



Update on Long Term Data Preservation @ LHCb

- October 16th, 2013 -

S.Amerio (U. Padova) for the LHCb LTDP task force

Official since March 2013.

It follows DPHEP levels of data preservation.

Level 1 (published data)	All scientific results are public. Data associated with the results will also be made available; format and repositories will be decided by the Editorial Board
Level 2 (samples for educational purposes)	LHCb already involved in outreach and education activities. Event displays and simple analysis level ntuples are already available and will continue to be provided to the public. The data are for educational purpose only, not suitable for publication
Level 3 (reconstructed data)	LHCb will make reconstructed data (DST) available to open public; 50% 5 years after data is taken, 100% after 10 years.
Level 4 (raw data)	Due to the complexity of the raw data processing stage, the extensive computing resources required and enormous access to tape resources, direct access to raw data is not permitted to individuals within the collaboration. Raw data processing is performed centrally. Due to this, the collaboration is currently not planning to allow open access to raw data

Adding information to published results: data to produce the published histograms.

CERN Document Server

Search Submit Help Personalize

me > CERN Experiments > LHC Experiments > LHCb > LHCb Papers > Measurement of the charge asymmetry in $B^\pm \rightarrow \phi K^\pm$ and search for $B^\pm \rightarrow \phi \pi^\pm$ decays

Information Discussion (0) Files Linkbacks

Preprint

Report number arXiv:1309.3742 ; LHCb-PAPER-2013-048 ; CERN-PH-EP-2013-166

Title **Measurement of the charge asymmetry in $B^\pm \rightarrow \phi K^\pm$ and search for $B^\pm \rightarrow \phi \pi^\pm$ decays**

Related data file(s):
ZIP

Related supplementary data file(s):
ZIP

External link:
Preprint

Fig3.png
Fig3.pdf
Fig3.eps
Fig3.C

```
TGraph *graph = new TGraph(61);
graph->SetName("graph_allcpv_p");
graph->SetTitle("");
graph->SetFillColor(1);
graph->SetLineColor(4);
graph->SetLineWidth(2);
graph->SetMarkerStyle(20);
graph->SetMarkerSize(0.6);
graph->SetPoint(0, -0.06221261421, 7.236171847);
graph->SetPoint(1, -0.05650273742, 6.969961101);
graph->SetPoint(2, -0.04744748815, 6.703843144);
graph->SetPoint(3, -0.03719151822, 6.437773779);
graph->SetPoint(4, -0.03218946446, 6.304733425);
graph->SetPoint(5, -0.02622174243, 6.171732888);
```

Soon available in Inspire too

Level 2: outreach and education

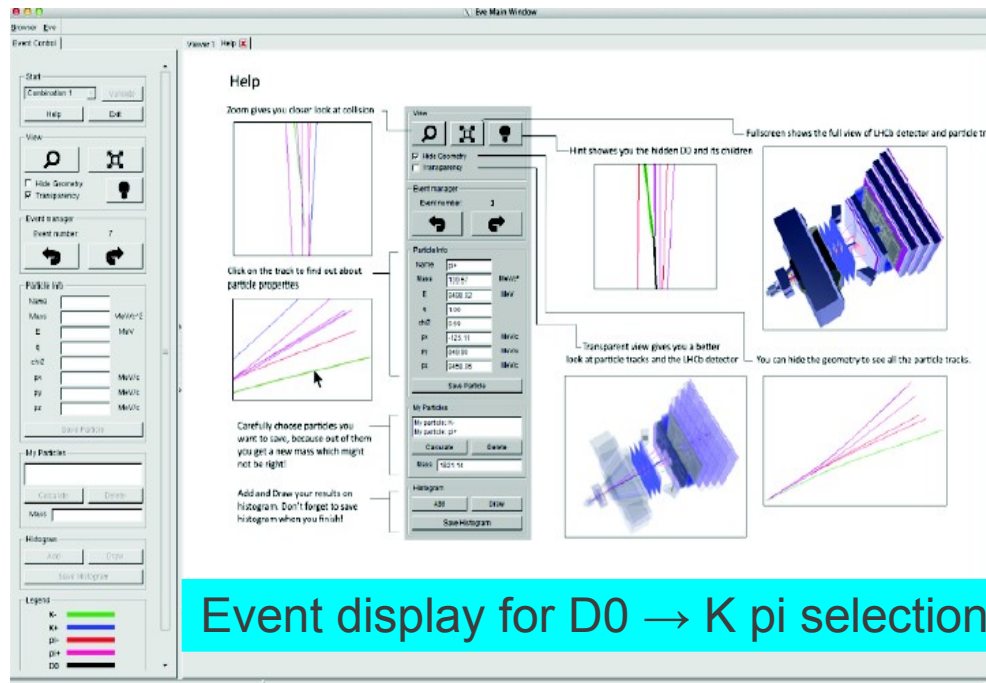
Samples of LHCb data available for educational purpose.

Example: $D^0 \rightarrow K\pi$ 2010 data for the International Masterclasses ($D^0 \rightarrow K\pi$ selection and D^0 lifetime measurement)

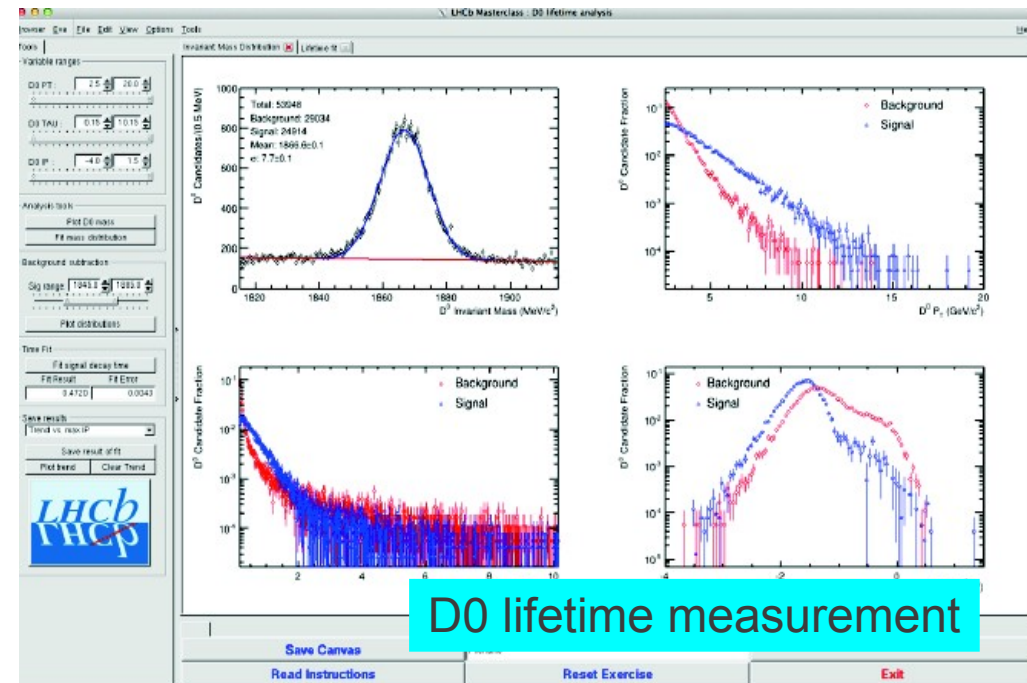
Exercise targeted for high school students.

Data stored in ROOT trees.

LHCb officially included in the International Masterclass program starting Spring 2014.



Event display for $D^0 \rightarrow K\pi$ selection



D^0 lifetime measurement

Level 3-4: preservation of the full analysis capability

Working on a specific use case: *production of new MC samples corresponding to old data*

Important use case for physics, e.g. it is necessary if we want to reanalyze old data to search for a signal predicted by a new theory.

Important for DP too. We are working on

- **Definition of requirements:**

- *DATA*: all data reprocessed with the latest version of the software
- *SOFTWARE*: full data processing chain: Generators, Detector simulation, Trigger simulation, Reconstruction and Stripping. *NB: We need to preserve the Trigger simulation corresponding to the specific data taking period!*

- **Documentation:** all steps for the current MC production (software versions, platforms, configuration files,...) → define documentation procedure for future productions.

- **Definition of a standard validation procedure.**

An efficient validation system is a fundamental ingredient of any long term data preservation plan.

Archived data and software need to be regularly checked against data loss and/or corruption, new operating systems and hardware.

Long term validation system should be built on top of the current LHCb validation tools.

Current status of LHCb validation framework: different tools, addressing specific needs, e.g,

- **nightly build system** to check the code against new platforms;
- set of **reference histograms defined and checked against old versions** to validate the physics output of the different softwares (MC generators, reconstruction, etc...)

LHCb is working towards a common framework, **LHCb Performance and Regression testing framework**, featuring

- automated run of reference tests for versions/nightly builds
- framework to gather results and send them to a central location
- application to display and compare results

Preservation of analysis level ntuples and software

LHCb authors currently preserve:

- **Analysis software** in ad-hoc repository (Erasmus)
- **Analysis level ntuples** in their private areas on Castor
- **Detailed instructions** on twiki pages

We are working to define a standard procedure for the archival of final ntuples and analysis software.

We should have a common, experiment-restricted area where users can upload for each analysis all the necessary material.

We are in contact with CMS and Cern-IT: can CDS offer a framework common to all experiments?

BACKUP