# Computing Performance and Software Status

during/for CERN07 data taking

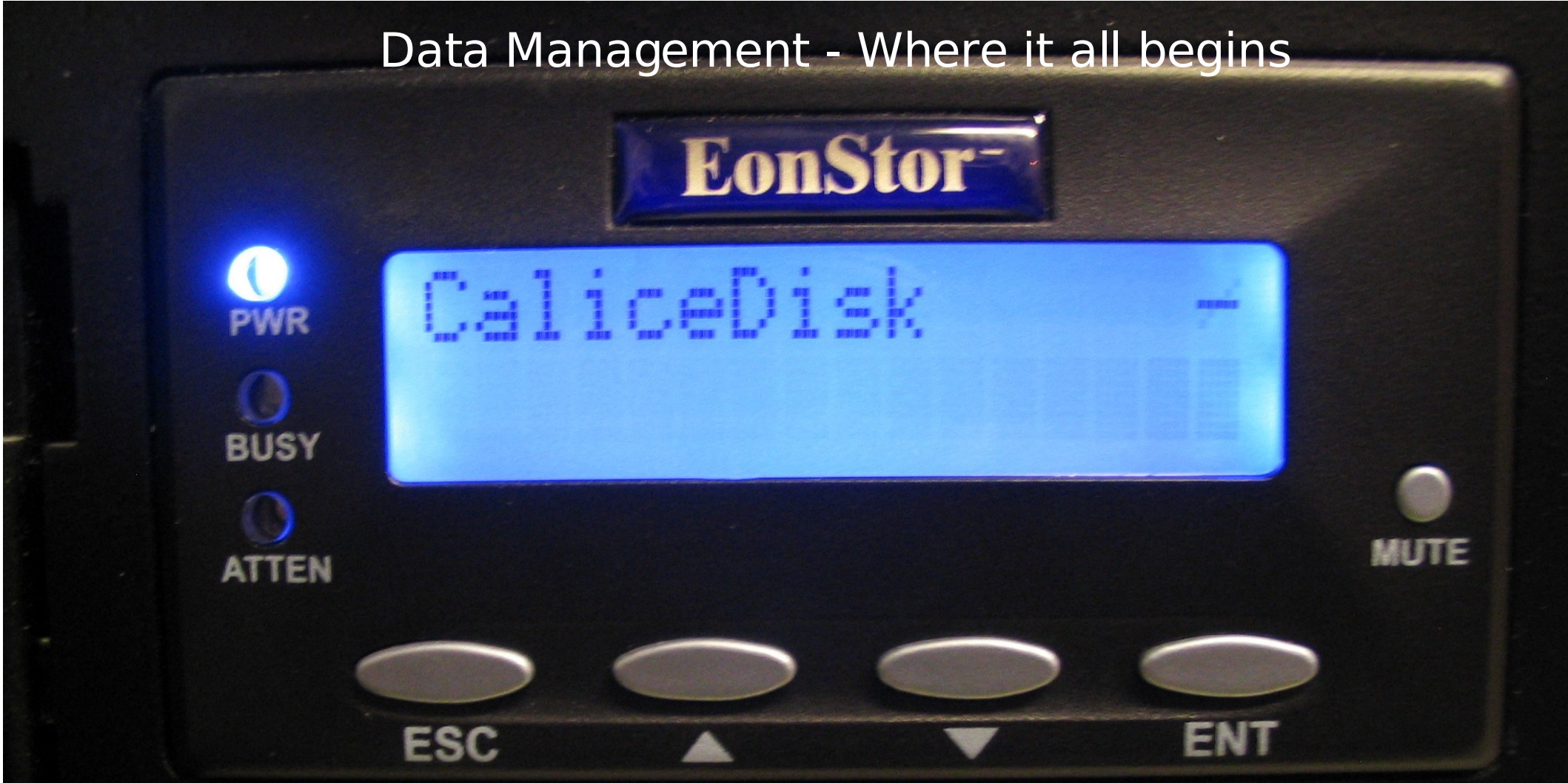
Roman Pöschl
LAL Orsay

Part I:
- Data Management
- Status of Conversion and Reconstruction Software
- Summary and Outlook

Part II:

- General Structure of Ecal Reconstruction

# Data Management - Where it all begins




- Local Buffer – 3 Tbyte
  First storage of data
  Enough to store ~ 1 week of data during cern07

*Picture courtesy of C. Rosemann DESY*

# CALICE “TIER 0” – Infrastructure in the Control Room

*Picture courtesy of C. Rosemann DESY*

Gigabit Uplink
- High Speed Connection to the outside world
- Serves all Calice Control Room Computers

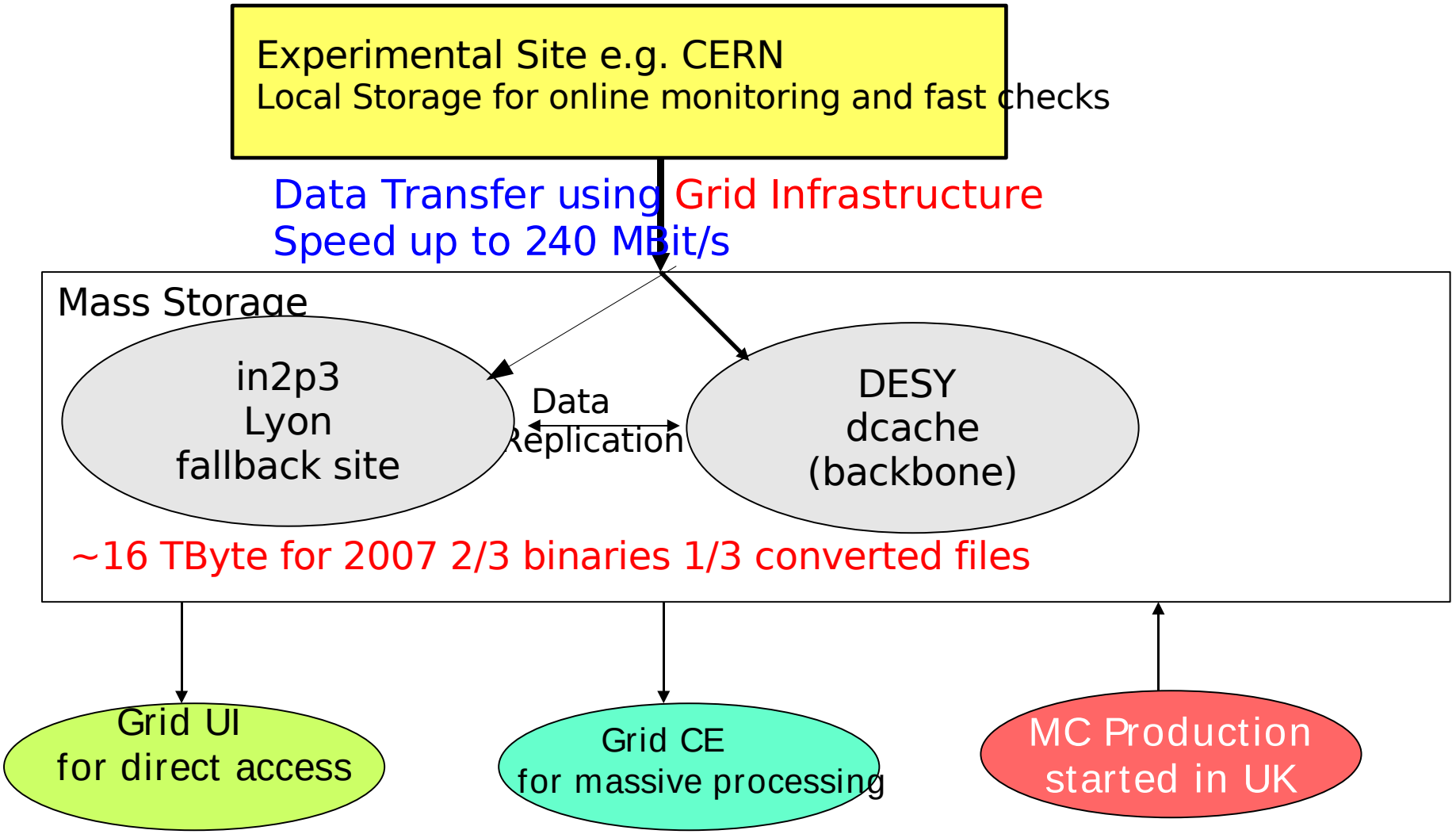
caliceserv.cern.ch
- Online Monitoring
- Grid Transfers

Disk Array

DAQ Computer

Well organized setup of computing
Thanks to B. Lutz

# Data Handling and Processing

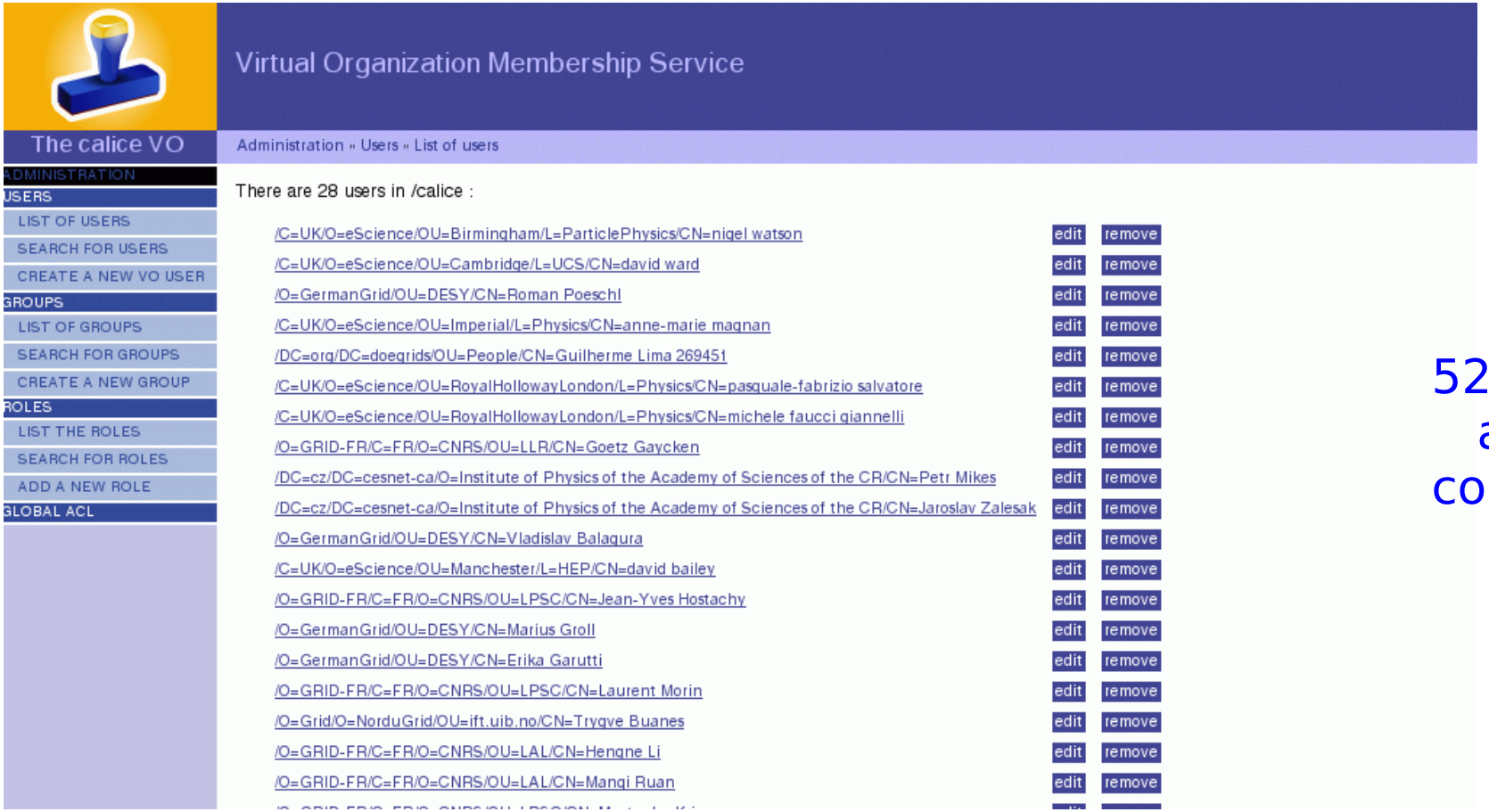



- Raw Data were (usually) available ~20 Min. After Run End
- Delay of Converted Files (usually) < 1 day

CALICE is the first Experiment which uses the grid systematically for real data!!!

# The Virtual Organisation - vo calice

Hosted by DESY:
Page for registration is https://grid-voms.desy.de:8443/voms/calice




52 Members
and
counting ..

VO Manager: R.P./ LAL, Deputy: A. Gellrich/ DESY

# Institutes which provide Grid support for Calice

| | |
|---|---|
| Supported by: DESY Hamburg | Hosting, Computing and Storage |
| LAL | Computing and Storage |
| LLR | Computing and Storage |
| DESY Zeuthen | Computing and Storage |
| Imperial College | Computing and Storage |
| Birmingham | Computing and Storage |
| cc in2p3 Lyon | Computing and Storage |
| Cambridge | Computing and Storage |
| Institute of Physics Prague | Computing and Storage (in preparation) |
| University College | Computing and Storage |
| KEK | Computing and Storage |
| Manchester | Computing and Storage |
| CIEMAT Madrid | Computing and Storage |
| Fermilab | Computing and Storage<br>Exploit started between Fermilab and NIU Colleagues |
| Univ. Liverpool | Resources Provided (not yet exploited) |
| Univ. Regina | Offer Received |

- Most of the sites have been involved in recent data and MC processing
  Smaller Problems at Manchester and KEK (about to be solved)

# The Grid kept me/us busy – Problems encountered

- Transfers blocked due to server overload
  Justification of having (at least) two major sites at hand
  One site (desy or cc lyon) usually was well performing

- Slow dcache doors at desy

- Hacker attack on desy

- Human Errors
  e.g. Corrupt mapping file at desy

Very fast a nd efficient response by experts
Thanks to their support

- Conversion sometimes failed due to several problems
  Connectivity problems
  Full disks at the various sites

- Conversion slowed in general down by 'poor' connectivity between sites
  It looks to me as if all sites are well connected to cern
  but badly connected among each other.

Status Summarized on Web Page – Created by B. Lutz

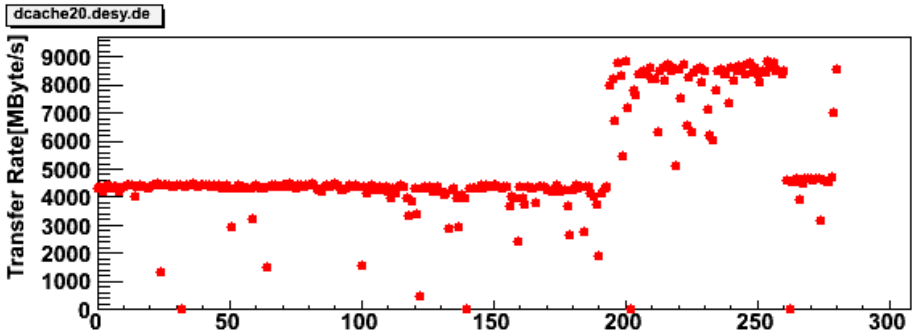
http://www.desy.de/~blutz/testbeam/dataStatus.html

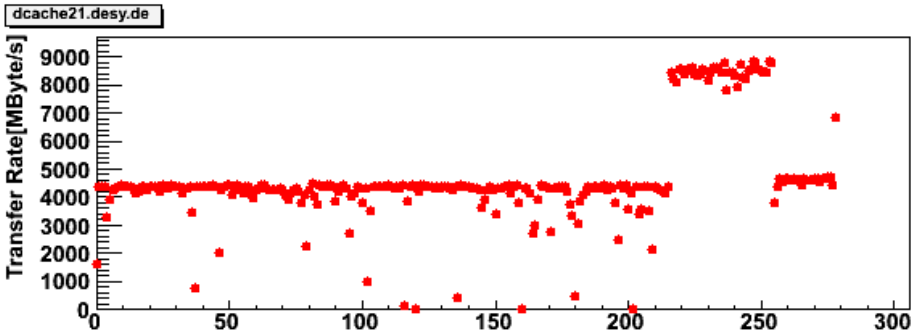
Last combined run taken: 331693
Number of combined runs: 1693
Current number of combined runs saved to GRID: 1691 (99%)
Current number of converted runs in GRID: 1687 (99%)
GRID disk space used by combined runs: 8274 GB
GRID disk space used by converted combined runs: 5991 GB
GRID disk space used by all combined data: 13 TB 953 GB

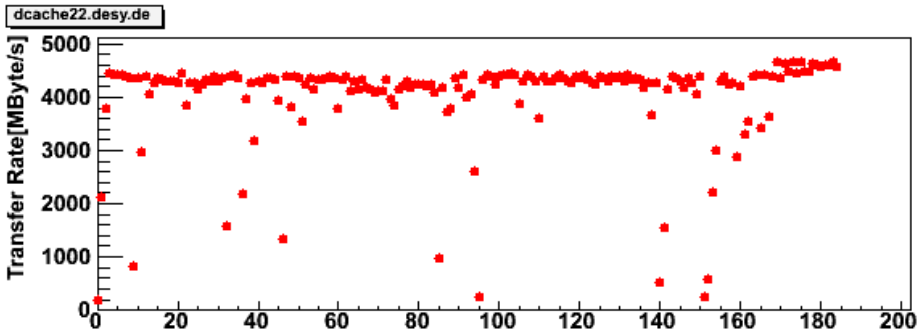
Last HCAL run taken: 350406
Number of HCAL runs: 406
Current number of HCAL runs saved to GRID: 407 (100%)
Current number of converted HCAL runs in GRID: 403 (99%)
GRID disk space used by HCAL runs: 606 GB
GRID disk space used by converted HCAL runs: 423 GB
GRID disk space used by all HCAL data: 1 TB 5 GB

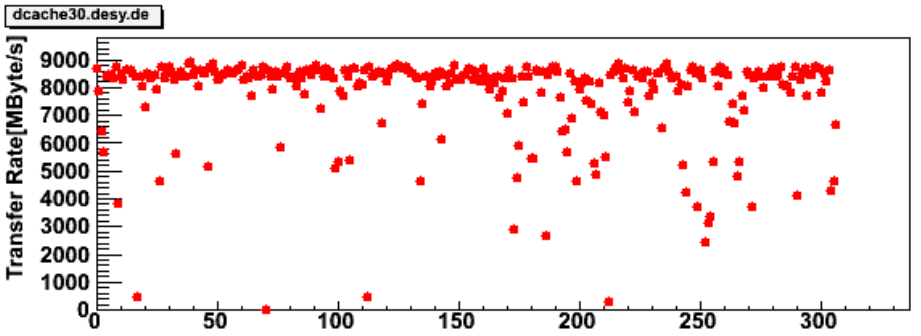
# Details of Data Transfers – DESY I

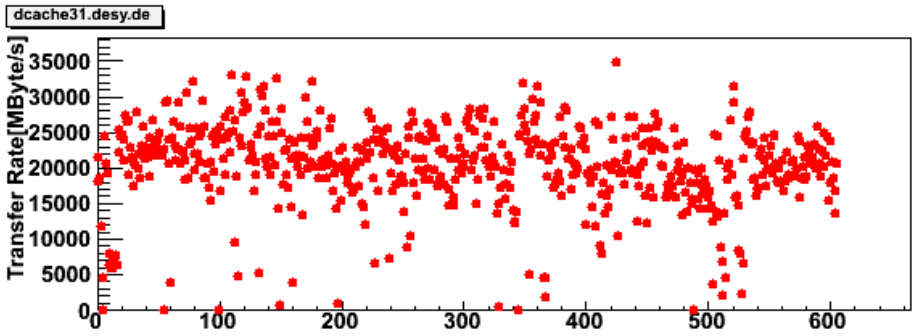
Transfer Rates Control Room -> Mass Storage obtained with different dcache doors
Transfers realized using lcg software and with 4 parallel streams

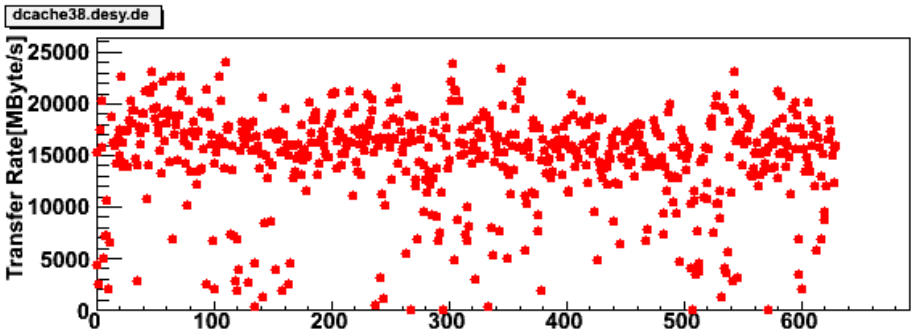


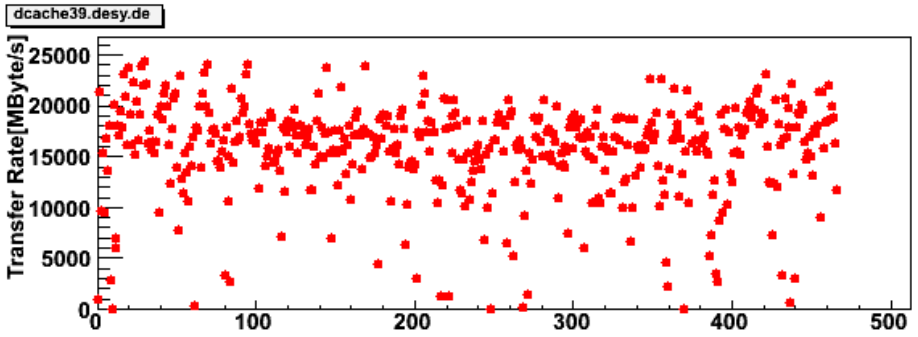



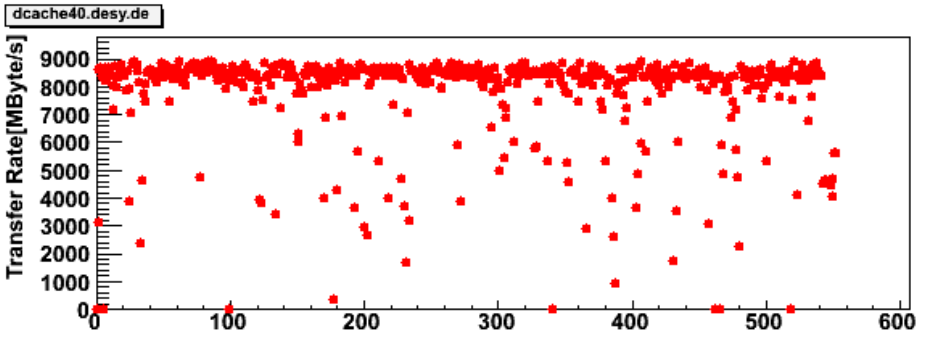



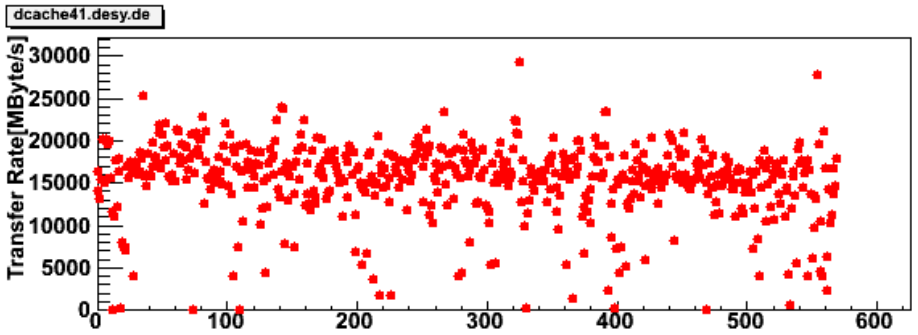

Average Rates between 4 and 22 MByte/s <- Proof of quality of our equipment
Different performance of different dcache doors not yet understod

# Details of Data Transfers – DESY II

Transfer Rates Control Room -> Mass Storage obtained with different dcache doors
Transfers realized using lcg software and with 4 parallel streams

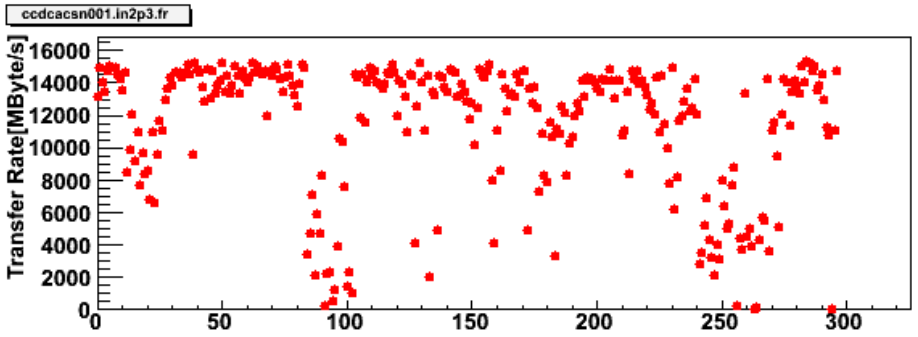


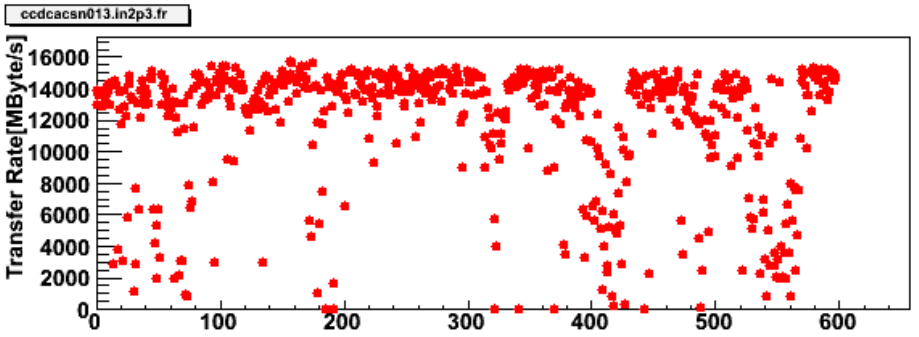



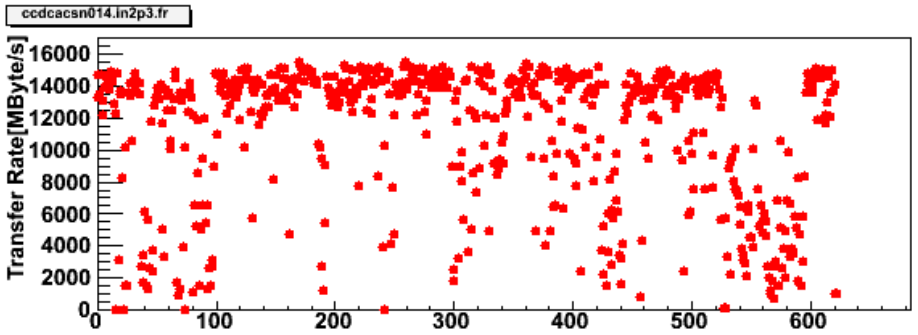



# Details of Data Transfers – CC in2p3 Lyon

Transfer Rates Control Room -> Mass Storage obtained with different dcache doors
Transfers realized using lcg software and with 4 parallel streams




Average Rates ~14 MByte/s - Independant of dcache door

# Outlook

- CALICE will continue data taking with fully equipped detector at Fermilab
  20000 cells in r/o again w/o zero suppression

- Data Taking of ScEcal and DHCAL(s)

- Mass Production of MC not yet started

Do have 30 Tbyte of data in stock !!!!
raw, converted and reconstructed data
MC files
Will grow beyond 2007!!!!

Need ~100 TByte until the end of 2008

Data Management

Need to have good network communications between 'major' sites

Investigations on bottlenecks between desy and cc in2p3 lyon are ongoing
(~ 2 Mbyte/s second transfer rate, therefore no mass replication so far)

Connectivity to KEK are under investigation
Tests with Fermilab should start soon
I see Manchester as another major site

# CALICE Software Packages

## - LCIO Conversion

All data of 2007 (CERN) have been converted using
the version v04-02-05 of the converter
New feature: new db entry RunTimeInfo for quick access to run times
db updated also for 2006 runs

## - Reconstruction

Nearly all runs of 2007 have been reconstructed for the Ecal using
Version v04-04 of the reco package
see lfn:/grid/calice/tb-cern/reco/monitor-cern07

'Unexperienced users' are encouraged to use these
as an entry point to the data analysis

DQ checks based on this version -> See talk by Manqi

Hcal Reconstruction Implemented into reco package
for 2007 only pedestal subtracted raw data are available

## - Userlib (Common to all packages)

Current version v04-07-02

# Under construction – Known Shortcomings and necessary updates

- DB issues

Trigger Delay for Cerenkov Counter not correctly in db (patch posted to s/w list)
Missing trigger Bit 9 Veto Counter
Validity times for rotated setups incorrect (at the beginning of a given period)

Cern beam info either missing or incorrect for 2007
Reason is restricted access to cern db
Correct settings are only available at run end – Conversion did fill info from run begin
-> refill of db needed

Will have a mixture of automatically created and hand written entries
for beam parameters
Makes the handling less trivial

Clear Agreement on these issues needed for FNAL (and other) testbeam(s)

# Under Construction – Known Shortcomings and necessary Updates II

- SiW Ecal reconstruction
  MC reconstruction and digitization revised by A.M. Magnan
  -> to be tested
  Who is doing the calibartion for 2007 data?
  Bottom slabs are only 'calibrated' with constant 1/45

- ScEcal Reconstruction
  No effort so far on a dedicated reconstruction package

- Hcal and TCMT Reconstruction
  Implementation into reconstruction package ongoing

- Drift chamber Reconstruction
  Progress for CERN DC?

- LCIO conversion
  Implementation of Hodoscope data ongoing
  Drivers for digital Hcal(s) needed

Expect new release of software packages for End of September

# Conclusions ...

- Successful running at CERN also from the software and computing point of view

  All data are available on the grid
  All runs have been converted
  Majority of runs available as reconstructed files

- Regular usage of the Grid for our data processing
  Did not run always smooth but proven to be a powerful tool

- Update of software packages under construction

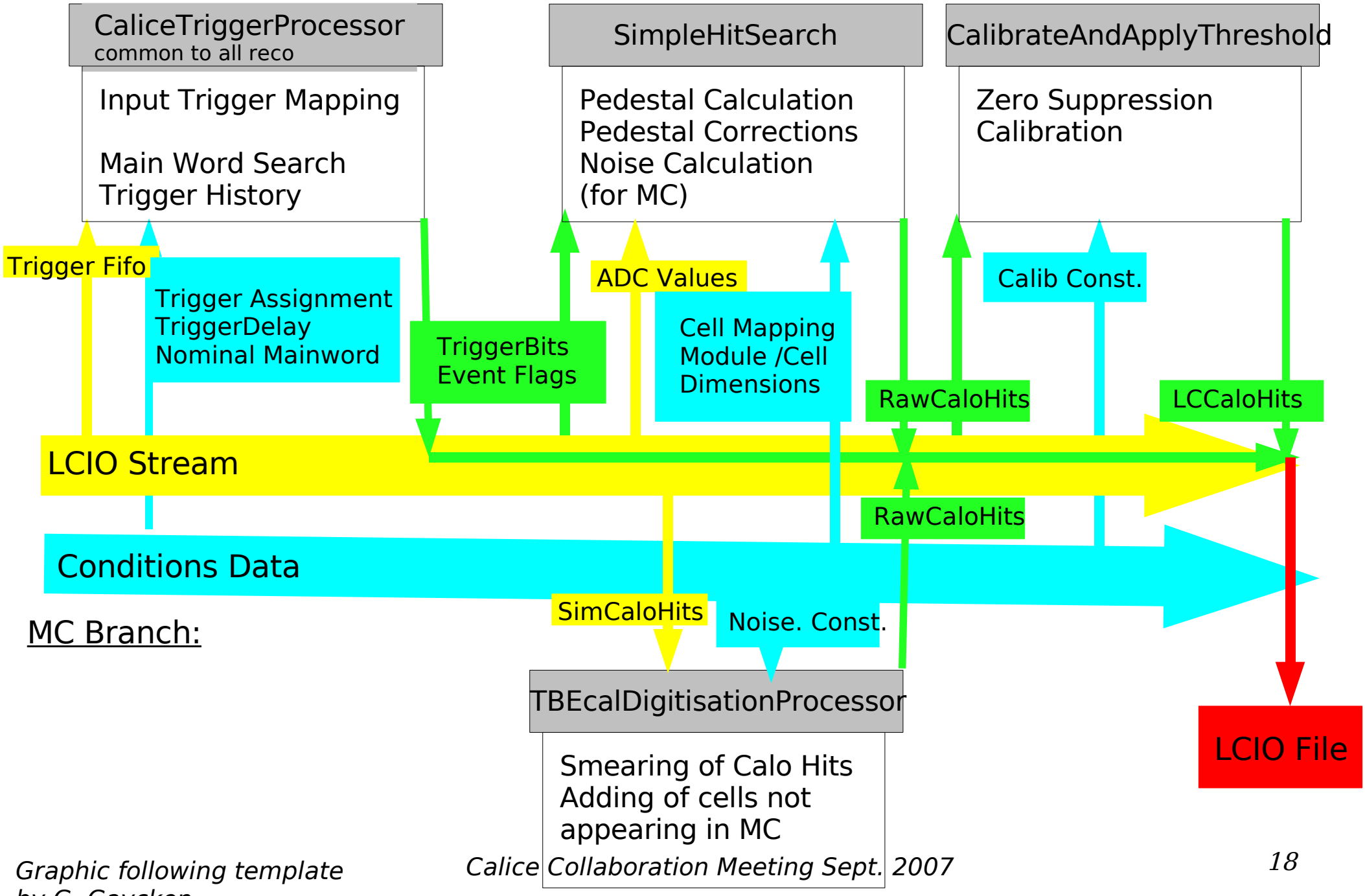
# Software Discussion

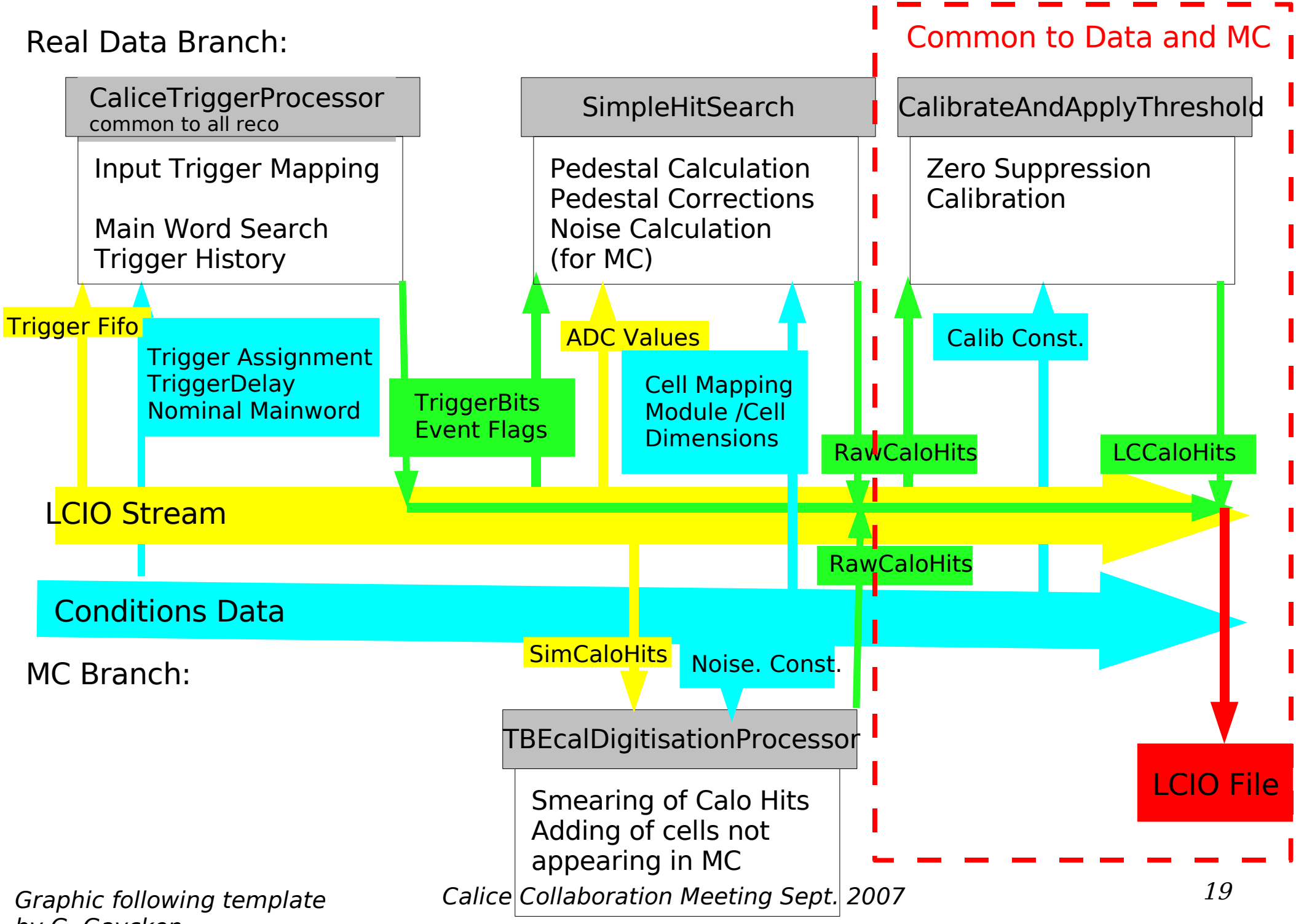
## General Structure of Ecal Reconstruction

- Will concentrate on core processors

  e.g. Square Event Finder or SimpleClusteriser
  will be discarded here

## Real Data Branch:




Real Data Branch:
Common to Data and MC
CaliceTriggerProcessor
common to all reco
Input Trigger Mapping
Main Word Search
Trigger History
SimpleHitSearch
Pedestal Calculation
Pedestal Corrections
Noise Calculation
(for MC)
CalibrateAndApplyThreshold
Zero Suppression
Calibration
Trigger Fifo
Trigger Assignment
TriggerDelay
Nominal Mainword
TriggerBits
Event Flags
ADC Values
Cell Mapping
Module /Cell
Dimensions
Calib Const.
RawCaloHits
LCCaloHits
LCIO Stream
RawCaloHits
Conditions Data
SimCaloHits
Noise. Const.
MC Branch:
TBEcalDigitisationProcessor
Smearing of Calo Hits
Adding of cells not
appearing in MC
LCIO File

# Gotcha's

- Calibration is based on the hardware structure of the cells
  Based on root file which contains the mean values of all cells

  For 'user defined' calibration
  - need to know the position of the Cell 0 in each module type
  - need to know the module names

  Once known user can create own calibration and feed it into the data stream as conditions data
  No code change but change in steering to define source of CD data

  Calibration can currently not be applied to reconstructed data!!!
  Needs RawCalorimeterHit
  Can be changed if necessary – However may put cells below thresholds

  Calibration is defined for whole running period

- Noise Thresholds can be set in steering file
  Pedestal Corrections need to be switched off/on in code
  -> to be changed

- Noise and digitization is (partially based) on real data and therefore realistic
  Averaged values for non existing cells when full Ecal is simulated
  No benefit seen so far from digitisation

# Generalities – Not only true for Ecal

- Need different steerings for different modes of data taking

Reason is (partially) parallel data taking
- e.g. @ desy and cern, tent and beam area at desy

- DAQ allows for parallel data taking with different detector components

Remember (CondDBMySQL) Database is (validity) time oriented

Steering files used for a given run can be extracted from the grid
see ../log directories in grid directory structure

Feasible (but not trivial) to hide these details from users