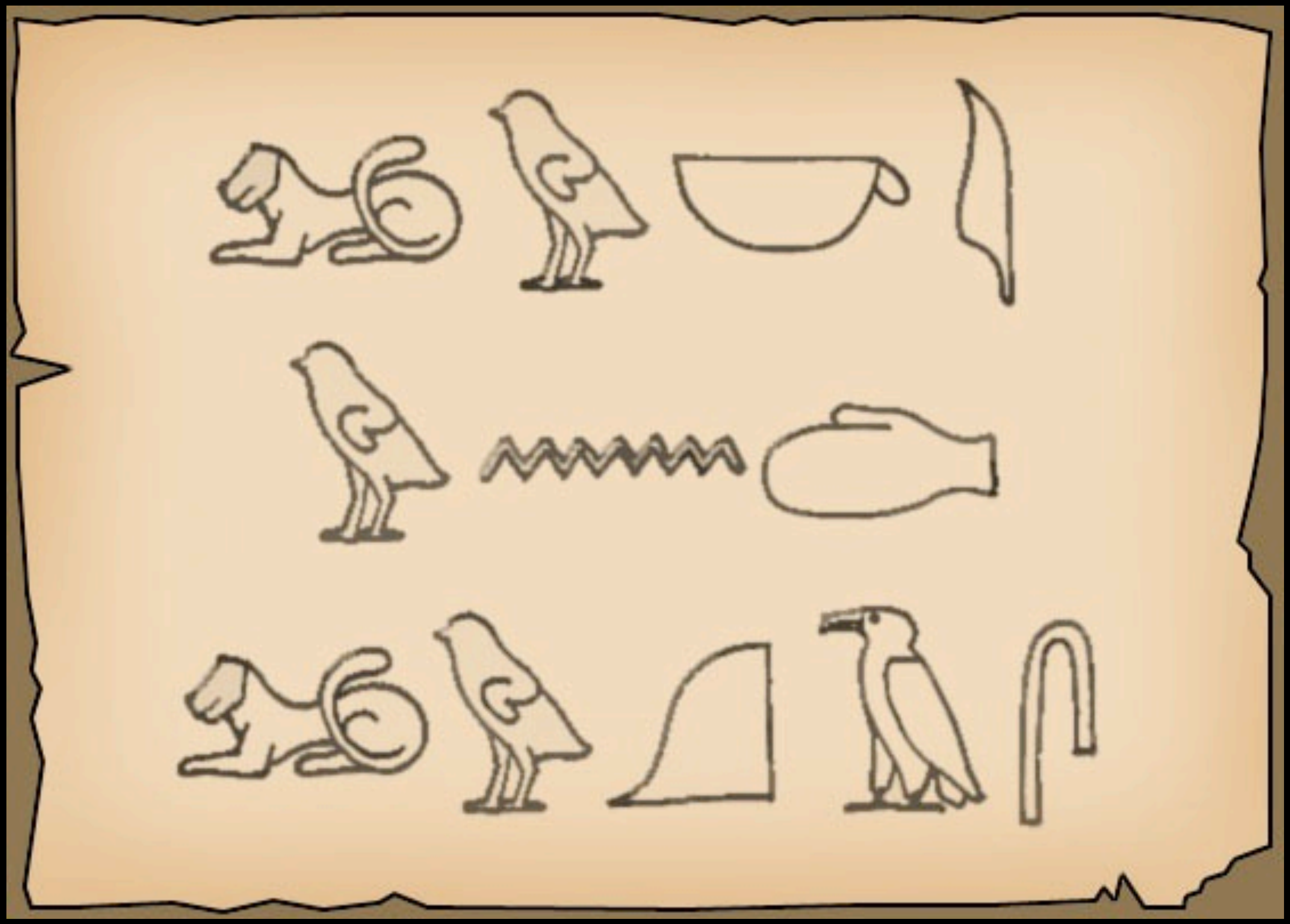
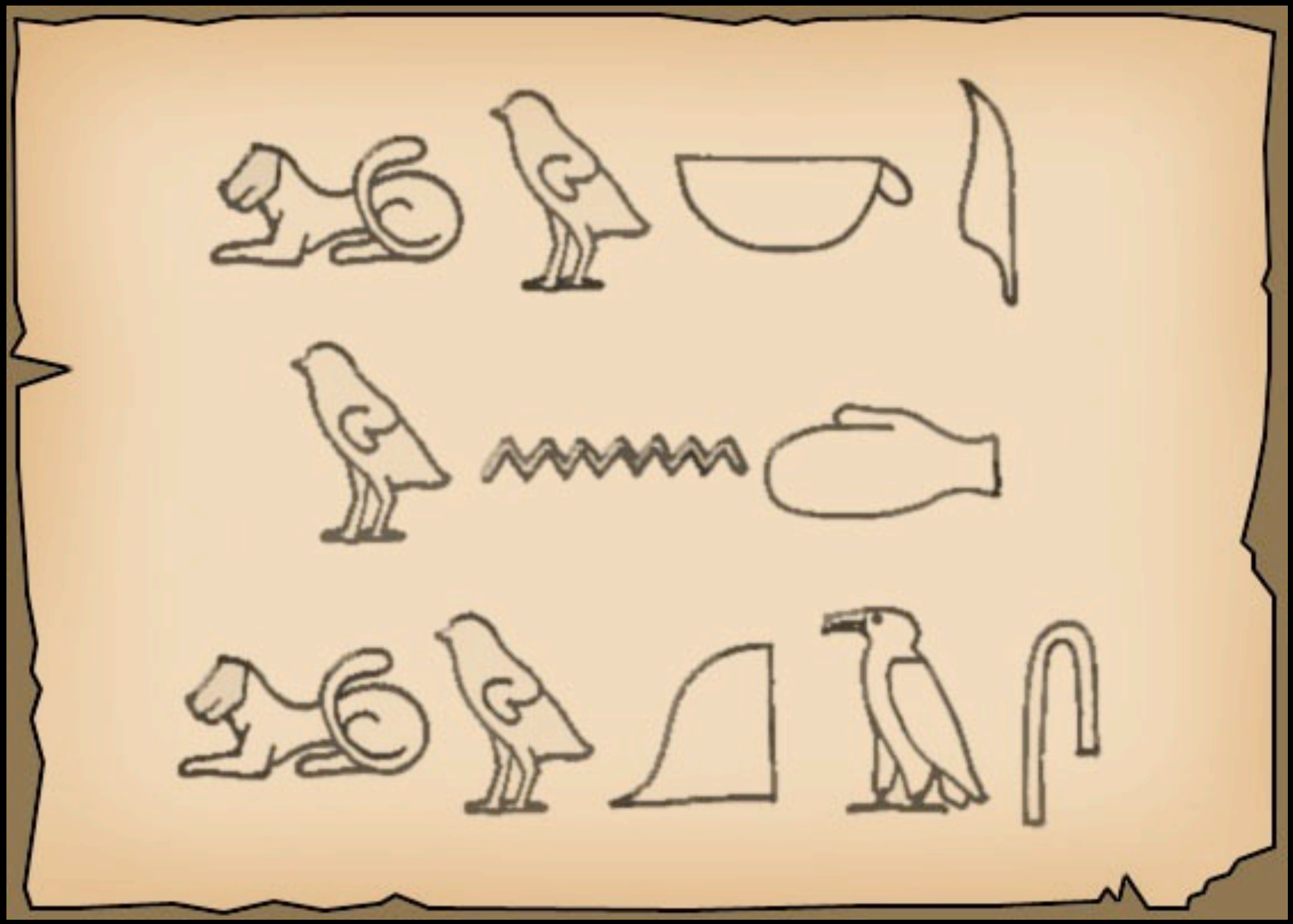
Data Preservation and Completed Experiments



DPHEP JADE OPAL H1 ZEUS

Data Preservation and Completed Experiments



DPHEP JADE OPAL H1 ZEUS

MAX-PLANCK-GESELLSCHAFT

Max-Planck-Institut für Physik
(Werner-Heisenberg-Institut)

Physics Cases for Data Preservation

- long-term completion and extension of scientific program
- cross-collaboration analyses
- Data re-use
- education, training and outreach

Physics Cases for Data Preservation

- long-term completion and extension of scientific program
- cross-collaboration analyses
- Data re-use
- education, training and outreach

„normal“ fate of HEP experiment‘s data:

- full use for completion of analyses (~5 years)
- periodic copy of data (and MC ?) to modern media (every 2 yrs, for 10-15 years)
- software code not maintained; unable to run after ~5-10 years

Study Group for Data Preservation and Long Term Analysis in High Energy Physics

recognised by ICFA

dphep.org

DPHEP

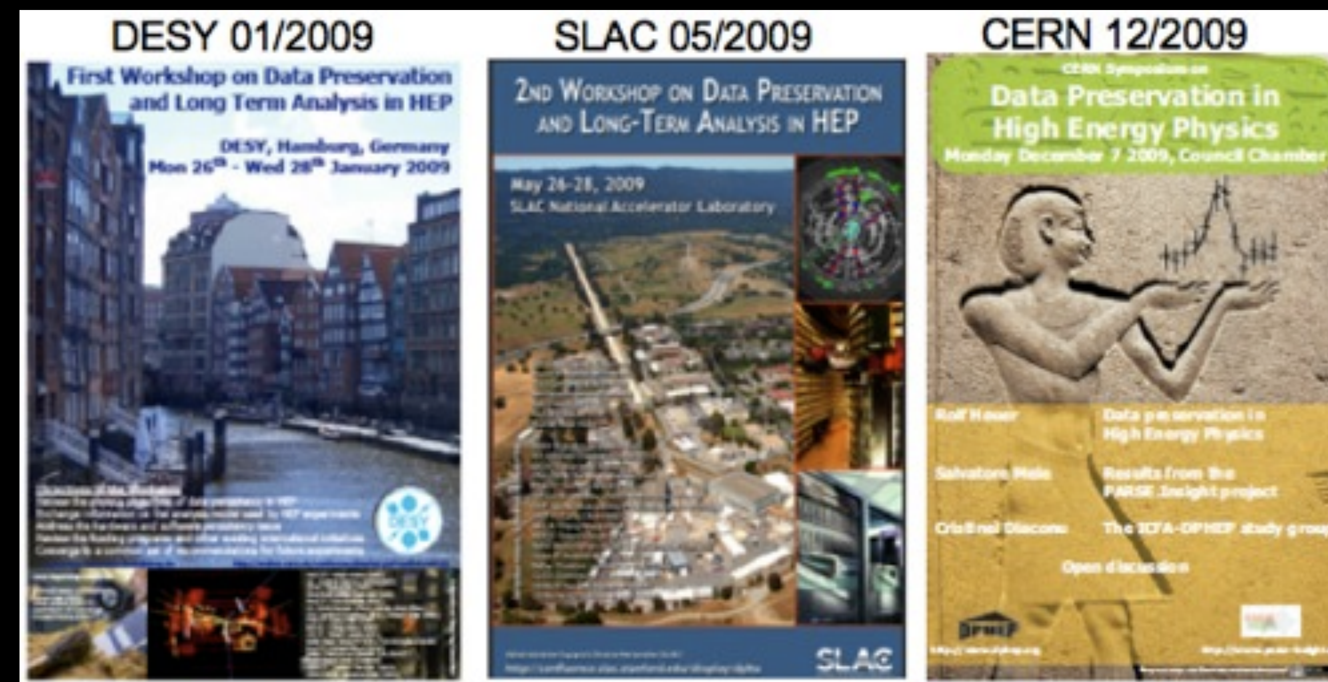
Study Group Committees

International Steering Committee

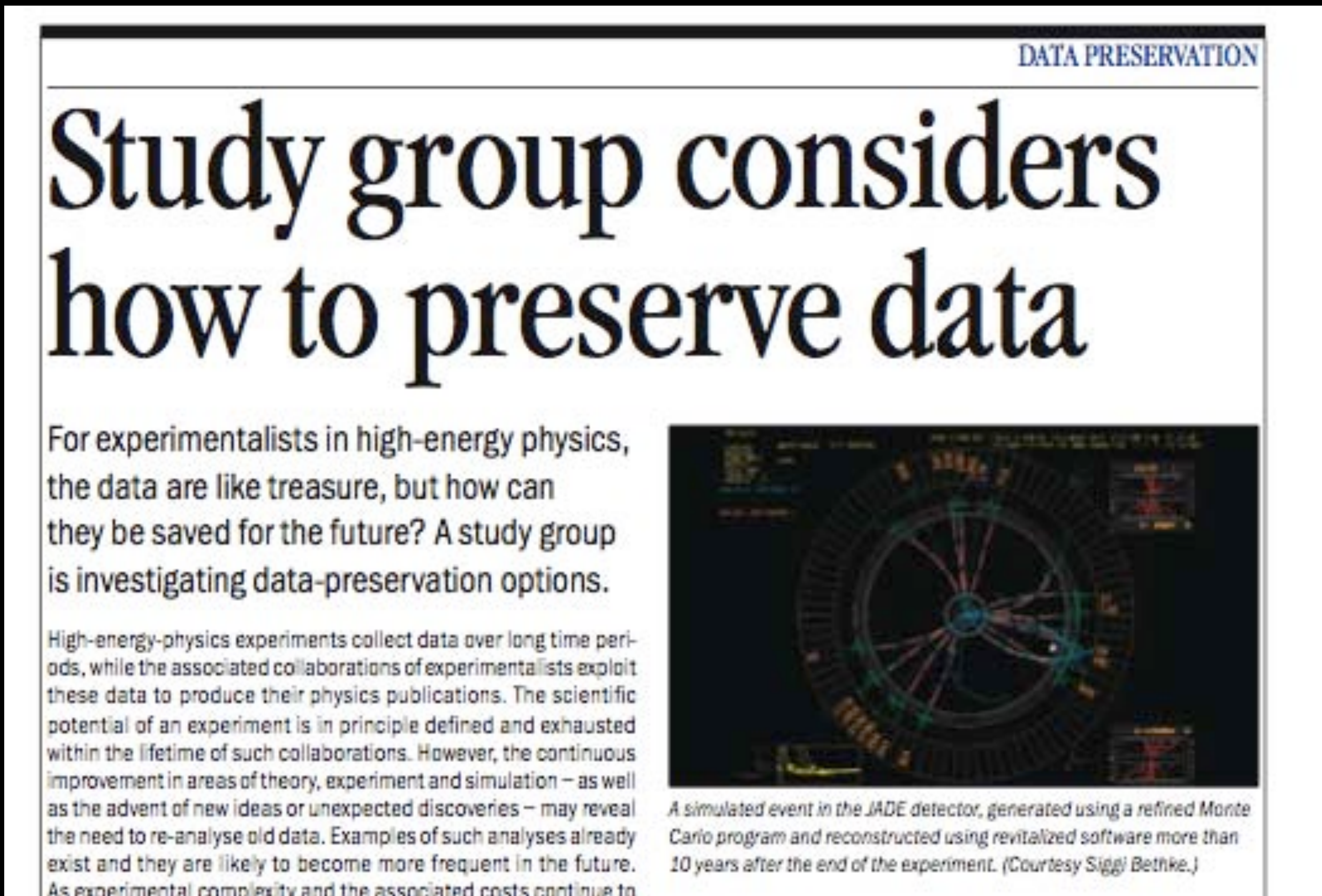
DESY-IT: Volker Gülzow (DESY)
H1: Cristinel Diaconu (CPPM/DESY) (Chair)
ZEUS: Tobias Haas (DESY)
FNAL/DoE: Amber Boehnlein (DoE)
FNAL-IT: Victoria White (FNAL)
D0: Dmitri Denisov, St. Soldner-Rembold (FNAL)
CDF: Jacobo Konigsberg, Robert Roser (FNAL)
IHEP-IT: Gang Chen (IHEP)
BES III: Yifang Wang (IHEP)
KEK-IT: Takashi Sasaki (KEK)
Belle: Masanori Yamauchi (KEK), Tom Browder (Hawaii)
SLAC-IT: Richard Mount (SLAC)
BaBar: Francois Le Diberder (LAL/SLAC)
CERN-IT: Frederic Hemmer (CERN)
CERN/PARSE: Salvatore Mele (CERN)
CLEO: David Asner (Carleton)
STFC: John Gordon (RAL)

International Advisory Committee

Jonathan Dorfan (SLAC) (co-chair)
Siegfried Bethke (MPI Munich) (co-chair)
Young-Kee Kim(FNAL)
Hiroaki Aihara (U.Tokio)
D. Boutigny (IN2P3)
M. Peskin (SLAC)
Gigi Rolandi (CERN)



Data Preservation – Press Coverage



DATA PRESERVATION

Study group considers how to preserve data

For experimentalists in high-energy physics, the data are like treasure, but how can they be saved for the future? A study group is investigating data-preservation options.

High-energy-physics experiments collect data over long time periods, while the associated collaborations of experimentalists exploit these data to produce their physics publications. The scientific potential of an experiment is in principle defined and exhausted within the lifetime of such collaborations. However, the continuous improvement in areas of theory, experiment and simulation – as well as the advent of new ideas or unexpected discoveries – may reveal the need to re-analyse old data. Examples of such analyses already exist and they are likely to become more frequent in the future. As experimental complexity and the associated costs continue to

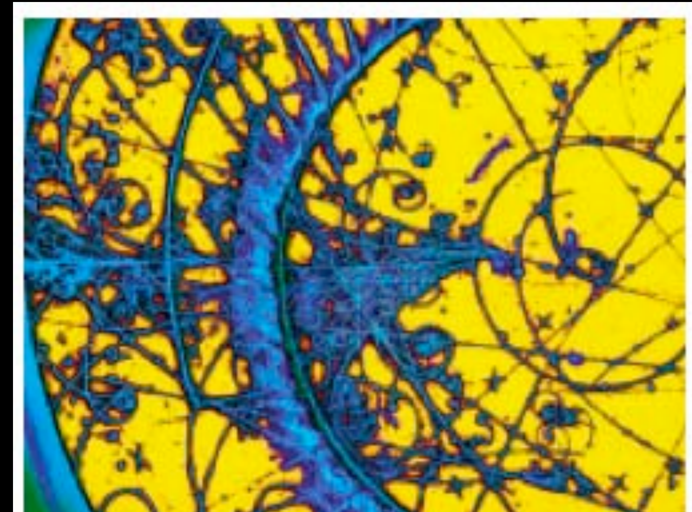
A simulated event in the JADE detector, generated using a refined Monte Carlo program and reconstructed using revitalized software more than 10 years after the end of the experiment. (Courtesy Siggi Bethke.)

CERN Courier. 04/2009

Symmetry magazine, Sept. 2009

Science Rescue of Old Data Offers Lesson for Particle Physicists

Hieroglyphen im Teilchenlabor Frankfurter Rundschau, February 2010



Berliner Zeitung, February 2010

Preservation Models

Preservation Model	Use case
1. Provide additional documentation	Publication-related information search
2. Preserve the data in a simplified format	Outreach, simple training analyses
3. Preserve the analysis level software and data format	Full scientific analysis based on existing reconstruction
4. Preserve the reconstruction and simulation software and basic level data	Full potential of the experimental data

(subsequent models are inclusive)

Preservation Models

Preservation Model	Use case
1. Provide additional documentation	Publication-related information search
2. Preserve the data in a simplified format	Outreach, simple training analyses
3. Preserve the analysis level software and data format	Full scientific analysis based on existing reconstruction
4. Preserve the reconstruction and simulation software and basic level data	Full potential of the experimental data

(subsequent models are inclusive)

- level 1 already implemented in many cases (physics notes, suppl. material)
- use of level 2 underestimated, but successfully used (master classes)
- level 3 can only use existing reconstruction/calibration and MC data
- level 4 requires full functionality of reconstruction and simulation software

Preservation Models

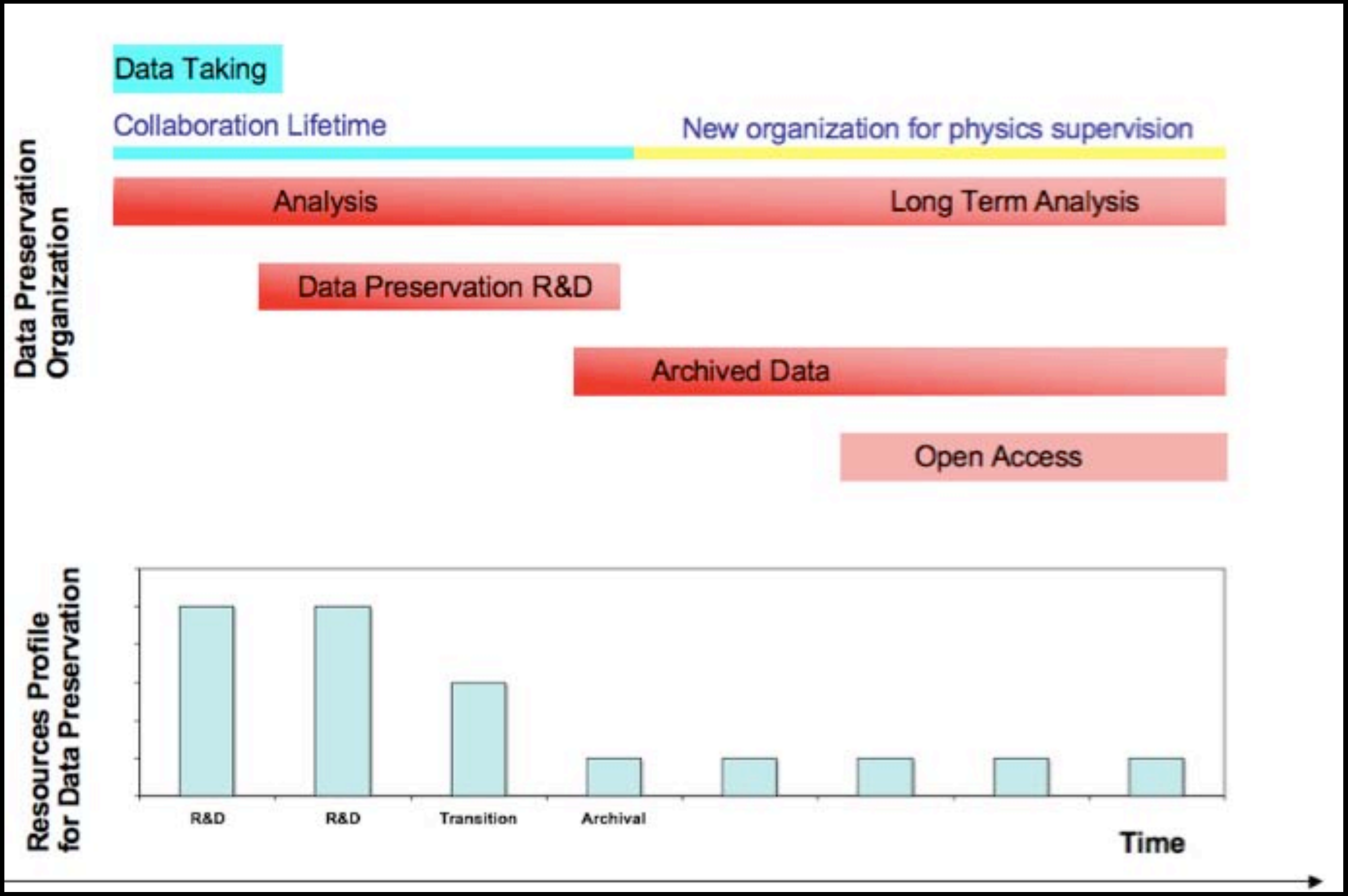
Preservation Model	Use case
1. Provide additional documentation	Publication-related information search
2. Preserve the data in a simplified format	Outreach, simple training analyses
3. Preserve the analysis level software and data format	Full scientific analysis based on existing reconstruction
4. Preserve the reconstruction and simulation software and basic level data	Full potential of the experimental data

(subsequent models are inclusive)

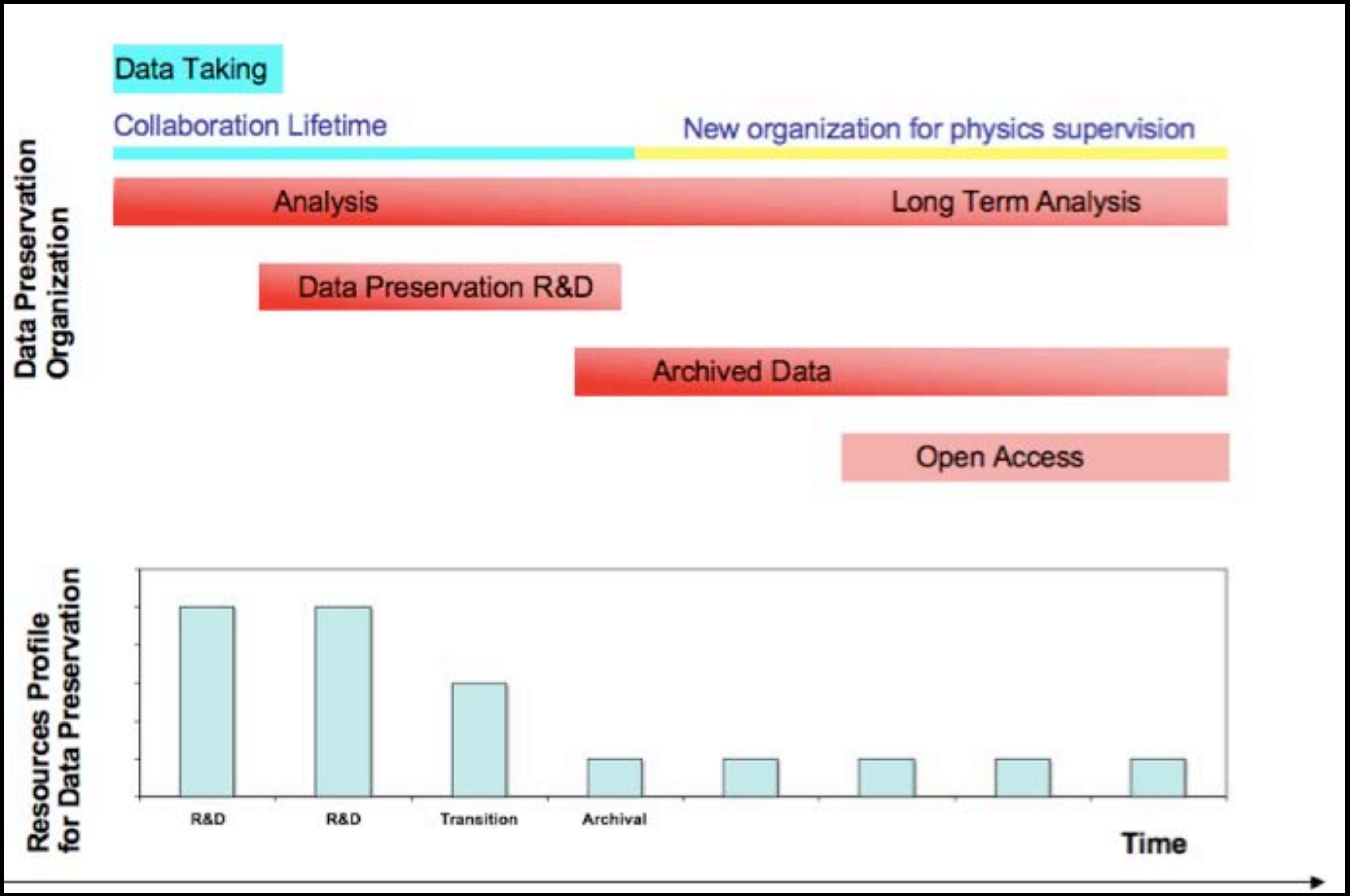
- level 1 already implemented in many cases (physics notes, suppl. material)
- use of level 2 underestimated, but successfully used (master classes)
- level 3 can only use existing reconstruction/calibration and MC data
- level 4 requires full functionality of reconstruction and simulation software

only level 4 allows full flexibility for future analyses,
but requires significant preparation, maintenance and validation

Preservation Models



Preservation Models



Further important issues (not covered here; see **arXiv:0912.0255** [hep-ex]):

- Technologies, facilities, funding
- Governance, supervision, authorship rights

actual preservation / long-term usage cases:
(relevant for MPP)

actual preservation / long-term usage cases:
(relevant for MPP)

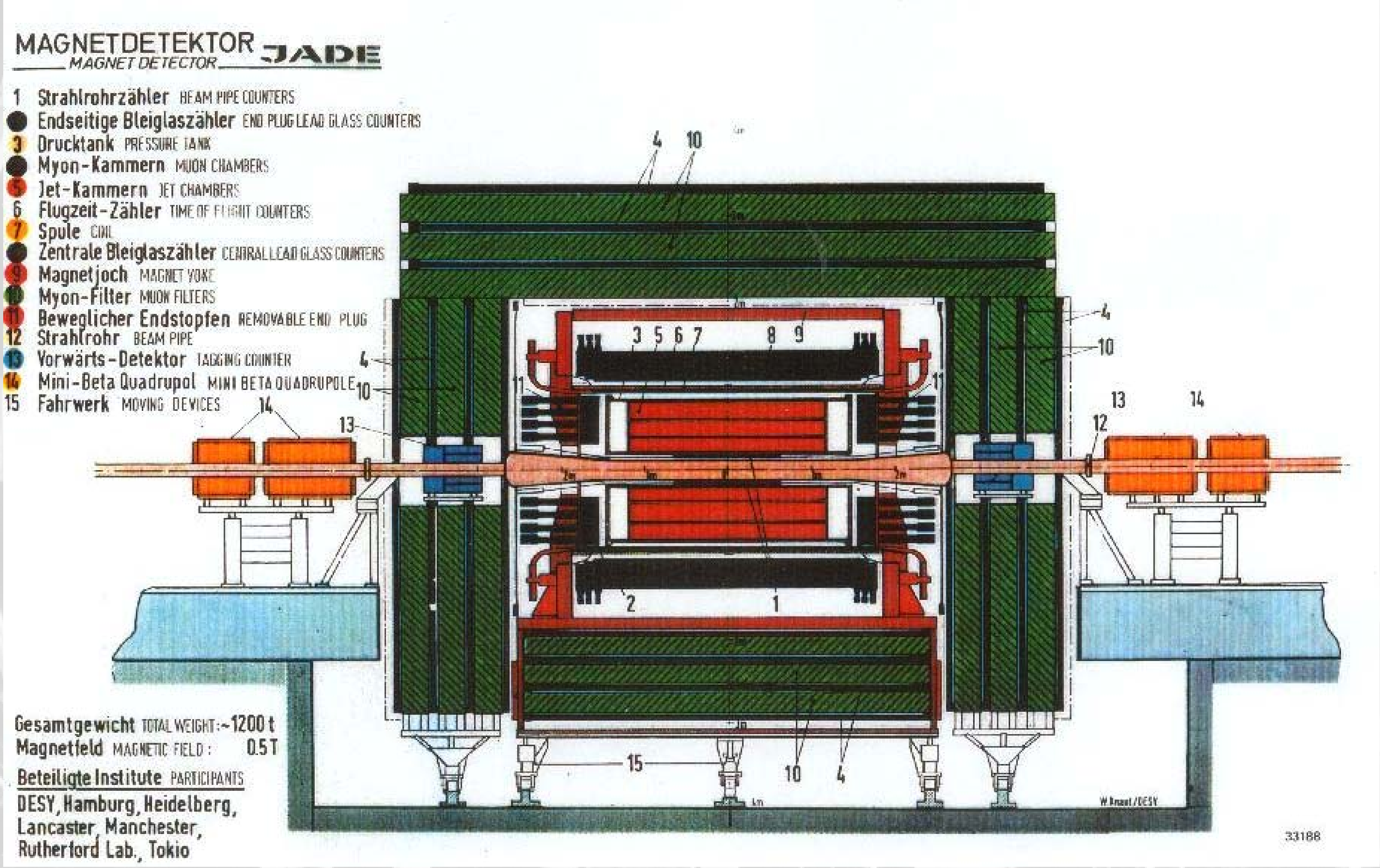
- JADE experiment at PETRA (1979-1986)
  - raw data recovery
  - (almost full) software revival
  - ability to regenerate modern MC data at detector level

- ALEPH, OPAL at LEP (1989-2000)
  - raw data on archive, Some DST copies at MPP,...
  - software barely usable (CERNLIB, ...)
  - few analyses in past years
  - no coordinated actions; slow loss of data, software...?

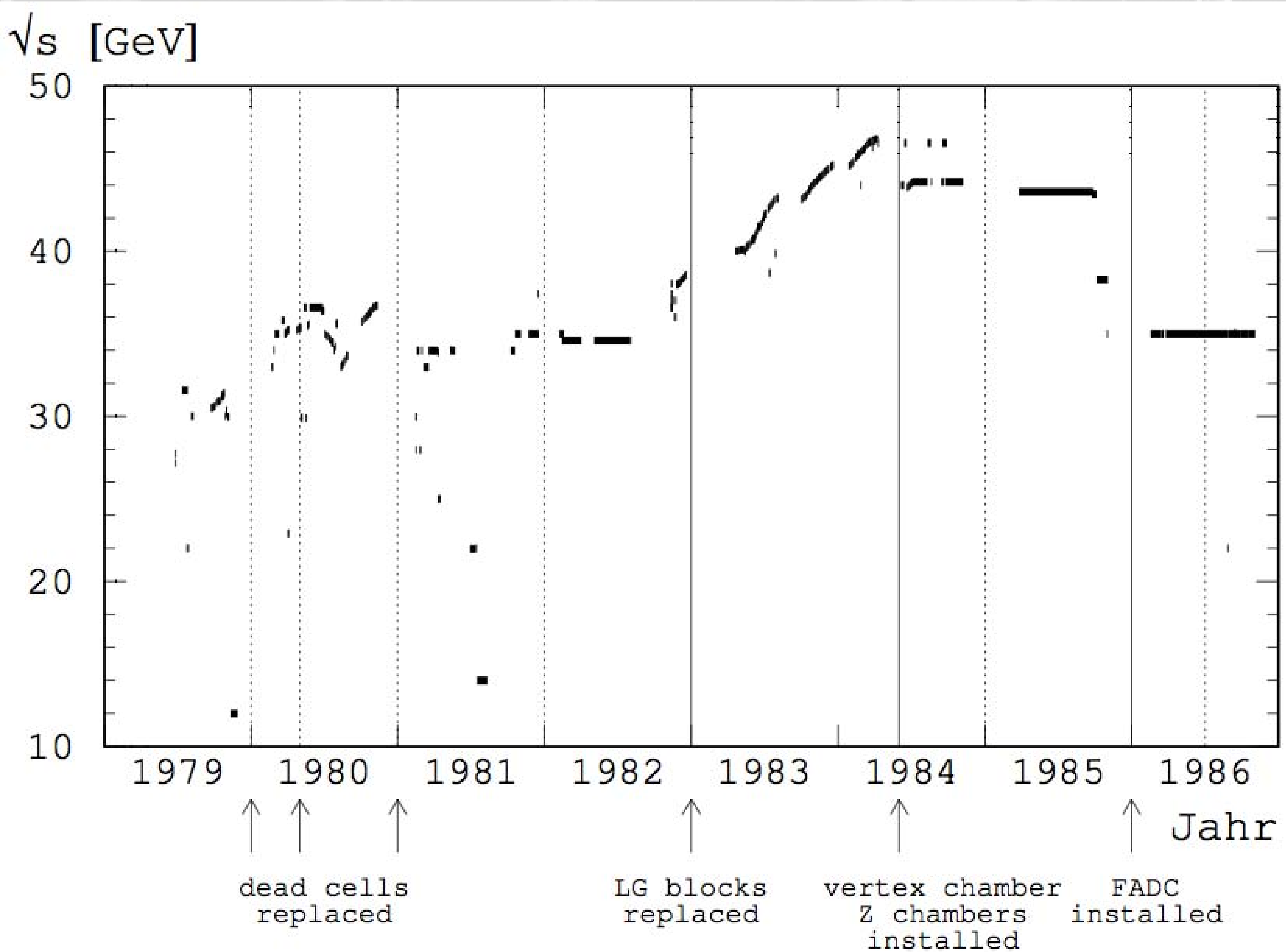
- H1, ZEUS at HERA (1992 - 2007)
  - ongoing analyses; combinations of results
  - data prepared for preservation
  - software prepared for preservation

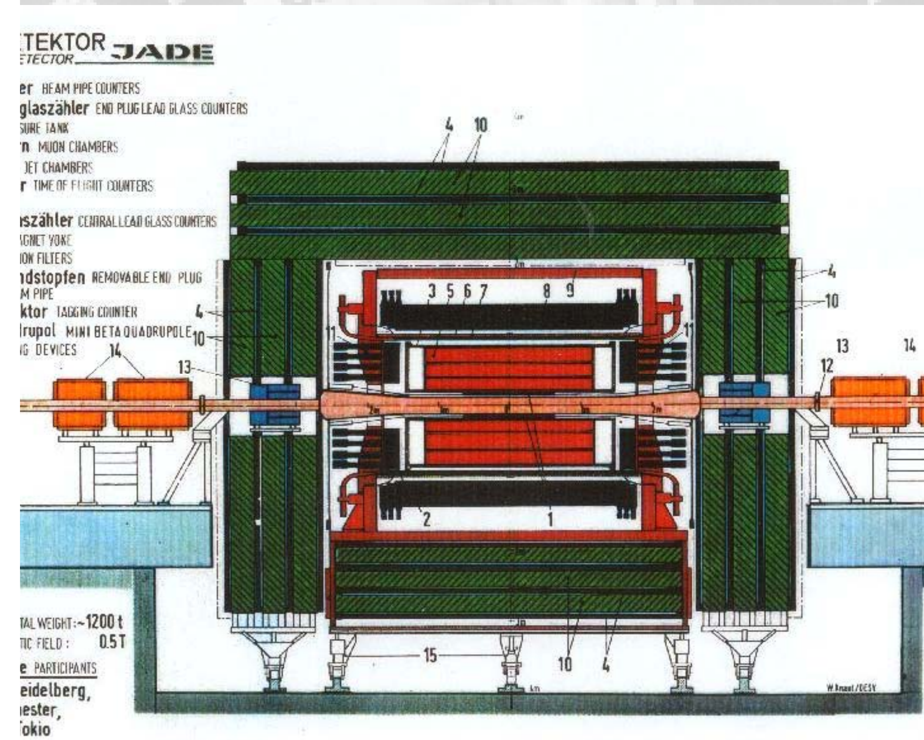
JADE @ PETRA (e+e- @ 11...46 GeV c.m.)

JADE @ PETRA (e+e- @ 11...46 GeV c.m.)



JADE @ PETRA (e+e- @ 11...46 GeV c.m.)



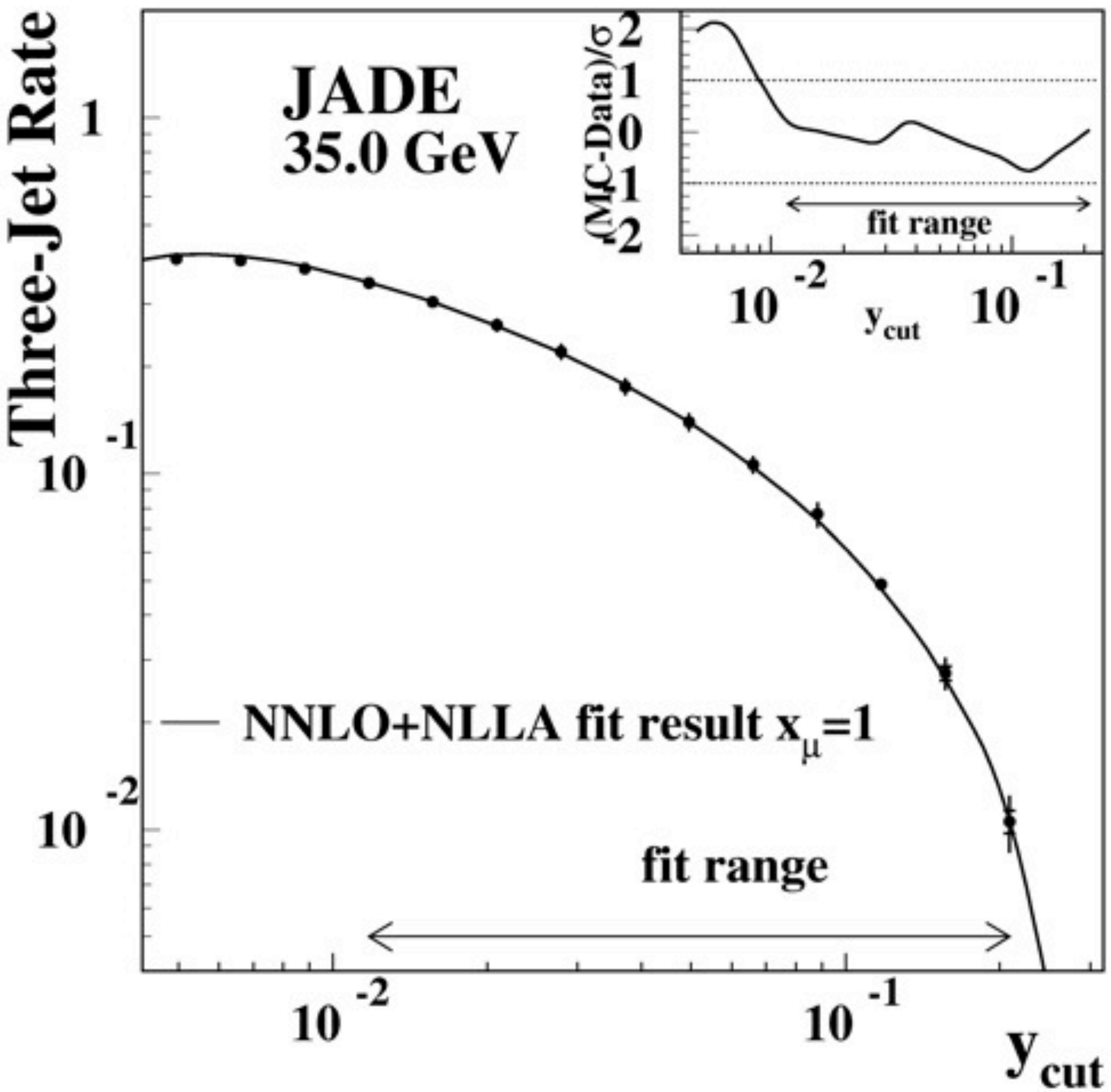


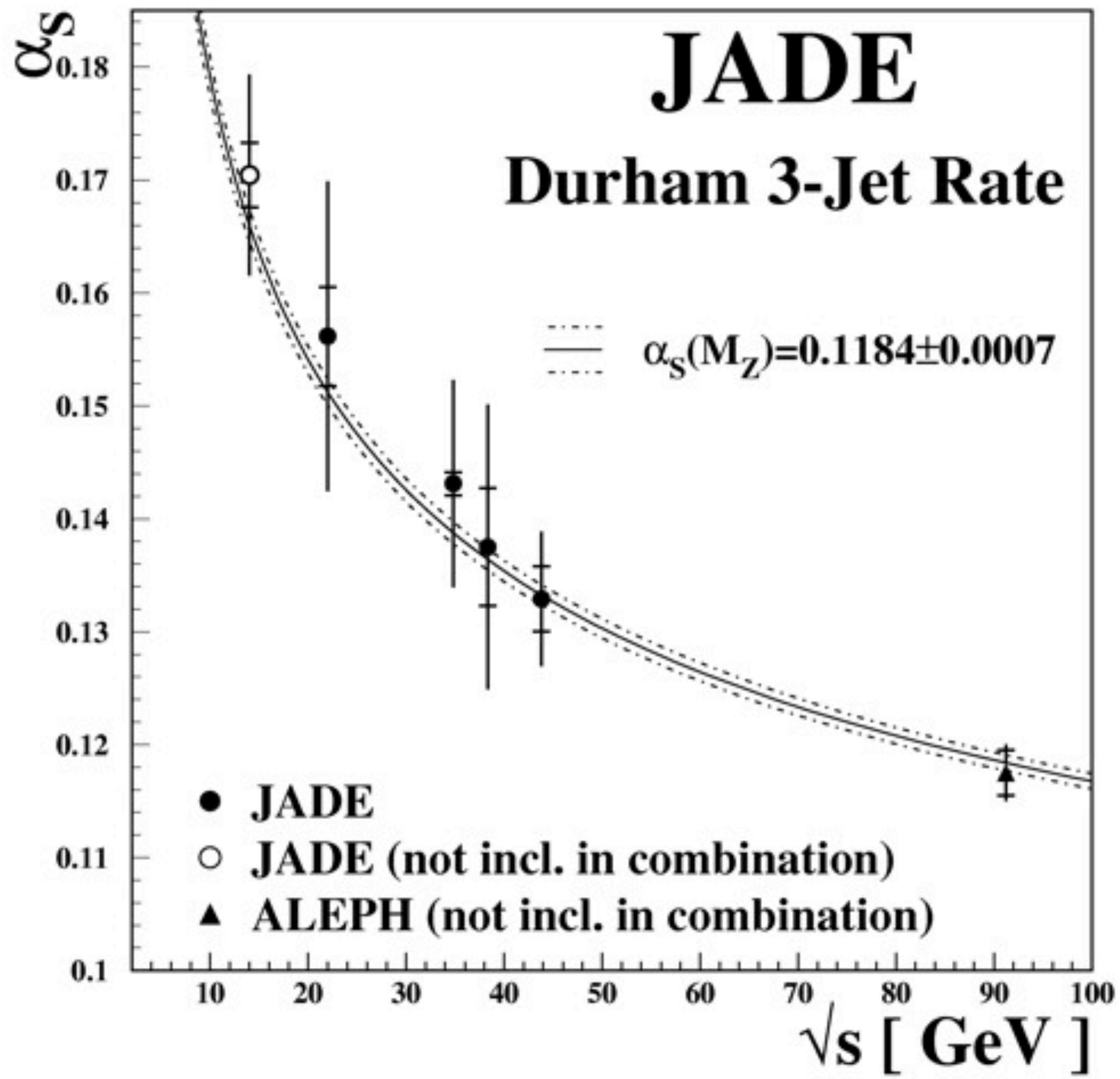
~200 pb^{-1} ; ~45.000 „good“ multihadronic events

NNLO+NLLA α_S from 3-jet rate

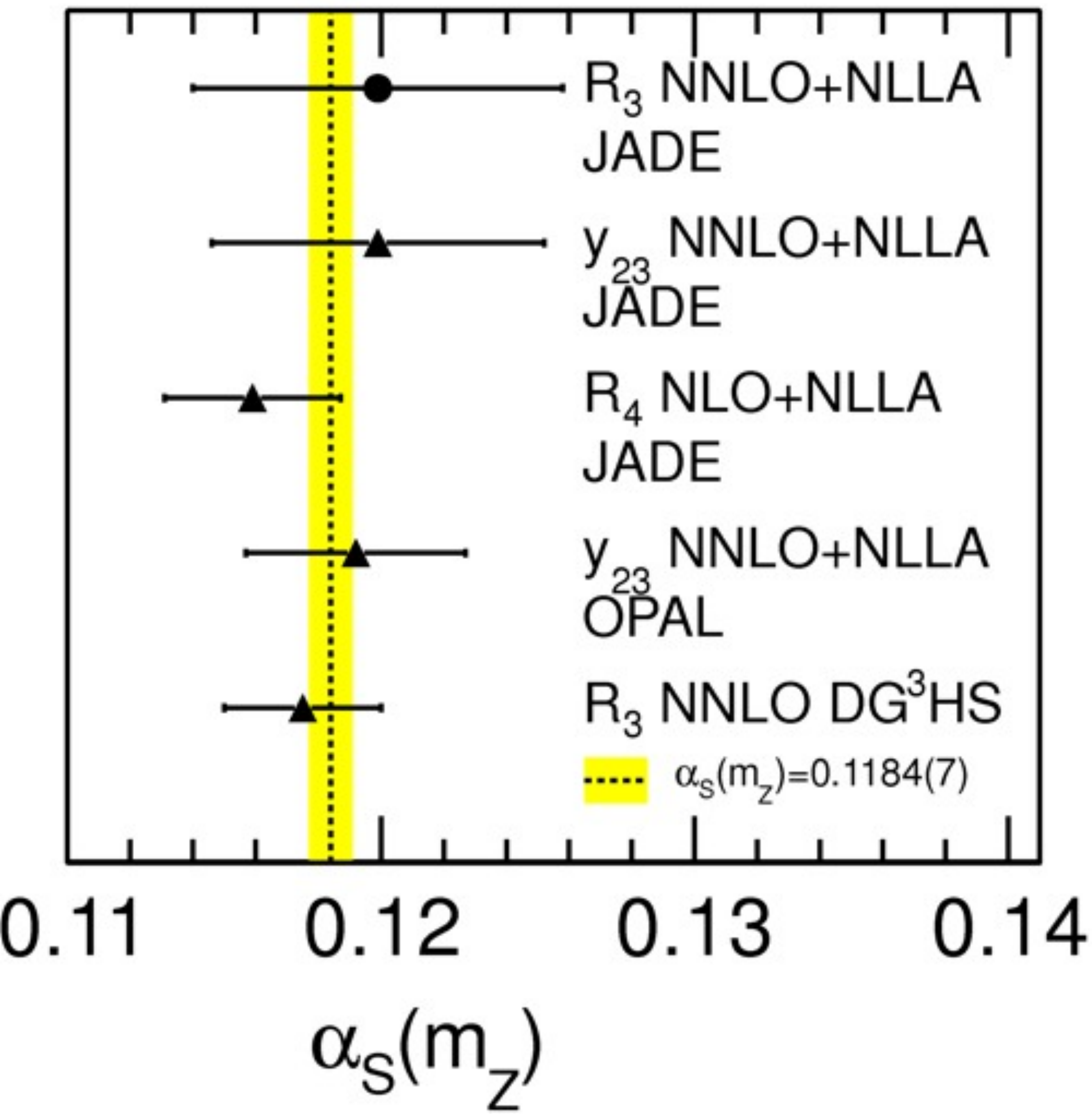
α_S (35GeV) = 0.147 ± 0.010

χ^2/d.o.f. = 11.7/10





NNLO+NLLA α_S from 3–jet rate



$\alpha_S(m_Z) = 0.1199 \quad \pm 0.0010(\text{stat})$
$\pm 0.0021(\text{exp})$
$\pm 0.0054(\text{had})$
$\pm 0.0007(\text{theo})$

NNLO or NNLO+free x_μ worse

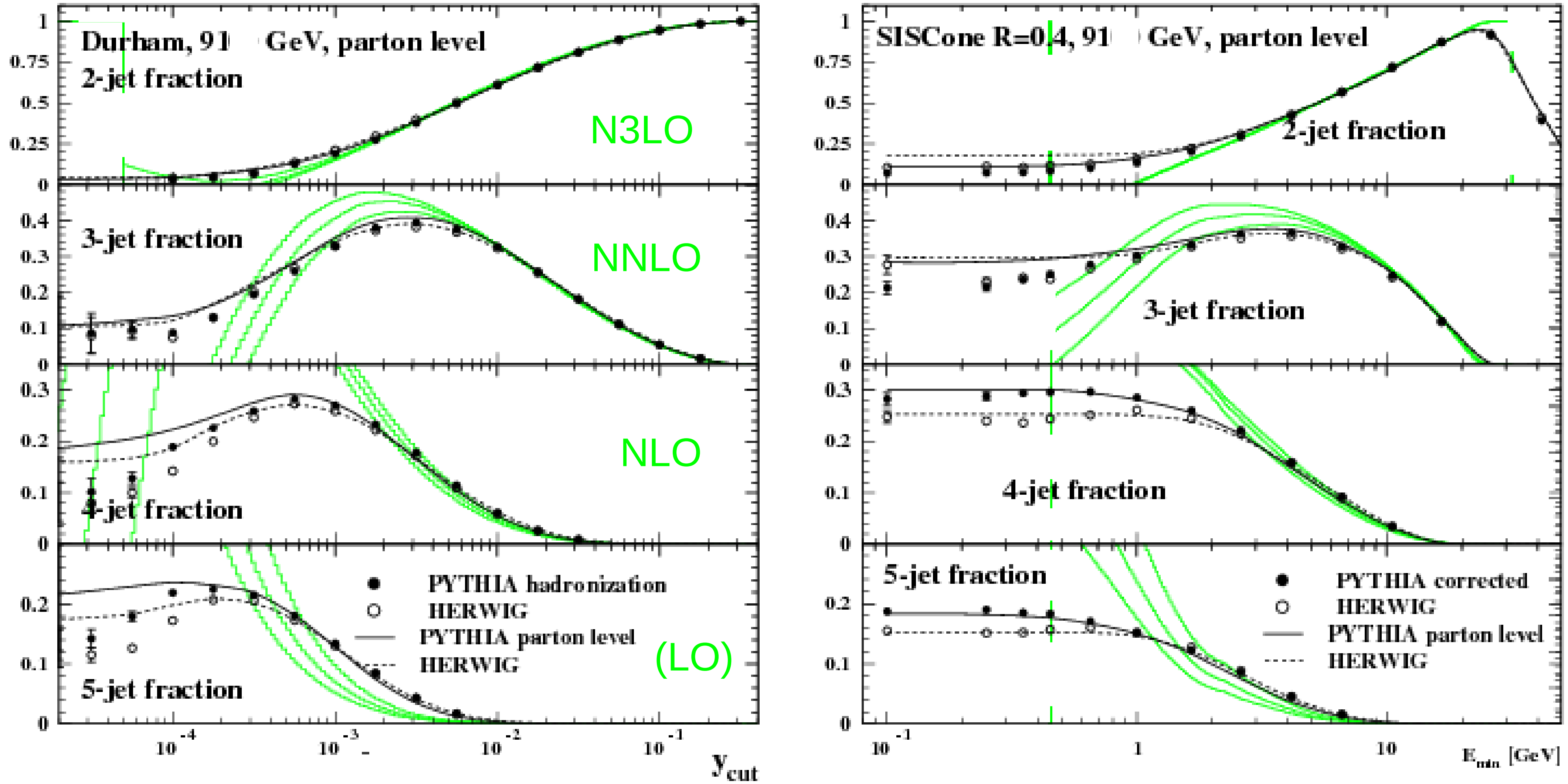
K–term important

Had. uncertainty now dominant, need better models (MC or analytic)

Bethke, Kluth, Pahl (MPP)
Schieck (LMU)

New jet algorithms/calculations with OPAL

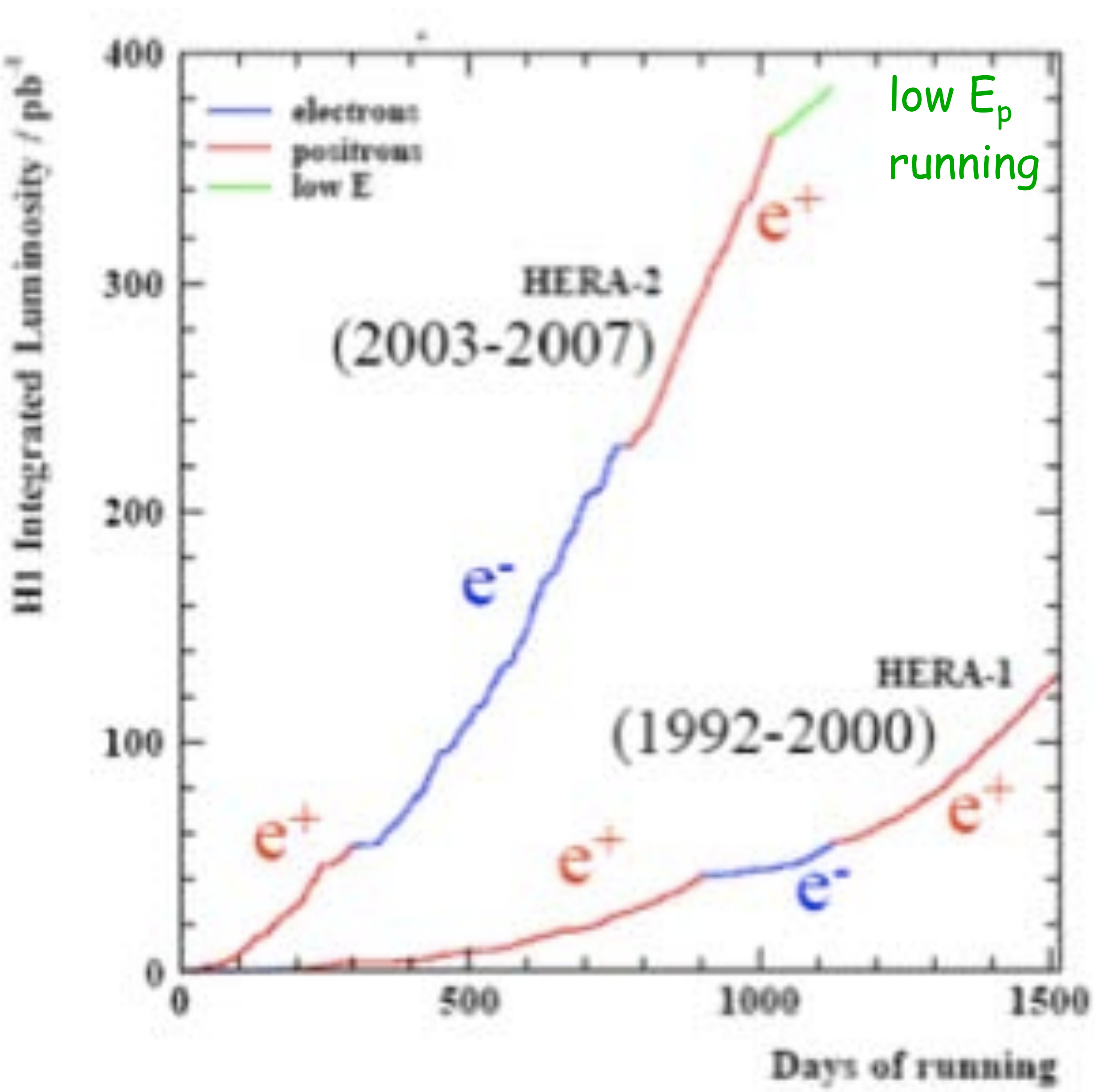
- Studied Algorithms: Durham, and Durham, Anti-kt, SISCONE with energy cut
- 2-jet fraction in N3LO from $\sigma_{Had,N3LO} = R_{2,N3LO}+R_{3,NNLO}+R_{4,NLO}+R_{5,LO}+...$



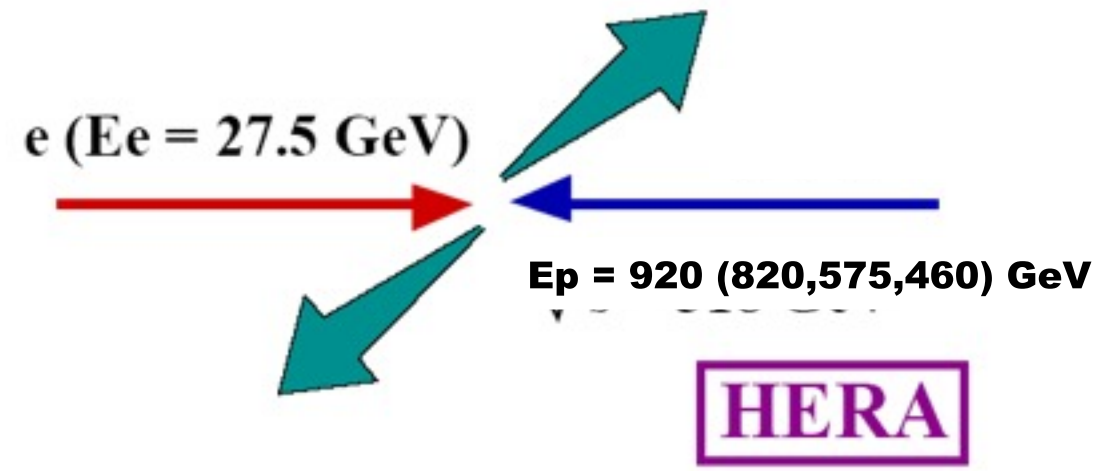
Theory : 2..5-jet rates 91 GeV. Fixed $\alpha_S(m_Z)$=0.1180, x_μ=0.5, 1, 2 .

Durham	**SISCone R=0.4**
Expect precise α_S measurement	Interested in tests of new jet algorithms in e^+e^-

HERA experiments (H1 & ZEUS)



HERA I 1992-2000 ~120 pb-1
HERA II 2003-2007 ~380 pb-1

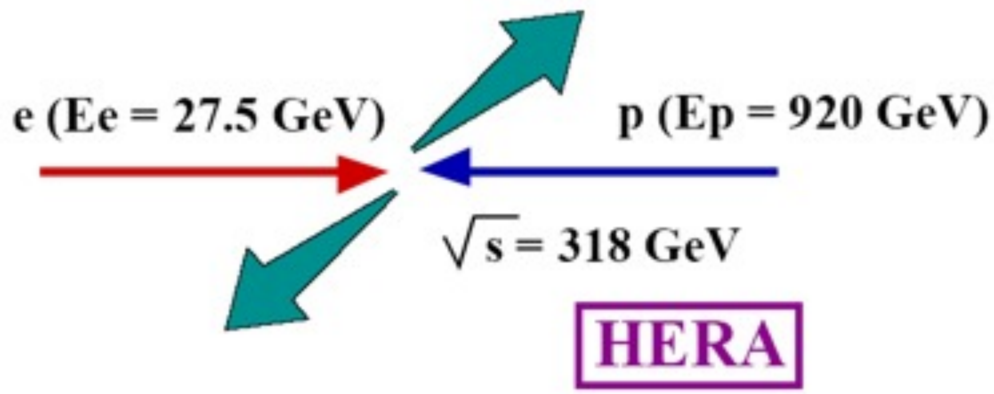


DESY, Hamburg
peak luminosity 5 10^{31} cm^{-2} sec^{-1}
$Q^2_{max} = 10^5\ GeV^2$
$\lambda_{min} \sim 1/1000\ r_{proton}$
longitudinal e-beam polarisation
$P_e=(N_R-N_L)/(N_R+N_L)$

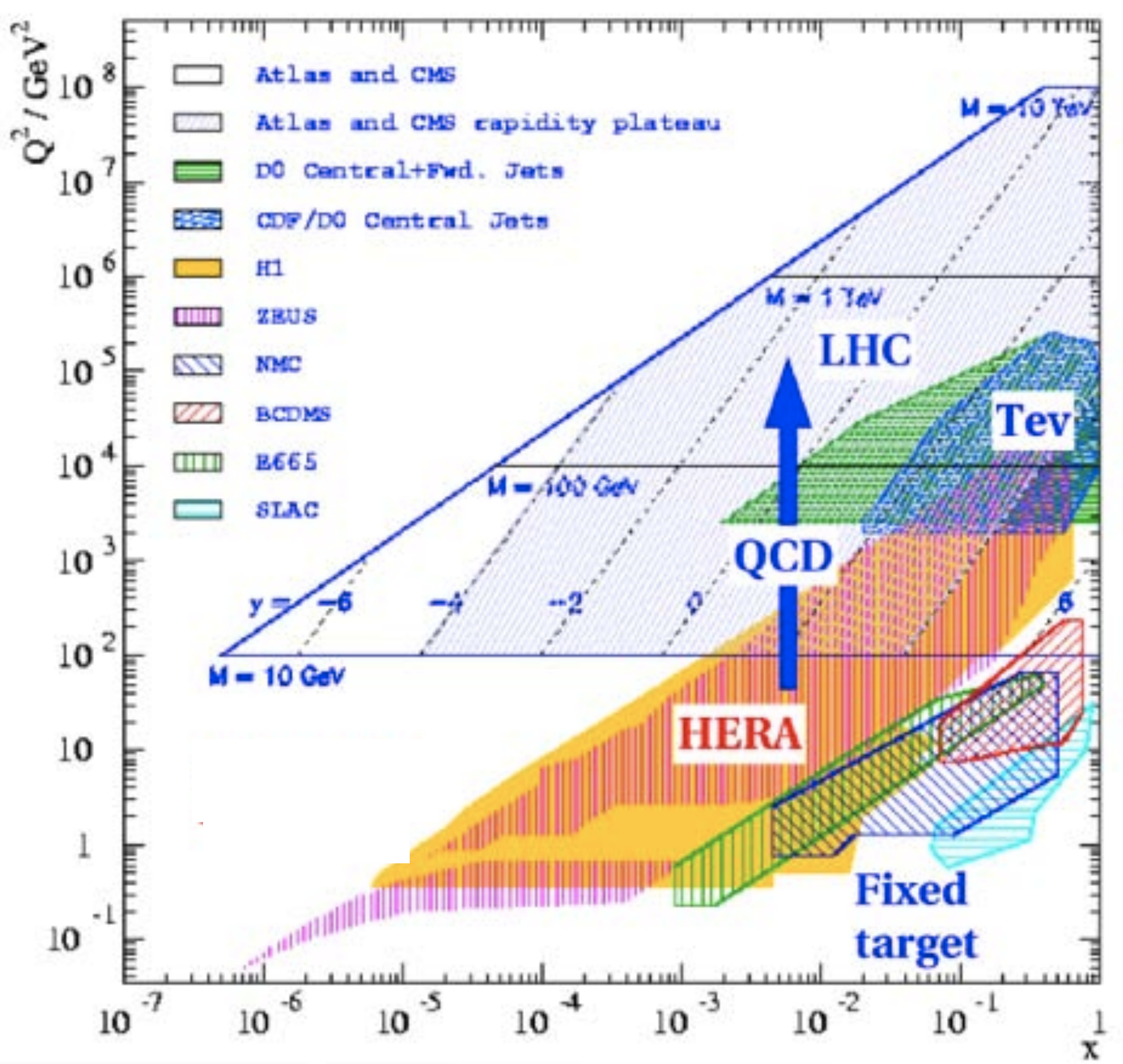
H1+ZEUS in total ~ 1 fb^{-1}
about equally shared between
- **experiments (H1, ZEUS)**
- **e^+ and e^-,**
- **positive and negative P_e**

low proton energy running for F_L

HERA experiments (H1 & ZEUS)



HERA: spans 6 orders of magnitude in x and Q^2



Progress in 2012 :

- H1 published $e^{\pm}p$ NC & CC HERA II inclusive data
- ZEUS published $e^{+}p$ NC HERA II inclusive data

→ Thus, inclusive NC and CC bulk measurements at HERA are completed and published

Plans:

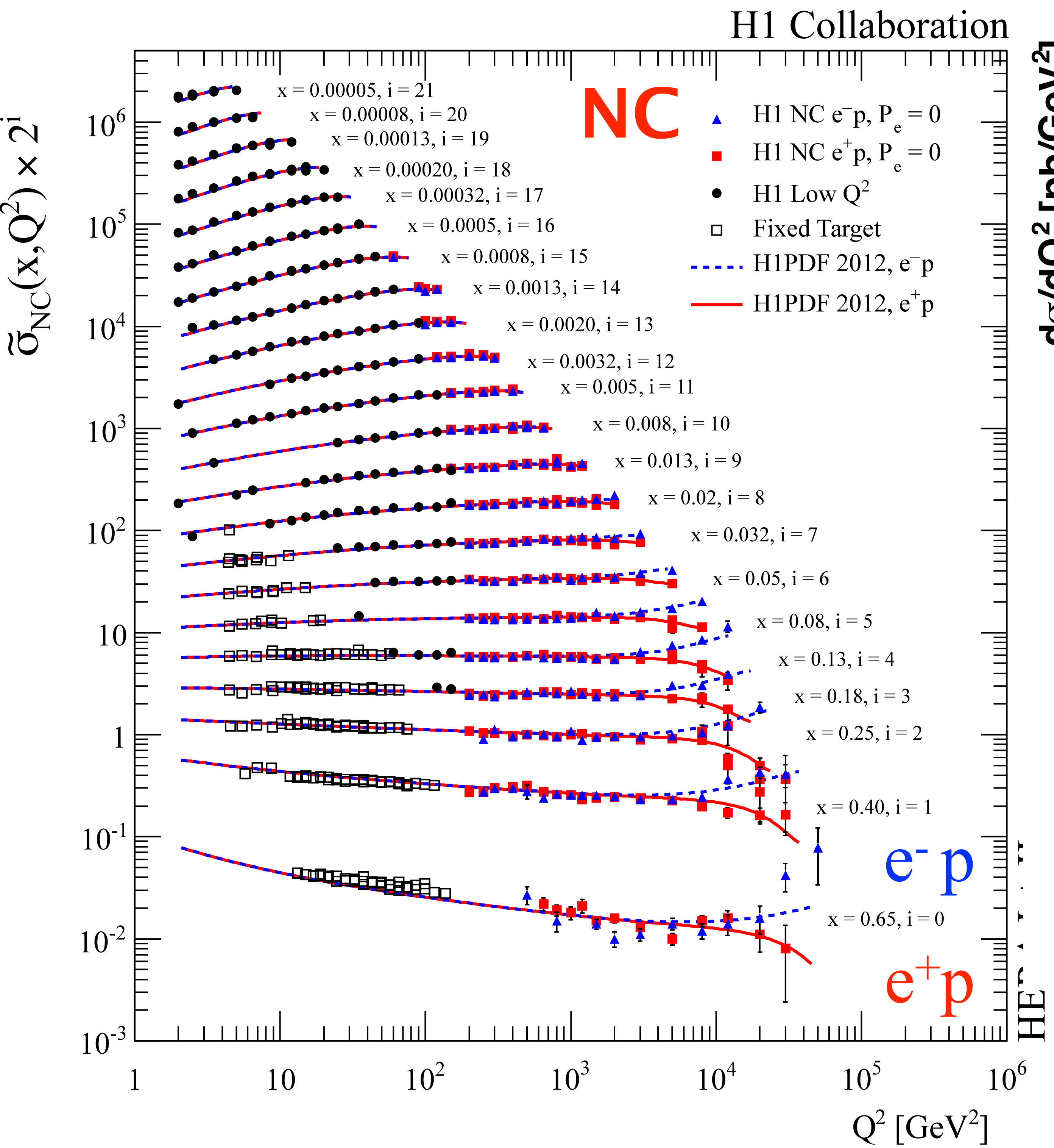
- combination of all HERA I and HERA II inclusive data
- extension of F_L measurement by ZEUS to lower Q^2
- extension of F_L measurement by H1 to high Q^2
- multi-jet data and strong coupling α_s (H1)
- NC at highest x (ZEUS)

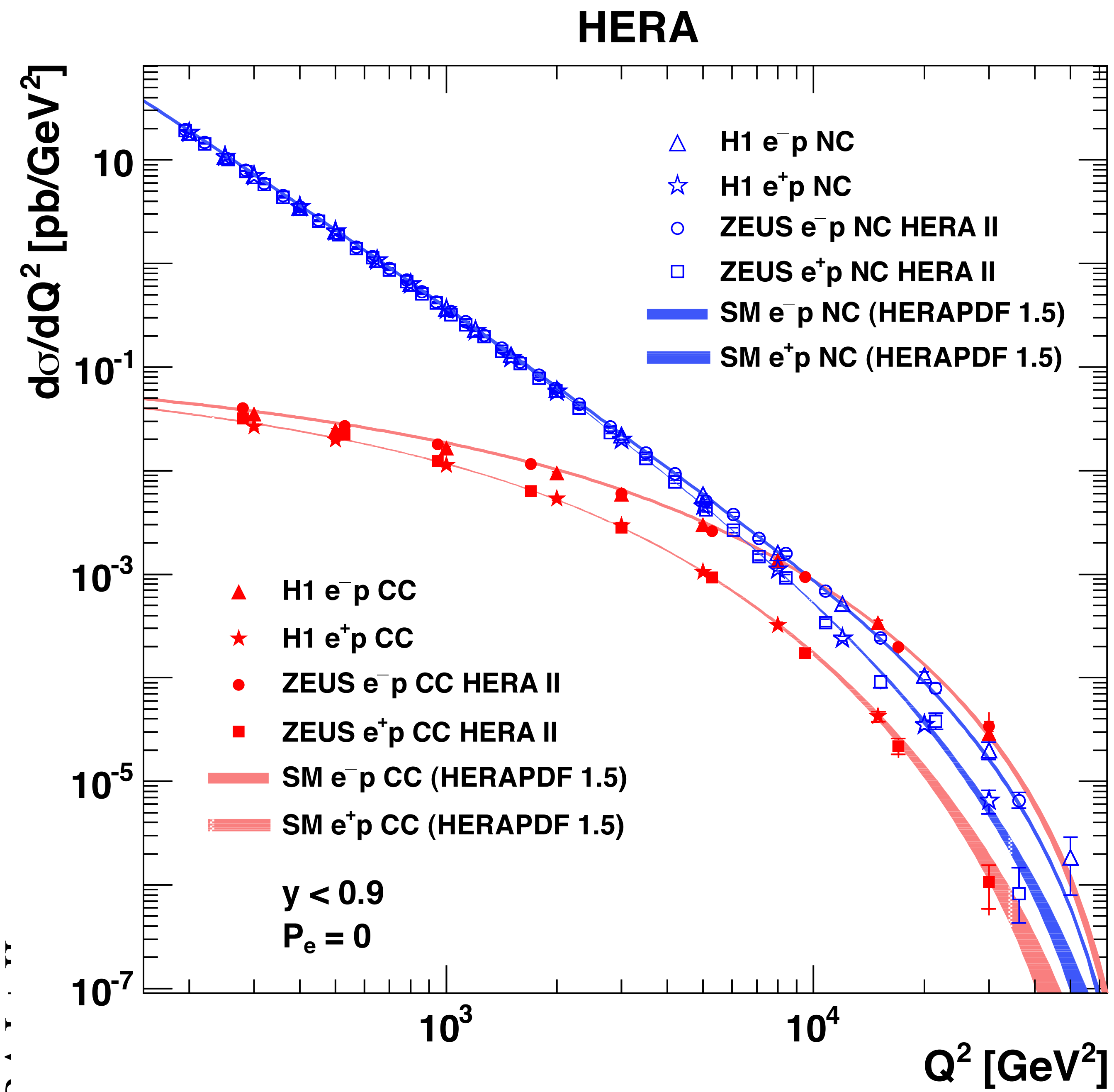
People:

Allen Caldwell (director)
Iris Abt (ZEUS, project leader, Physics Chair)
Vladimir Chekelian (H1, project leader)
Günter Grindhammer (H1, QCD convener)
Christian Kiesling (H1)
Aziz Dossanov (H1)
Daniel Britzger (H1, DESY, Hamburg U.)

in close collaboration with

Halina Abramowicz (MPP ext. scientific member, ZEUS)
Aharon Levy (Tel Aviv U., ZEUS spokesperson)
Stas Shushkevich (DESY)
Roman Kogler (Hamburg U.)





Precise measurements (best accuracy ±1%) of inclusive e+p & e-p NC & CC polarised & unpolarised cross sections at HERA

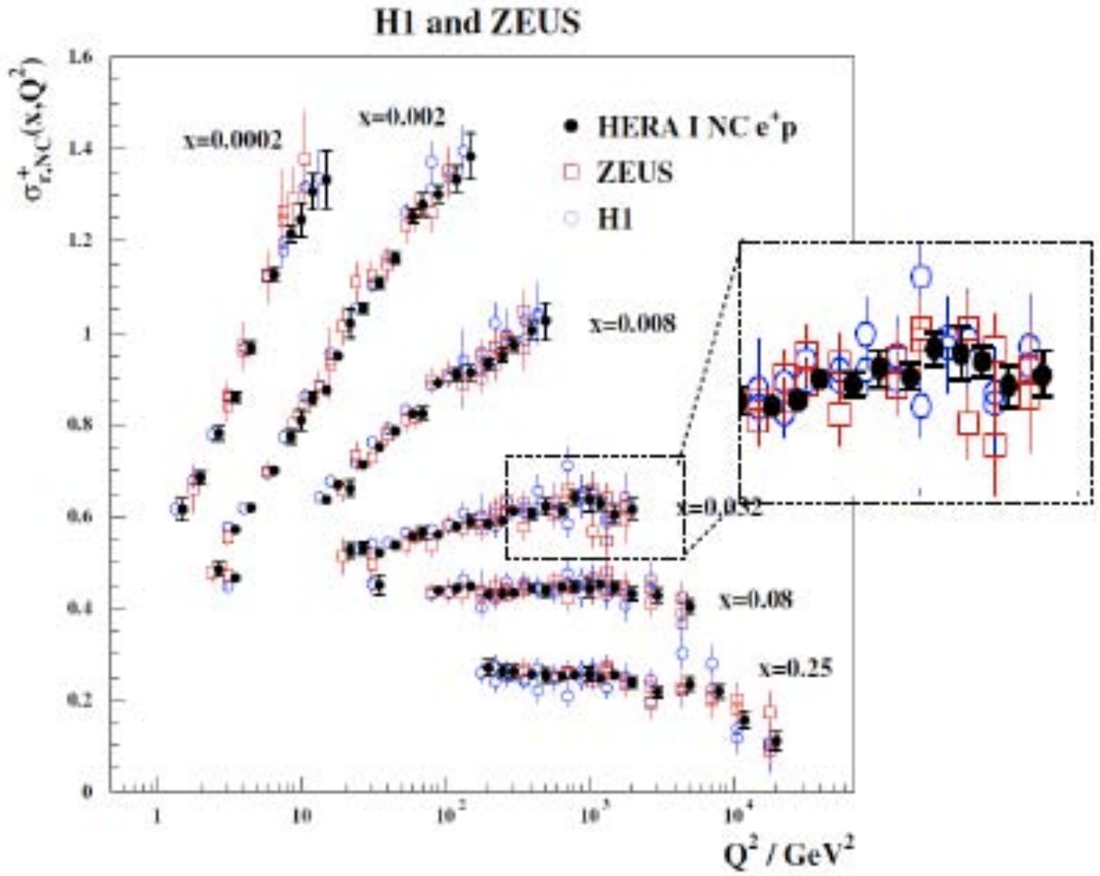
→ HERA data are an indispensable input to modern QCD PDF fits

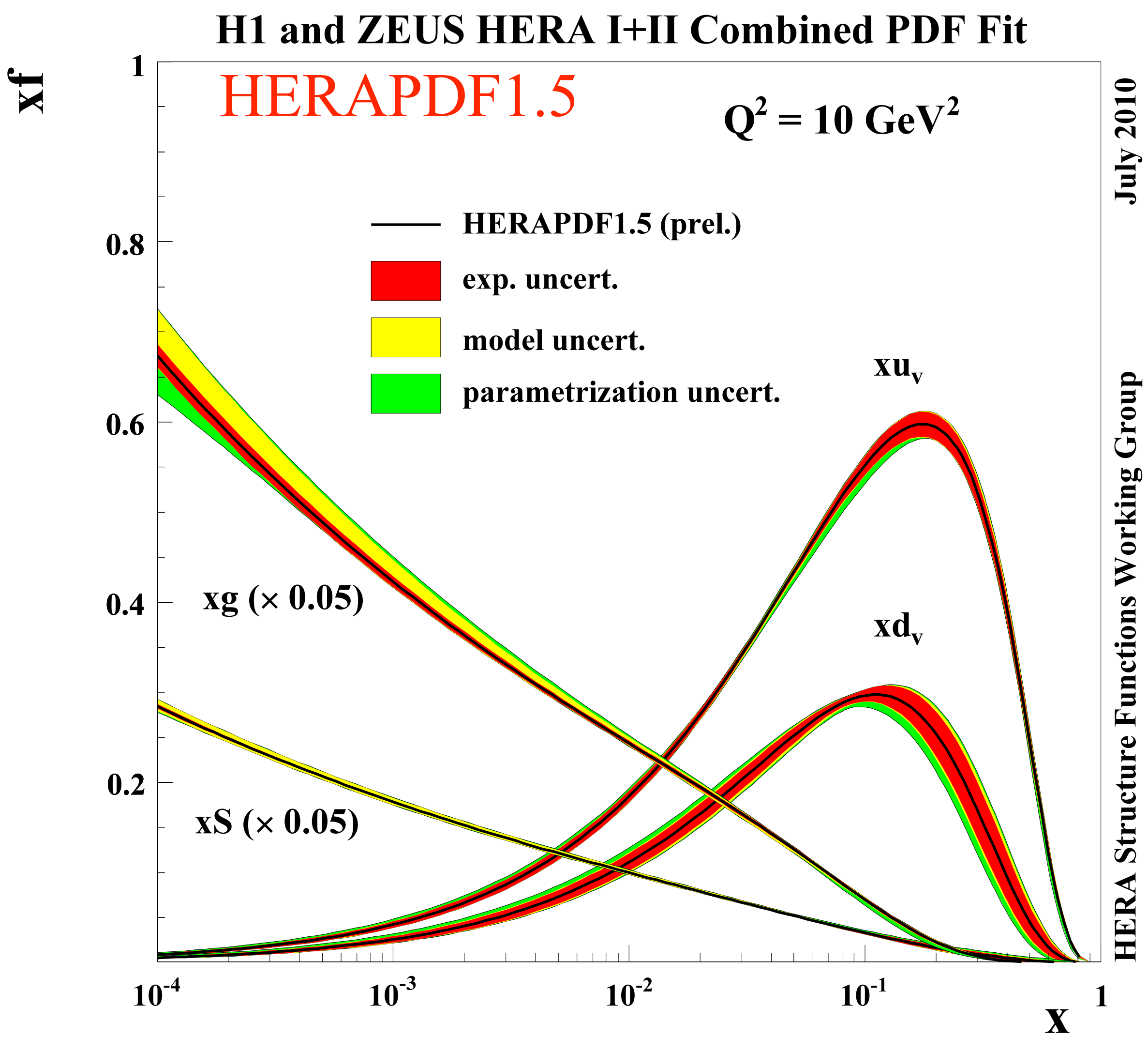
Combination of H1 & ZEUS incl. NC and CC ep data

→ includes expert knowledge in the treatment of the correlations between many individual data sets.
→ precise, complete and easy in use
→ reduction of statistical and systematic uncertainties

1. HERA I data: JHEP 1001:109,2010 HERAPDF 1.0
2. HERA I and preliminary HERA II data HERAPDF 1.5

→ to combine HERA I+II final data and get HERAPDF 2.0



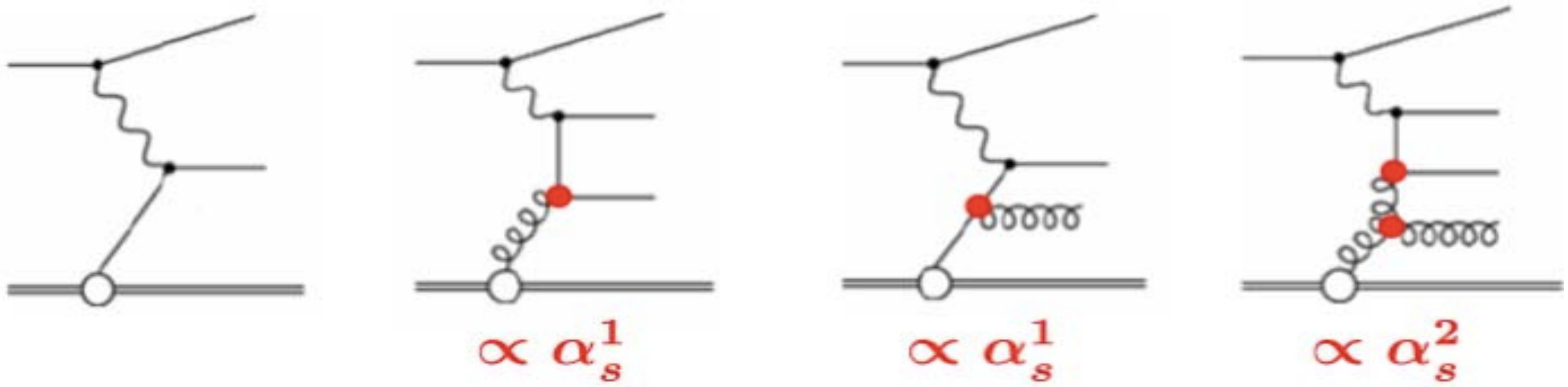


HERAPDF: QCD Fits using HERA data only

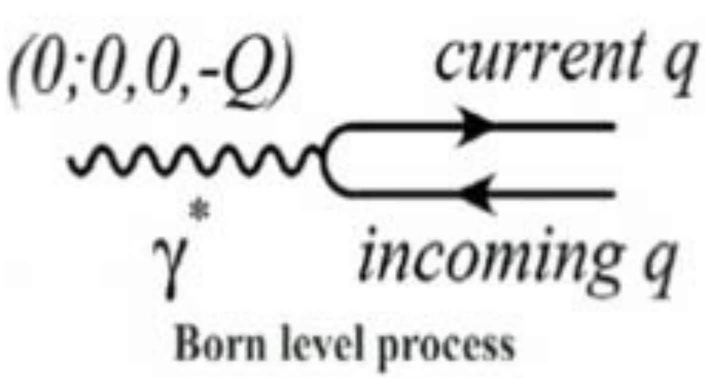
- no nuclear corrections
- no heavy target correction
- $\Delta\chi^2$ = 1 criterion for exp. errors
- parametrise $xg(x)$, xu_v, xd_v. xUbar, Dbar at starting scale Q_o^2
- apply quark number and momentum sum rules
- NLO/NNLO DGLAP evolution
- different schemes for heavy flavor treatment
- uncertainty bands:
  - experimental
  - model (variations of Q^2_{min}, f_s, m_c, m_b)
  - parameterisation (variation of param. assumptions)

Jet production in DIS

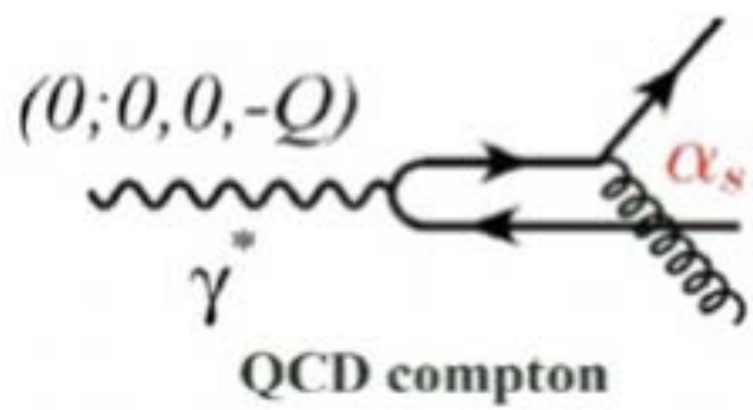
DIS Incl. Jets, dijets Incl. Jets, trijets



in Breit frame only QCD processes with $P_{T,jet} > 0$ are measured

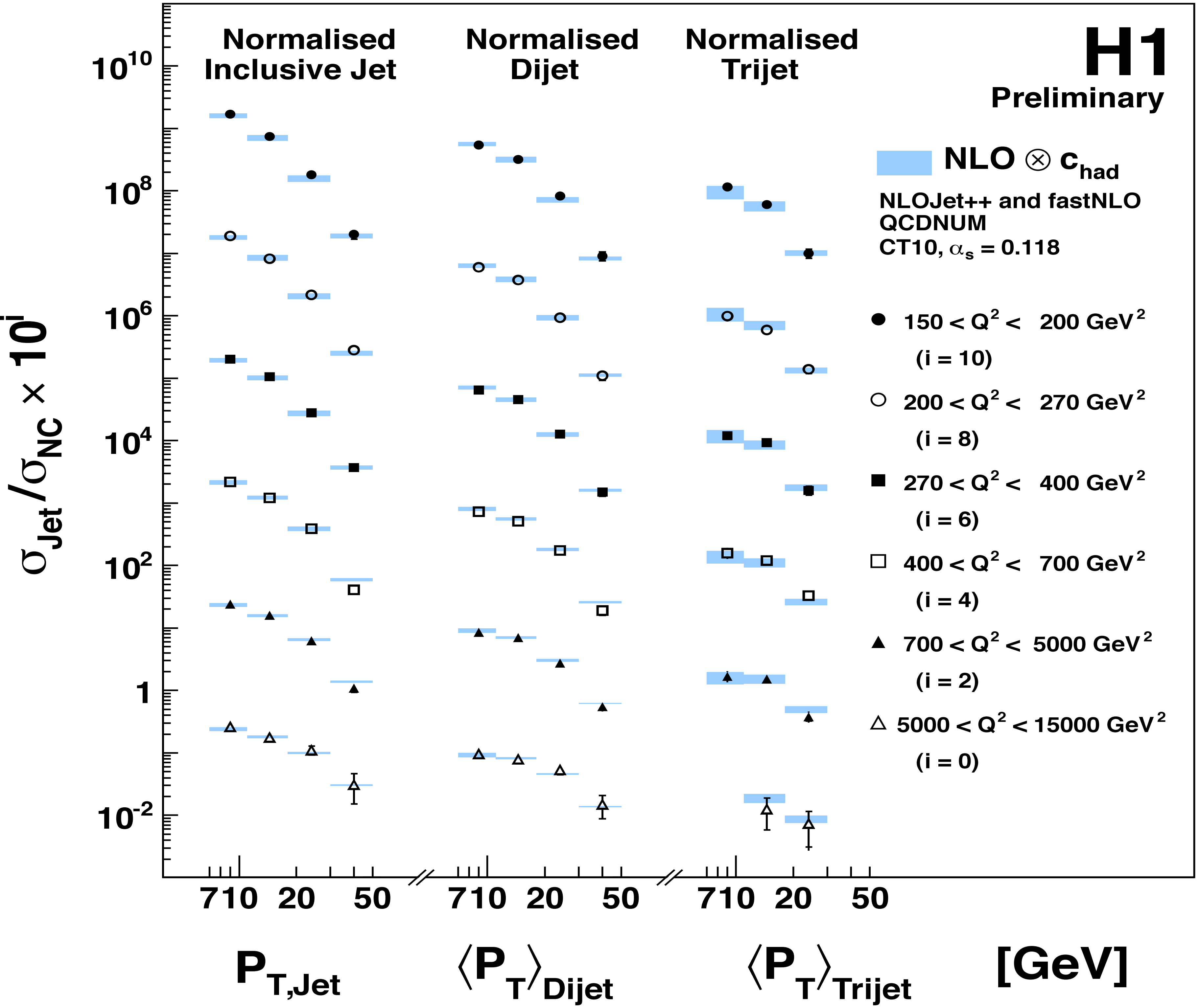


for example:



Direct sensitivity to α_s and the gluon pdf

Normalized multi-jet cross sections



H1-Prelim-12-031

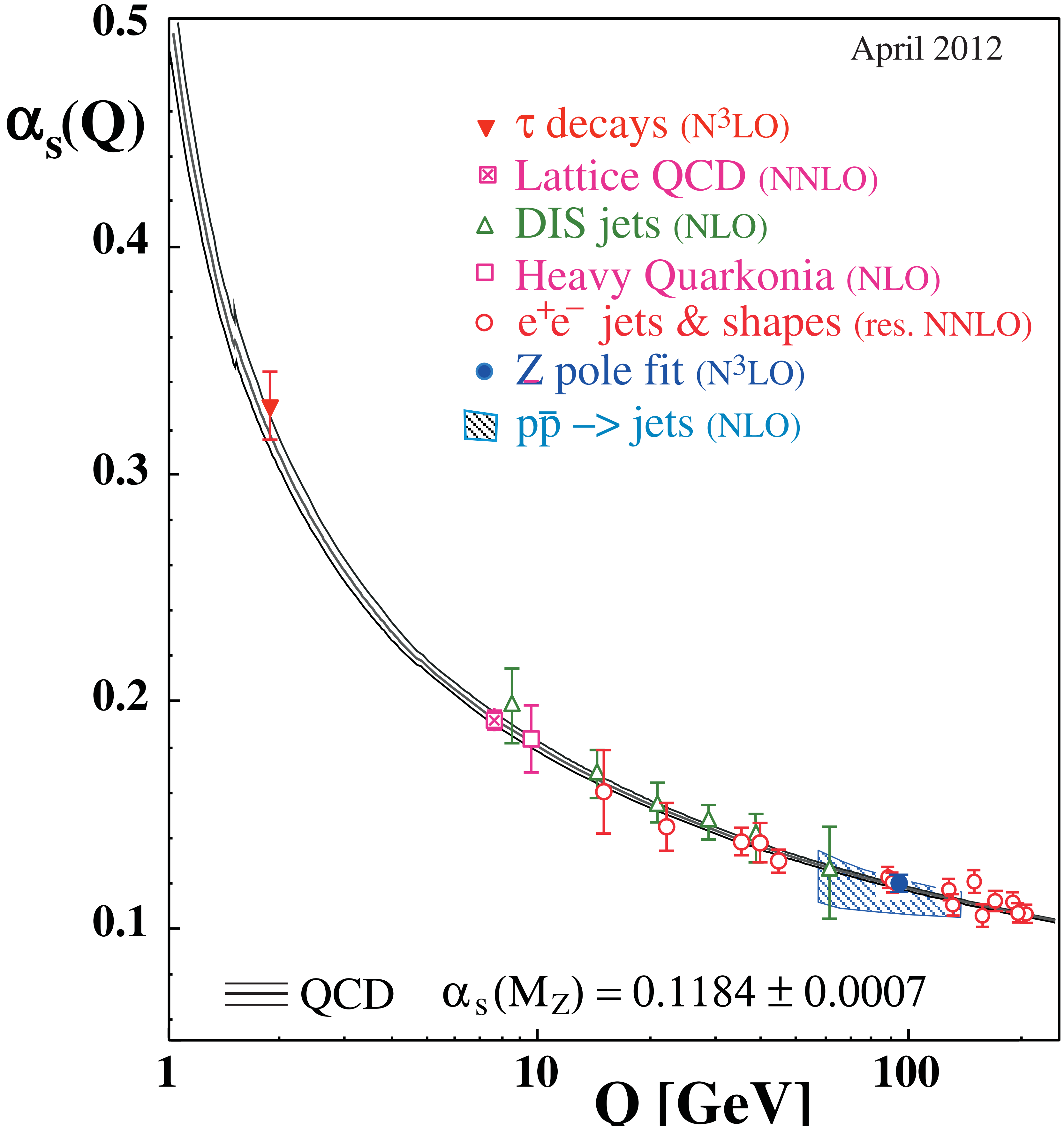
NLO calculation: NLOJet++ and fastNLO, corrected for hadronization effects, and QCDNUM

Scale choice:

$$\mu_r = \sqrt{(Q^2 + P_T^2)/2}$$
$$\mu_f = Q$$

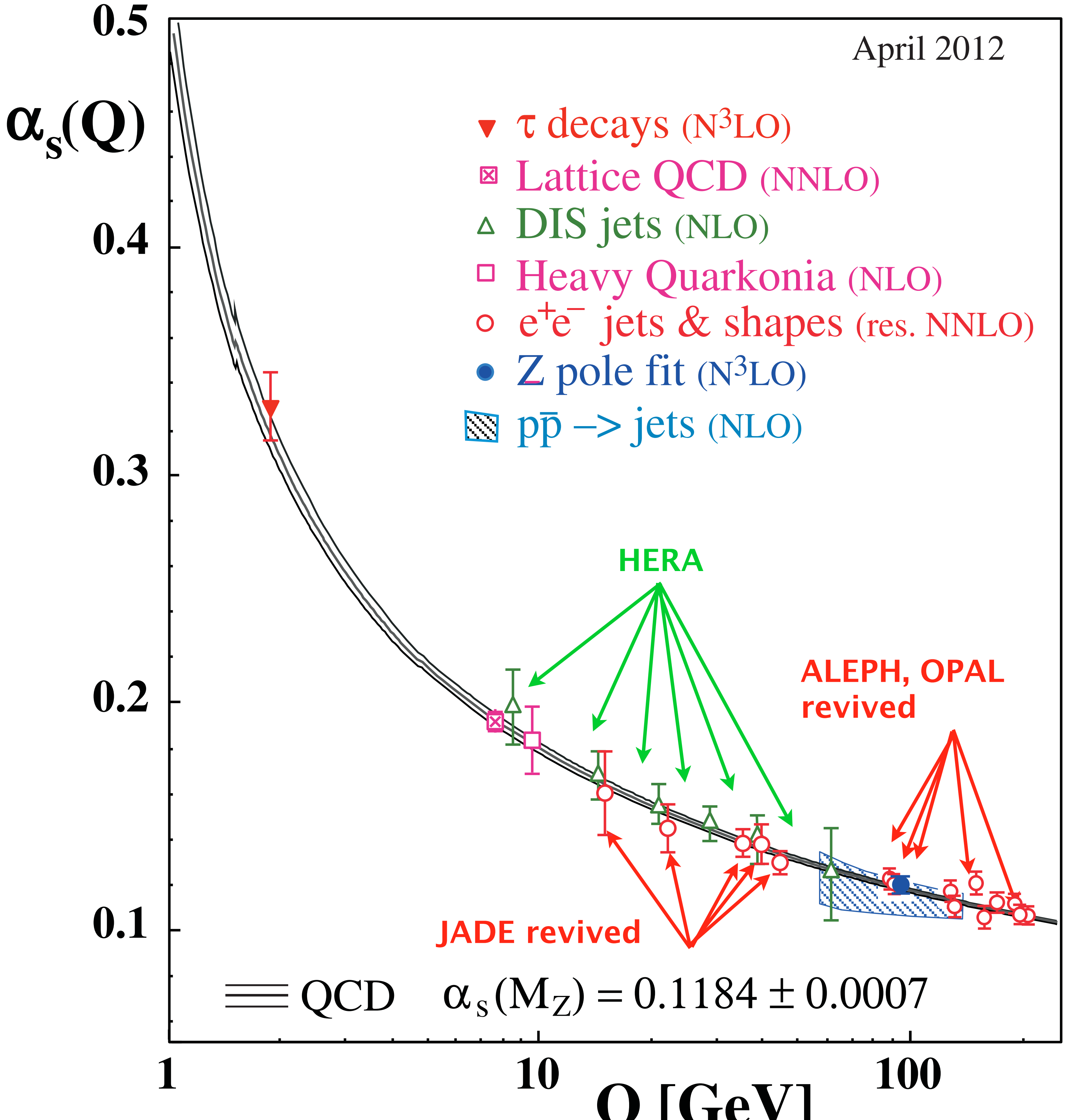
- Next steps:
- Extraction of αS(MZ)
- Include jet cross sections in determination of PDFs & from NC and CC and jet data from H1 and ZEUS.

World Summary of α_s 2011:



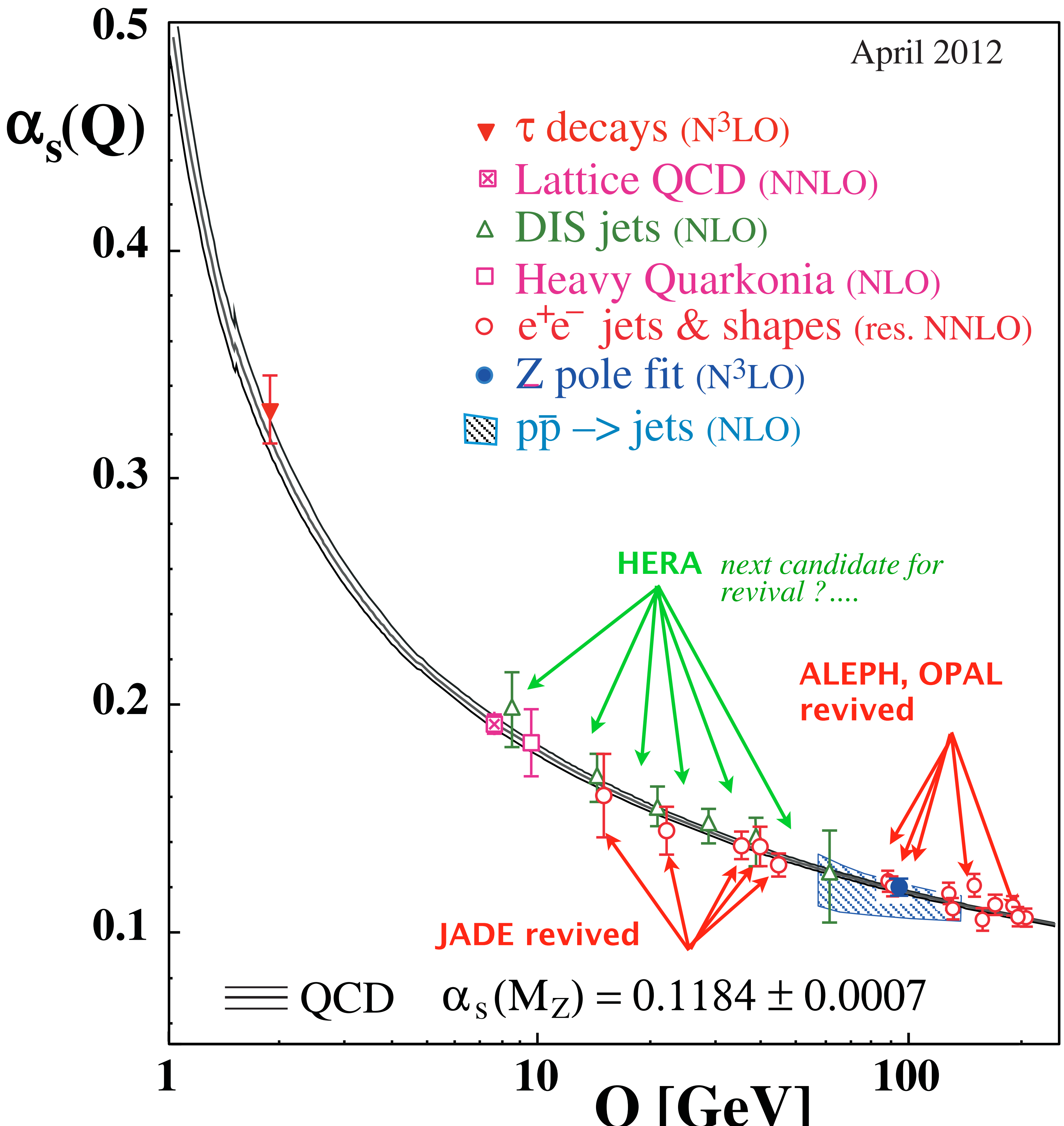
S.B. in: PDG 2012
J. Beringer *et al*.
(Particle Data Group),
Phys. Rev. D**86**, 010001 (2012)

World Summary of α_s 2011:



S.B. in: PDG 2012
J. Beringer *et al*.
(Particle Data Group),
Phys. Rev. D**86**, 010001 (2012)

World Summary of α_s 2011:



S.B. in: PDG 2012
J. Beringer *et al*.
(Particle Data Group),
Phys. Rev. D**86**, 010001 (2012)

necessaty to re-use unique data grows with time

necessaty to re-use unique data grows with time

these data have cost a lot of €/$/£/¥/ !

necessaty to re-use unique data
grows with time

these data have cost
a lot of €/$/£/¥/ !

not preserving them for long-time
and future use
would be a crime

necessaty to re-use unique data

grows with time

these data have cost

a lot of €/$/£/¥/ !

not preserving them for long-time

and future use

would be a crime

„If the same (loss of PETRA data)

will happen with LEP data,

I will sue the CERN DG“

(A well-known theorist after having seen reanalysed JADE results)

Data Preservation in HEP – Status and Problems:

- increased general awareness
- international study group, recommendations
- activities at SLAC, DESY, Fermilab, DESY experiments have started
- LHC collaborations joined DPHEP

Data Preservation in HEP – Status and Problems:

- increased general awareness
- international study group, recommendations
- activities at SLAC, DESY, Fermilab, DESY experiments have started
- LHC collaborations joined DPHEP

- no really coordinated efforts yet
- slow progress in many cases
- big labs need increased efforts (personnel)
- LEP: data & software fading away ...
- LHC: huge amount of challanges ...

Data Preservation in HEP – Status and Problems:

- increased general awareness
- international study group, recommendations
- activities at SLAC, DESY, Fermilab, DESY experiments have started
- LHC collaborations joined DPHEP

- no really coordinated efforts yet
- slow progress in many cases
- big labs need increased efforts (personnel)
- LEP: data & software fading away ...
- LHC: huge amount of challanges ...

We're not yet there … !