Experience from re-analysis of PETRA (and LEP) data

S. Bethke, O. Biebel[0], P.A. Movilla Fernandez[1], S. Kluth, J. Olsson[2], C. Pahl, J. Schieck, Max-Planck-Institute of Physics, Munich 0):now at LMU 1):now at Fermilab 2):DESY

- the JADE experiment
- physics benefits: new results from old data
- JADE data & software formats
- revitalisation of JADE data & software
- some anecdotes along the line...
- lessons to be learned

The JADE Experiment
at the PETRA e^+e^- storage ring @ DESY

MAGNETDETEKTOR JADE
MAGNET DETECTOR

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
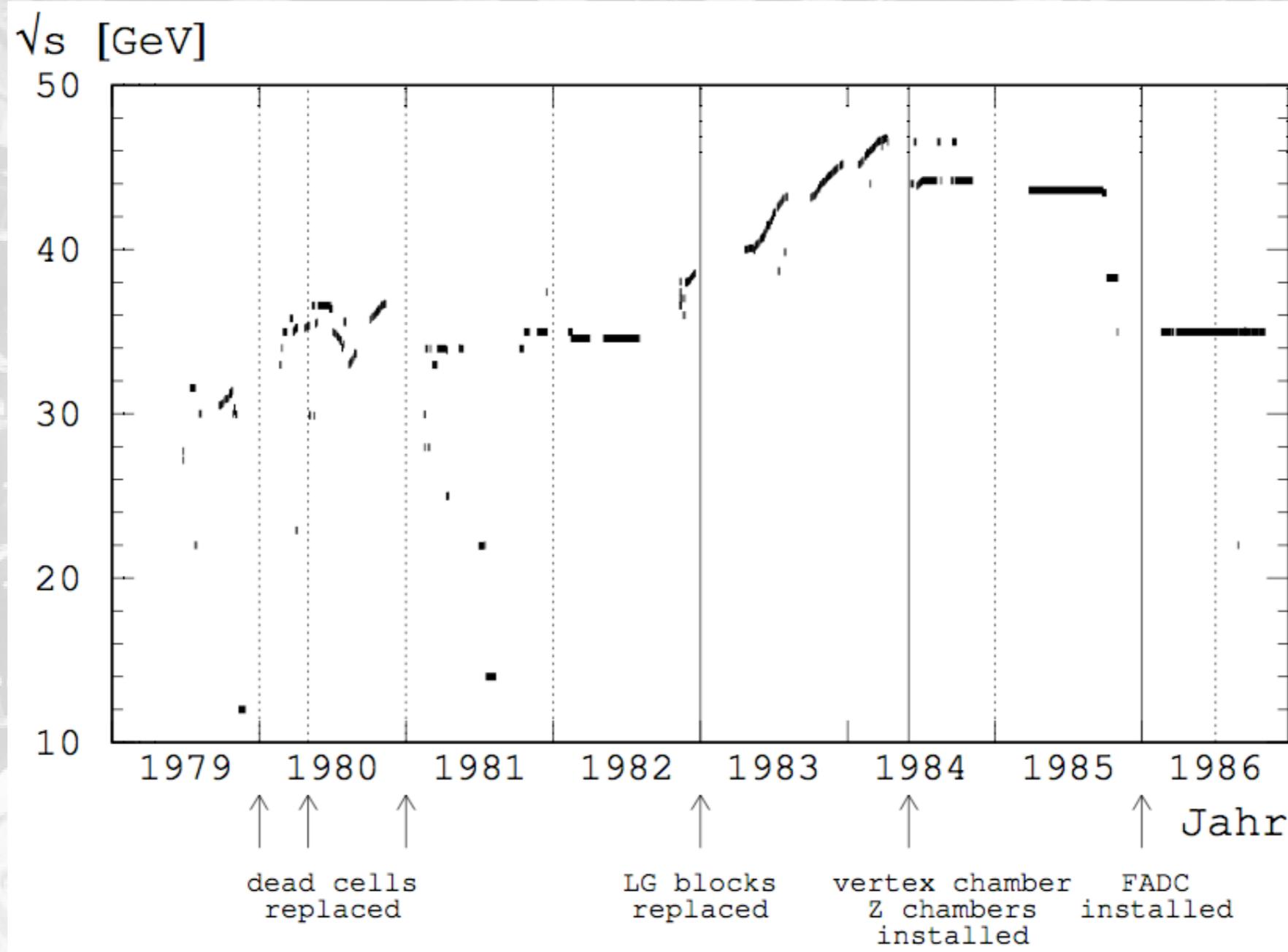15 Fahrwerk MOVING DEVICES
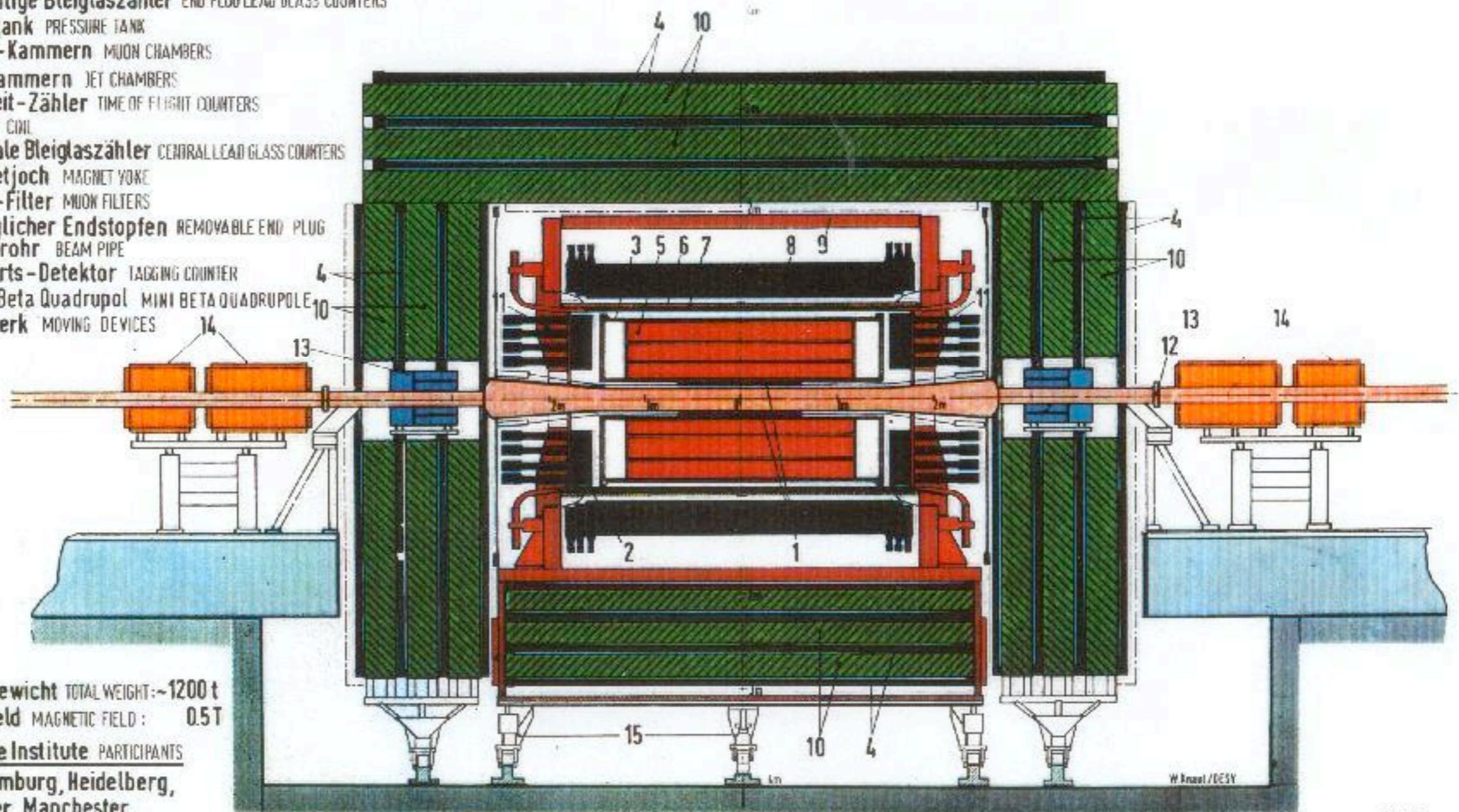
Gesamtgewicht TOTAL WEIGHT: ~1200 t
Magnetfeld MAGNETIC FIELD: 0.5 T

Beteiligte Institute PARTICIPANTS
DESY, Hamburg, Heidelberg,
Lancaster, Manchester,
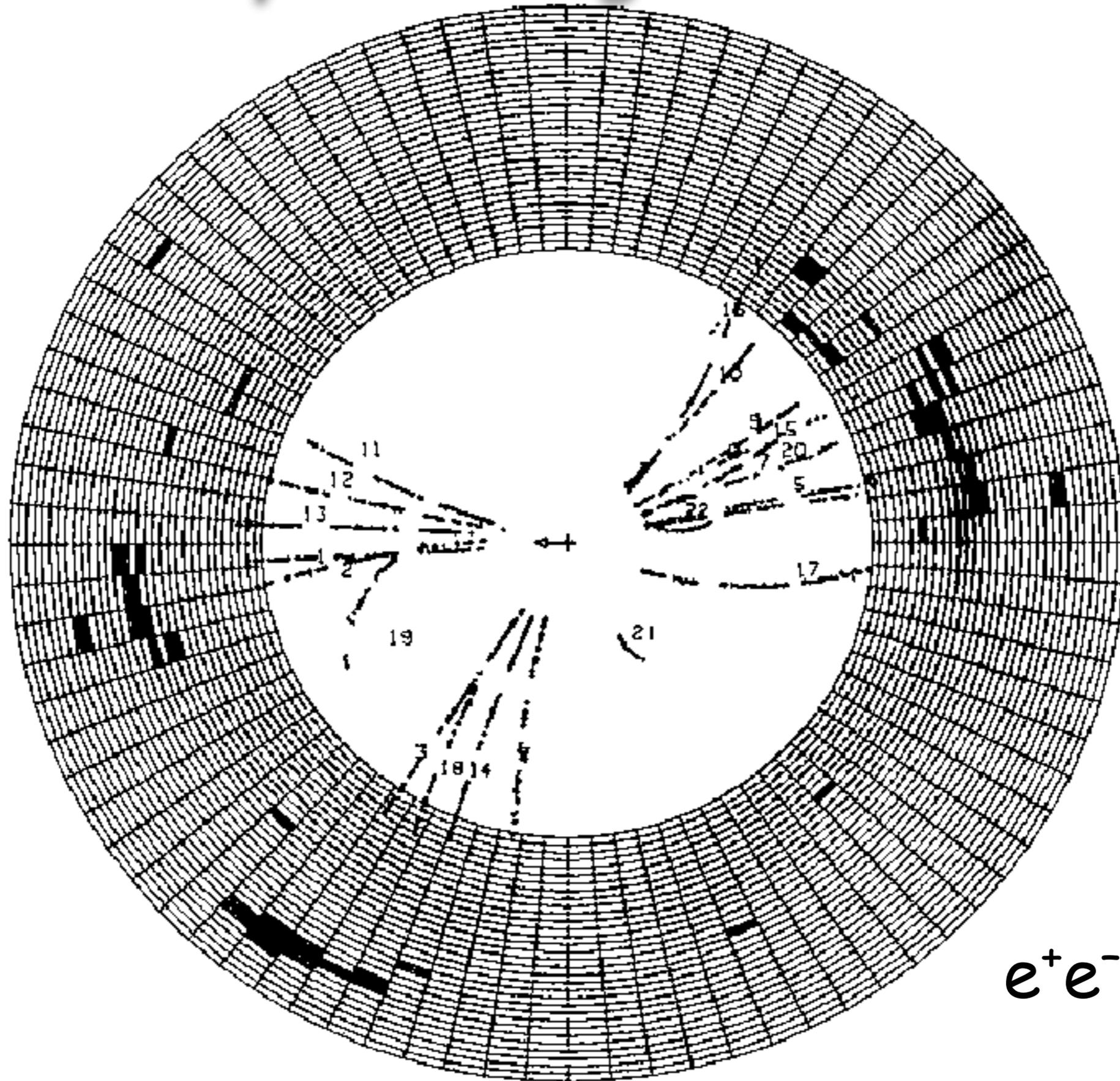Rutherford Lab., Tokio

operation time: 1978 - 1986

operation mode: $e^+ e^-$ annihilation; E_{cm} ~ 14 ... 46 GeV

JADE data taking

~200 pb^{-1} ; ~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

two examples:

Universality of QCD and Hadronisation

$1/\sigma \, d\sigma/d(1-T)$

- 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 α_s .
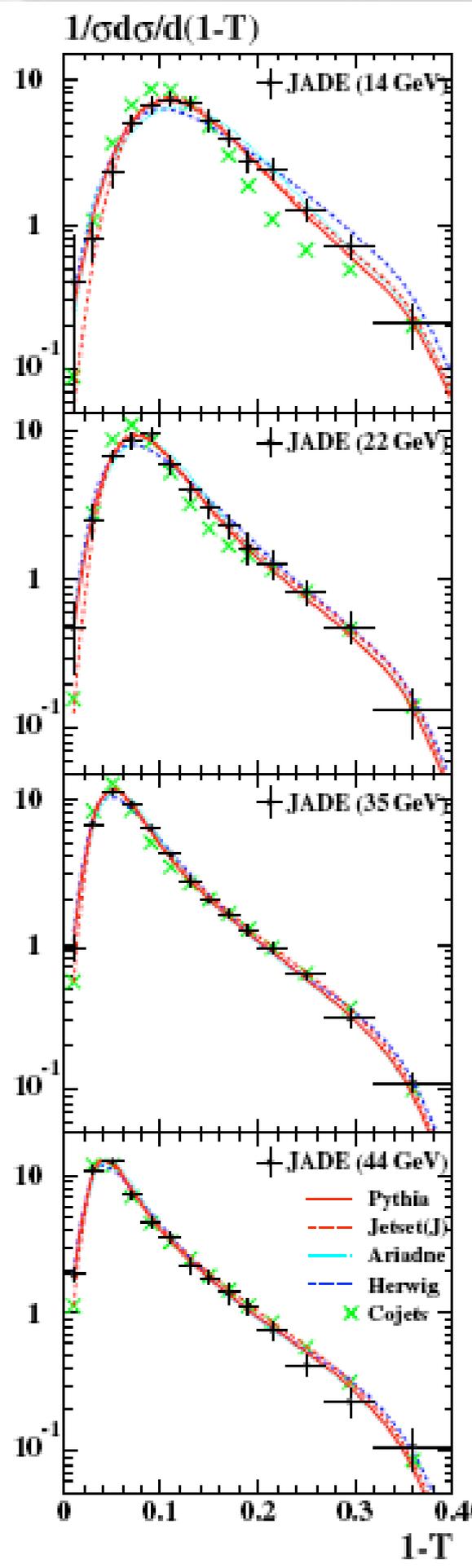
Precision measurements of α_s and proof of Asymptotic Freedom
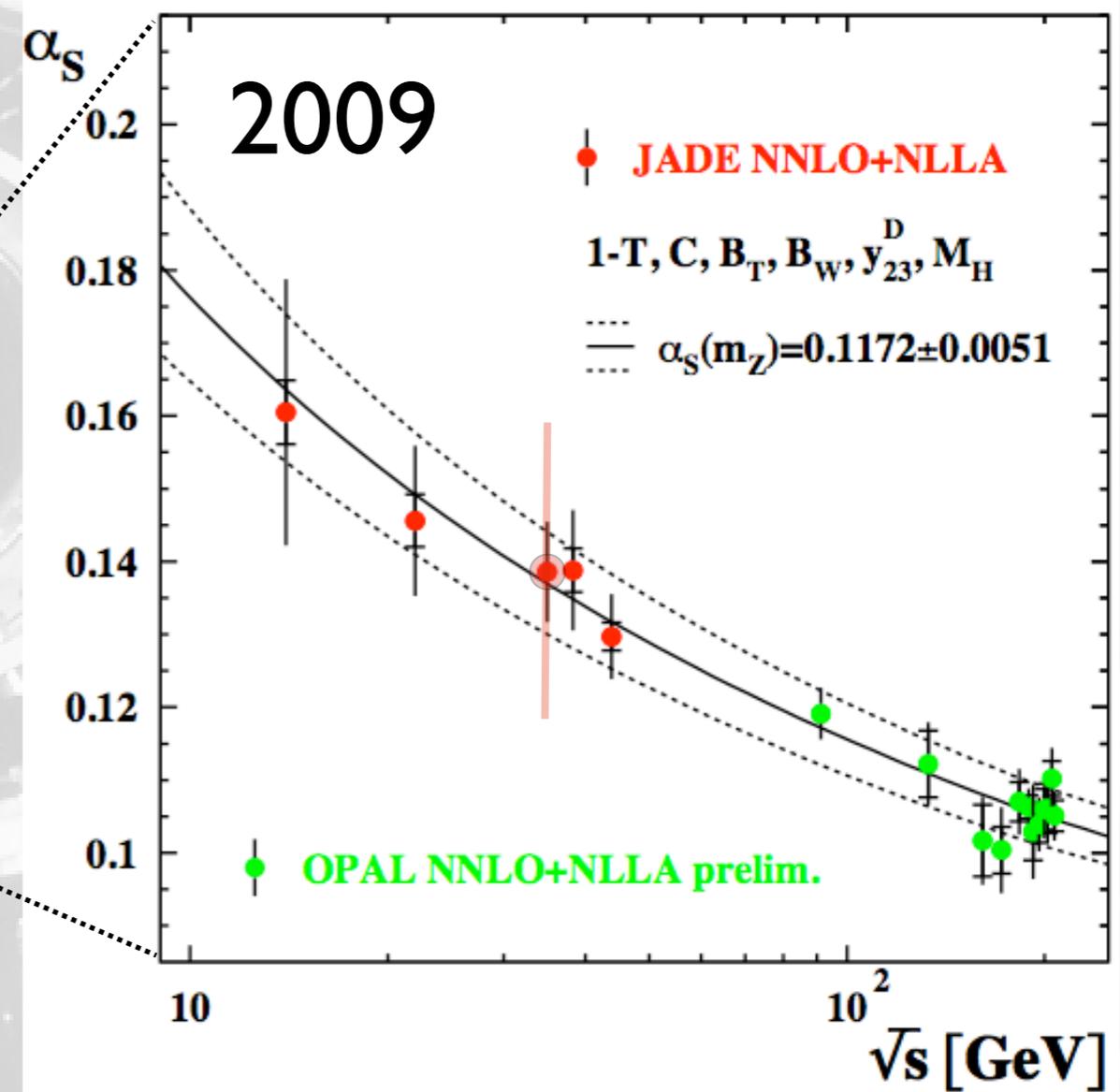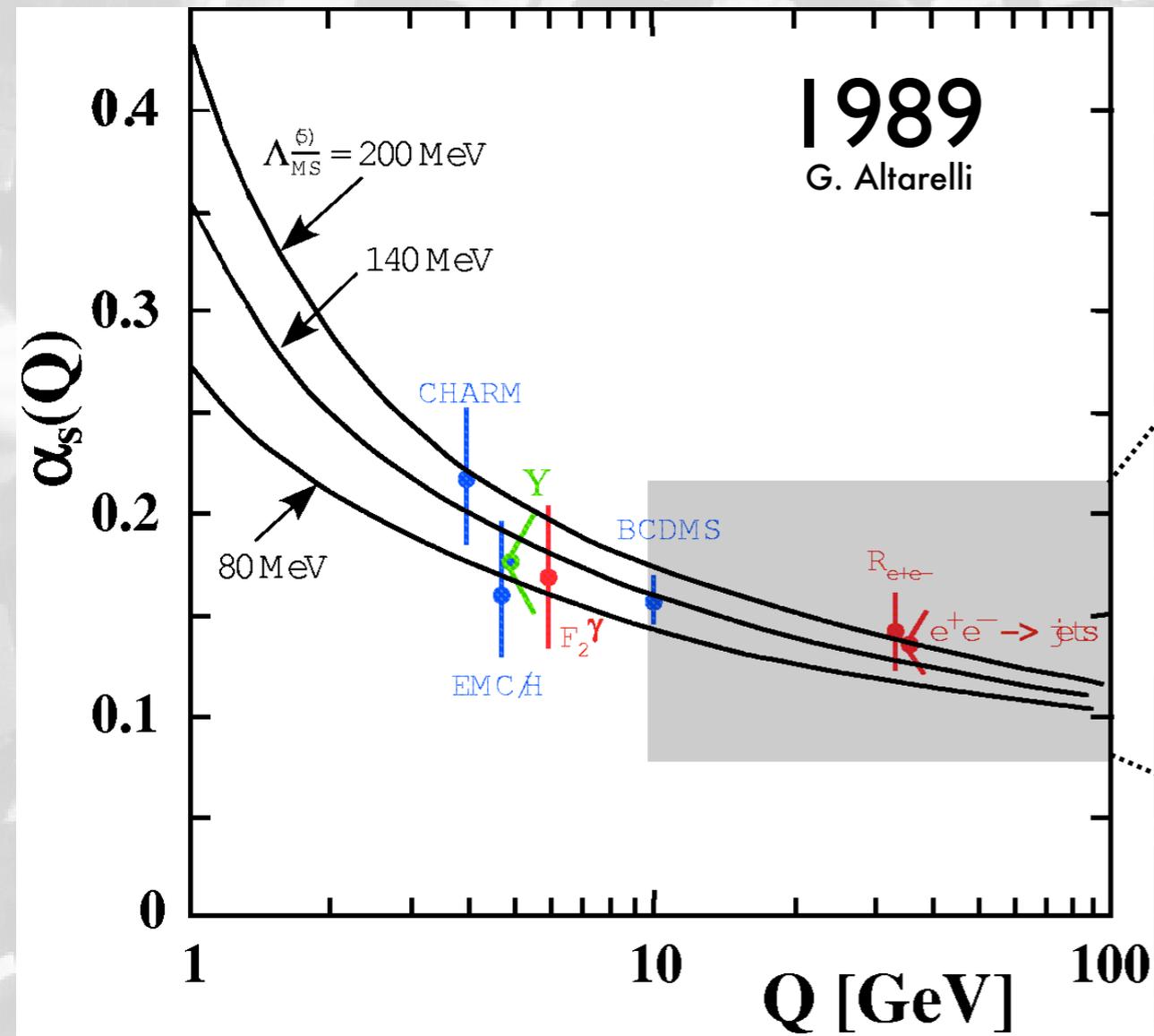
- in NLO QCD: $\alpha_s(35\ \text{GeV}) = 0.14 \pm 0.02$

- no running α_s signature

- in resummed NNLO: $\alpha_s(M_Z) = 0.1172 \pm 0.0051$

- significant proof of running α_s and asympt. freed.

JADE data and software

- 1995: „private" (*neither* collaboration *nor* lab) initiatives to :

  – rescue data from original archive tapes and copy them onto more modern media (IBM cartridges & Exabyte)
    (J. Olsson @ DESY)

  – reanalyse data using modern (LEP-like) methods and observables plus improved theoretical calculations
    (S. B. and P. Movilla-Fernandez @ RWTH Aachen

  – revitalise JADE software on modern computer platforms to enable generation of new MC data files
    (P. Movilla Fernandez, J. Olsson)

- so far, the only example of reviving and still using 25-30 year old data & software in HEP

- since 1996, O(10) publications, O(10) conf. contributions; no competition in e^+e^- data analysis at E_{cm} ~14 ... 200 GeV

JADE data and software formats

Software:

- Fortran IV (1974), Fortran 77, Sheltran, Mortran, Assembler; on IBM/370 (offline) and Nord 10S/50 (online)

- need to minimize memory and processing time:
  - compact programming, extensive use of features and functions which nowadays are dispraised

    (use of `ENTRY`; alternate `RETURN` statements etc.)

- core software archived and saved at DESY, but routines kept on *private user accounts* were inevitably *lost* when removing and destroying old archive tapes. (e.g. JADE muon system libraries gone forever ... !)

JADE data and software formats

Data:

- BOS (version of 1979) binary format (V. Blobel)

- compact mixture of 16-bit integers, 32-bit integers and floating point numbers
  –> correct access is platform dependend (byte ordering)!

- about 1 TB, originally stored on 1600 IBM 9-track tapes: ~15 kB/event; ~$20 \cdot 10^6$ events

- 1995/96: data files (but *not MC* generated data files !!) copied onto EXABYTE cassettes (using FPACK; J. Olsson)

- autonomous „mini-DST" or „n-tuple" like formats for fast and efficient data analyses.

- of those, only ZE4V format (E. Elsen) and files (also some MC) survived trashing of old archive tapes.

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
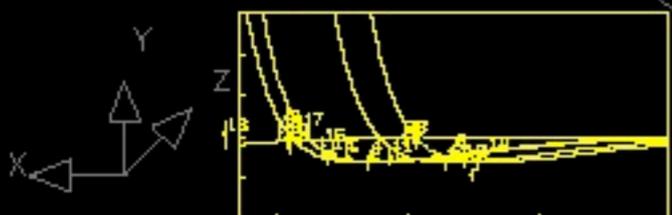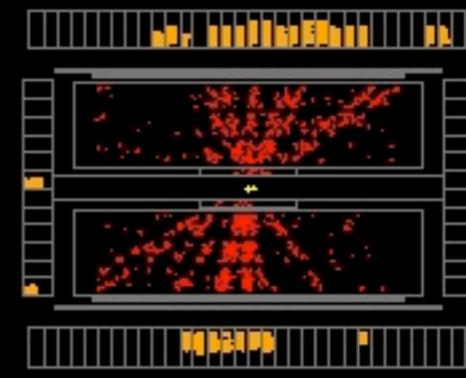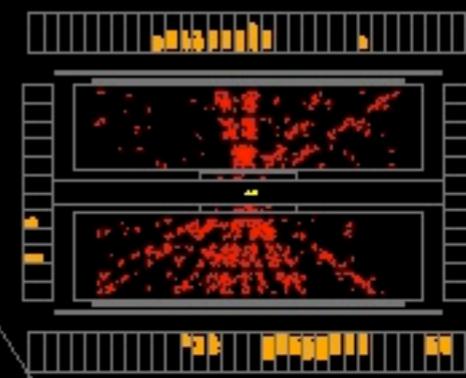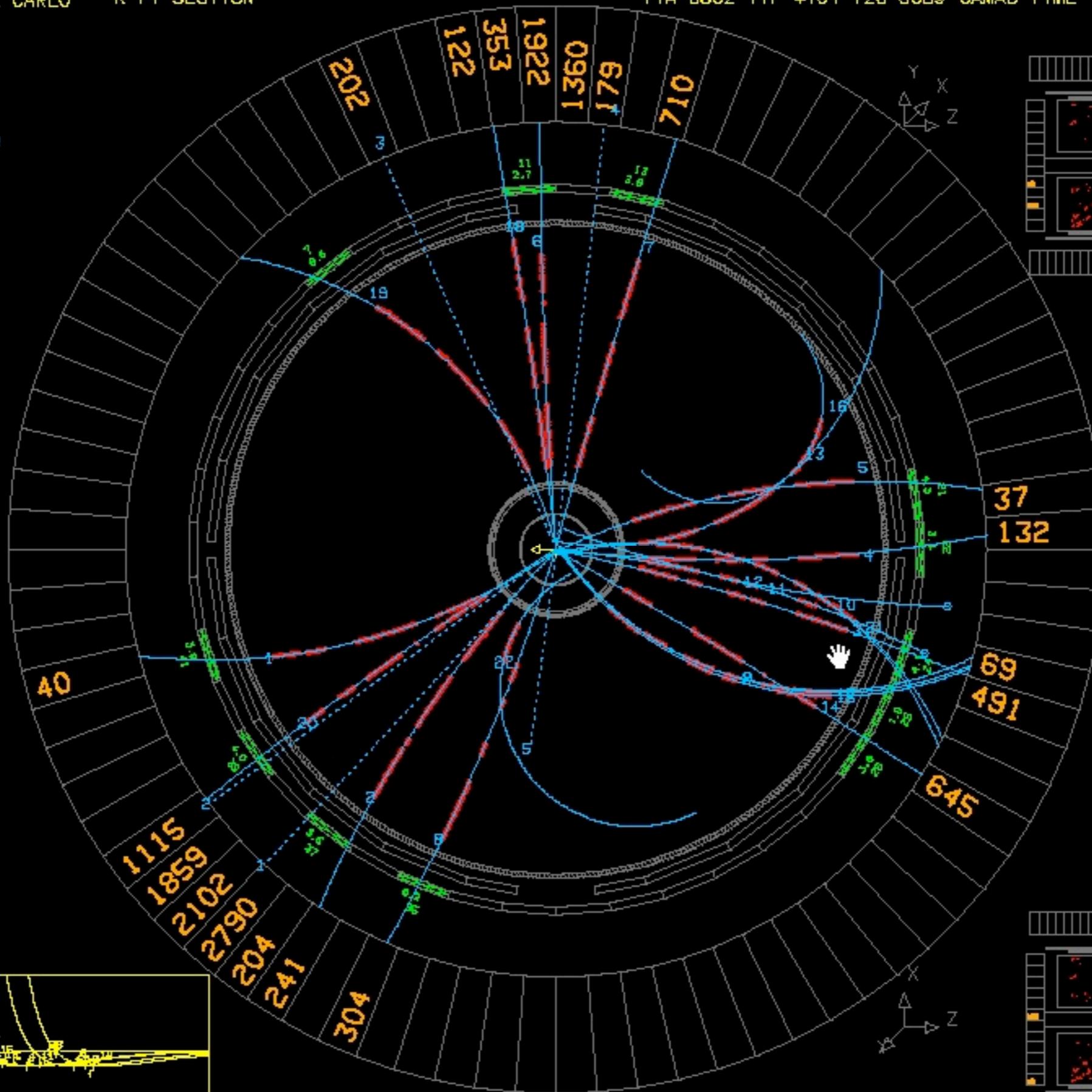
Revitalisation of JADE data

- 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 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

Future prospects

- develop (semi automatic) portability of JADE software
  to other platforms (Mac OS, Linux/PC)

- studies of topics which involve access to raw data
  (inclusive lepton production; b-jet and vertex tagging; take advantage of FADC resolution; ....)

- however: question of available manpower (LHC running!)

Lessons to be learned ....

- keeping data and software from old experiments for future (improved and new) analyses is absolutely neccessary and beneficial

- especially vital for the time long after initial analyses terminated!

- especially vital for processes and energies not being repeated at future projects

- in order to minimise efforts and to separate physics from archeology, several requirements should be met:

  - *data and software formats must comply to sound, internationally accepted and widely supported standards*

  - *data format and software must be concisely and consitently documented*

  - *same formats and archiving procedures for different projects and experiments*

  - *=> archiving & re-use of data & software must be planned while experiment is still in running mode!*

  - *proper and sustainable archiving must be assured (at accel. labs plus 1 loc. outside)*

  - *dedicated manpower to assure transfer of data and software to new platforms and hardware will be vital*

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)