

Why A Software Review?

- Now have experience of real data and first major analysis results
 - What have we learned?
 - How should that change what we do next time?
- Roman is pivotal to whole software effort
 - Software requires his expertise to develop
 - And also to run the code, particularly the reco
 - Should Roman really be running the reco jobs?
 - How do we reduce the burden from him so his skills can be used for higher level tasks?
- An open and detailed discussion of the software
 - Get collaboration agreement to decisions which affect everyone
 - Generate some level of documentation for code developers

Base on Collaboration Goals

- For software, the aims of CALICE are
 - To use the physics prototypes to compare data to MC physics models and find the “best” agreement
 - To use these models to develop with confidence an optimised detector both for global detector and for technological prototype design
- We must be able to compare data and MC with confidence
 - As much code as possible must be common to both
- We must be able to handle data from different conditions and places (DESY/CERN/FNAL, missing layers, etc)
 - As much as possible should be automated
- We must be confident we can apply the tuned models to the global detector simulations without extra differences
 - As much as possible should be common

Some examples of issues

- Some random things which came to mind
 - Give some idea of what we would discuss
 - But there will be plenty more
- You may have strong (good!) ideas on what to do with these
 - Informal discussions show these are not unique 😊
 - We don't want to solve these issues today; that is what the review is for
 - I hope people will come along and give their opinions...

Reconstruction

- Code from several/many developers
 - Different solutions to similar issues
 - Should be more commonality?
- Needs to be more automated
 - Dependence on steering files is very high
 - Difficult to be sure it is right (even for Roman); almost impossible for users
 - Move all parameters into database or reco files?
- Data and MC reco should be common as far as possible
 - Need same “semi-raw data” format for data and MC for every detector?
 - Need MC reco for all items, e.g. trigger?
 - Should data be converted (=mapped) to physical position indices or MC to electronics indices?

Systematics

- We should be systematics limited
 - Else we should have taken more data!
- Every analysis will need to do systematic studies
 - Many will involve varying parameters used in reconstruction, e.g. what is effect of uncertainty of ECAL calibration on analysis result?
 - There could be many such effects to be studied in each analysis
- Must be guaranteed to reproduce original reco results if no change
 - Needs all cuts, constants, etc, used in reco to be available
 - Many cuts in steering files; others in database (but they could have been updated after reco was done)
- How should this all be organised?
 - Users rerun whole reco from raw data for each systematic?
 - Rerun just specific part of reco for each?
 - Produce centrally standard variations for each run?

Databases

- We are using a database with a “conditions” data structure
 - Entries are organised as values (e.g. temperature) at time t
- Tools implicitly assume only one experiment
 - Only one value for each item at time t ; unique database folder set at start of job
 - We have simultaneous values because working at different sites (DESY/CERN/FNAL) so must subdivide database and use correct folder; not known until data file read in after start of job
 - We also use it for MC values; also not known if data or MC until data file read
- Currently handled by specifying folders by hand depending on run being used
 - Should we rewrite tools?
 - Read file and restart job?
 - Should we use a different database? ☹

Databases (Cont)

- We are using the database also for storing “configuration” data
 - Related to run structures, i.e. more like how we work
 - Imposing differently structured data is not always easy
 - E.g. Beam energy for run x? Need to scan times to find when run x happened and then find beam energy; takes minutes per run
- Should this be changed? How?
 - Have second database with “configuration” data structure?
 - Do lookup once for every run and store in run header in reco files?

Global detector studies

- We have (uniquely?) the whole breadth of knowledge from the detector hardware data to PFA implementation within CALICE
 - How do we optimise our contribution to concept groups?
 - LDC is clear, SiD is not (GLD I don't know...)
- Use common (Mokka) simulation?
 - One implementation of detector model but may have less impact in concept meetings
- Use concept group “native” simulations?
 - Requires two independent implementations; are they guaranteed to be equivalent to each other and the beam test results?
- This may become more critical as detector concepts move to collaborations...