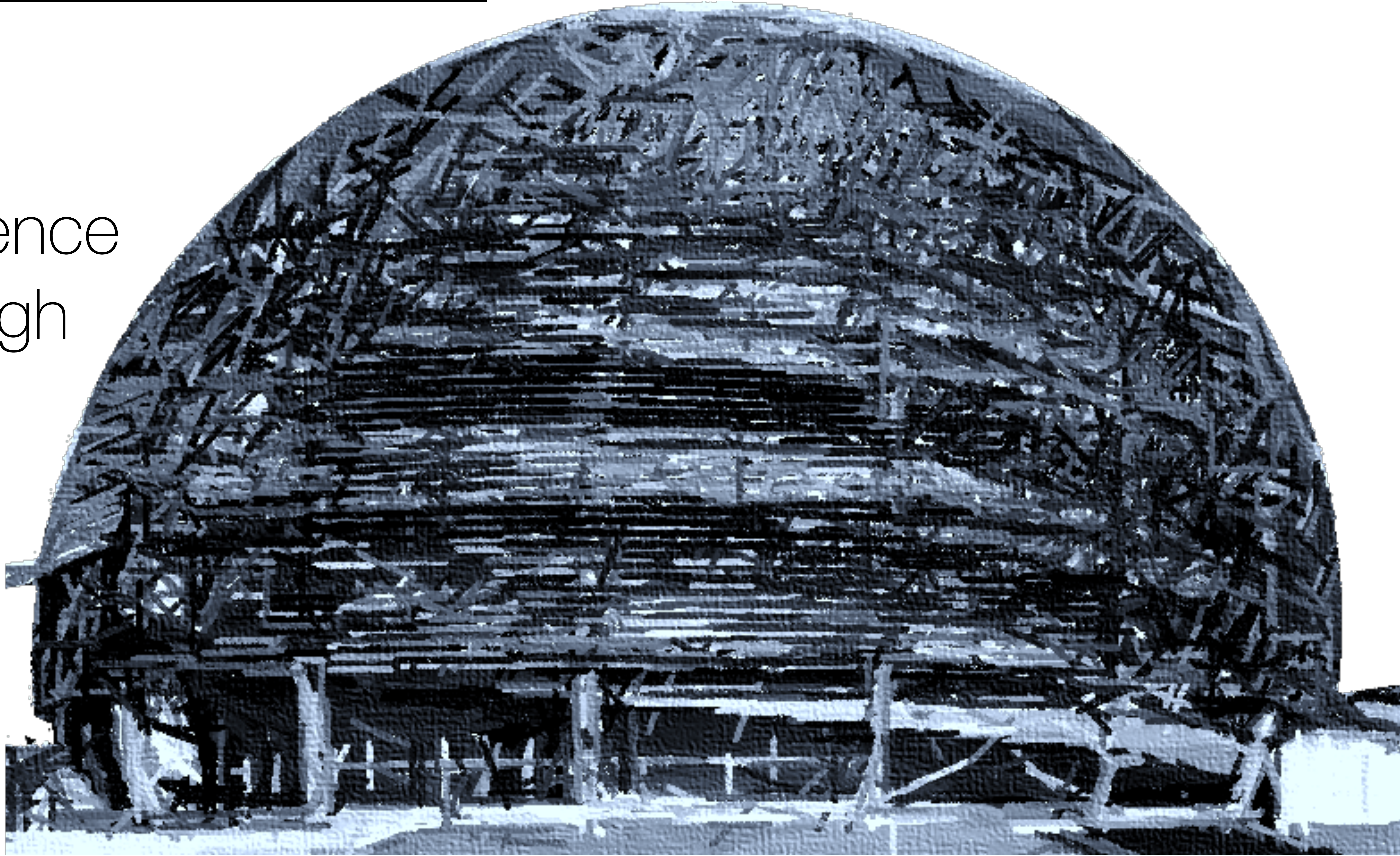
DATA DRIVEN EXTRACTION OF PROTON'S VALENCE QUARK RATIO

DANIELE BINOSI
ECT* - FONDAZIONE BRUNO KESSLER

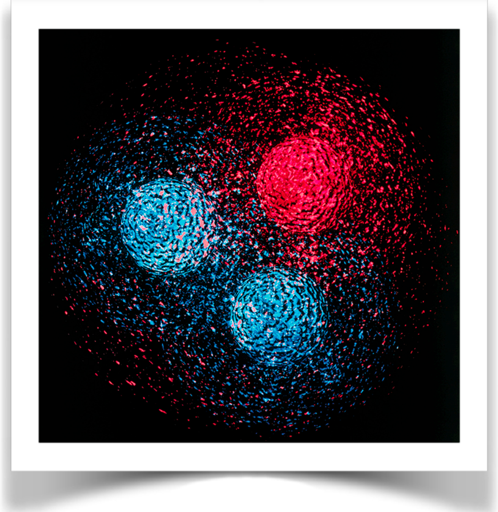
Perceiving the Emergence of Hadron Mass through AMBER@CERN ❻

SEPTEMBER 27 - SEPTEMBER 29 2021



ECT*
EUROPEAN CENTRE
FOR THEORETICAL STUDIES
IN NUCLEAR PHYSICS AND RELATED AREAS
FONDAZIONE BRUNO KESSLER

Proton STRUCTURE



Nature's lowest-mass **bound-state** in the scattering of two *u*-quarks and one *d*-quark

---

Quantum ChromoDynamics

describes the proton's wave function Ψ

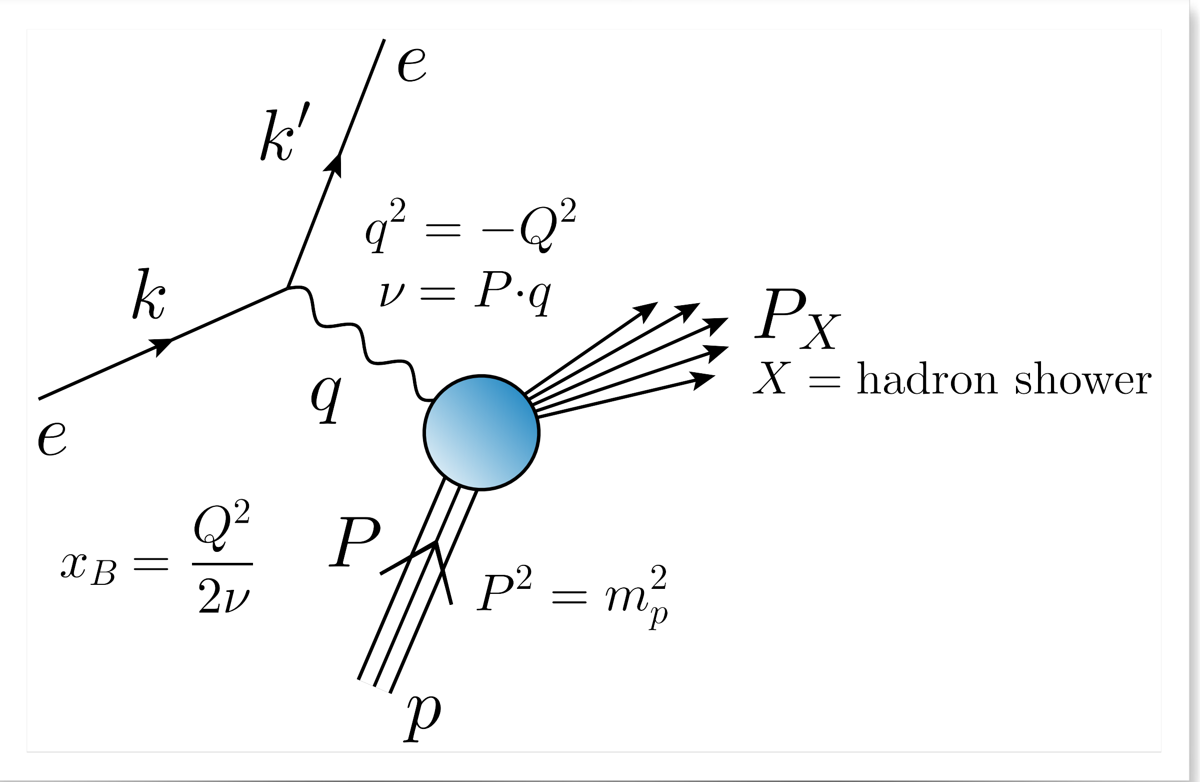
Knowledge of Ψ yields proton's gluons and quarks *number density distributions*

DISTRIBUTION FUNCTIONS

Powerful discriminator between competing descriptions of proton structure:

ratio of *u* and *d* quark densities in the far-valence domain

① THEORY



Consider DIS reactions from neutron and proton targets ratio of Poincaré-invariant structure functions

$$\frac{F_2^n(x)}{F_2^p(x)} = \frac{u_v(x) + 4d_v(x) + 6d_s(x) + \Sigma(x)}{4u_v(x) + d_v(x) + 6u_s(x) + \Sigma(x)}$$

valence   sea

Sea distributions negligible on $x \gtrsim 0.2$

$$\frac{F_2^n(x)}{F_2^p(x)} \overset{x\gtrsim 0.2}{\approx} \frac{1 + 4d_v(x)/u_v(x)}{4 + d_v(x)/u_v(x)}$$

d/u ratio fixed point under QCD evolution at $x = 1$

Nachtmann bounds:

$$1/4 \le F_2^n(x)/F_2^p(x) \le 4$$

Jefferson Lab

EXPERIMENT ②

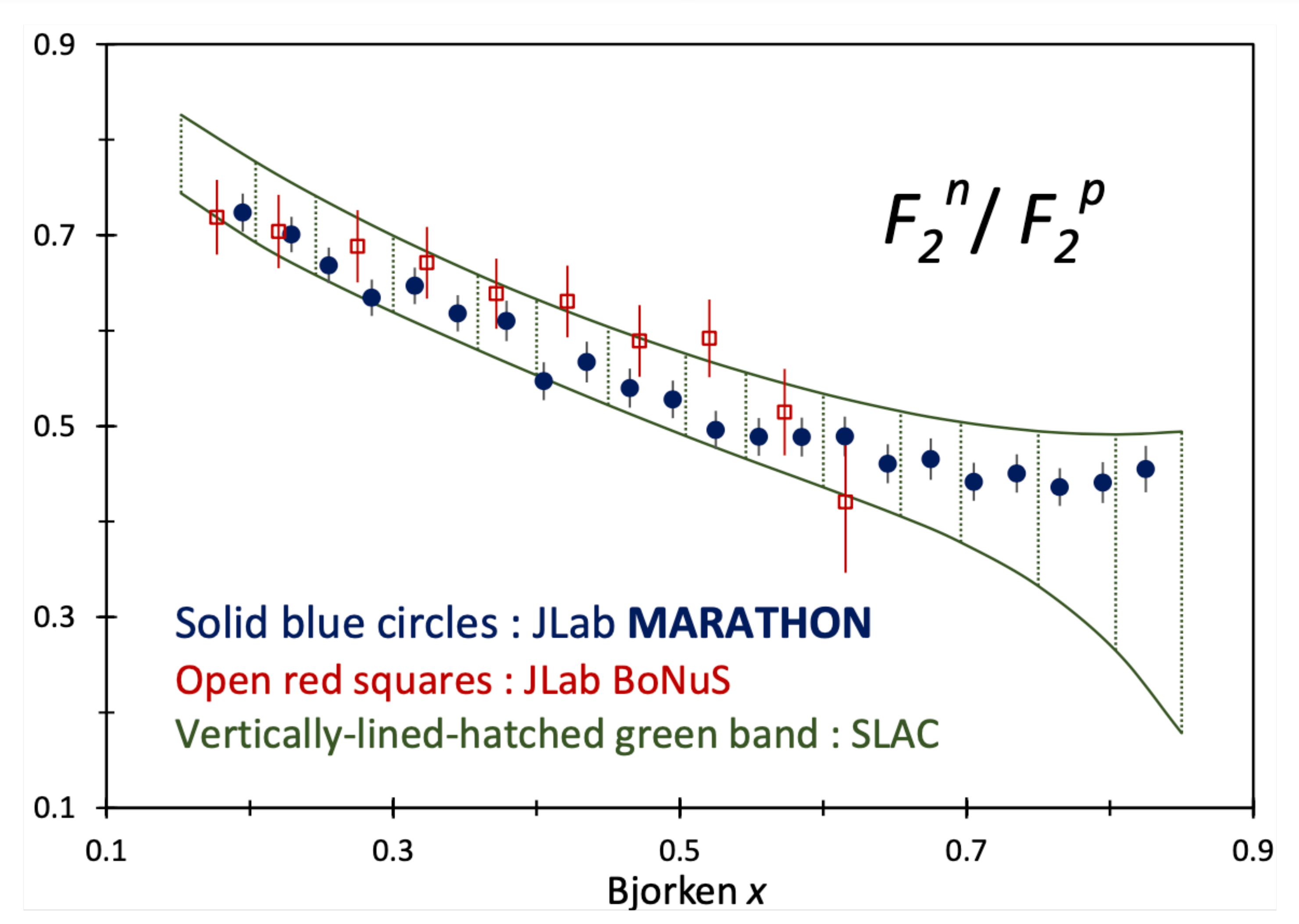
Main problem: construction of a **free neutron target**

Use the *deuteron* (BONuS); large systematic uncertainties beyond $x \gtrsim 0.7$ due to characterisation of proton-neutron interactions

---

Perform DIS measurement on ^{3}He and ^{3}H and take ratio of scattering rates (MARATHON)

Nuclear interaction effects largely cancel; handling of radioactive ^{3}He target challenging



WHAT DO *DATA* ENCODE ABOUT PROTON'S STRUCTURE?

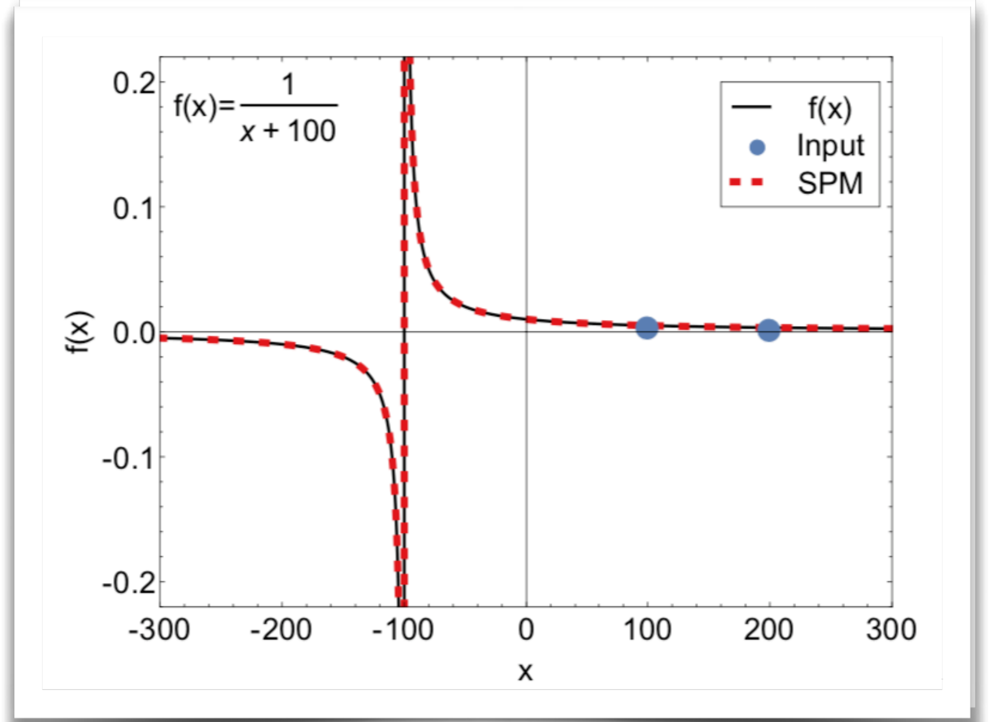
SPM AND SMOOTHING

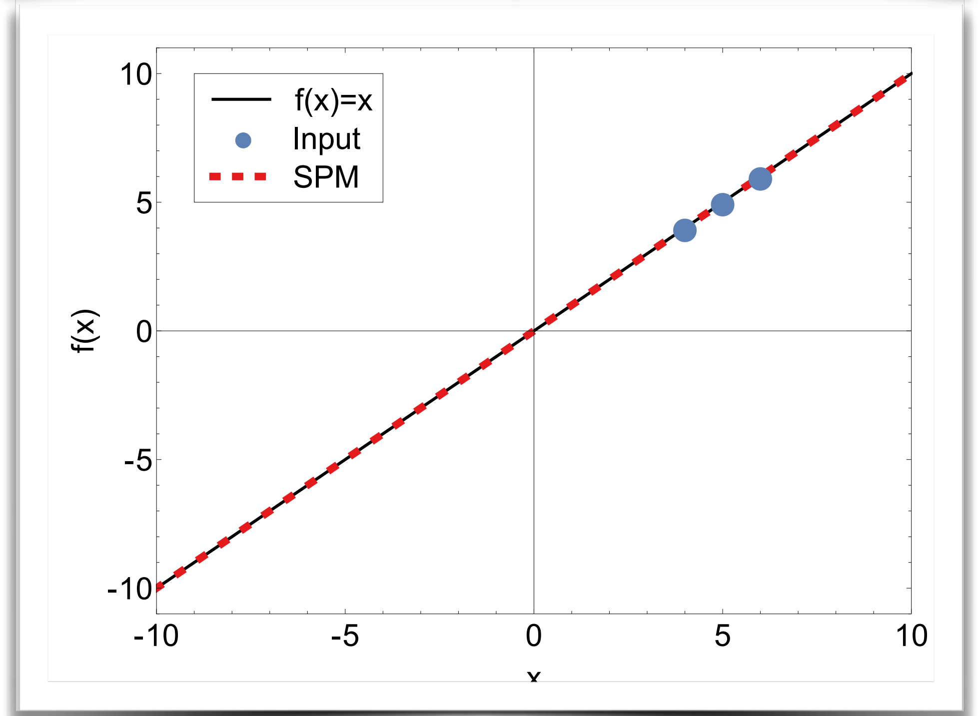
$D = \{(x_i, y_i = f(x_i)),\ i = 1, \ldots, N\}$

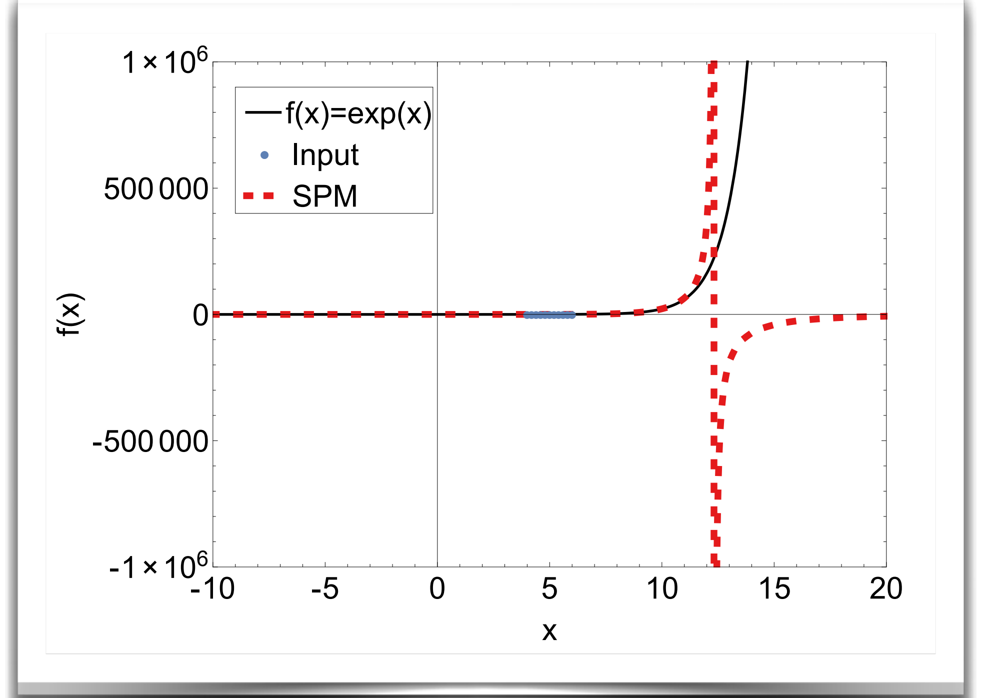
$$C_N(x) = \frac{y_1}{1+}\frac{a_1(x-x_1)}{1+}\frac{a_2(x-x_2)}{1+}\cdots\frac{a_{N-1}(x-x_{N-1})}{1}$$

Schlessinger, PR 167 (1968)

elementary (functions) examples







LARGE DATASETS

randomly choose $4 < M \lesssim N/2$ points
reduce (binomial) number of interpolators
introducing **physical constraints**
(absence of poles)

Chen et al., PRD 99 (2019)

IN THE PRESENCE OF ERRORS?

direct interpolation does not work
requires **smoothing** with **roughness penalty**:
seek $g \in \mathbb{S}$ minimising

$$P(g, \lambda) = \lambda \sum_{i=1}^{\ell} [y_i - g(x_i)]^2 + (1-\lambda) \int_a^b dx\, [g''(x)]^2$$

smoothing par. data fidelity roughness penalty

THEOREM: g is the *natural spline* interpolant of nodes $\{x_i\}$

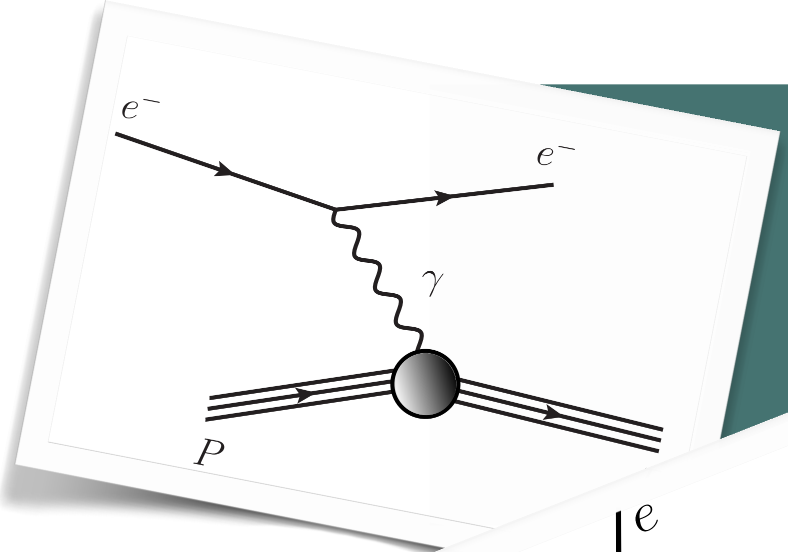
Reinsch, NM 10 (1967)

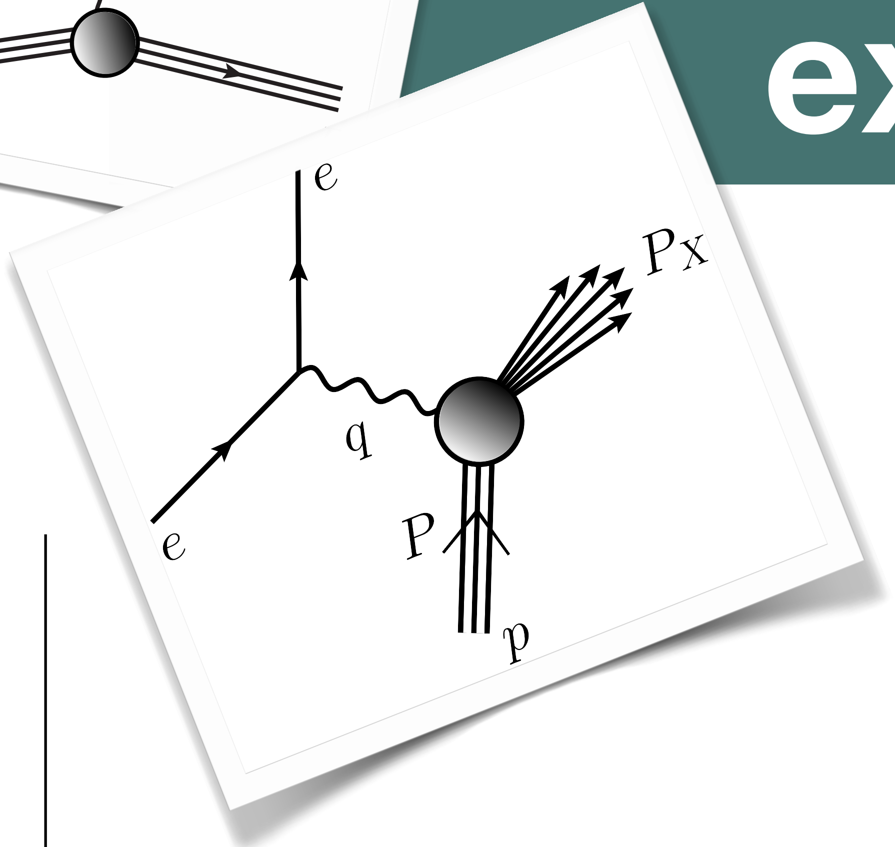
optimal smoothing parameter determined via generalised cross validation

Craven and Wahba, NM 31 (1978)

use **bootstrap** procedure to generate replicas accounting for statistical errors in data when extrapolating

$$(x_i, y_i, \sigma_i) \to (x_i, \mathcal{N}(y_i, \sigma_i))$$





SPM parameter extraction

General algorithm to extrapolate a certain parameter from given noisy experimental datasets.

generate (10^3) replicas for the given experimental central values and error
smooth each replica with associated optimal λ
set $\{M_j = 5 + j \mid j = 1, \ldots, n_M\}$ for a suitable n_M
fix M_j and get a number of monotonic SPM interpolators for each replica
determine the replicas' parameter value (averaging over the obtained curves)
construct the (normal) distribution of the replicas' (10^3) extracted parameter;
extract the mean p^{M_j} and standard deviation $\sigma_p^{M_j}$
final result

$$p \pm \sigma; \qquad p = \sum_{j=1}^{n_M} \frac{p^{M_j}}{n_M}; \qquad \sigma_p = \left[\sum_{j=1}^{n_M} \frac{(\sigma_p^{M_j})^2}{n_M^2} + \sigma_{\delta M}^2\right]^{\frac{1}{2}}.$$

standard deviation of p^{M_j} distribution

DOES IT WORK?

SPM AND SMOOTHING VALIDATION

does it really work? ✓
is it robust? ✓

If you want to disprove a result, show first you can replicate it

build **replicas of the observable** of **known parameter values** p^*

DATA GENERATORS

Generate data from a variety of models: functional forms; parametrisations of experimental data; “real-world” calculations

CHECKS

$\forall$ *M*/generators/kinematics:

Gaussianity of *p* distribution ✓
robustness of *p* extraction ✓

bias

Root Mean Square Error

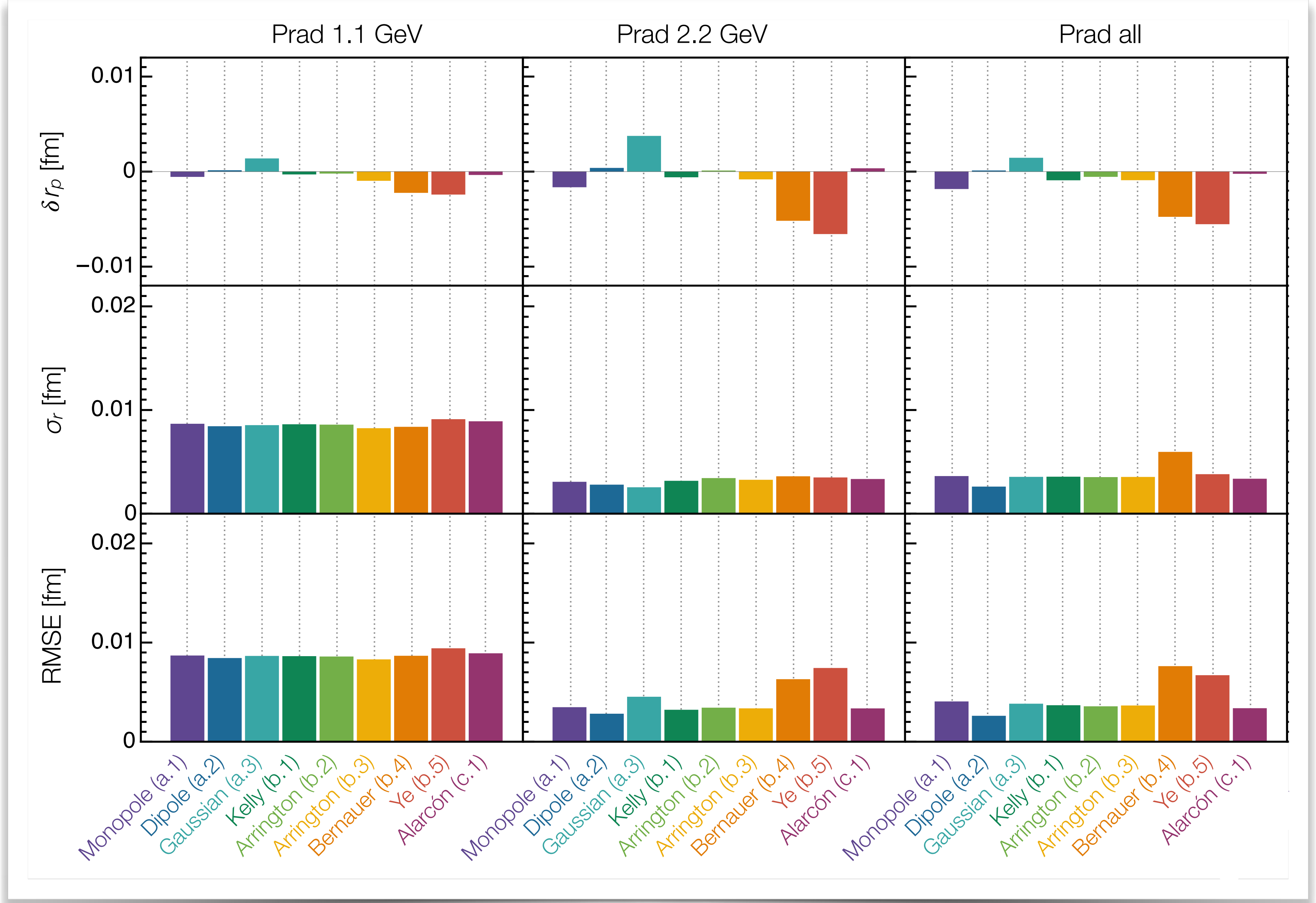
$$\delta p = p - p^* \qquad \sigma_p \qquad \mathrm{RMSE} = \sqrt{\delta p^2 + \sigma_p^2}$$

standard deviation

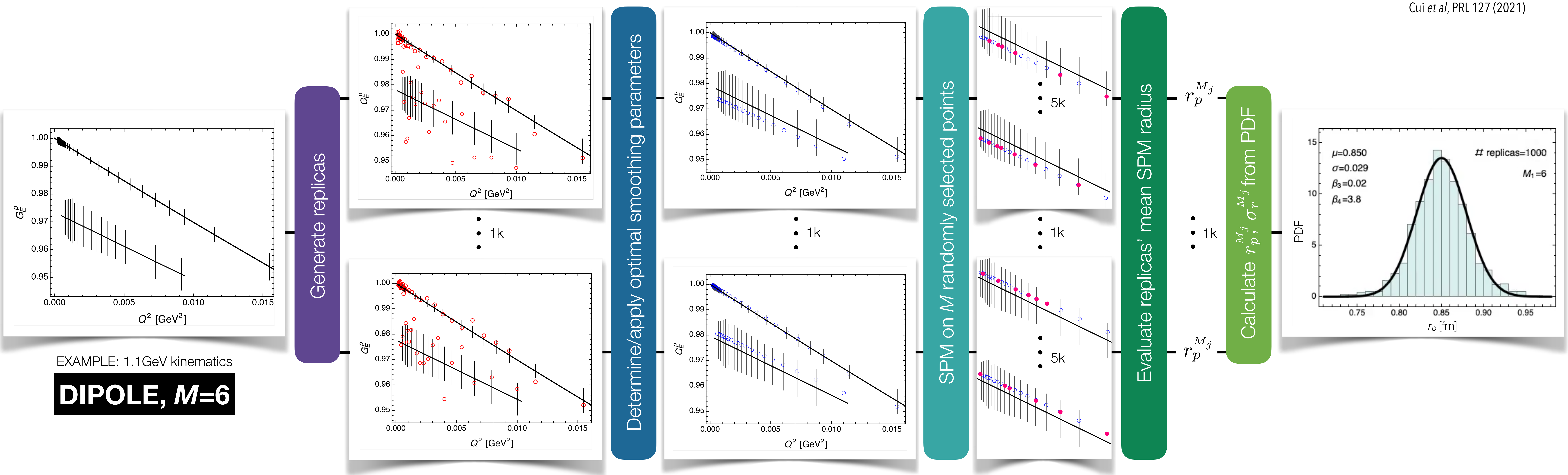
p extraction robust if

$|\delta p| < \sigma_p$ ✓

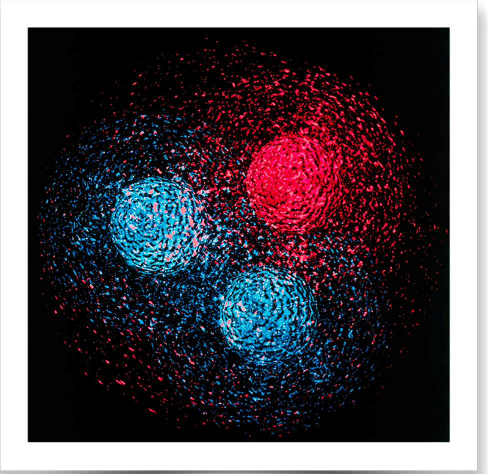
RMSE independent from generator ✓



Cui *et al*, PRL 127 (2021)



Proton
CHARGE RADIUS



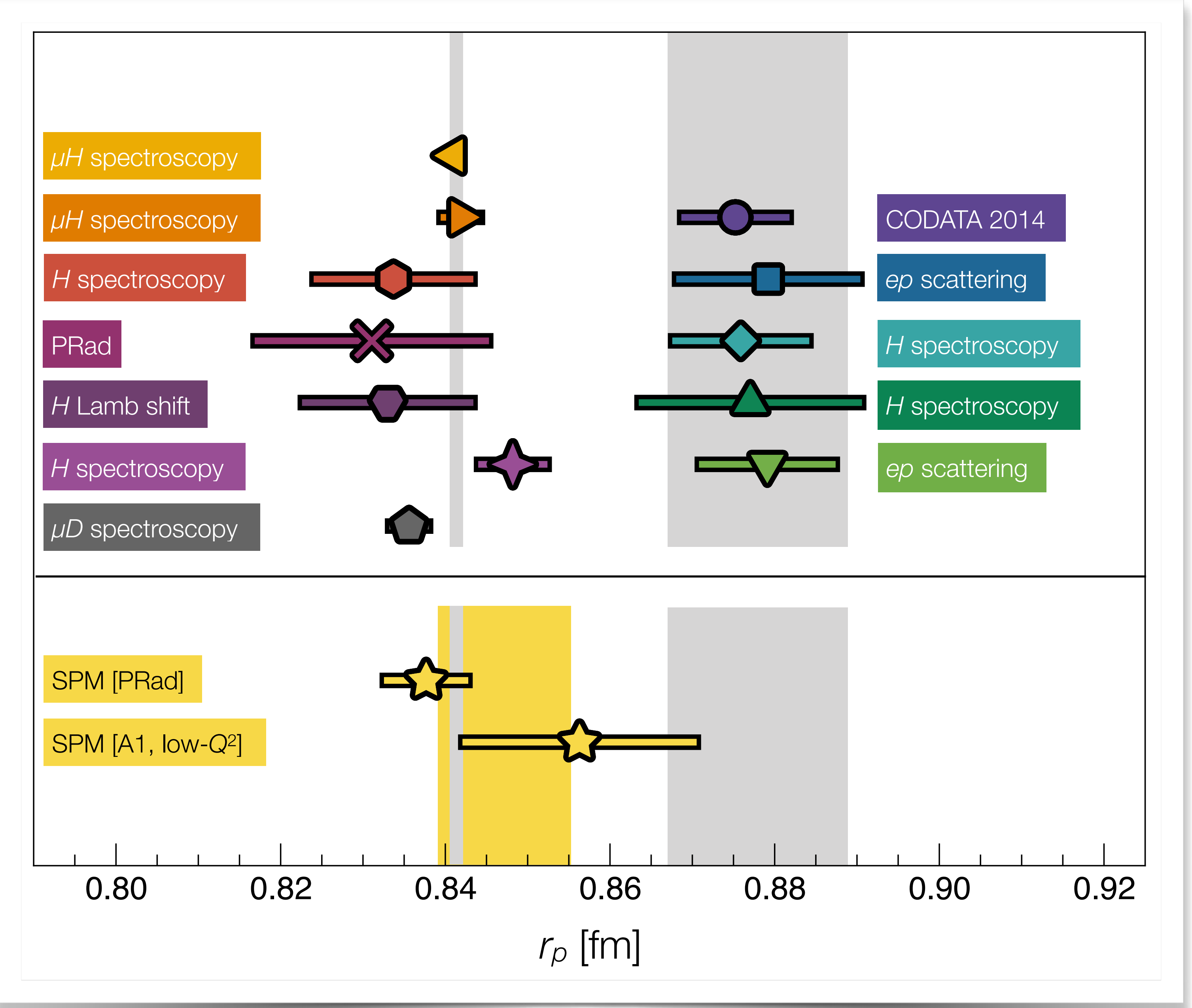
① PRad DATA

$$r_p^{\mathrm{PRad}} = 0.838 \pm 0.005_{\mathrm{stat}}\ [\mathrm{fm}]$$

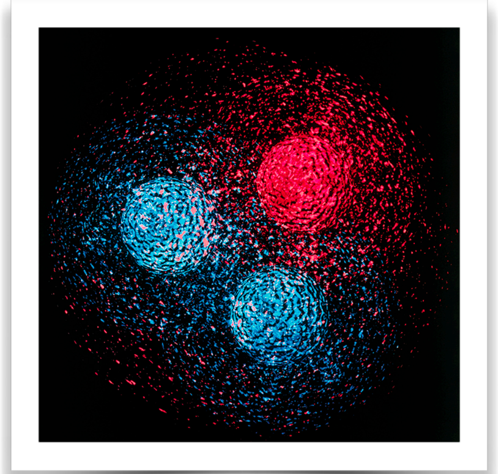
❷ A1 DATA

$$r_p^{\mathrm{A1-low}Q^2} = 0.856 \pm 0.014_{\mathrm{stat}}\ [\mathrm{fm}]$$

Cui *et al*, PRL 127 (2021)



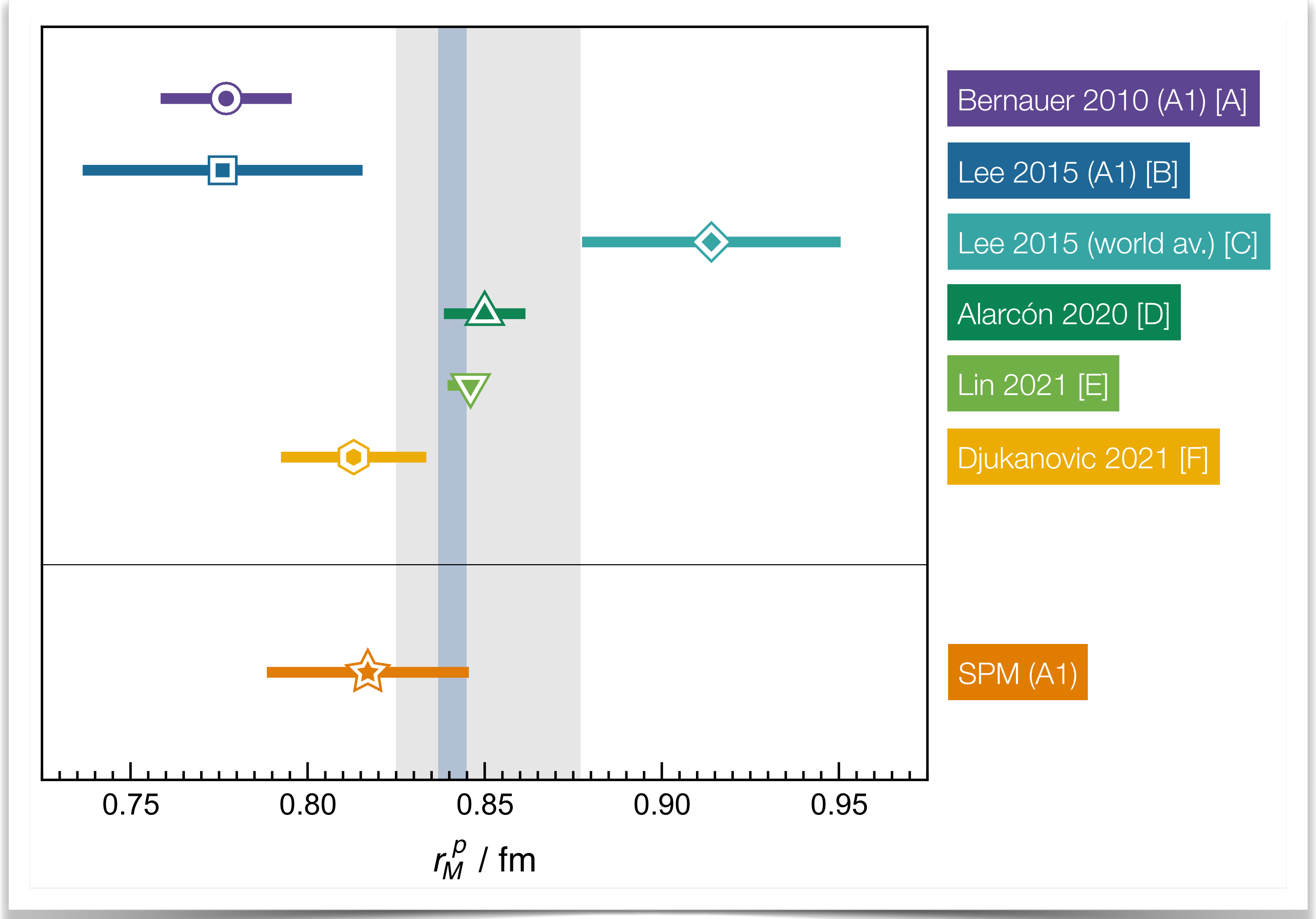
Proton
PAULI RADIUS



① A1 DATA

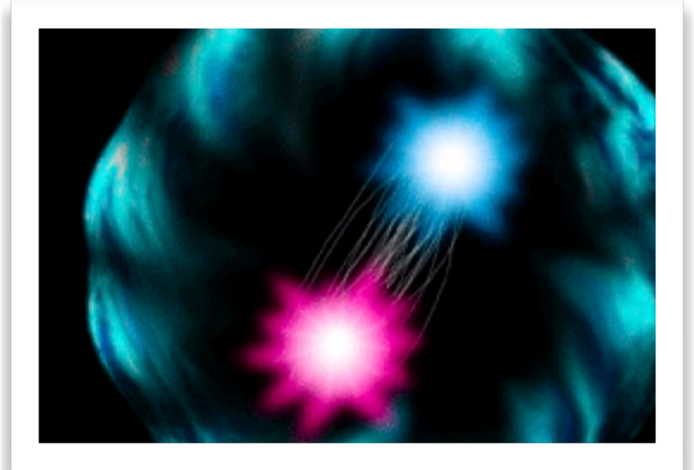
$$r_M^{\mathrm{A1}} = 0.817 \pm 0.027_{\mathrm{stat}}\ [\mathrm{fm}]$$

Cui *et al*, 2109.08768



$$\frac{F_2'(0)}{\kappa_p} \simeq F_1'(0) - \frac{\mu_p}{4m_p^2}$$

Pion
CHARGE RADIUS

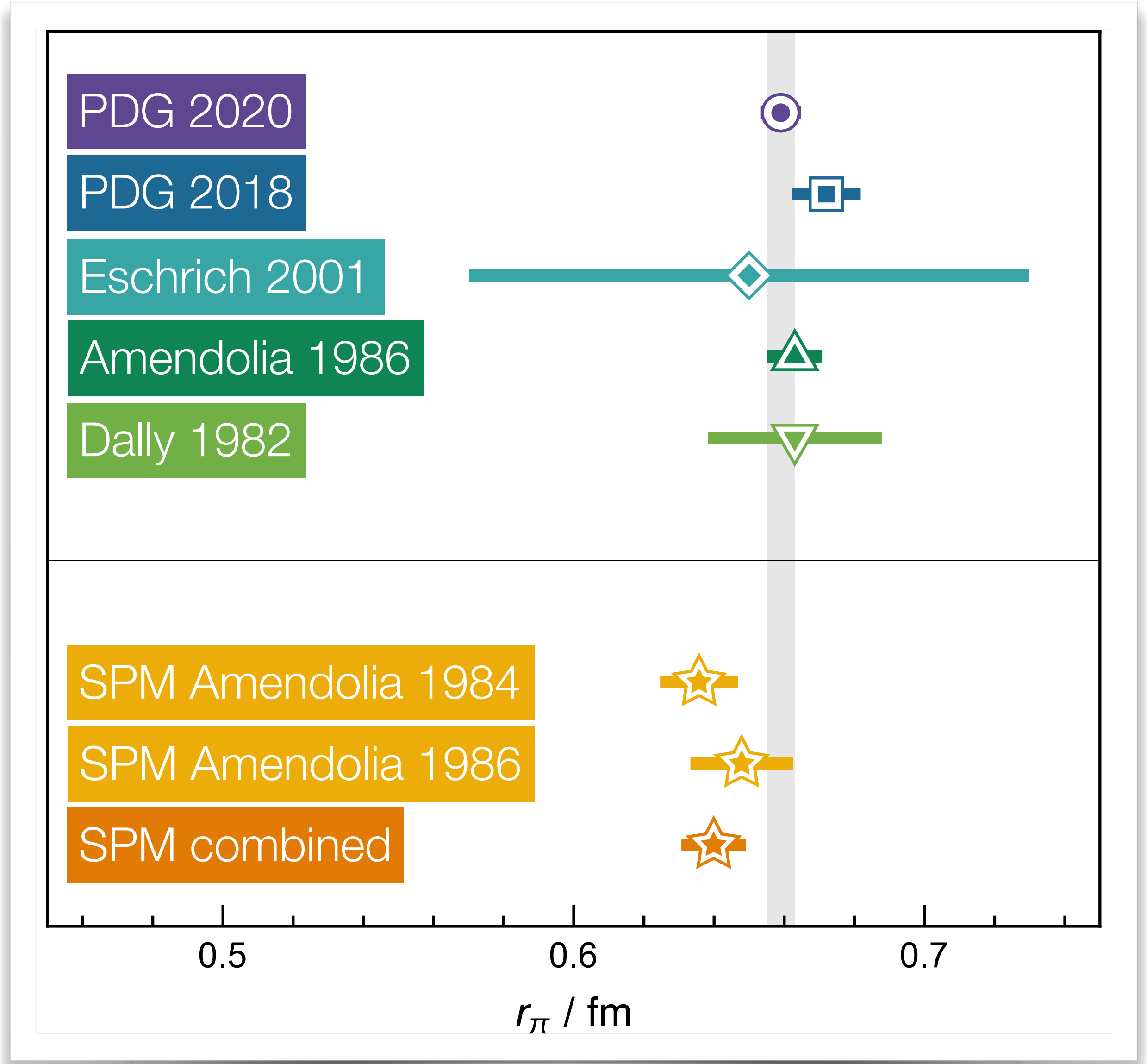


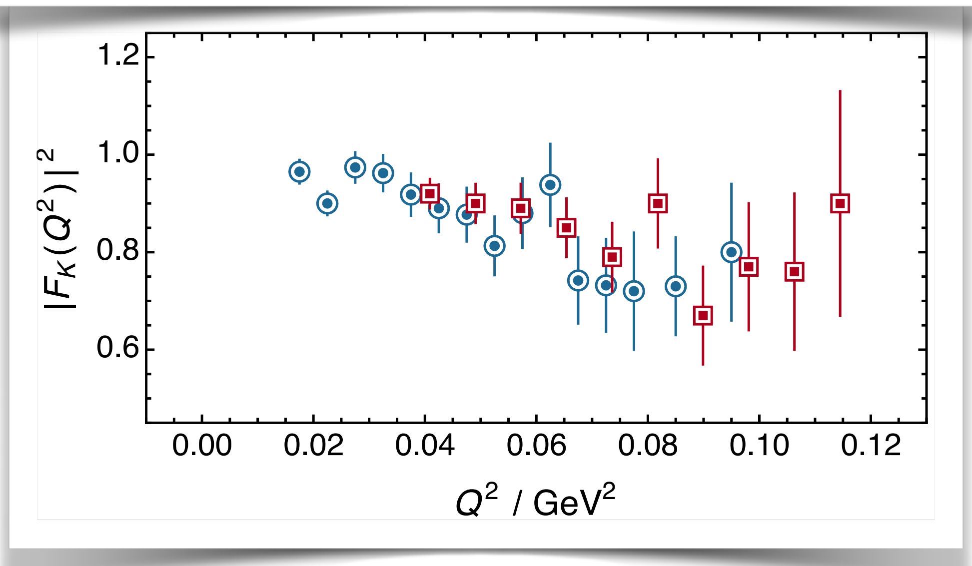
① NA7 DATA

$$r_\pi^{\mathrm{NA7-86}} = 0.648 \pm 0.013_{\mathrm{stat}}\ [\mathrm{fm}]$$

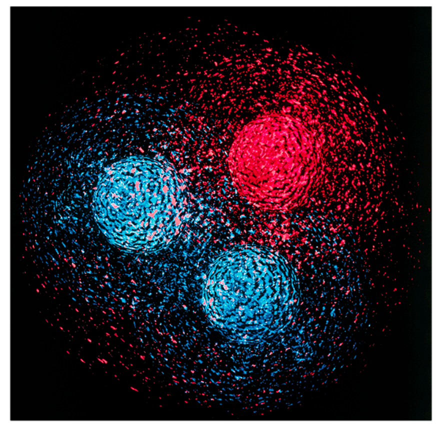
$$r_\pi^{\mathrm{NA7-84}} = 0.636 \pm 0.009_{\mathrm{stat}}\ [\mathrm{fm}]$$

Cui *et al*, PLB 822 (2021)





Proton STRUCTURE



MARATHON dataset
21 data equally spaced in x

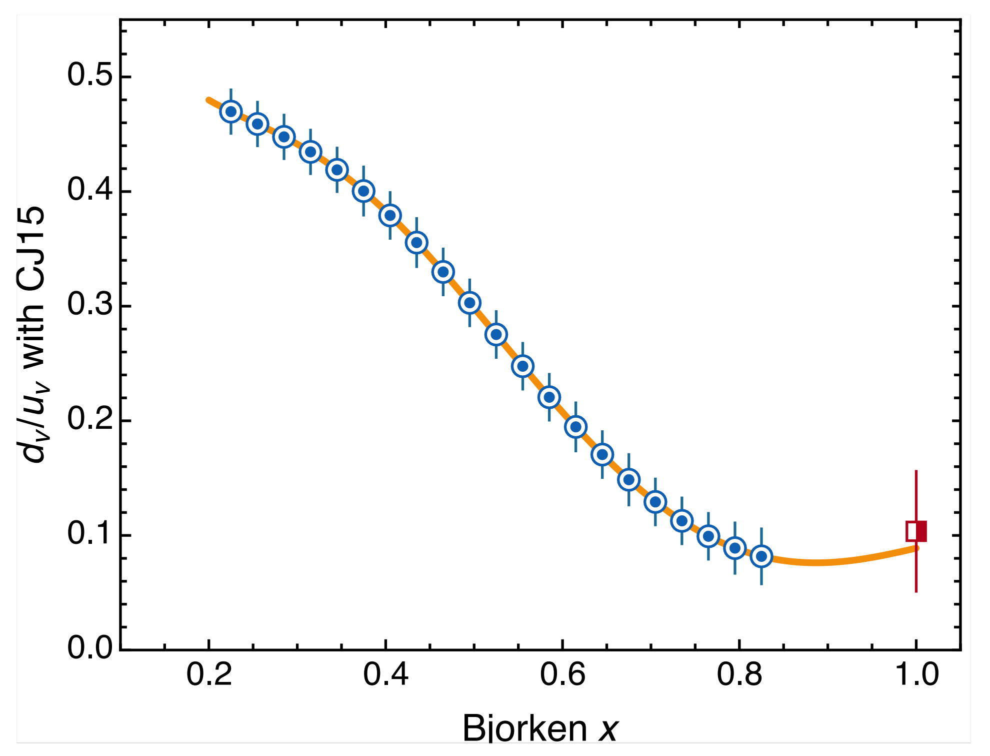
$0.225 \leq x \leq 0.825$

SPM set up

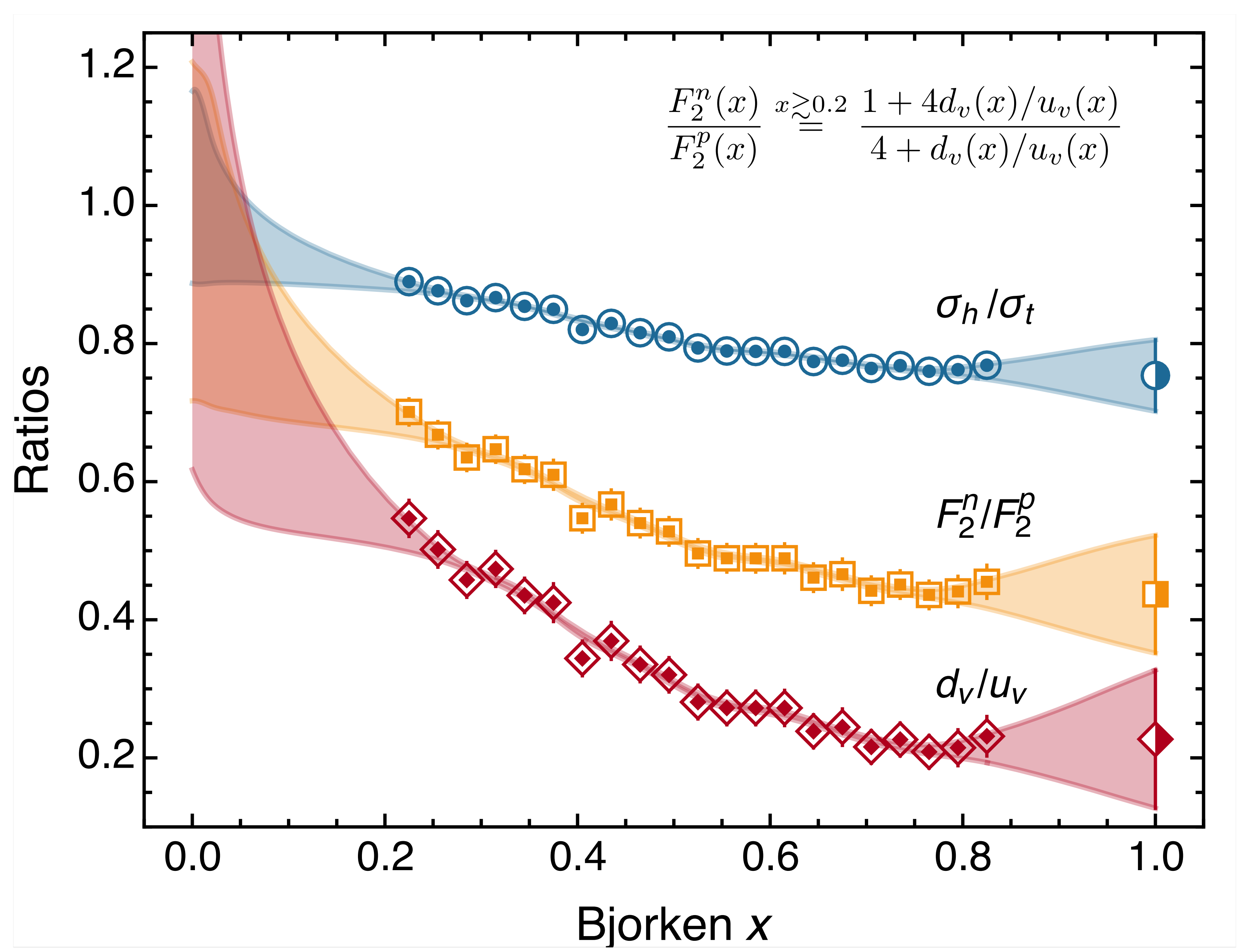
replicas	1,000	×
interpolators/replica	1,000	×
M_j	10	=
total interpolators	10,000,000	

Validation (CJ15)

$\lim_{x\to1} d_v^{\rm CJ15}/u_v^{\rm CJ15} \stackrel{\rm SPM}{=} 0.10(5)$ exp: $0.09(3)$



Accardi *et al*, PRD 93 (2016)



$$\left.\begin{array}{c}\sigma_h/\sigma_t\\ F_2^n/F_2^p\\ d_v/u_v\end{array}\right\} \stackrel{x\simeq0}{=} \left\{\begin{array}{l}1.026\pm0.139\,,\\ 0.962\pm0.245\,,\\ 1.323\pm0.706\,.\end{array}\right.$$

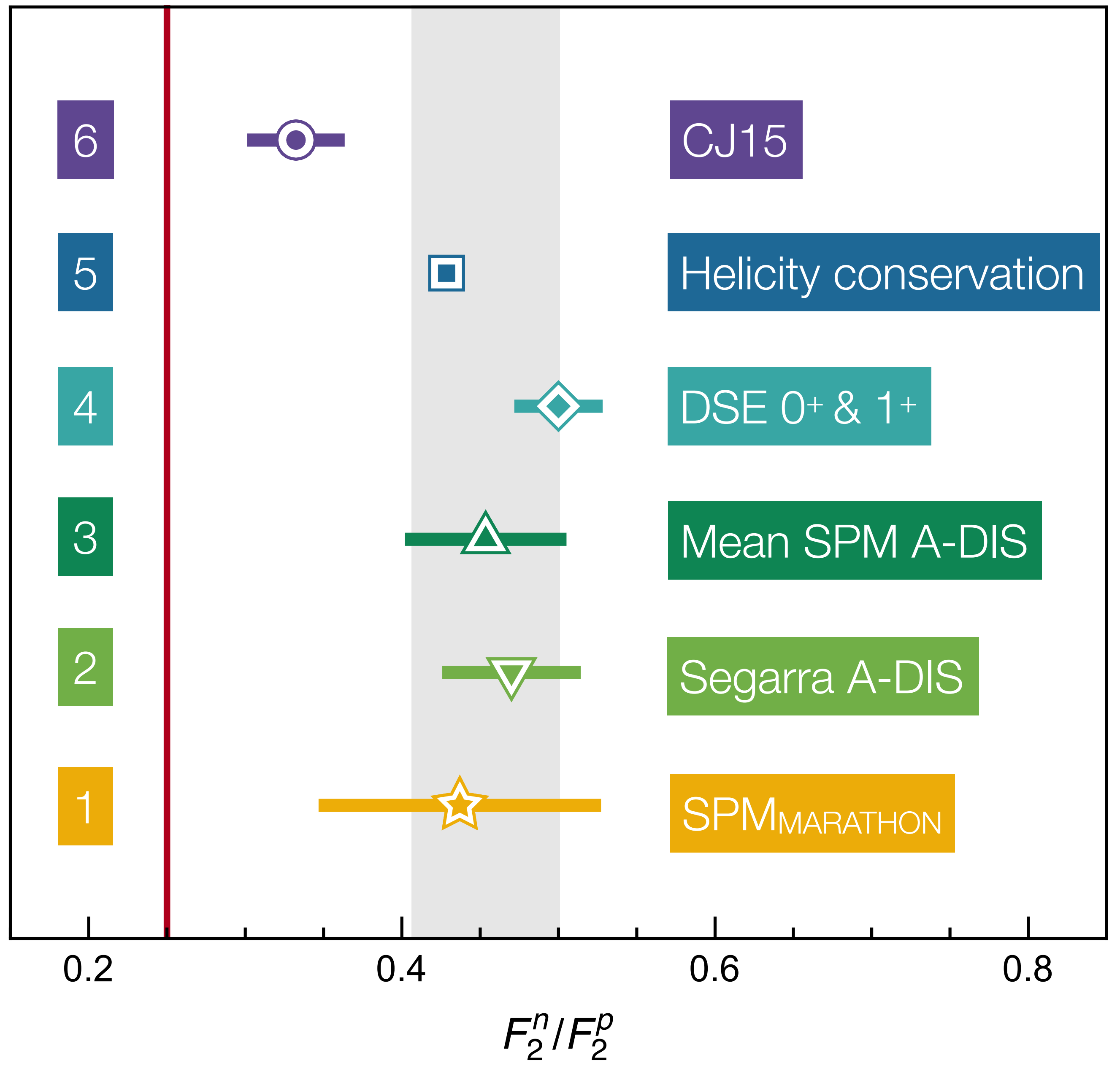
consistent with sea-quark dominance at low x

$$\left.\begin{array}{c}\sigma_h/\sigma_t\\ F_2^n/F_2^p\\ d_v/u_v\end{array}\right\} \stackrel{x\simeq1}{=} \left\{\begin{array}{l}0.754\pm0.052\,,\\ 0.437\pm0.085\,,\\ 0.227\pm0.100\,.\end{array}\right.$$

$$\frac{F_2^n(x)}{F_2^p(x)} = \frac{2\mathcal{R}_{ht} - F_2^h/F_2^t}{2F_2^h/F_2^t - \mathcal{R}_{ht}}$$

$$\mathcal{R}_{ht}(x=1) = 1.019(13)$$

$$\lim_{x\to1}\frac{d_v(x)}{u_v(x)} = 0.230 \pm 0.057$$



$$F_2^n/F_2^p\big|_{x\to1}^{\rm SPM\,\&\,DIS-A} = 0.454\pm0.047$$

Lower Nachtmann bound excluded at a fairly high level of accuracy: Ψ cannot contain scalar-diquarks only!

Agrees with:

Farrar *et al*, PRL 35 (1975); Brodsky *et al*, NPB 441 (1995)

uncorrelated SU(4) spin-flavour proton wave function and helicity conservation in high-Q^2 interactions

Poincaré-covariant Faddeev equation approach giving rise to scalar and vector diquarks correlations with dynamically determined strengths

Segovia *et al*, FBS 55 (2014); Xu *et al*, PRD 92 (2015)