

# Some CMS Grid analyses

- *Computing, Simulation and Analysis challenge (CSA) 2006*
- *Tracker Integration*

Stijn De Weirdt (VUB)

Pascal Vanlaer (ULB)

Giacomo Bruno (UCL)

# CMS analysis exercises during CSA06

## ■ Goals

- Test various analysis workflows and data formats
  - Calibration and alignment / AlCaReco formats
  - Physics studies / Analysis Object Data (AOD) format
- Train physicists to CMS Grid computing tools
  - Data Bookkeeping System / Data Location Service (DBS/DLS)
  - CRAB: CMS Remote Analysis in Batch
    - User Interface to grid scheduler (LCG, gLite and OSG) and to DBS/DLS
  - Tier-2 computing tools
    - Data transfers to Tier-2
    - Tier-2 file system (dCache etc.)
    - Interactive analysis of root files on Tier-2 store
- In total ~30 exercises, ~70 physicists involved during 2 months, quasi-online

# List of exercises

## Calibration/Alignment

### ■ ECAL calibration

- E/p of isolated electrons
- $\Phi$ -Symmetry of ECAL
- $Z \rightarrow e^+e^-$  intercalibration

### ■ HCAL calibration

- By isolated tracks
- $\Phi$ -Symmetry

### ■ Tracker, Muon alignment

- $Z \rightarrow \mu^+\mu^-$

## Physics Analyses

### ■ Extraction of $W \rightarrow \mu\nu$

### ■ Di- $\mu$ spectrum

### ■ Tau-tagging efficiency

### ■ Minbias-underlying event

### ■ Leptonic top decays

### ■ Lepton isolation in $t\bar{t}$

### ■ W mass

### ■ SUSY LM1 benchmark point studies

### ■ Excited quarks / $Z' \rightarrow \text{jets}$

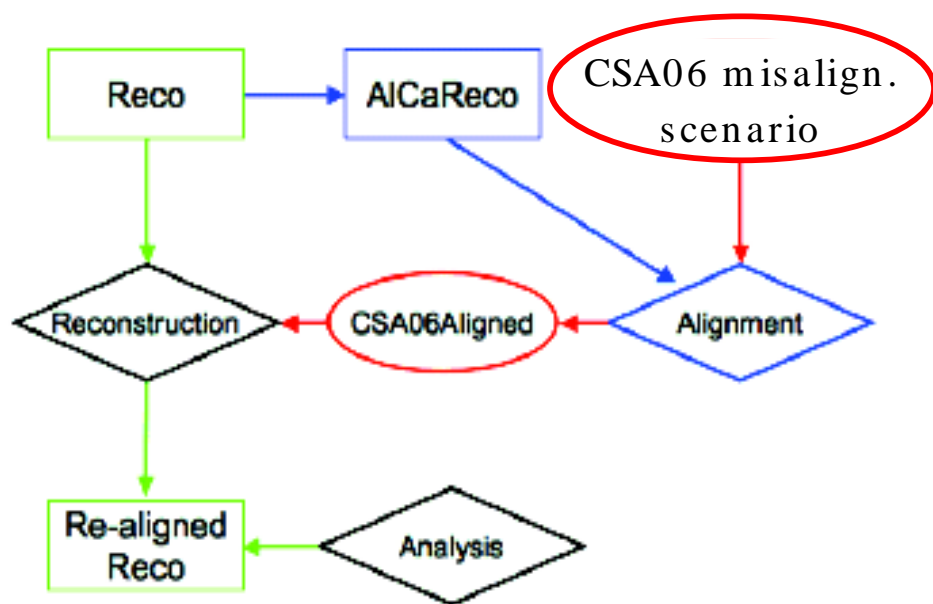
### ■ High-energy di-electron pairs

### ■ ...

# Example: tracker+ muon alignment workflow

## Goal

Provide and apply alignment corrections for reconstruction, **in a production environment.**



- Read constants from DB at prompt reco. (PR) [T0]
- Write dedicated stream during PR [T0]
- Run alignment on these streams [T0]
- Store and distribute DB object [T1/2]
- Re-reconstruct with new constants [T1]
- Run analysis [T2]

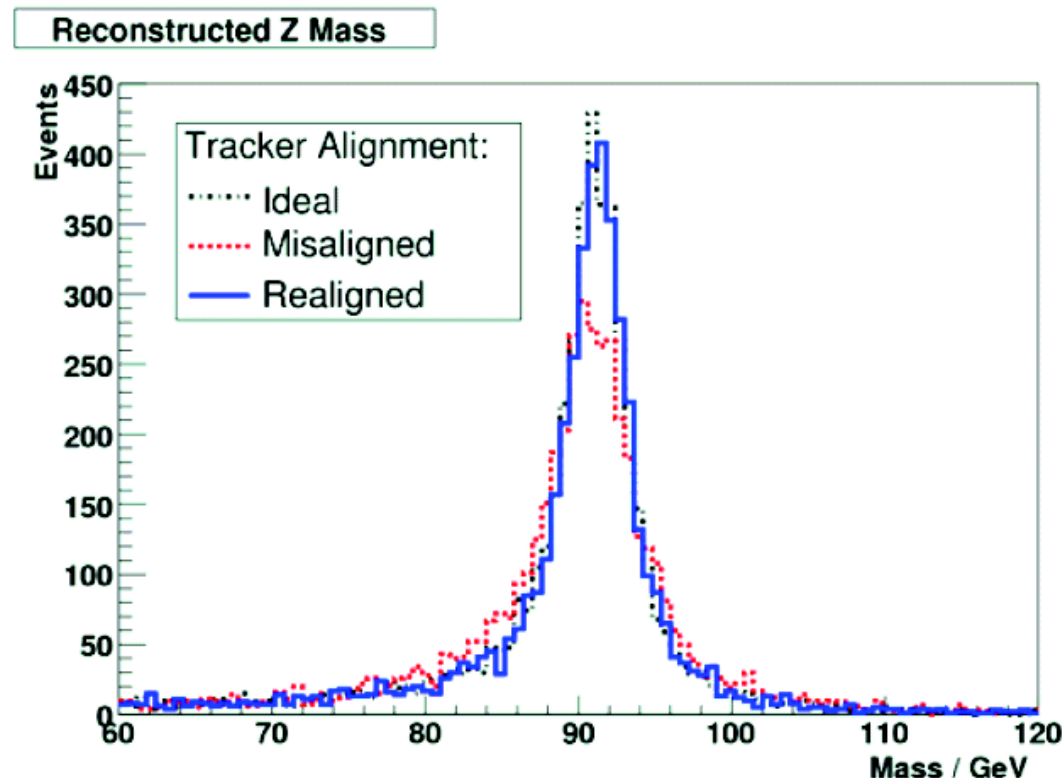
# Achievements

- **AlCaReco formats defined, e.g. For alignment:**
  - 1 million  $Z \rightarrow \mu^+ \mu^-$  events simulated
  - AlCaReco with only (used tracker hits + full track fit info)
    - Residuals etc.; allows alignment + track refit

- **Full chain exercised**

- Data flows T0-T1-T2
- Alignment
- Geometry DB update
- Re-reconstruction
- Publication of Re-RECO data
- Re-analysis of  $Z \rightarrow \mu^+ \mu^-$

...in 2 days after publication  
of AlCaReco data file



# Two “real- scale” analyses on the Grid in Belgium

- **During CSA06: High-Energy Electron and Photon Pairs**  
working group (HEEPP – Brussels et al.) within CMS  
SUSY/BSM Physics group
- **During Tracker Integration test: tracker time**  
synchronization and data quality monitoring (UCL et al.)

## ■ Goals

- Learn and prepare tools for early data analysis
  - CMS computing tools
    - Was very useful
    - Lessons made into tutorials for Belgian Tier-2 users
      - » <http://mon.iihe.ac.be/trac/t2b/wiki>
      - Getting started with CMS software on the Grid
  - Data and workflow for early analysis
    - Data validation
    - Event selection technology, specific HEEPP trigger criteria

## ■ The exercise

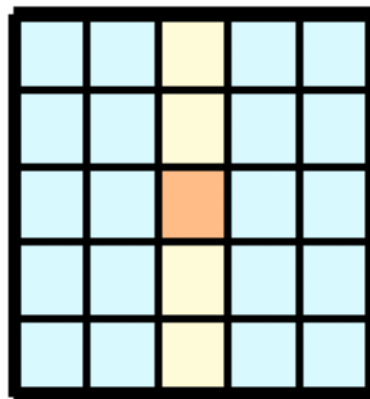
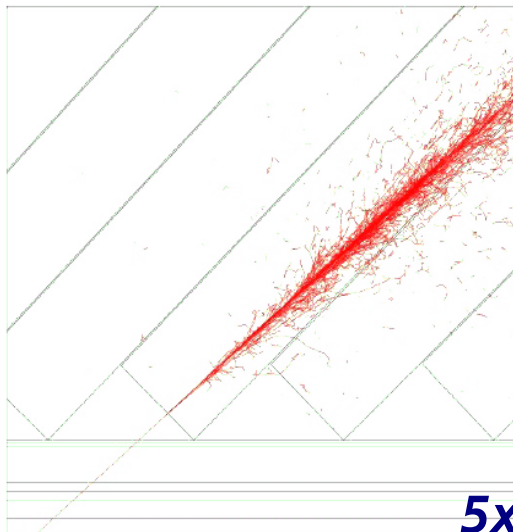
- Select  $Z' (m=700 \text{ GeV}) \rightarrow e^+e^-$  in “ExoticSoup” event mixture
  - 1M evts, 39%  $Z'$  all decays; 39% SUSY LM1, 22% excited quarks

# CSA06 HEEPP analysis workflow

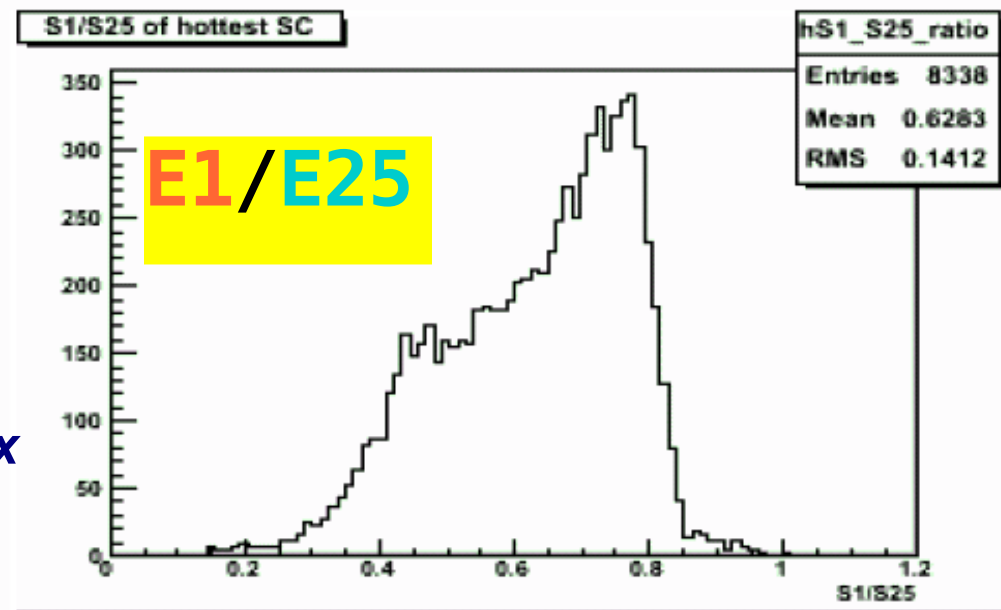
*Validate objects used  
to filter events  
(e.g. ECAL clusters)  
@ Tier- 1*

- Wait for publication of official ExoticSoup RECO files in DBS/DLS
- Run Ecal cluster analysis on host Tier-1 using CRAB

• e.g. Histos of lateral shower shape



*5x5 crystal matrix  
E1 = energy of  
hottest crystal*





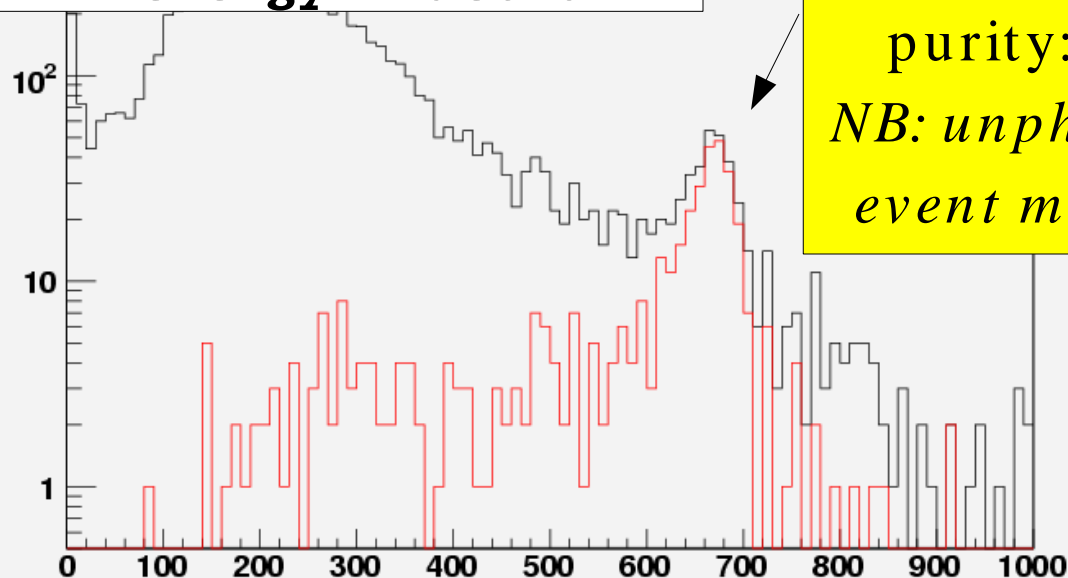
# CSA06 HEEPP analysis workflow (2)

*Validate objects used  
to filter events  
(e.g. ECAL clusters)  
@ Tier- 1*

*Implement event filter  
in CMS software  
framework*

*Validate filter  
(efficiency - purity)  
on test sample  
@ Tier- 2*

**$M(e^+e^-)$  before / after  
cut on Hadron/EM  
energy fraction**



*$Z' \rightarrow e^+e^-$  signal  
efficiency: 81%  
purity: 75%  
NB: unphysical  
event mixture*

*Publish filter in cvs  
for prod. team to run*

*Subscribe to  
filtered data set*

*Perform analysis  
on filtered data set  
transferred @ Tier- 2*

# Lessons learned

*Validate objects used  
to filter events  
(e.g. ECAL clusters)  
@ Tier- 1*

*Implement event filter  
in CMS software  
framework*

*Validate filter  
(efficiency - purity)  
on test sample  
@ Tier- 2*

*Publish filter in cvs  
for prod. team to run*

*Subscribe to  
filtered data set*

*Perform analysis  
on filtered data set  
transferred @ Tier- 2*

- Lot of preparation work before large-scale batch analysis on Grid

- Preparation = mostly interactive analysis

- i.e. non-Grid
- Need some moderate Tier-3 CPU capacity

- » 8 CPU's during CSA06

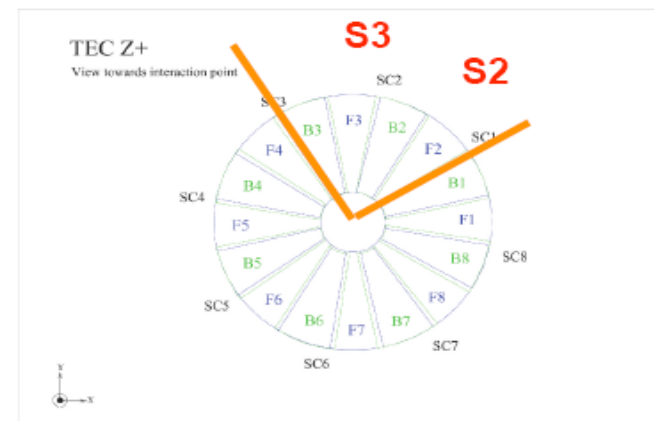
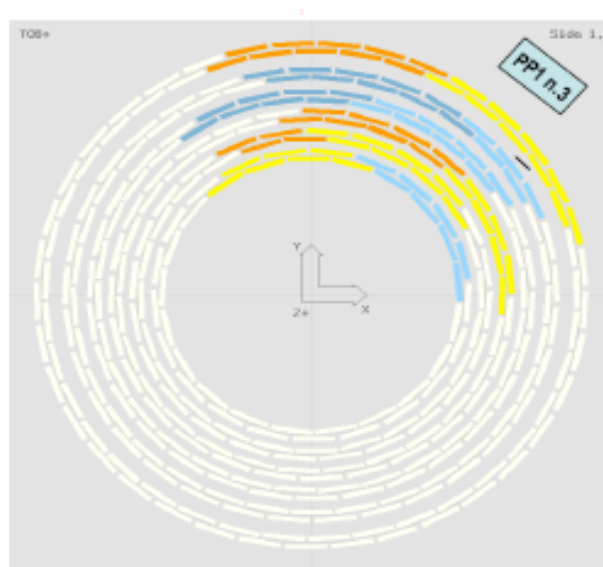
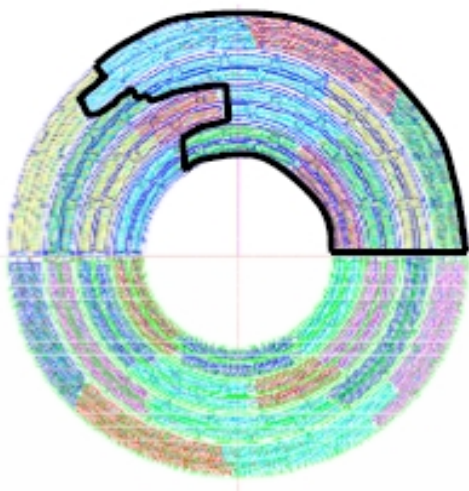
- » direct access to Tier-2 disks

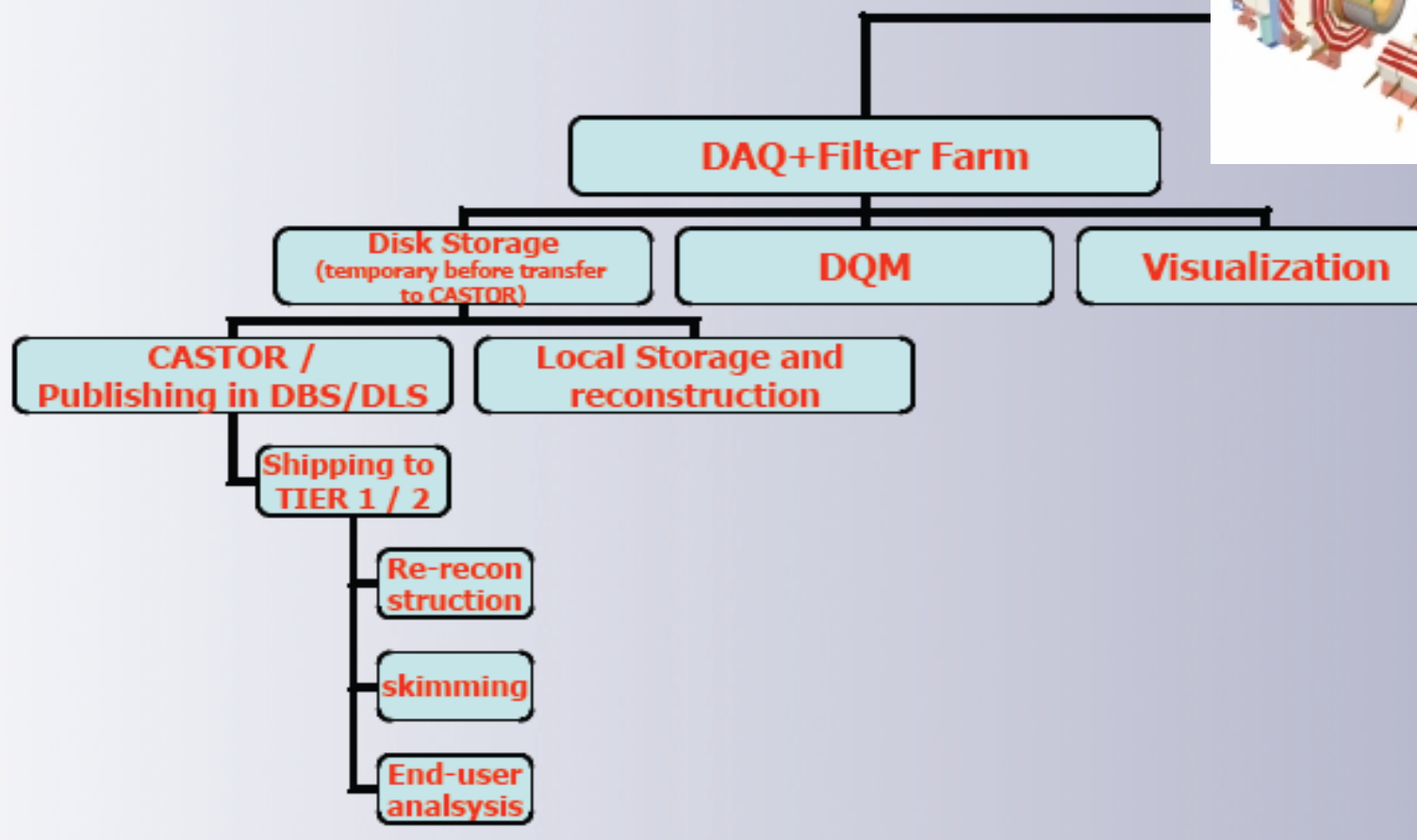
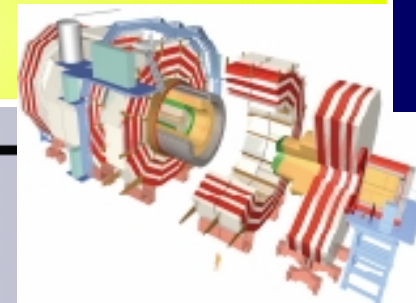
- HEEPP exercise in CSA06: **extreme case**

- **Data and filter validation (time-critical) done on Tier-3 on local copy of official sample**

# CMS Tracker Slice Test at the TIF (Tracker Integration facility)

- **>2000 silicon strip modules (15% final system)**
  - Tracker Inner Barrel / Inner Disks: 640 modules (36% of TIB/TID+)
  - Tracker Outer Barrel: 720 modules (28% of TOB+)
  - Tracker EndCap: 800 modules (25% of TEC+)





## *Tracker Analysis Center (TAC):*

- Tracker Control Room + Tier 0 + Central Analysis Facility
- Use standard CMS computing tools

## ■ Storage at TAC:

- Filter Farm Storage manager stores raw data files on local disk server (2.8TB - no redundancy)
- Only temporary storage

## ■ Processing at TAC:

- Cabling info, pedestals and noise moved from online DB to offline DB via 020 programs
- Process Raw data with CMSSW + ProdAgent
  - Currently only file conversion: RU (old Tk custom online output format)→EDM or DAT (standard Storage Manager output format)→EDM

## ■ Files copied to CASTOR (RAW data + EDM ROOT tuples)

## ■ EDM files (only) registered in Data Bookkeeping Service (DBS) and Data Location Service (DLS)

## ■ Data published in DBS/DLS is ready to be transferred via PhEDEx (injection currently done in Bari)

- T1/T2 center (e.g. FNAL) simply subscribe to the data sets

# Conclusions

- CMS data and work-flows exercised in many analysis use-cases
  - Full data and work-flow in Alignment/Calibration use cases
  - Physics analyses
    - Main output is training to software and computing tools
  - Detector integration
    - Standard data and work-flow successfully used for Tracker
- Many problems identified and solved
  - e.g. Undesireable couplings between CRAB, DBS/DLS and reconstruction software versions
- Next
  - Study implications of data distribution according to trigger streams (CSA07)
  - Improve involvement and reliability of Tier-2's (CSA07)
  - Improve support to analyses (in progress)