

**NEURAL NETWORKS,**  
**PROBABILITY DISTRIBUTIONS,**  
**AND STRUCTURE FUNCTIONS**

STEFANO FORTE  
UNIVERSITÀ DI MILANO

LECCE, APRIL 24, 2003

## SUMMARY

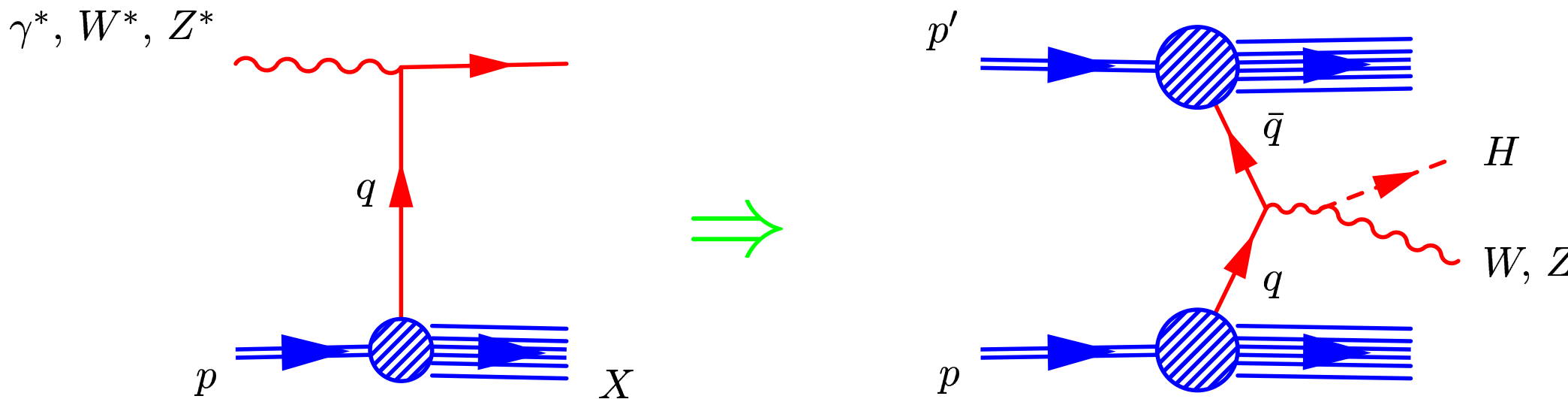
- MOTIVATION: PARTON DISTRIBUTIONS WITH ERRORS
- THE PROBLEM: FUNCTIONAL INTEGRALS FROM DATA
- THE SOLUTION: THE NEURAL MONTECARLO

## CREDITS

- NEURAL NETWORK PARAMETRIZATION OF STRUCTURE FUNCTIONS  
S. F., Lluís Garrido, José I. Latorre and Andrea Piccione, *JHEP* **205**, 62 (2002)
- TRUNCATED MOMENTS OF PARTON DISTRIBUTIONS  
S. F. and Lorenzo Magnea, *Phys. Lett.* **B448**, 295 (1999); S. F., Lorenzo Magnea, Giovanni Ridolfi and Andrea Piccione, *Nucl. Phys.* **B594**, 46 (2001); Andrea Piccione, *Phys. Lett.* **B518**, 207 (2001)
- UNBIASED DETERMINATION OF  $\alpha_s$   
S. F., José I. Latorre, Lorenzo Magnea and Andrea Piccione, *Nucl. Phys.* **B643** (2002) 477

# THE CONTEXT: FACTORIZATION

THE ACCURATE COMPUTATION OF PHYSICAL PROCESS AT A HADRON COLLIDER  
REQUIRES GOOD KNOWLEDGE OF PARTON DISTRIBUTIONS OF THE NUCLEON



IN ORDER TO EXTRACT THE RELEVANT PHYSICS SIGNAL,  
WE NEED TO KNOW THE ERROR ON THE PARTON DISTRIBUTION

# THE MOTIVATION: AN EXAMPLE: THE “NUTeV ANOMALY”

THE “PASCHOS-WOLFENSTEIN RATIO” RELATES TOTAL NEUTRINO-NUCLEON DIS  
CROSS-SECTIONS TO THE WEAK MIXING ANGLE:

$$\frac{\sigma_{NC}(\nu) - \sigma_{NC}(\bar{\nu})}{\sigma_{CC}(\nu) - \sigma_{CC}(\bar{\nu})} = \frac{1}{2} - \sin^2 \theta_W + \left( \frac{1}{2} - \frac{7}{6} \sin^2 \theta_W \right) \left[ -2 \frac{s - \bar{s}}{u - \bar{u} + d - \bar{d}} \right]$$

$u, d, \dots$  denote the fraction of the nucleon's momentum carried by the respective quarks

# THE MOTIVATION: AN EXAMPLE: THE “NUTeV ANOMALY”

THE “PASCHOS-WOLFENSTEIN RATIO” RELATES TOTAL NEUTRINO-NUCLEON DIS CROSS-SECTIONS TO THE WEAK MIXING ANGLE:

$$\frac{\sigma_{NC}(\nu) - \sigma_{NC}(\bar{\nu})}{\sigma_{CC}(\nu) - \sigma_{CC}(\bar{\nu})} = \frac{1}{2} - \sin^2 \theta_W + \left( \frac{1}{2} - \frac{7}{6} \sin^2 \theta_W \right) \left[ -2 \frac{s - \bar{s}}{u - \bar{u} + d - \bar{d}} \right]$$

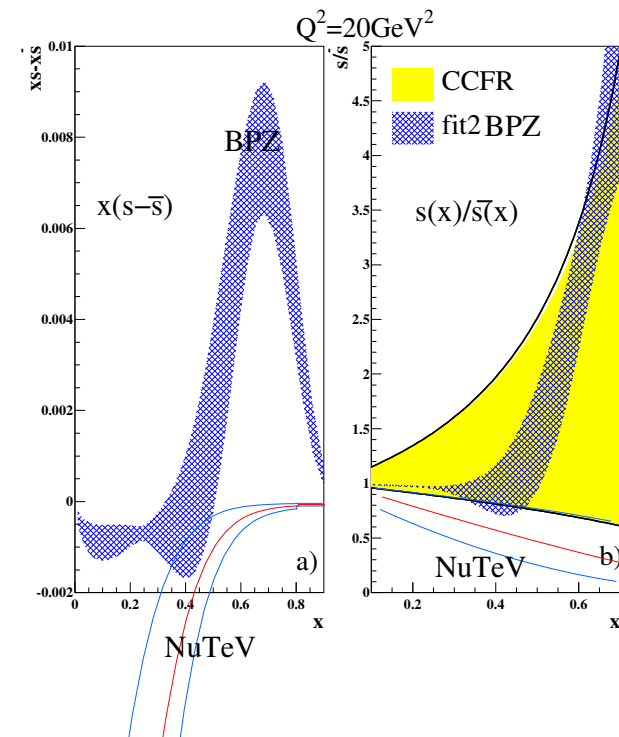
$u, d, \dots$  denote the fraction of the nucleon’s momentum carried by the respective quarks

CAN ONE NEGLECT THE  $s - \bar{s}$  CONTRIBUTION?

— NUTeV (2001) **NEGLECTS IT & GETS**  $\sin^2 \theta_W$  **THAT DISAGREES BY  $3\sigma$  WITH SM FIT**

—  $s - \bar{s} = 0.003$  **REMOVES THE DISCREPANCY** (DAVIDSON ET AL. (2002))

- $q - \bar{q}$  HARD TO DETERMINE IN DIS BECAUSE  $\gamma^*$  COU- PLES THROUGH (ELECTRIC CHARGE)<sup>2</sup>
- BARONE ET AL. (2000) GET  $s - \bar{s} = +0.002$ ,  
NUTeV (2002) CLAIM  $s - \bar{s} = -0.003$



# THE NAME OF THE GAME

DIS DATA → STRUCTURE FUNCTIONS (FORM FACTORS, DEP. ON KIN. VARIABLES  $x, Q^2$ )

STRUCTURE FUNCTION = HARD COEFF.  $\otimes$  PARTON DISTN.

$$F_2^{\text{NC}}(x, Q^2) = x \sum_{\text{flav. } i} e_i^2 (q_i + \bar{q}_i) + \alpha_s [C_i[\alpha_s] \otimes (q_i + \bar{q}_i) + C_g[\alpha_s] \otimes g]$$

- TRIVIAL COMPLICATIONS: DISENTANGLE INDIVIDUAL QUARK & GLUON CONTRIBUTION TO STRUCTURE FUNCTION; EVOLVE TO COMMON SCALE; DECONVOLUTE see below: truncated moms.
- SERIOUS COMPLICATION: DETERMINE ERROR ON FUNCTIONS  $f(x)$ ,  $f = q_i, \bar{q}_i, g$

# THE NAME OF THE GAME

DIS DATA → STRUCTURE FUNCTIONS (FORM FACTORS, DEP. ON KIN. VARIABLES  $x$ ,  $Q^2$ )

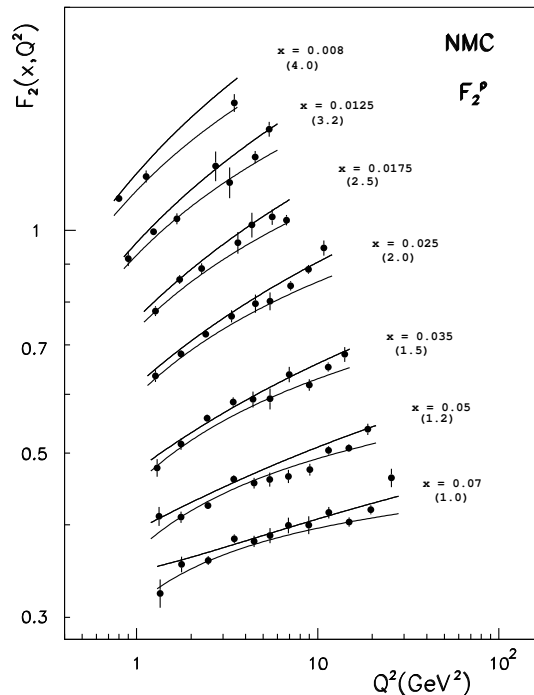
STRUCTURE FUNCTION=HARD COEFF.⊗PARTON DISTN.

$$F_2^{\text{NC}}(x, Q^2) = x \sum_{\text{flav. } i} e_i^2 (q_i + \bar{q}_i) + \alpha_s [C_i[\alpha_s] \otimes (q_i + \bar{q}_i) + C_g[\alpha_s] \otimes g]$$

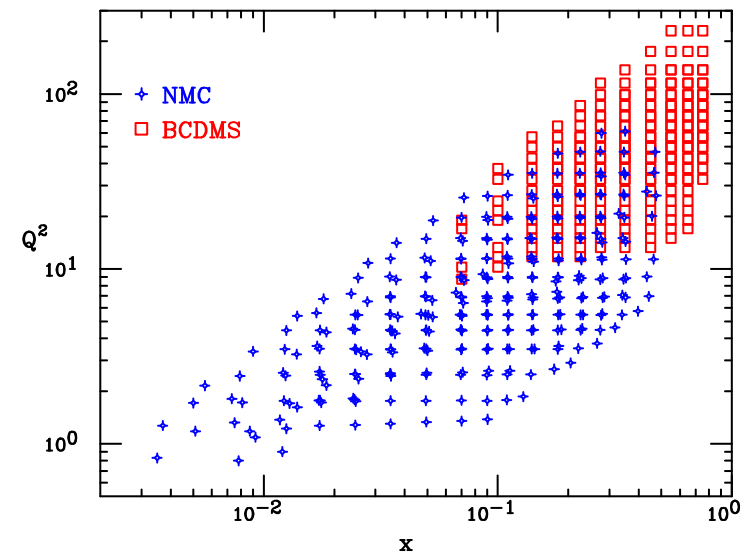
- TRIVIAL COMPLICATIONS: DISENTANGLE INDIVIDUAL QUARK & GLUON CONTRIBUTION TO STRUCTURE FUNCTION; EVOLVE TO COMMON SCALE; DECONVOLUTE see below: truncated moms.

- SERIOUS COMPLICATION: DETERMINE ERROR ON FUNCTIONS  $f(x)$ ,  $f = q_i, \bar{q}_i, g$

A (MARGINALLY) SIMPLER PROBLEM: DETERMINE THE STRUCTURE FUNCTION



GIVEN A BUNCH OF EXPERIMENTAL DATA  $F_2(x, Q^2)$  AT POINTS  $(x_i, Q_i^2)$ , WITH STAT. ERRORS (fig.→bars) AND CORRELATED SYST. ERRORS (fig.→bands) DETERMINE THE STRUCTURE FUNCTION AND ASSOCIATE ERROR



# WHAT'S THE PROBLEM? D. Kosower, 1999

- FOR A SINGLE QUANTITY, WE QUOTE 1 SIGMA ERRORS: VALUE  $\pm$  ERROR
- FOR A PAIR OF NUMBERS, WE QUOTE A 1 SIGMA ELLIPSE
- FOR A FUNCTION, WE NEED AN “ERROR BAR” IN A SPACE OF FUNCTIONS

MUST DETERMINE THE PROBABILITY DENSITY (MEASURE)  $\mathcal{P}[F_2]$

IN THE SPACE OF FUNCTIONS  $F_2(x, Q^2)$



# WHAT'S THE PROBLEM? D. Kosower, 1999

- FOR A SINGLE QUANTITY, WE QUOTE 1 SIGMA ERRORS: VALUE  $\pm$  ERROR
- FOR A PAIR OF NUMBERS, WE QUOTE A 1 SIGMA ELLIPSE
- FOR A FUNCTION, WE NEED AN “ERROR BAR” IN A SPACE OF FUNCTIONS

MUST DETERMINE THE PROBABILITY DENSITY (MEASURE)  $\mathcal{P}[F_2]$

IN THE SPACE OF FUNCTIONS  $F_2(x, Q^2)$

EXPECTATION VALUE OF  $\mathcal{F}[F_2(x, Q^2)] \Rightarrow$  FUNCTIONAL INTEGRAL

$$\left\langle \mathcal{F}[F_2(x, Q^2)] \right\rangle = \int \mathcal{D}F_2 \mathcal{F}[F_2(x, Q^2)] \mathcal{P}[F_2],$$

# WHAT'S THE PROBLEM? D. Kosower, 1999

- FOR A SINGLE QUANTITY, WE QUOTE 1 SIGMA ERRORS: VALUE  $\pm$  ERROR
- FOR A PAIR OF NUMBERS, WE QUOTE A 1 SIGMA ELLIPSE
- FOR A FUNCTION, WE NEED AN “ERROR BAR” IN A SPACE OF FUNCTIONS

MUST DETERMINE THE PROBABILITY DENSITY (MEASURE)  $\mathcal{P}[F_2]$

IN THE SPACE OF FUNCTIONS  $F_2(x, Q^2)$

EXPECTATION VALUE OF  $\mathcal{F}[F_2(x, Q^2)] \Rightarrow$  FUNCTIONAL INTEGRAL

$$\langle \mathcal{F}[F_2(x, Q^2)] \rangle = \int \mathcal{D}F_2 \mathcal{F}[F_2(x, Q^2)] \mathcal{P}[F_2],$$

MUST DETERMINE AN INFINITE-DIMENSIONAL OBJECT  
FROM A FINITE SET OF DATA POINTS

# SOLUTIONS. . .

- CHOOSE A FIXED FUNCTIONAL FORM, E.G. (SMC, 1998)

$$F_2(x, Q^2) = x^{a_1} f(x, Q^2)$$

$$f(x, Q^2) = A(x) \left[ \frac{\log Q^2 / \Lambda^2}{\log Q_0^2 / \Lambda^2} \right]^{B(x)} \left[ 1 + \frac{C(x)}{Q^2} \right]$$

$$A(x) = (1-x)^{a_2} [a_3 + a_4(1-x) + a_5(1-x)^2 + a_6(1-x)^3 + a_7(1-x)^4]$$

$$B(x) = b_1 + b_2 x + \frac{b_3}{x+b_4}$$

$$C(x) = c_1 x + c_2 x^2 + c_3 x^3 + c_4 x^4$$

PROBLEM PROJECTED ONTO THE FINITE-DIMENSIONAL SPACE OF PARAMETERS

WHAT IS THE BIAS (THEOR. ERROR) DUE TO THE CHOICE OF FUNCTIONAL FORM?

# SOLUTIONS. . .

- CHOOSE A FIXED FUNCTIONAL FORM, E.G. (SMC, 1998)

$$F_2(x, Q^2) = x^{a_1} f(x, Q^2)$$

$$f(x, Q^2) = A(x) \left[ \frac{\log Q^2 / \Lambda^2}{\log Q_0^2 / \Lambda^2} \right]^{B(x)} \left[ 1 + \frac{C(x)}{Q^2} \right]$$

$$A(x) = (1-x)^{a_2} [a_3 + a_4(1-x) + a_5(1-x)^2 + a_6(1-x)^3 + a_7(1-x)^4]$$

$$B(x) = b_1 + b_2 x + \frac{b_3}{x+b_4}$$

$$C(x) = c_1 x + c_2 x^2 + c_3 x^3 + c_4 x^4$$

PROBLEM PROJECTED ONTO THE FINITE-DIMENSIONAL SPACE OF PARAMETERS

WHAT IS THE BIAS (THEOR. ERROR) DUE TO THE CHOICE OF FUNCTIONAL FORM?

- EXPAND OVER A FINITE SET OF BASIS FUNCTIONS, E.G. ORTHOGONAL POLYNOMIALS (Yndurain 1975, Parisi, Sourlas 1976, Furmański, Petronzio, 1982)

PROBLEM PROJECTED ONTO THE FINITE-DIMENSIONAL SPACE OF EXPANSION

COEFFICIENTS

WHAT IS THE BIAS (THEOR. ERROR) DUE TO THE CHOICE OF TRUNCATION?

E.g. assume a periodic f. is expanded over a basis of ortho. polynomials, or a non-periodic f is Fourier-expanded . . .

# SOLUTIONS. . .

- CHOOSE A FIXED FUNCTIONAL FORM, E.G. (SMC, 1998)

$$F_2(x, Q^2) = x^{a_1} f(x, Q^2)$$
$$f(x, Q^2) = A(x) \left[ \frac{\log Q^2 / \Lambda^2}{\log Q_0^2 / \Lambda^2} \right]^{B(x)} \left[ 1 + \frac{C(x)}{Q^2} \right]$$
$$A(x) = (1-x)^{a_2} [a_3 + a_4(1-x) + a_5(1-x)^2 + a_6(1-x)^3 + a_7(1-x)^4]$$
$$B(x) = b_1 + b_2 x + \frac{b_3}{x+b_4}$$
$$C(x) = c_1 x + c_2 x^2 + c_3 x^3 + c_4 x^4$$

PROBLEM PROJECTED ONTO THE FINITE-DIMENSIONAL SPACE OF PARAMETERS

WHAT IS THE BIAS (THEOR. ERROR) DUE TO THE CHOICE OF FUNCTIONAL FORM?

- EXPAND OVER A FINITE SET OF BASIS FUNCTIONS, E.G. ORTHOGONAL POLYNOMIALS (Yndurain 1975, Parisi, Sourlas 1976, Furmański, Petronzio, 1982)

PROBLEM PROJECTED ONTO THE FINITE-DIMENSIONAL SPACE OF EXPANSION COEFFICIENTS

WHAT IS THE BIAS (THEOR. ERROR) DUE TO THE CHOICE OF TRUNCATION?

E.g. assume a periodic f. is expanded over a basis of ortho. polynomials, or a non-periodic f is Fourier-expanded . . .

- GENERATE A MONTE-CARLO SAMPLE OF FCTS. W. “REASONABLE” PRIOR DISTN., AND UPDATE FROM DATA USING BAYESIAN INFERENCE (Giele, Kosower, Keller 2001)  
PROBLEM IS MADE FINITE-DIMENSIONAL BY THE CHOICE OF PRIOR, BUT RESULT DO NOT DEPEND ON THE CHOICE IF SUFFICIENTLY GENERAL

HARD TO HANDLE “FLAT DIRECTIONS” (Monte Carlo replicas which lead to same agreement with data); COMPUTATIONALLY VERY INTENSIVE

# THE NEURAL MONTE CARLO APPROACH

**BASIC IDEA:** USE NEURAL NETWORKS AS UNIVERSAL UNBIASED INTERPOLANTS

# THE NEURAL MONTE CARLO APPROACH

**BASIC IDEA:** USE NEURAL NETWORKS AS UNIVERSAL UNBIASED INTERPOLANTS

- GENERATE A SET OF MONTE CARLO REPLICAS  $F_2^{(k)}(x_i, Q^2)$  OF THE ORIGINAL DATASET  $F_2^{(\text{data})}(x_i, Q^2)$  WHICH IS LARGE ENOUGH TO REPRODUCE CENTRAL VALUES (AS AVERAGES), ERRORS (AS VARIANCES) AND CORRELATIONS (AS COVARIANCES)

⇒ REPRESENTATION OF  $\mathcal{P}[F_2]$  AT DISCRETE SET OF POINTS  $(x_i, Q_i^2)$

- TRAIN A NEURAL NET ON EACH REPLICA, THUS OBTAINING A NEURAL REPRESENTATION OF THE FUNCTION  $F_2^{(\text{net})^{(k)}}(x, Q)$

- THE SET OF NEURAL NETS IS A REPRESENTATION OF THE PROBABILITY DENSITY:

$$\left\langle \mathcal{F} [F_2(x, Q^2)] \right\rangle = \frac{1}{N_{rep}} \sum_{k=1}^{N_{rep}} \mathcal{F} [F_2^{(\text{net})^{(k)}}(x, Q^2)]$$

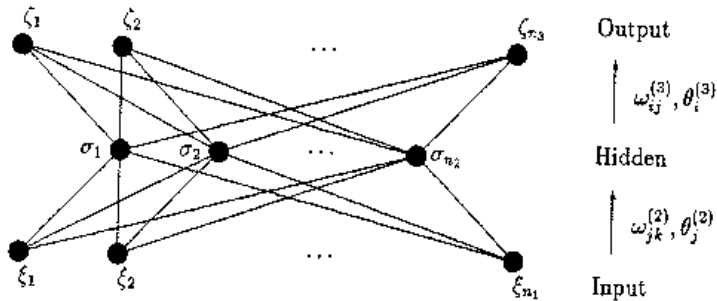
**EXAMPLE: MELLIN MOMENT**

$$\left\langle \int_0^1 dx x^{N-1} F_2(x, Q^2) \right\rangle = \frac{1}{N_{rep}} \sum_{k=1}^{N_{rep}} \int_0^1 dx x^{N-1} F_2^{(\text{net})^{(k)}}(x, Q^2)$$

- CHECK GOODNESS OF FIT THROUGH STATISTICAL INDICATORS ( $\chi^2$ , CORRELATION, . . .)

# NEURAL NETWORKS

## STRUCTURE



## MULTILAYER FEED-FORWARD NETWORKS

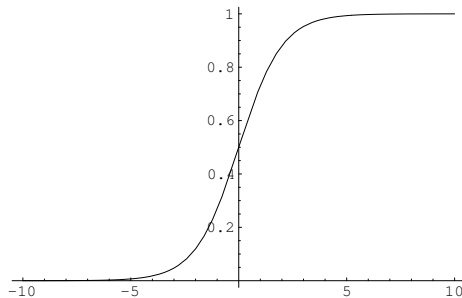
- Each neuron receives input from neurons in preceding layer and feeds output to neurons in subsequent layer

- Activation determined by **weights** and **thresholds**

$$\xi_i = g \left( \sum_j \omega_{ij} \xi_j - \theta_i \right)$$

- Sigmoid activation function

$$g(x) = \frac{1}{1 + e^{-\beta x}}$$



- WEIGHTS & THRESHOLDS CAN BE ADJUSTED SO THAT SIGMOIDS ARE IN CROSSOVER NONLINEAR REGION
- THANKS TO NONLINEAR BEHAVIOUR, ANY FUNCTION CAN BE EXPANDED OVER BASIS OF  $g(x), g(g(x)), g(g(g(x))) \dots$
- CAN CHOOSE REDUNDANT ARCHITECTURE (NO. OF LAYERS & NODES) TO MAKE SURE NO SMOOTHING BIAS IS INTRODUCED



# NEURAL NETWORKS

## TRAINING

### TRAINING BY BACK-PROPAGATION

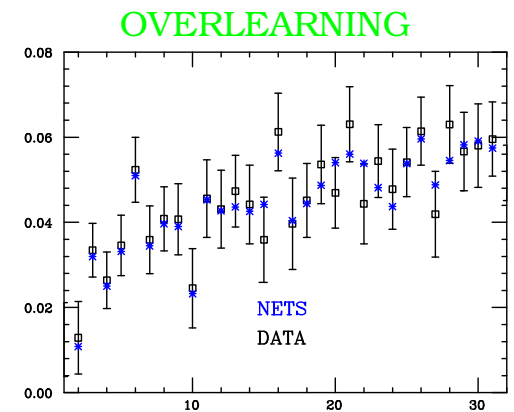
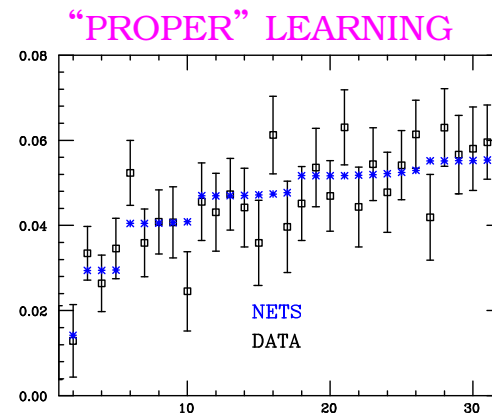
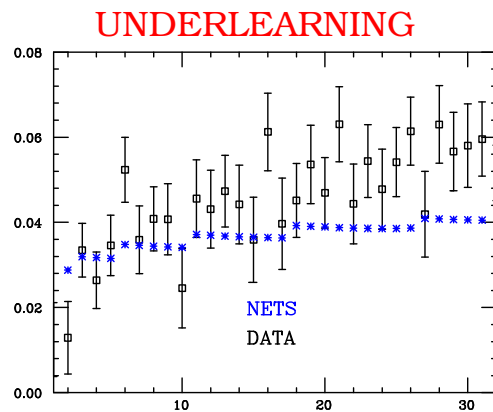
- START WITH RANDOM NETWORK & COMPUTE OUTPUT FOR GIVEN INPUT ( $F_2$  FOR GIVEN  $(x, Q^2)$ )
- COMPARE COMPUTED OUTPUT TO DESIRED OUTPUT BY MEANS OF ENERGY FUNCTION (*e.g.*  $\chi^2$ )
- VARY WEIGHTS AND THRESHOLDS ALONG DIRECTION OF STEEPEST DESCENT OF ENERGY FUNCTION  $\Rightarrow$  CAN BE DONE BY BACK-PROPAGATION
- ITERATE

# NEURAL NETWORKS

## TRAINING

### TRAINING BY BACK-PROPAGATION

- START WITH RANDOM NETWORK & COMPUTE OUTPUT FOR GIVEN INPUT ( $F_2$  FOR GIVEN  $(x, Q^2)$ )
- COMPARE COMPUTED OUTPUT TO DESIRED OUTPUT BY MEANS OF ENERGY FUNCTION (*e.g.*  $\chi^2$ )
- VARY WEIGHTS AND THRESHOLDS ALONG DIRECTION OF STEEPEST DESCENT OF ENERGY FUNCTION  $\Rightarrow$  CAN BE DONE BY BACK-PROPAGATION
- ITERATE



WHEN SHOULD TRAINING STOP?

WHICH IS THE APPROPRIATE ENERGY FUNCTION?

# OPTIMAL TRAINING

WITH LONG ENOUGH TRAINING & BIG ENOUGH NETWORK,  
PREDICTION GOES THROUGH ALL POINTS

any error function proportional to (data-nets) will do: vanishes at minimum.

**Q: DO WE REALLY WANT THIS?**

**NAIVE A: SURE!** Then when averaging over MC sample, at  $(x, Q^2)$  of datapoints averaging over nets is *identical* to averaging over data

**OBJECTION: WHAT IF WE HAVE TWO MEASUREMENTS AT THE SAME  $(x, Q^2)$ ?**

**PERFORM WEIGHTED AVERAGE**  $\frac{F_2^{(1)}/\sigma_1 + F_2^{(2)}/\sigma_2}{1/\sigma_1 + 1/\sigma_2}$  **BEFORE DATA GENERATION.**

**BUT WHAT IF WE HAVE TWO MEASUREMENTS AT  $(x_i, Q_i^2)$  WHICH ARE VERY CLOSE?**

**$F_2$  IS NOT A FRACTAL!**

# OPTIMAL TRAINING

WITH LONG ENOUGH TRAINING & BIG ENOUGH NETWORK,  
PREDICTION GOES THROUGH ALL POINTS

any error function proportional to (data-nets) will do: vanishes at minimum.

**Q: DO WE REALLY WANT THIS?**

**NAIVE A: SURE!** Then when averaging over MC sample, at  $(x, Q^2)$  of datapoints averaging over nets is *identical* to averaging over data

**OBJECTION: WHAT IF WE HAVE TWO MEASUREMENTS AT THE SAME  $(x, Q^2)$ ?**

**PERFORM WEIGHTED AVERAGE**  $\frac{F_2^{(1)}/\sigma_1 + F_2^{(2)}/\sigma_2}{1/\sigma_1 + 1/\sigma_2}$  **BEFORE DATA GENERATION.**

**BUT WHAT IF WE HAVE TWO MEASUREMENTS AT  $(x_i, Q_i^2)$  WHICH ARE VERY CLOSE?**

$F_2$  IS NOT A FRACTAL!

**CLEVER A: ● ERROR FUNCTION → USUAL LOG-LIKELIHOOD**

$$E^{(k)}[\omega, \theta] = \sum_{i=1}^{N_{dat}} \frac{\left( F_i^{(art)(k)} - F_i^{(net)(k)} \right)^2}{\sigma_{i,s}^{(exp)2}}$$

**● ESTABLISH FIXED TRAINING LENGTH SUCH THAT  $\frac{E^{(k)}[\omega, \theta]}{N_{dat}} \approx 1$**

**WHAT ABOUT SYST. ERRORS? TAKEN CARE OF BY MC DATA GENERATION!**

$$F_i^{(net)} \text{ provide best fit of } F_i^{(sys)(k)} \equiv F_i^{(exp)} + \sum_{p=1}^{N_{sys}} r_{i,p}^{(k)} \sigma_{i,p}.$$

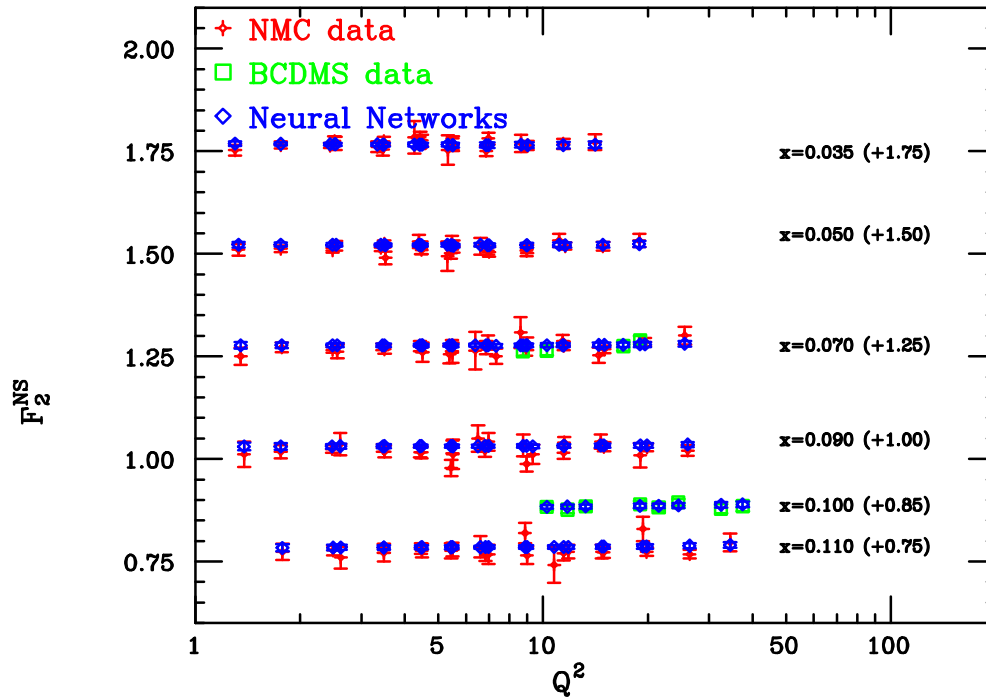
Including systematics in likelihood not practical (nonlocal back-propagation).

**⇒ TRAIN 1000 PROTON, 1000 DEUTERON & 1000 NONSINGLET NETS**

# NEURAL INFORMATION HANDLING I

## COMBINING DATA

NS data vs. neural nets  
 $0.03 < x < 0.12$



IN NONSINGLET CASE,

AVERAGE VARIANCE OF NETS  $\ll$  STAT.

ERROR OF DATA (FACTOR 3-4)

IS IT DUE TO SMOOTHING BIAS?

OR IS IT DUE TO COMBINING DATA?

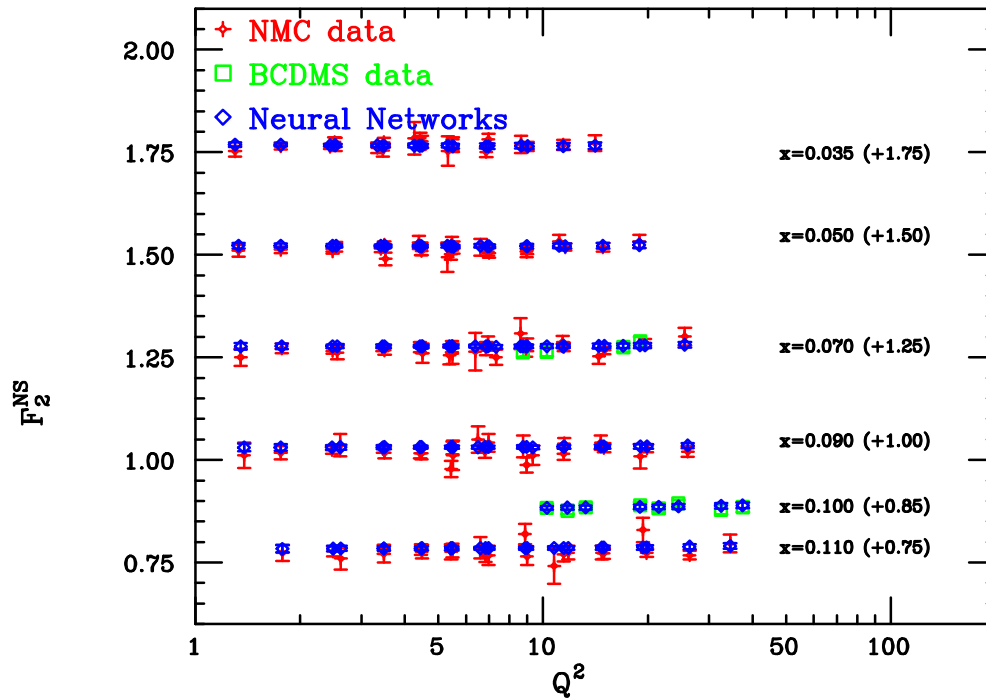
recall error on weighted average

$$\sigma = \frac{1}{1/\sigma_1^2 + 1/\sigma_2^2} < \sigma_i$$

# NEURAL INFORMATION HANDLING I

## COMBINING DATA

NS data vs. neural nets  
 $0.03 < x < 0.12$



IN NONSINGLET CASE,

AVERAGE VARIANCE OF NETS  $\ll$  STAT.

ERROR OF DATA (FACTOR 3-4)

IS IT DUE TO SMOOTHING BIAS?

OR IS IT DUE TO COMBINING DATA?

recall error on weighted average

$$\sigma = \frac{1}{1/\sigma_1^2 + 1/\sigma_2^2} < \sigma_i$$

CAN CONSTRUCT A STATISTICAL  
 INDICATOR TO TELL!

Average error  $\langle E \rangle = \frac{1}{N_{rep}} \sum_{n=1}^{N_{rep}} \sum_{i=1}^{N_{dat}} \frac{\left( F_i^{(art)(n)} - F_i^{(net)(n)} \right)^2}{\sigma_{i,s}^{(exp)2}}$  ( $n \rightarrow$  replica;  $i \rightarrow$  datapoint)

“Central” error  $\langle \tilde{E} \rangle = \frac{1}{N_{rep}} \sum_{n=1}^{N_{rep}} \sum_{i=1}^{N_{dat}} \frac{\left( F_i^{(exp)} - F_i^{(net)(n)} \right)^2}{\sigma_{i,s}^{(exp)2}}$

Bias indicator  $\mathcal{R} \equiv \langle \tilde{E} \rangle / \langle E \rangle$ : if  $\sigma_{net} \ll \sigma_{exp}$  then

$\mathcal{R} \approx 1 \Rightarrow$  BIAS;  $\mathcal{R} \approx 1/2 \Rightarrow$  ERROR REDUCTION      HERE  $\mathcal{R} = 0.58$  (0.53 NMC only)

# NEURAL INFORMATION HANDLING II

STUDY DEPENDENCE OF ERROR FCTN  $E^{(0)} = \frac{1}{N_{dat}} \sum_{i=1}^{N_{dat}} \frac{\left(F_i^{(exp)} - F_i^{(net)(0)}\right)^2}{\sigma_{i,s}^{(exp)2}}$  ON  
TRAINING LENGTH FOR NET TRAINED ON CENTRAL VALUES

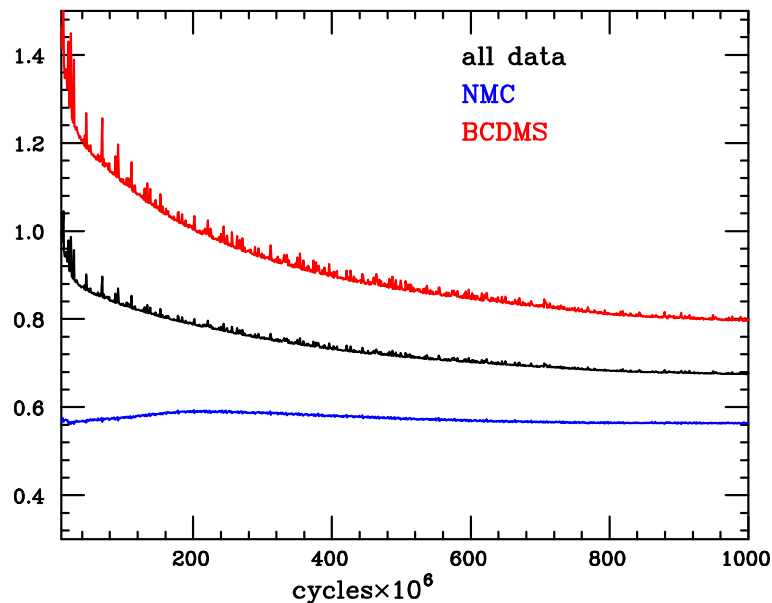
# NEURAL INFORMATION HANDLING II

STUDY DEPENDENCE OF ERROR FCTN  $E^{(0)} = \frac{1}{N_{dat}} \sum_{i=1}^{N_{dat}} \frac{(F_i^{(exp)} - F_i^{(net)(0)})^2}{\sigma_{i,s}^{(exp)2}}$  ON TRAINING LENGTH FOR NET TRAINED ON CENTRAL VALUES

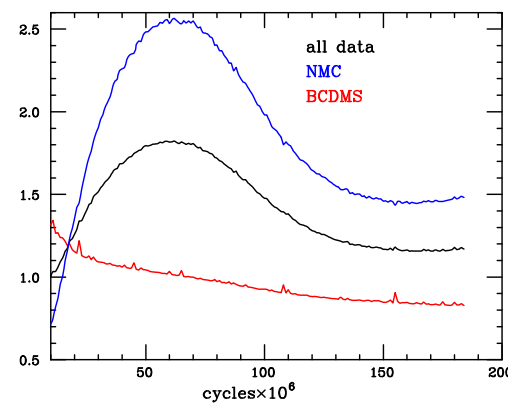
## INHOMOGENEOUS ERRORS

**NS:** AFTER  $\sim 10^7$  TRAINING CYCLES,  $E^{(0)} \approx 1$  BUT WIDE SPREAD BETWEEN DATASETS  
 $\Rightarrow$  NMC OVERLEARNT & BCDMS UNDERLEARNT

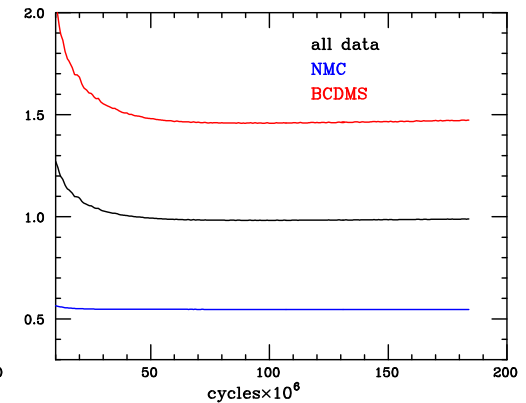
training on all data



training on BCDMS



training on NMC



- EACH DATASET PREDICTS THE OTHER  
 $\Rightarrow$  FULL COMPATIBILITY
- BCDMS HARDER TO LEARN THAN NMC  
(SMALLER ERRORS)

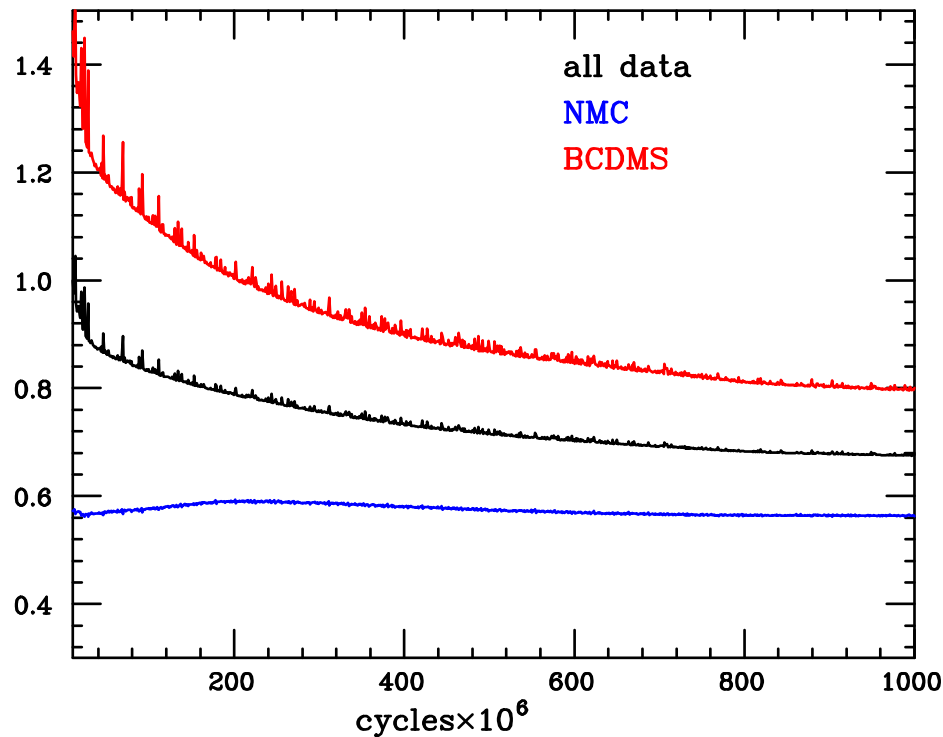


# INHOMOGENEOUS ERRORS cont'd

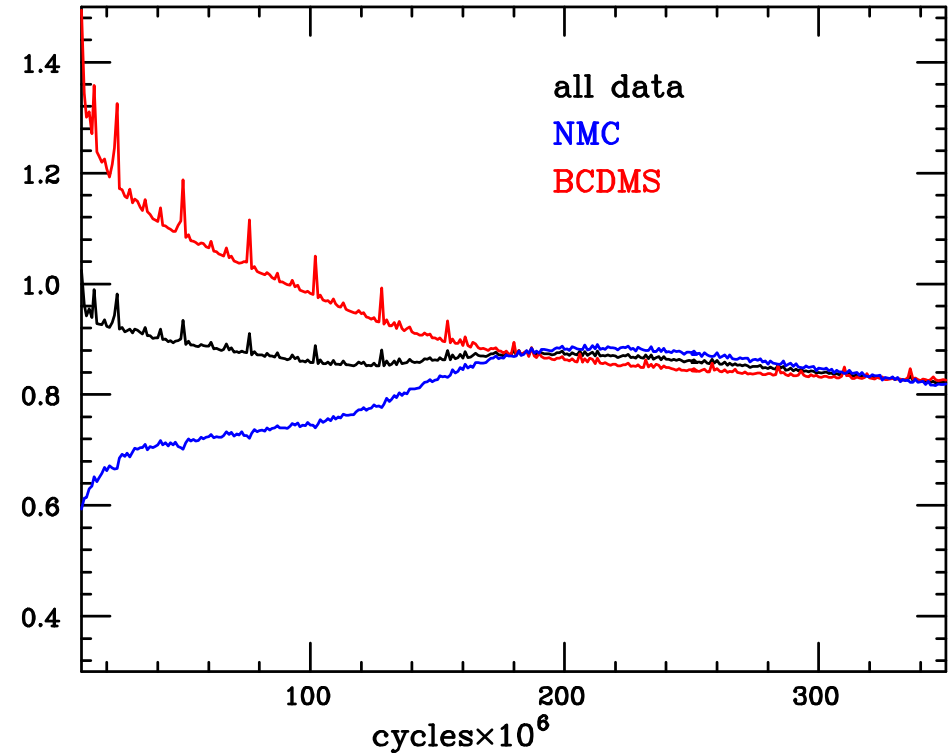
NETS ARE GETTING TRAPPED IN LOCAL MIN. OF THE DATA WHICH ARE LEARNT FASTER  
global min. can only be reached at overlearning point

## SOLUTION: WEIGHTED TRAINING

uniform training



90% BCDMS 10 % NMC



- convergence of two experiments reached fast by weighted training
- at convergence,  $E^{(0)} \approx 1$
- after convergence,  $E^{(0)}$  for two experiment slowly improve at same rate, oscillating about each other  $\Rightarrow$  global minimum found

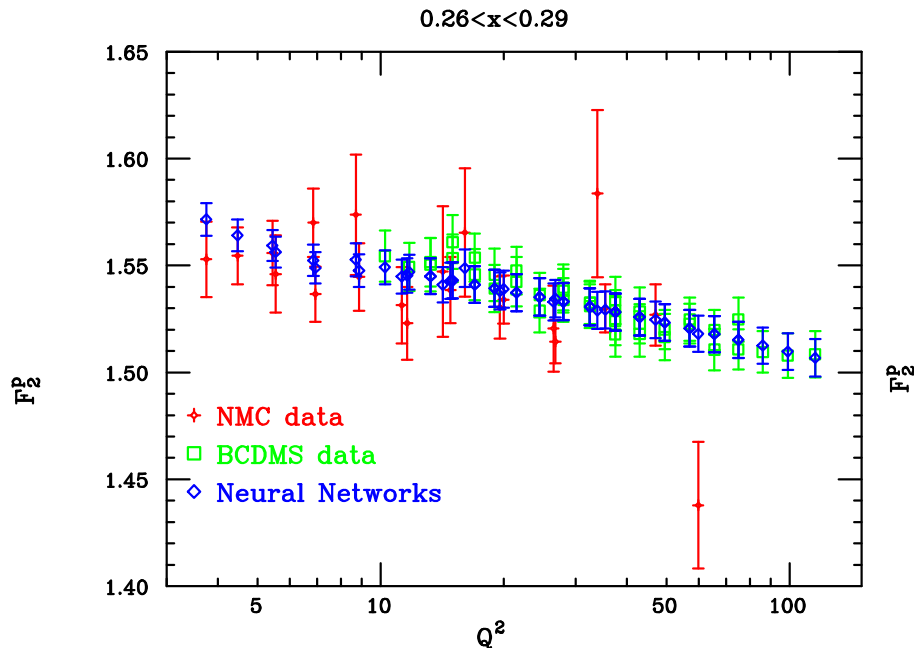
# NEURAL INFORMATION HANDLING III

## INCOMPATIBLE DATA

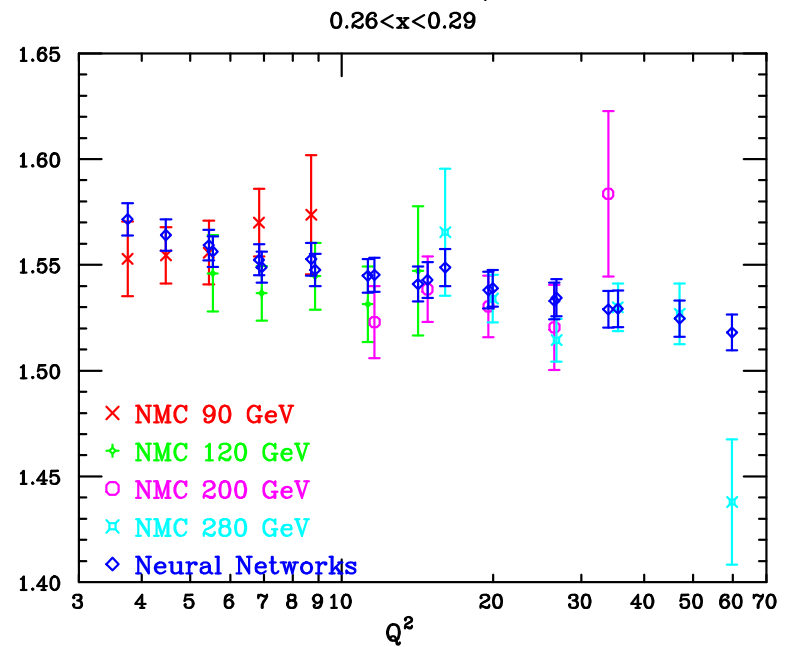
- FOR PROTON FITS, CONVERGENCE ACHIEVED, BUT  $E^{(0)} \gtrsim 1.4$  EVEN W. VERY LONG TRAINING
- for NMC data  $E^{(0)} \gtrsim 1.6$  (training with all data)
- for NMC data  $E^{(0)} \gtrsim 2.2$  (training with NMC only)
- ALL OTHER STATISTICAL INDICATORS OK

## SOME NMC DATA ARE INCOMPATIBLE WITH OTHER DATA

Blow-up of proton data/nets



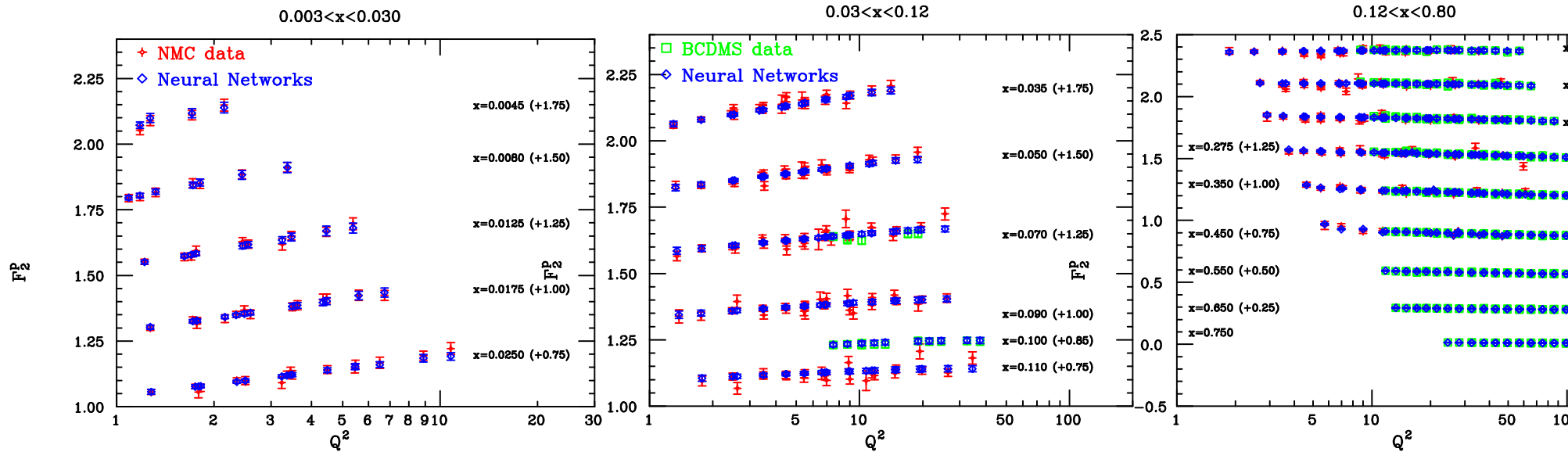
NMC proton data/nets



NEURAL NET DISCARDS INCONSISTENT DATA & PROVIDES GOOD FIT TO THE REST

# RESULTS

## NEURAL FIT TO PROTON $F_2$ DATA



- FULL NEURAL FIT TO  $F_2$  FOR PROTON, DEUTERON & NONSINGLET AVAILABLE
- ERRORS AND CORRELATIONS FAITHFULLY REPRODUCED, BUT STAT. UNCERTAINTIES OPTIMALLY COMBINED
- ⇒ FIT CAN BE USED IN LIEU OF DATA, BUT BETTER THAN THEM
- SOURCE CODE, DRIVER PROGRAM & GRAPHIC WEB INTERFACE FOR  $F_2$  PLOTS & NUMERICAL COMPUTATION AVAILABLE @

<http://sophia.ecm.ub.es/f2neural>

# AN APPLICATION: $\alpha_s$ FROM SCALING VIOLATIONS

NONSINGLET  $F_2 \Rightarrow$  NONSINGLET QUARK DISTRIBUTION

IN THE “DIS” FACTORIZATION SCHEME

$$F_2^{NS}(x, Q^2) \equiv F_2^p(x, Q^2) - F_2^d(x, Q^2) = \sum_{i=1}^{n_f} e_i^2 [q_i(x, Q^2) + \bar{q}_i(x, Q^2)]_{p-n}$$

SO  $F_2^{NS}$  EVOLVES MULTIPLICATIVELY

$$\mu^2 \frac{d}{d\mu^2} F_2^{NS}(x, \mu^2) = \frac{\alpha_s(\mu^2)}{2\pi} \int_x^1 \frac{dy}{y} P\left(\frac{x}{y}, \alpha_s(\mu^2)\right) F_2^{NS}(y, \mu^2)$$

$P$ : DIS-scheme Altarelli-Parisi NS splitting function

GIVEN DATA FOR  $F_2^{NS}$  CAN DETERMINE  $\alpha_s$  FROM ITS SCALING VIOLATIONS

# AN APPLICATION: $\alpha_s$ FROM SCALING VIOLATIONS

NONSINGLET  $F_2 \Rightarrow$  NONSINGLET QUARK DISTRIBUTION

IN THE “DIS” FACTORIZATION SCHEME

$$F_2^{NS}(x, Q^2) \equiv F_2^p(x, Q^2) - F_2^d(x, Q^2) = \sum_{i=1}^{n_f} e_i^2 [q_i(x, Q^2) + \bar{q}_i(x, Q^2)]_{p-n}$$

SO  $F_2^{NS}$  EVOLVES MULTIPLICATIVELY

$$\mu^2 \frac{d}{d\mu^2} F_2^{NS}(x, \mu^2) = \frac{\alpha_s(\mu^2)}{2\pi} \int_x^1 \frac{dy}{y} P\left(\frac{x}{y}, \alpha_s(\mu^2)\right) F_2^{NS}(y, \mu^2)$$

$P$ : DIS-scheme Altarelli-Parisi NS splitting function

GIVEN DATA FOR  $F_2^{NS}$  CAN DETERMINE  $\alpha_s$  FROM ITS SCALING VIOLATIONS

**PROBLEM:** HARD TO DEAL WITH CONVOLUTIONS...

**SOLUTION:** INTRODUCE A PARAMETRIZATION OF  $F_2$ , TAKE MELLIN MOMS.

$$\mu^2 \frac{d}{d\mu^2} F_{2,N}^{NS}(\mu^2) = \frac{\alpha_s(\mu^2)}{2\pi} \gamma_N(\alpha_s(\mu^2)) F_{2,N}^{NS}(\mu^2);$$
$$\gamma_n(\alpha_s(\mu^2)) \equiv \int_0^1 x^{N-1} P(x, \alpha_s(\mu^2)), \quad F_{2,N}^{NS}(\mu^2) \equiv \int_0^1 x^{N-1} F_2^{NS}(x, \mu^2)$$

NO PARAMETRIZATION BIAS  $\Rightarrow$  NEURAL NETS

NO EXTRAPOLATION BIAS  $\Rightarrow$  TRUNCATED MOMENTS  $F_{2,N}^{NS}(x_0, \mu^2) \equiv \int_{x_0}^1 dx x^{N-1} F_2^{NS}(x, \mu^2)$

# DETERMINATION OF $\alpha_s$

- MOMENTS CAN BE COMPUTED AT ANY SCALE IN TERMS OF MOMS. AT REF. SCALE  $Q_0^2$  through evolution matrix  $M(x_0; Q_0^2, Q_i^2; \alpha_s)$  determined by an. dim. and  $\alpha_s$ :

$$q_n^{th}(x_0, Q_i^2) \equiv \sum_{p=n_{min}}^M M_{np}(x_0; Q_0^2, Q_i^2; \alpha_s) q_p(x_0, Q_0^2)$$

- CAN DETERMINE  $\alpha_s$  BY MINIMIZING  $\chi^2$  with covariance matrix  $V^{-1}$  from neur. nets
- $$\chi^2 = \sum_{n,i} \sum_{m,j} \left[ q_n^{exp}(x_0, Q_i^2) - q_n^{th}(x_0, Q_i^2) \right] V_{ni;mj}^{-1} \left[ q_m^{exp}(x_0, Q_j^2) - q_m^{th}(x_0, Q_j^2) \right]$$

## MOMENTS AND CORRELATIONS

IN PRINCIPLE FIT  $\alpha_s$  & ALL MOMENTS AT REF. SCALE

IN PRACTICE NEIGHBOURING MOMENTS HIGHLY CORRELATED;

OFF-DIAGONAL ANOMALOUS DIMS. SMALL  $\Rightarrow$  FIT ONLY A SUBSET OF MOMENTS

(NMC + BCDMS)

### SINGLE MOMENT

$n$	$\alpha_s$
2	0.085 $\pm$ 0.070
3	0.106 $\pm$ 0.030
4	0.115 $\pm$ 0.019
5	0.123 $\pm$ 0.015
6	0.127 $\pm$ 0.014
7	0.129 $\pm$ 0.014
8	0.129 $\pm$ 0.016
9	0.129 $\pm$ 0.018

purple: minimal error

### MORE MOMENTS

FITTED MOMENTS	$\alpha_s$
2+3+4	0.126 $\pm$ 0.010
2+4+6	0.140 $\pm$ 0.008
3+5+7	0.138 $\pm$ 0.009
2+4+6+8	0.142 $\pm$ 0.009
3+5+7+9	0.124 $\pm$ 0.007
2+4+5+7	0.141 $\pm$ 0.009
3+4+5+6+7	0.1256 $\pm$ 0.0049
3+4+5+6+8	0.1247 $\pm$ 0.0050
2+4+5+6+8	0.1242 $\pm$ 0.0042
2+4+5+7+8	0.1254 $\pm$ 0.0044

red: optimal fit

### OPTIMAL FIT

AS THE NUMBER OF FITTED MOMENTS IS INCREASED

ERROR DECREASES,

STABILITY OF CENTRAL VALUES IMPROVES

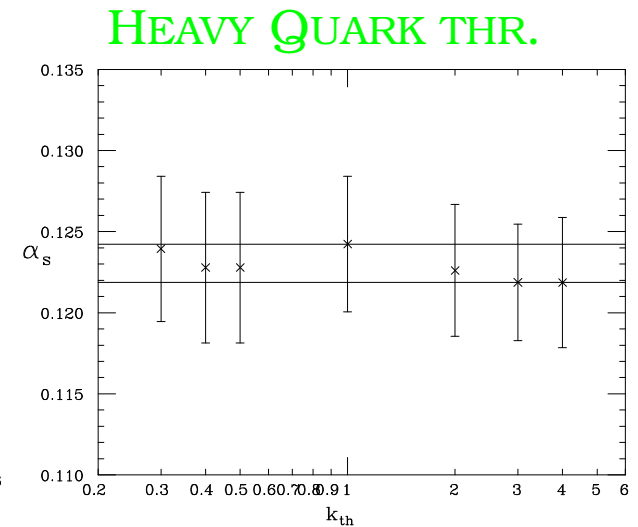
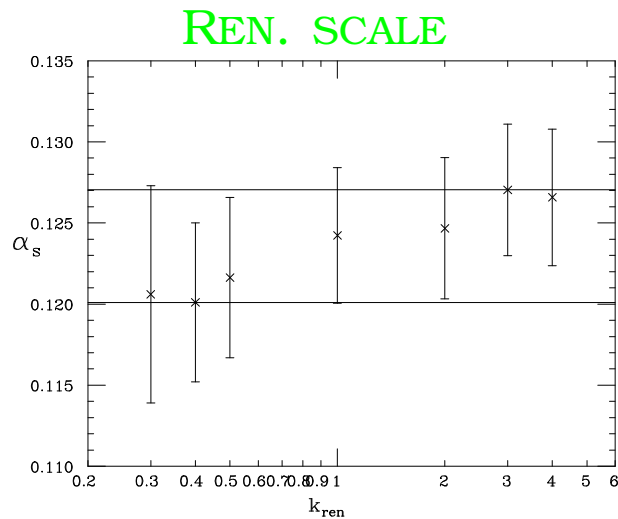
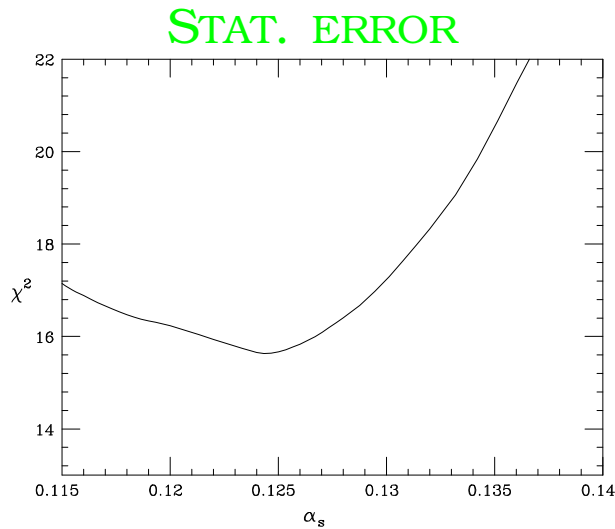
BUT IF CORRELATIONS LARGE, FIT UNSTABLE

- $20 \leq Q^2 \leq 70 \text{ GeV}^2$ , THREE SCALES correlns. larger if  $Q^2$  values closer

- $x_0 = 0.03$  correlns. larger if  $x_0$  larger

- 2+4+5+6+8 higher moments less reliable and more correlated

# UNCERTAINTIES



- **ASYMMETRIC  $\chi^2$ :  $\sigma(\text{STAT.}) = \begin{matrix} +0.004 \\ -0.007 \end{matrix}$**
- **HIGHER ORDER CORRNS FROM  $\mu_{ren}^2 = k_{ren} Q^2$ ,  $0.3 \leq k_{ren} \leq 4$ :  $\sigma(\text{REN.}) = \begin{matrix} +0.003 \\ -0.004 \end{matrix}$**
- **POSITION OF HQ THRESH.  $Q_{th}^2 = k_{th} M_q^2$ ,  $0.3 \leq k_{th} \leq 4$   $\sigma(\text{THRESH.}) = \begin{matrix} +0.000 \\ -0.002 \end{matrix}$**
- **POWER CORRNS. VARY  $Q_{min}^2$  FROM 20 TO 30  $\text{GEV}^2$   $\sigma(\text{HT}) < 0.001$**

$$\alpha_s(M_Z) = 0.124 \begin{matrix} +0.004 \\ -0.007 \end{matrix} (\text{EXP.}) \begin{matrix} +0.003 \\ -0.004 \end{matrix} (\text{TH.}) = 0.124 \begin{matrix} +0.005 \\ -0.008 \end{matrix} (\text{TOTAL})$$

**ERROR:** DOMINATED BY EXP. ERROR, TH. BIAS & UNCERTAINTY MINIMIZED

**CENTRAL VALUE:** CONSISTENT WITH WORLD AVERAGE BUT HIGH

**EVIDENCE FOR SUDAKOV?** High moments dominate the fit,  $Q_{\text{eff}}^2 = Q^2/N$ ;

$\alpha_s$  from a single moment increases with  $N$

# OUTLOOK

- RELIABLE DETERMINATION OF STRUCTURE FUNCTIONS  
⇒ GLUON SPIN FRACTION (AND MORE...)



# OUTLOOK

- RELIABLE DETERMINATION OF STRUCTURE FUNCTIONS
  - ⇒ GLUON SPIN FRACTION (AND MORE...)
- SUCCESSFUL IMPLEMENTATION OF NEURAL FITTING
  - ⇒ NEURAL PARTON DISTRIBUTIONS!

# OUTLOOK

- RELIABLE DETERMINATION OF STRUCTURE FUNCTIONS

⇒ GLUON SPIN FRACTION (AND MORE...)

- SUCCESSFUL IMPLEMENTATION OF NEURAL FITTING

⇒ NEURAL PARTON DISTRIBUTIONS!

...A WHOLE NEW SET OF TOOLS IN THE BOX!