A *Personal* View of Particle Physics Data

Francesca Di Lodovico, QMUL

Interoperability of Digital Repositories
QMUL, 2-4 December 2009

Queen Mary University of London

erc

Outline

- Physics, experiments and whereabouts
- Physics analysis
- Use of data after the end of the experiments
- Open questions
- Conclusions

Queen Mary University of London

erc

Outline

- Physics, experiments and whereabouts
- Physics analysis
- Use of data after the end of the experiments
- Open questions
- Conclusions

Queen Mary University of London

erc

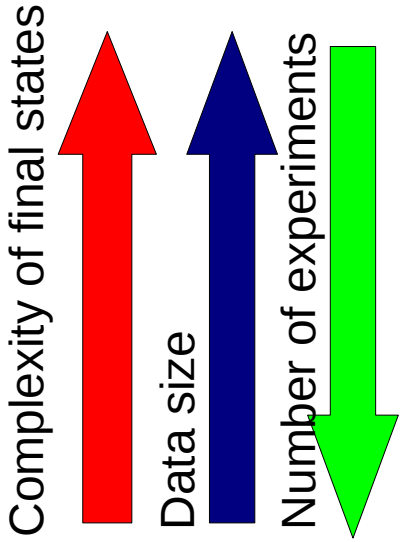
Experiments are spread all around

Basic Differences

•Underlying processes for producing elementary particles are different:

- Hadron Collider (proton against proton)
- Hadron Lepton Collider (proton against electron/positron)
- e^+e^- Collider (electron versus positron)
- Neutrino beam

....many more, eg kaon, tau experiments, etc.



•Experiments are: taking data, stopped taking data, planned.

Queen Mary University of London

erc

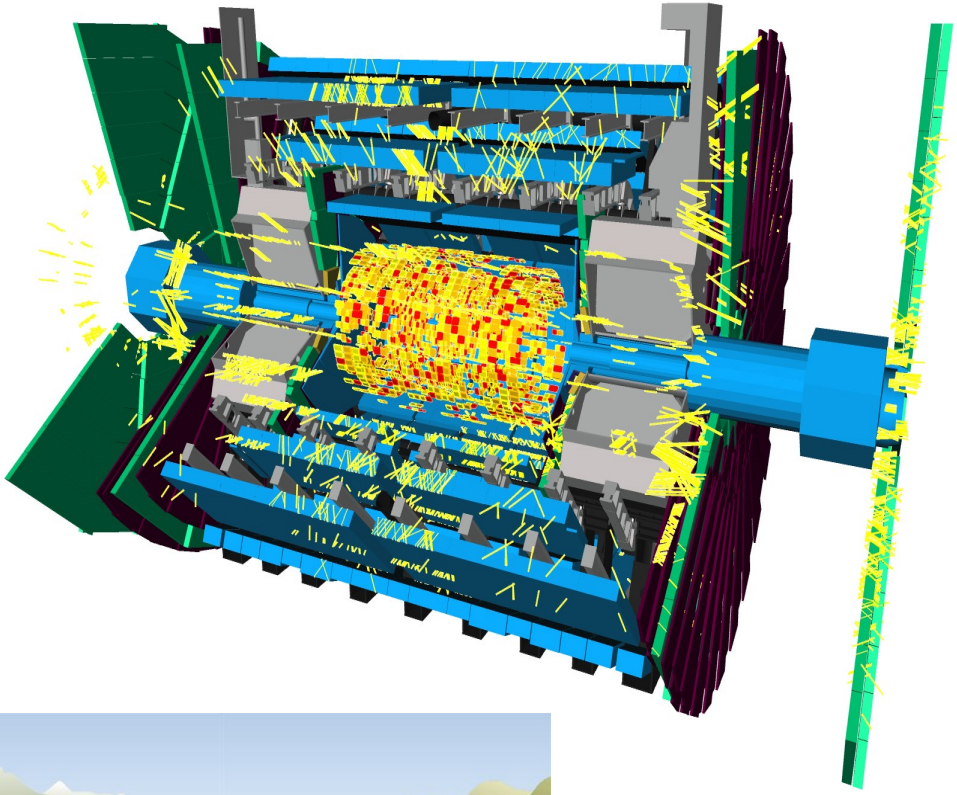
Hadron Colliders

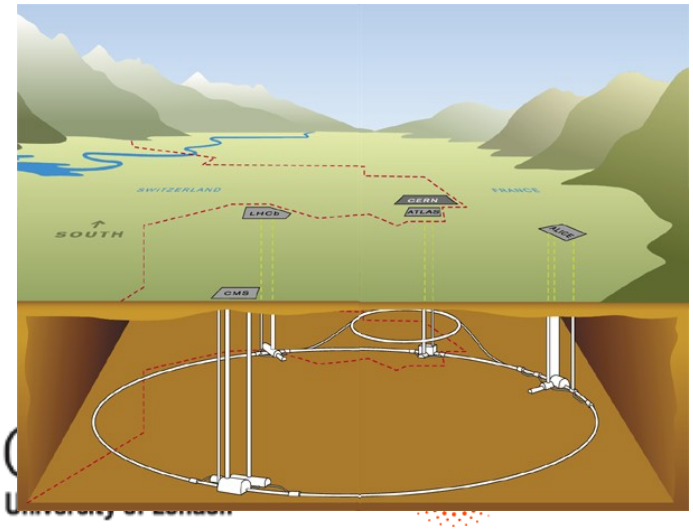
- Starting to take data: ATLAS, CMS, ALICE, LHCb @ LHC, CERN

- Taking data now: CDF, D0 @ Tevatron, Fermilab

- Past: eg UA1, UA2, etc @ CERN

First Atlas event

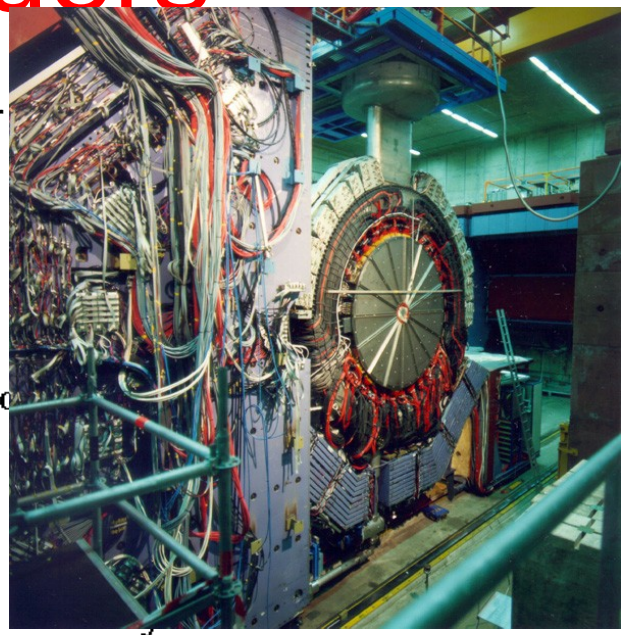




Lepton-Proton Colliders

Finished data taking:
H1 and ZEUS @ HERA

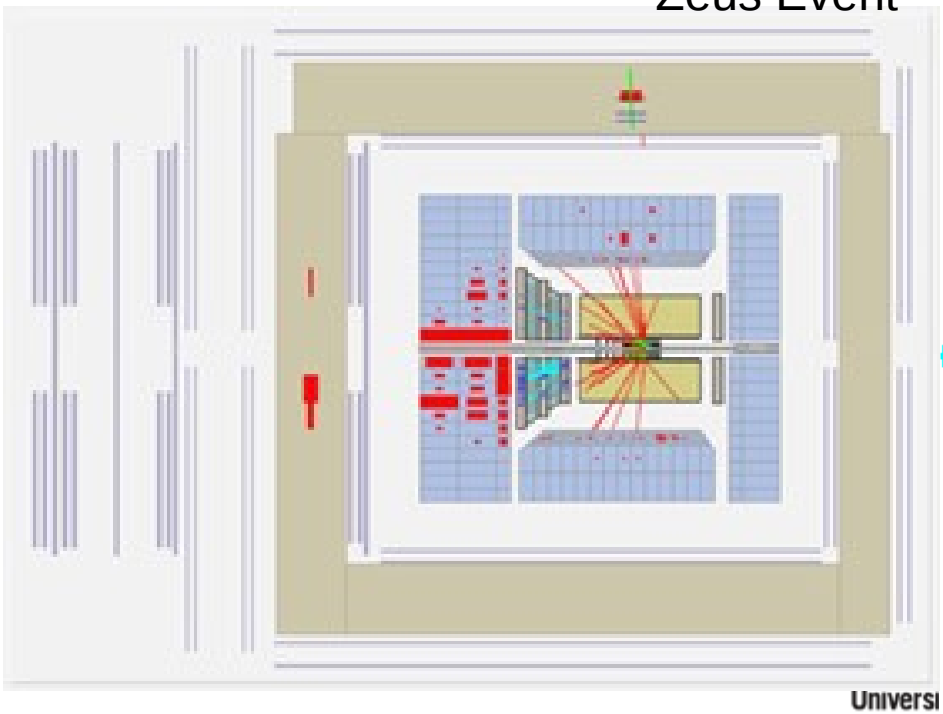
H1 detector

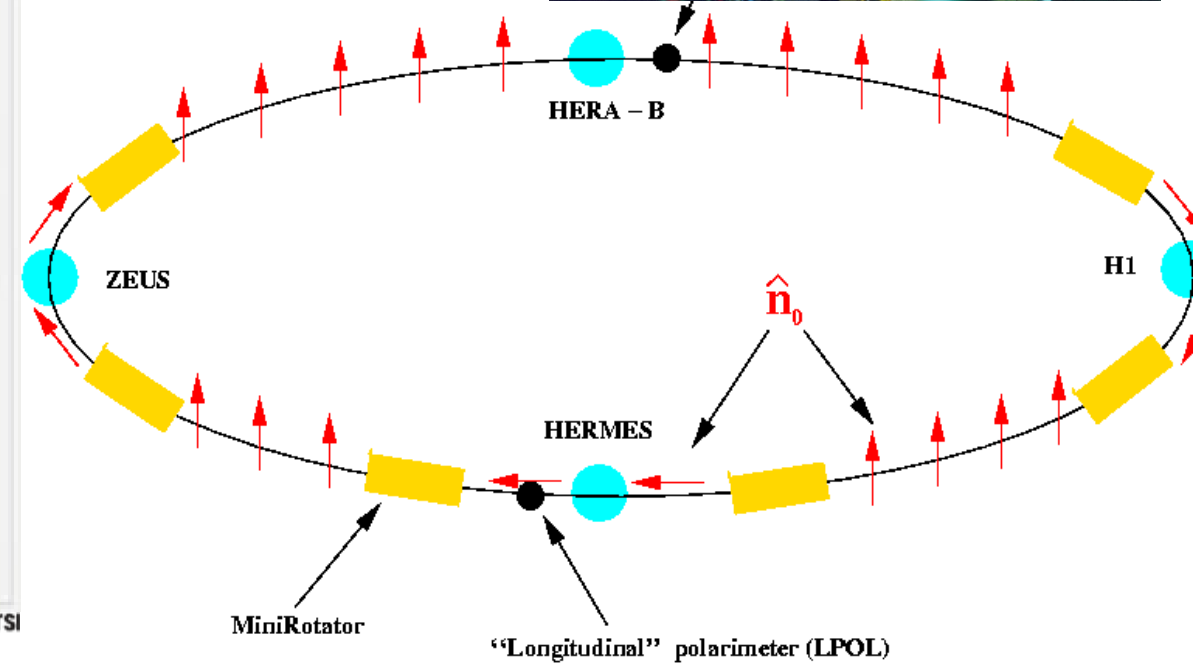


HERA electron/po

$\vec{P}_{meas} \,//\, \hat{n}_0$

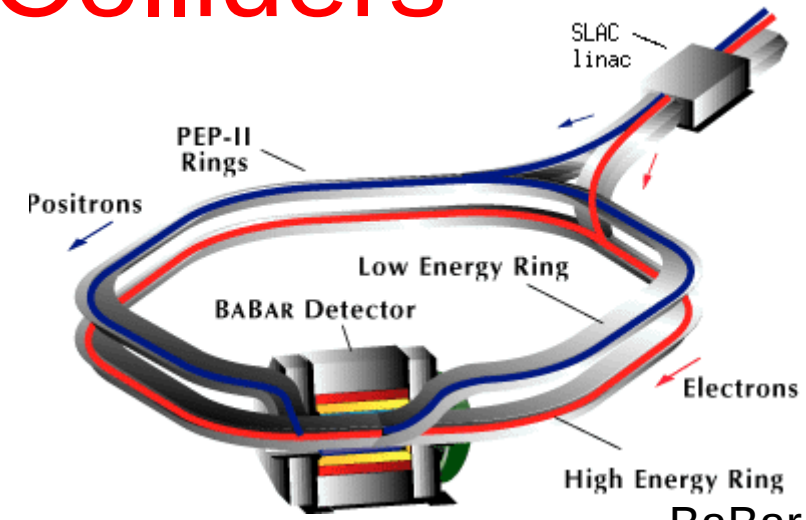
Zeus Event



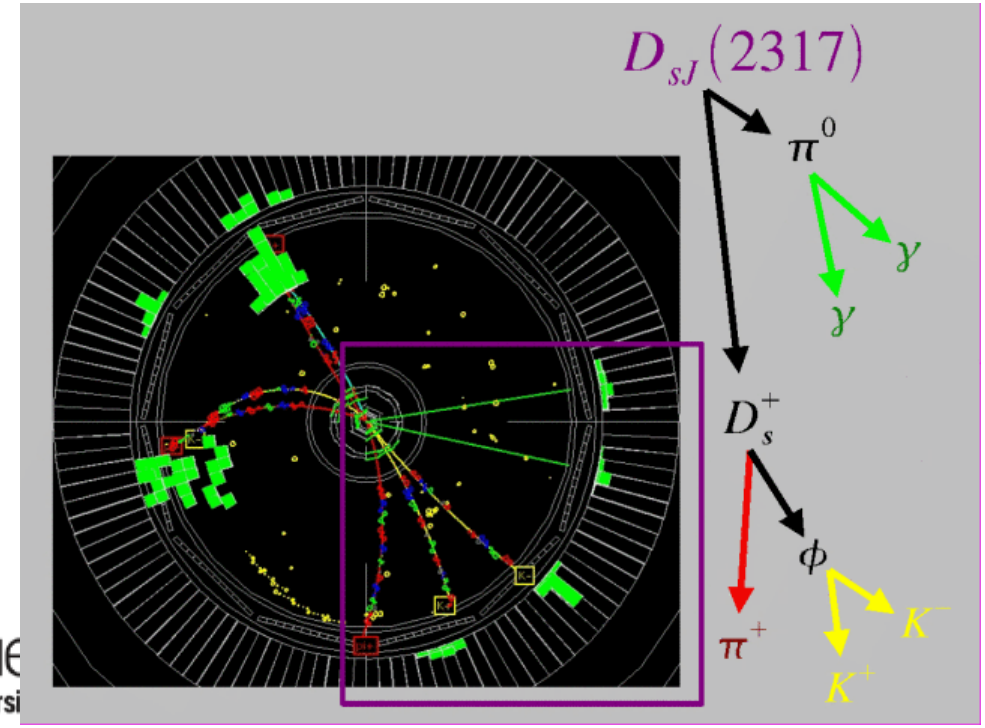


e^+e^- Colliders

- Running: Belle @KEKB, BES @ BEPC
- Finished data taking: BaBar @ PEPII, CLEO @ CESR, KLOE @ DAFNE
- Future: SuperKEK, SuperBelle, KLOE2
- Past: LEP @ CERN (Aleph, Delphi, L3, Opal), SLD@SLC



BaBar Event

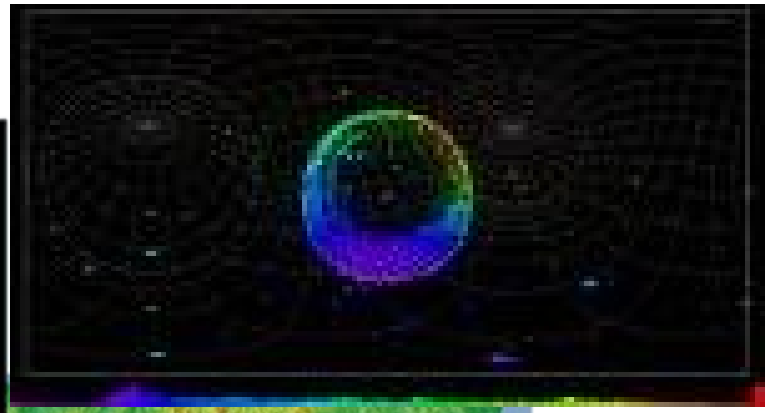


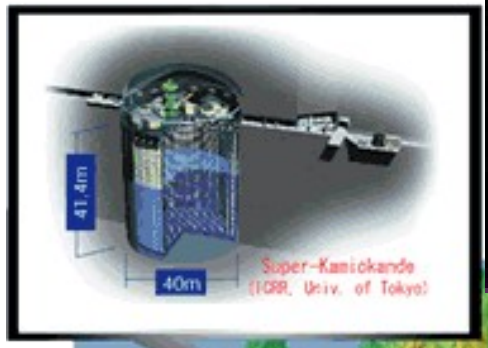
Neutrinos

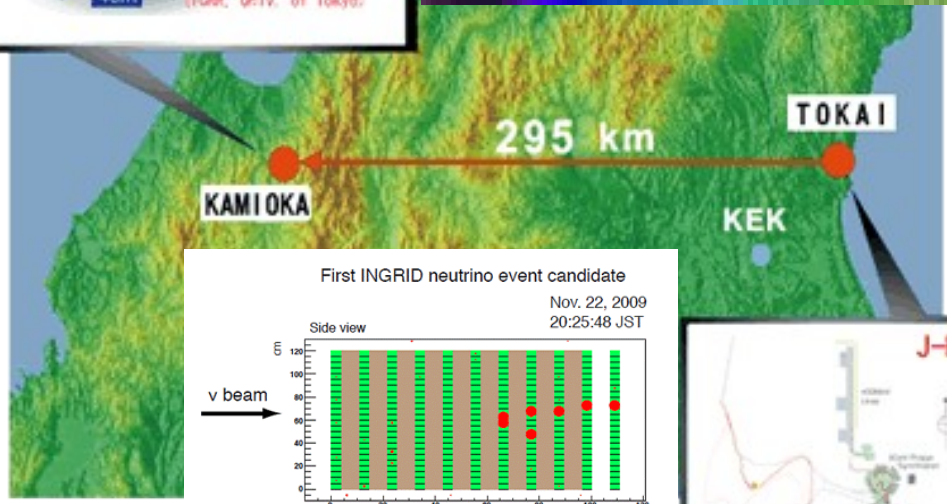
from natural or human-made sources

- Running: T2K, MINOS, MiniBoone, SciBoone, SK...
- Past: K2K, SNO, Kamland, Chooz...
- Future: Neutrino Factor, Double Chooz, Daya Bay, Nova...

SK event







INGRID event

Super K
Sonic Boooum
By Nelly Ben Hayoun

Queen Mary University of London

Outline

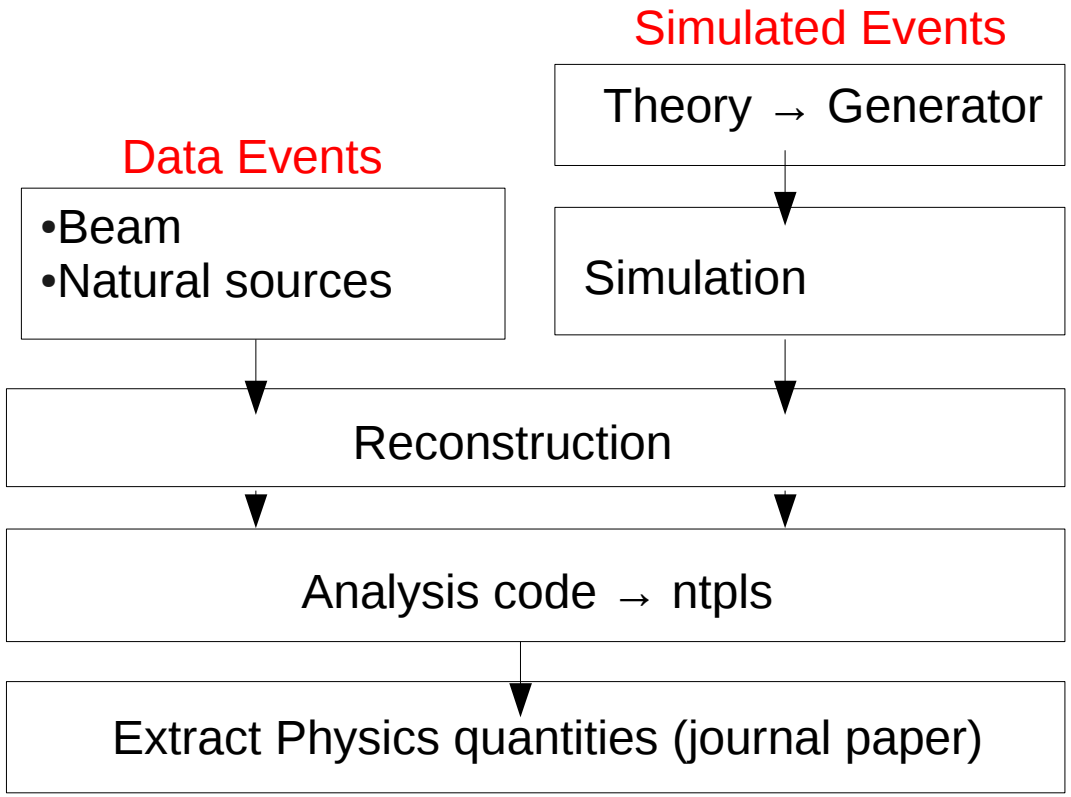
- Physics, experiments and whereabouts
- Physics analysis
- Use of data after the end of the experiments
- Open questions
- Conclusions

Queen Mary University of London

erc

Analysis of data

Schematic diagram of how we get our physics quantities



Possible level of storage (data & code & metadata):

- All
- Intermediate
- Journal-paper info

Queen Mary University of London

erc

Outline

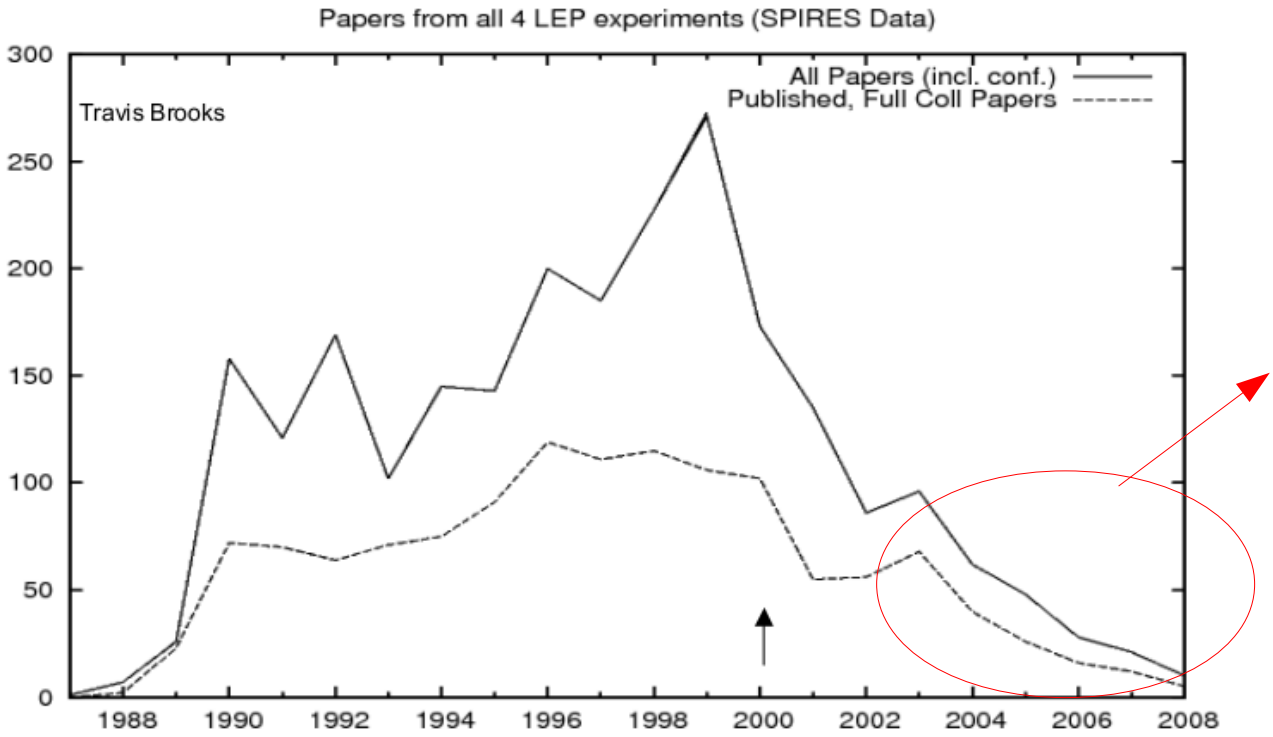
- Physics, experiments and whereabouts
- Physics analysis
- Use of data after the end of the experiments
- Open questions
- Conclusions

Queen Mary University of London

erc

Data analysis longevity

Significant analysis for 2-3y after the end of data taking, but still papers after 8y



What kind of analysis is performed after few years from end of data taking? For how many years will it continue?

Let's look at different uses of data and distinguish them according to the difficulty degree.

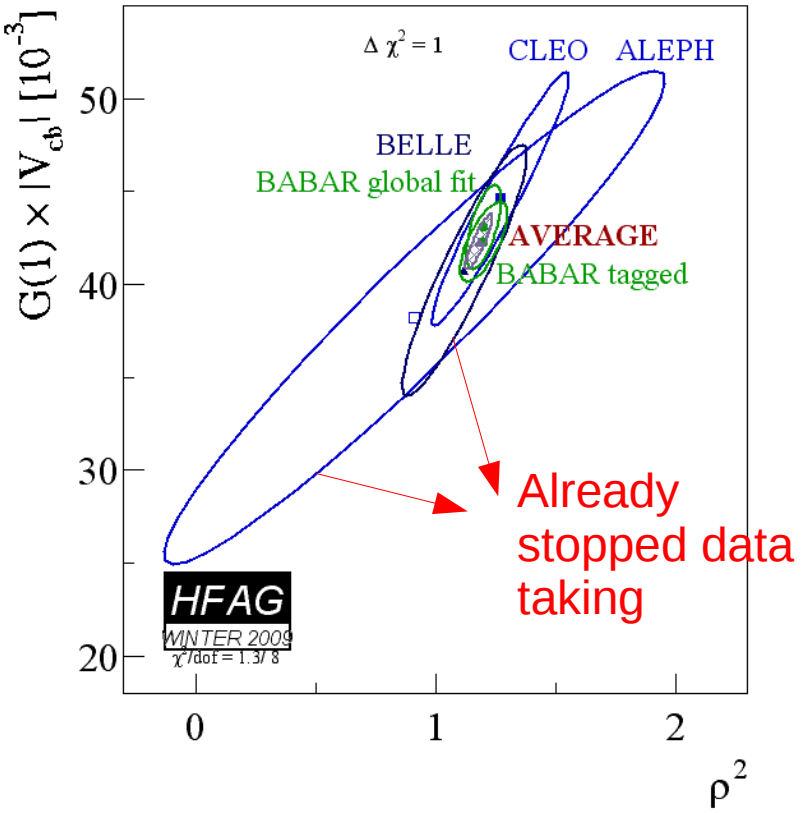
Queen Mary University of London

erc

The “easy” case: interpreting the results

“Easy” case means changing analysis code at most and run analysis, eg use published data for combining them with other results from other experiments. Only Analysis code and reconstructed data needed.

- Basic info from journal papers
- More results, eg correlation matrices, used for fit (eg HFAG) → it's basically more docs to keep.
- Use functions, eg likelihoods, for fits → ability to rerun the analysis code.



Queen Mary University of London

erc

Still “easy” cases

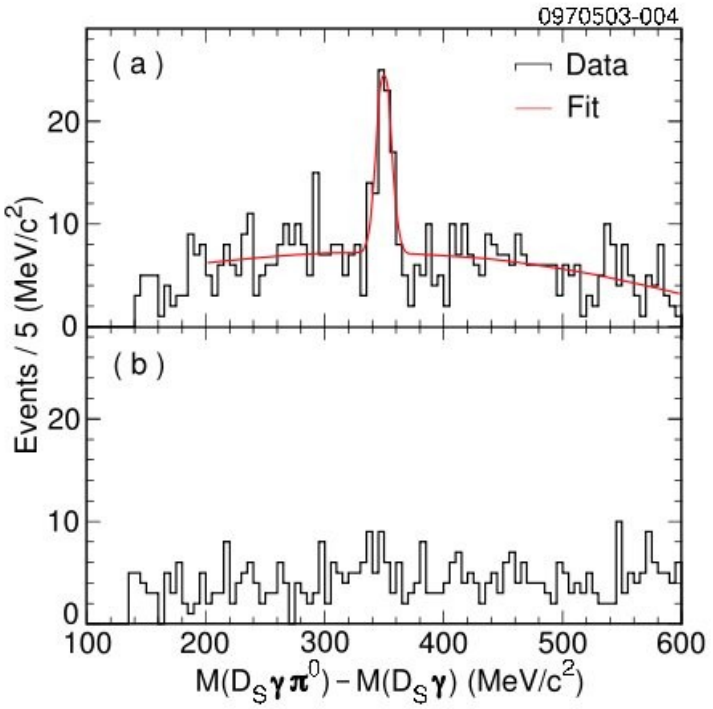
- Develop new analysis technique with advantages wrt the past (eg less errors) → modify and rerun analysis code
- There is something unexpected suggested by other experiments: create new analysis code and run it. Recent examples from CLEO:

Phys.Rev.D68:032002,2003, Erratum-ibid.D75:119908,2007.

Abstract

Using 13.5 fb^{-1} of e^+e^- annihilation data collected with the CLEO II detector, we have observed a narrow resonance decaying to $D_s^{*+}\pi^0$, with a mass near 2.46 GeV/c^2. The search for such a state was motivated by the recent discovery by the BaBar Collaboration of a narrow state at 2.32 GeV/c^2, the $D_{sJ}^*(2317)^+$, that decays to $D_s^+\pi^0$. Reconstructing the $D_s^+\pi^0$ and $D_s^{*+}\pi^0$ final states in CLEO data, we observe peaks in both of the corresponding reconstructed mass difference distributions, $\Delta M(D_s\pi^0) = M(D_s\pi^0) - M(D_s)$ and $\Delta M(D_s^*\pi^0) = M(D_s^*\pi^0) - M(D_s^*)$, both of them at values near 350 MeV/c^2. We interpret these peaks as signatures of two distinct states, the $D_{sJ}^*(2317)^+$ plus a new state, designated as the $D_{sJ}(2463)^+$. Because of the similar ΔM values, each of these states represents a source of background for the other if photons are lost, ignored or added. A quantitative accounting of these reflections confirms that both states exist. We have measured the mean mass differences $\langle\Delta M(D_s\pi^0)\rangle = 350.0 \pm 1.2$ [stat.] ± 1.0 [syst.] MeV/c^2 for the $D_{sJ}^*(2317)^+$ state, and $\langle\Delta M(D_s^*\pi^0)\rangle = 351.2 \pm 1.7$ [stat.] ± 1.0 [syst.] MeV/c^2 for the new $D_{sJ}(2463)^+$ state. We have also searched, but find no evidence, for decays of the two states via the channels $D_s^{*+}\gamma$, $D_s^+\gamma$, and $D_s^+\pi^+\pi^-$. The observations of the two states at 2.32 and 2.46 GeV/c^2, in the $D_s^+\pi^0$ and $D_s^{*+}\pi^0$ decay channels respectively, are consistent with their interpretations as $c\bar{s}$ mesons with orbital angular momentum $L = 1$, and spin-parity $J^P = 0^+$ and 1^+.

PACS numbers: 14.40.Lb, 13.25.Ft, 12.40.Yx



“medium” cases: new inputs

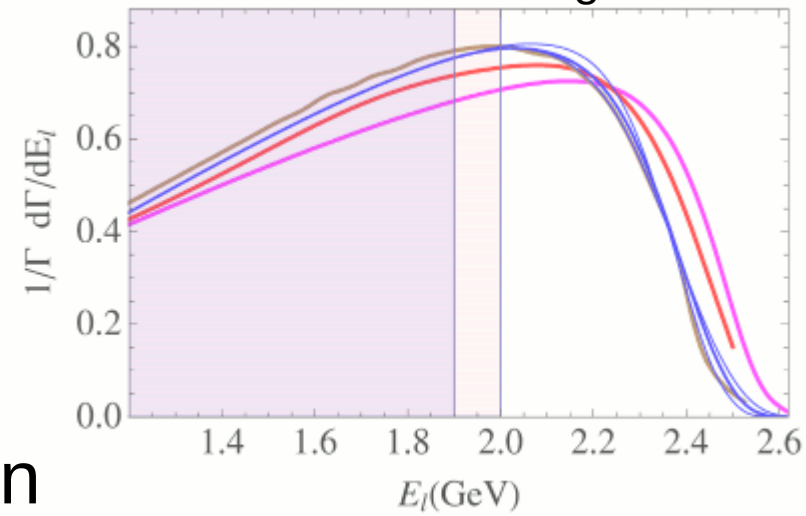
- Variation of input parameters. Different kinds of params:
  - Theoretical parameters in generator
  - Constants in simulation/ reconstruction code.
    - Eg in 2004, we reran the simulation for our analysis, and a variation of the central value of a parameters in Geant, changed the ratio of 2/3 jets in our b-physics sample. We had to change the weights according to data/MC comparison in our analysis to take care of it.
  - Calibration constants, eg after further detector studies.
- The code can be run as it is, but different parameters may bring differences in the final results to be dealt with.

“difficult”: code changes

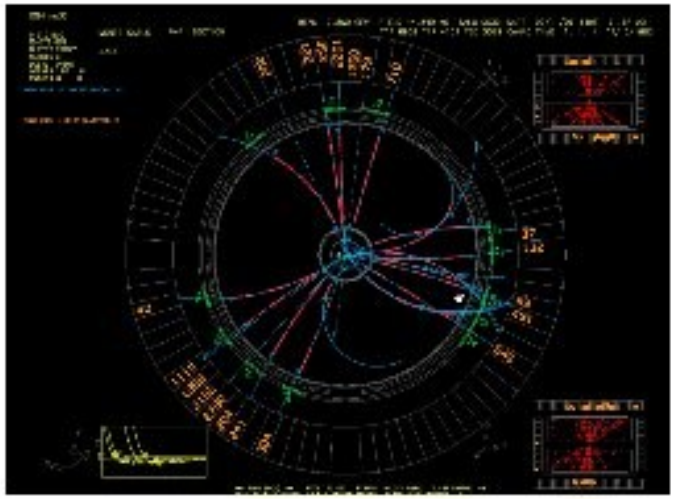
Deep code changes, eg:

- MC generator, eg using a different model
- MC simulation, improved simulation eg using different models.
- Reconstruction, eg better understanding of the detector

Different theoretical models for generator:



Example from JADE



Queen Mary University of London

What used the most & for how long?

- Use input from papers or analysis eg for combination w/ other experiments or testing new theoretical models → always done.
- From the end of data taking lot's of analysis activity for few years → continue exploiting physics potential.
- In longer term, the usage of data strongly depends on what kind of future experiments we will have.
  - New experiments with much more stat and more precision. I would envisage all the old results to be superceeded.
  - New experiments we partial overlap with the past data: yes we will need to have a subset of data available.
  - No new experiments: yes, we need to keep all.

Outline

- Physics, experiments and whereabouts
- Physics analysis
- Use of data after the end of the experiments
- Open questions
- Conclusions

Queen Mary University of London

erc

Open Questions

- Collaboration Policy: who can look at the data? How this policy will change with years?
- Can (part of) data be distributed for outreach?
- Should the code kept up to date with new technologies? If so, are there tools in place to check that the results still make sense? Eg. “monitoring” analyses?
- How do we deal with the plethora of experiments and different needs?
- What information needs to be kept along w/ the data? (docs, code, etc)
- How is expertise of the experiment retained during the years?
- How early should we start thinking about the future of our data after the end of data taking?

Outline

- Physics, experiments and whereabouts
- Physics analysis
- Use of data after the end of the experiments
- Open questions
- Conclusions

Queen Mary University of London

erc

Conclusion

- Many kinds of experiments, ie many kinds of data and needs to deal with.

- Use of data after the data taking can be distinguished in: “easy”, “medium”, “difficult”
→ different preservation needs

- Many open questions to be answered to define strategy

Queen Mary University of London

erc