Calibration Work-Flow and Data-Flow in CMS

Luca Malgeri on behalf of the CMS offline group

CHEP'07 - Victoria (CA)

Outline

- What is intended by "calibration"
- Storage of calibrations (a.k.a. DataBase)
- Producers/consumers of calibrations (online/offline)
- Commissioning of the calibration workflow
- Few results

Calibration: a definition

Whatever is needed to go from electronic readout to physics quantities:

Energy in calorimeters: $$E_i = \sum_{i}^{channel} c_i \cdot ADC_i$$

Alignment in tracker: $$x, y, z = \Delta_{x,y,z} + x_{nom}, y_{nom}, z_{nom}$$

see *OC2 "Alignment strategy for the CMS tracker", M.Weber*

By extension it includes:

Low-level: pedestals, noise, gains,.....
High-level: jet energy scale, electron energy, tau jets, ...

(CMS) Calibration workflow design

Requirements:

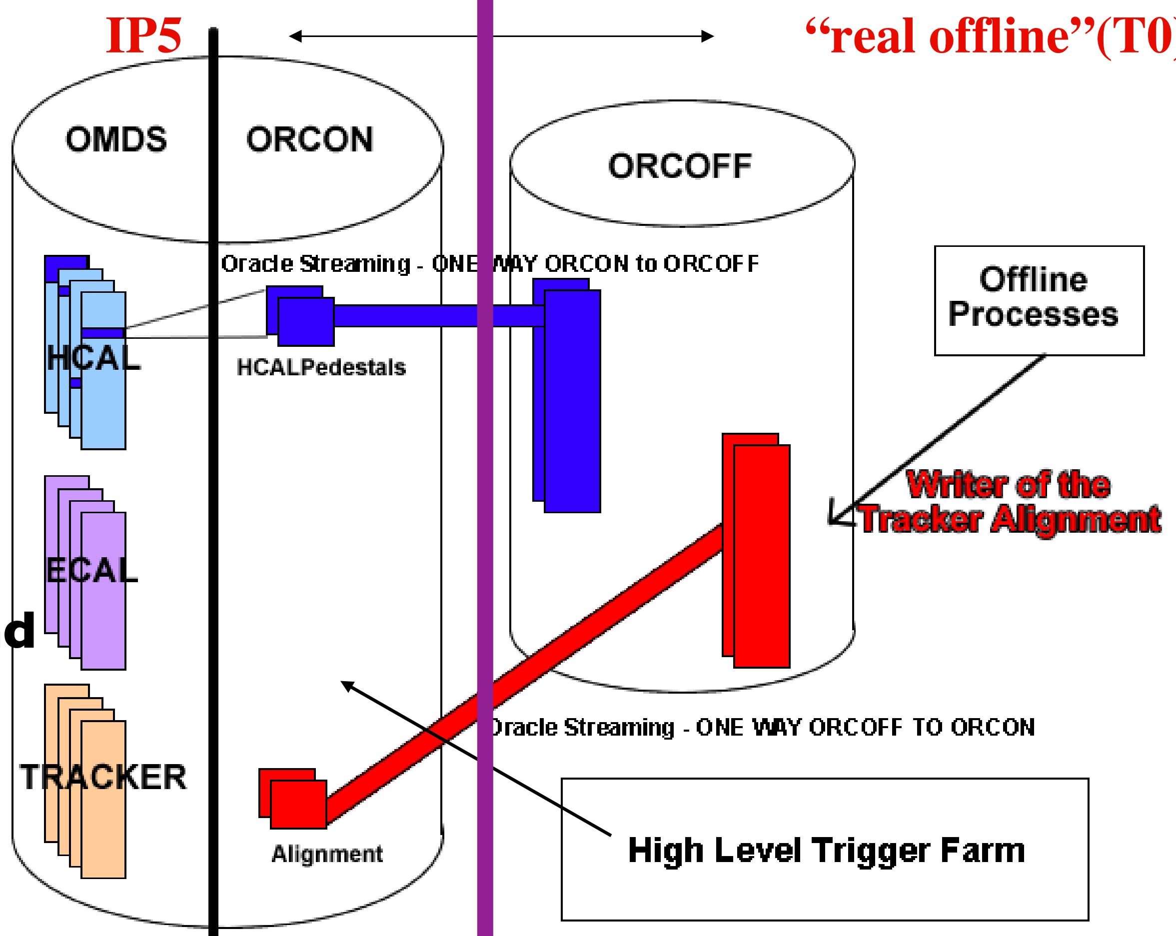
- Full traceability → need to know when constants are produced and when have to be used
- Handle large data sets (several millions channels)
- Cope with potentially fast turnaround of constants
- Ability to work in online as well as in offline world (including remote Tier1/Tier2 sites)
- Ease of use is a envisageable

Storage of constants: DataBases

CMS experimental area

Tier0 - Meyrin site

OMDS:
Full Oracle DB
Serves Lv1 trigger



ORCON:
Oracle Back-end
Pool-ORA
Serves HLT trigger

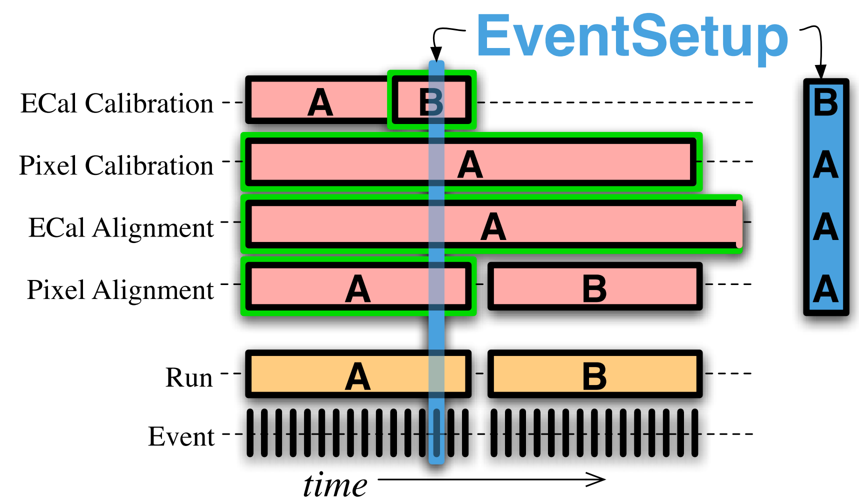
ORCOFF:
Oracle Back-end
Pool-ORA
Serves offline reco

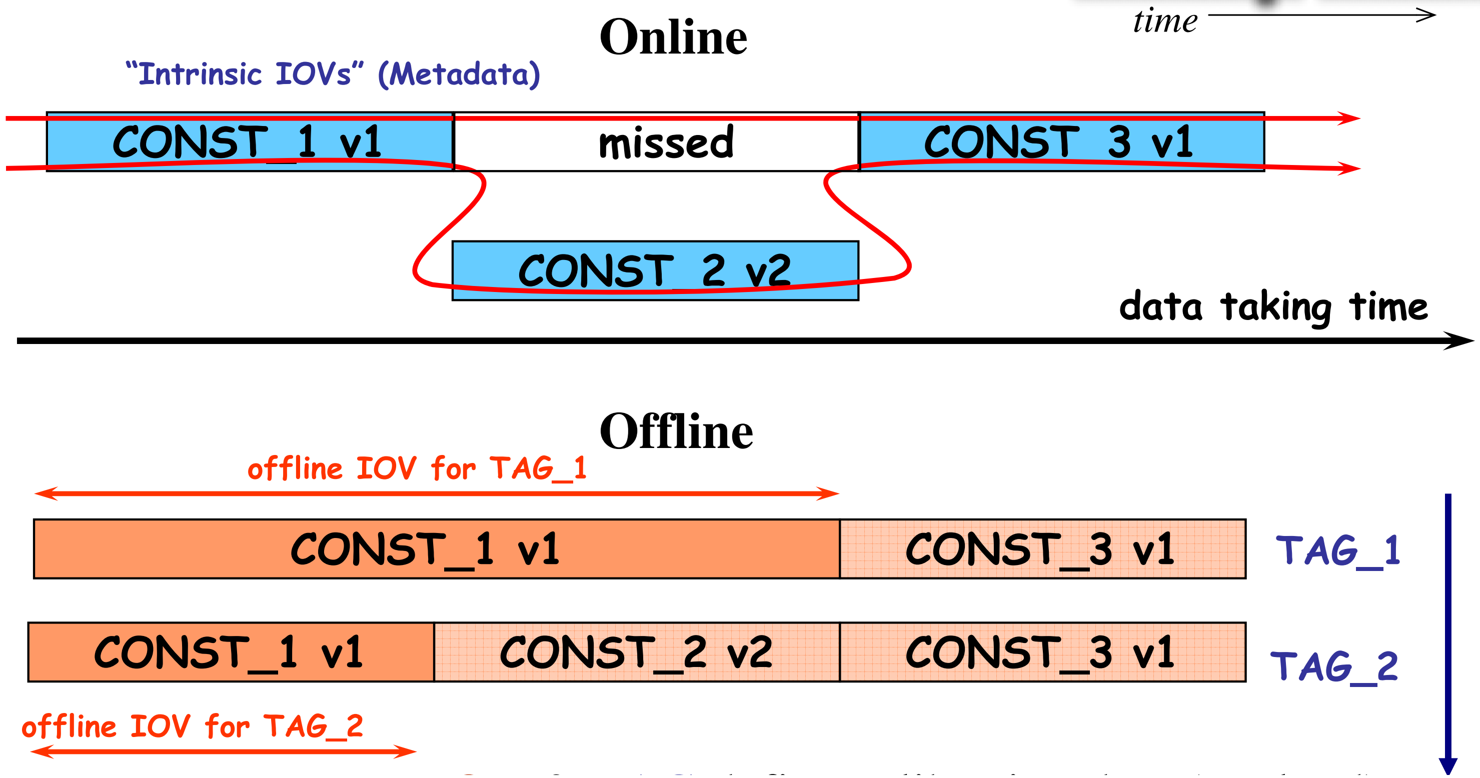
DataBase technology

- Oracle (full support from CERN IT) is used as back-end for all DB layers.
- Pure "online" tasks have direct access to OMDS to continuously store calibration and monitoring constants: pedestals, noise, temperatures, firmwares, etc.
- ORCON is, topologically, a subset of ORCOFF needed for HLT functioning ensuring CMS autonomy from network disruption.
- ORCON and ORCOFF are accessed through POOL-ORA:
  POOL-ORA is a LCG-AA product which provides a mapping from the relational DB to a C++ object.
- The constants tracking, and interval of validity, is integrated in the Event Data Model (CMS software).
- Constants are transferred from OMDS to ORCON and from ORCON to ORCOFF through C++ applications or direct DB-link.
- DB deployment to external sites provided through Frontier technology

Interval of Validity: concept

The Interval Of Validity in CMS is an **intrinsic** characteristic of the Event Data Model. It is "attached" to all Event Setup objects, including the calibration DB objects.

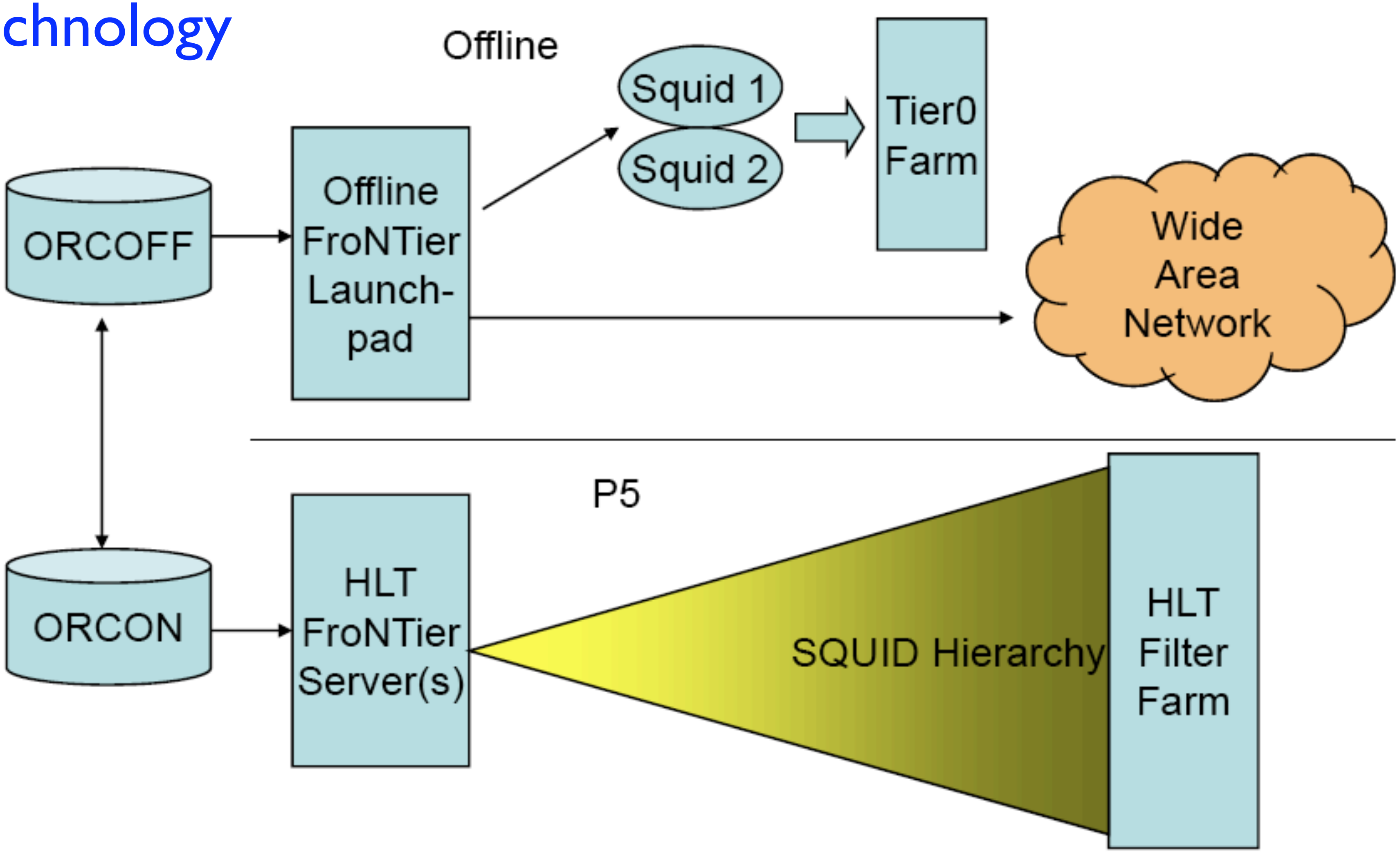




DataBase deployment

Frontier:

intermediate caching (read-only) layer based on http-proxy technology



see *DD4 "CMS Conditions Data Access using FroNTier", L. Lueking*

DataBase deployment

Frontier technology will be used also to distribute constants online

- HLT
  - Startup time for Cal/Ali < 10 seconds.
  - Simultaneous
  - Uses hierarchy of squid caches
- Tier0
  - Startup time for Cal/Ali < 1% of total job time.
  - Usually staggered
  - DNS Round Robin should scale to 8 squids

Parameter	**HLT**	**Tier0**
# Nodes	1000	1000
# Processes	~8k	~3k
Startup	<10 sec all clients	<100 sec per client
Client Access	Simultaneous	Staggered
Cache Load	< 1 Min.	< 1 Min
Tot Obj Size	150 MB*	150 MB*
New Objects	100% / run*	100% / run*
# Squids	1 per node	Scalable (2-8)

Producers of “calibration”

Four different “calibration data” typologies have been individuated:

1) Online defined:
   calibrations executed in the pit and directly used in Level1/HLT: pedestals, gains, etc.

2) Non-event runs:
   special runs with calibration sources (lasers, injected pulses, etc.) or physics runs with special conditions (without zero suppression,etc.)

3) “Pre-HLT” defined:
   calibrations needing reduced information from large datasets not saved in physics streams: minimum bias, $\pi 0$, Level 1 primitives.

4) Physics-event runs:
   calibration streams taken during normal runs (single electron, $Z \rightarrow \mu\mu$, $Z \rightarrow ee$,etc.)

Producers of "calibration"

Four different "calibration data" topologies have been individuated:

1)Online defined:

Little CPU usage and immediate availability.
→Run at the pit.

2)Non-event runs:

3)"Pre-HLT" defined:

4)Physics-event runs:

Producers of "calibration"

Four different "calibration data" topologies have been individuated:

1)Online defined: Little CPU usage and immediate availability. →Run at the pit.

2)Non-event runs:

3)"Pre-HLT" defined:

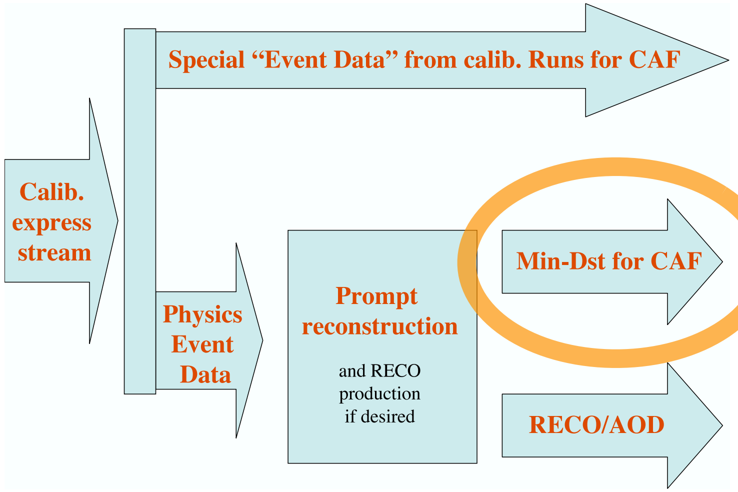
4)Physics-event runs:

Moderate to high CPU usage.

Need to:

- reduce amount of data
- use Tier0/offline resources

Producers of calibration (inputs)



•Reduce amount of data

Store only the event information needed to perform the calibration while keeping the compatibility with the event data format.

Size of event greatly reduced
Access rate greatly improved
→ fast turn-around

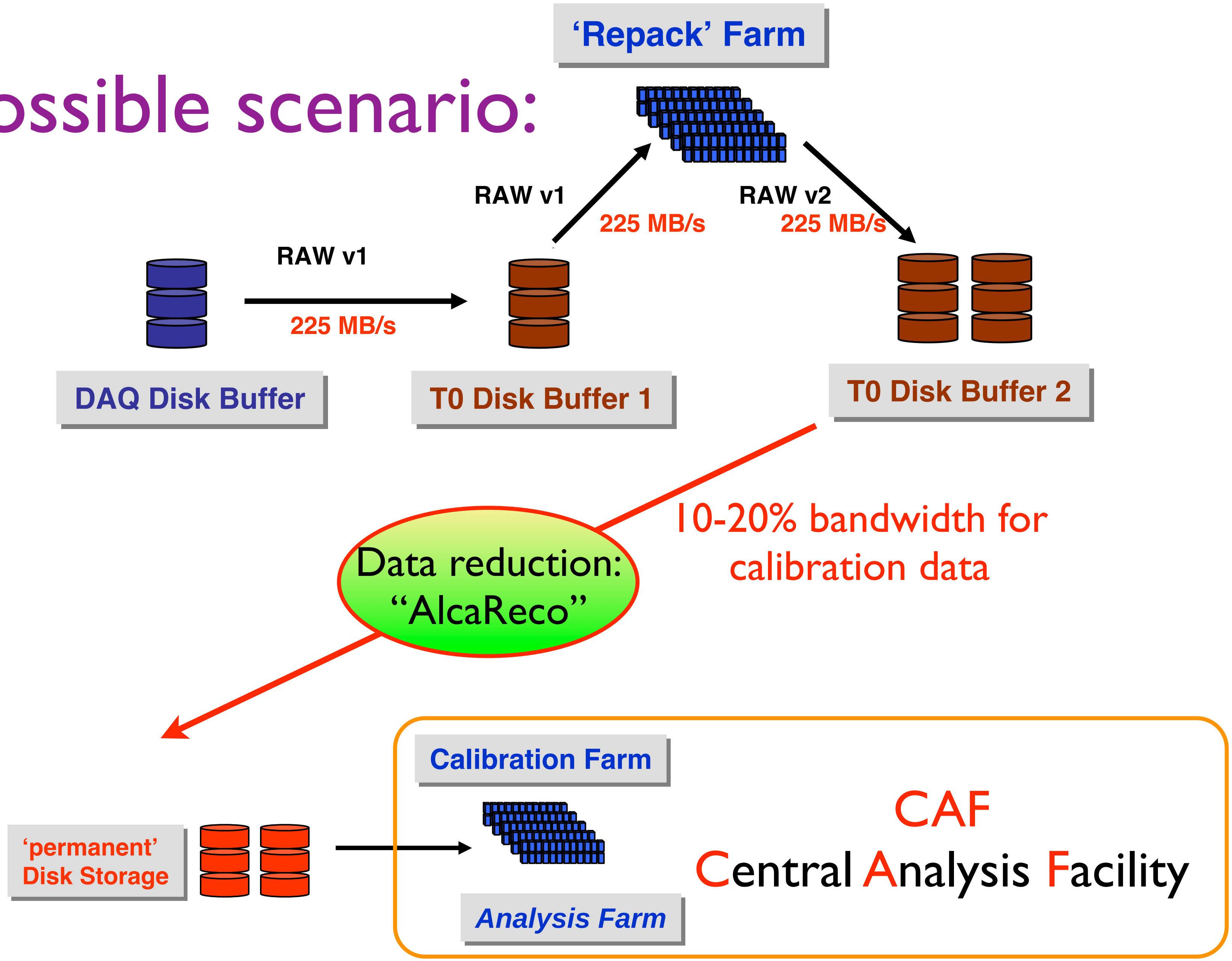
ECAL physics events calibration					
Method	Input	Size/ev (kB)	Rate (kB/s)	Running Freq.	Update Freq.
Phi Symmetry	Pre-HLT	0.2	200	Continous	Every day at start-up
Low-M res.	Pre-HLT	0.2	100	Continous	Every day at start-up
Single electron	AlCa-RECO	3.0	7.5	Days/Week	Week/Month
$Z \rightarrow ee$	AlCa-RECO	6.0	0.6	Days/Week	Week/Month
$Z \rightarrow \mu\mu\gamma$	AlCa-RECO	5.0	0.2	Days/Week	Week/Month

•Use Tier0/offline resources

A dedicated cluster for fast analysis/calibration: CAF
Express streams (including calibration streams) readily available.

Calibration stream(s) data-flow

A possible scenario:



Commissioning of the calibration workflow

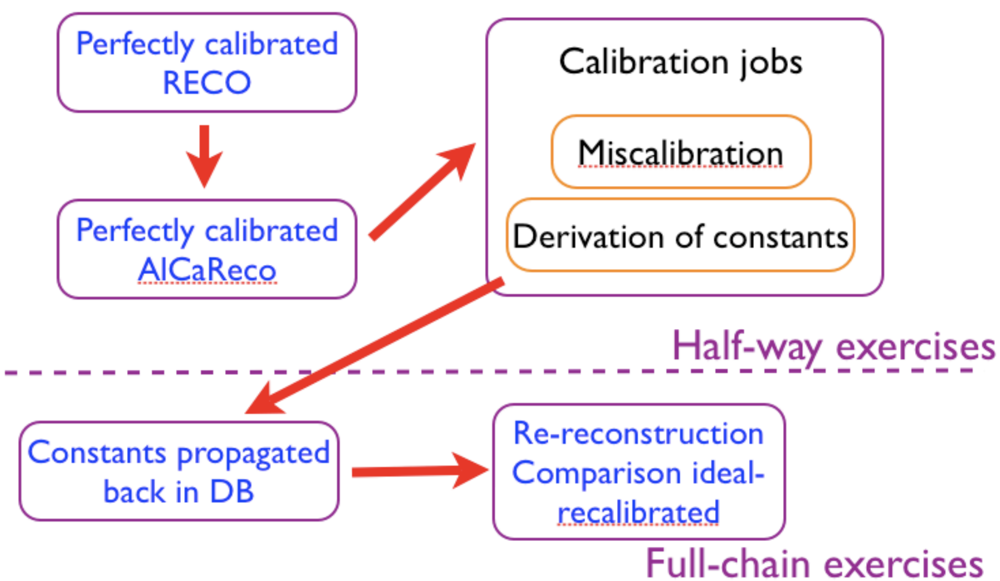
Several "staged" tests of the workflow have been (and are being) exercised:

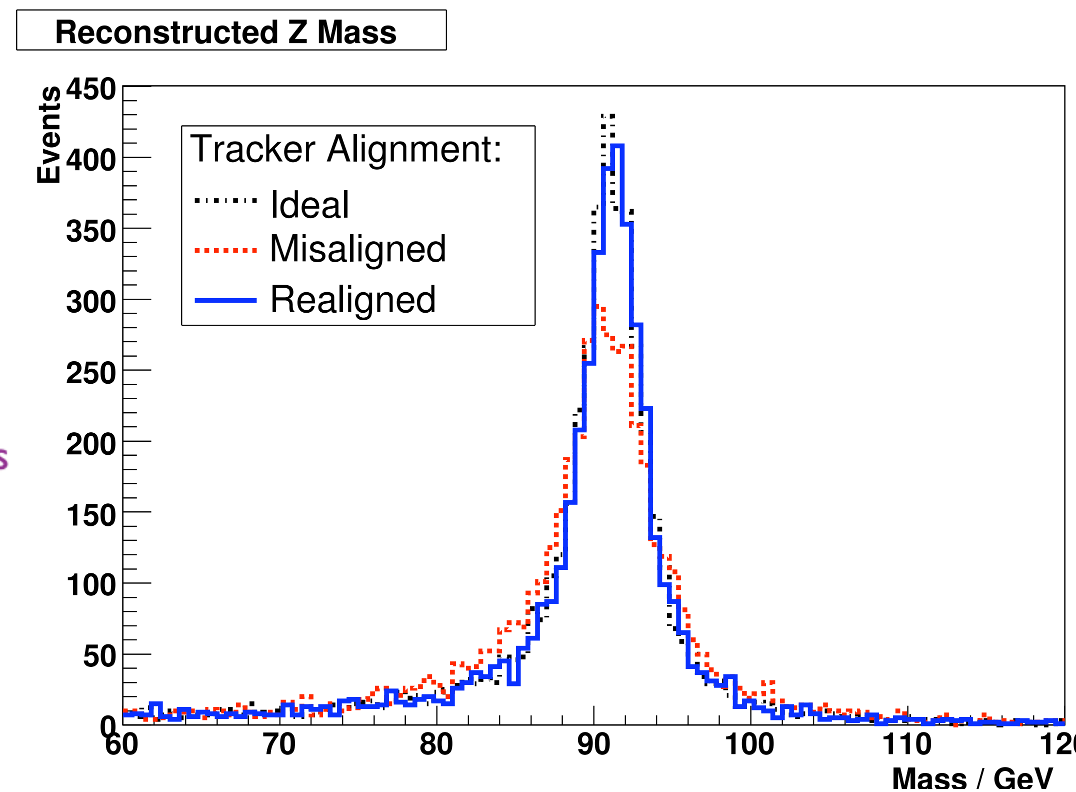
- CSA2006: Computing, Software and Analysis challenge 2006
  A 25% scale prototype of CMS data processing
- MTCC: Magnet Test and Cosmic Challenge
  Magnet switched on and several millions cosmics taken in global runs
- ECAL and HCAL test-beams: test of calibration algorithms
- TIF: Tracker Integration Facility (full tracker DAQ on cosmics)
- Large MC samples for algorithm improvement
- CSA07: large (>50%) scale prototype

Few selected results in the following →

CSA2006: several components in place

- Mis-calibrated samples (reproducing expected accuracy)
- AlCaReco production for selected streams
- DB usage and deployment "limited"
- Algorithm performances
- Re-Reconstruction after calib/align

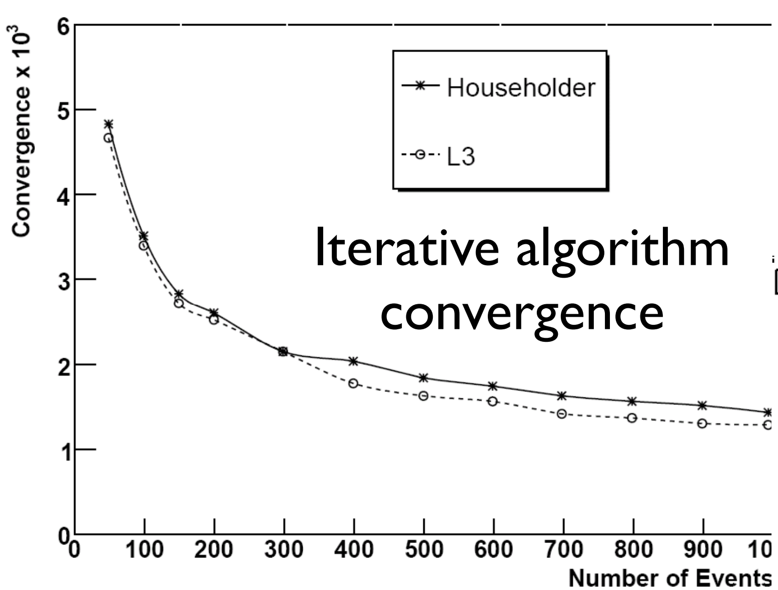


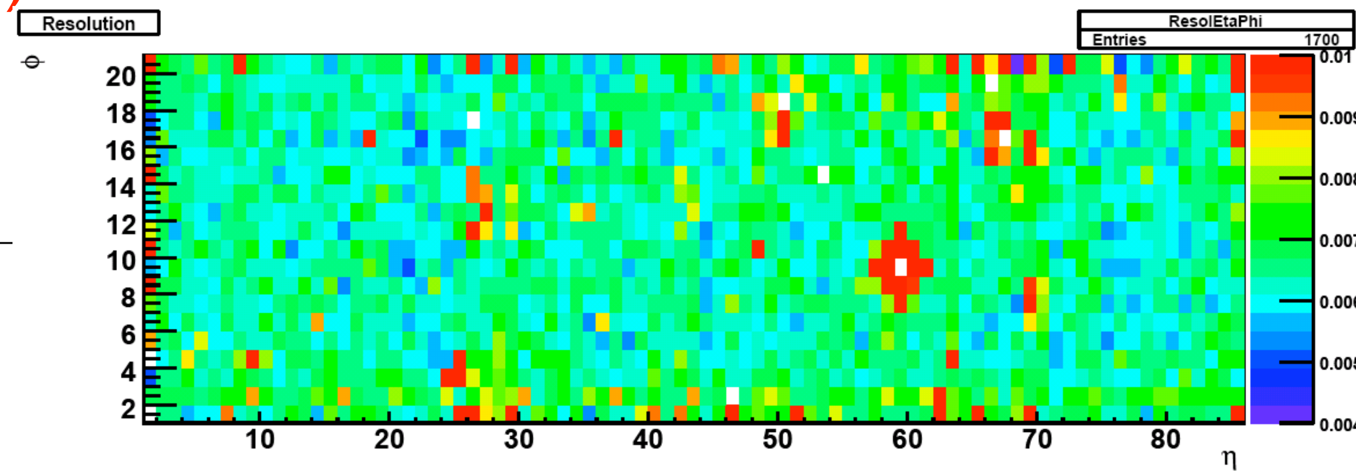


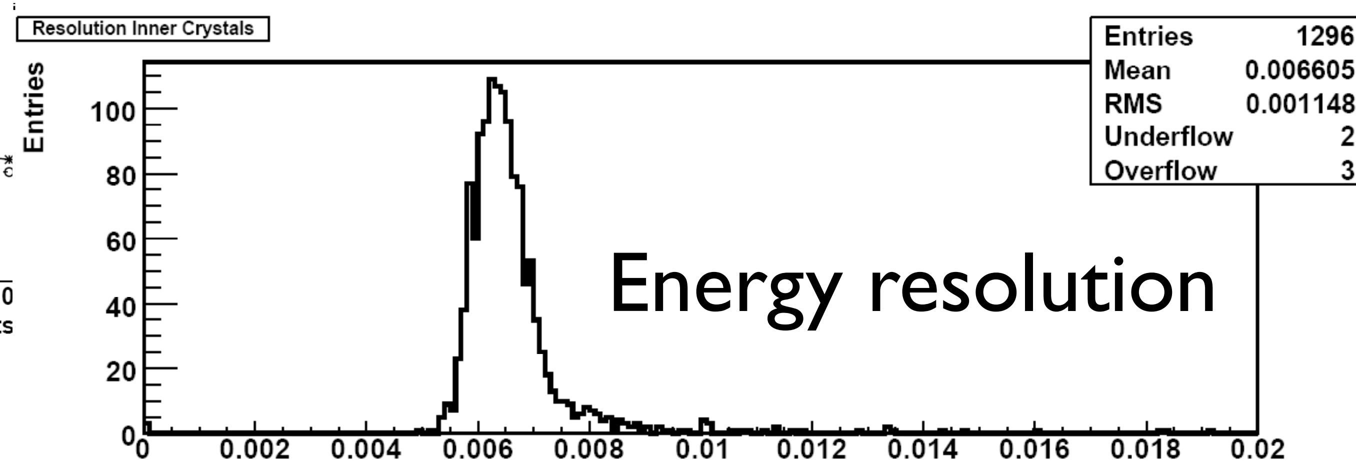
Commissioning of the calibration workflow

ECAL test-beam 2006: >15000 ECAL crystals calibrated

- Electrons of 120 GeV used to calibrated 25% of the entire ECAL (calibration at the startup).
- Maximum performance (0.6% resolution) reached.
- Same algorithms (code) as for "in-situ" calibration have been used.







Commissioning of the calibration workflow

CSA2007: large scale prototype

- More than 100M events will be generated
- All calibration constants read from DB and applied in reconstruction as for real data-taking
- Simulation of start-up mis-calibration/ alignment
- Work-flow and data-flow (AlCaReco and CAF) as for real data taking. $\mathcal{O}$(day) for first round-up and re-reconstruction.

ECAL

- Calibration from single electrons (W->ev, full workflow)
- Calibration from electrons in Z->ee
- Calibration from phi-symmetry (dataflow/workflow)
- Calibration from Pi0s (study and initial dataflow/workflow)

HCAL

- Calibration/monitoring from Phi Symmetry
- Calibration from isolated tracks (energy scale and energy dependence)
- Calibration from dijet events (calibration regions outside tracker coverage)
- Calibration from gamma+jet events (second order corrections)
- Calibration from muons for HO (CSA07 includes filter, analyzer outside CSA07)
- Calibration from Z->ee (development outside CSA07)

Tracker

- Alignment using muons from Z->mumu (full workflow)
- Alignment using charged pions (study)
- Laser alignment simulated data (study)

Muon

- Alignment using muons from Z->mumu (common AlCaReco for DT and CSC)
- RPC calibration stream for monitoring detector performance
- DT drift velocity calibration exercise (contacts: Sara Bolognesi, Silvia Maselli)

The challenge is starting in a few days, results expected soon.

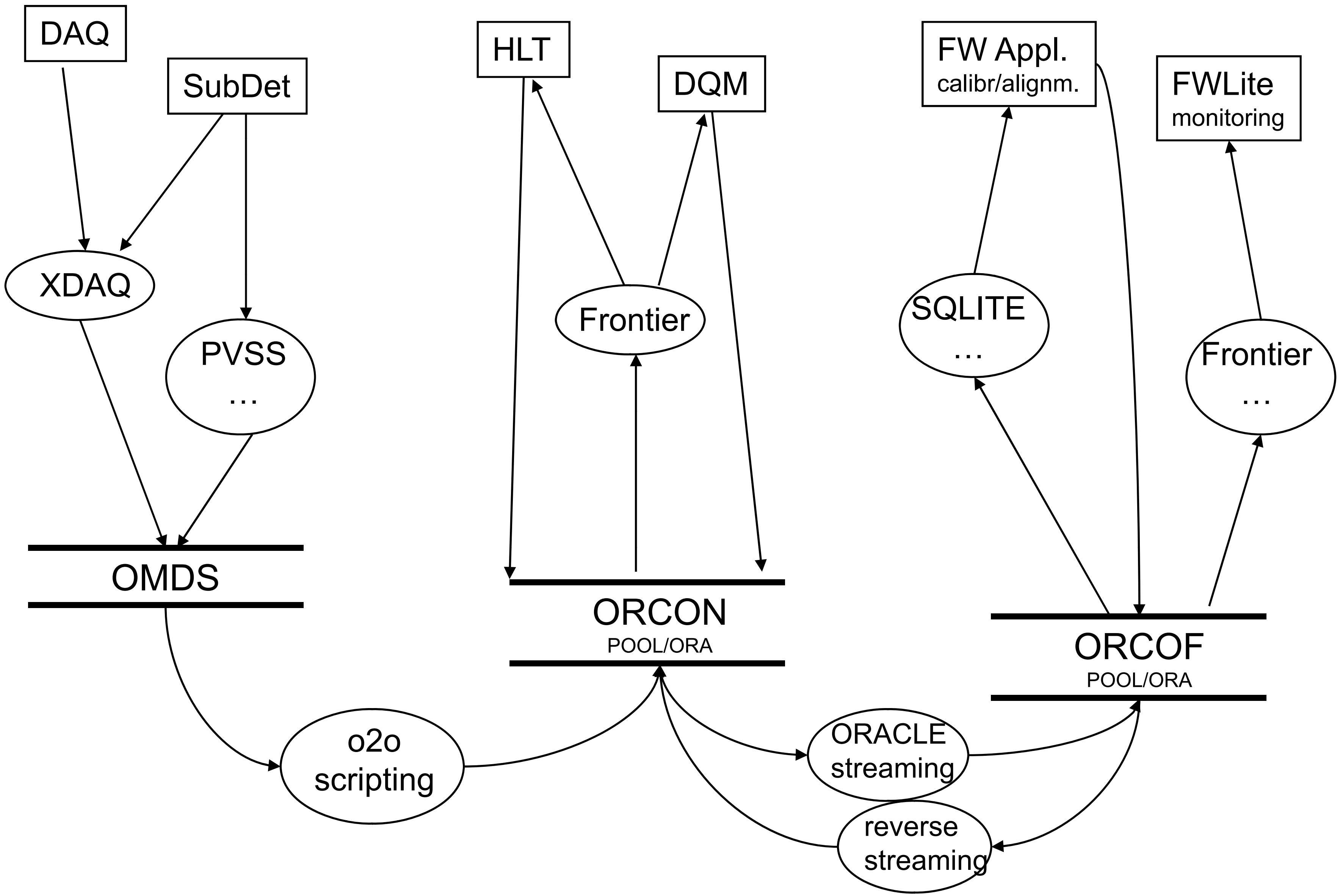
Conclusions

- A complete work- and data-flow of calibration procedures has been designed in CMS having in mind:
  - flexible access (driven by usage/physics needs)
  - robust deployment to remote sites
  - full traceability and integration in the Event Data Model
- Several staged test of the workflow have been successfully performed
- System is getting ready (from prototype to production)

Thanks for contributions: O. Buchmüller, L. Lueking, V. Innocente, F. P. Schilling, S. Beauceron

BACKUP

DB Data-Flow



DAQ
SubDet
XDAQ
PVSS
...
OMDS
HLT
DQM
Frontier
ORCON
POOL/ORA
FW Appl.
calibr/alignm.
FWLite
monitoring
SQLITE
...
Frontier
...
ORCOF
POOL/ORA
o2o
scripting
ORACLE
streaming
reverse
streaming

DB predicted storage

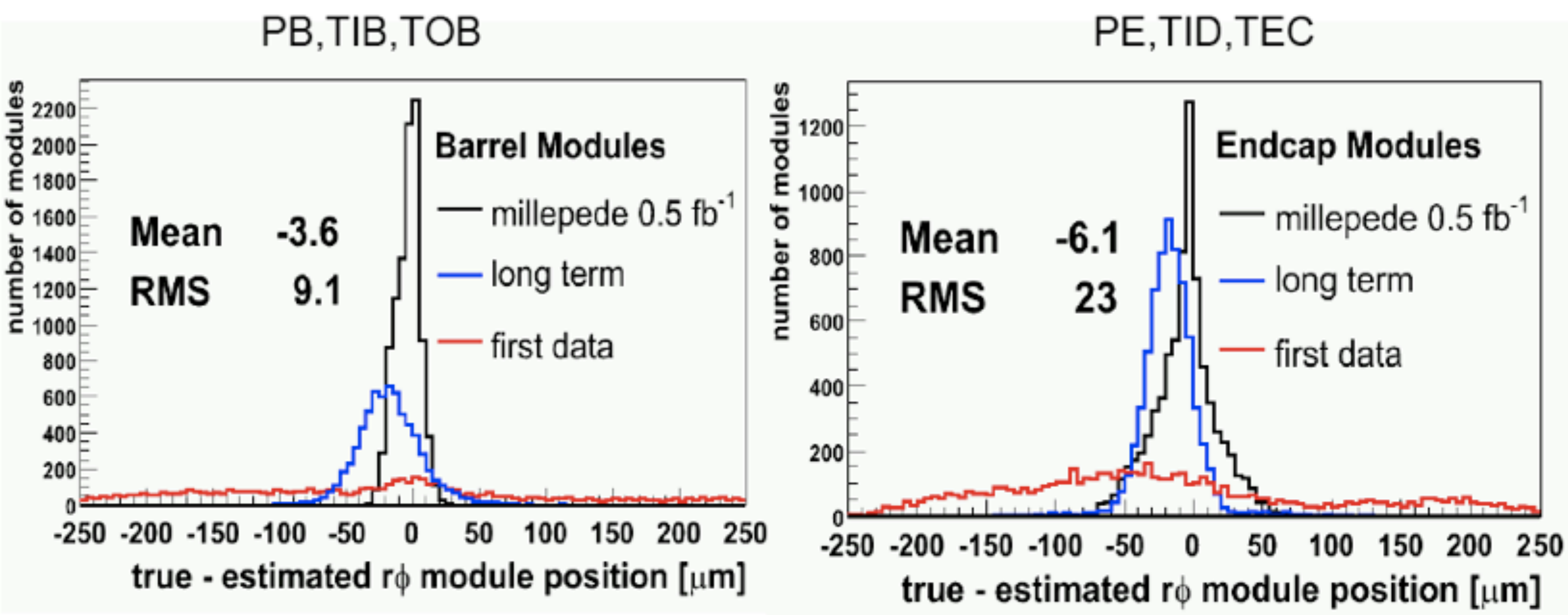
OMDS

	CondDB	Config	ORCON	ORCOFF
PIXEL	< 80	< 20	-	**< 50**
STRIP	**1.800**	43	10	10
ECAL	460	1	31	31
HCAL	23	0.2	0.5	0.5
CSC	60	0.1	**60**	**60**
DT	300	1	15	15
RPC	30+?	1+?	-	-
Lumi	200	?	-	-

Commissioning of the calibration workflow

Full tracker alignment using millepede iterative algorithm

Cosmics and single muons of 2 mio. Z^0 events used.



Barrel Modules RMS = 9 μm

Pixel barrel alignment significantly better than in the long term scenario.

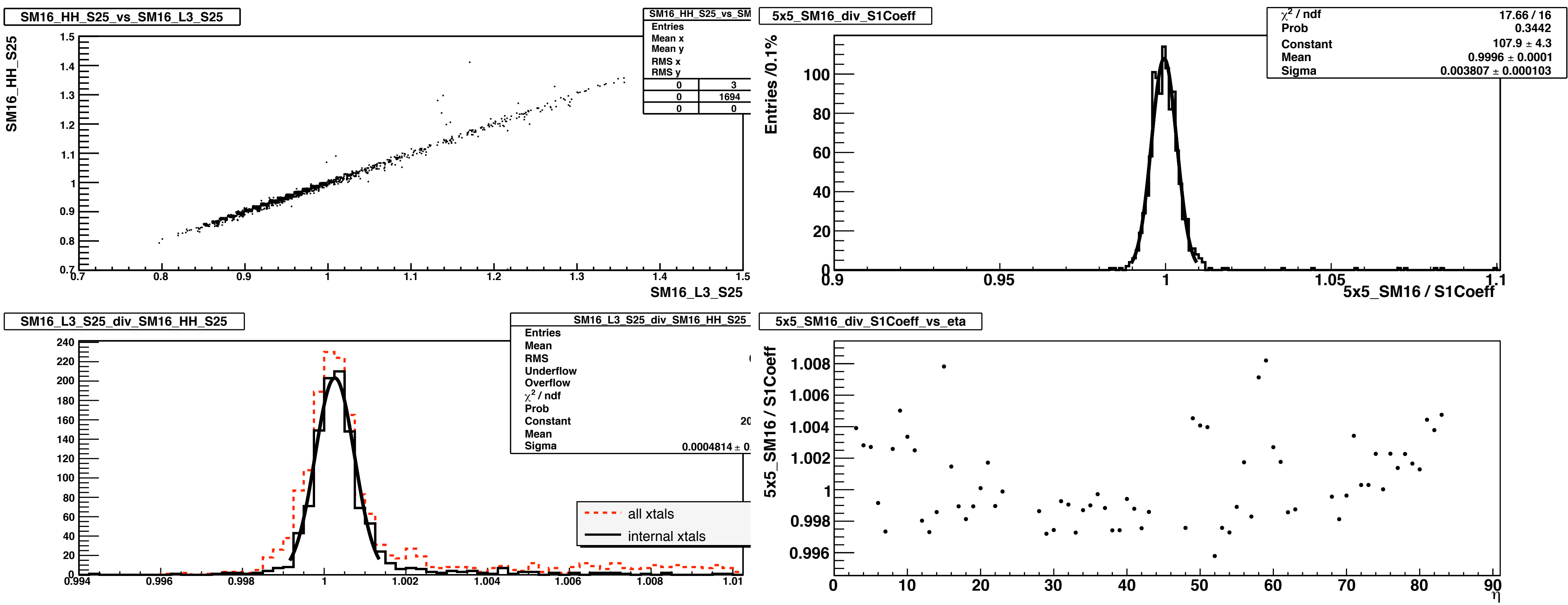
Endcap Modules RMS = 22 μm

The mean is better than the longterm scenario! The RMS is similar.

Milestone reached!

Markus Stoye, Hamburg

ECAL calibration in test-beam



ECAL calibration in CSA2006