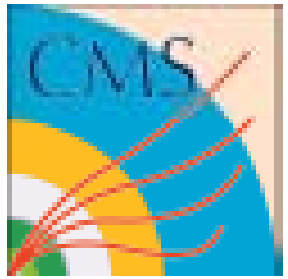


Laser Monitoring Data Processing

ECAL DPG Meeting, CERN

Adi Bornheim
California Institute of Technology

May 07, 2007

CMS Laser Monitoring Procedure

See talk March 12 2007

1. **Online Data Flow**
   Everything from the hardware level to handing the data over to the Laser Farm.

2. **Laser Farm**
   Processing of the raw data and inserting results into the online DB

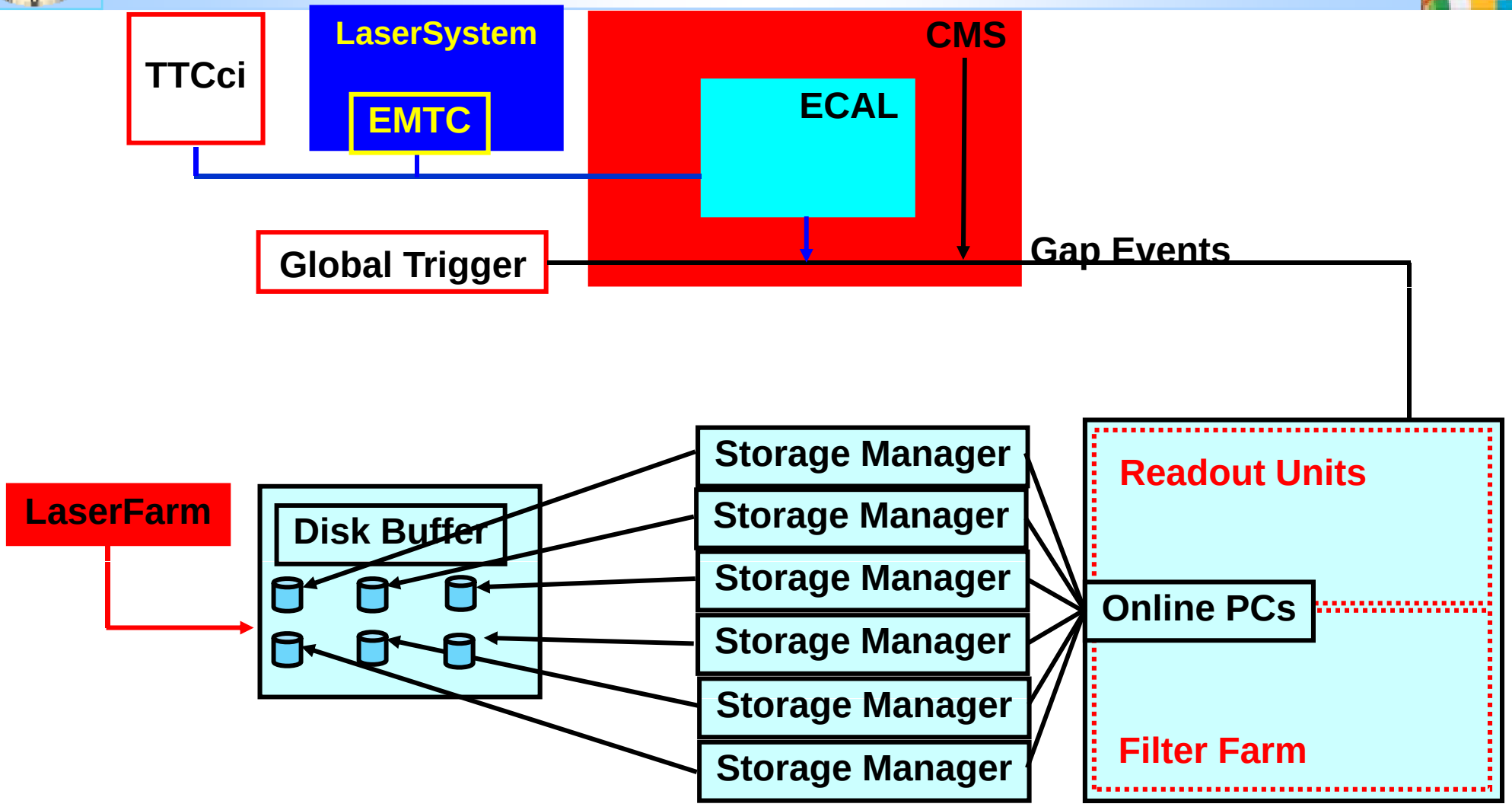
Today mostly this …

… and a little bit about this.

3. **Offline Transparency Correction**
   Online to Offline transfer of APD/PN ratios and physics event correction of the transparency change.

+ some discussion on the sharing of tasks to be run at P5 and elsewhere, like TIER0/CAF.

Recap : Online Monitoring Data Flow



Gap Event Data is temporarily stored on disk buffer at P5, distributed over multiple files - one per storage manager per stream. In this format it is shipped out of P5.

Structure of the Monitoring Data

- Current standard is 600 events per crystal on each half module. These events correspond to one transparency measurement. They have to averaged to get one APD/PN ratio per channel.
- At 100 Hz it would take 6 seconds to record that many laser events.
- ~88 ‘half modules’ in the entire ECAL.
- To scan the entire ECAL in 20 minutes we have to scan one half module every ~15 seconds.
- At the test beam one half module worth of laser events corresponds to 13 Mb raw data. This would correspond to ~3.2 Gb/h which is ~40Gb/12h for the entire ECAL.
- For processing the laser data these 600 events have to accumulated in a single place eventually.
- Beyond this, there is no natural structure. In particular, completing one cycle over the entire ECAL does not correspond to any significant change in the data flow or processing need.
- I strongly favor to process the data in chunks of 600 events, corresponding to one transparency measurement.

Matching Data Structures to Data Processing

- **Raw data online on disk buffer is distributed over multiple files. The data needs to be merged to get it processed.**
- **Offline, for processing on TIER0, a 'repackaging' tool exists which merges the files back together to one stream. This stream can then be written to files. Details need to be investigated – possibly this tool can be used for our purpose. Using an existing tool possibly helps in keeping the processing flexible.**
- **Since merging and repackaging is done anyway it would be best to tailor the output of this step to our needs : Separate ECAL laser event data from the gap stream and write it in files corresponding to the amount of data we want to process in one job. For 12 hours of data taking this corresponds to 4x60x12 = 2880 files.**

Processing Laser Monitoring Data

- Processing will very likely be CPU limited – it always was at the TB. Presumably recent improvements in the code (eg. unpacking) will not change this.
- Main task of this CMSSW job : Reconstructing raw amplitudes for APDs, PNs and average over 600 events. The result of this I call 'Raw APD/PN ratios'.
- Process each file in a separate CMSSW job to avoid creating any dependency.
- These results should be stored in the online data base.
- The stability of the raw APD/PN ratios was found to be good (few ‰) in 2006 TB, but not optimal (below 1 ‰).

This part consumes the bulk of the CPU, will likely not change significantly in the future, has no correlations among channels, SMs, vs. time etc.

There are (is a) known correction(s) to these APD/PN ratios :

- Pulse Width correction.
- Others ? Given the experience from the test beam over many years we should foresee the possibility of additional corrections – even if they are not fully defined yet.

Further manipulation of the raw APD/PN ratios :

- They have to be normalized to each other for all channels.
- Interpolation between laser runs, beginning and end of an LHC fill, missing laser runs.

This part consumes very little CPU, will likely change in the future, has correlations among channels, SMs, vs. time, luminosity etc.

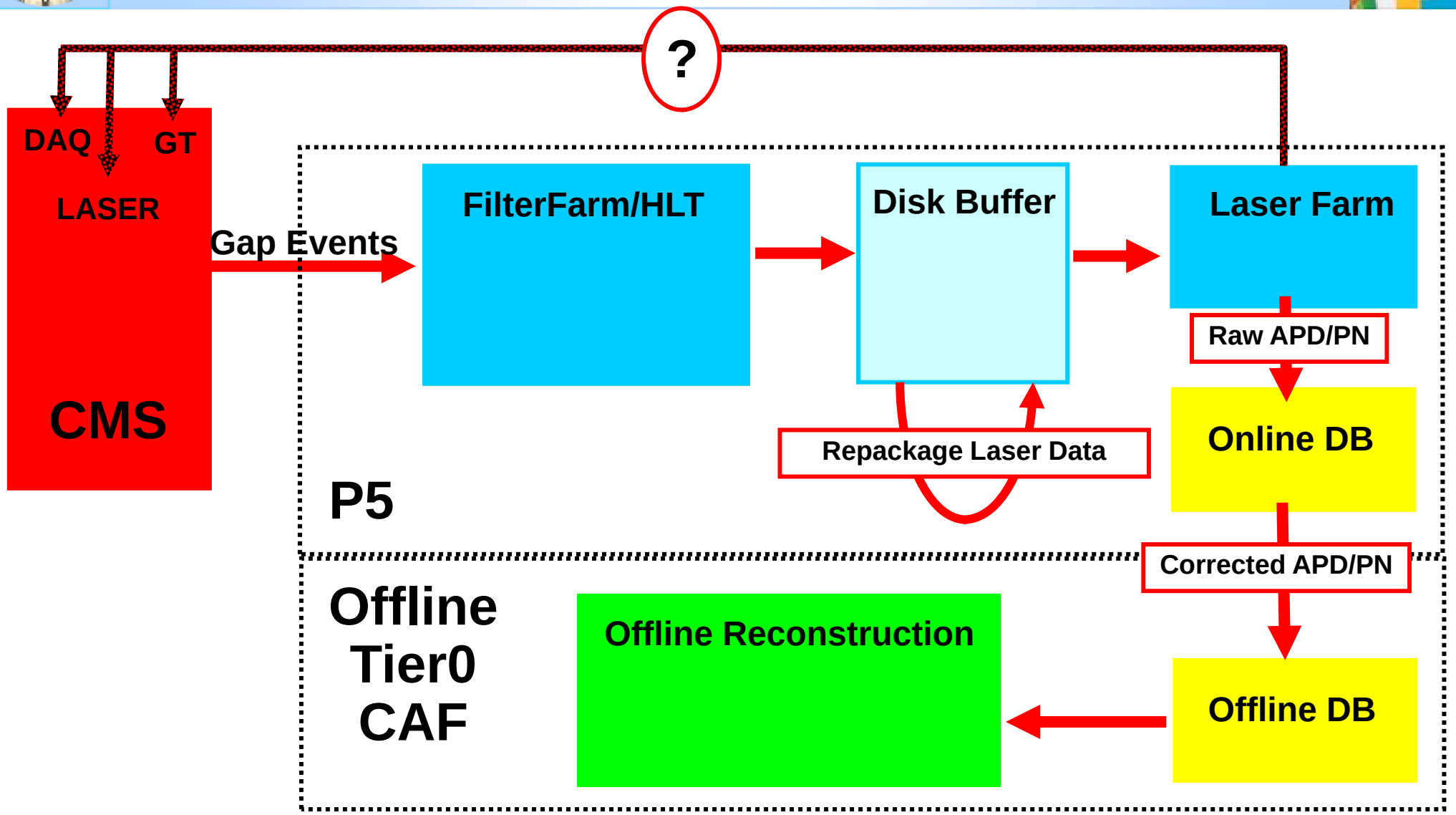
⇒ Decouple both steps !

Laser Monitoring and CMS data bases

Current plan :

- **Store raw APD/PN ratios in the online DB immediately after processing, possibly even after each single CMSSW has finished. Procedure is currently being optimized.**
- **The structure of the online data base is well suited for monitoring the monitoring results (history plots etc.). Partial copies of the online DB can easily be made for more detailed offline studies. The offline data base is quite flexible regarding the granularity of writing and reading. Corrections could be applied by reading and writing back corrected APD/PN ratios.**
- **The online-to-offline copy of the data bases will involve a step in which the data is available in a c++ structure. Global corrections and interpolations to the APD/PN ratios can be applied in this step – possibly more efficiently than reading and writing back to the online DB.**
- **The offline data base is optimized to retrieve all the information for a single physics event. The monitoring results have to be tailored such that the correction for each physics event can be calculated from the DB entry valid for this physics event.**

Laser Monitoring Workflow



Laser Monitoring and DQM

Using the laser monitoring for ECAL DQM tasks :

- Laser monitoring is a powerful tool to monitor general ECAL performance.
- Dividing line between DQM and transparency monitoring task is not fully defined.
- ECAL DQM will only need a fraction of the laser events – which fraction will depend on what is still considered a DQM task : Channel sees laser (y/n) ?, channel changed response by x% ?
- DQM will run 'real online' – possibly in event consumer mode.

Currently it seems to me that the way the DQM is being planed it will have no impact on the way we process the laser monitoring data - and vice versa.

Using the CMS DQM tools to monitor the monitoring :

Since there is a centrally developed DQM service at P5 we should investigate to use this. This requires to run the processing at P5. We need to have some direct feedback at P5 to monitor the proper running of the monitoring – both hardware and software.

Only the laser monitoring processing results will tell us if the monitoring system is working properly – the ECAL DQM using the laser will not !

Laser Monitoring and DAQ

Laser monitoring and central DAQ :

The details of the monitoring sequence are only know to the TTCi board and the EMTC card, not to central DAQ. As far as I know there is no centralized record of which SM has been monitored when with which laser setting etc.. Some bookkeeping has to be implemented to keep a record of all this. After a data taking interruption, the monitoring cycle probably should continue on the same SM it has stopped.

Updating HLT constants online ?

Needs to be studied if this will ever be needed. For initial implementation not an issue.

Laser monitoring and Laser DAQ/EMTC :

- So far, laser intensities etc. have been set manually to a suitable value. This should probably be automated at some point.
- Feedback for timing ?
- Safety feedback ?
- Quality feedback to raise alarms if the monitoring hardware performance deteriorates.

Local DAQ operation :

We need the ability to run the monitoring in an ECAL standalone mode out of LHC runs.

Laser Monitoring Farm Hardware

- We estimate that ~12 PCs are needed to do all the processing of the laser monitoring data. This is a small fraction of either the PC farm at P5 or TIER0.
- To optimize the processing, the PCs would run in a quasi standalone mode at P5, taking advantage of the connectivity to the raw data stream and other online applications.
- How the special needs of the processing would fit into the TIER0 scheme has not been looked at. There are certainly some special arrangements needed, eg to guarantee that the laser data is processed before any physics data. There is no connectivity to online applications.
- The availability of the farm at P5 is presumably higher than the availability of the TIER0 cluster. Whenever data is taken P5 has to be available.
- The core functionality of the CMSSW code should not depend on whether the code runs at P5 or TIER0.
- We must have the possibility to analyse the raw data at P5 as well as TIER0.

Summary

- **The base line for the laser data processing is to run multiple jobs in parallel at P5 over files tailored to the needs of the laser monitoring tasks. This processing is done on a dedicated set of PCs. Results get stored in the online data base. Corrections get applied in the online-to-offline copy.**
- **We should aim to keep monitoring code such that it can run either at P5 or at TIER0. Functionality only needed at P5 should thus be modular so that it can be disabled.**
- **In order to work efficiently towards a functioning system we should focus on one base line.**
- **We should aim to implement as much of the entire scheme as possible as soon as possible. Only running the system will tell us where the bottlenecks really are.**