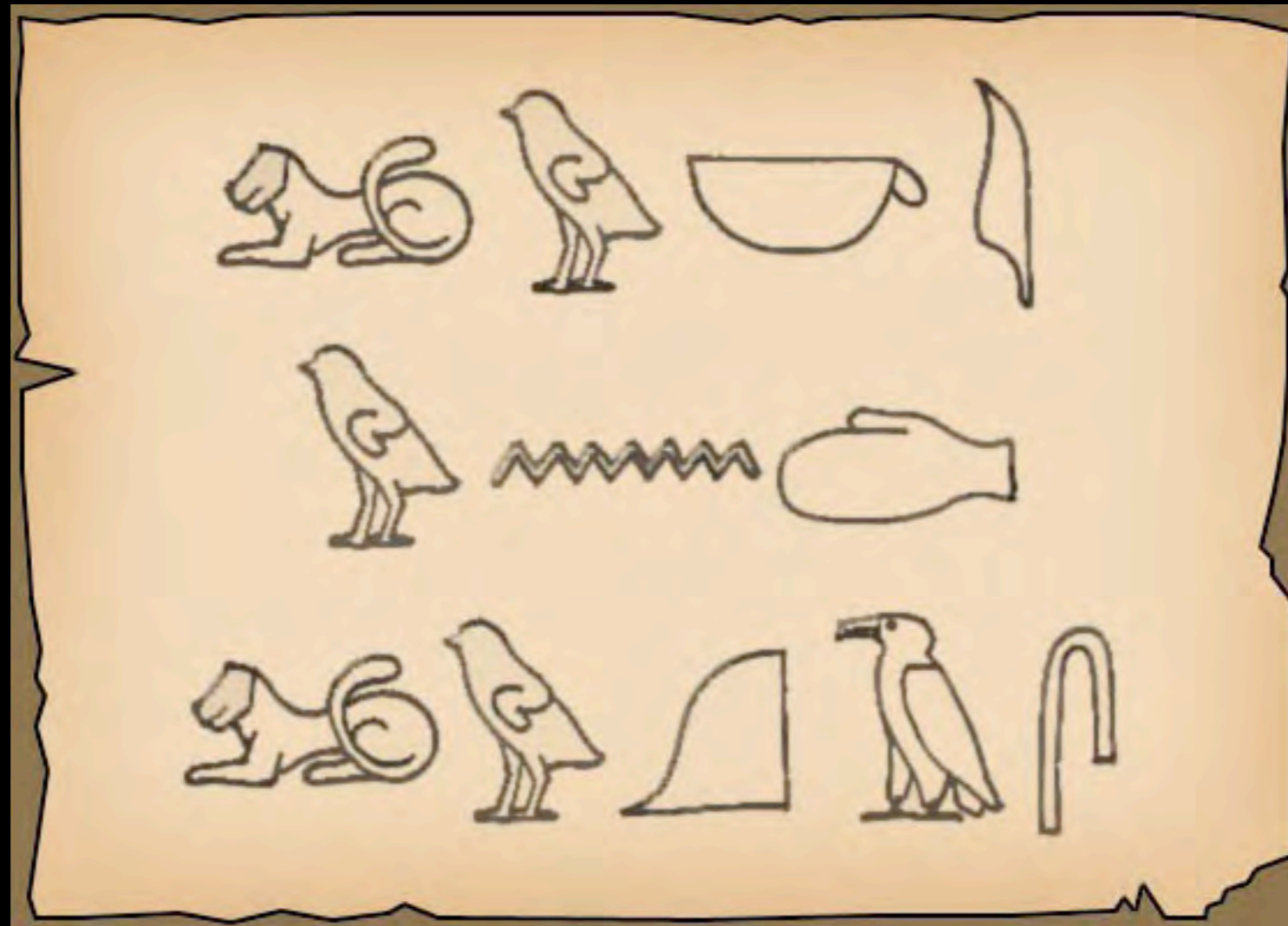


# An Experience with Data Preservation in High Energy Physics



measurements of  $\alpha_s$  over three decades

S. Bethke



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

# Physics Case for Data preservation

- long-term completion and extension of scientific program
- Cross-collaboration analyses
- **Data re-use**
- Education, training and outreach

## An example of successful reanalysis of HEP data:

JADE experiment @ PETRA, 1979 - 1986

resurrection of data & software: 1995 ~ 2003

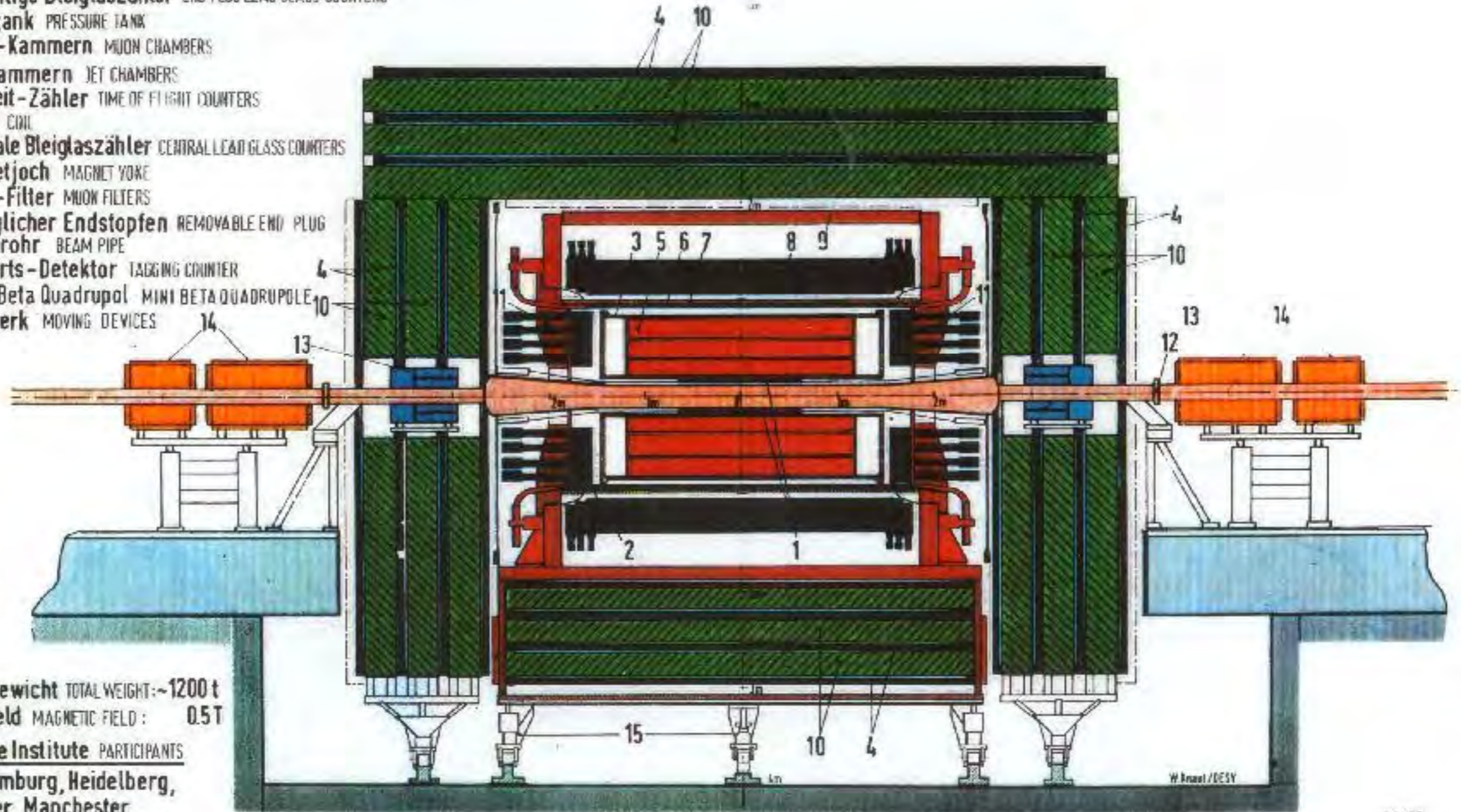


# The JADE Experiment

MAGNETDETEKTOR **JADE**  
MAGNET DETECTOR

at the PETRA  $e^+e^-$  storage ring @ DESY

- 1 Strahlrohrzähler BEAM PIPE COUNTERS
- 2 Endseitige Bleiglaszähler END PLUG LEAD GLASS COUNTERS
- 3 Drucktank PRESSURE TANK
- 4 Myon-Kammern MUON CHAMBERS
- 5 Jet-Kammern JET CHAMBERS
- 6 Flugzeit-Zähler TIME OF FLIGHT COUNTERS
- 7 Spule COIL
- 8 Zentrale Bleiglaszähler CENTRAL LEAD GLASS COUNTERS
- 9 Magnetjoch MAGNET YOKE
- 10 Myon-Filter MUON FILTERS
- 11 Beweglicher Endstopfen REMOVABLE END PLUG
- 12 Strahlrohr BEAM PIPE
- 13 Vorwärts-Detektor TAGGING COUNTER
- 14 Mini-Beta Quadrupol MINI BETA QUADRUPOLE
- 15 Fahrwerk MOVING DEVICES



Gesamtgewicht TOTAL WEIGHT: ~1200 t

Magnetfeld MAGNETIC FIELD: 0.5 T

Beteiligte Institute PARTICIPANTS

DESY, Hamburg, Heidelberg,  
Lancaster, Manchester,  
Rutherford Lab., Tokio

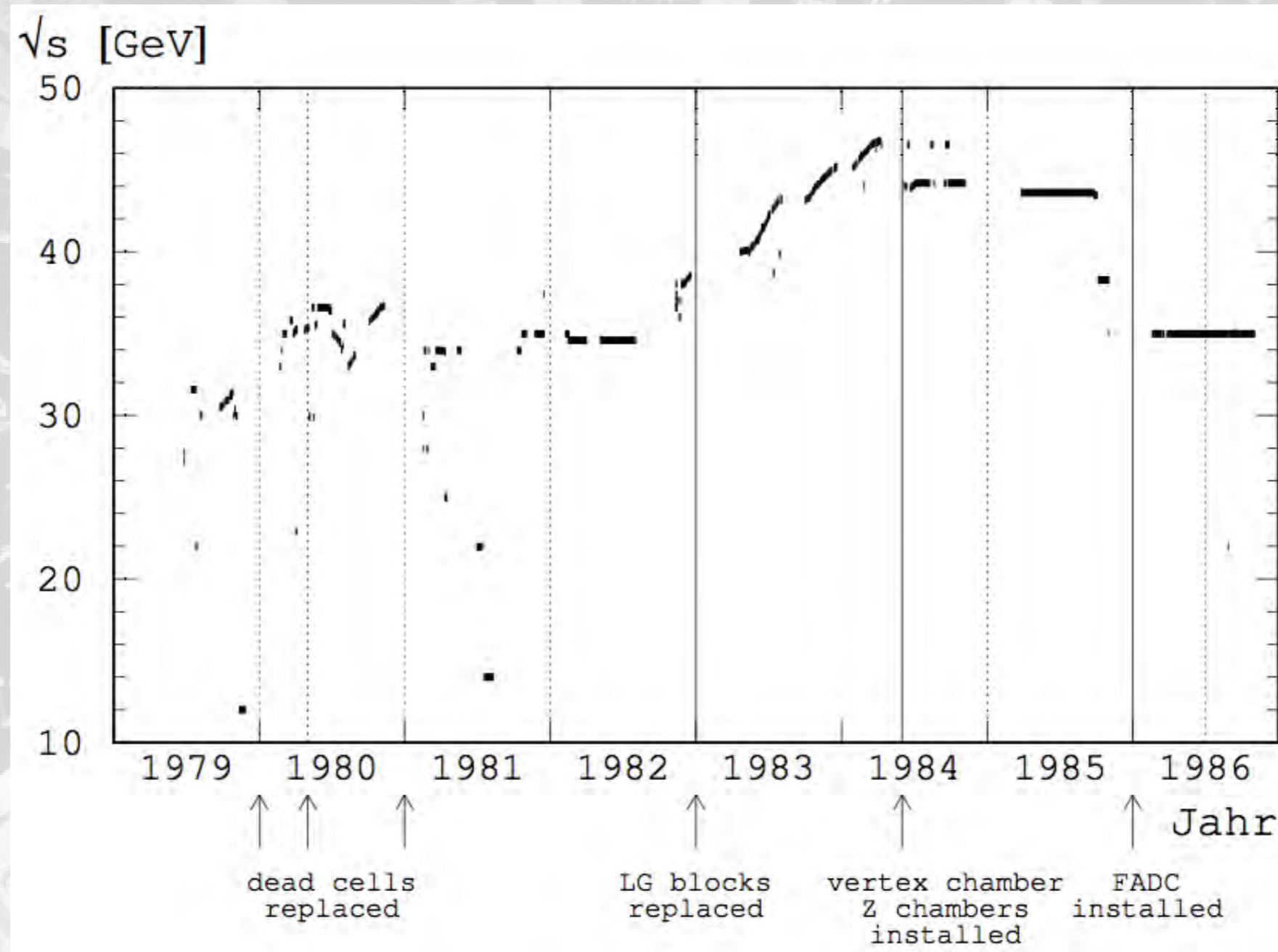
33188

operation time: 1978 - 1986

operation mode:  $e^+e^-$  annihilation;  $E_{cm} \sim 14 \dots 46 \text{ GeV}$

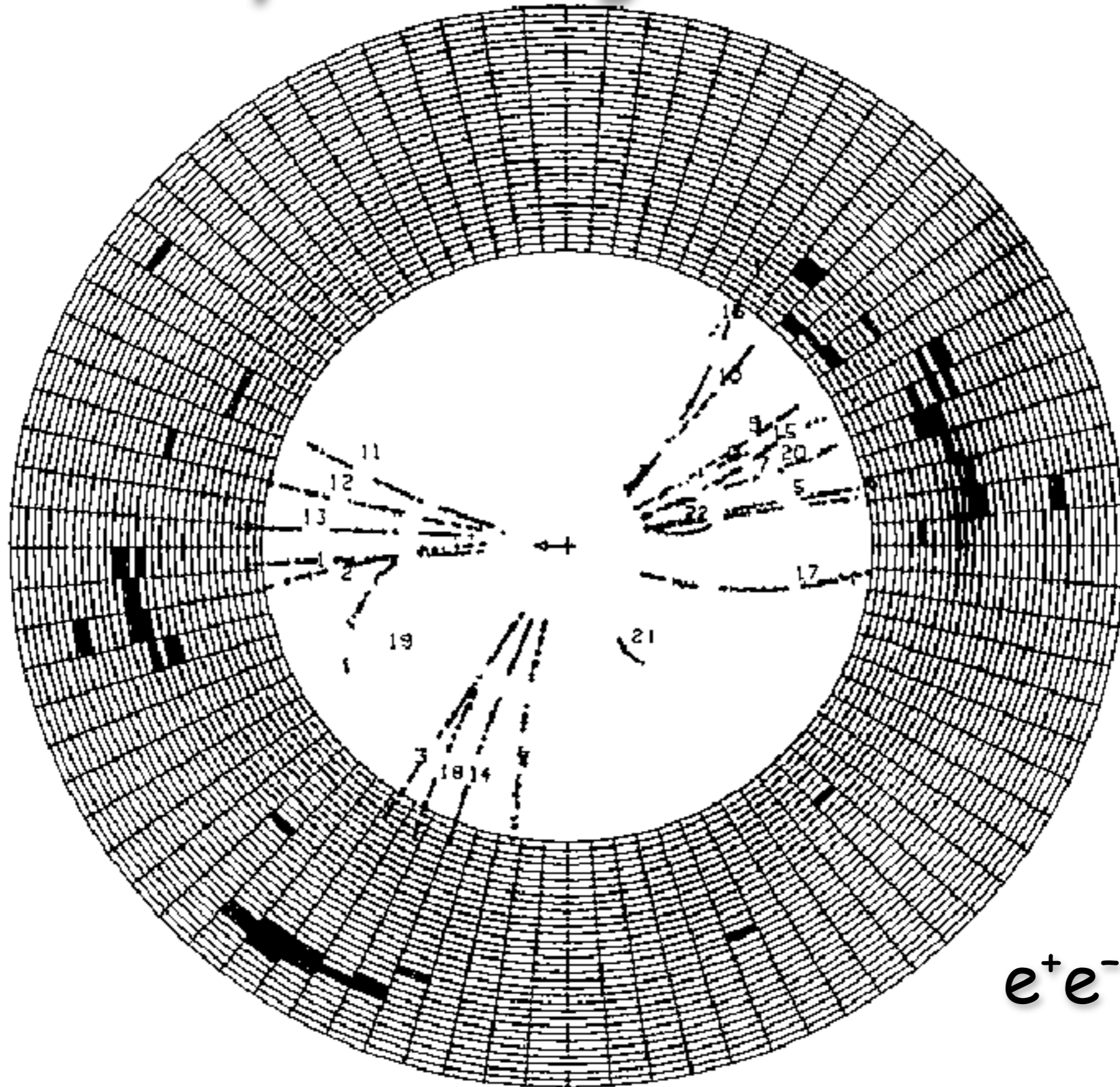


# JADE data taking



$\sim 200 \text{ pb}^{-1}$  ;  $\sim 45.000$  „good“ multihadronic events

# discovery of the gluon at PETRA



$$e^+e^- \rightarrow q\bar{q}g$$



# physics benefits: new results from old data

	now (after LEP)	then (PETRA)
• new and improved theoretical calculations	NNLO QCD	(N)LO QCD
• new and improved MC models	NLLA+NLO shower	(N)LO fixed order
• new and optimised observables	$B_w$ , $B_t$ , $D_3$ , Durham jets	event shapes: T, S, O, ...
• more complete knowledge of Standard Model	top, W, Z	-----

-> re-do previous measurements:

- increased precision
- reduced systematics

-> perform new measurements:

- at Energies and processes where no other data are available today (and in future)

-> if new phenomena found today:

- go back and check at lower E

# Revitalisation of JADE software

- started in 1995 at RWTH Aachen (P. Movilla-Fernandez; Diploma- and PhD thesis), proceeded until ~2003 at MPP.
  - conversion, translation, partly rewriting of Fortran-IV, Mortran, Sheltran, assembler routines
  - complete installation on IBM RS6000 AIX platform using xlf compiler
  - successfully revitalised and validated entire JADE core software:
    - reconstruction software
    - simulation software
    - event display and JADE graphics package (now in colour !)
- generation of full-simulation MC events, using modern MC generators plus the experience from LEP



DSN mc35

BEAM 17.500 GEV FIELD -4.849 KG TALC 0039 DATE 27/11/99 TIME 01.37.20  
T1A 0802 T1P 4101 T2C 0089 CAMAC TIME 1. 1. 1 17/ 5/1985

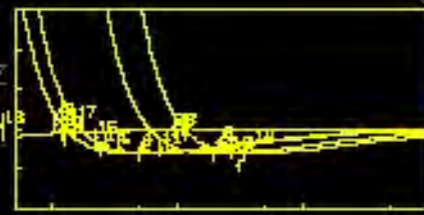
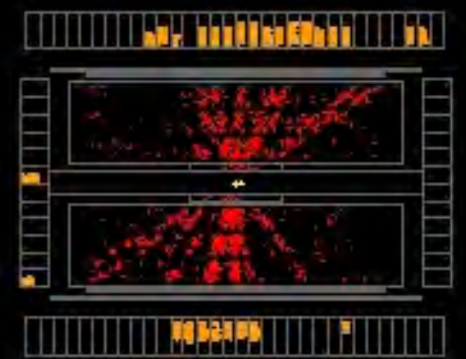
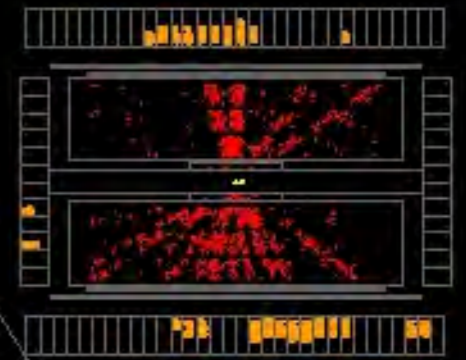
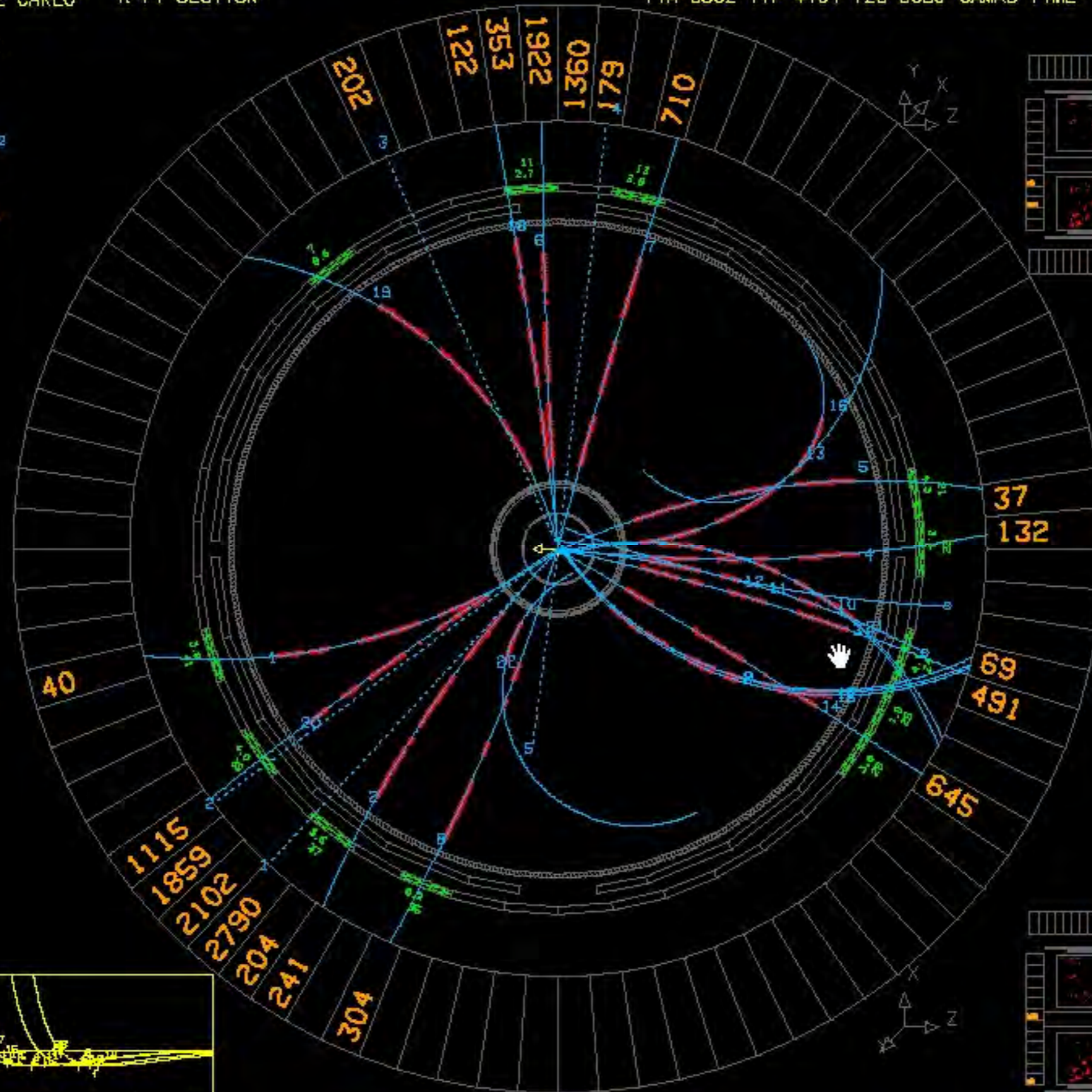
0 B10 B10  
IDHITS 053  
ELGTOT 15047  
MUHITS 0  
LGCYL 14897  
LGCAPS 150 0  
FWDAPS 0 0

MONTE CARLO R-FI SECTION

JADE

BANK PATR 10 NR OF TRACKS 22

BANK LGCL 1 NR OF CLUSTERS 14



\*\*\* SUMS (GEV) \*\*\* PTOT 20.037 PTRANS 18.012 PLONG 8.488 CHARGE -2  
 TOTAL CLUSTER ENERGY 16.017 PHOTON ENERGY 8.152 NR OF PHOTONS 5



# 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 Siggj Bethke.)*

CERN Courier, May 2009

# Revitalisation of JADE data

- primarily used original ZE4V format data files for new analyses, plus newly generated MC data converted to ZE4V format
- task to convert FPACK generated copies of raw data files back to readable BOS files accomplished in 2005 and 2008 (J. Olsson)
- data (600 GB) now reside on file servers at MPG RZ Garching

## continued analyses of OPAL (LEP) data

- use of NTUPLES (data and MC) generated for QCD studies during running time of OPAL
- JADE data (and new „LEP-like“ MC) converted to same format of NTUPLES → use of identical analysis software for JADE and OPAL data



# Some anecdotes along the line ....

- one important „calibration“ file, containing the recorded luminosities of each run and fill, was stored on a private account and therefore lost when DESY archive was cleaned up.

*Jan Olsson, when cleaning up his office in ~1997, found an old ASCII-printout of the JADE luminosity file. Unfortunately, it was printed on green recycling paper - not suitable for scanning and OCR-ing. A secretary at Aachen re-typed it within 4 weeks. A checksum routine found (and recovered) only 4 typos.*

- an old version of the original BOSlib 1979 version was found, on our request, at the Univ. of Tokyo computer centre.
- Peter Bock, when cleaning out an old lab at the Physics Institute at Heidelberg University, found a few 9-track tapes containing original JADE MC files which were very valuable for validating results of our first re-analyses in ~1997

# Universality of QCD and Hadronisation

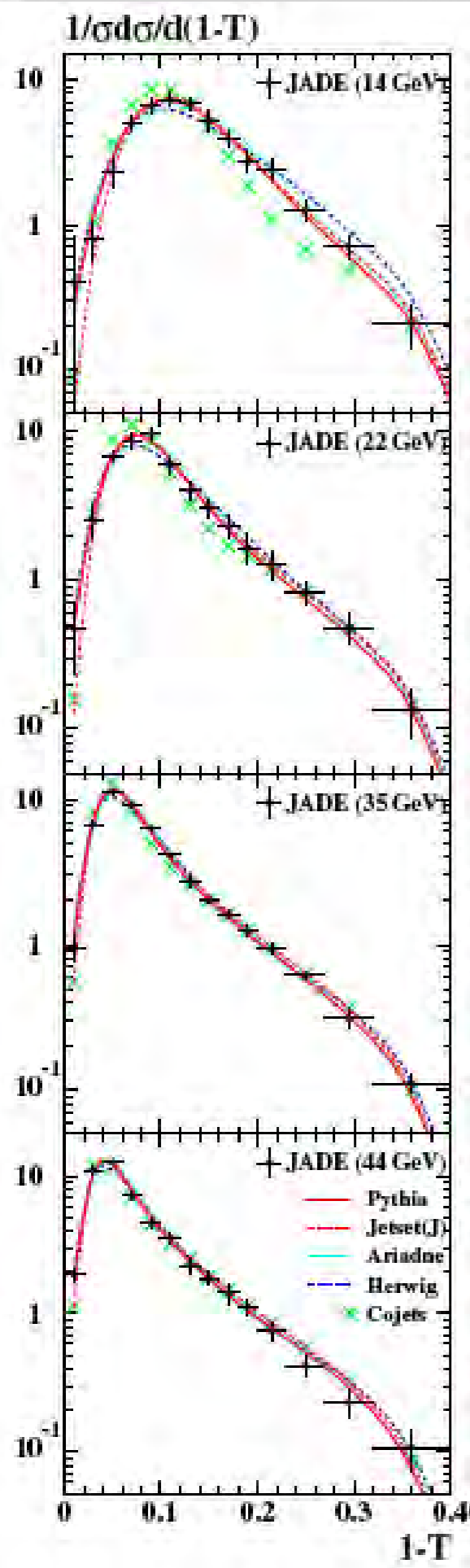
- at Petra times: not possible to describe data at all energies using QCD MC models with one consistent set of parameters

-> lowest energy data (14 GeV, 22 GeV) hardly used

- (new) QCD models with parameters tuned at LEP describe PETRA data down to lowest energies

-> confirms QCD concept of running coupling (Asymptotic Freedom) plus universality of hadronisation process

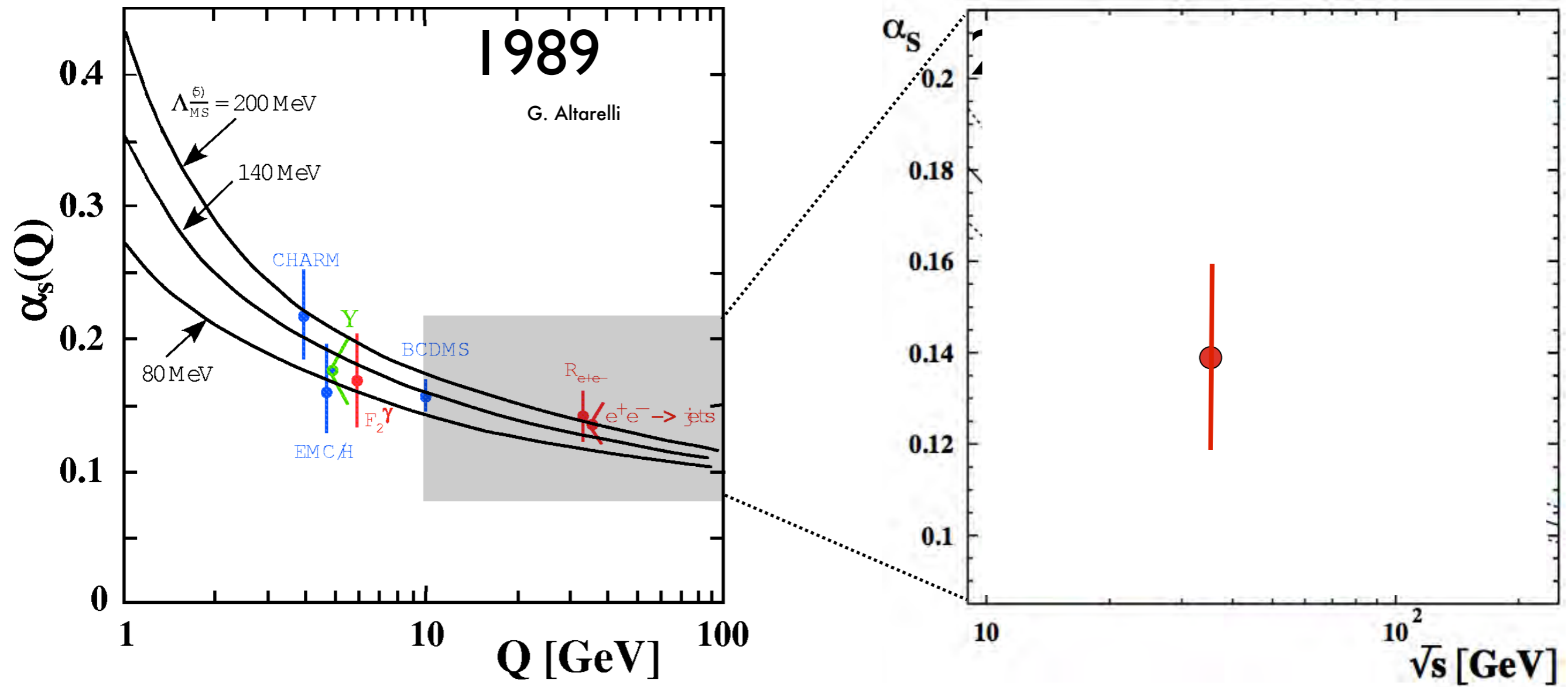
-> now allows to use lowest energy data for precision measurements, e.g. of  $\alpha_s$ .



Eur.Phys.J.C1:461-478,1998;  
hep-ex/0305023



new results from old e+e- data (JADE at PETRA):  
precision  $\alpha_s$  and proof of Asymptotic Freedom



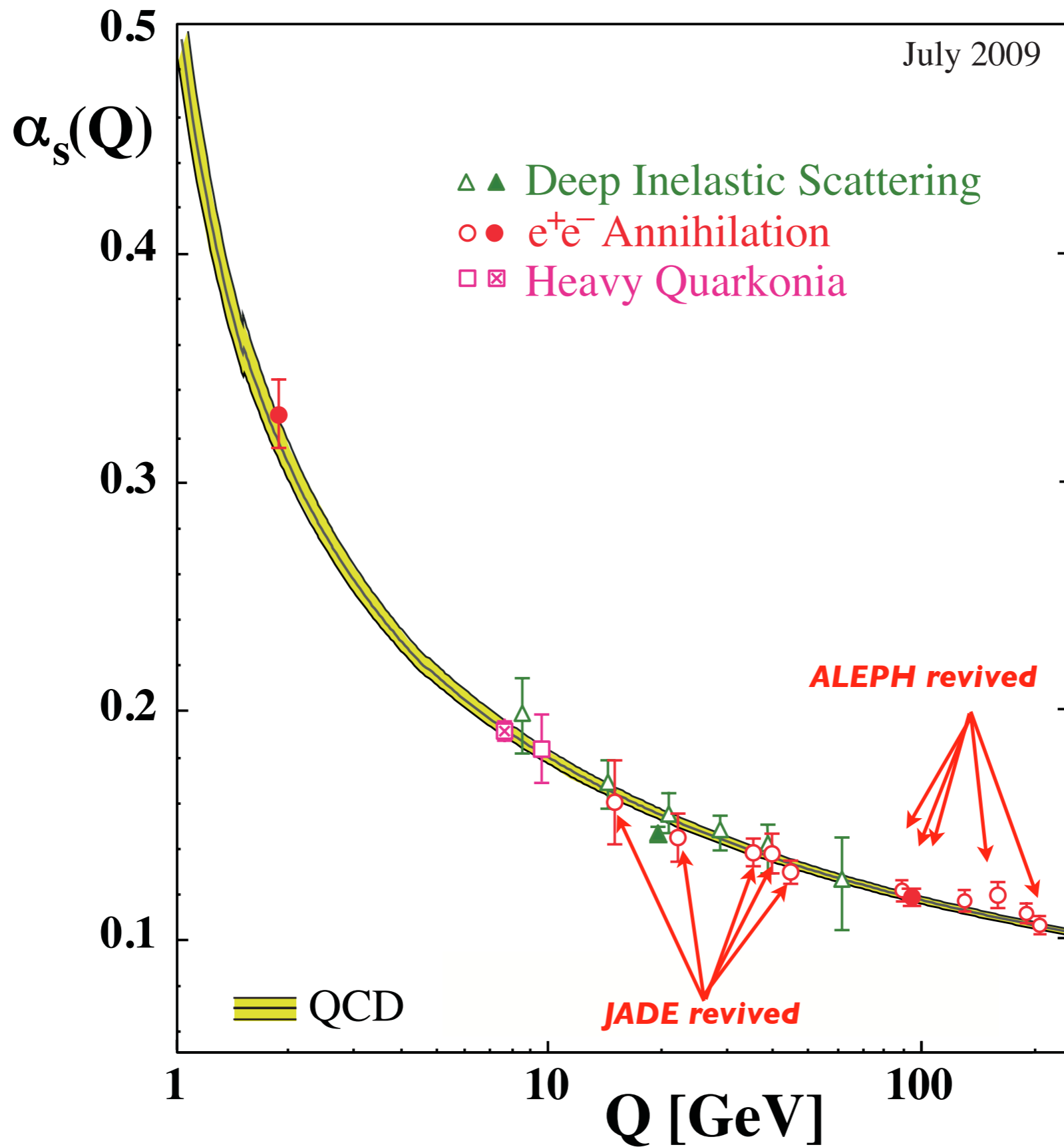
$\rightarrow \alpha_s(M_Z) = 0.1172 \pm 0.0020_{\text{exp.}} \pm 0.0046_{\text{th.}}$

(Bethke et al., arXiv:0810.1389  
 Eur.Phys.J.C64:351-360,2009)

- in NLO QCD:  $\alpha_s(35 \text{ GeV}) = 0.14 \pm 0.02$
- no running  $\alpha_s$  signature

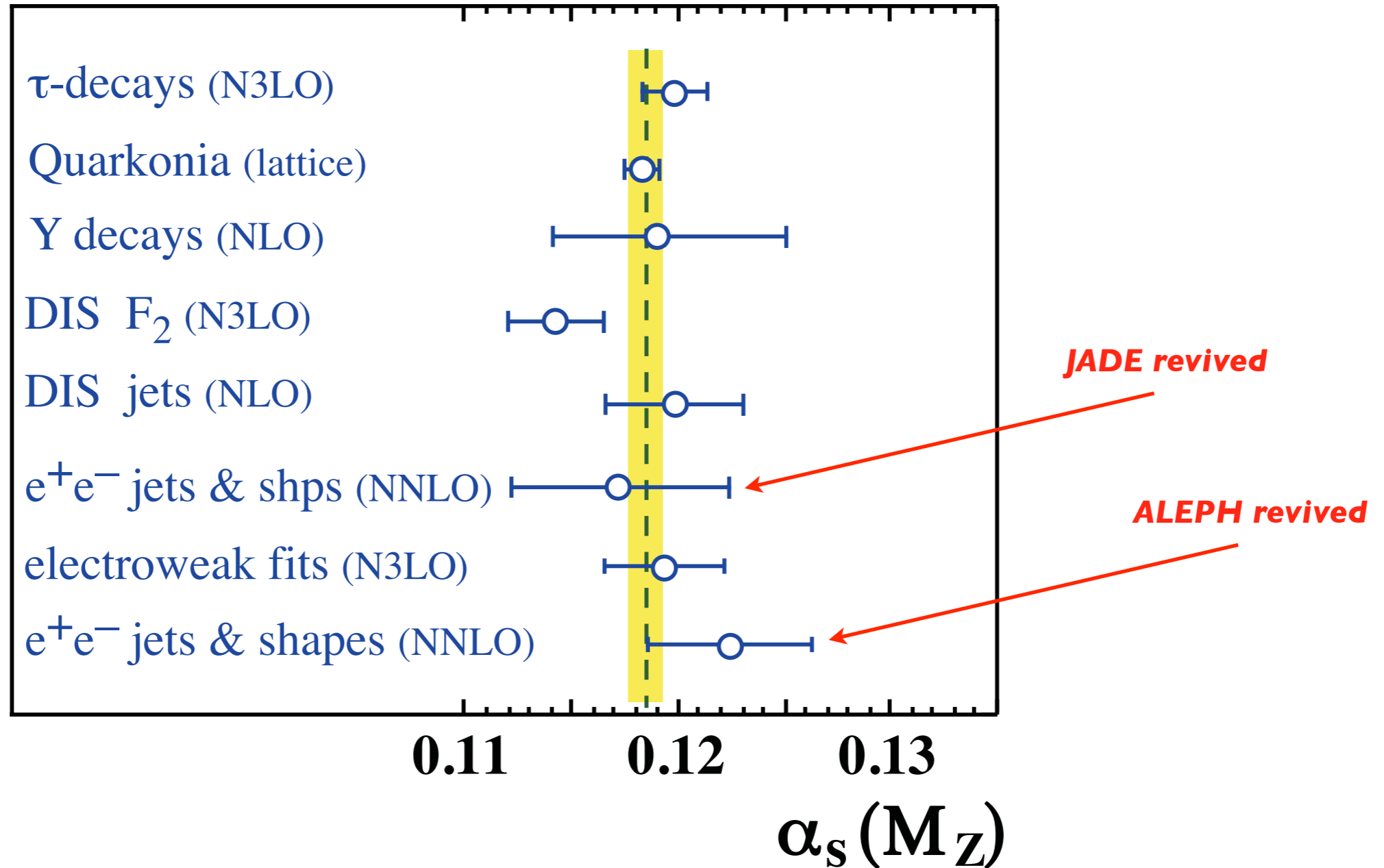
- significant proof of running  $\alpha_s$  and asymptotic freedom from e+e- data alone

# World Summary of $\alpha_s$ 2009:





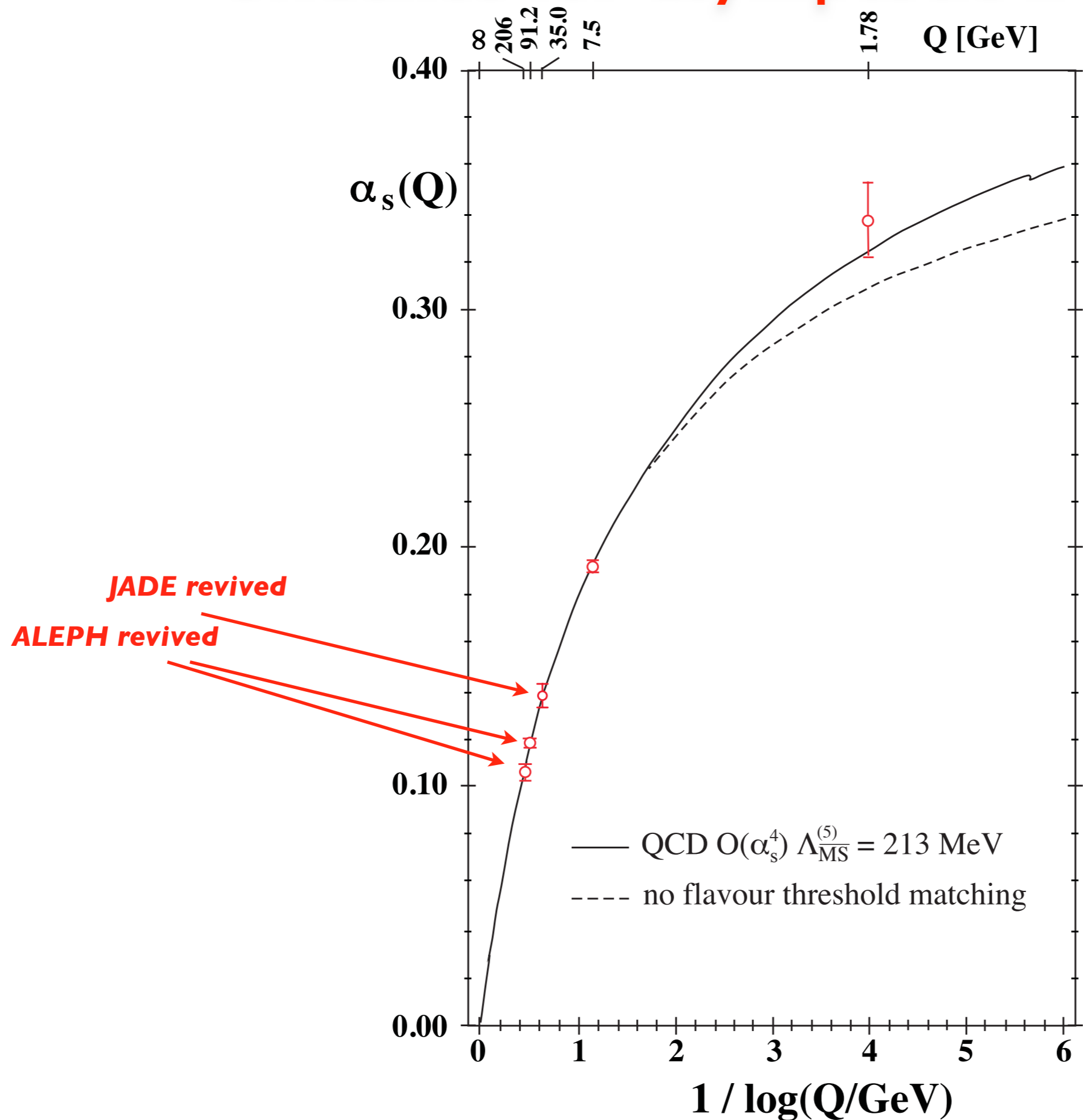
# World Summary of $\alpha_s$ 2009:



$$\rightarrow \alpha_s(M_Z) = 0.1184 \pm 0.0007$$

(Bethke, arXiv:0908.1135  
Eur.Phys.J.C64:689-703,2009)

# evidence for asymptotic freedom:



**exp. HEP data have cost  
a lot of €/\$/£/¥/🍏 !**

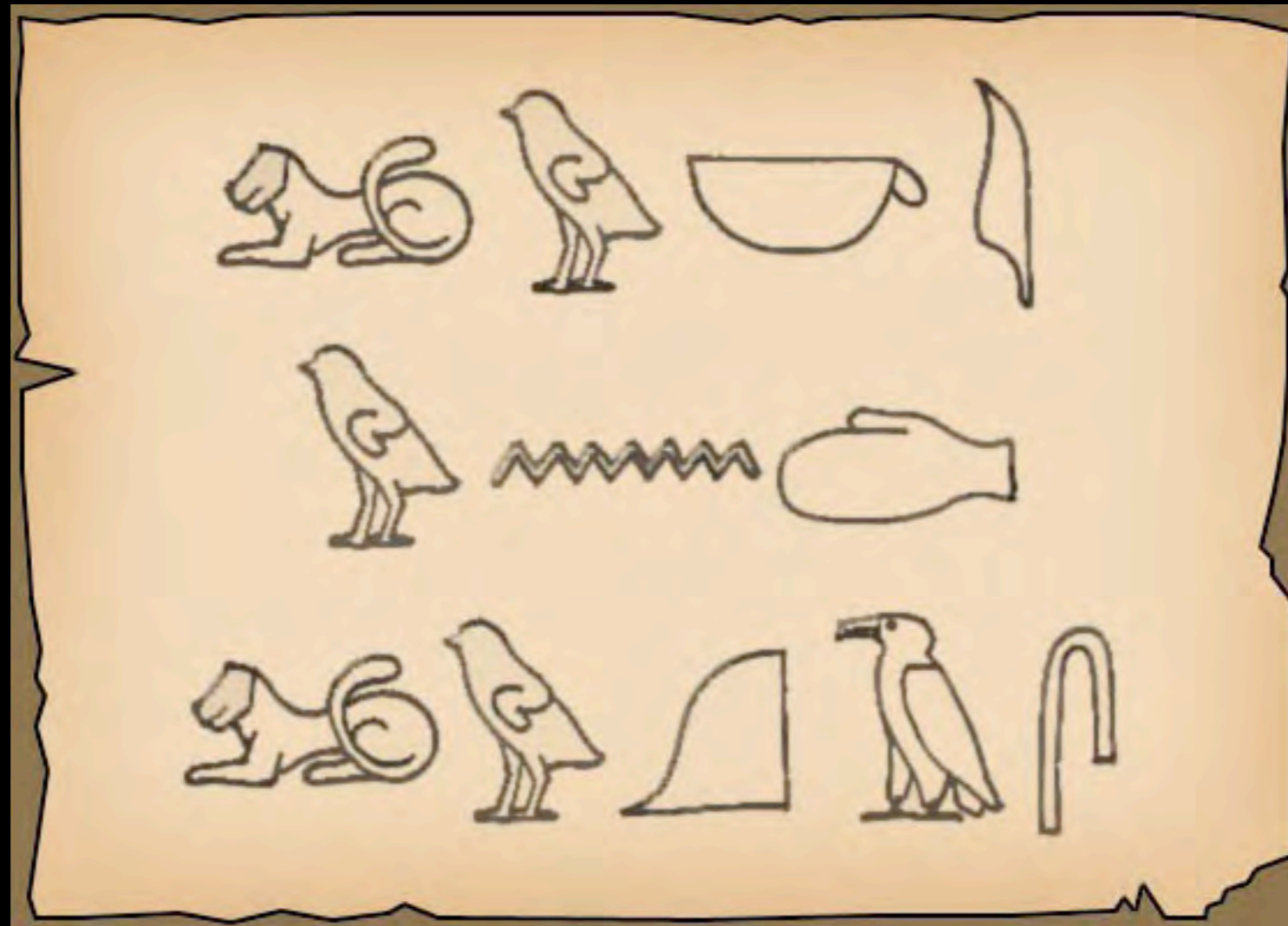
**not preserving them for long-time  
and future use  
would be a crime**

**„If the same (loss of data as at PETRA)  
will happen with LEP data,  
I will sue the CERN DG“**

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



# Data Preservation in High Energy Physics



we just *have* to do it !

