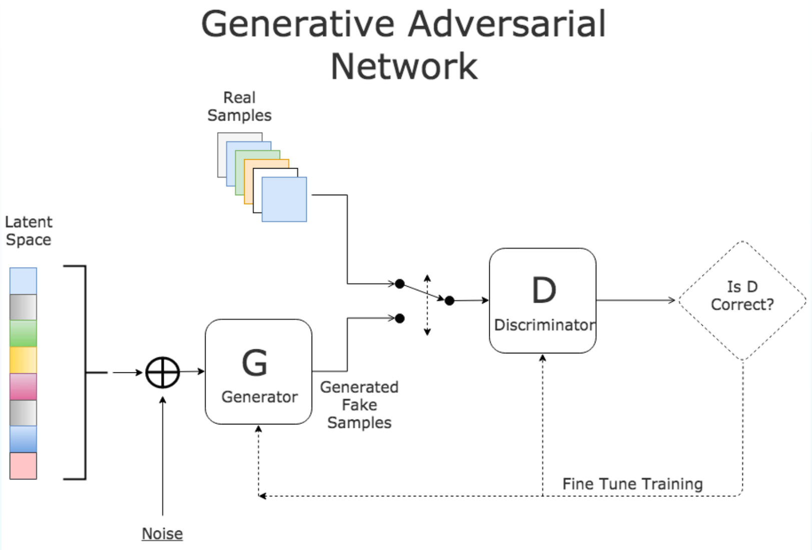
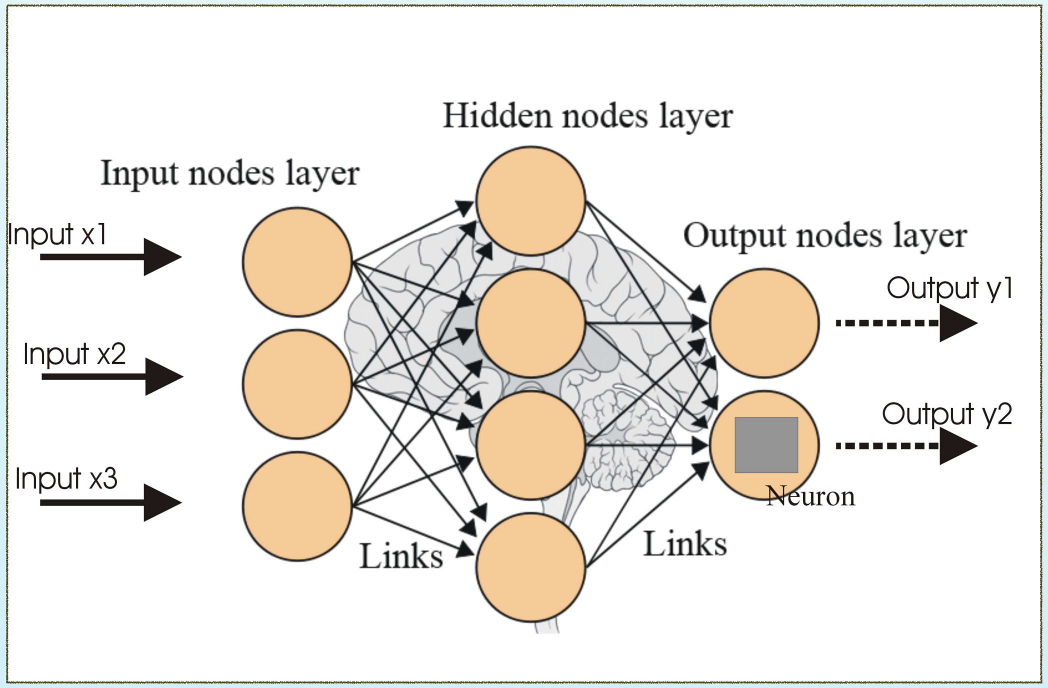
Machine Learning Tools for Global PDF Fits

Juan Rojo
VU Amsterdam & Nikhef

Quark Confinement and the Hadron Spectrum XIII
Maynooth University, 02/08/2018



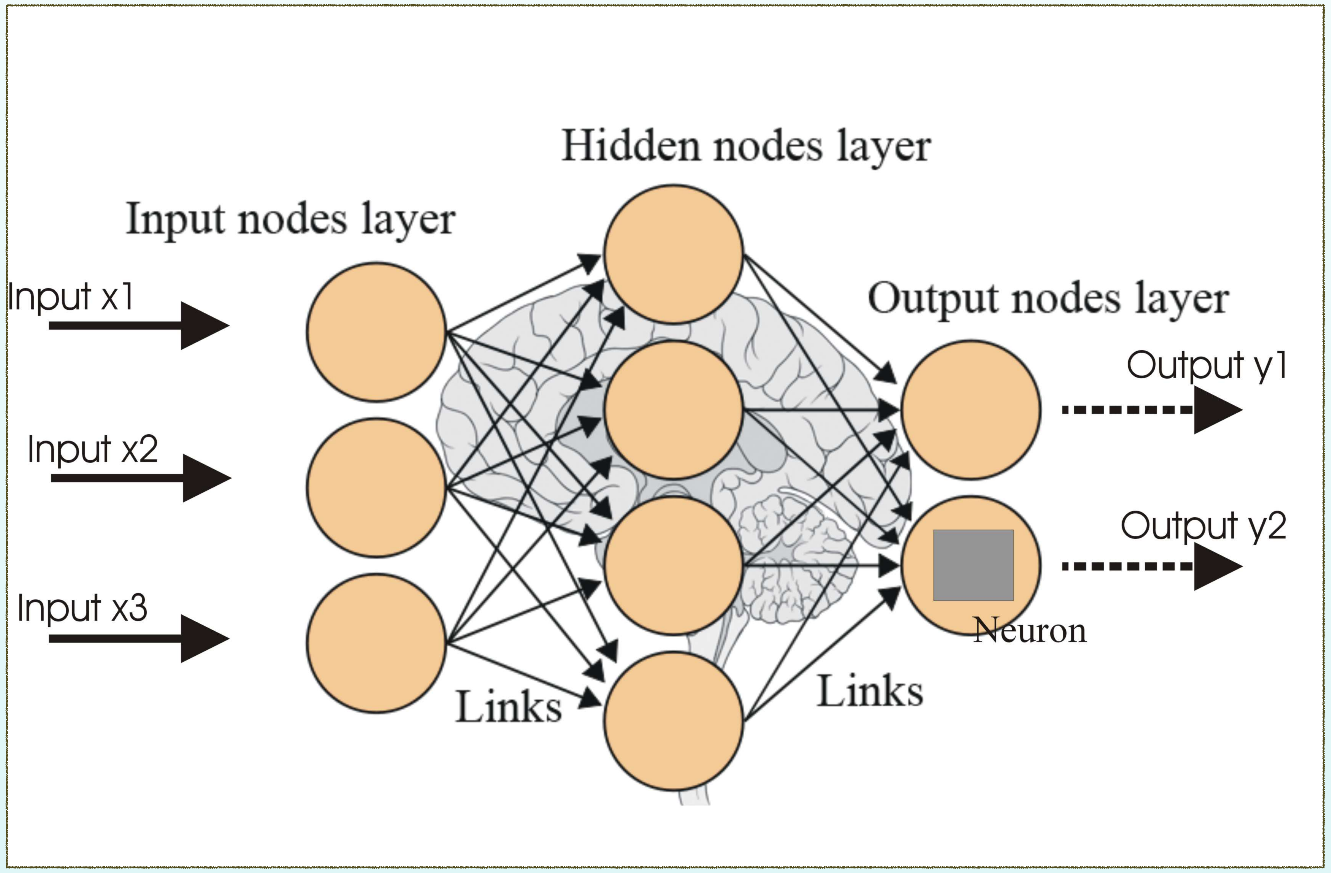
Artificial Neural Networks



Artificial Neural Networks

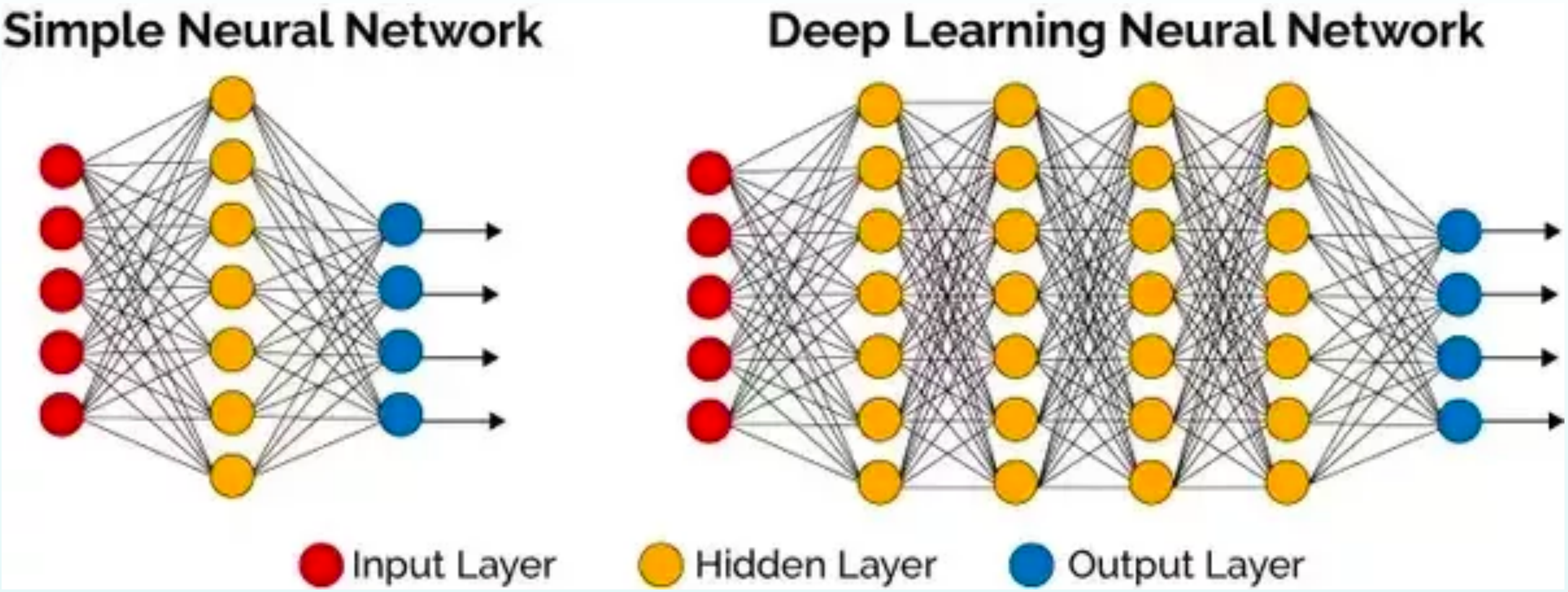
Inspired by **biological brain models**, **Artificial Neural Networks** (ANNs) are **mathematical algorithms** widely used in a wide range of applications, from **HEP** to **targeted marketing** and **finance forecasting**

From biological to artificial neural networks



Artificial neural networks aim to excel where domains as their **evolution-driven counterparts outperforms traditional algorithms in tasks such as pattern recognition, forecasting, classification**, …

Deep Neural Networks



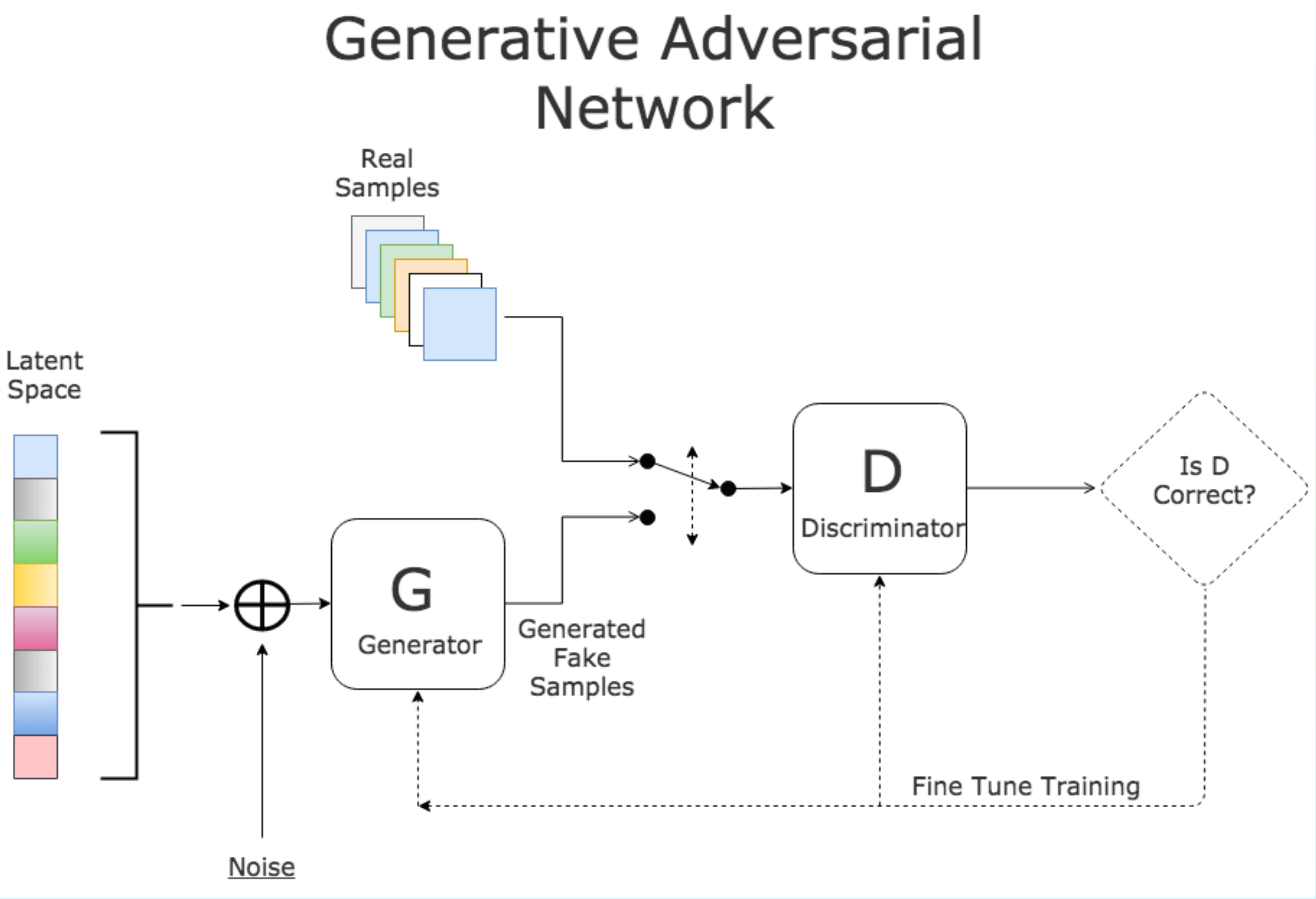
- A **Deep Neural Network (DNN)** is a standard multi-layer feed-forward perceptron with a large number of internal layers

- All types of neural nets eg **Recursive, Convolutional, Parametrised** etc can be made **"deeper"** by adding more hidden layers

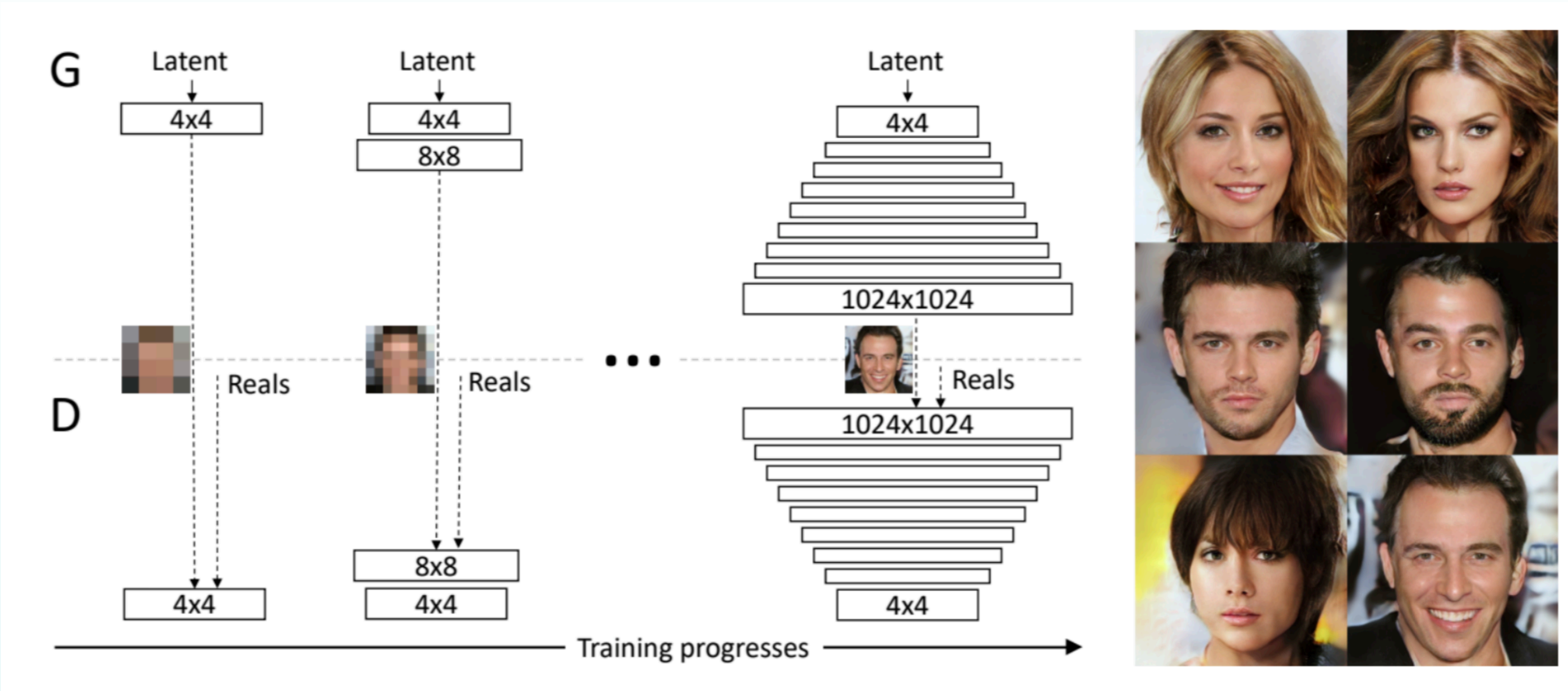
- For several applications, the **increased complexity** achieved this way leads to a significant improvement in performance

Generative Adversarial Networks

- New architecture for an **unsupervised neural network training** (unlabelled samples)
- Based on two **independent nets** that work separately and act as adversaries:
    - the **Discriminator (D)** undergoes training and plays the role of classifier, and
    - the **Generator (G)** and is tasked to generate random samples that **resemble real samples** with a twist rendering them as fake samples.



The many uses of GANs

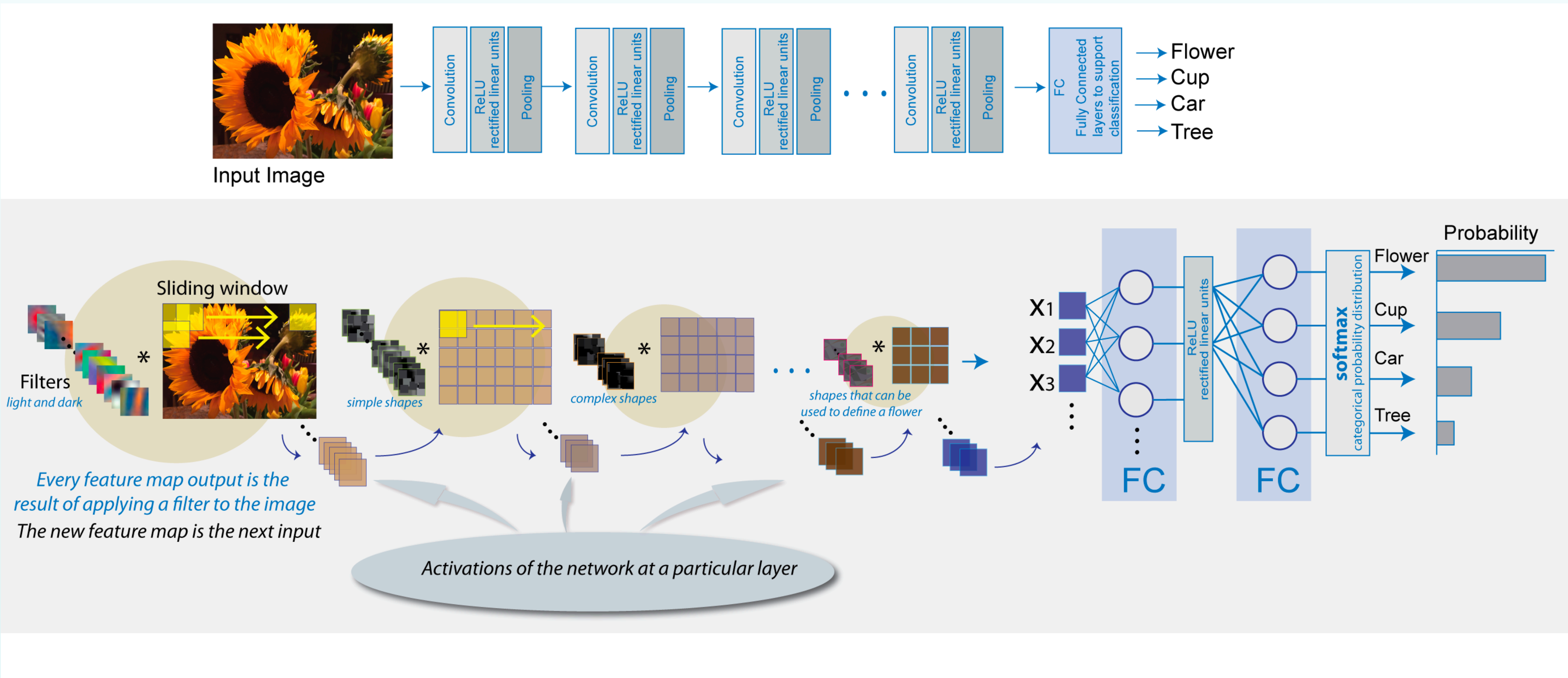


arXiv:1710.10196

Which one of these images are real and which ones are fake (generated by the GANs)?

Convolutional Neural Networks

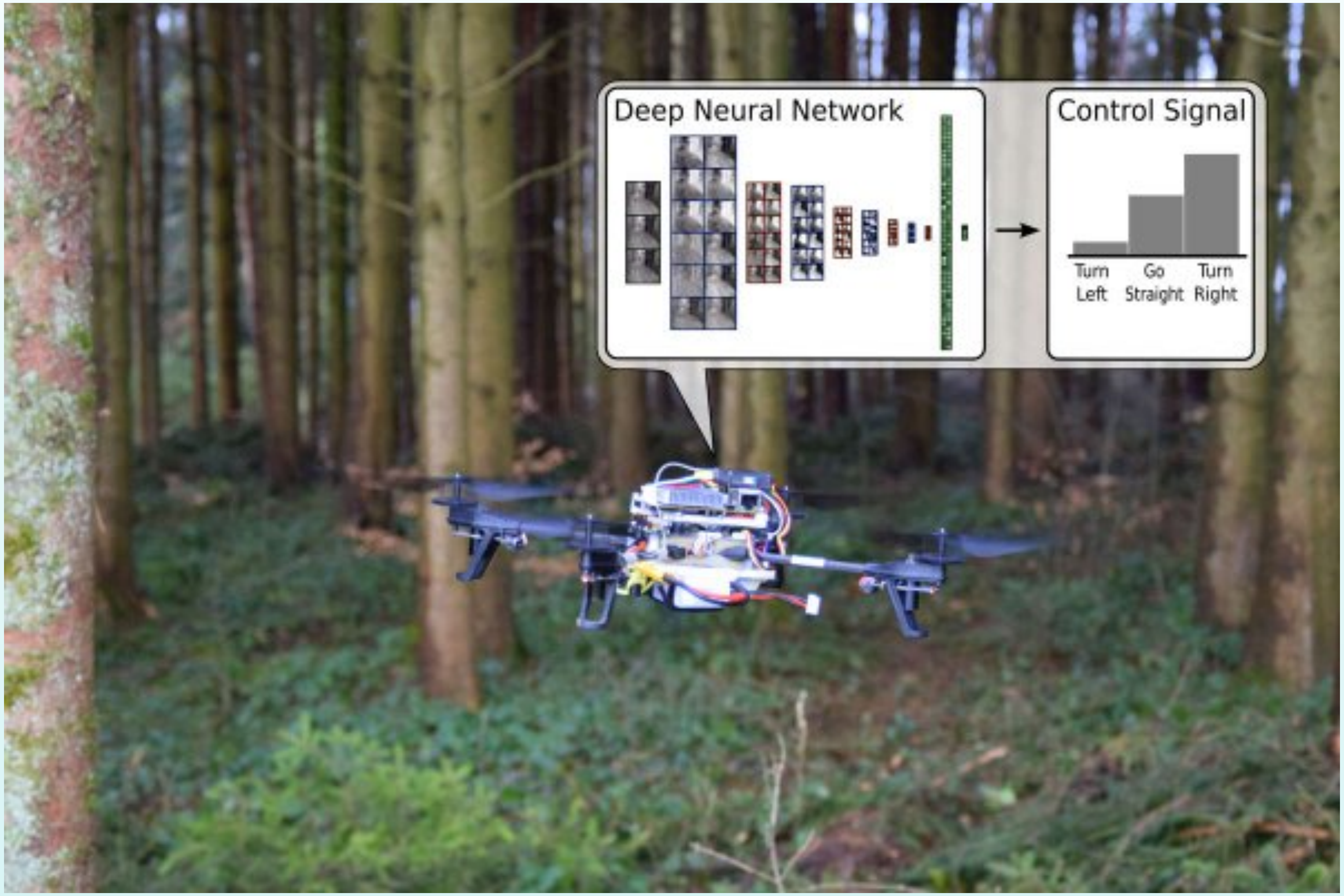
- Convolutional Neural Networks (CNNs) have convolutional layers based on **filters**
- Each **filter** maps a group of numbers into a number, reducing the dimensionality of the data
- Specially useful for **pattern recognition** (eg for self-driving vehicles)



mathworks.com

Convolutional Neural Networks

- ANNs can enable an **autonomous vision-control drone** to recognise and follow forest trails
- Image classifier operates directly on **pixel-level image intensities**
- If a trail is visible, the **software steers the drone** in the corresponding direction

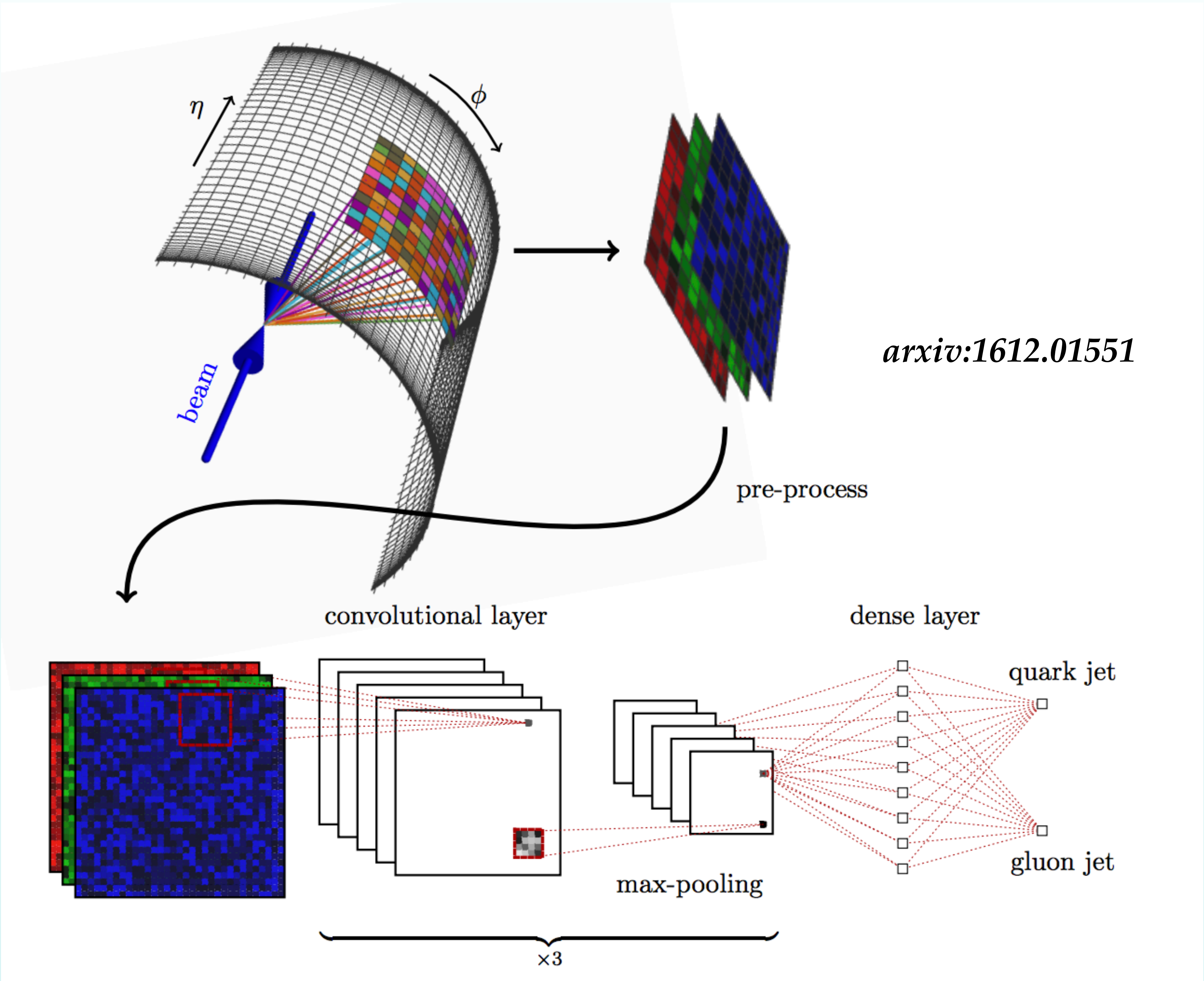


Giusti et al, IEEE Robotics and Automation Letters, 2016

Similar algorithms at work in self-driving cars!

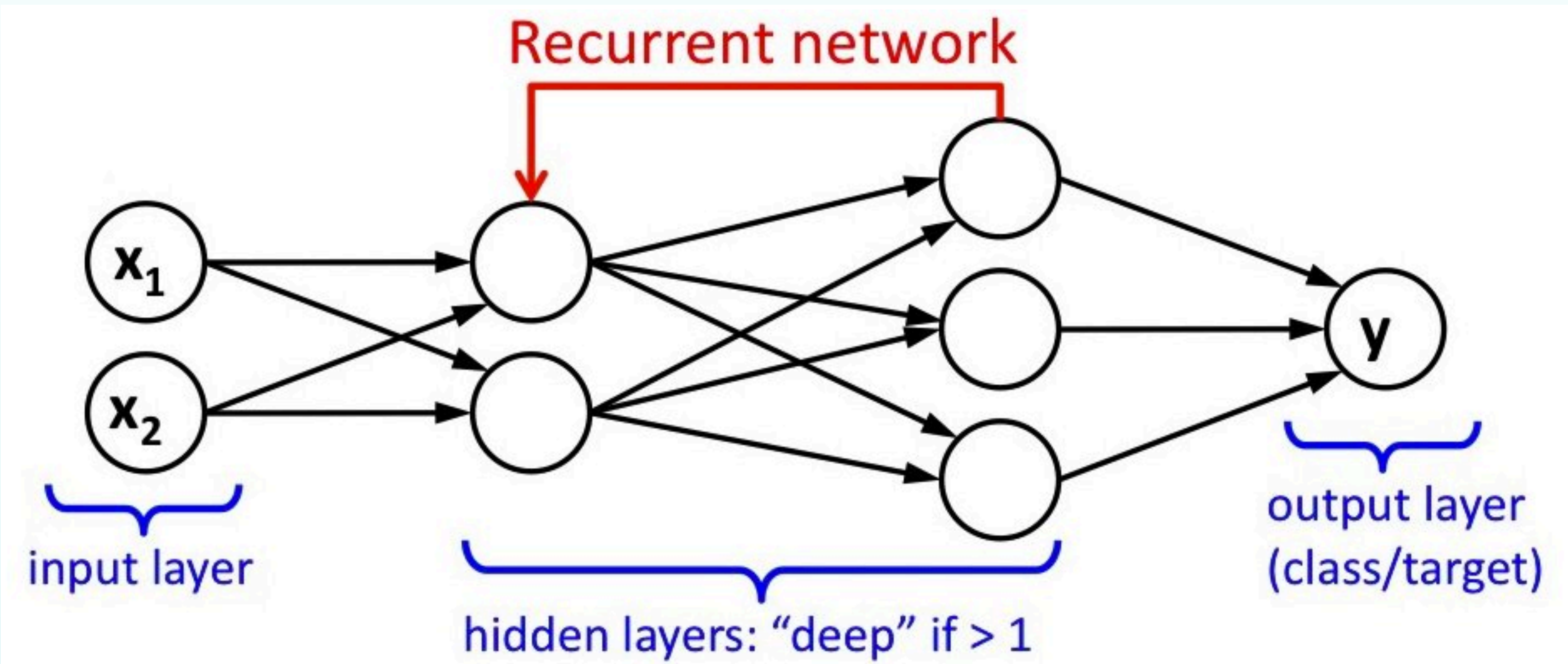
Convolutional Neural Networks

The results of the **collisions of high-energy particles** can be treated analogously to **image processing** using Convolutional Neural Networks



Recurrent Neural Networks

RNNs use as inputs not just the current "training examples" but also **what they have perceived previously**: they have a **built-in notion of time ordering** useful for time-dependent functions



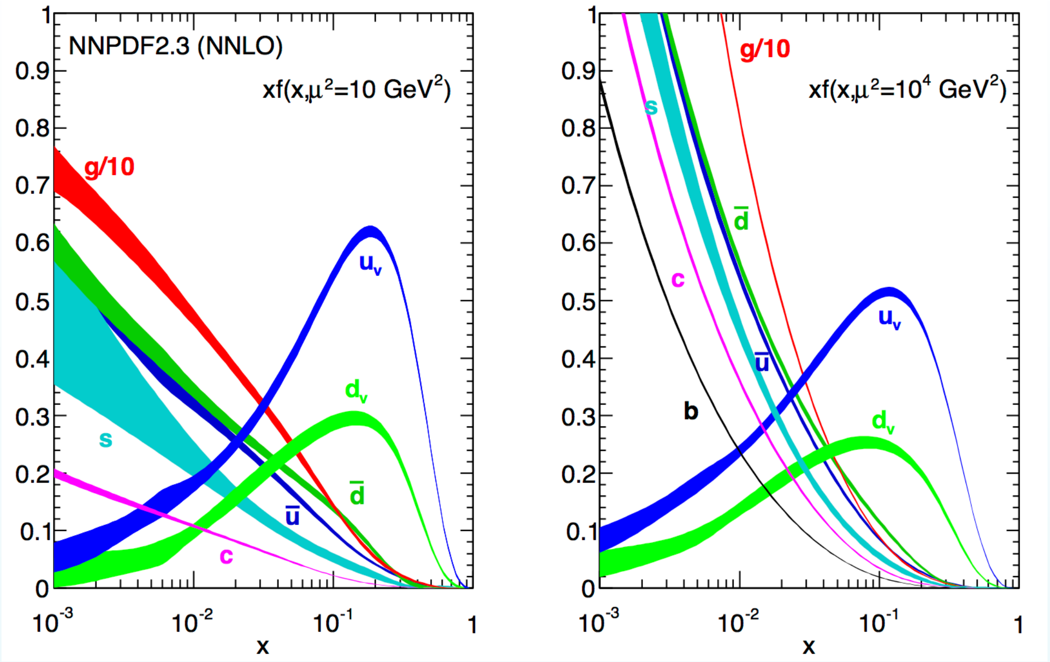
The output of a RNN at time time, ***y(t),*** depends both on the current input example ***x(t)*** as well as of its previous output ***y(t-1)*** (or activation states of hidden neutrons at ***t-1***)

Recurrent Neural Networks

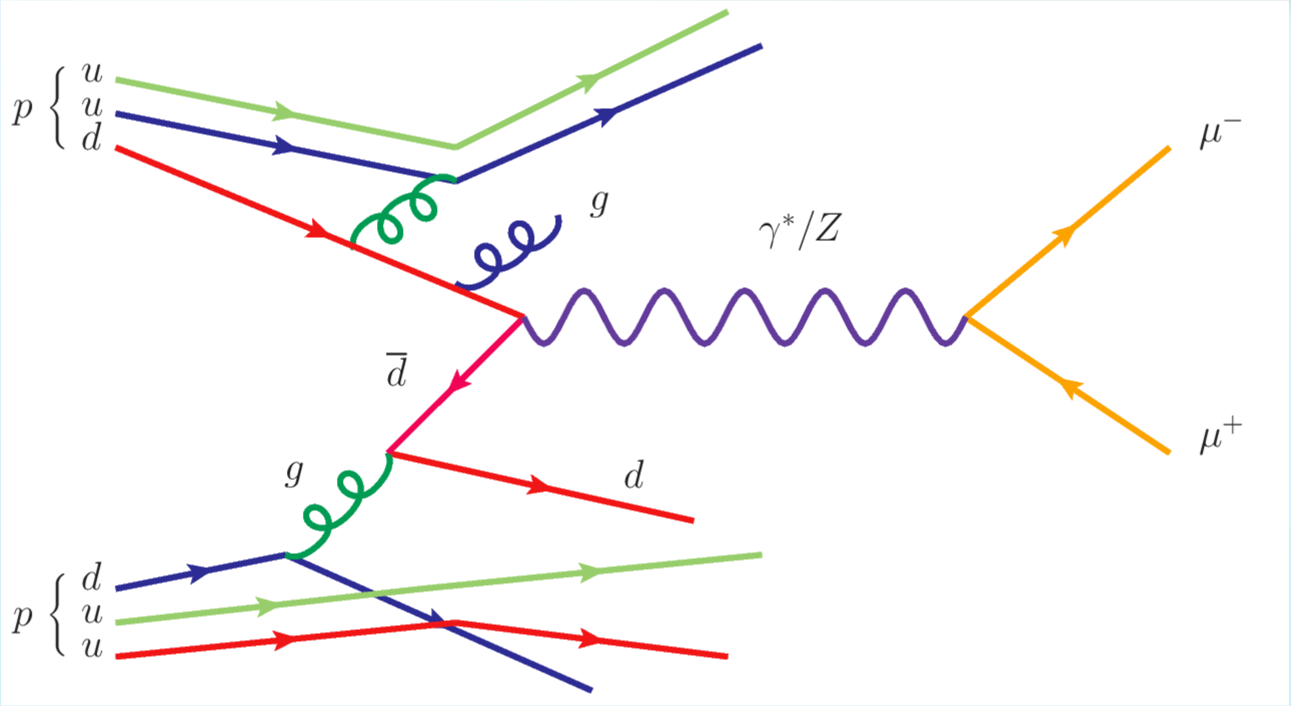
Lead to truly game-changer applications, such as **random generation of country song lyrics**

```
Tied right now
I got life now he never thought I got by the all
Going up like a house four boy
Nothing his thing out of hands
No one with the danger in the world
I love my black fire as I know
But the short knees just around me
Fun the heart couldnes fall to back
I see a rest of my wild missing far
When I was missing to wait
And if I think
It's a real tame
I say I belong is every long night
Maybe lovin' you
```

http://www.mattmoocar.me/blog/RNNCountryLyrics/



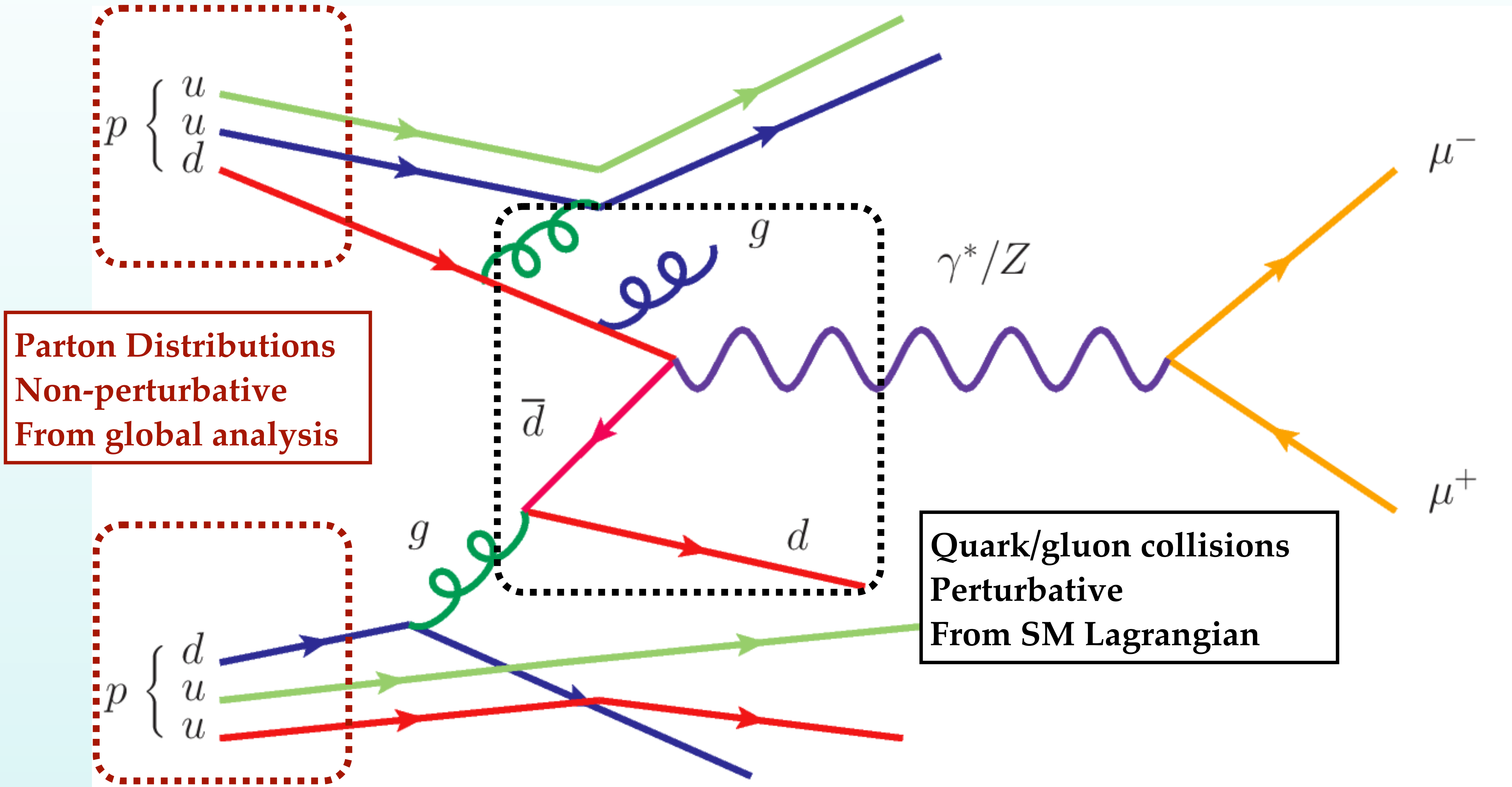
Neural Nets and PDF fits



See also slides of my talk in the ``QCD and New Physics'' session!

anatomy of hadronic collisions

In high-energy **hadron colliders** the collisions involve **composite particles** (protons) with internal substructure (quarks and gluons): the LHC is actually a **quark/gluon collider!**



Calculations of **cross-sections** in hadron collisions require the combination of **perturbative cross-sections** with **non-perturbative parton distribution functions (PDFs)**

the inner life of protons

Distribution of energy that **quarks and gluons carry inside proton** quantified by **Parton Distributions**

$$g(x,Q)$$

$g(x,Q)$: **Probability** of **finding a gluon inside a proton**, carrying a fraction ***x*** of the proton momentum when probed at energy ***Q***

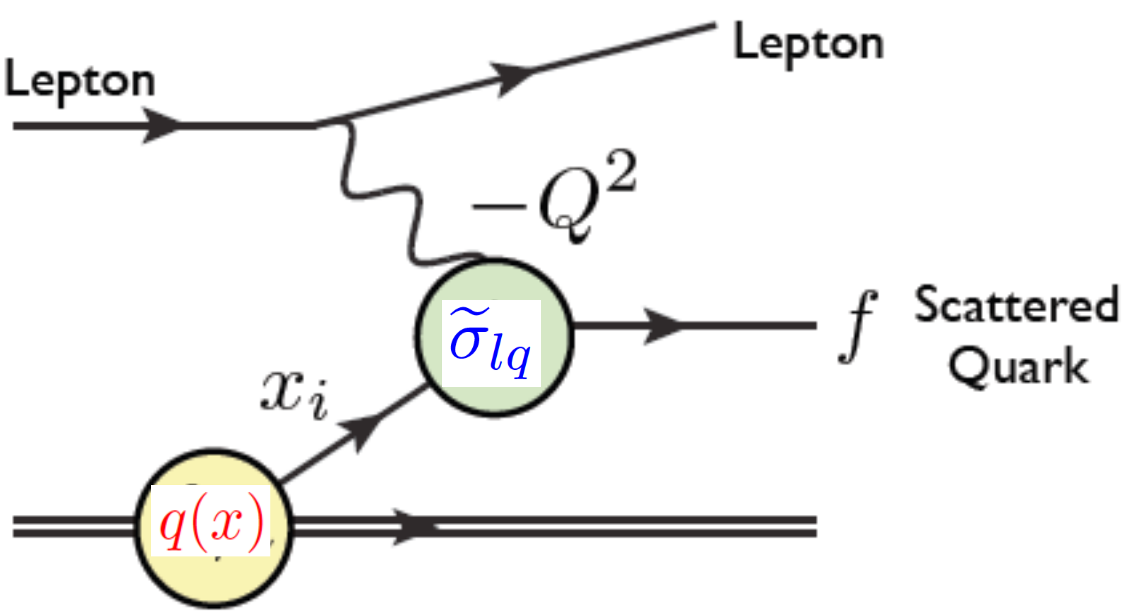
***Q*: Energy of the quark/gluon collision**
Inverse of the resolution length

***x*: Fraction of the proton's momentum**

PDFs** determined by **non-perturbative QCD dynamics

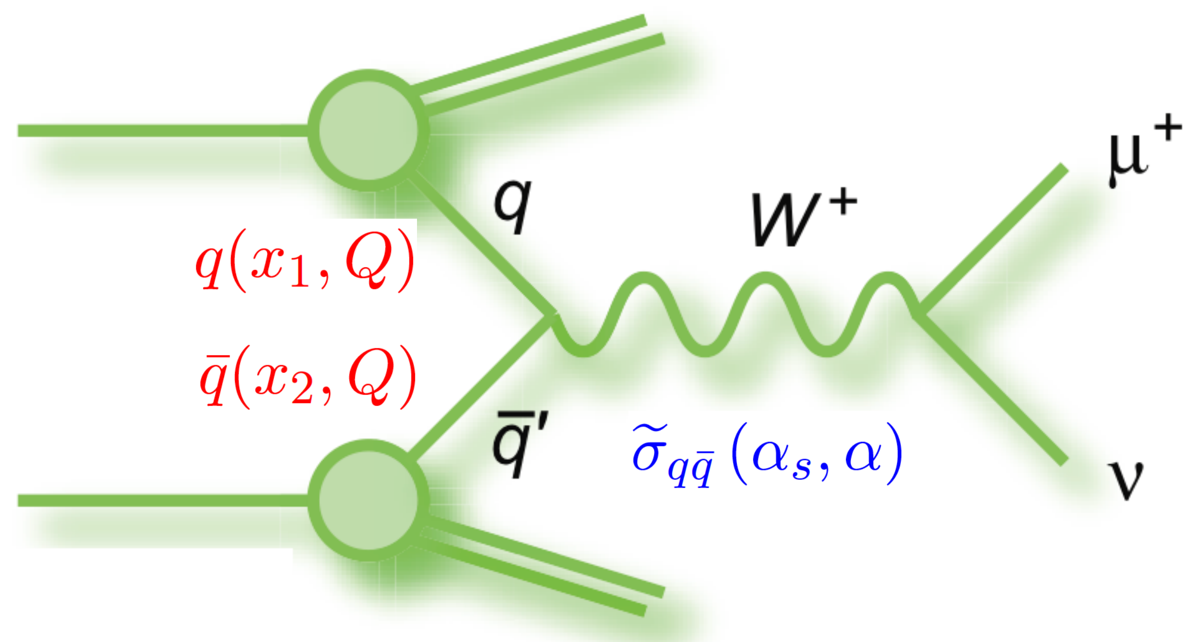
*Extract from experimental data within a **global analysis***

$$\sigma_{lp} \simeq \widetilde{\sigma}_{lq}\,(\alpha_s,\alpha) \otimes q(x,Q)$$



Extract PDFs from lepton-proton collisions

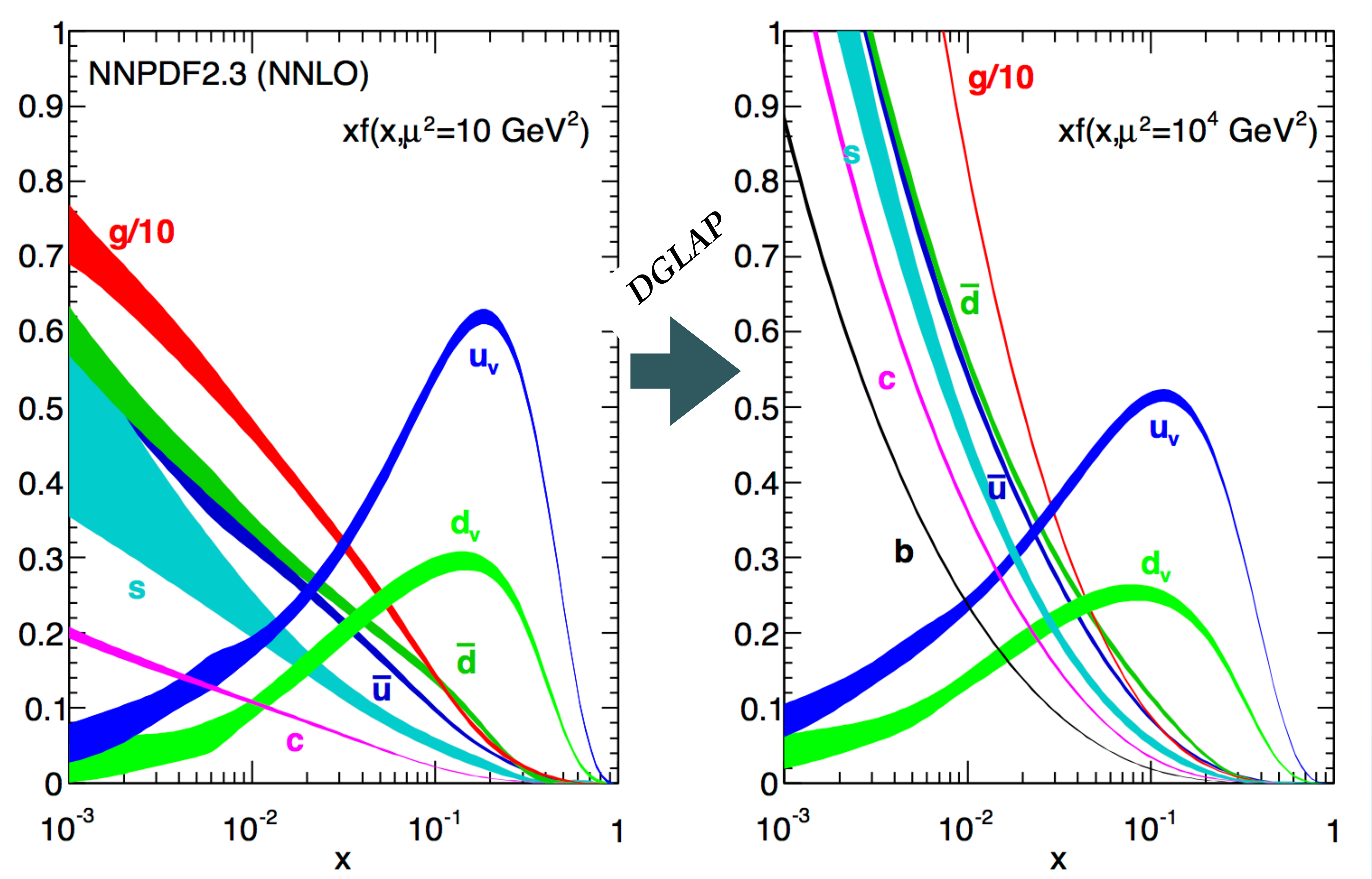
$$\sigma_{pp} \simeq \widetilde{\sigma}_{q\bar{q}}\,(\alpha_s,\alpha) \otimes q(x_1,Q) \otimes \bar{q}(x_2,Q)$$



Use PDFs to predict proton-proton cross-sections

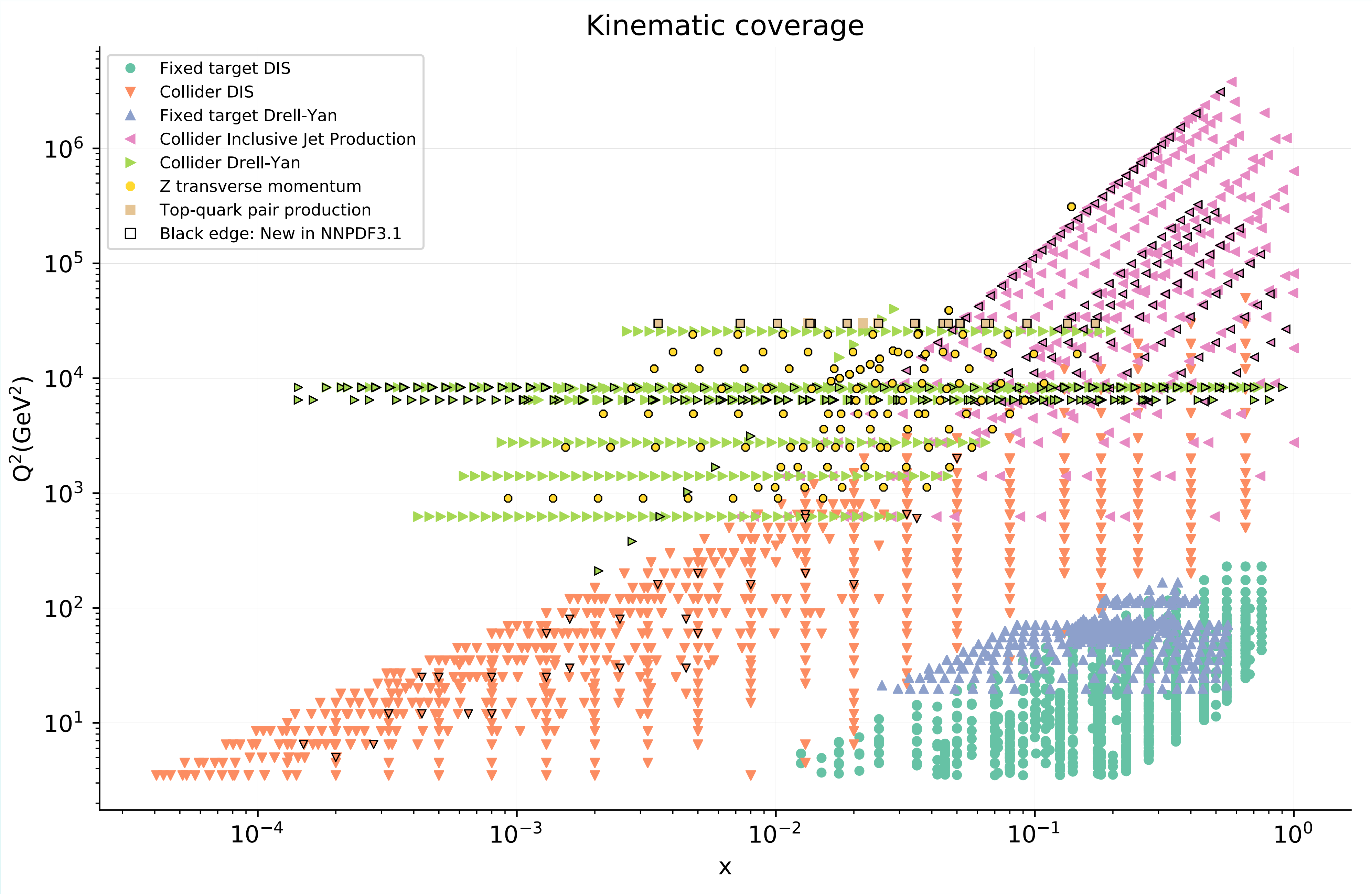
the inner life of protons

Determine the PDFs at some low scale $Q_0 \simeq m_p \simeq 1$ GeV



… and then evolve upwards using DGLAP to predict LHC cross-sections

The global QCD fit machinery



Highly non-trivial validation of the **QCD factorisation framework**:

- Including **O(5000) data** points ,
- from **O(40) experiments**,
- some of them with ≈**1% errors**,

yet still $\chi^2/N_{dat} \approx \mathbf{1}$!

The global QCD fit machinery

Experimental data

Fixed-target & collider DIS
Tevatron and LHC measurements
Jets, DY, top, Z pT, ...

Statistical framework

PDF parametrisation,
PDF uncertainties and propagation
Model and theory uncertainties

Theory calculations

The global QCD fit

Minimise figure of merit (*) and determine PDF parameters

fit validation, statistical estimators, diagnosis tools

APFEL WEB

http://apfel.mi.infn.it/

on-line plotting toolbox

LHAPDF

lhapdf.hepforge.org

standard interface for public PDF delivery

(*)

$$\chi^2\left(\{a_i\}\right) = \sum_{m,n=1}^{N_{\rm dat}} \left(\sigma_m^{(\rm exp)} - \sigma_m^{(\rm th)}\left(\{a_i\}\right)\right)\left({\rm cov}_{\rm exp} + {\rm cov}_{\rm th}\right)^{-1}_{mn}\left(\sigma_n^{(\rm exp)} - \sigma_n^{(\rm th)}\left(\{a_i\}\right)\right)$$

APPLgrid, FastNLO, aMCfast....

NNLO DGLAP evolution
DIS structure functions

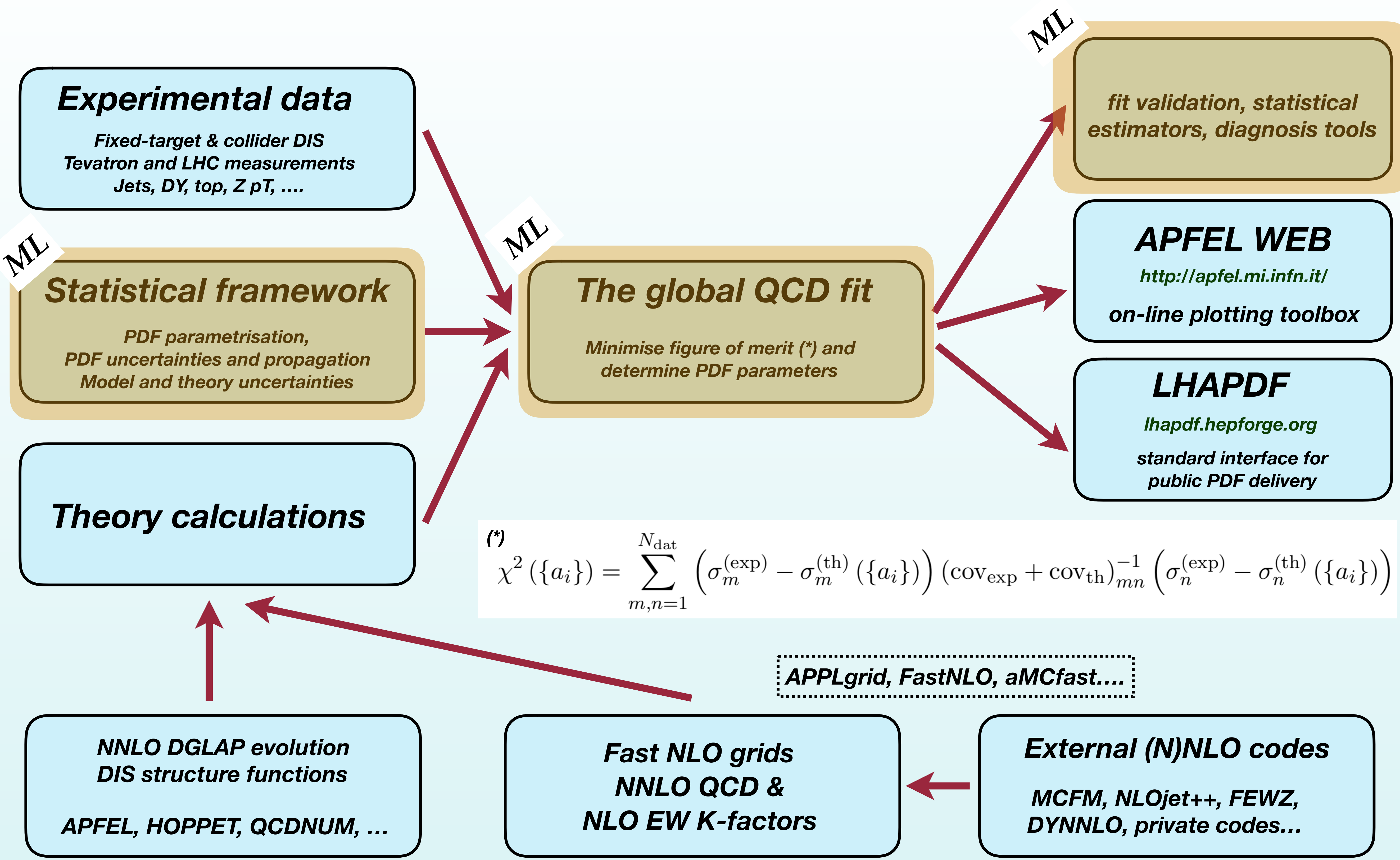
APFEL, HOPPET, QCDNUM, ...

Fast NLO grids
NNLO QCD &
NLO EW K-factors

External (N)NLO codes

MCFM, NLOjet++, FEWZ, DYNNLO, private codes...

The global QCD fit machinery



ANNs as universal unbiased interpolants

ANNs provide **universal unbiased interpolants** to parametrise the non-perturbative dynamics that determines the **size and shape of the PDFs** from experimental data

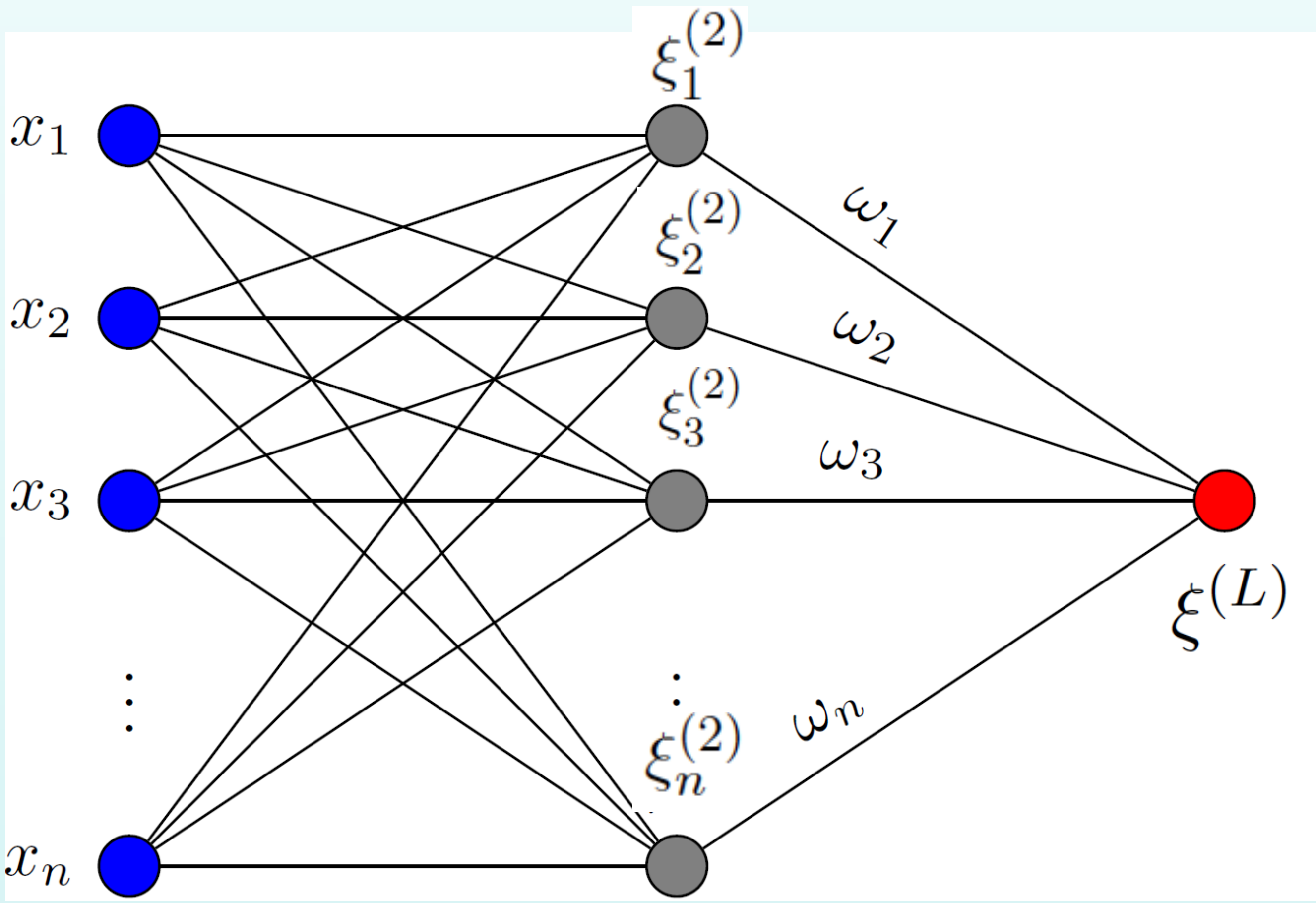
ad-hoc ansatz

Traditional approach

$$g(x,Q_0) = A_g(1-x)^{a_g}x^{-b_g}\left(1 + c_g\sqrt{s} + d_g x + \ldots\right)$$

NNPDF approach

$$g(x,Q_0) = A_g \mathrm{ANN}_g(x)$$



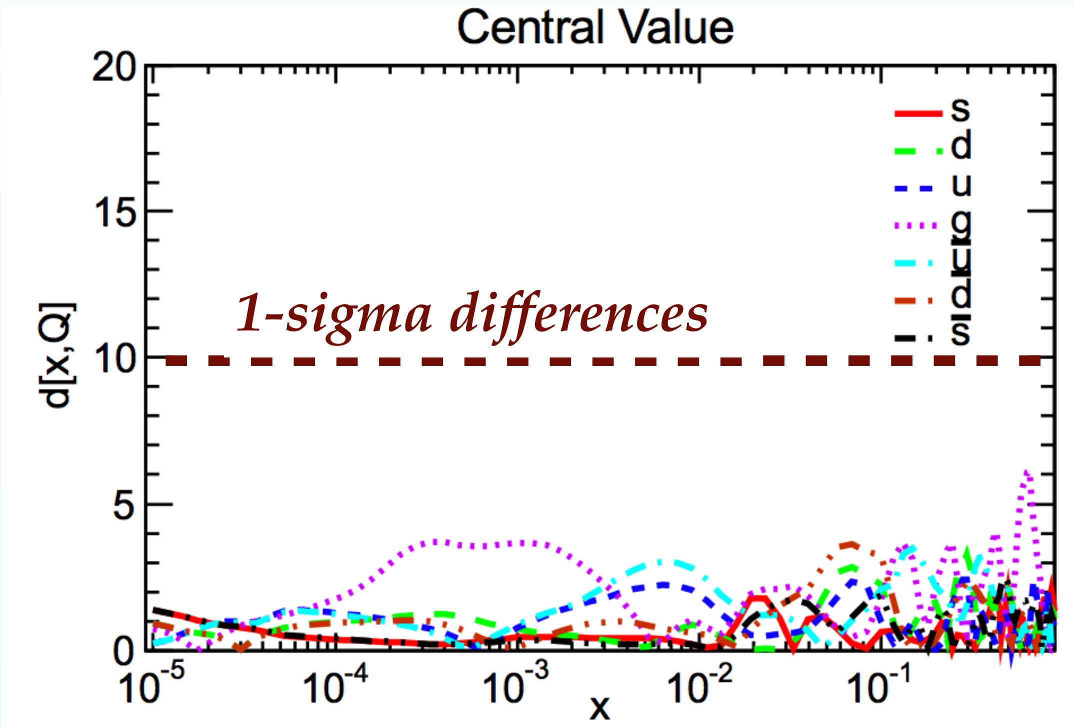
$$\mathrm{ANN}_g(x) = \xi^{(L)} = \mathcal{F}\left[\xi^{(1)}, \{\omega_{ij}^{(l)}\}, \{\theta_i^{(l)}\}\right]$$

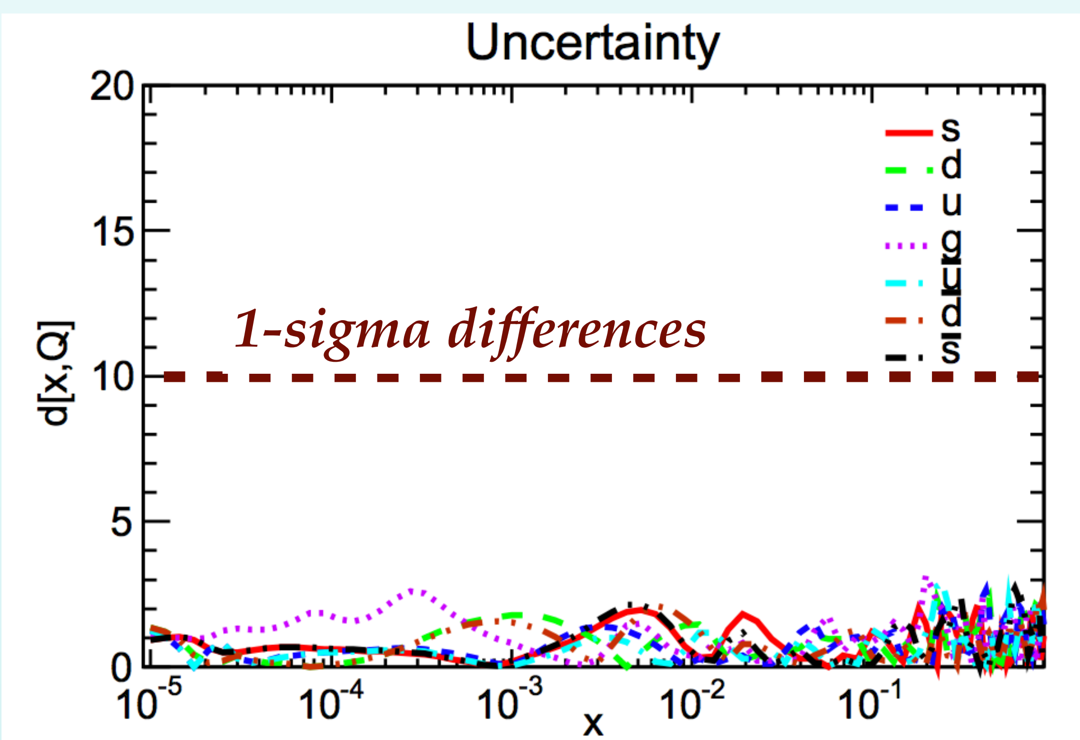
$$\xi_i^{(l)} = g\left(\sum_{j=1}^{n_{l-1}} \omega_{ij}^{(l-1)}\xi_j^{(l-1)} - \theta_i^{(l)}\right)$$

- ANNs eliminate **theory bias** introduced in PDF fits from choice of *ad-hoc* functional forms
- NNPDF fits used **O(400) free parameters**, to be compared with O(10-20) in traditional PDFs. Results stable if **O(4000) parameters used!**

ANNs as universal unbiased interpolants

Compare two global PDF fits, one based on **2-5-3-1 architecture** and another based on **2-20-15-1**





- The results of the fit are very similar (differences << than 1-sigma ranges) if a **huge ANN is used**
- Implies **results are driven by the input experimental data** rather than by the methodological assumptions
- Same level of agreement if **fitting basis is changed**

Parametrisation with ANNs of strange sea in NNPDF2.3 …

$$s^+(x,Q_0^2) = (s+\bar{s})(x,Q_0^2)$$
$$s^-(x,Q_0^2) = (s-\bar{s})(x,Q_0^2)$$

… and in NNPDF3.1

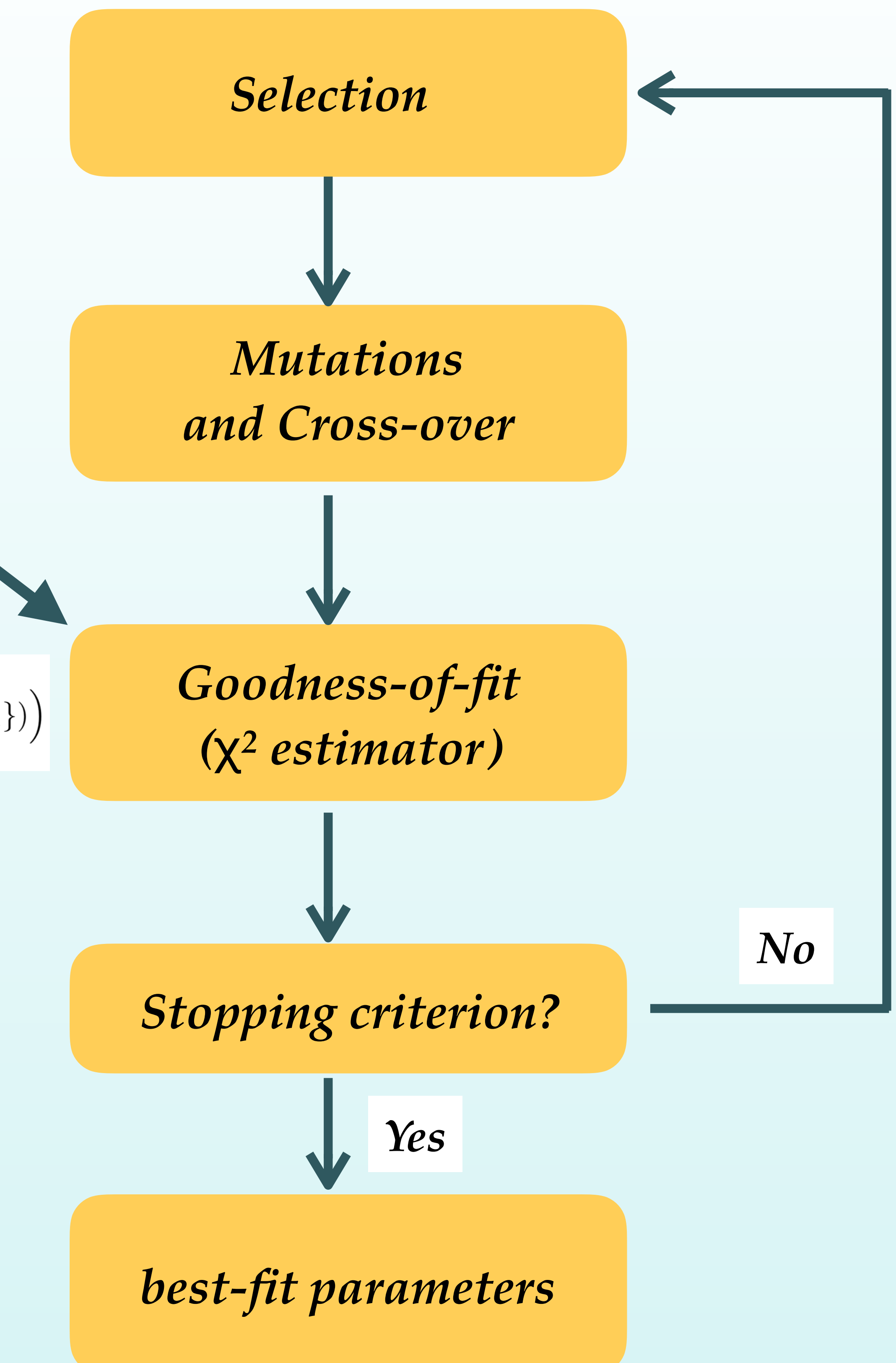
$$\begin{aligned}
\Sigma(x,Q_0^2) &= \left(u+\bar{u}+d+\bar{d}+s+\bar{s}\right)(x,Q_0^2) \\
T_3(x,Q_0^2) &= \left(u+\bar{u}-d-\bar{d}\right)(x,Q_0^2) \\
T_8(x,Q_0^2) &= \left(u+\bar{u}+d+\bar{d}-2s-2\bar{s}\right)(x,Q_0^2) \\
V(x,Q_0^2) &= \left(u-\bar{u}+d-\bar{d}+s-\bar{s}\right)(x,Q_0^2) \\
V_3(x,Q_0^2) &= \left(u-\bar{u}-d+\bar{d}\right)(x,Q_0^2) \\
V_8(x,Q_0^2) &= \left(u-\bar{u}+d-\bar{d}-2s+2\bar{s}\right)(x,Q_0^2),
\end{aligned}$$

ANN training: genetic algorithms

Initial population (random sampling)

Fit parameters: ***weights*** *and* ***thresholds*** *of the ANNs*

$$\chi^2\left(\{a_i\}\right) = \sum_{m,n=1}^{N_{\rm dat}} \left(\sigma_m^{(\rm exp)} - \sigma_m^{(\rm th)}\left(\{a_i\}\right)\right) \left({\rm cov}_{\rm exp} + {\rm cov}_{\rm th}\right)^{-1}_{mn} \left(\sigma_n^{(\rm exp)} - \sigma_n^{(\rm th)}\left(\{a_i\}\right)\right)$$



- GAs specially useful to **avoid getting stuck in local minima**
- No need to **compute analytically χ^2 gradients**
- Low efficiency close to global minima

ANN training: covariance matrix adaptation

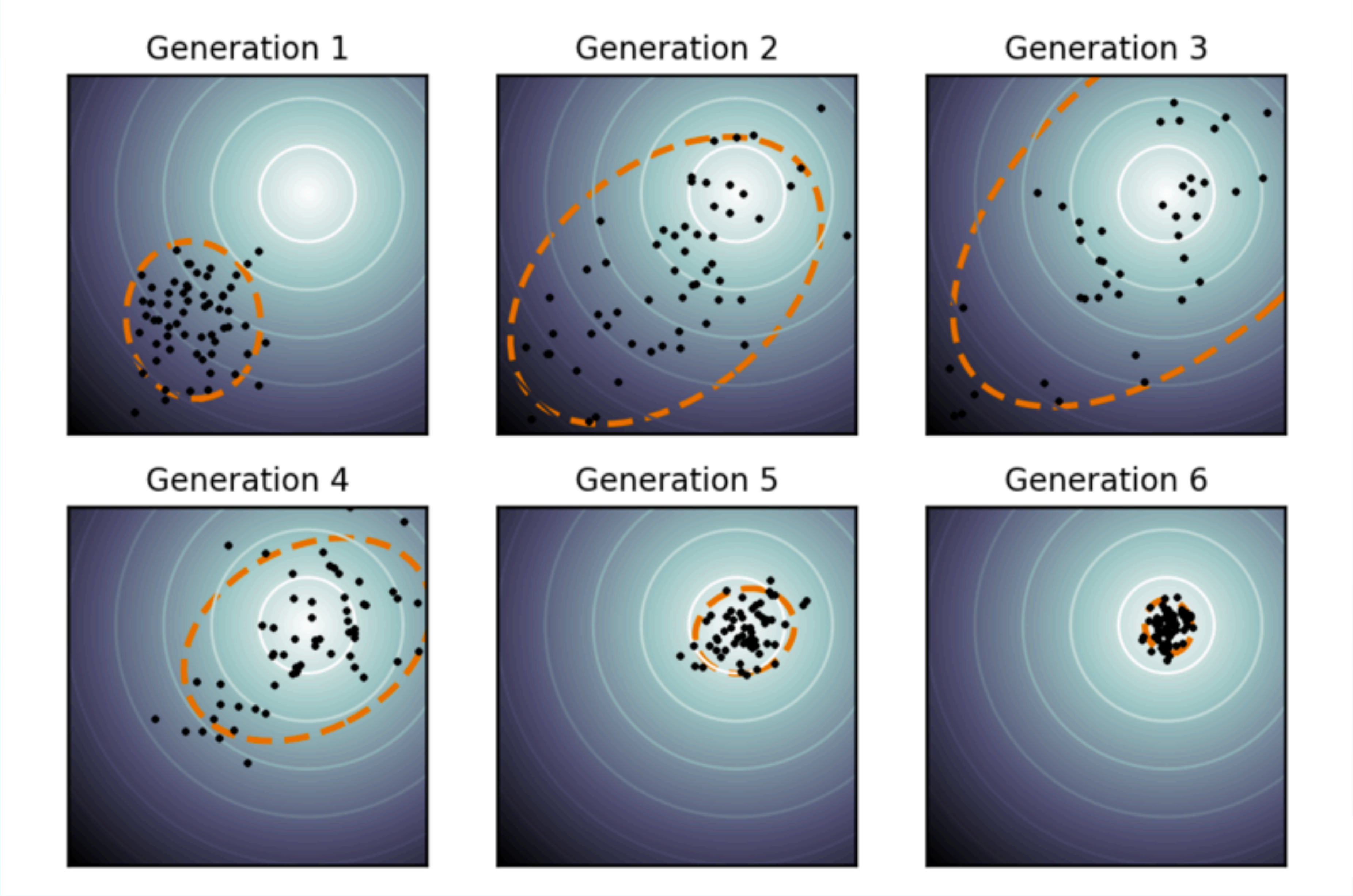
Improved exploration of parameter space by using **information from previous iterations when generating the mutants**, rather than fully random variations as in GA

$$\mathbf{x}_k^{(i)} \sim \mathbf{a}^{(i-1)} + \sigma^{(i-1)}\mathcal{N}(0,\mathbf{C}^{(i-1)})\,, \quad \text{for } k=1\,,\ldots,\lambda\,.$$

mutants

best-fit previous generation

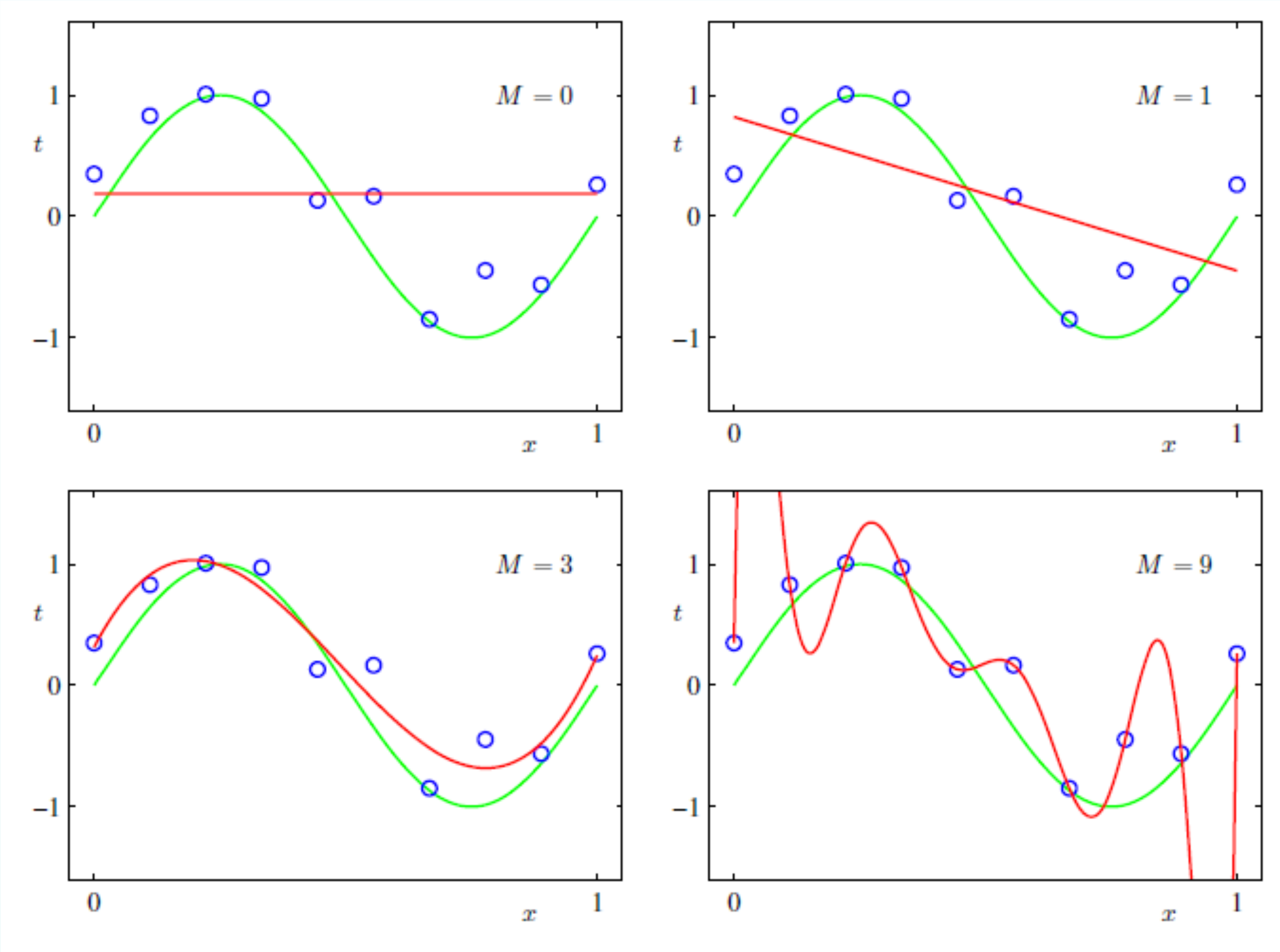
Multigaussian in parameter space



Creative Commons

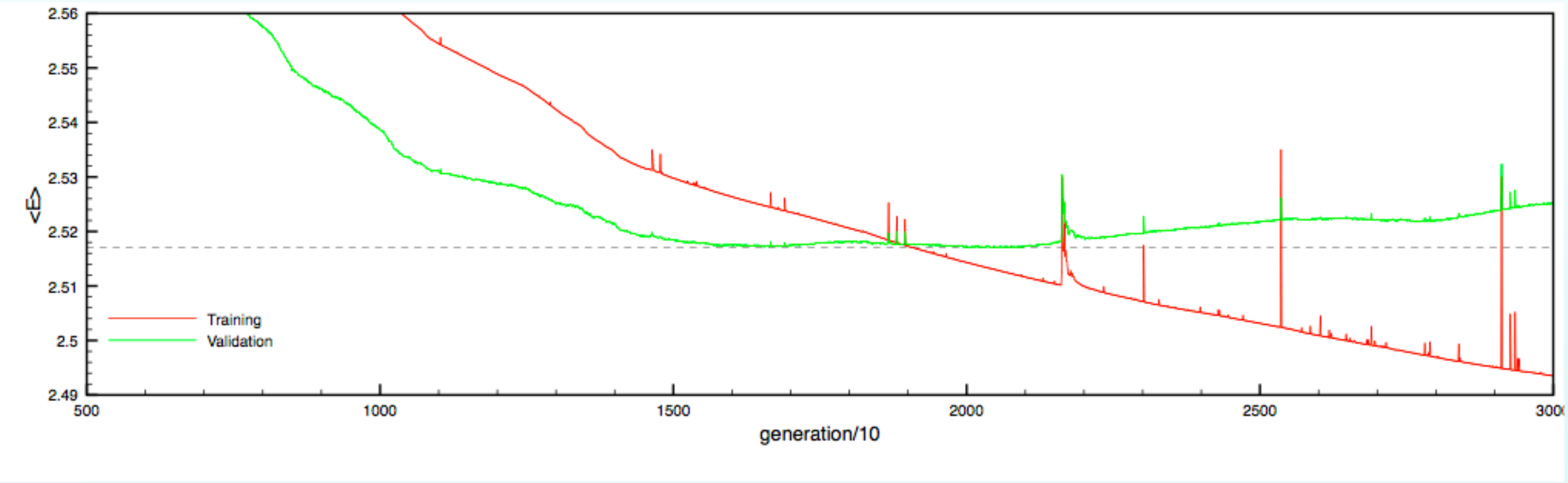
Avoiding overfitting

For a **flexible enough input functional form for the Parton Distributions**, one might end up **fitting statistical fluctuations** rather than the underlying physical law!



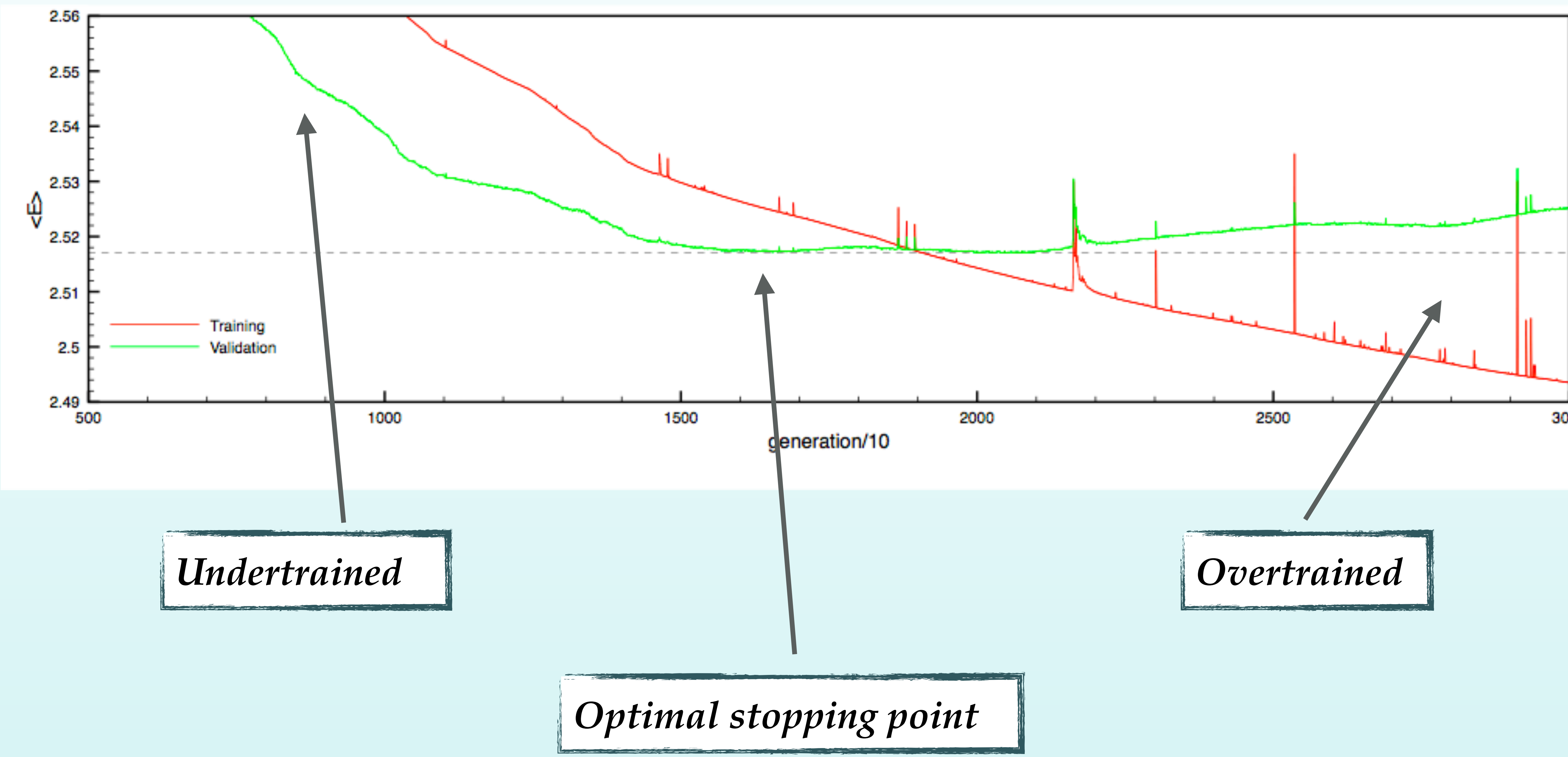
Avoiding overfitting

- Separate the input measurements into a **training** and a **validation** sample



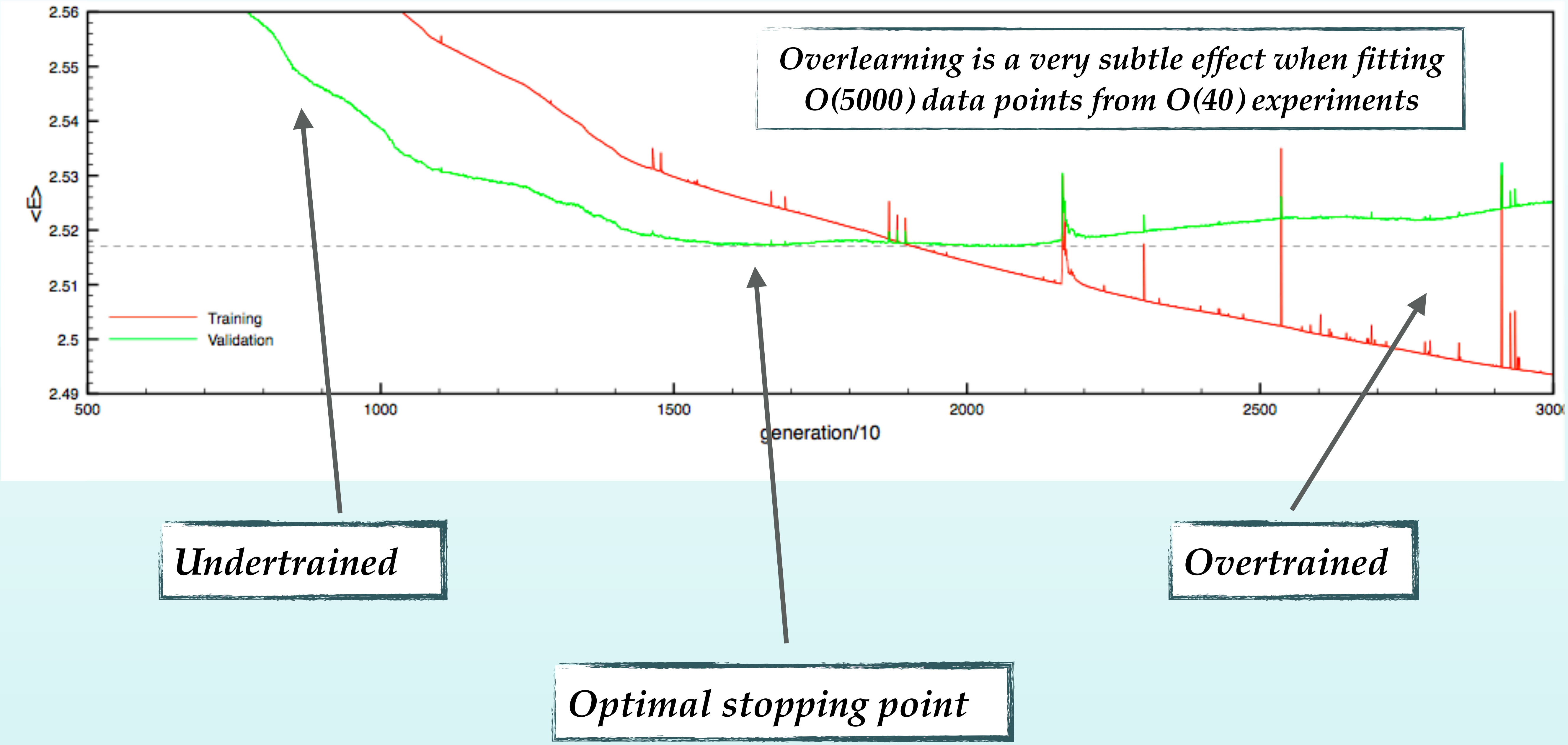
Avoiding overfitting

- Separate the input measurements into a **training** and a **validation** sample
- The validation sample is never trained, only used to monitor the quality of the fit to the training sample
- The optimal stopping point is at the **global minimum of the validation χ^2**



Avoiding overfitting

- Separate the input measurements into a **training** and a **validation** sample
- The validation sample is never trained, only used to monitor the quality of the fit to the training sample
- The optimal stopping point is at the **global minimum of the validation χ^2**



The Monte Carlo method

- Construct a **sampling of the probability distribution in the space of experimental data** based on all available information on central values, uncertainties, and correlations

$$\mathcal{O}_i^{(\mathrm{art})(k)} = S_{i,N}^{(k)} \mathcal{O}_i^{(\mathrm{exp})} \left(1 + \sum_{\alpha=1}^{N_{\mathrm{sys}}} r_{i,\alpha}^{(k)} \sigma_{i,c}^{(\mathrm{sys})} + r_i^{(k)} \sigma_i^{(\mathrm{stat})} \right) , \quad k = 1, \ldots, N_{\mathrm{rep}} \ , \quad i = 1, \ldots, N_{\mathrm{dat}}$$

MC pseudo-data replicas

Correlated Gaussian random numbers

Uncorrelated Gaussian random numbers

number of MC replicas

- A **full global PDF fit** is then performed for each MC replica
- This results into a sampling of the **probability distribution in the space of PDFs** (or LHC cross-sections …)
- From this **any statistical estimator** of the sample can be computed using textbook statistics, *e.g.*

Central values

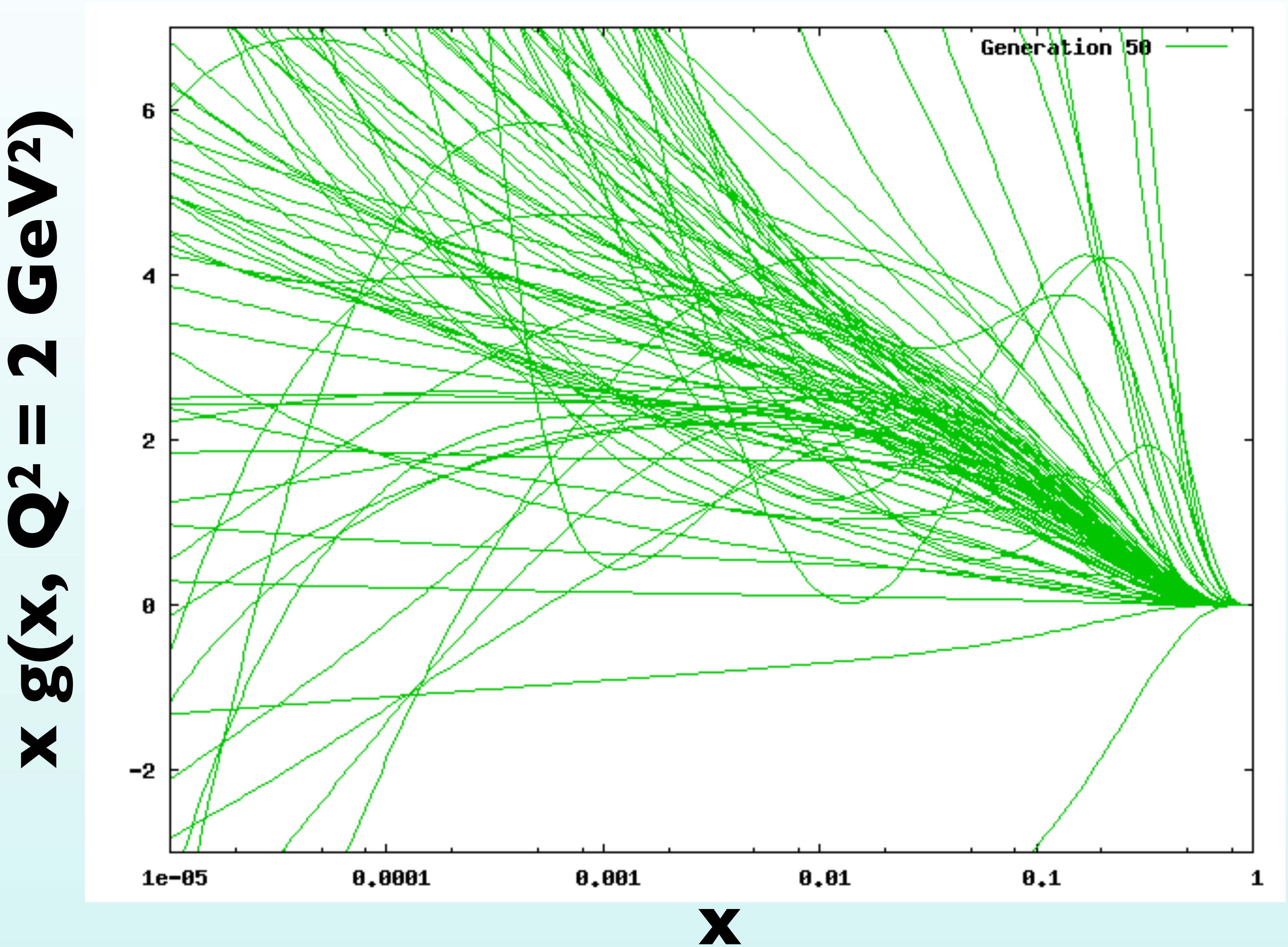
Variances

$$\langle F \rangle_{\mathrm{rep}} = \frac{1}{N_{\mathrm{rep}}} \sum_{k=1}^{N_{\mathrm{rep}}} F^{(k)} , \qquad \delta F = \sqrt{\frac{\sum_{k=1}^{N_{\mathrm{rep}}} \left(\langle F \rangle_{\mathrm{rep}} - F^{(k)} \right)^2}{N_{\mathrm{rep}} - 1}}$$

Cross-section computed from the k-th replica PDF set

Neural network training

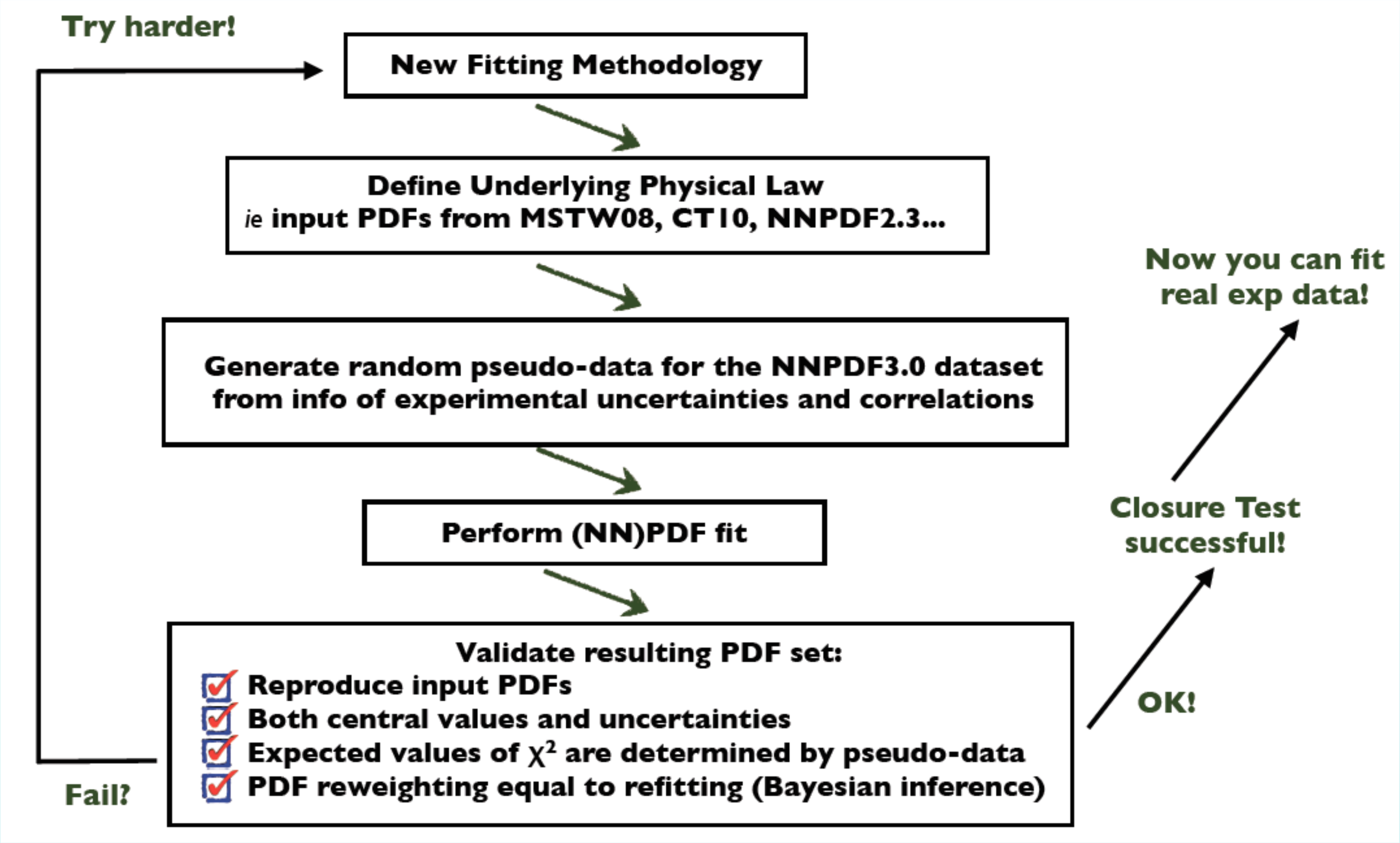
Starting from **random boundary conditions** for the N_{rep} replicas, the ANN training ensures that only those functional forms **minimising the χ^2** are selected



Closure testing the methodology

Methodology validated on closure tests applied to **pseudo-data** generated from a **known underlying theory**

Inadequacies in fitting methodology can be disentangled from *e.g.* data inconsistencies or theory limitations

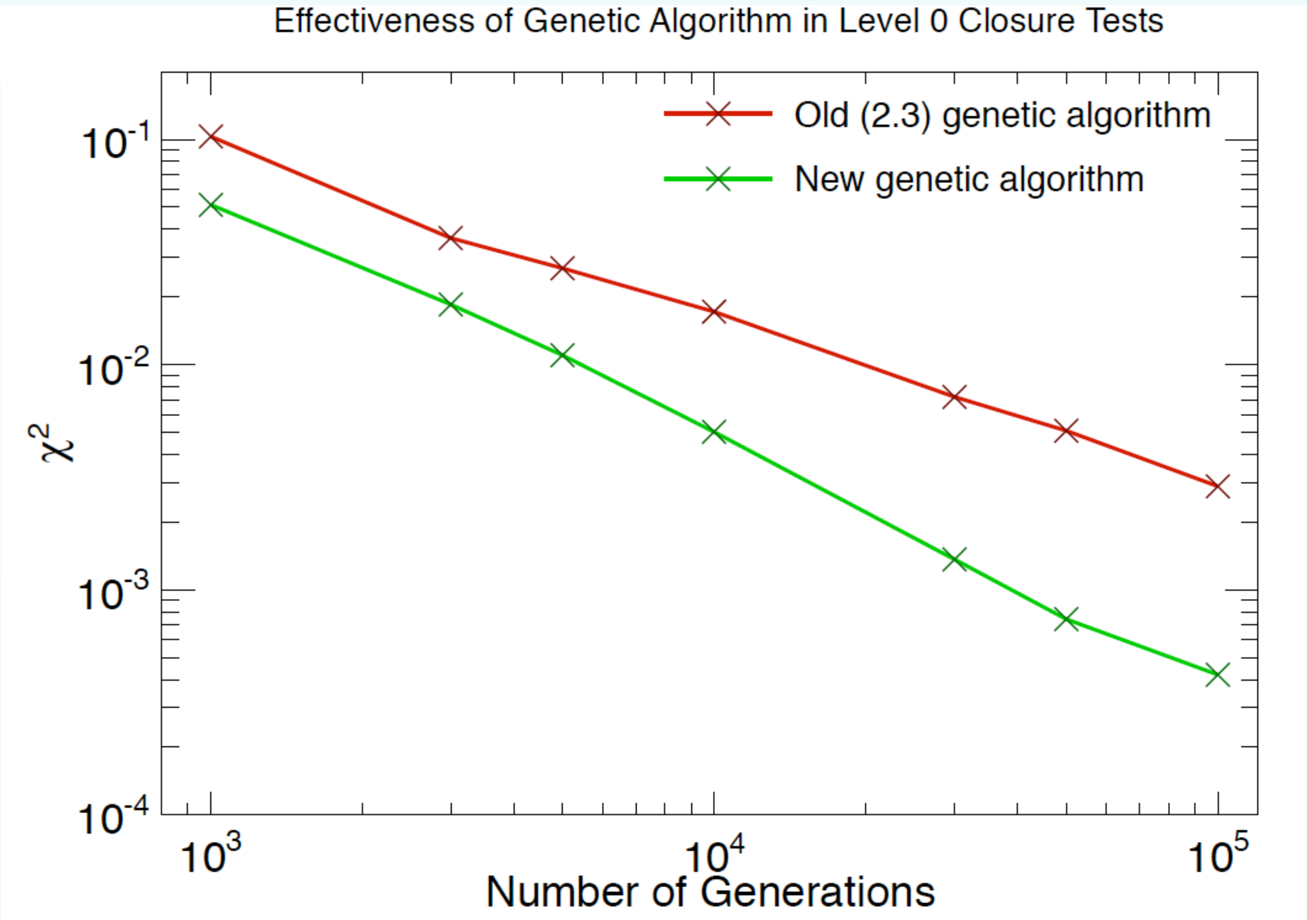


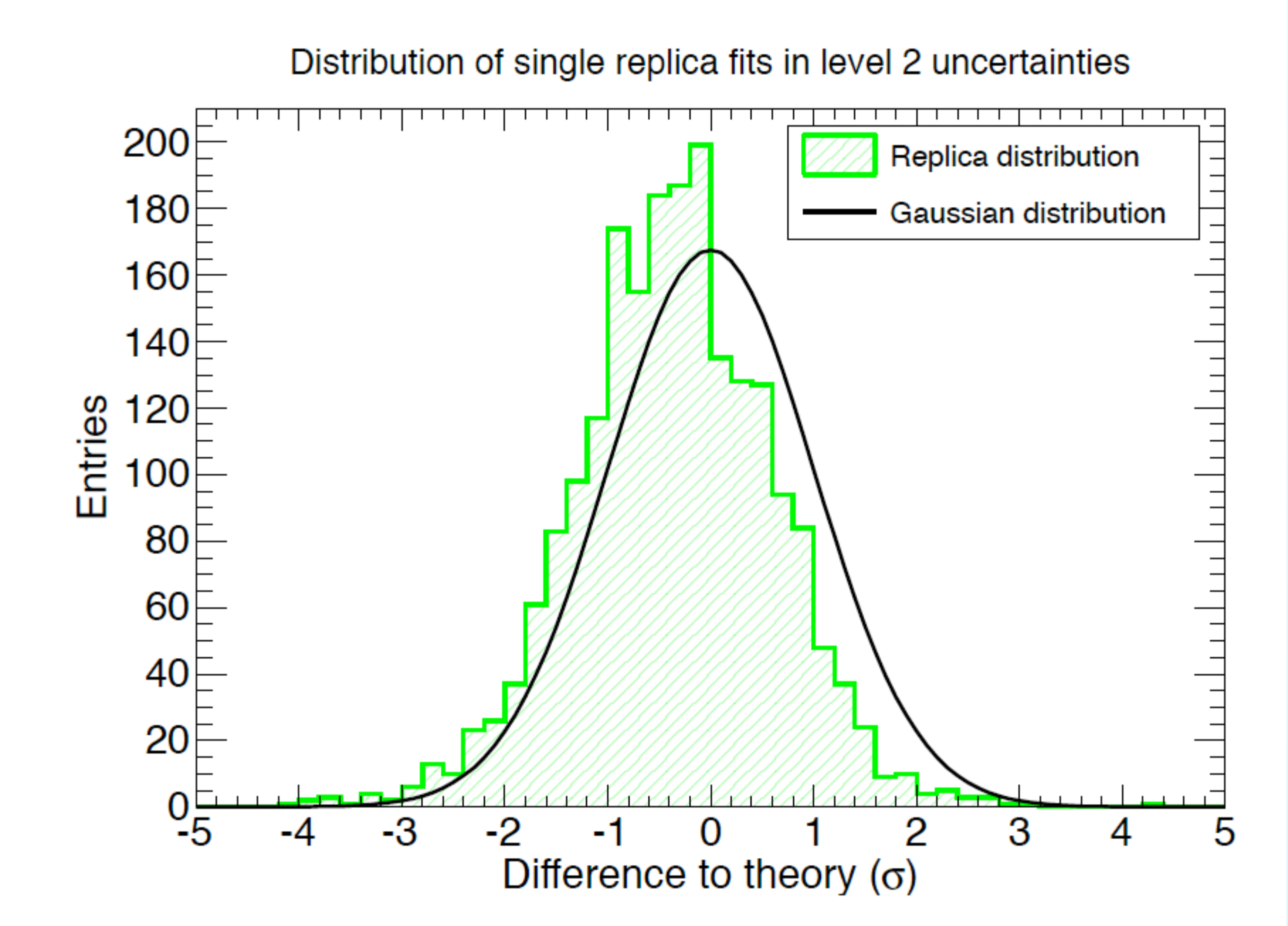
Closure testing the methodology

Methodology validated on closure tests applied to **pseudo-data** generated from a **known underlying theory**

Inadequacies in fitting methodology can be disentangled from *e.g.* data inconsistencies or theory limitations

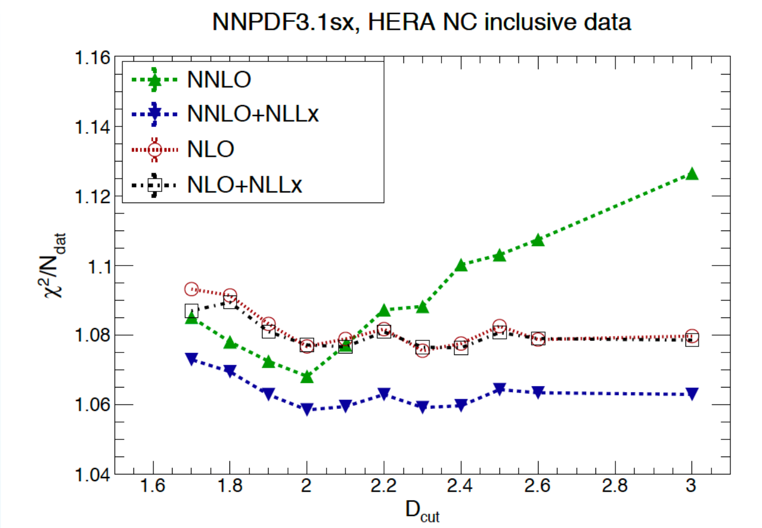
- ☑ **Level 0: no fluctuations** on pseudo-data, **no Monte Carlo** replica generation
- ☑ **Level 1**: **with fluctuations** on pseudo-data, **no Monte Carlo** replica generation
- ☑ **Level 2**: **with fluctuations** on pseudo-data, **with Monte Carlo** replica generation



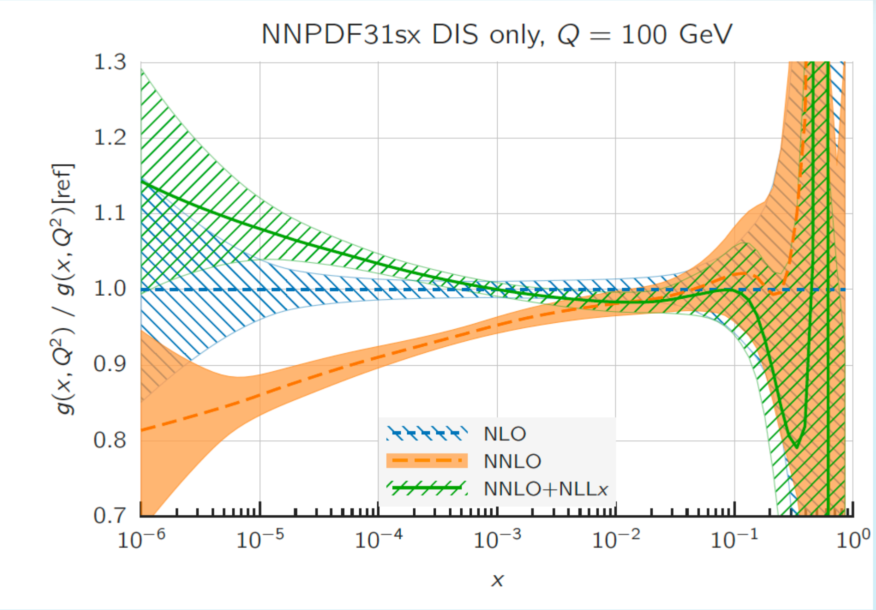


χ^2 can become arbitrarily small

*Central values of **68% of the fits** should fall within the **1-sigma PDF errors***



Discovering ``New Physics" from the global QCD analysis



AKA when exploiting a robust PDF fitting methodology really pays off!

Discovering New Physics within QCD

How we can ensure that we are not ``**fitting way" BSM effects** in the global PDF analysis?

Our recent discovery of **BFKL effects in HERA data** illustrates how this can be achieved!

- At small-x, **logarithmically enhanced terms in *1/x* become dominant** and need to be resummed
- **BFKL/high-energy/small-x resummation** can be matched to the **DGLAP collinear framework**
- Until recently**,** no conclusive evidence for the onset of BFKL dynamics had been provided

DGLAP Evolution in Q^2

$$\mu^2 \frac{\partial}{\partial \mu^2} f_i(x,\mu^2) = \int_x^1 \frac{dz}{z} P_{ij}\left(\frac{x}{z}, \alpha_s(\mu^2)\right) f_j(z,\mu^2),$$

BFKL Evolution in x

$$-x\frac{d}{dx} f_+(x,\mu^2) = \int_0^\infty \frac{d\nu^2}{\nu^2} K\left(\frac{\mu^2}{\nu^2}, \alpha_s\right) f_+(x,\nu^2)$$

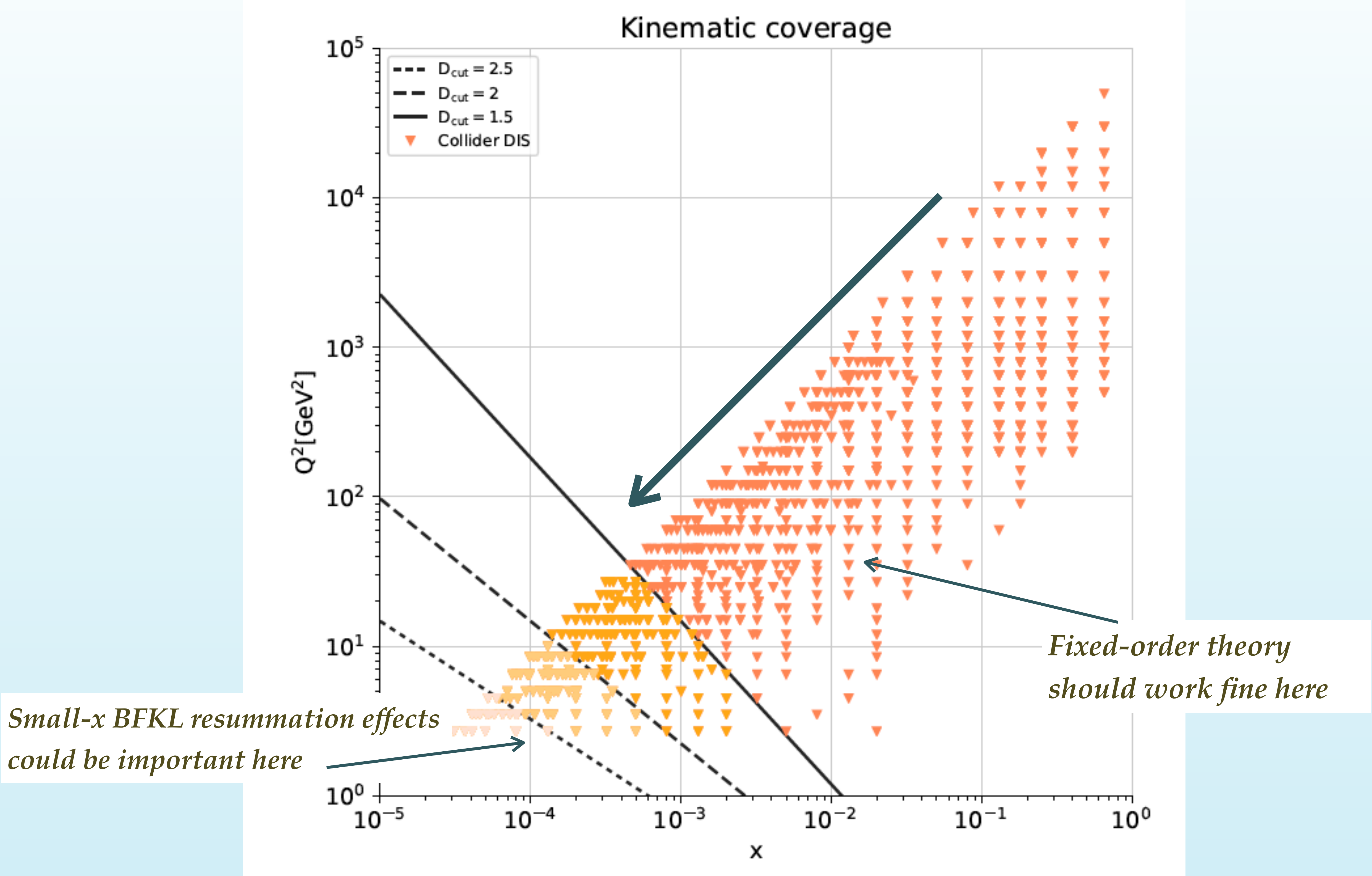
Within small-*x* resummation, the **N^kLO fixed-order DGLAP splitting functions** are complemented with the **N^hLL*x* contributions from BKFL**

$$P_{ij}^{\mathrm{N}^k\mathrm{LO}+\mathrm{N}^h\mathrm{LL}x}(x) = P_{ij}^{\mathrm{N}^k\mathrm{LO}}(x) + \Delta_k P_{ij}^{\mathrm{N}^h\mathrm{LL}x}(x),$$

ABF, CCSS, TW + others, 94-08

Evidence for BFKL dynamics in HERA data

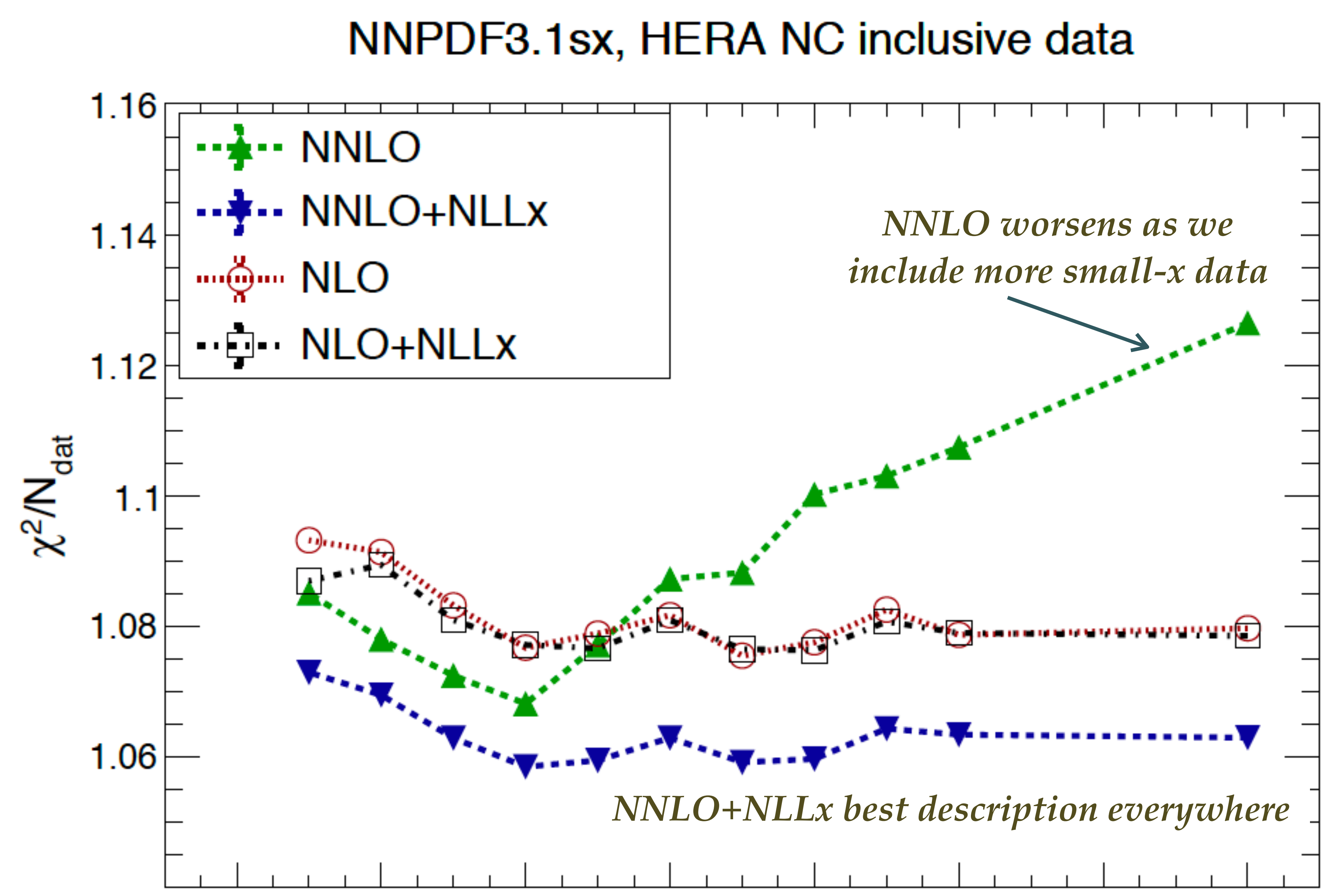
In order to assess the impact of small-x resummation for the description of the **small-x and Q^2 HERA data,** compute the χ^2 **removing data points in the region where resummation effects are expected**



Evidence for BFKL dynamics in HERA data

Using **NNLO+NLL*x* theory, the NNLO instability at small-*x* of the χ^2 disappears**

Excellent fit quality to **inclusive and charm HERA** data achieved in the **entire (x,Q²) region**



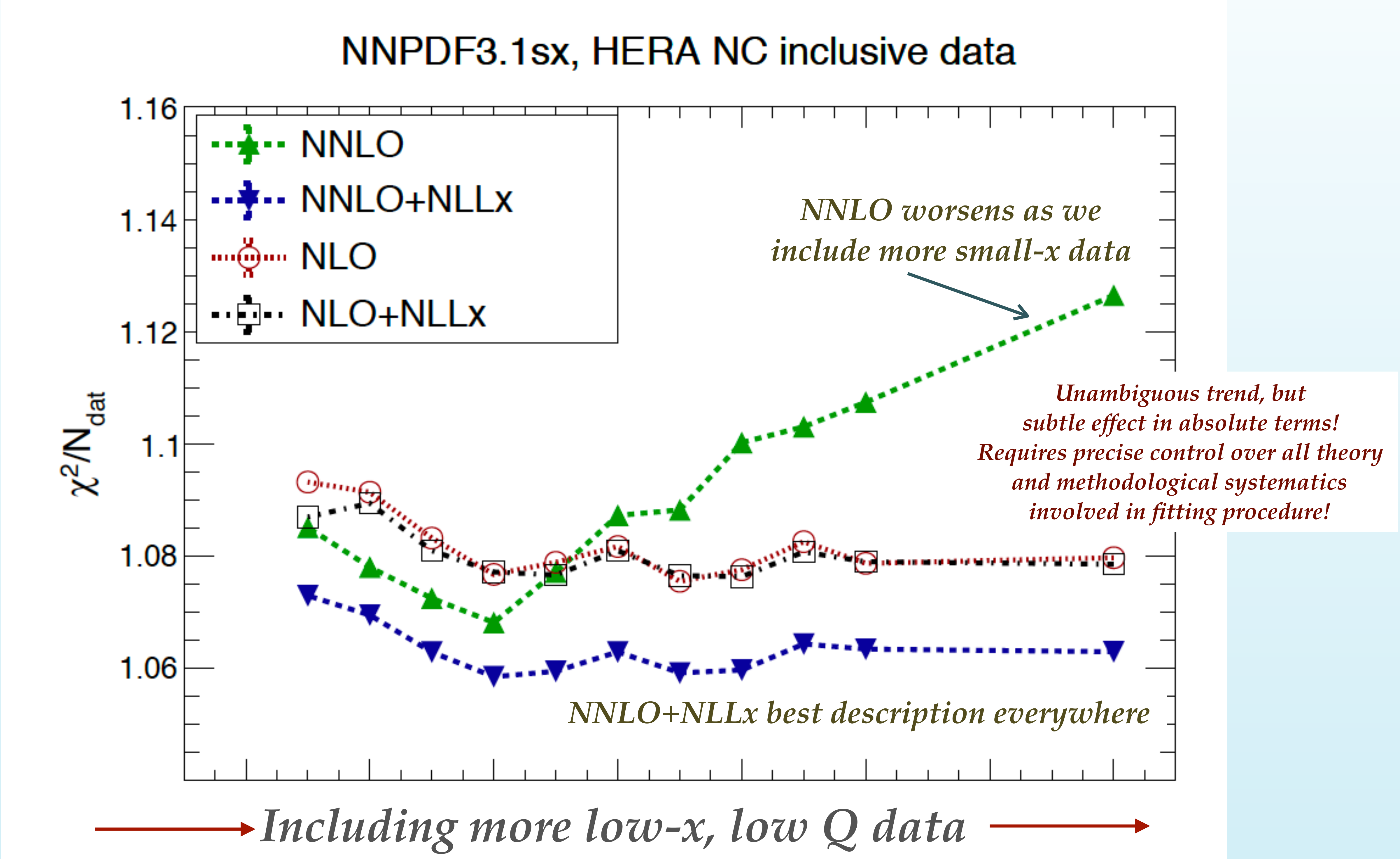
Including more low-x, low Q data

NNPDF 17

Evidence for BFKL dynamics in HERA data

Using **NNLO+NLLx theory, the NNLO instability at small-x of the χ^2 disappears**

Excellent fit quality to **inclusive and charm HERA** data achieved in the **entire (x,Q²) region**



NNPDF 17

Science
Life and Physics

After 40 years of studying the strong nuclear force, a revelation

This was the year that analysis of data finally backed up a prediction, made in the mid 1970s, of a surprising emergent behaviour in the strong nuclear force

Jon Butterworth

@jonmbutterworth
Thu 28 Dec 2017 17.30 GMT

529 | 59

In the mid 1970s, four Soviet physicists, Batlisky, Fadin, Kuraev and Lipatov, made some predictions involving the strong nuclear force which would lead to their initials entering the lore. "BFKL" became a shorthand for a difficult-to-

Jon Butterworth, The Guardian

Machine Learning for Global PDF fits

- ☑ Machine Learning tools are becoming increasingly important in the **toolbox of HEP physicists**

- ☑ ML algorithms are relevant in various aspects of the global PDF fit, from **unbiased parametrisation of the boundary conditions** to **efficient exploration of high-dimensionality parameter spaces**

- ☑ The validation of novel fitting strategies should ideally be tested at the **closure test level,** to avoid interference with unrelated issues such as data incompatibilities

- ☑ Parton distributions could be the key for **unravelling new physics at the LHC!**