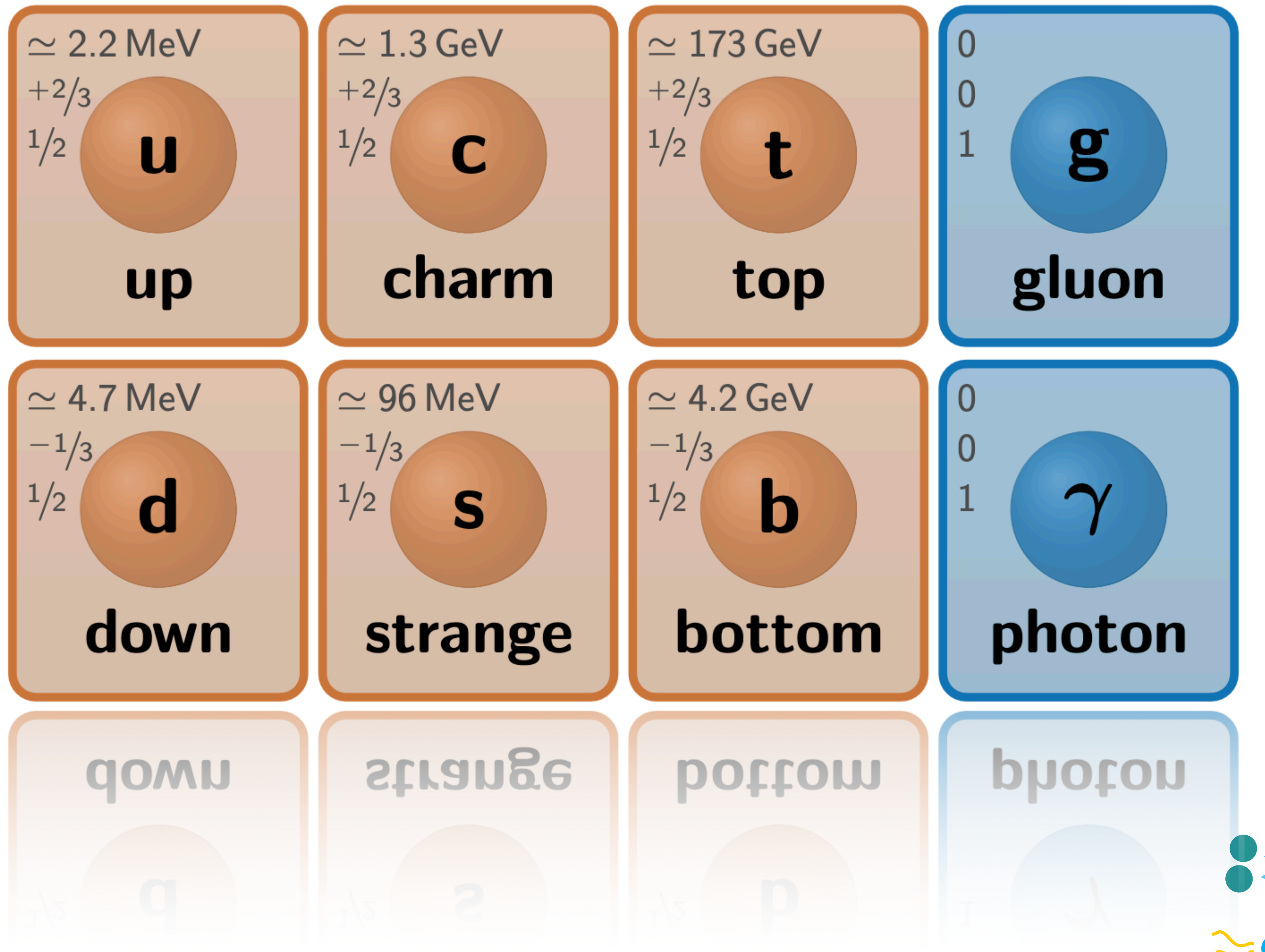
From quark soup to ordinary matter

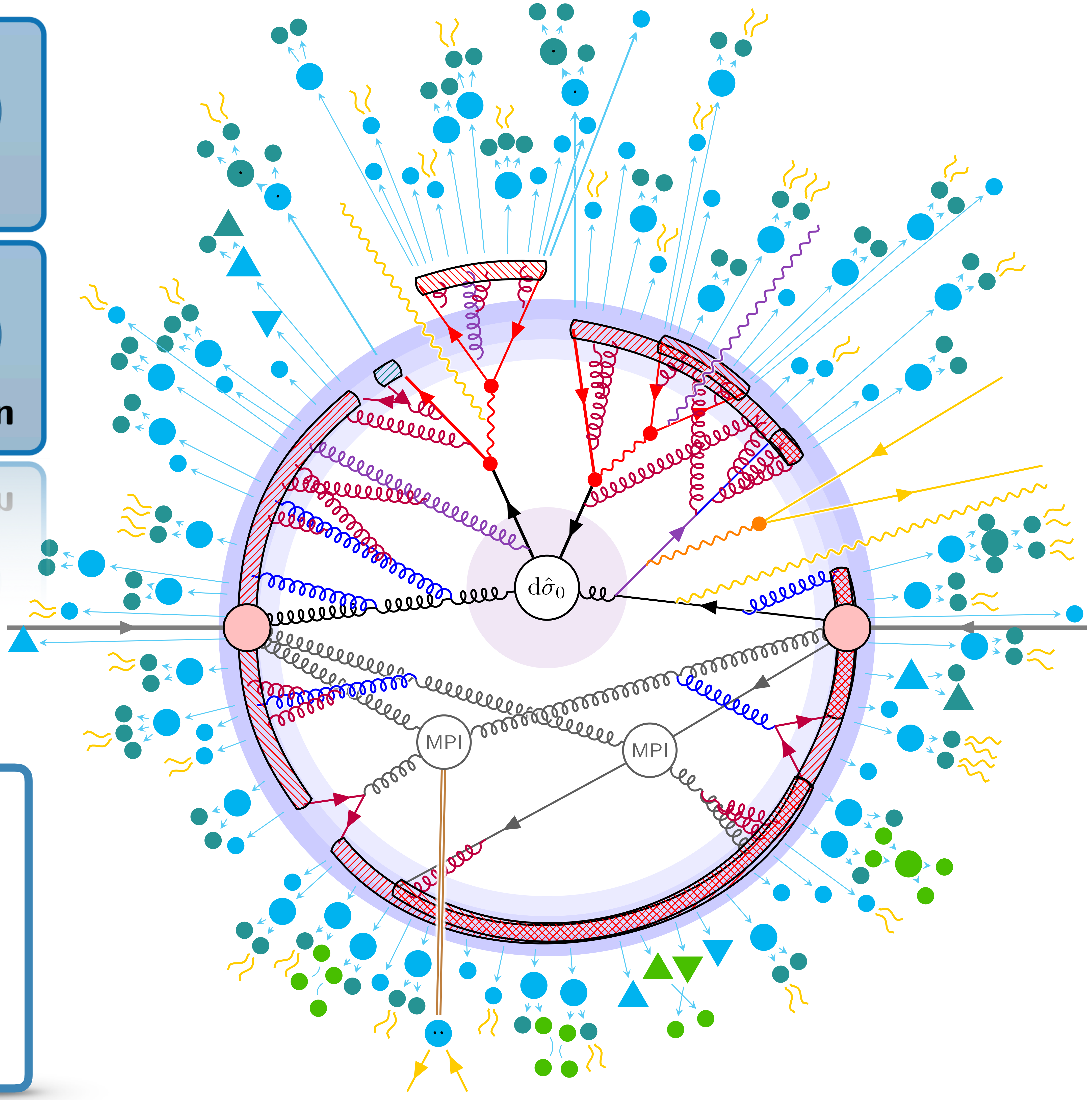
04/10/14

| By Karen McNulty Walsh, Brookhaven National Laboratory

Scientists have gained new insight into how matter can change from a hot soup of particles to the matter we know today.

Predictions in High-Energy Physics

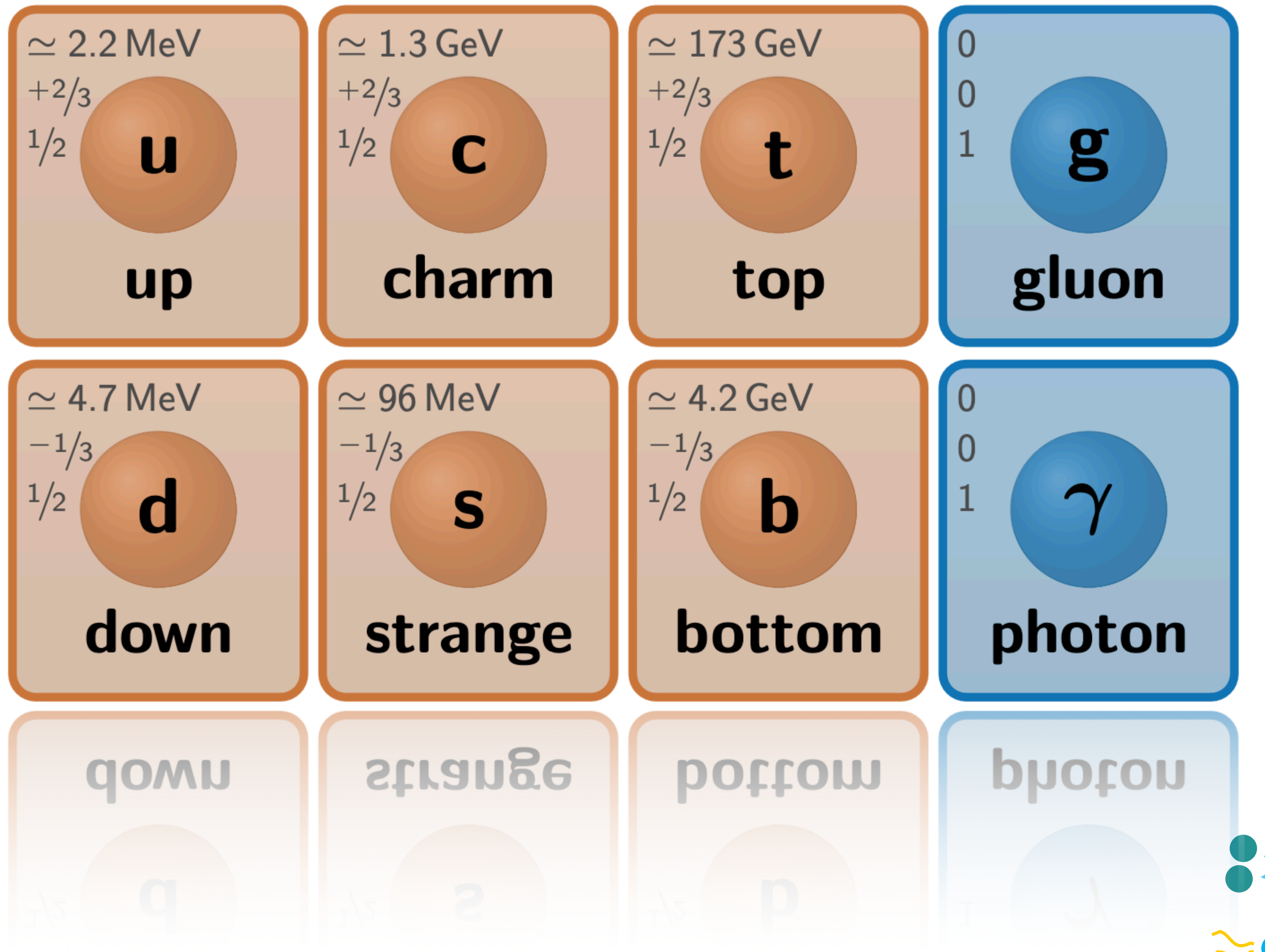


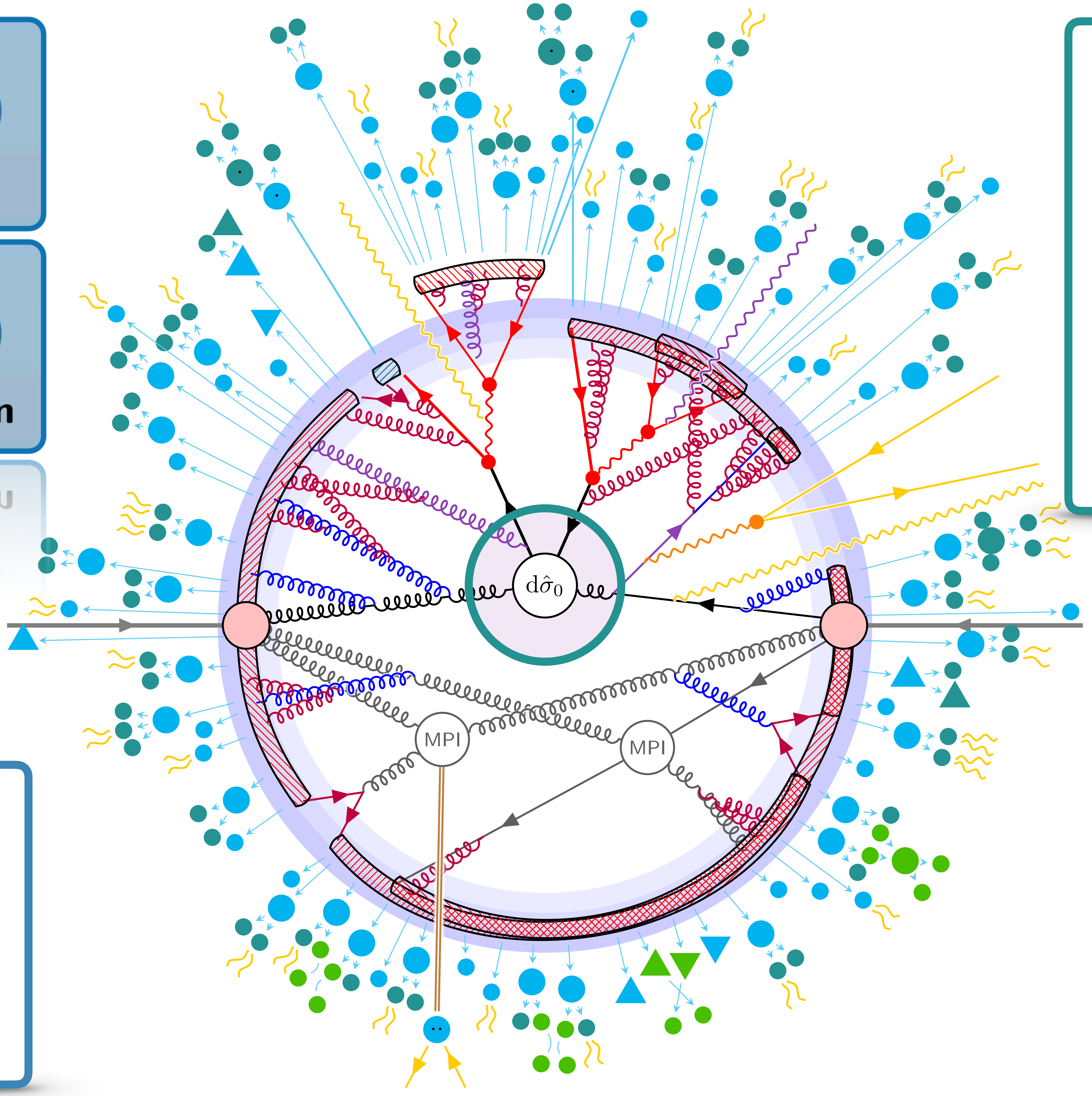


Experiments measure an Observable $\mathcal{O}$ (cross-section, decay rates, etc.):

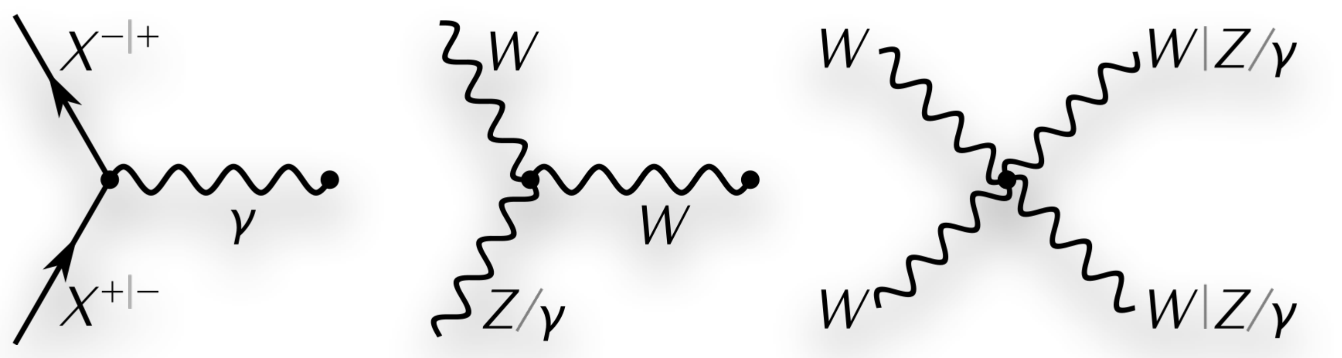
$$\mathcal{O} = \sum \hat{\sigma}_0 \otimes \text{PDF}$$

Predictions in High-Energy Physics





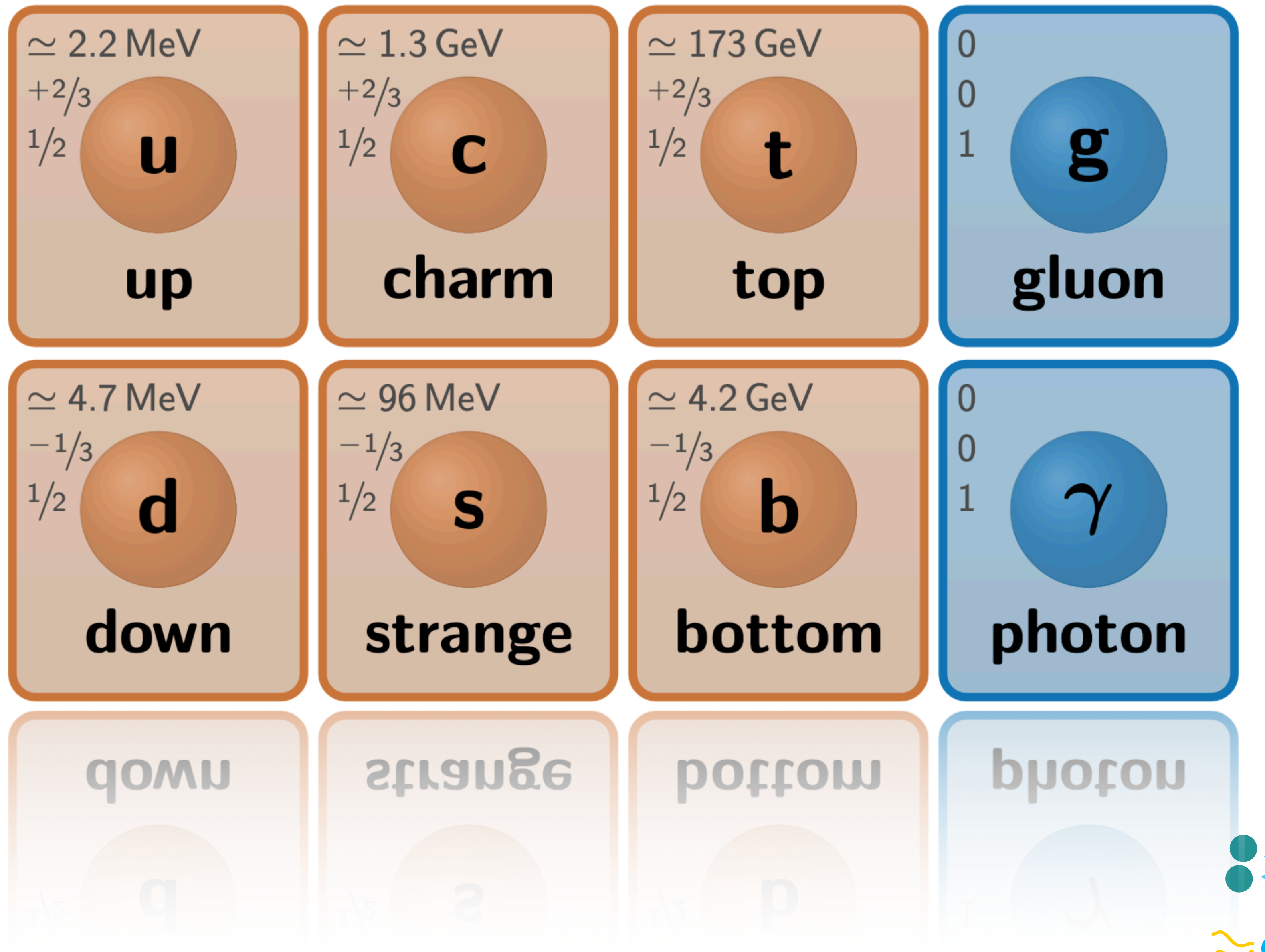
Hard scattering $\hat{\sigma}_0$: encodes short-range interactions; computed from first principles.



Experiments measure an Observable $\mathcal{O}$ (cross-section, decay rates, etc.):

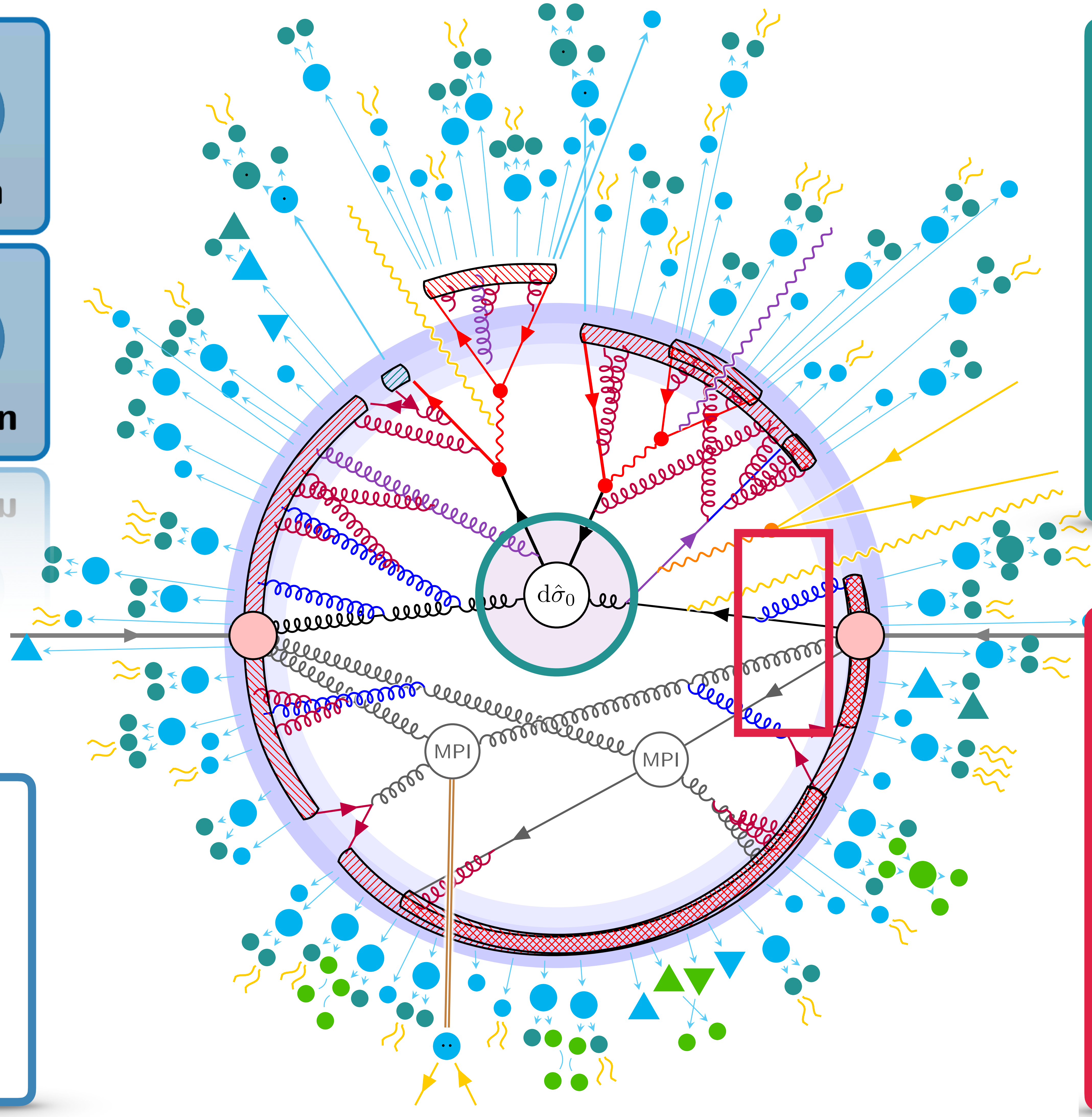
$$\mathcal{O} = \sum \hat{\sigma}_0 \otimes \text{PDF}$$

Predictions in High-Energy Physics

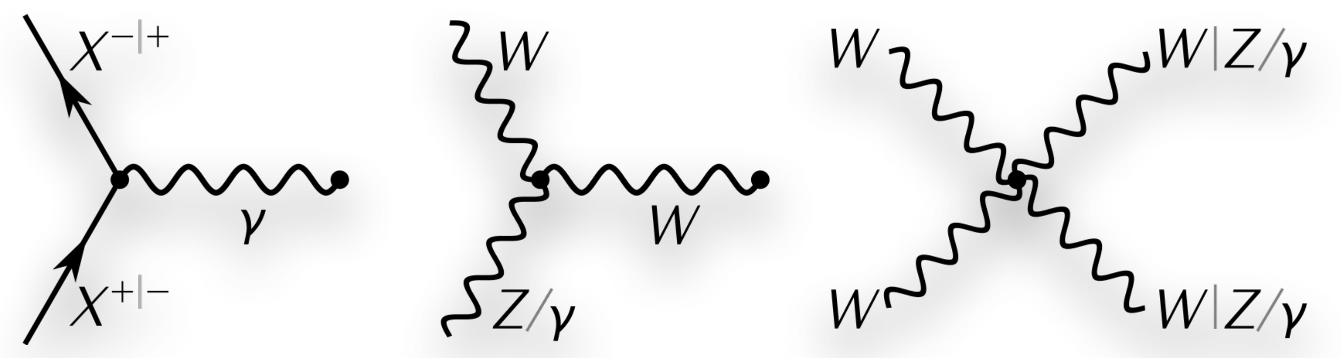


Experiments measure an Observable $\mathcal{O}$ (cross-section, decay rates, etc.):

$$\mathcal{O} = \sum \hat{\sigma}_0 \otimes \text{PDF}$$



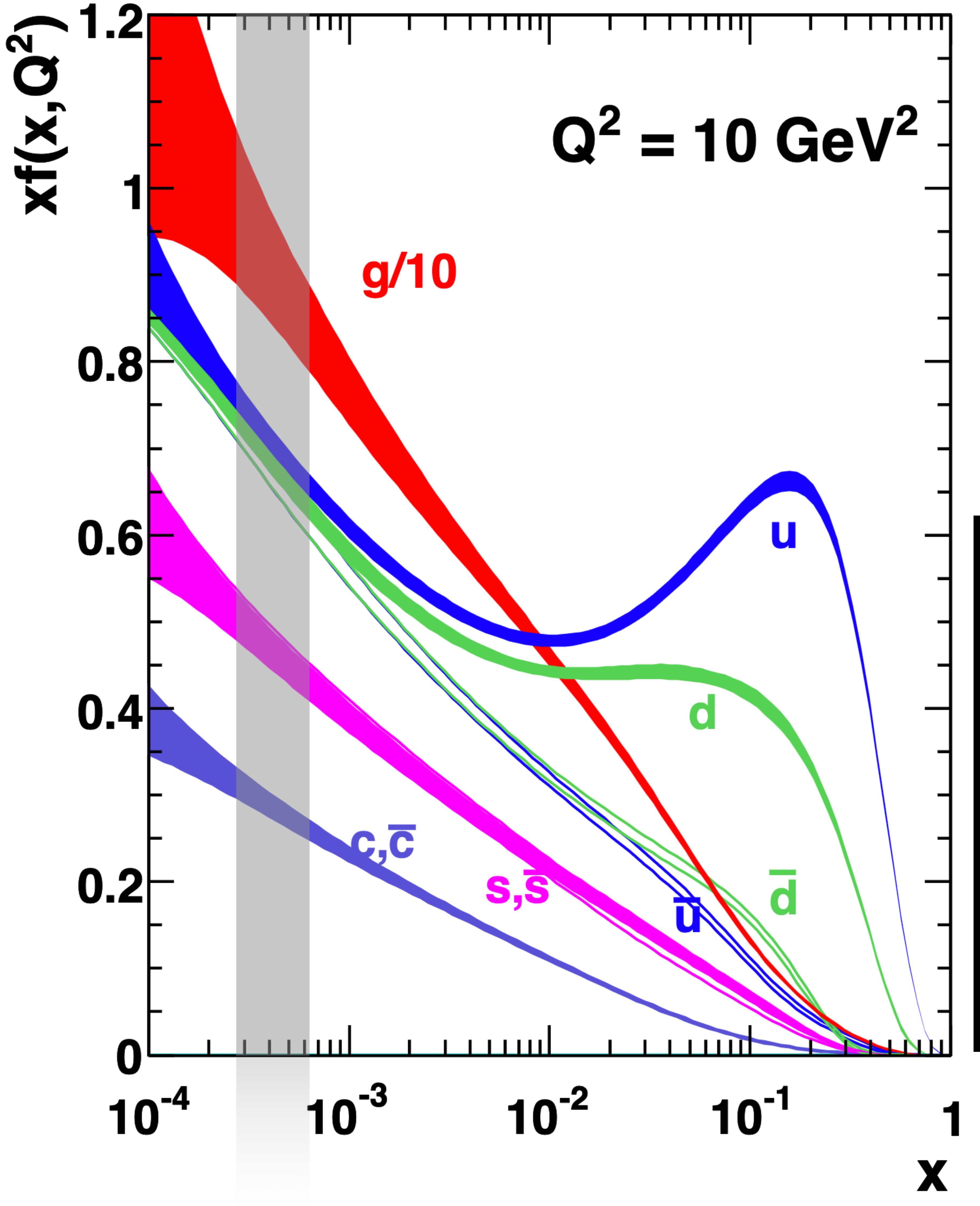
Hard scattering $\hat{\sigma}_0$: encodes short-range interactions; computed from first principles.



Parton Distribution Functions (PDFs): encodes long-range non-perturbative interactions; cannot be computed from first principle and have to be determined from experimental Data.

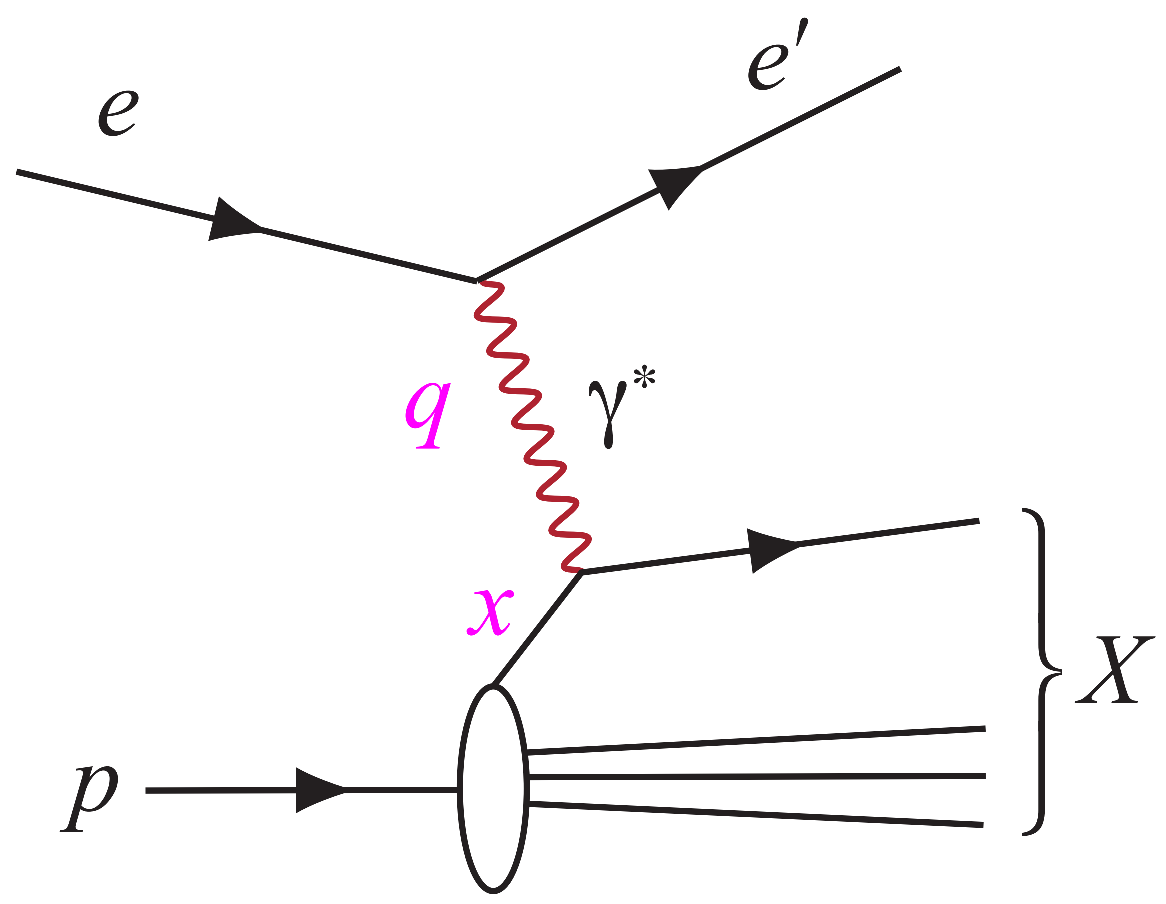
PDFs are Universal

How does the inside of a Proton look like?

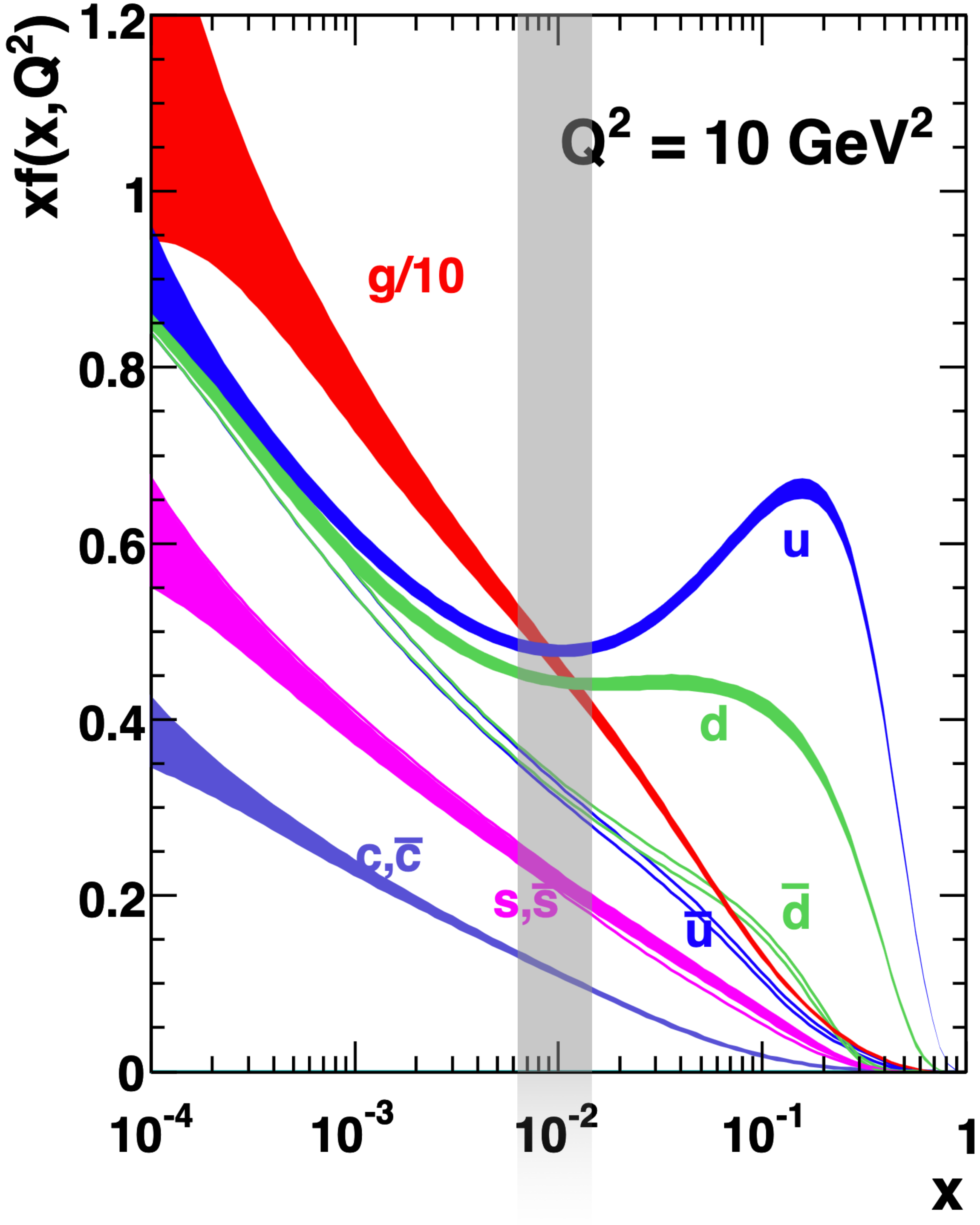


The content of a proton depends on how we look at it.

- **(Spatial) Resolution** $(Q^2 = -q^2)$: energy/momentum transfer
- **Shutter speed** $(1/x)$: fraction of momentum of the proton carried out

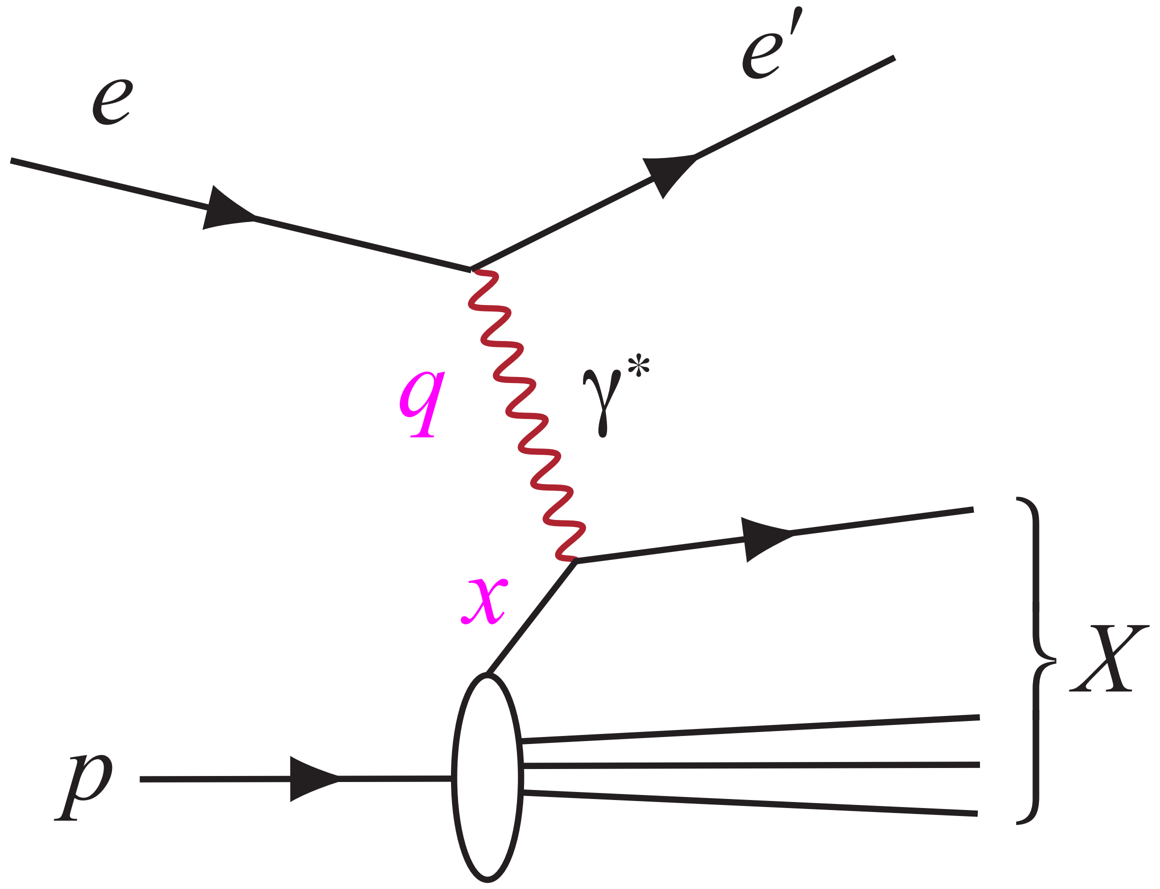


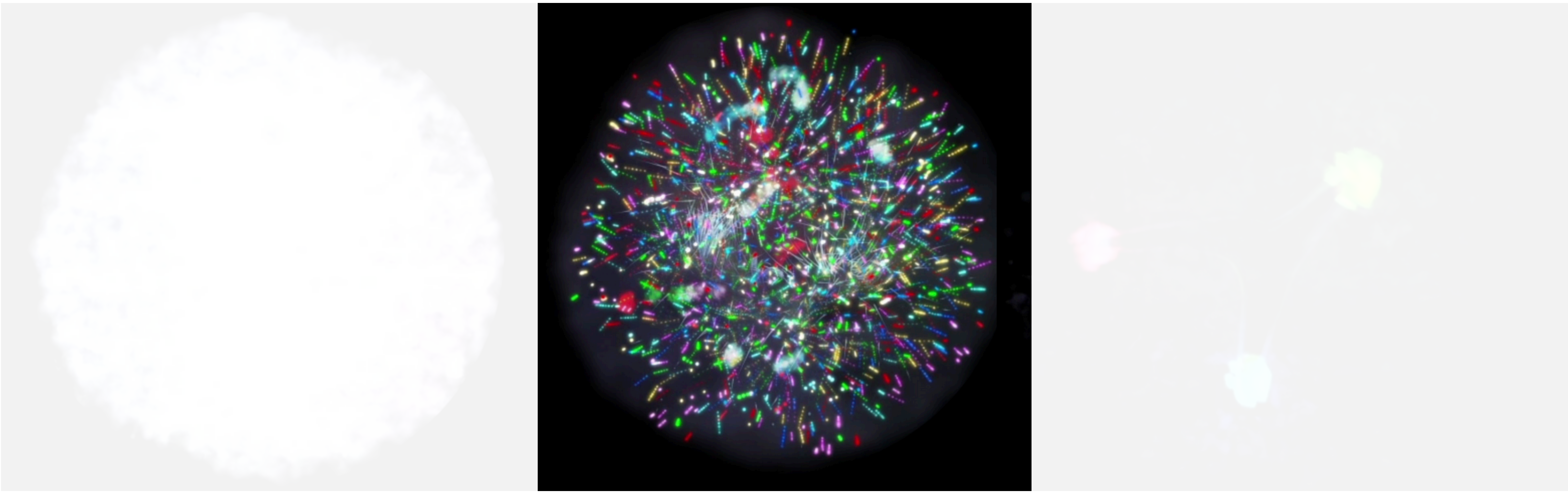
How does the inside of a Proton look like?



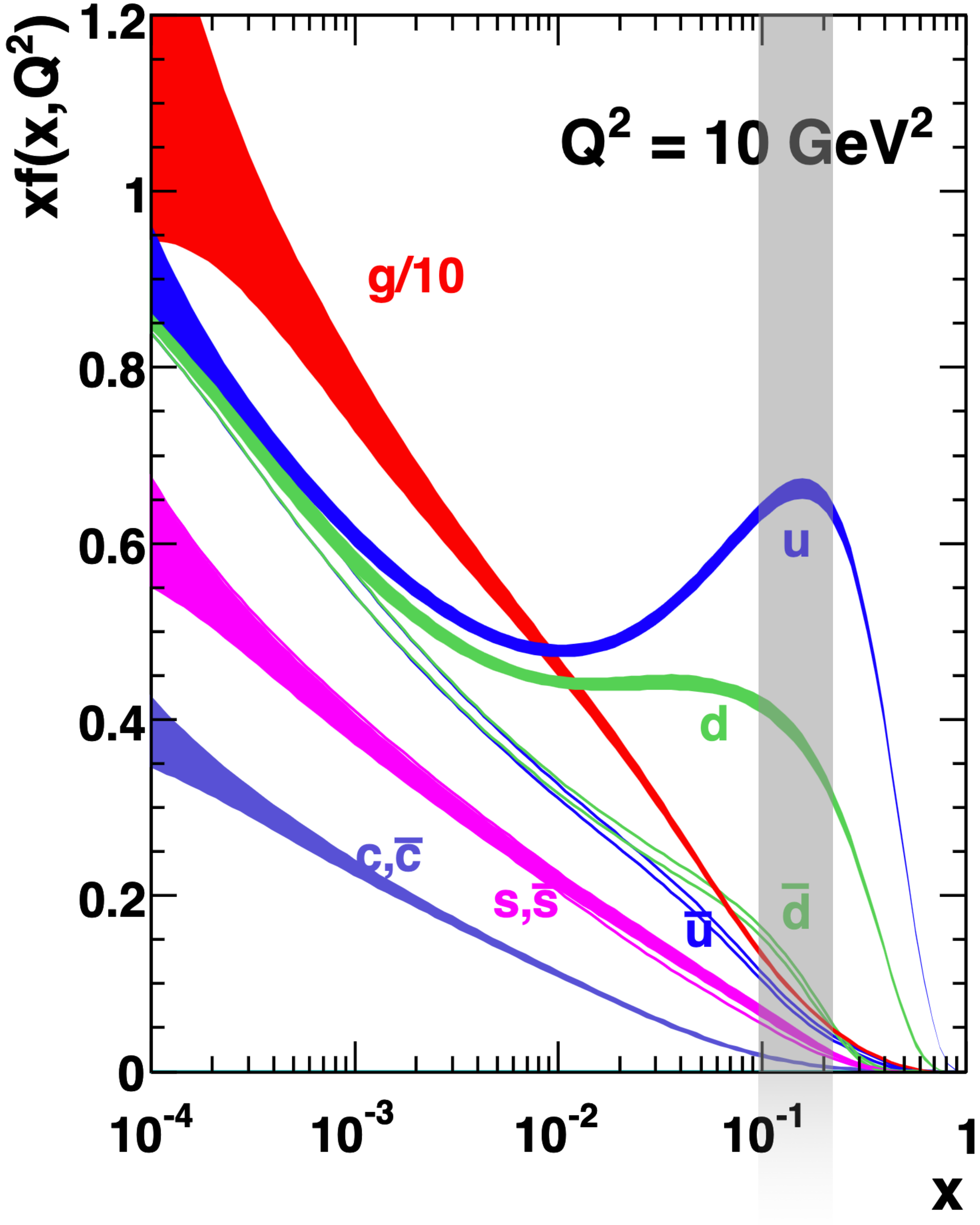
The content of a proton depends on how we look at it.

- (Spatial) Resolution ($Q^2 = -q^2$): energy/momentum transfer
- Shutter speed ($1/x$): fraction of momentum of the proton carried out



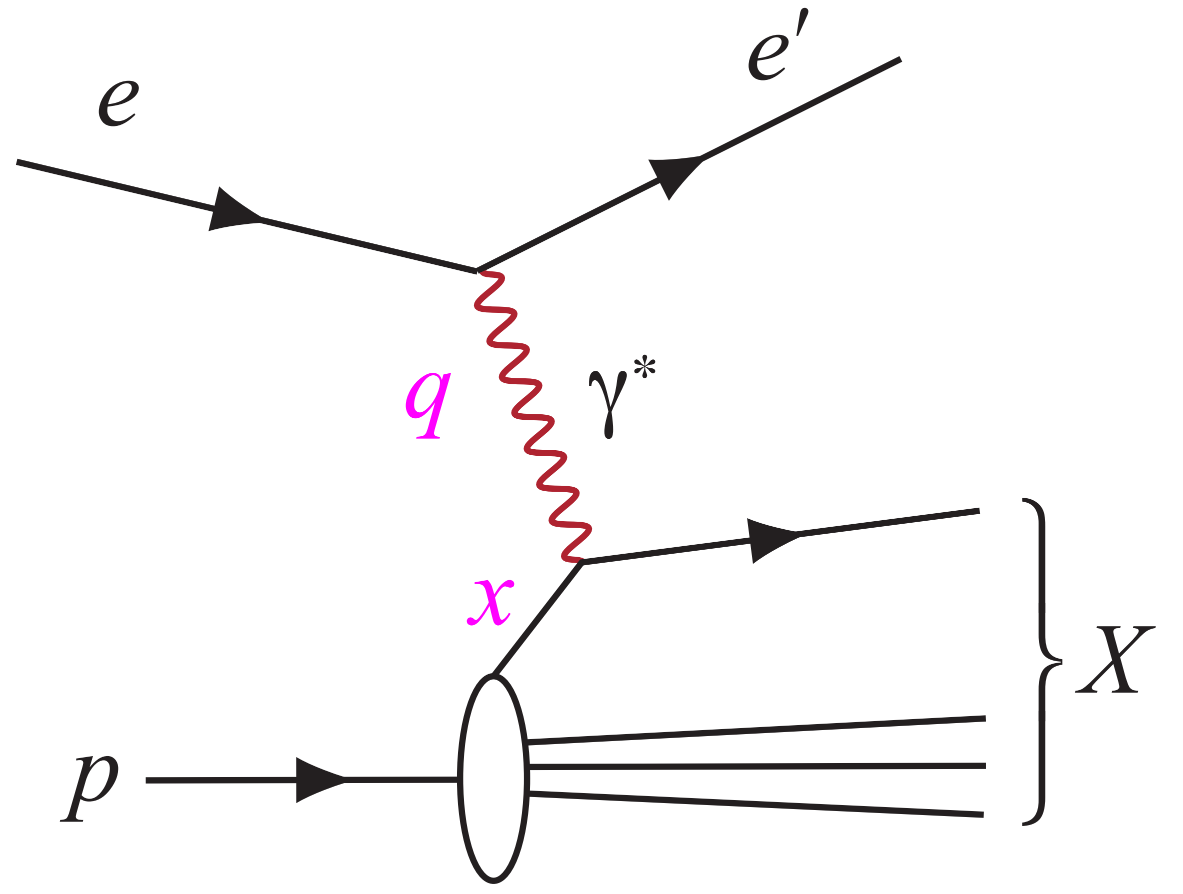


How does the inside of a Proton look like?



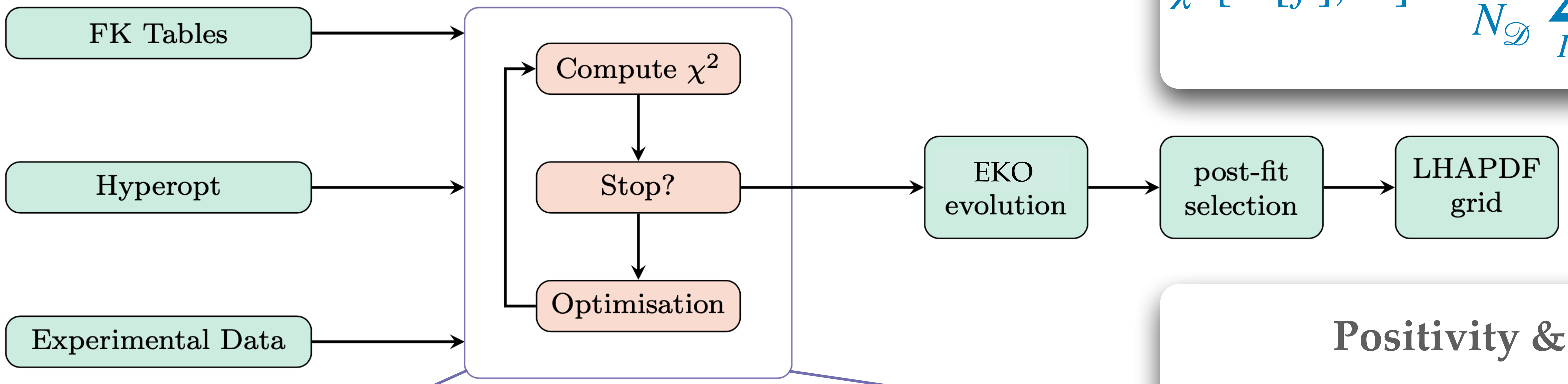
The content of a proton depends on how we look at it.

- (Spatial) Resolution ($Q^2 = -q^2$): energy/momentum transfer
- Shutter speed ($1/x$): fraction of momentum of the proton carried out



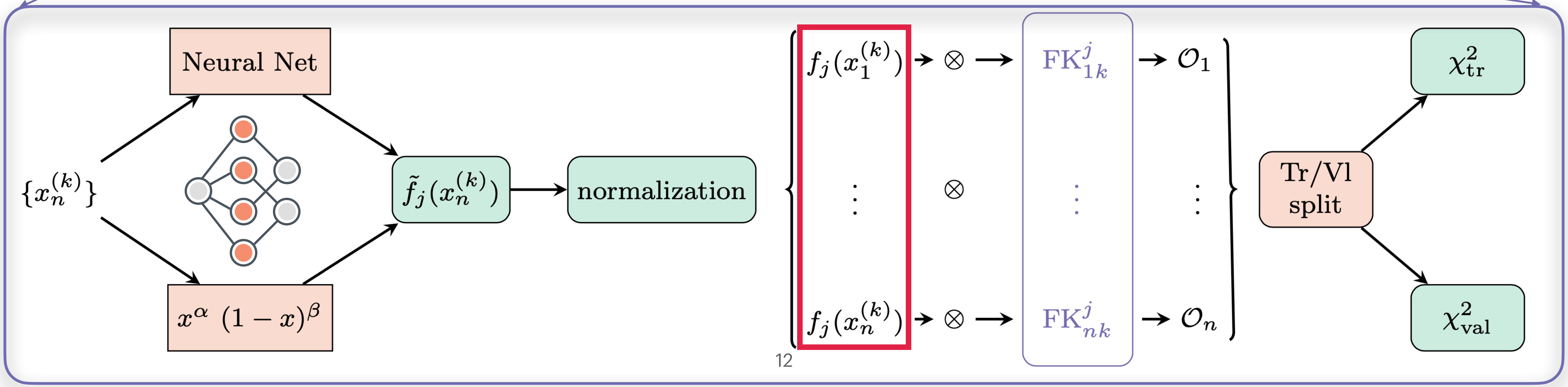
NNPDF Methodology in a Nutshell

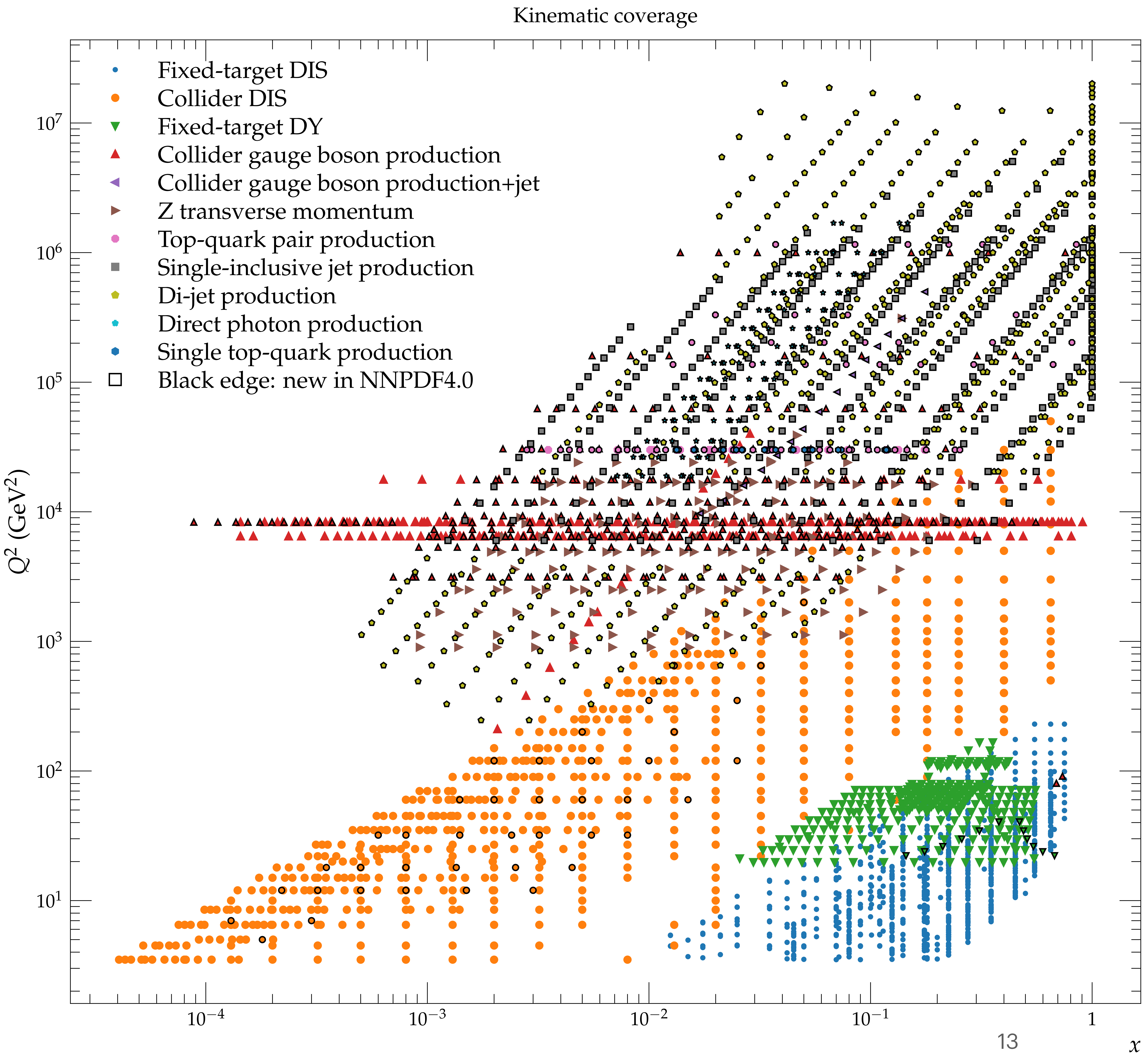
$$\chi^2[\mathcal{T}[f],\mathcal{D}] = \frac{1}{N_{\mathcal{D}}}\sum_{I,J}\left(\mathcal{T}_I[f]-D_I\right) C^{-1}_{IJ}\left(\mathcal{T}_J[f]-D_J\right)$$



Positivity & Integrability constraints:

$$\chi^2_{\mathrm{tot}} \to \chi^2_{\mathrm{tot}} + \sum_{k=1}^{8}\Lambda_k \sum_{i=1}^{n_i}\mathcal{F}\left(\tilde{f}_k\left(x_i,Q^2\right)\right)$$





Snapshot of the Experimental Datasets

- $\mathcal{O}(5000)$ **datapoints** that span a wide range of kinematic regions and probe various channels $\Longrightarrow$ **Large space of functional forms**
- Precision of the data reach the **percent level accuracy**; mostly from correlated systematic uncertainties
- Significant amount of the datasets ($\mathcal{O}(500)$ **datapoints**) were introduced in the NNPDF4.0 release (LHC Run II data)

Uncertainty Propagation

Monte Carlo Representation

Experimental uncertainties are propagated into the proton PDF fit by **fluctuating the central data** w.r.t. the uncertainties coming from the experimental inputs

$$\mathscr{D}_k = \mathscr{D}_k^{(0)} + \sum_{\ell=1}^{n_D} \sqrt{\mathrm{Cov}_{k\ell}} \times \delta_\ell$$

Instances of such samplings are called **"Pseudodata Replicas"**. Each of the pseudodata replica is then fitted to a NN with different training/validation random seeds.

The final output - which defines the PDF distribution - is an ensemble of PDF replicas.

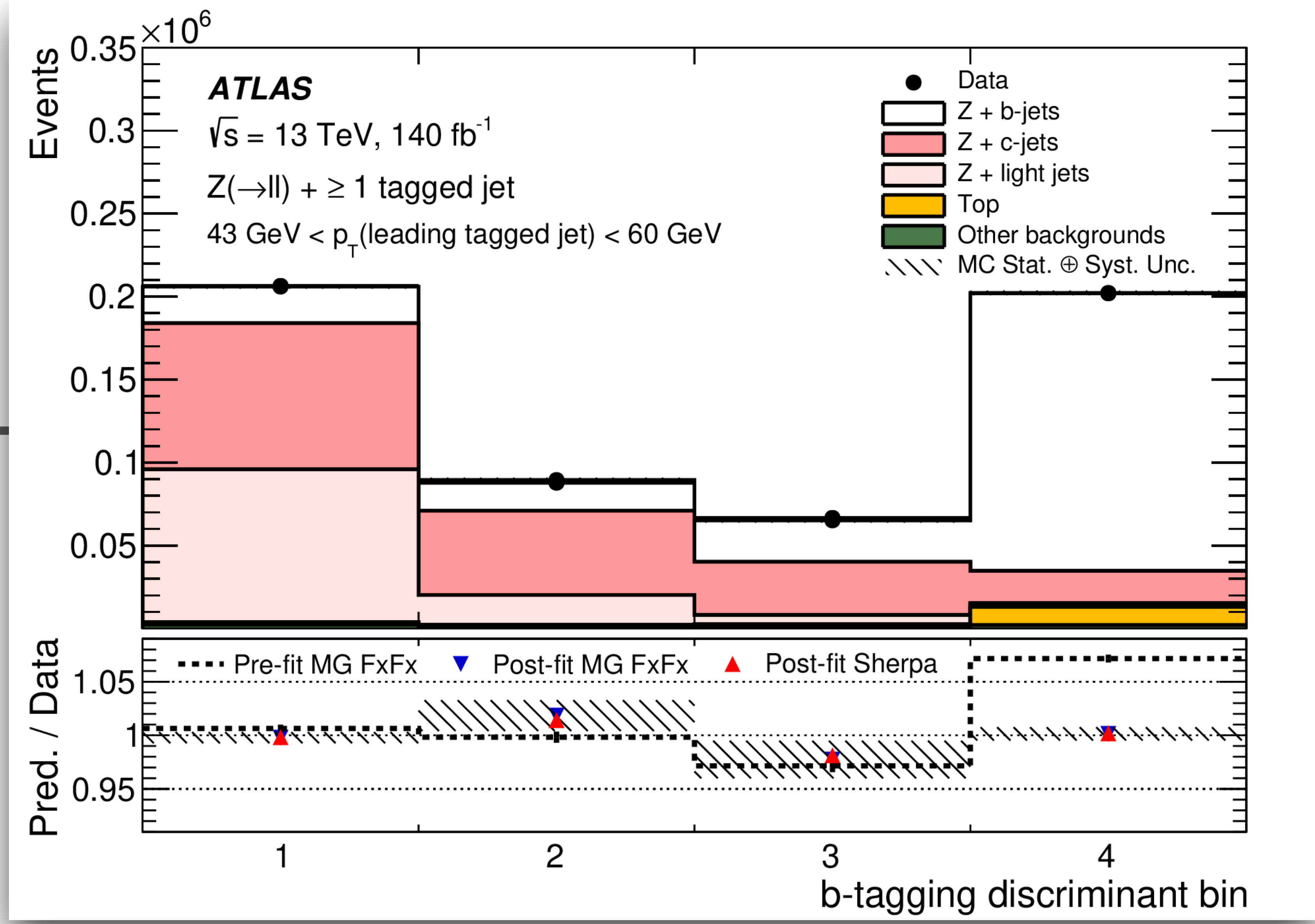
$r_1 \rightarrow \mathrm{NN}(\theta_1) \rightarrow \mathrm{PDF}_1$

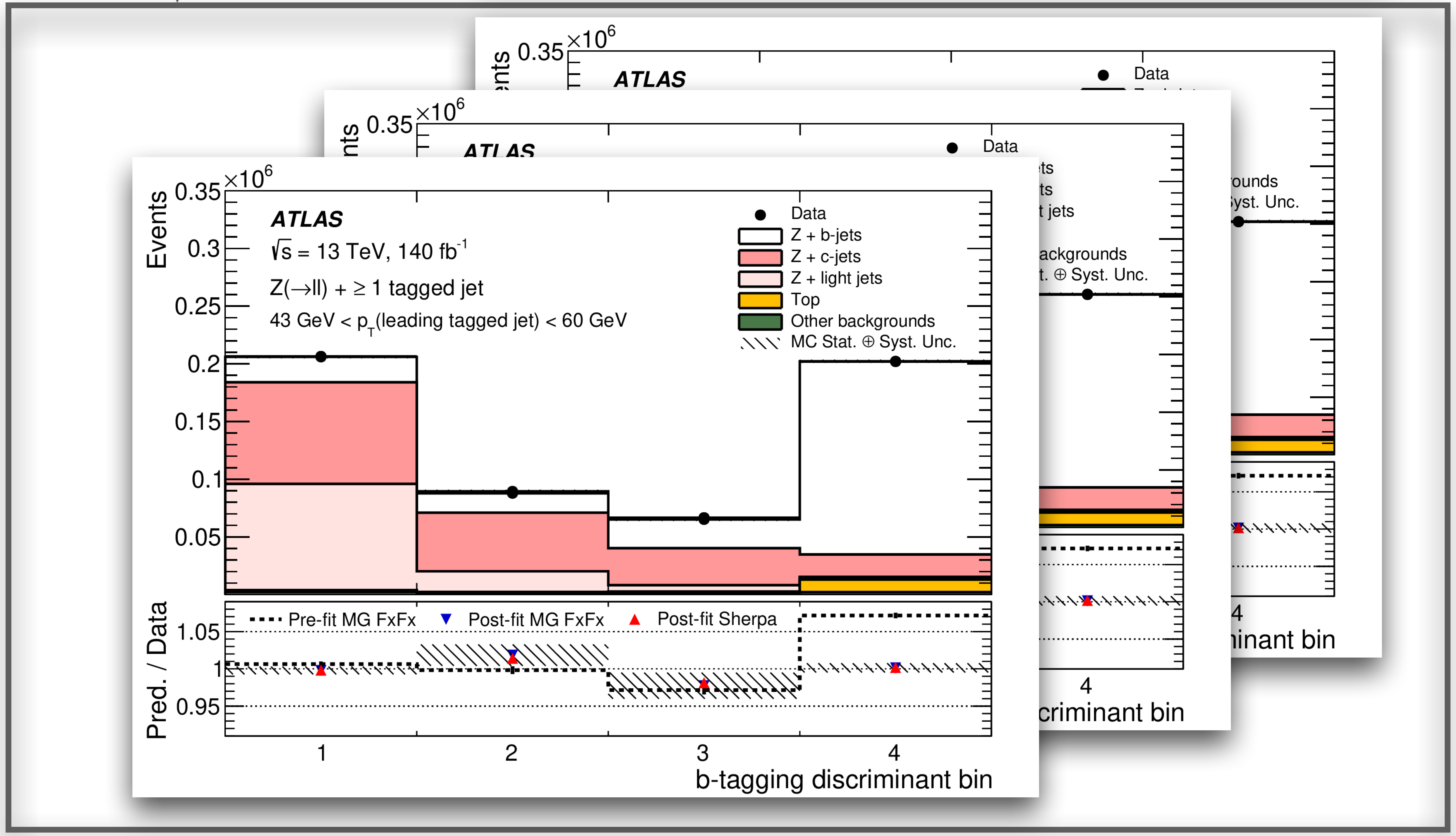
$r_2 \rightarrow \mathrm{NN}(\theta_2) \rightarrow \mathrm{PDF}_2$

$\vdots$

$r_n \rightarrow \mathrm{NN}(\theta_n) \rightarrow \mathrm{PDF}_n$

Generate Replicas of the Datasets





Hyperparameter Optimisation

One of the main reasons to resort to Neural Network (NN) was **to reduce biases** in defining a functional form, however:

❌ The hyperparameters that define the NN have to be chosen
❌ Random and/or Manual selection of hyperparameters are tedious and not guaranteed

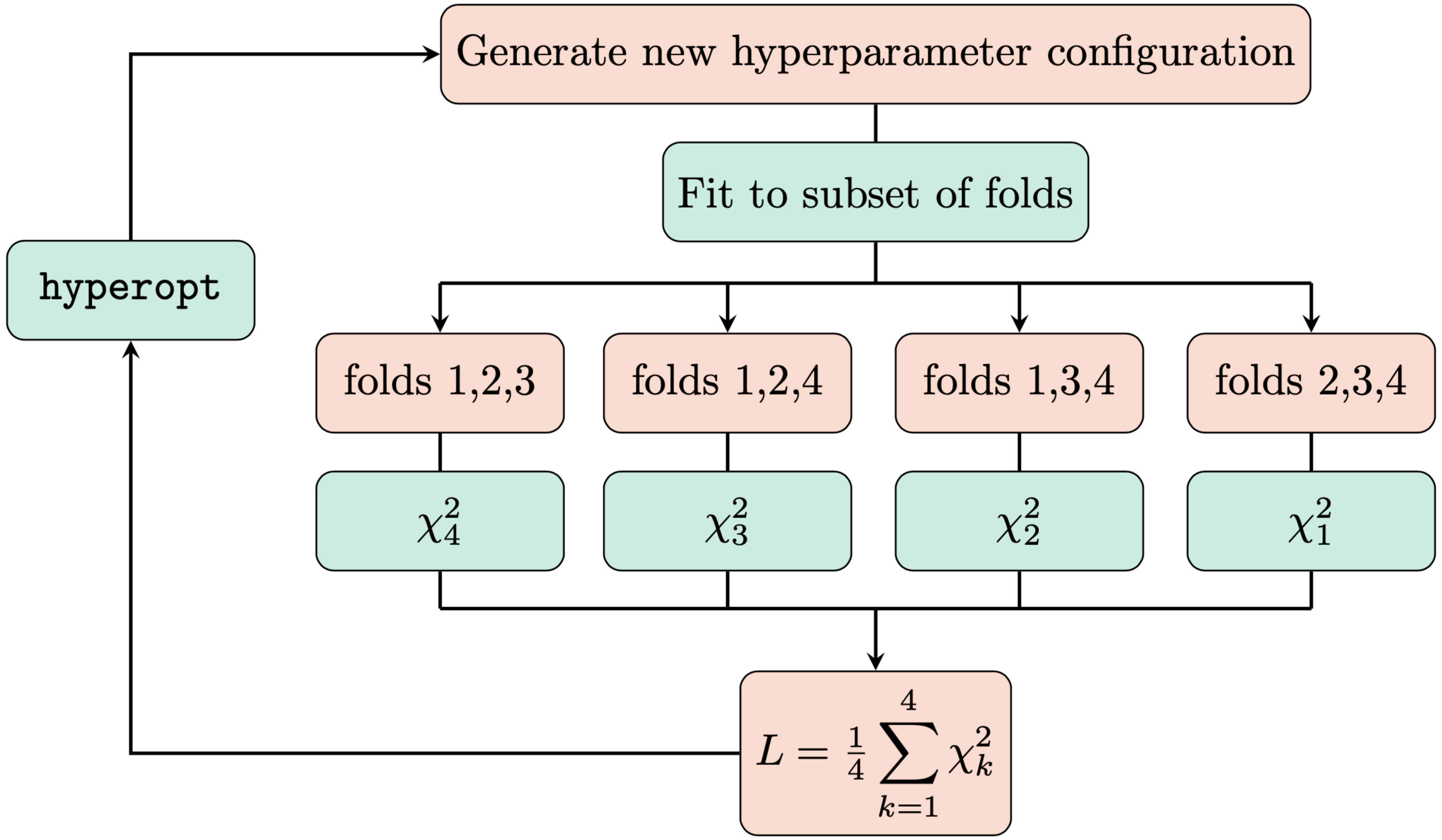
✅ Perform an **automated scan of the search space** by running fits with thousands of hyperparameter combinations using a **suitable metric** ‼

"When a measure becomes a target, it ceases to be a good measure" Goodhart's law

The choice of **figure of merit is crucial in obtaining a "Good Fit"** (smoothness of the PDFs, generalisation power to future experimental data, time/iterations it takes to complete a fit, etc.)

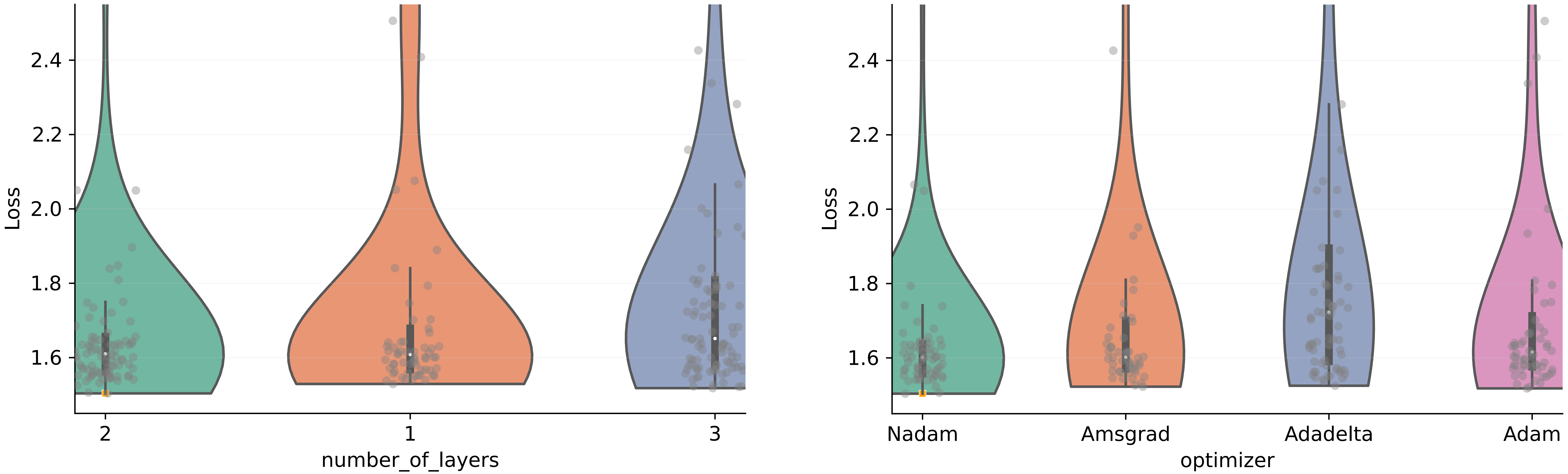
In NNPDF4.0, the figure of merit is defined in terms of **k-fold cross validation method**. For each hyperparameter configuration, we run 4 fits to the **central** experimental data, and in each of these fits, **the n-th fold is left out**.

The metric is then defined as the χ^2-averaged of the left-out folds.



Hyperparameter Optimisation à la NNPDF4.0

- Large parameter space that takes into account all possible hyperparameters: optimiser, initialiser, number of layers, activation functions , learning rates, number of epochs, stopping patience, Lagrange multiplier.
- Due to the computationally intensive nature of the fit $\otimes$ hyperoptimisation, only **one single** replica was considered during the hyperparameter tuning. A single replica fit requires about **~4 hrs (4×4 folds=16 hrs)** and **~16 GB of memory**).



But what if we want to perform hyperparameter tuning at the level of the PDF distribution?

Hyperopt on PDF Distribution using GPUs

Significant improvements **on two main fronts**, namely the hyperparameter optimisation procedure and at the level of the fits themselves.

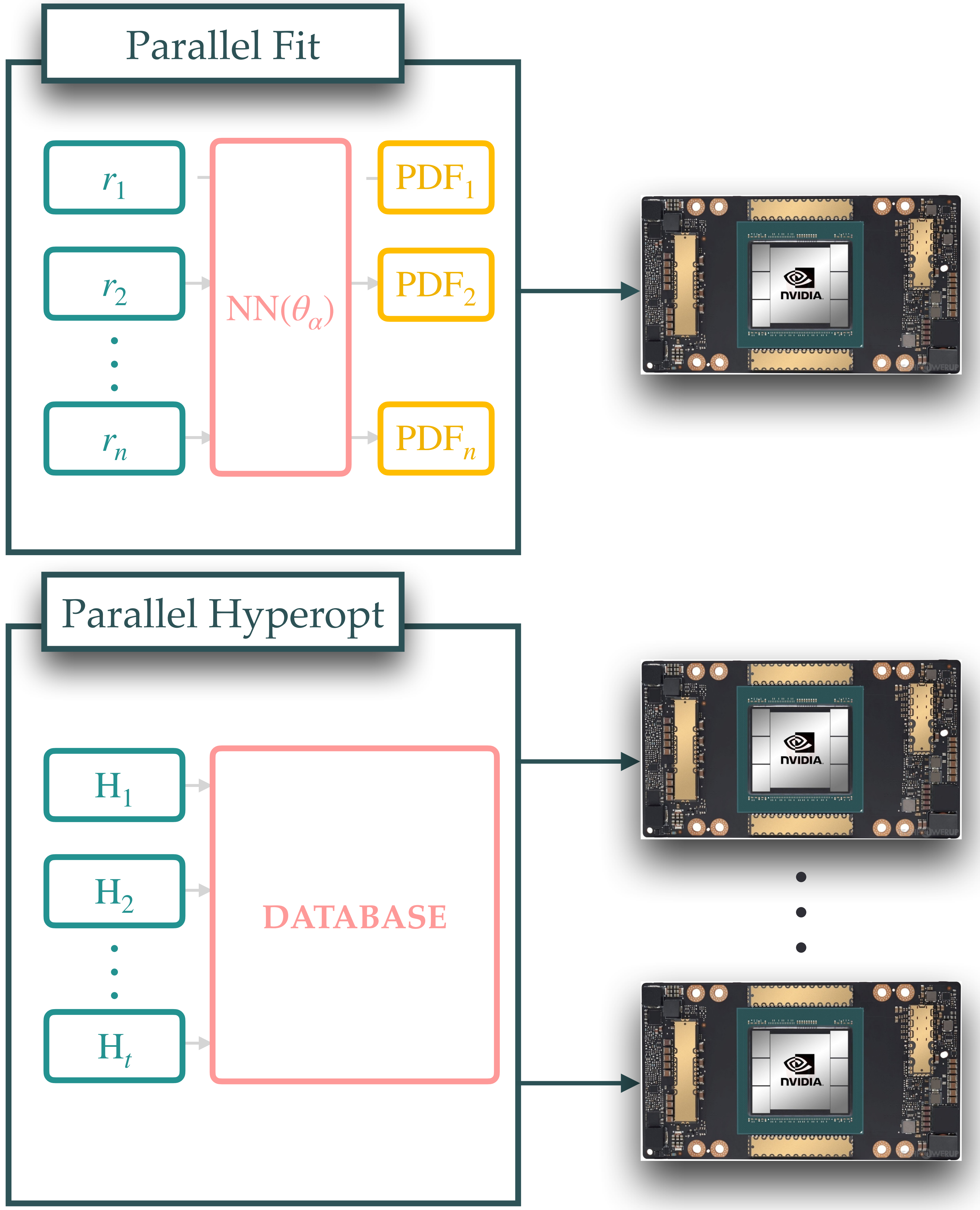
✦ Simultaneous fit of multiple replicas:

- Tensorflow allows the exact same codebase to be used for both CPU and GPU
- Redesign of the framework in order to share memory-heavy objects across all the replicas
- Resort to single PDF neural network model

$\Longrightarrow$ Running ~150 replicas at once on a A100 Nvidia GPU is now as fast as a fit of one single replica.

✦ **Distributed asynchronous Hyperparameter Optimisation:**

- Evaluate trials in parallel across many different GPUs
- Each instance of the worker shares the same database (MongoDB)



Hyperopt on PDF Distribution using GPUs: Figure of Merit

The difficult question: which figure of merit(s) should be considered? It turns out that defining what the perfect metric is a very challenging task (should the ensuing metric be just a combination of various metrics?).

We can define the properties of a **"Good Fit"**:

- Not under-learned nor overfitted: smoothness vs wiggles
- Generalisable to accommodate for future experiments
- Provides a faithful representation of the data uncertainties **?**

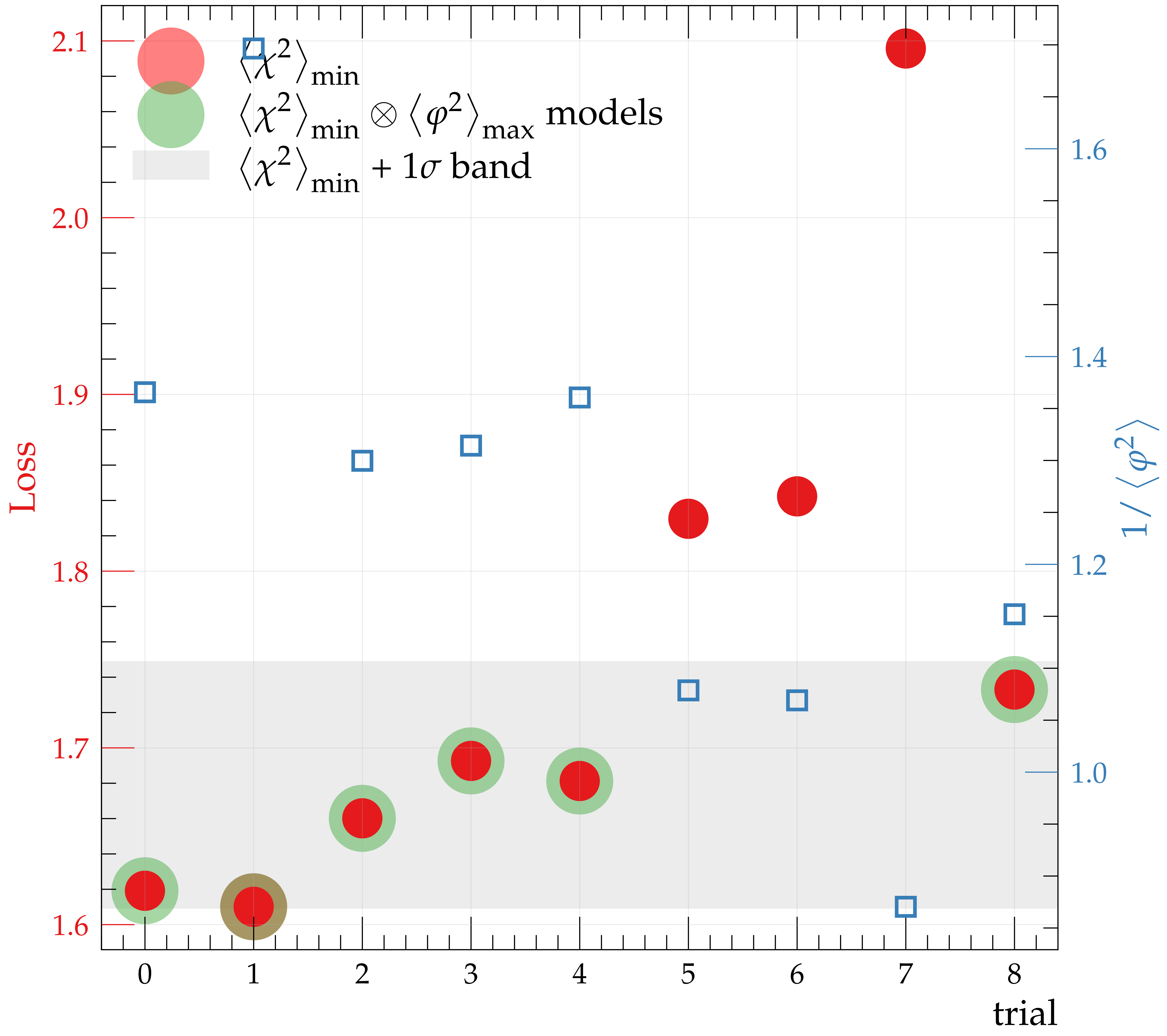
$\Longleftrightarrow$ k-fold loss

$$\mathscr{L} = \frac{1}{n_{\rm fold}} \sum_{k}^{n_{\rm fold}} \chi_k^2$$

A possible metric that accounts for all these criteria is a combination of the k-fold loss function with an indicator that assesses the PDF uncertainties w.r.t the ones from experimental data.

$$\varphi^2_{\chi^2_k} = \langle \chi^2_k [\mathscr{T}[f_{\rm fit}], \mathscr{D}] \rangle_{\rm rep} - \chi^2_k [\langle \mathscr{T}[f_{\rm fit}] \rangle_{\rm rep}, \mathscr{D}]$$

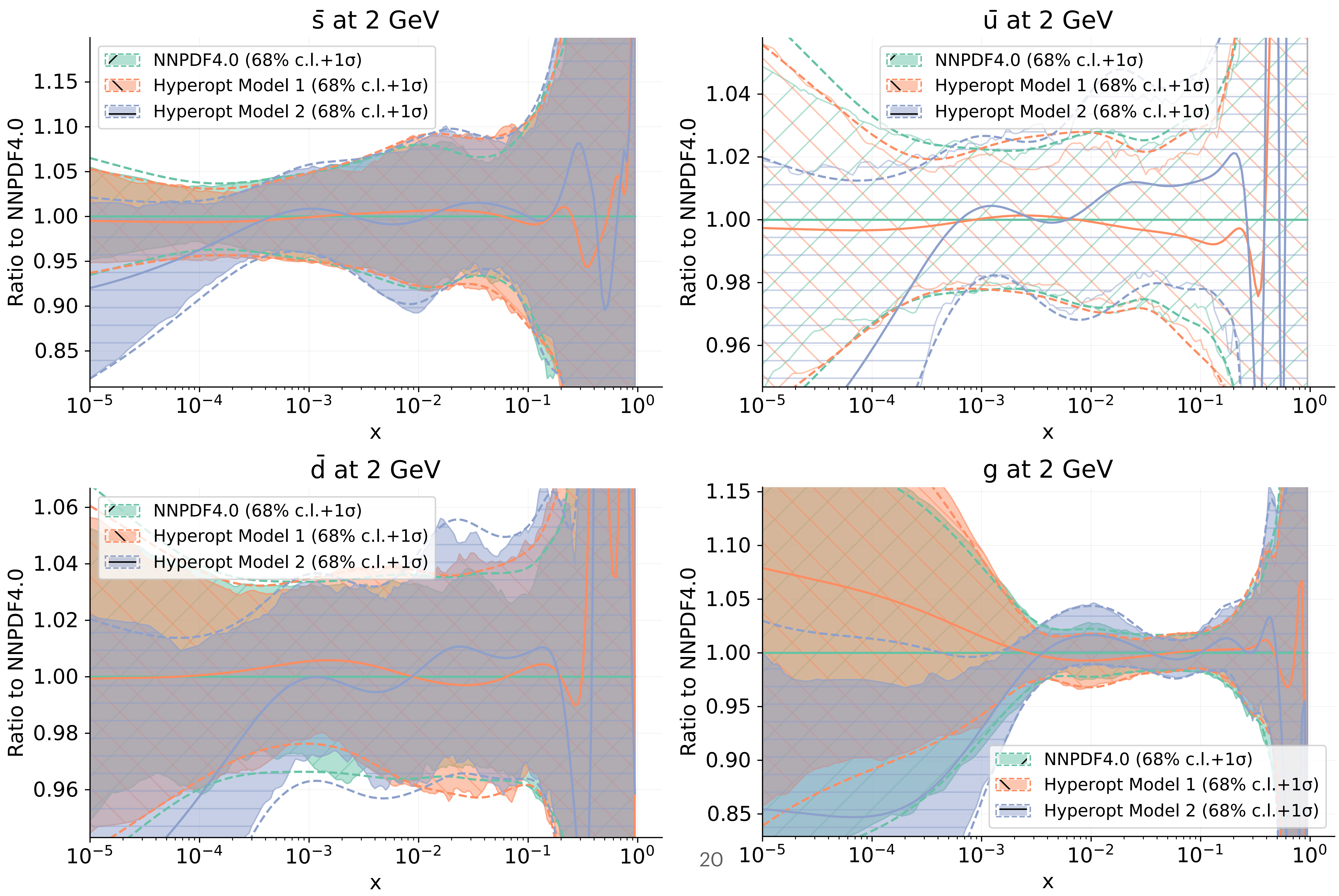
$\varphi^2_{\chi^2_k}$ measures the standard deviation over the replicas in units of data uncertainties.



What does such a Hyperopt look like?

- Define some hyperopt loss threshold in terms of the $\langle\chi^2\rangle$ of non-fitted folds
- Compute one-sigma standard deviation and select a range defined by $[\langle\chi^2\rangle_{\min}, \langle\chi^2\rangle_{\min} + 1\sigma]$
- Select n configurations based on the smallest $1/\varphi^2$

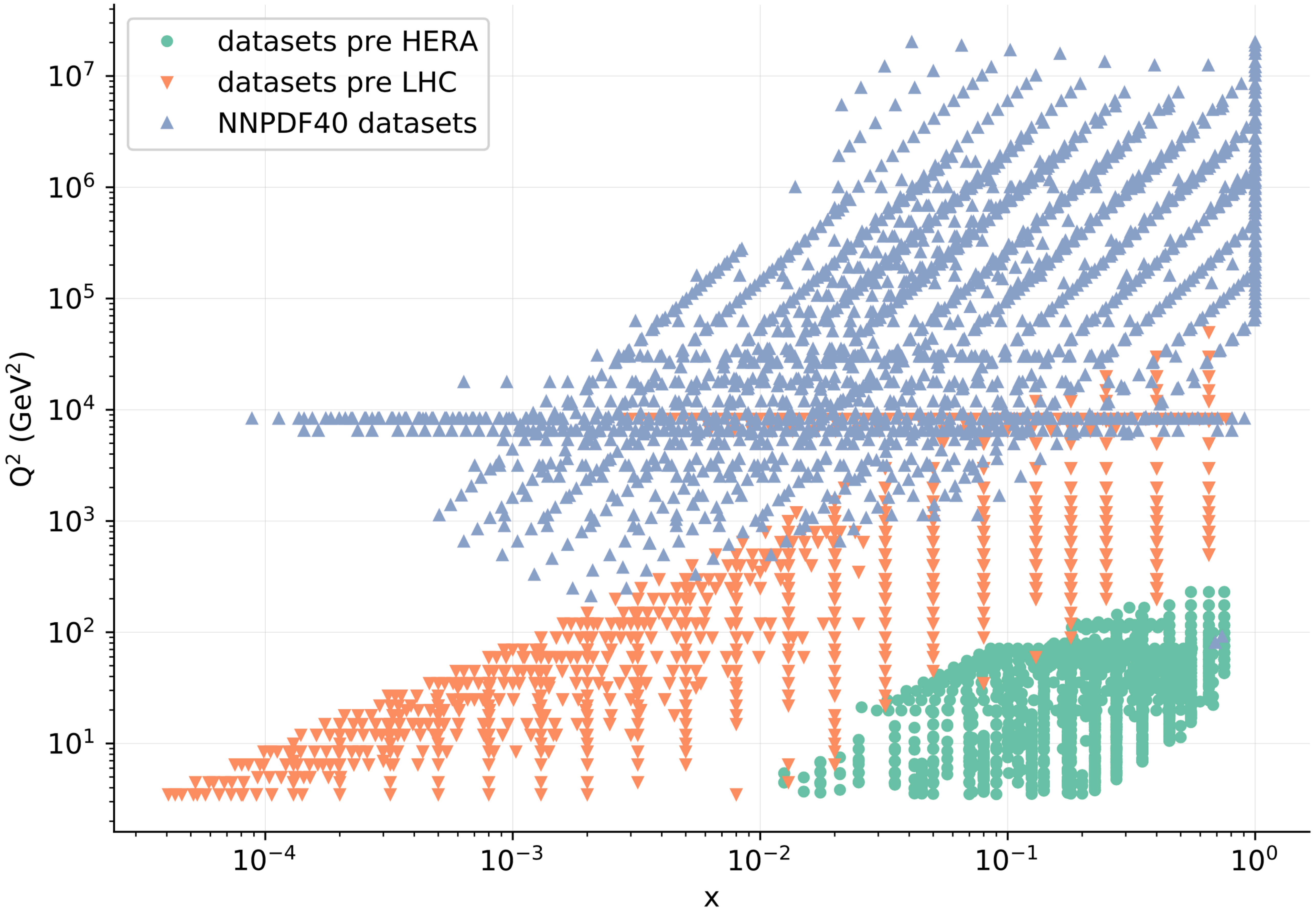
Hyperopt Models: PDF Distributions

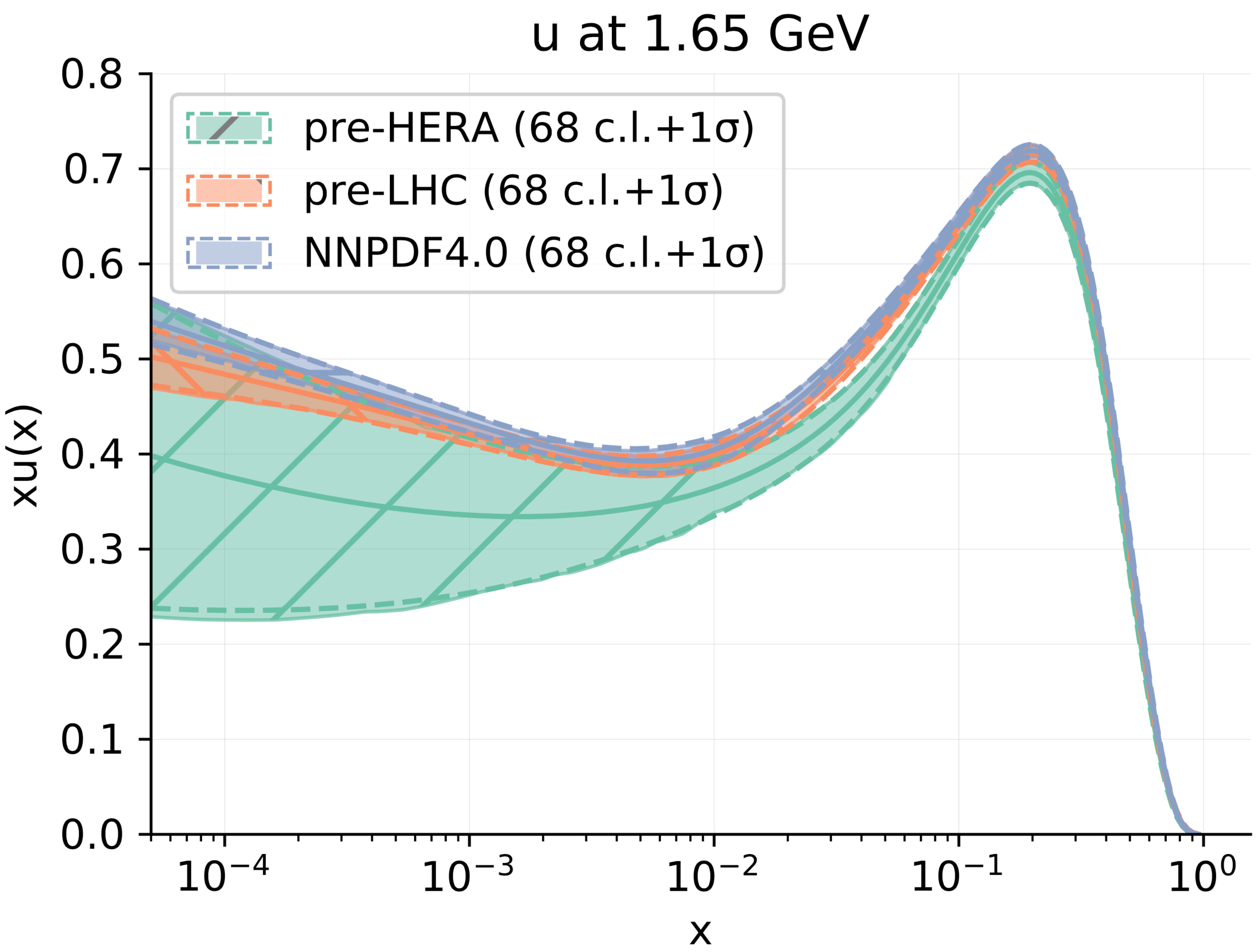


- Representative tuned hyper-parameter from the selected configurations
- All models are consistent with the published baseline NNPDF4.0 (in the data region) with one-sigma uncertainties
- Should the different PDF fits combined to account for the methodological uncertainties?

Uncertainty Validations: Future Tests

Fit Data to specific kinematic regions, and then checks the generalisation (extrapolation) to unseen experimental data:



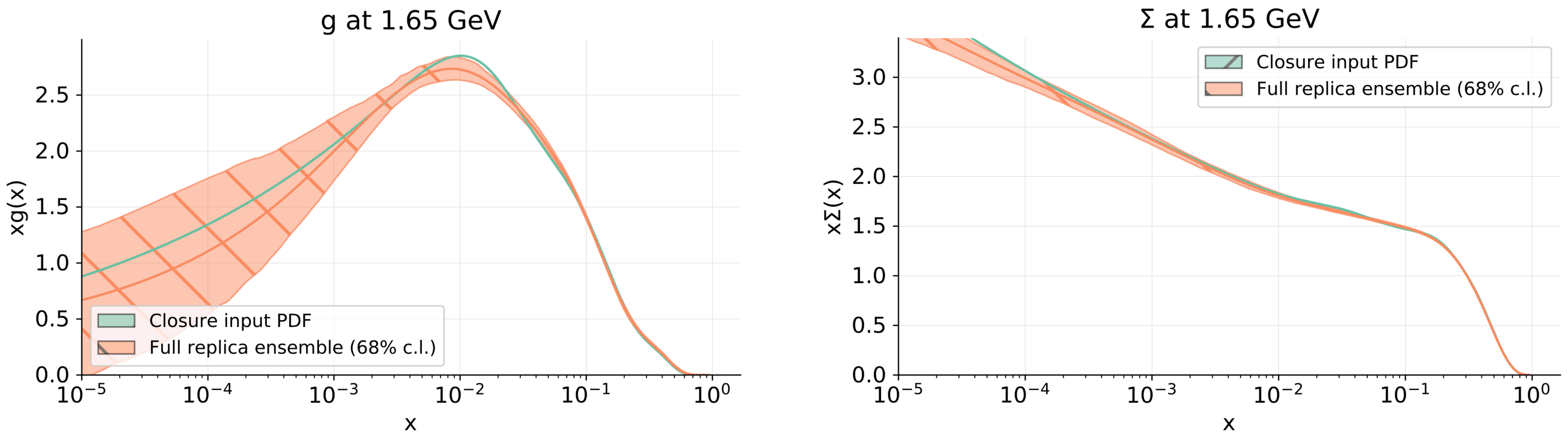


Uncertainty Validations: Closure Tests

Generate **"toy data"** based on some known PDF and check a posteriori that the true **underlying law** $\mathscr{F}$ is reproduced within errors. Fit replicas to pseudodata in the standard way according to:

$$\mathscr{Y} = \mathscr{F} + \eta + \epsilon, \text{ where } \eta \sim \mathcal{N}(0,C) \text{ and } \epsilon \sim \mathcal{N}(0,C)$$

If the uncertainty associated to the PDF replicas is faithfully reproduced, then the **bias-to-variance ratio** should be unity, ie. $\mathscr{R}_{bv} \equiv \sqrt{\mathbf{E}_\eta[\text{ bias }]/\mathbf{E}_\eta[\text{ variance }]} = 1$.



Open Source Framework

NNPDF

Test conda package passing DOI 10.5281/zenodo.10730835

NNPDF: An open-source machine learning framework for global analyses of parton distributions

The NNPDF collaboration determines the structure of the proton using Machine Learning methods. This is the main repository of the fitting and analysis frameworks. In particular it contains all the necessary tools to reproduce the NNPDF4.0 PDF determinations.

Documentation

The documentation is available at https://docs.nnpdf.science/

Github: https://github.com/NNPDF/nnpdf

Documentation: https://docs.nnpdf.science/

NNPDF

Search docs

Tutorials

This section contains tutorials for common things you might want to do using the code. Adding to the Documentation and Reviewing pull requests).

Running fits

- How to run a PDF fit
- How to run an iterated fit
- How to run a QED fit
- How to run a Polarized fit
- Including a general theory covariance matrix in a fit
- How to include a theory covariance matrix in a fit

Analysing results

- How to compare two fits
- How to generate a report
- How to run an analysis in parallel
- Using dask without a Scheduler
- How to plot PDFs, distances and luminosities
- Plotting non-trivial combinations of PDFs
- How to do a data theory comparison
- Interpreting the \(\mathcal{R}_O\) overfit metric

Closure tests

- How to run a closure test
- How to analyse a closure test

Special PDF sets

- Bundle PDFs with \(\alpha_s\) replicas
- How to transform a Monte Carlo PDF set into a Hessian PDF set

Conclusions & Outlook

- NNPDF4.0 studies the proton PDFs by fitting to experimental datasets and achieves **high accuracy** in an unprecedentedly broad kinematic range thanks to **deep learning models**

- Hyperparameter tuning is an important part in selecting good Machine Learning models; the definition of the figure of merit is crucial

- GPU optimisation allows for a tuning of the hyperparameters at the level of PDF distributions thanks to parallelisations

- The full NNPDF frameworks is Open Source and contains documentations and tutorials

"Wanderer above the Sea of Fog" by Caspar David Friedrich