A Tender for the Scheduled Analysis Train on GRID

Jacek Otwinowski
Silvia Masciocchi
GSI Darmstadt
June 24, 2009

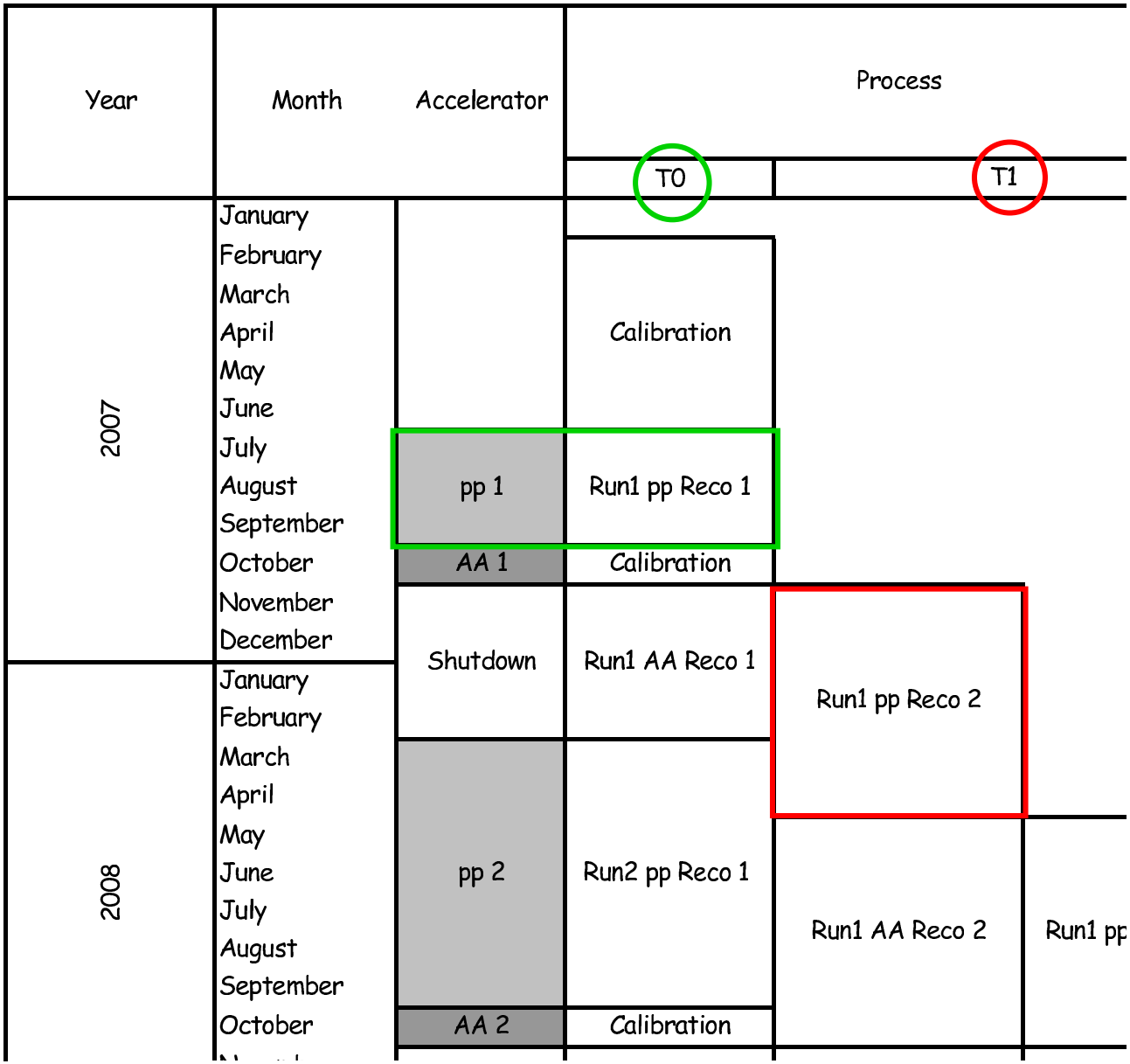
ALICE

The Topic

When shall we have reconstructed data (ESDs) where sufficiently good calibration is used to calculate quantities of interest?

Note:
This discussion concerns all analyses which rely on high statistics (months), NOT the day-1 or week-1 papers
(for which multiple reconstruction passes with improved calibration and alignment can be done)

The ALICE Computing Model



Computing TDR, 2005
Chapter 7, p.13

Reconstruction:

Pass 1: during the pp run, at Tier0

Pass 2: after the end of the pp run, at Tier1s

Times of Calib/Align

C. Cheskov, C. Zampolli

e.g. http://indico.cern.ch/conferenceDisplay.py?confId=50502

1. Partial calib/align from cosmics (summer 2009)

2. Calibration Framework with Collision Data

   2.0 Online Calibration (DAQ/DCS/HLT) $\rightarrow$ OCBD
   2.1 Reconstruction pass 1 (per chunk) at Tier0
   2.2 Offline Calibration (Step 1 + Step 2)
   improved calibration, validation $\rightarrow$ OCBD
   2.3 Alignment: similar scheme (eventually). Larger data samples needed, careful validation $\rightarrow$ OCBD

3. Reconstruction pass 2 will use the improved calibration and alignment.
   But WHEN will this happen? At the end of run 1?
   Winter 2010/2011??

Move pass 2 earlier!

- **Pass 2 is done at Tier1s**

  Different resources are used

  It can be done!!
  (needs preparation, probably)

- **the validation of the output of calibration and alignment will anyhow require a few months**

Proposal for the train

On-the-fly improvement at the analysis train

By the time when the scheduled analysis train will run on GRID, the OCDB will already contain the information to improve some of the ESD content.

Some examples:

- **TPC gain $\rightarrow$ dE/dx $\rightarrow$ TPC PID**
- **TRD reference samples $\rightarrow$ NN training $\rightarrow$ TRD PID**
- **beam diamond $\rightarrow$ primary vertex**

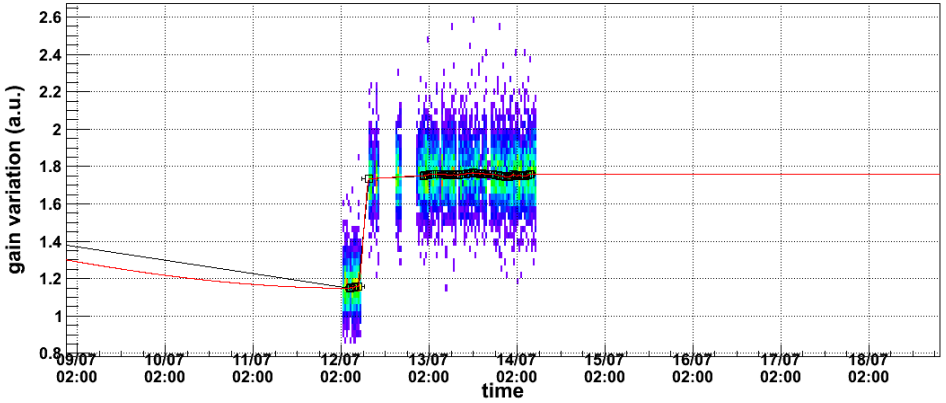
It would be a pity to ignore it!!

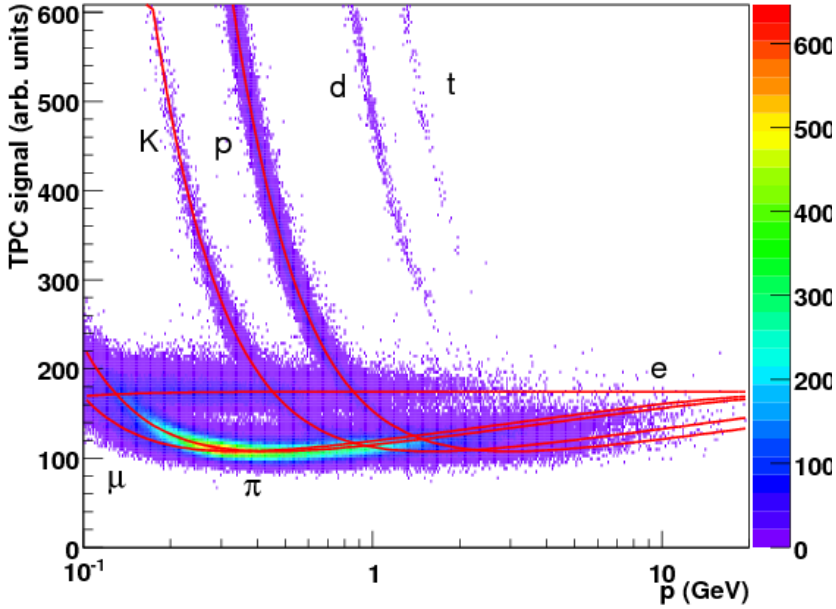
Example: the TPC PID

M. Ivanov, A. Kalweit, J. Wiechula

- the TPC gain varies in time
- the calibration takes care of most of the effects and saves the gain variation in the OCDB
- using this information allows to improve the dE/dx resolution (≈5.7%)
- re-cook the particle hypothesis likelihoods
- better TPC PID

This is already used successfully in the cosmics analysis





Example: the TRD PID

A. Andronic, A. Bercuci, M. Heide, A. Wilk

The TRD PID can be calculated both with neural networks (NN, default) and with a 2-dimensional likelihood method:

- **at pass 1, the NN training and the reference histograms for the 2-dim method will be from simulation mostly (partially from test beam data)**
- **only during pass 1, we will select a reference sample of electrons (conversions, TPC only) and pions (K^0_s)**
- **those will be used (offline) for the NN training and for new reference histograms → OCDB**
- **at the time of the analysis train, the results can be used to recalculate on-the-fly the TRD PID**
- **ALSO: the TRD gain varies in time and corrections need be taken into account**

One Obstacle

Access to the OCDB ?

During analysis, access to OCDB:
- is not foreseen
- is discouraged
to limit dependencies, to let analysis be lighter and not bound to databases

BUT the scheduled train is running on GRID (Tier0? Tier1s?)
The OCDB is available there without overhead.

Right here: we accept an exception and have OCDB available

The Tender

Tender (rail)

From Wikipedia, the free encyclopedia

A tender or coal-car is a special rail vehicle hauled by a steam locomotive containing the locomotive's fuel (wood, coal, or oil) and water.

The ALICE Tender

We propose to add in the scheduled analysis train an
ALICE tender
(BEFORE all other wagons) which:

- **re-cooks for example the TPC and TRD PID, the primary vertex w/ and w/o the actual beam diamond constraint**
- **is prepared by experts, and avoids fiddling by non specific experts (which can be incorrect and additionally can vary from analysis to analysis)**
- **the re-cooked values overwrite the ESD content in memory, and is then available to all wagons**

Can we avoid the exception?

What about the access to the OCDB?

The re-cooking happening in the tender might need only a few numbers or few small objects from OCDB

E.g. a histogram for the TPC gain factor, with time dependence

Can we save this information INSIDE the ESD file itself?

It can be considered for sure!

But care about the limitation: that will ONLY be the knowledge at the end of pass 1, and no later improvement will be considered

One example: this would not consider the TRD neural network training, which is done offline on larger statistics → serious limitation for the TRD PID!

Outlook

- evaluate if the tender is of interest and is accepted in the scheme
- check what are ALL improvements which can be re-cooked in the tender
- implement the code (experts)

- consider to move the pass 2 to an earlier time