Users' Analysis Experience Part 2

- HCAL and TCMT -

Outline:

- Scope of this Talk
- 'Typical' Analyses using HCAL and TCMT
- Data Reconstruction / Analysis Chain
- Monte Carlo Simulation / Digitisation / Reconstruction Chain
- Run 'caliceMarlin'
- Conclusions

Scope of this Talk

Following the idea for this talk given by Paul:

How do people work (analyse data) **right now** (concerning HCAL and TCMT)?

- What is **easy**?
- What is **hard** to do?
- What are main issues **limiting** the analysis at the moment?
- What is **missing**?

and estimate if

- these issues improve with the 'evolving' software model
- needs in the analysis will change with larger datasets in the future

restrict to analyses based on **reconstructed data** files

impossible, since **'nobody'** is working on the reconstructed data files or 'official' Monte-Carlos

- reconstructed data of 2006 came 'too late', 'out-of-date' calibration/corrections
- reconstructed 2007 data 'not' available yet
- documentation is missing: Where do I find the files? What is in the files? What processors and what processor parameters have been used? Which version of these processors?

'Typical' Analyses using HCAL + TCMT

e.m. studies: need HCAL, plus beam instrumentation, but no ECAL and TCMT

muon studies: need ECAL, HCAL and TCMT (?), plus beam instrumentation

hadron studies: need ECAL, HCAL and TCMT, plus beam instrumentation

- data **and** Monte-Carlo reconstruction/digitisation chain necessary
  - → study response, energy resolution, shower profiles, particle separation, ...
  - → compare data and Monte-Carlo
- people use **very different** software for these studies
  - → official CALICE software
  - → 'semi-official' software packages, e.g. HCAL software by Sebastian
  - → private software, might be advanced and contain more features, e.g. stand-alone HCAL software by Niels, but also others

→ there is no unique way of using **'the'** CALICE software, partly people only 'dump' the information into a ROOT file and do the analysis there

→ stick here to the official CALICE software/'semi-official' software package

'Typical' Analyses using HCAL + TCMT

e.m. studies: need HCAL, plus beam instrumentation, but no ECAL and TCMT

muon studies: need ECAL, HCAL and TCMT (?), plus beam instrumentation

hadron studies: need ECAL, HCAL and TCMT, plus beam instrumentation

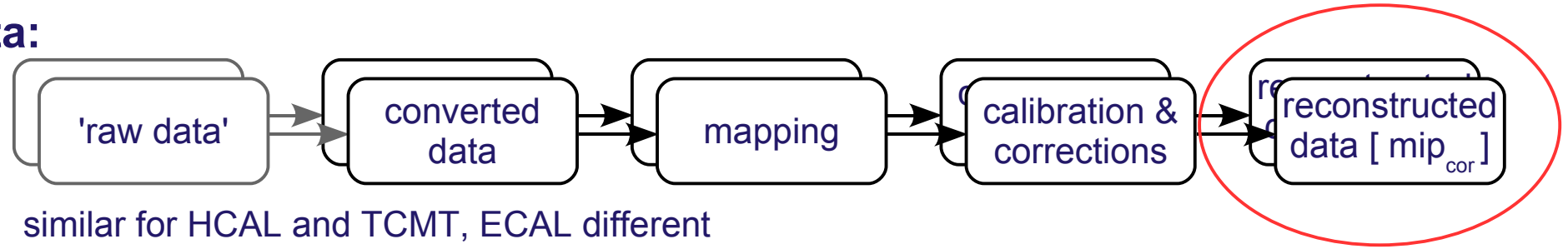
- data **and** Monte-Carlo reconstruction/digitisation chain necessary
  - → study response, energy resolution, shower profiles, particle separation, ...
  - → compare data and Monte-Carlo
- people use **very different** software for these studies
  - → however, more and more people 'converge on' using the official CALICE software

→ there is no unique way of using **'the'** CALICE software, partly people only 'dump' the information into a ROOT file and do the analysis there

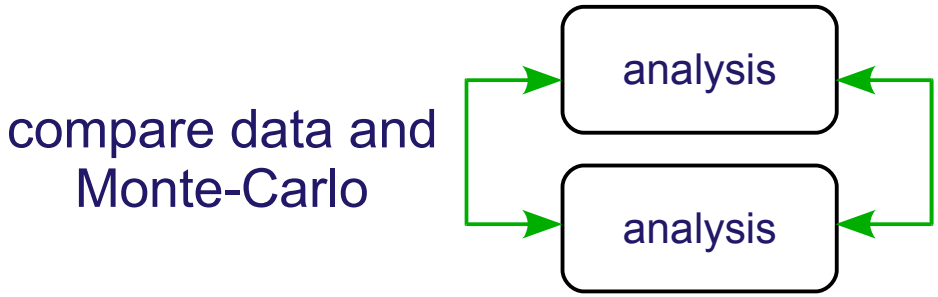
→ stick here to the official CALICE software/'semi-official' software package

Data Reconstruction / Analysis Chain and Monte-Carlo / Digitisation Chain

Data:



- similar for HCAL and TCMT, ECAL different
- additional: beam instrumentation

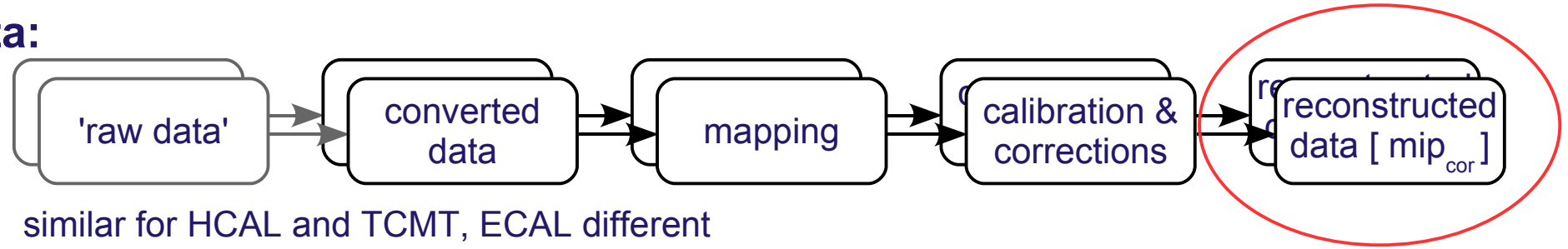


Monte-Carlo:

- similar for HCAL and TCMT, ECAL?
- beam instrumentation 'missing' (Trigger)

Data Reconstruction / Analysis Chain and Monte-Carlo / Digitisation Chain

Data:

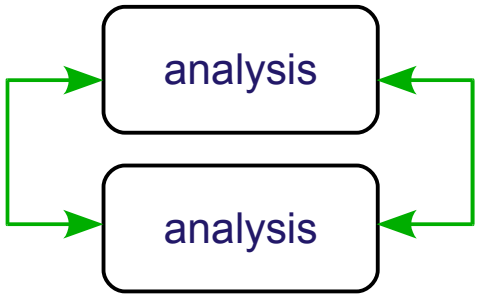


- similar for HCAL and TCMT, ECAL different
- additional: beam instrumentation

additional steps necessary:

- Conditions Processor
- beam instrumentation, trigger

compare data and Monte-Carlo

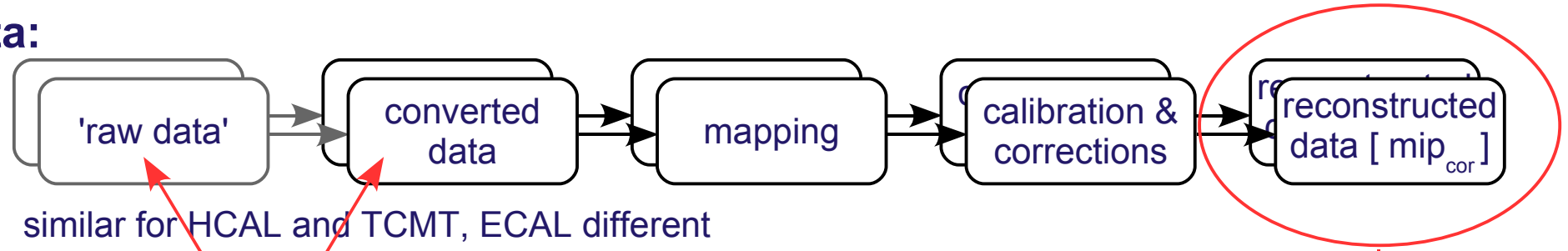


Monte-Carlo:

- similar for HCAL and TCMT, ECAL?
- beam instrumentation 'missing' (Trigger)

Data Reconstruction / Analysis Chain and Monte-Carlo / Digitisation Chain

Data:

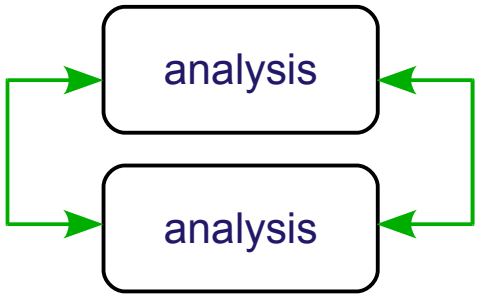


- similar for HCAL and TCMT, ECAL different
- additional: beam instrumentation

start from 'converted data'

- but people also/still use 'raw data'

compare data and Monte-Carlo



Monte-Carlo:

- similar for HCAL and TCMT, ECAL?
- beam instrumentation 'missing' (Trigger)

Data Reconstruction / Analysis Chain

first step: Conditions Processor

- processor itself: **easy** to use, no problem to access database, but many DBHandler needed (>150), **not very clear** (→ long steering files)
- which DBHandlers are necessary
- conditions database: not very user friendly
  - far too **many** folders
  - what is stored in which folder, which folders are obsolete, which not
  - consistent naming conventions would be nice
  - information in the folders is not obvious, need to understand the structure of the corresponding CALICE generic objects first
  - conddbview.tcl-script not much of a help

Data Reconstruction / Analysis Chain

first step: Conditions Processor

- processor itself: **easy** to use, no problem to access database, but many DBHandler needed (>150), **not very clear** (→ long steering files)
- which DBHandlers are necessary
- conditions database: not very user friendly
  - far too **many** folders
  - what is stored in which folder, which folders are obsolete, which not

- need naming conventions and version management of these folders
- perhaps 'condense' the information
- need more documentation and a **central** point to access it

Future estimate: converge on 'standardised' system, for larger data sets it gets more complex (more, different run conditions)

→ usage of **centrally reconstructed** data would be helpful

Data Reconstruction / Analysis Chain

second step: set-up beam instrumentation, mainly Trigger Processors

- Triggers: easy to use, but need a set of **reference** parameters / **reference** steering file
- Drift-Chambers: code is available, for 2006 all the tracks are already reconstructed, for 2007 missing (?)
  - need a set of **reference** parameters / **reference** steering file
  - two (?) different approaches for tracking (both are used)

Data Reconstruction / Analysis Chain

second step: set-up beam instrumentation, mainly Trigger Processors

- Triggers: easy to use, but need a set of **reference** parameters / **reference** steering file
- Drift-Chambers: code is available, for 2006 all the tracks are already reconstructed, for 2007 missing (?)
  - need a set of **reference** parameters / **reference** steering file

- Triggers: more detailed documentation would be nice plus a **central** point to access it
- Drift-Chambers: not really used in HCAL analyses yet

Future estimate: Triggers are in good shape, need a bit of polishing, Drift-Chambers: tracks for 2007 data should be available soon

for both: larger data sets shouldn't give a problem here

→ usage of **centrally reconstructed** data would be helpful

Data Reconstruction / Analysis Chain

third step: Mapping and Calibration Processors

- easy to use, but again need set of **reference parameters / reference steering file**
- temperature depending corrections missing / **clear** documentation of the Cell-IDs missing (especially in context of Cell-IDs from Monte-Carlo)
- more documentation needed (for the processor parameters as well as for the code)
- a few, optional benchmark plots would be nice

Data Reconstruction / Analysis Chain

third step: Mapping and Calibration Processors

- easy to use, but again need set of **reference parameters / reference steering file**
- temperature depending corrections missing / **clear** documentation of the Cell-IDs missing (especially in context of Cell-IDs from Monte-Carlo)
- more documentation needed (for the processor parameters as well as for the code)
- a few, optional benchmark plots would be nice

- more documentation necessary plus a **central** point to access it

Future estimate: issues will be settled within the next few releases, larger data sets shouldn't give a problem here

→ usage of **centrally reconstructed** data would be helpful

Data Reconstruction / Analysis Chain

fourth step: run analysis on fully reconstructed data

- people use **private processors** to 'dump' to ROOT files or
- run a few steps of the analysis (e.g. cuts) in Marlin, switch then to **ROOT**
- main part is done in **ROOT**, outside the CALICE software

Data Reconstruction / Analysis Chain

fourth step: run analysis on fully reconstructed data

- people use **private processors** to 'dump' to ROOT files or
- run a few steps of the analysis (e.g. cuts) in Marlin, switch then to **ROOT**
- main part is done in **ROOT**, outside the CALICE software

- processor directly producing some 'standard (check) plots' would be nice

Future estimate: ?

Monte-Carlo / Digitisation Chain

first step: run Mokka

- (more or less) easy to use, for a new-comer confusing (data-base)
- much **more** documentation needed
- detector models for 2007 **not released** yet, models for 2006 are fine
- how to implement **realistic** beam profiles?
  - → particle gun with Gaussian smearing is simple, but no (?) other profiles available
  - → simulate beam independently and use StdHep or HEPEvt input of Mokka
  - → at the moment StdHep or HEPEvt input is limited to vertices at (0,0,0) (?)

Monte-Carlo / Digitisation Chain

first step: run Mokka

- (more or less) easy to use, for a new-comer confusing (data-base)
- much **more** documentation needed
- detector models for 2007 **not released** yet, models for 2006 are fine
- how to implement **realistic** beam profiles?
  - ➔ particle gun with Gaussian smearing is simple, but no (?) other profiles available

- improve naming conventions for the detector models, data bases and drivers
- more documentation necessary plus a **central** point to access it, perhaps also small tutorials
- overview of **all** parameters, ... which can be used in the steering files

Future estimate: documentation has increased and will increase further but Mokka is still **complex,**

simulations need to be done on the GRID to produce full statistic

- ➔ **full central simulation** would help a lot

Monte-Carlo / Digitisation Chain

second step: ganging, mapping^{-1}, noise overlay, digitisation

- similar for HCAL and TCMT, processors are available, partially still unreleased code
- issues with CellIDs after ganging (?)
- HCAL: **revision** of the whole digitisation chain nearly finished, will be released **soon**
- set of **reference parameters / reference steering file** necessary

Monte-Carlo / Digitisation Chain

second step: ganging, mapping^{-1}, noise overlay, digitisation

- similar for HCAL and TCMT, processors are available, partially still unreleased code
- issues with CellIDs after ganging (?)
- HCAL: **revision** of the whole digitisation chain nearly finished, will be released **soon**
- set of **reference parameters / reference steering file** necessary

- more documentation necessary plus a **central** point to access it

Future estimate: revision ongoing, running on full statistic shouldn't be a problem

→ **full central simulation** would help a lot

Monte-Carlo / Digitisation Chain

third and fourth step: calibration and analysis

- apply the same processors as for the data chain (see p. 11ff)

Run 'caliceMarlin'

- full data and Monte-Carlo reconstruction chains lead to **very long** and **not very clear** steering files (> 1200 lines)
- ASCII **and** XML-based steering files in use
  - → difficult to compare and to exchange
- for systematic studies, people set-up scripts to generate these steering files automatically (kind of 'meta-layer')
- **default** processor parameters often **not useful**
  - → need more documentation
  - → need central source for these parameters
- writing of ROOT files in path ./histos might cause crashes
- usage of different ROOT versions in different packages cause problems
- people run 'caliceMarlin' on the GRID, local Clusters and local machines
  - → might cause problems

Conclusions

- people are using **heterogeneous** software
  - official CALICE software, semi-official versions, private software
  - very **difficult** to compare results
- we need more **documentation** and a **central** point to access it
  - perhaps a kind of wiki, where everybody can contribute
  - and/or a portal to collect all the links to the available documentation
  - but this needs maintenance to keep it up-to-date
- usage of **central** reconstruction and simulation/digitasation would **help a lot,** if
  - everybody can access information about the processors and parameters which have been used
  - there is enough documentation ('versioning') to understand what have been done
  - this documentation is kept up-to-date
  - it is still possible to access 'low-level' information to perform studies on the mapping, calibration, ganging, digitisation etc.

Some final Remarks

Some final remarks, in **my opinion**:

- our software system is **not** in such a bad shape, all the processors needed for simulation and reconstruction are basically available
- **all** our software is based on common 'tools' (lcio, Marlin, lccd, ...)
- we mainly have problems on **information/documentation:**
  - lack of documentation itself
  - reference steering files/parameters
  - accessibility of information/documentation
  - tutorials

backup slides ...