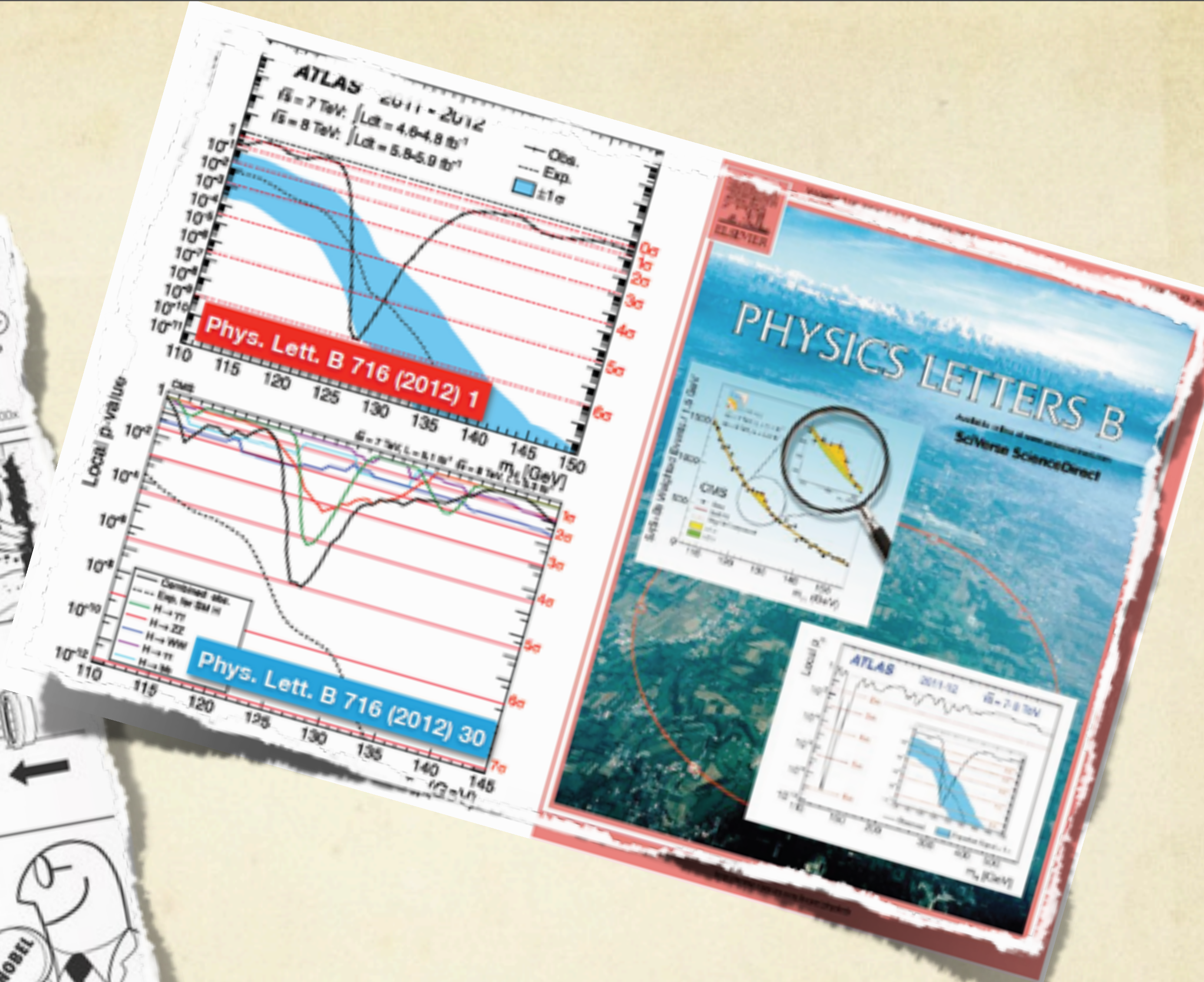
From detector building to physics publication: the real story of the data

Patrizia Azzi - INFN Padova

DISCLAIMER & FOREWORD

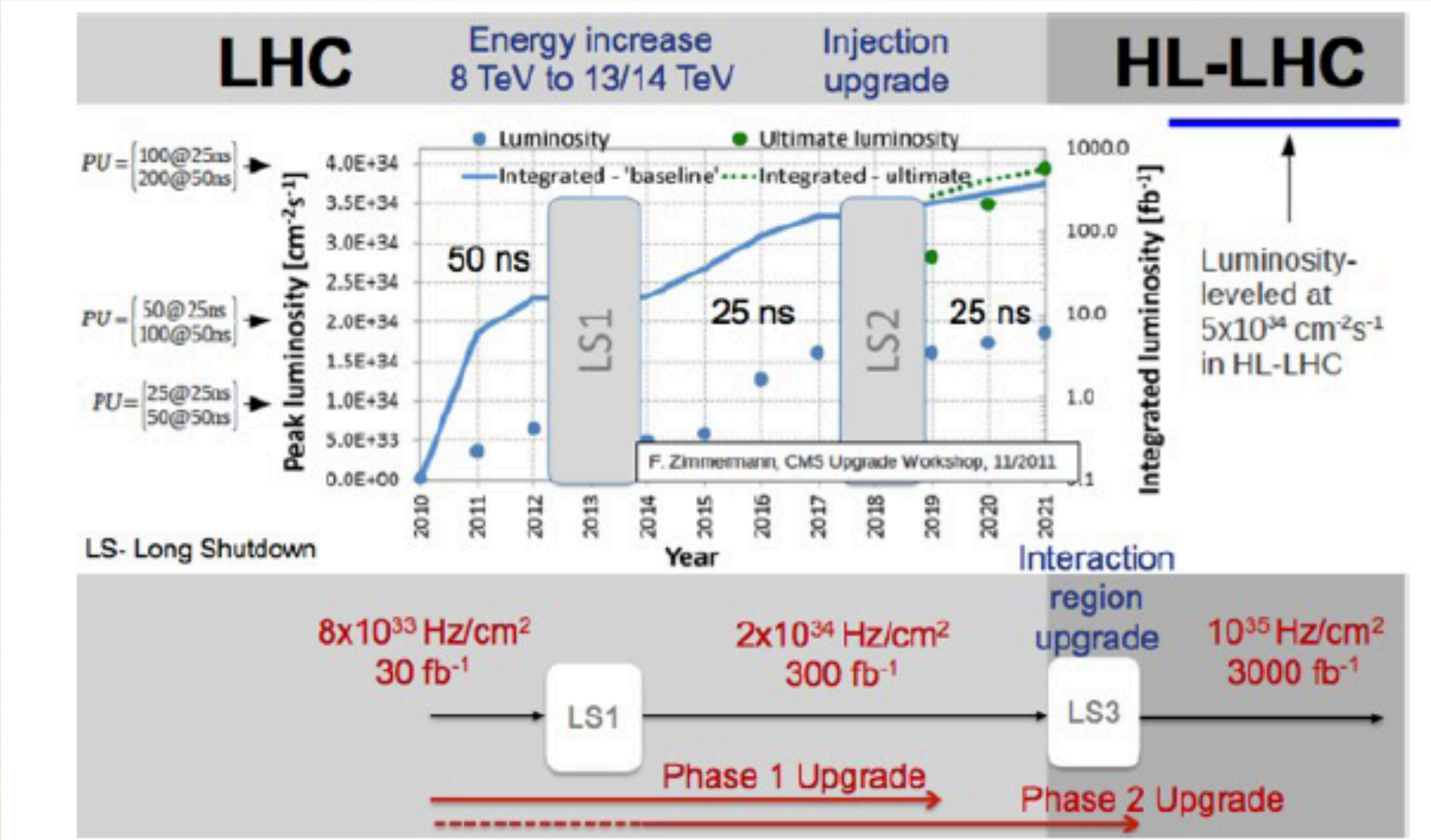
- BIG thank you to all those that I stole material from, especially my ATLAS counterparts that took time to answer my many questions.
- Attempting here a list (in random order): R. Van Kooten, S. Banjeree, JR Vlimant, L. Malgeri, H. Jung, L. Fiorini, G. Unal, L. Silvestris, G. Cerminara, M. Rovere, P. Govoni, P. Elmer, A. Giammanco, J. Boyd, A. Bocci, M. Hildredth, B. Mangano, C. Bernet, F. Cossutti, D. Lange, G. Franzoni .....
- This is a lecture and not a conference: experiments (CMS & ATLAS) are quoted and used as examples only. Comparisons are made to show you how similar/different are the ways and solution found to the same problems.
- Summarizing here what we have learned over the past 3 years. It was a long ride.
- Many very different topics! Not enough time to go in all details.
- I will be present at the Discussion sessions in the afternoon to answer all(?) your questions!

...that was last year...and now?



...that was last year...and now?

The road ahead



Outline

- Life during operations: get the physics out as fast as possible (really really fast)
  - Taking data
  - Calibrate your data
  - Certify your data
- Life during a shutdown: prepare for next data taking
  - Improve your simulation
  - Improve your reconstruction
  - Improve your computing

Life during operations

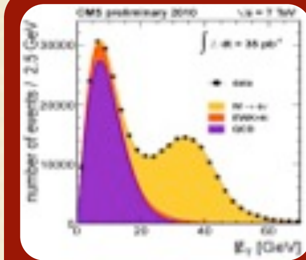
«Data preparation»

Collisions!

Detector Response

Trigger

Event Reco.



Physics Analysis

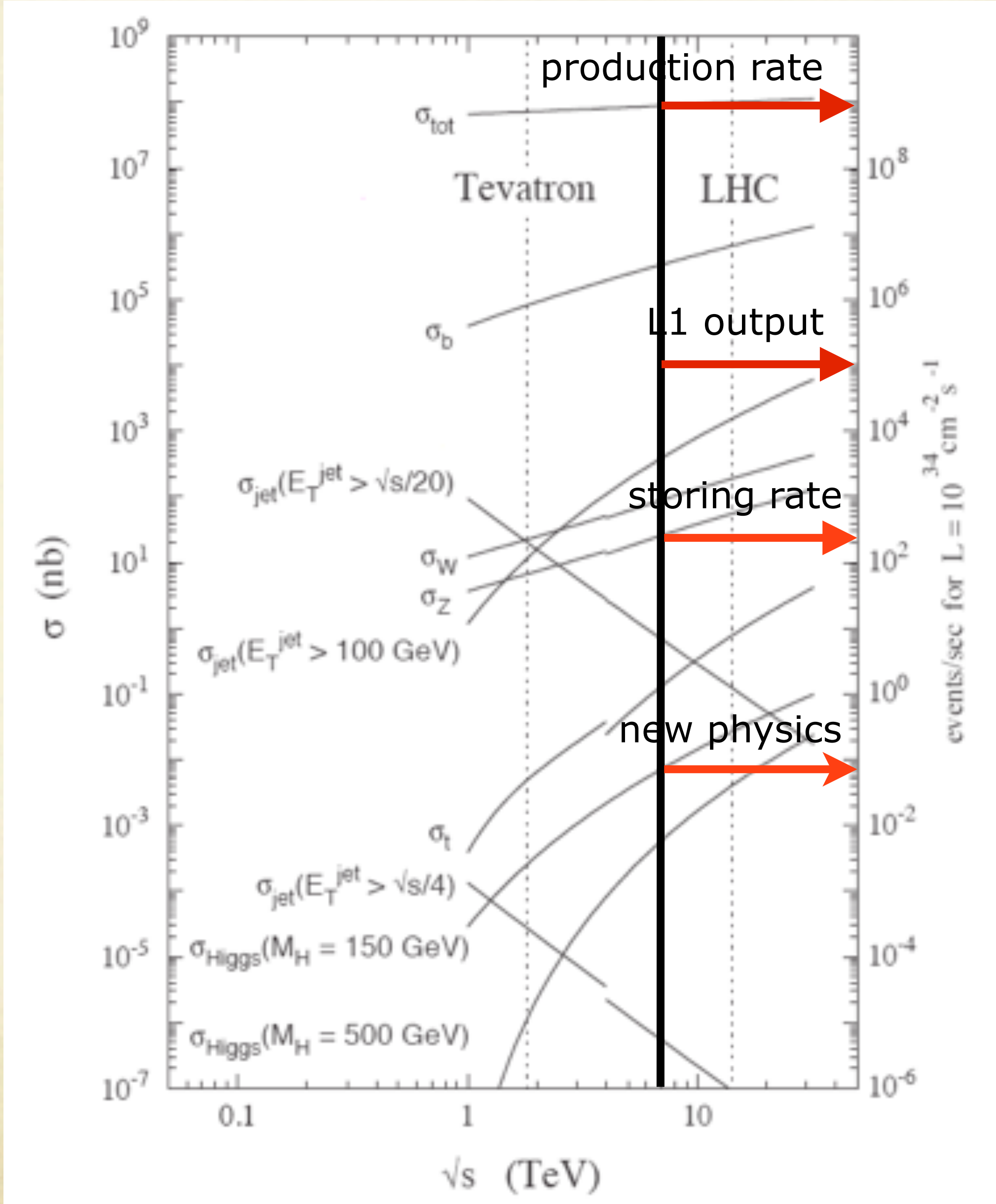
Taking data

Do we trust or do we check?

- Beautiful physics results are the final chapter of a long journey. They are the result of a complex recipe that involves diverse skills and tools.
  - Assume that we are working with an excellent accelerator producing a large statistics of collision data and we have a *beautiful detector* and our reconstruction software is *fast, efficient and pure.*
- The final physics analysis assumes as an input *data that are optimally calibrated and reconstructed* (both for real and simulated samples) hoping to minimize as much as possible the sources of systematical uncertainties
- **«data preparation» is the missing link often neglected when you take physics lectures, but were you might end up spending most of your time!**
  - **However, it is another way to _really_ know the physics of the detectors you are using.**

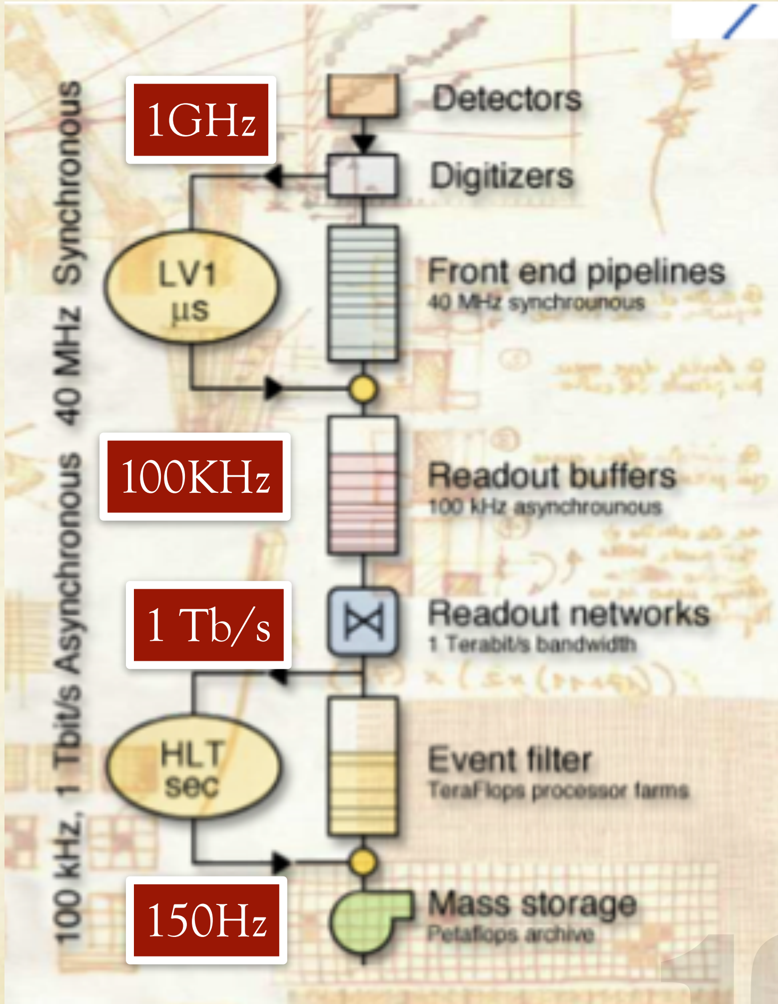
why do we need a trigger?

- Collision rate at the LHC is heavily dominated by large cross section QCD processes not interesting for the physics program of CMS. Interesting physics has rates <10Hz
- Not possible to write all the events and select later on. Final bandwith limited up to O(100-400 Hz)
- Need to select events beforehand very quickly: bunch crossing rate =1/25 ns
- Physics driven choices: select process with large transverse energy, and one or more interesting objects such as high momentum leptons or jets

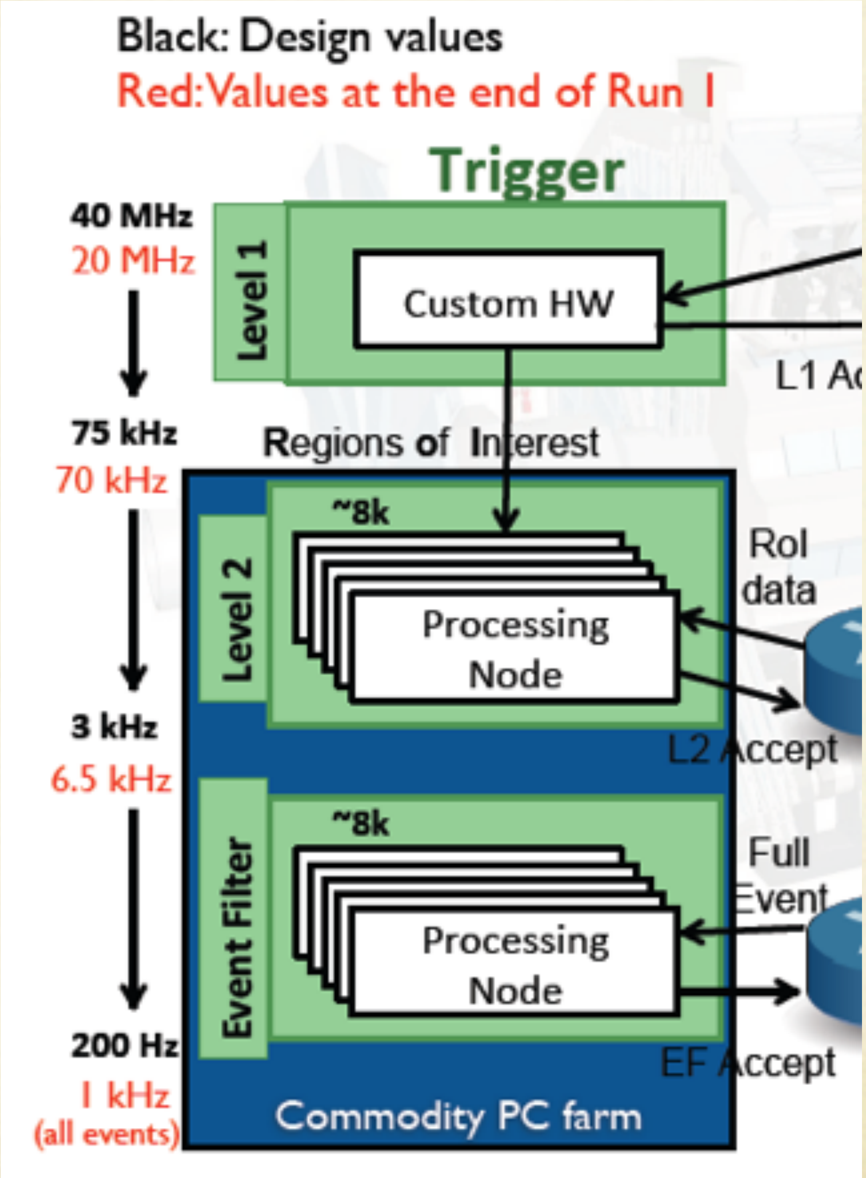


High Level Trigger - A CMS Example

- The CMS trigger system is structured in two levels:
  - **The Level1 trigger (L1)**, implemented in hardware, running at the nominal LHC rate of 40MHz. Based on regional information.
  - **The High Level Trigger (HLT)** implemented as a dedicated (simplified) configuration of the CMS reconstruction software, running on the DAQ filter farm, at the L1 output rate of 100KHz
    - There are 4704 processes in parallel, each of them with about 47ms to take the decision and stream out the data
  - The nominal output rate was ~200Hz
  - *For comparison offline reconstruction takes about 5s per event.*

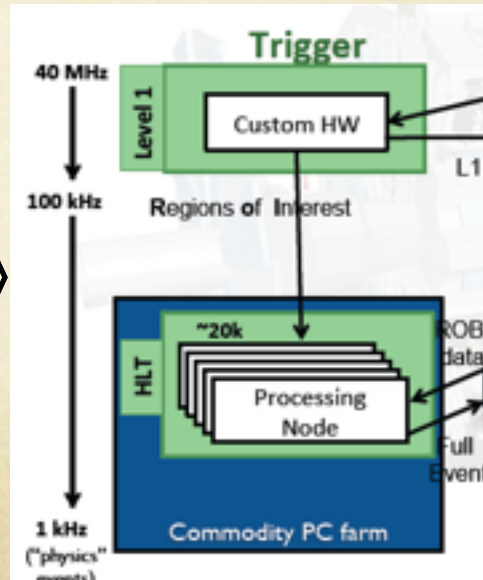


The High Level Trigger - the ATLAS example



- Trigger separated into 3 levels:
  - **The Level1 trigger:** implemented in hardware, running at the nominal LHC rate of 40MHz. Provides information on Region Of Interest.
  - **The Level 2 trigger:** dedicated trigger algorithms running on commodity PC hardware, making decision about events only accessing parts of the full event.
  - **Event Filter:** mixture of trigger specific and offline algorithms selecting interesting fully-built events. Receives input from Level1 and Level2

Future plans for RunII : merge L2+EF

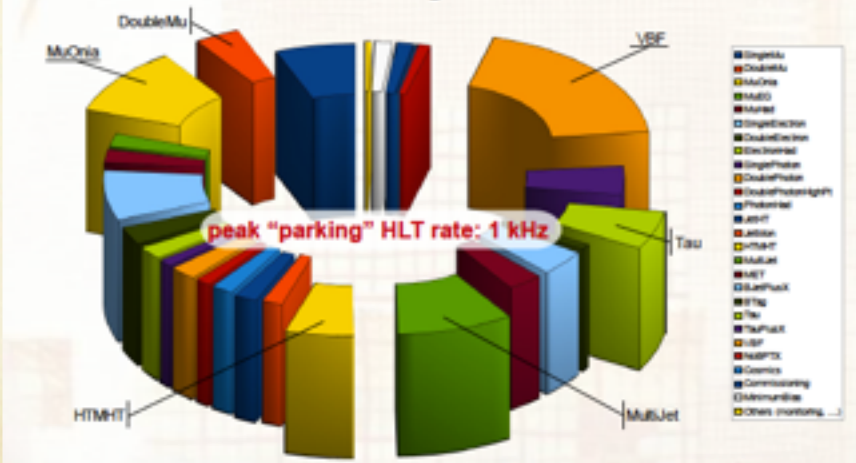


2012 Trigger table (CMS)

Rate	Stream / dataset	Purpose
500 Hz	ALCALUMIPIXELS	Alignment, Calibration and Luminosity
10 kHz	ALCAP0	Alignment, Calibration and Luminosity
1.5 kHz	ALCAPHISYM	Alignment, Calibration and Luminosity
1.8 kHz	RPCMON	Alignment, Calibration and Luminosity
100 Hz	Calibration	Alignment, Calibration and Luminosity
100 Hz	EcalCalibration	Alignment, Calibration and Luminosity
10 Hz	TrackerCalibration	Alignment, Calibration and Luminosity
1 kHz	Stream A	Physics
40 Hz	Express	Physics
1.2 kHz	PhysicsDST	Physics
10 kHz	NanoDST	Trigger studies
25 Hz	Stream B	Data Quality and Monitoring
75 Hz	DQM	Data Quality and Monitoring
1 kHz	HLTDQMResults	Data Quality and Monitoring
150 Hz	HLTDQM	Data Quality and Monitoring
20 Hz	HLTMON	Data Quality and Monitoring

“core” data for analysis in 2012

inclusive “core” and “parked” data for 2013

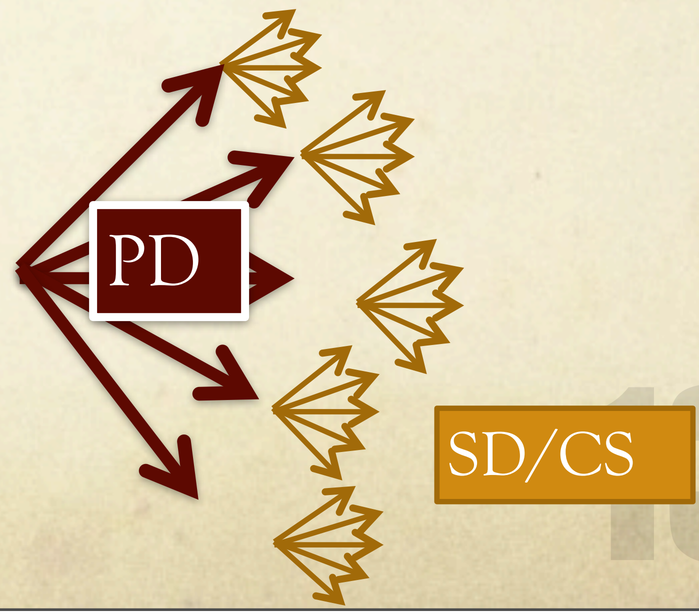


The highlighted dataset are the inclusive version of the corresponding "core" datasets, and collect extra "parked" events, to be analysed only in 2013. The others are unchanged with respect to the "core" datasets in the top plot.

Organization of Datasets for Physics

- Each event that passes one or more HLT selection is then collected into Primary Datasets.
  - **Primary Datasets: defined on the basis of the trigger bits.**
    - Events triggered by more then one trigger appear in more then one PD
- Further splitting philosophy different:
  - CMS: **Secondary Datasets(**subset of a PD based on trigger bit selection) and **Central Skim(**Subset of a PD based on reconstruction information)
  - ATLAS: **Derived Secondary Datasets** (subset based on reco information but with also a reduced event content matched to the analysis interest)

- **Parked/Delayed Datasets:** data collected is stored in RAW format for subsequent reconstruction at a later time. Useful for large rate triggers with lower thresholds for precision measurement (B physics etc)



Designing a «trigger path»

- HLT looks for the «interesting physics events» that usually contain **«interesting physics objects»**, such as:
  - leptons: high pt, isolated, multiple
  - large missing energy
  - large transverse energy, of high Pt jets, or many jets
  - mixture of different objects or specific topologies (rapidity gaps)

A recent HLT "menu" is composed by

- 440 logically independent paths and 16 streams
- over 580 reconstruction modules, organised in 240 sequences
- over 1900 event selection modules
- over 200 shared configuration services (access to database, geometry, magnetic field, ...)

CMS

- The trigger paths are defined starting from the analysis one is interested into:
  - however, signals with low p_T, loose ID, few leptons are more difficult to trigger in a pure and efficient way
  - when conditions become harder lot of effort is put to avoid raising the thresholds to reduce the rate. Rather improving path logic and/or reconstruction

Same for ATLAS!

Data Parking and Data Scouting

INFN

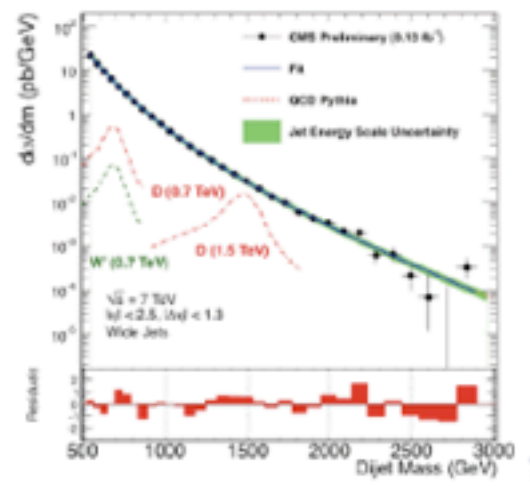
- Core Data (300-350 Hz)
  - Produce the datasets we need for our main physics program
- Parked Data (300-600 Hz)
  - Triggers are either a looser version of the physics triggers or brand-new triggers with small overlap with the rest
  - They complement and greatly enhance the physics program to be processed during the 2013-2014 LHC shutdown
    - For example, special Higgs production, Supersymmetry channels and B Physics

- Data Scouting
  - Typical use case: recover sensitivity for new physics searches in hadronic final states at " low jet $P_T/H_T/$ ..."
  - Novel trigger and data acquisition strategy applied to physics analysis
    - Trigger: $H_T > 250$ GeV
    - High event rate (~10^3 Hz)
    - Reduced event content (i.e. store only calorimeter jets reconstructed during HLT, no raw data, no offline reconstruction possible)
    - Bandwidth (rate x size) under control

Scouting approach extended the di-jet search below 1 TeV

Test Feasibility of Data Scouting in 2011: Dijet Resonance Search (0.13 fb^{-1})

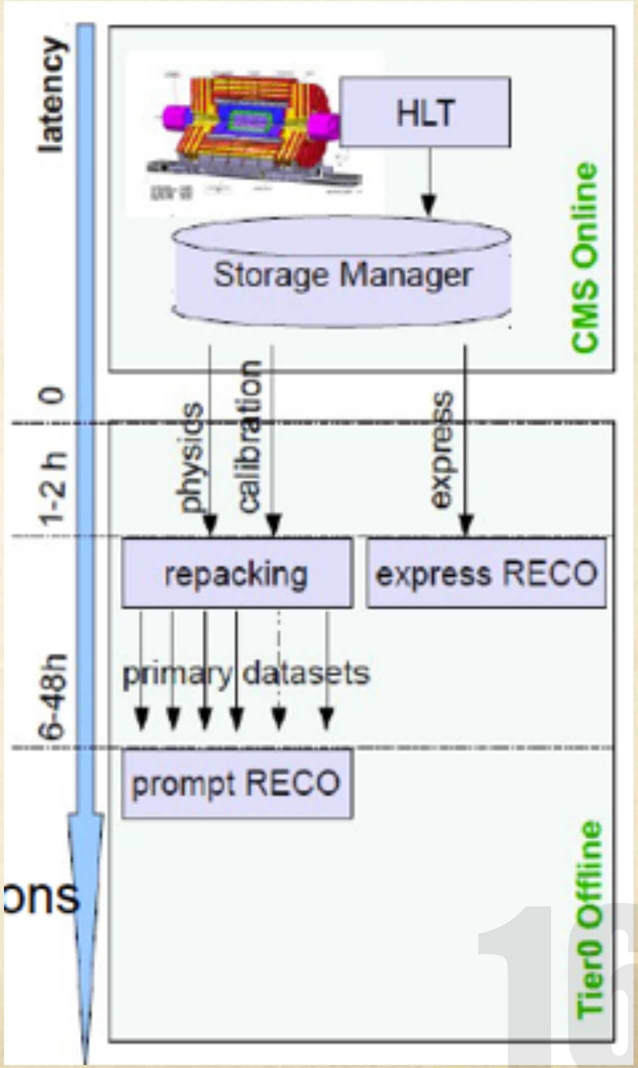
EXO-11-094 PAS



December 5, 2012 CMS Report, LHCC Open 112 L. Silvestris INFN-Bari

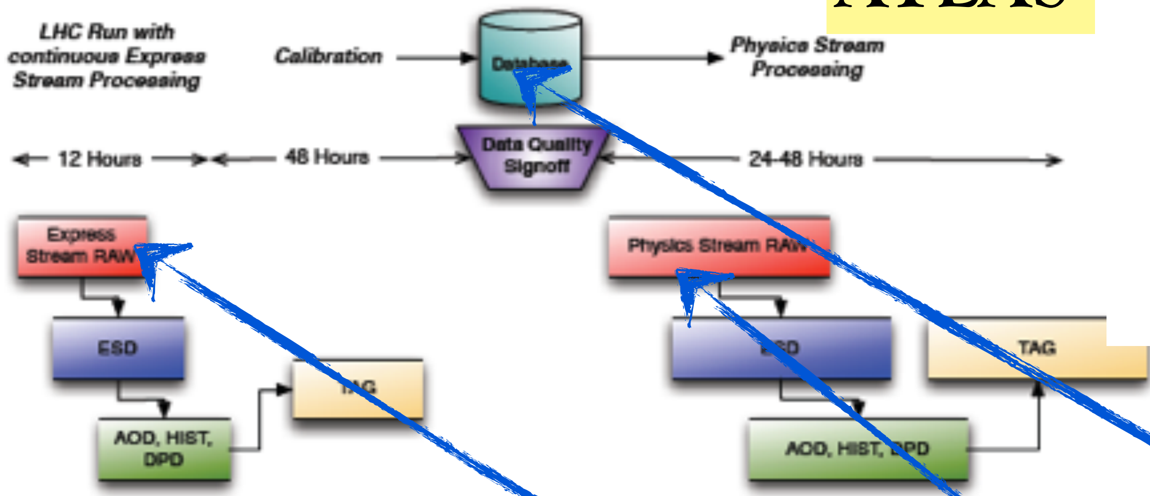
Data Streams & T0 processing

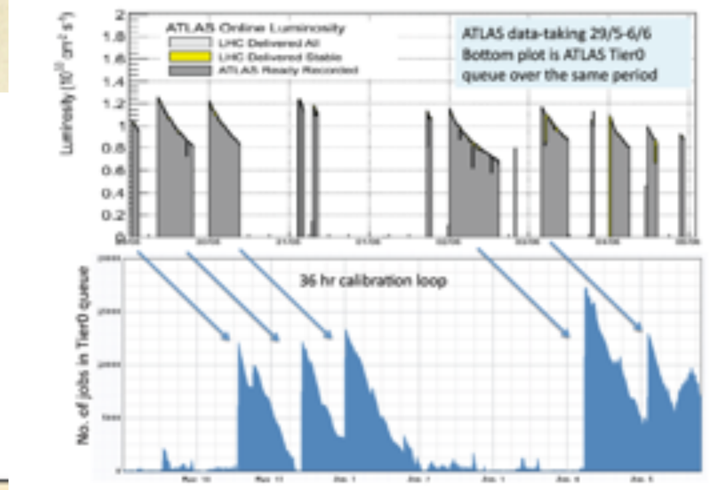
- Data Streams and the Tier0 workflows are specialized in different tasks.
- Depending on the latency:
  - **EXPRESS→PromptFeedback & calibrations**
    - Short latency: 1-2 hours
    - ~40Hz bandwith shared by calibration(1/2), detector monitoring(1/4) and physics monitoring(1/4)
  - **Alignment & Calibration Streams**
  - **Datasets for Physics:**
    - Split in Primary Dataset
    - Reconstruction delayed by 48h to get the latest calibration & conditions



The Prompt Reco flow

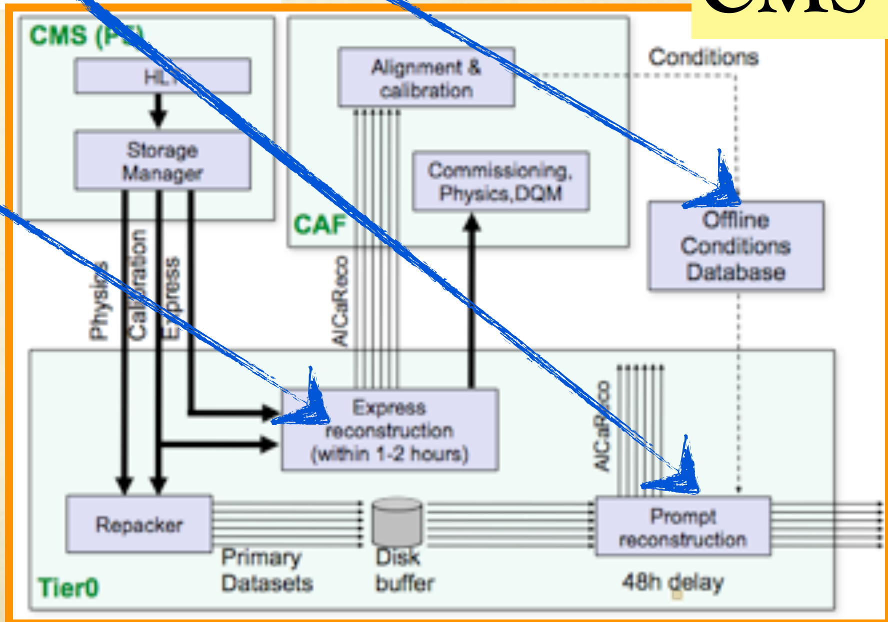
ATLAS





Visualization of the 48h delay

CMS



Some ATLAS # from 2012 to give the scale:

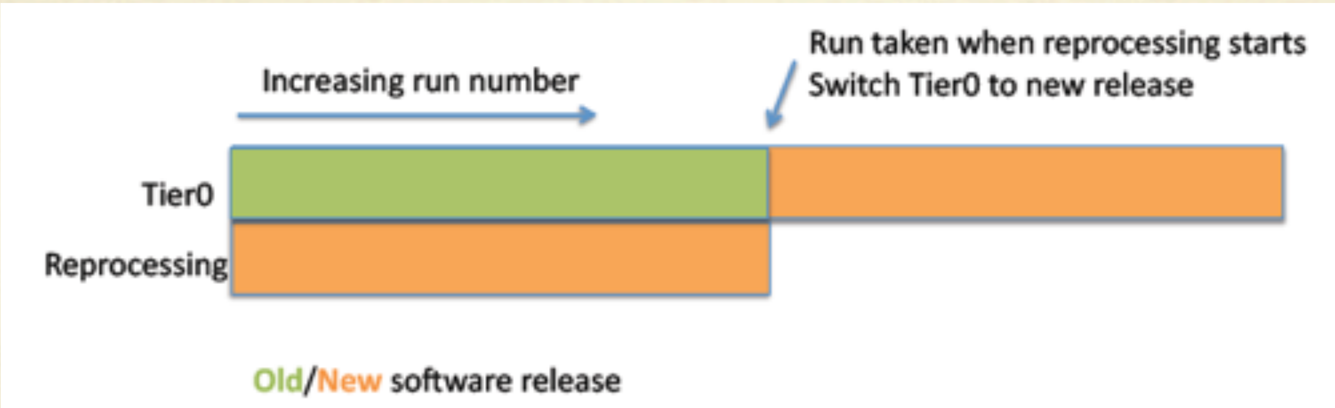
- 3.9 Billions of events (0.9 delayed/parked streams)
- 400Hz Prompt Physics
  - 3 physics streams: electron/Photon, Muon, JetTauMiss
- 130 Hz delayed Physics
- 15 Hz Zero Bias
- 10 Hz non colliding bunches

The Express Stream

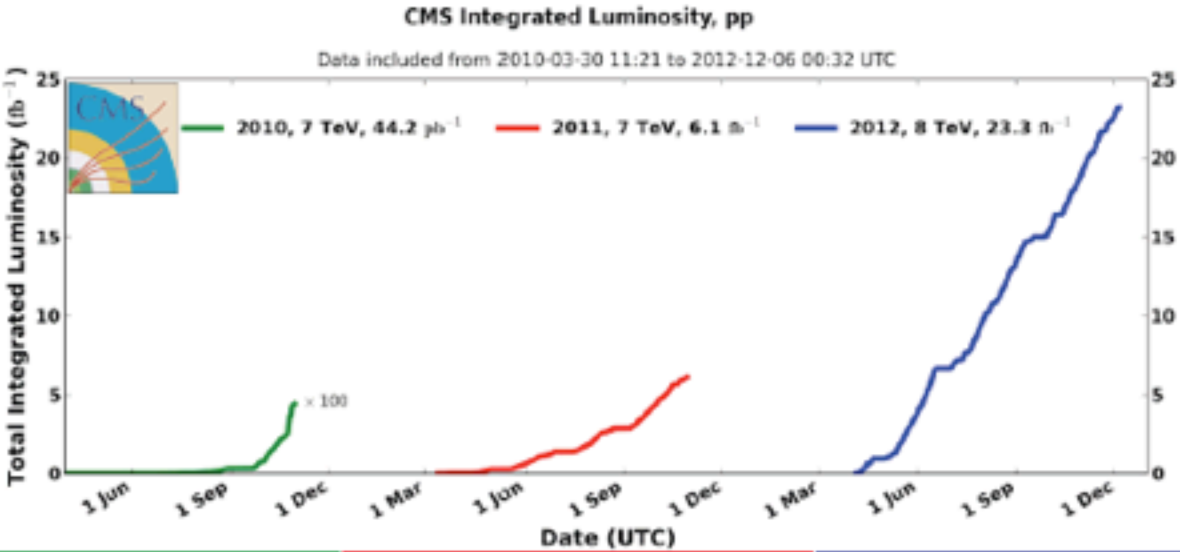
- This is a special stream used for data quality monitoring and calibrations.
- The RAW data are reconstructed very quickly in order to have feedback for the quality of the data taking --> ONLINE MONITORING OF DATA QUALITY
- For this reason the reconstruction of the Express stream does not have access to the latest and greates calibrations (beam spot, noisy/dead channels)
  - These calibration are extracted from these data (plus specific Alignement& Calibration streams in CMS) and later fed to the PromptReconstruction jobs at the T0.
  - PromptReco happens with a delay of 48 hours from the end of the run, to allow all the calibration jobs to be completed.
- This allows the data reconstructed at the T0 to be already very high quality for Physics analysis, but...

The data reprocessing

- Once updated reconstruction code or conditions are available the experiment can decide to «reprocess» all the data taken until then (usually targeting a major conference or for legacy).
  - when this happens, if the data taking is still ongoing, the experiment will also change the release that to process data at the T0
  - This allows analysers to always have complete and consistent analysis dataset.
- If the changes affect also the simulated data (such in the case of improved reconstruction) then a «reprocessing» of the MonteCarlo is performed as well.
  - in general there is no need for re-simulation, but only re-digitization and re-reconstruction (which are faster)
- During commissioning phase in 2010 many reprocessings/exp. Once stable data taking (2011/2012) only few per year



Reprocessing story



- 2010
  - ~15 Re-reco passes
  - 3 production releases (3.6, 3.8, and 3.9)
  - No prompt calibration loop

- 2011
  - ~3 Re-reco passes of MC and targeted reco)
  - 3 releases (4.1 used briefly, 4.2, and 4.4 used mostly in 2012)
  - PCL commissioned

- 2012
  - ~3 Re-reco passes and targeted reco)
  - 5_2 and 5_3
  - Good Management of calibration and validation

2

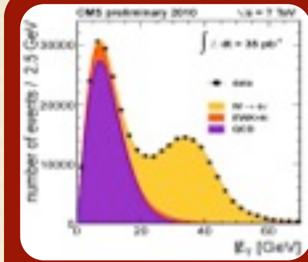
2013
- 1 rereco pass for full 8TeV data 5_3
- 1 rereco pass for full 7TeV data (legacy to be in the same release as 8TeV much easier for Data Preservation)

Collisions!

Detector Response

Trigger

Event Reco.



Physics Analysis

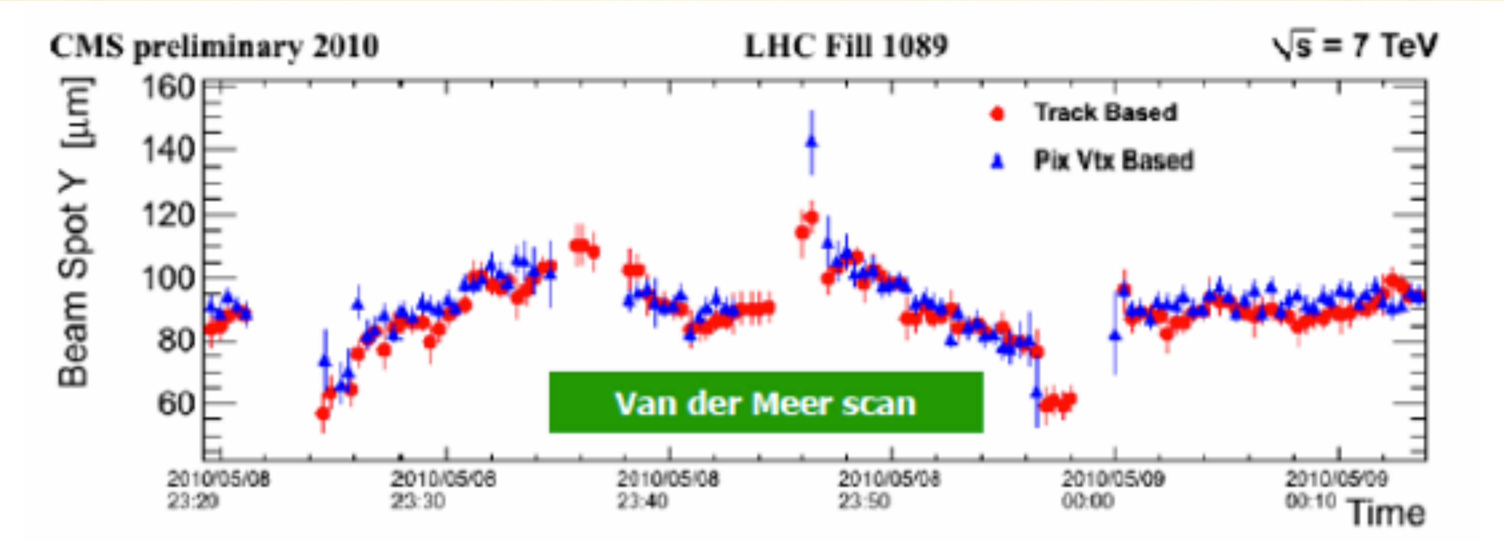
Calibration

Calibration Workflows Generalities

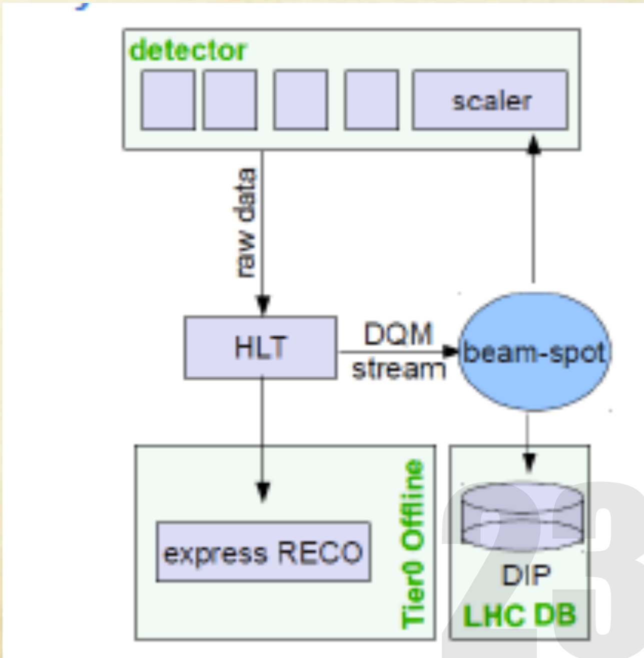
- The conditions and the environment surrounding the detector change continuously as we keep taking data. As a function of time, temperature, humidity, noise, accelerator conditions, etc etc… change and impact the response of our sensors.
  - Need to provide most up –to-date conditions @all stages of the data processing

- Different workflows exist depending on the time scales of the updates:
  - **Quasi-online calibration for HLT and Express:** **Immediate**
    - Beam-spot→quick online determination
  - **Prompt calibration**: monitor/update conditions expected to vary run-by-run (or less) **48Hours**
    - Updated conditions must be ready before prompt-reconstruction
  - **Offline re-reco and Analysis:** **~Weeks**
    - More stable conditions (i.e. alignement)
    - Workflows that need higher statistics: run on specific AlCa streams, in a specific reduced format to optimize speed and disk space.

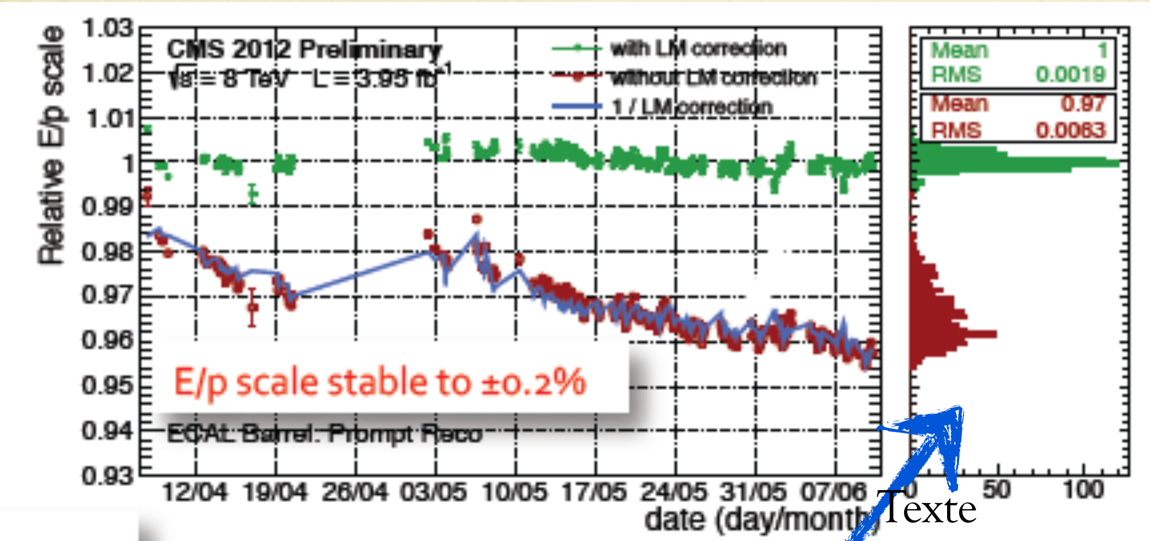
Online Beam Spot Calculation

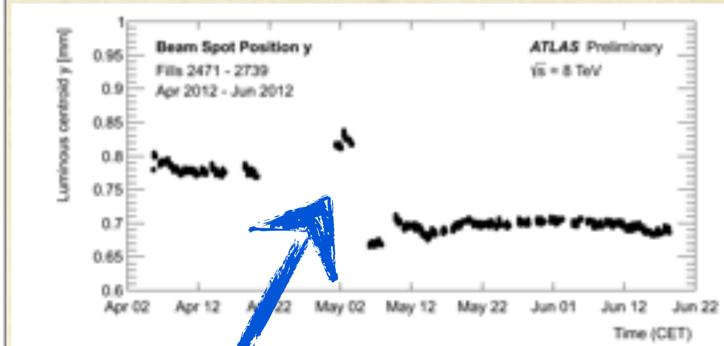


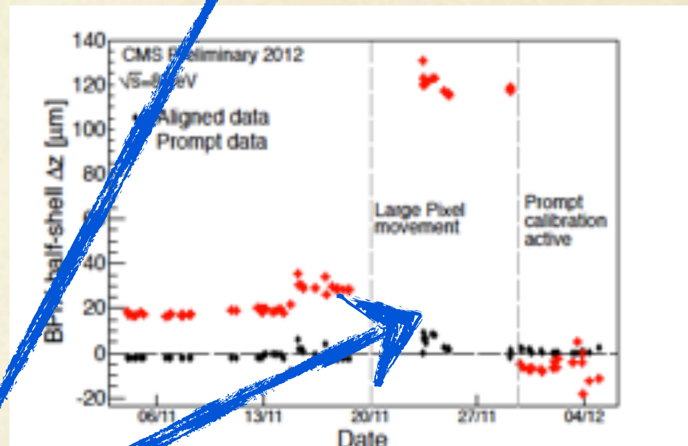
- Need to track the BS position as a function of time QUASI-ONLINE for HLT!
  - BS delivered every ~2min (CMS), 10min (ATLAS)
  - Use track based pixel-vertexing only (very fast)



Prompt-Calibration Loop





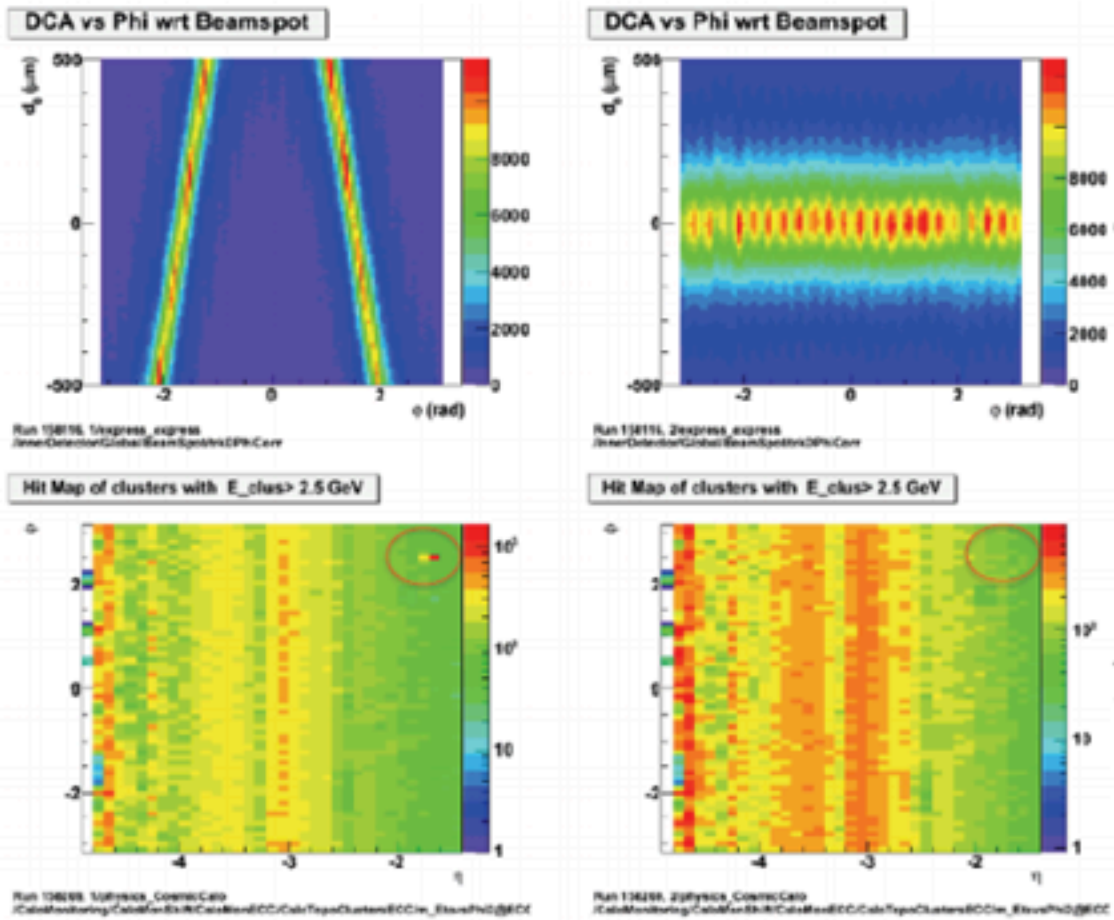


Relative shift between the 2 half-shells of BPIX

- Conditions that need continuous updates:
  - beam-spot calculation: every LS(23s)
  - tracker problematic channels: follow HV trips/noise
  - ECAL laser corrections (CMS)
- Conditions which need monitoring:
  - Calorimeter problematic channels: mask hot channels
  - tracker alignement: monitor movements of large structures affecting vertexing and btagging
- Dedicated streams (ALCARECO) out of Express (slightly different from ATLAS):
  - compute conditions in time for PromptReconstruction (48hours from end of run)

PCL: Beam Spot & hot Calorimeter channels

(ATLAS) 48 hr calibration loop



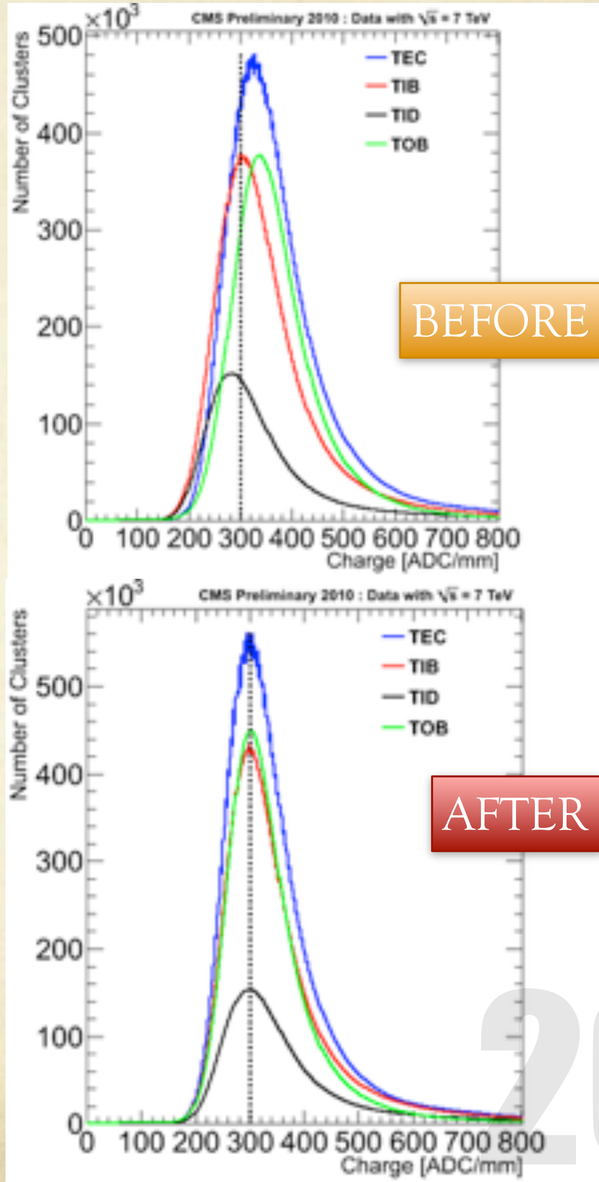
Beam spot before/after calibration loop

Hot channel in the calorimeter is masked in the physics stream processing.

20

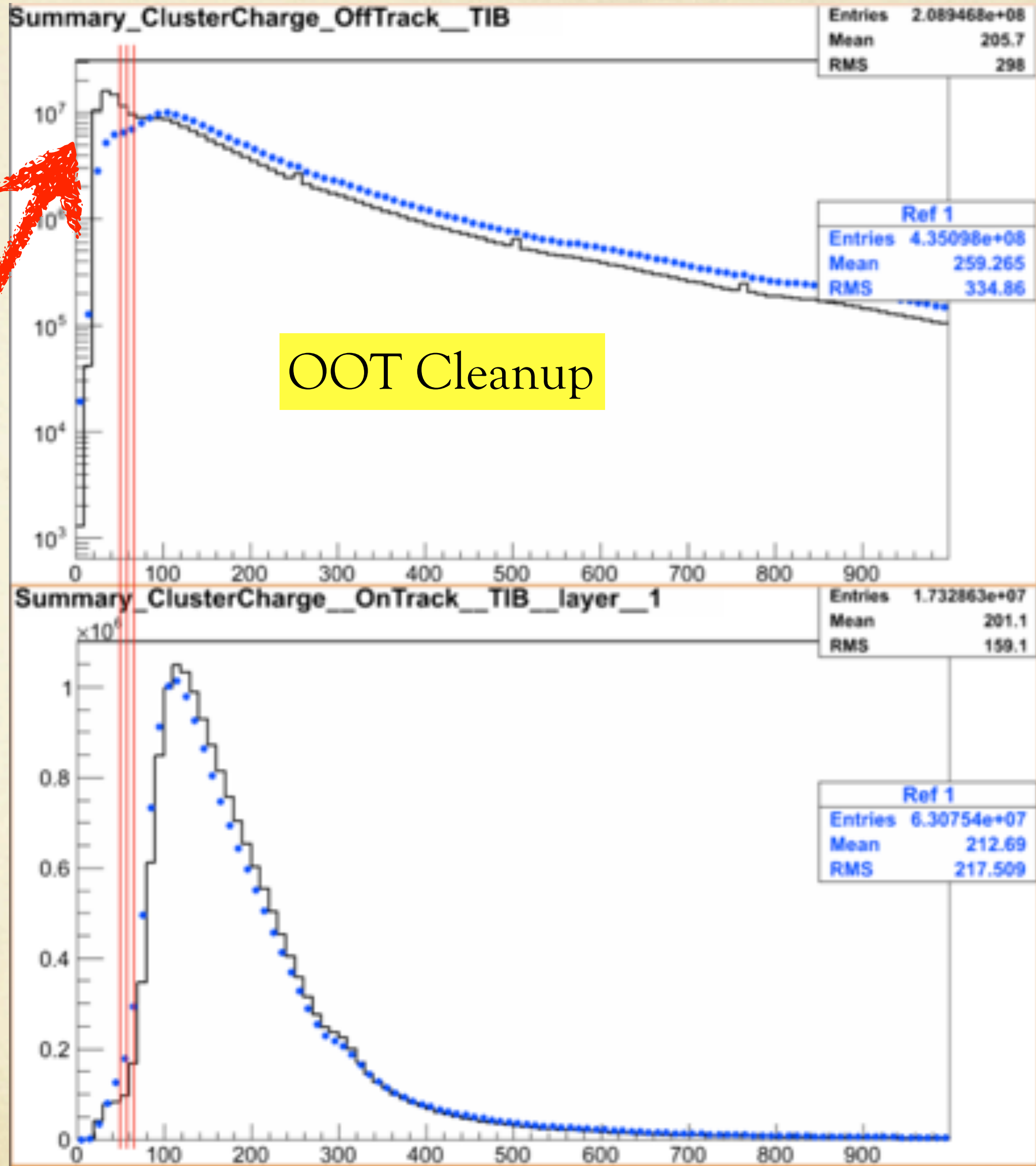
Future PCL workflows : Strip Tracker Gain Calibration

- CMS has an all Silicon Tracker, the largest ever built.
- The charge released in a silicon sensor by a charged particle is digitized into ADC counts assigned to a set of channels making up a cluster hit.
  - Non uniformities in the charge collection and in the readout chain can affect the correct amplification and linearity of the response: need to calibrate. Now being done by hand.
- This is a candidate for a new workflow in the PCL since un Run2 we can use the cluster charge to cleanup hits from PU



Ideas for Tracking in Run2 & connections to PCL

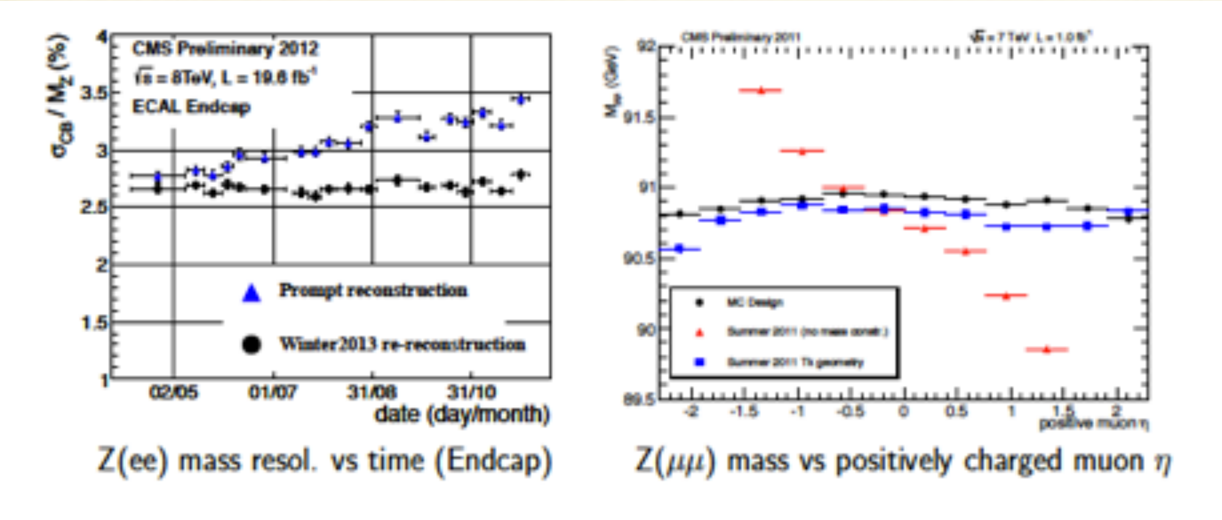
- Track reconstruction is the heart of CMS reconstruction
- need to keep it: efficient up to lowest possible pt, while keeping it as pure as possible, and achieving this in the shortest processing time.
- In RunII we expect scenarios: from ok (25ns-25PU) to nightmare (50ns-80PU)
- In particular the tracker will not be able to fully integrate the charge of the hits coming from particles belonging to «early/later» bunches OOT-PU
- studies show that a cluster charge cut to track reconstruction can help minimize OOT.
- Very important to have an automatized PCL calibration workflow!



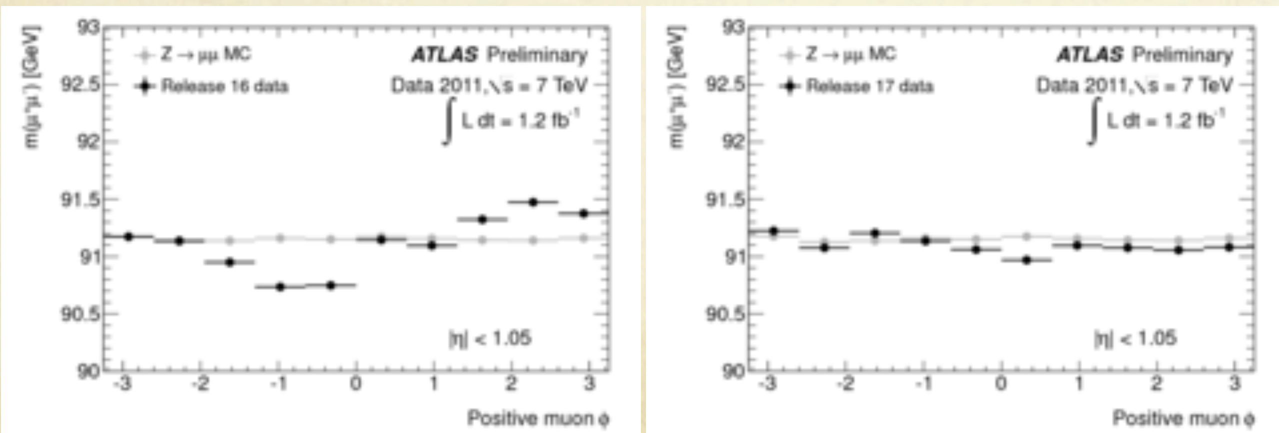
Alignment(I): the inner detectors

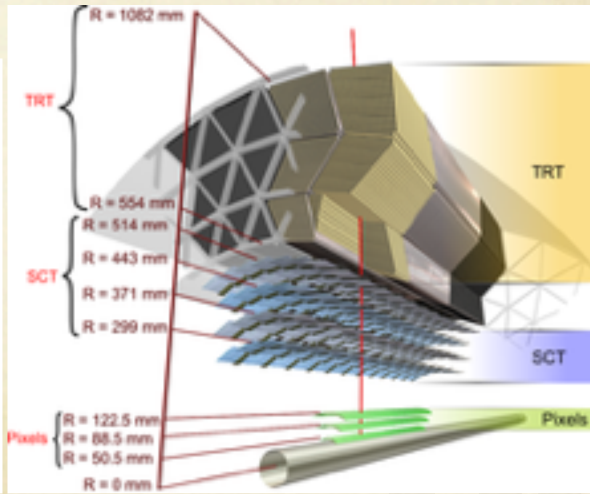
Not discussing here the specific methods!

- Alignements are released before major reprocessings.
- They affect the Data and the MC.
- Several other workflows depend on the Inner Detector Alignement (complex validation scheme)



Improvement on tracker alignement from PromptReco (CMS)

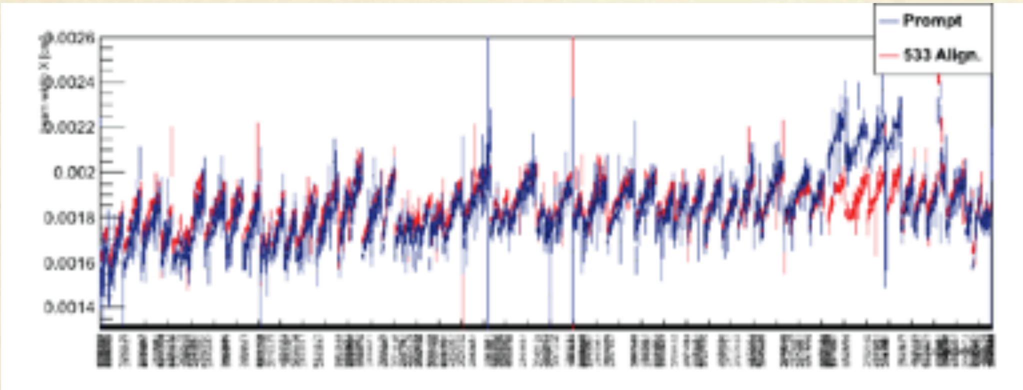


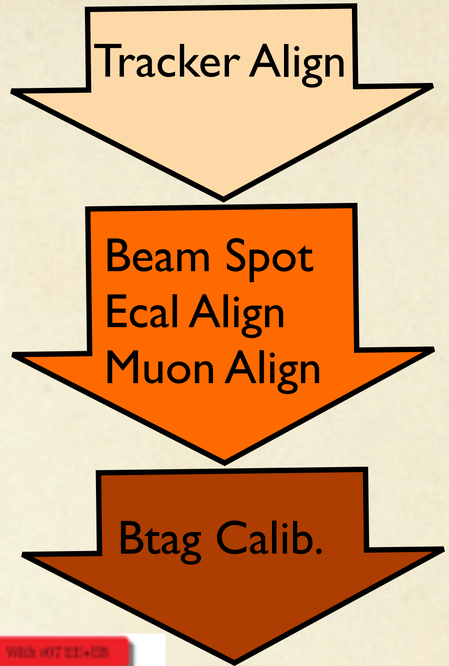


Improvement of inner detector alignement compared to Ideal MC(ATLAS)

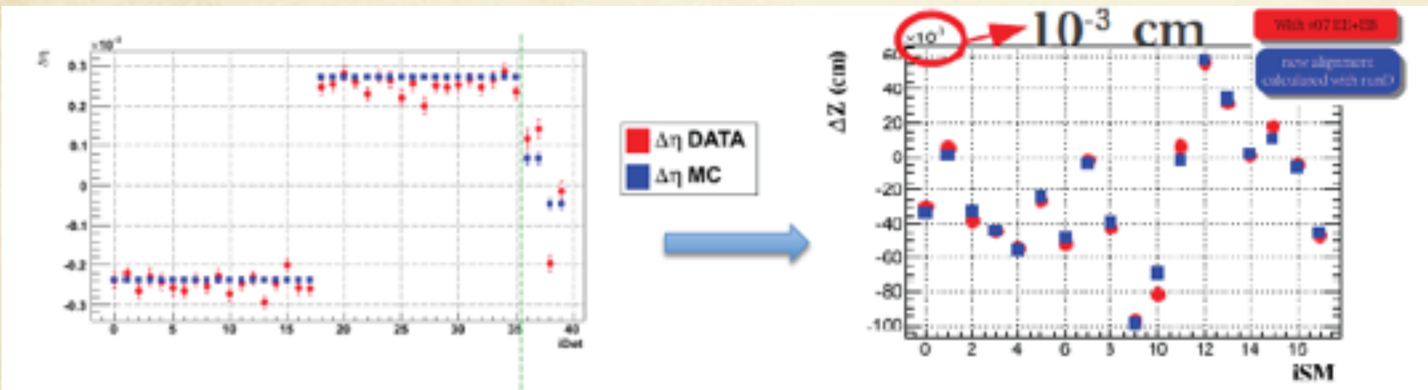
Alignement(II): Dependent Workflows

Recalculate the Beam Spot





Check the ECAL/ES alignement

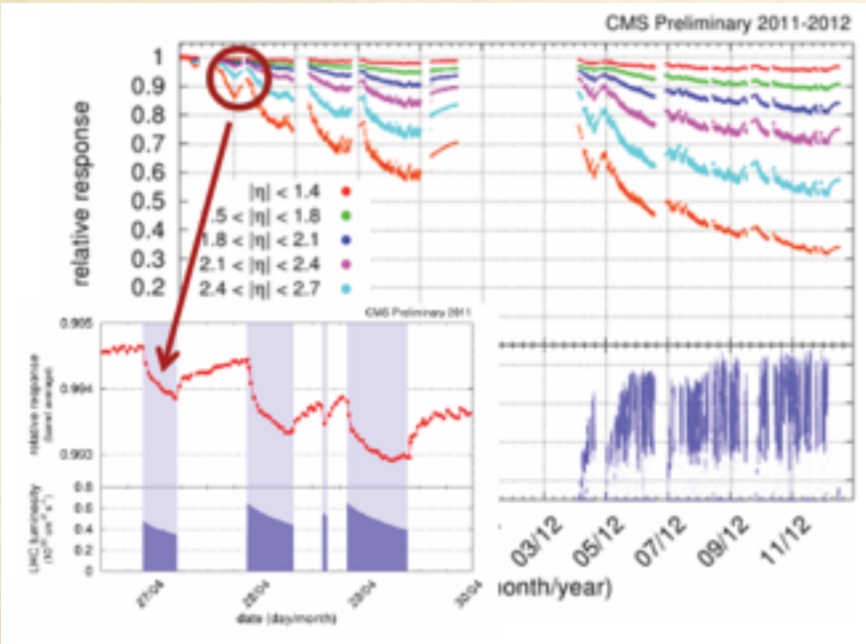


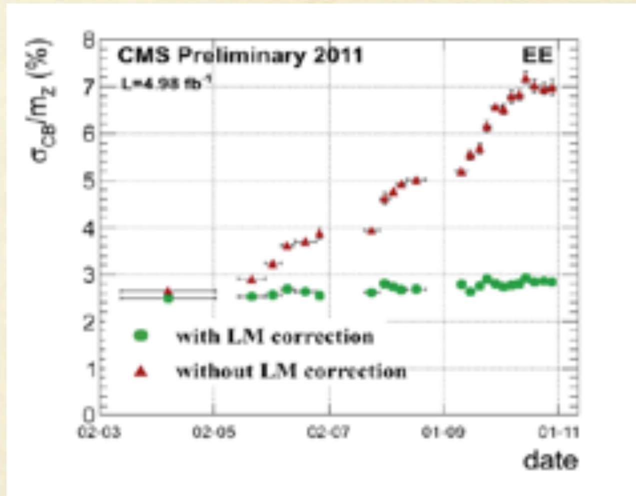
Check also: Muon alignement (as tracks are used to derive the Muon alignement), and Btag parameters as well (as they are obviusly very sensitive to tracking,vertexing performances)

The Calibrations of the CMS ECAL

$$\frac{\sigma}{E} = \frac{2.7\%}{\sqrt{E}} + 0.5\% + \frac{150MeV}{E}$$

- ECAL is the first crystal ($PbWO_4$) calorimeter installed at a hadron collider. Hermetic and homogeneous.
- To maintain the constant term of 0.5% in situ calibration and monitoring is needed. Exposure at the nominal LHC luminosity cause loss in crystal transparency due to radiation induced absorption.
- Laser monitoring system: 1 measurement/channel/40min
- Phi-symmetry Intercalibration: 1 measurement/4 days (use special Alca Stream in ZeroBias events)
- pi/eta intercalibration: 1 measurement/1.5 months (use the γγ peak)

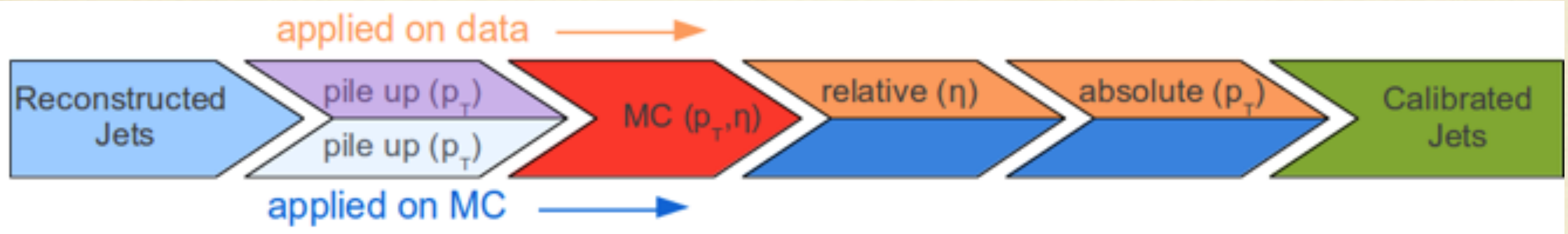




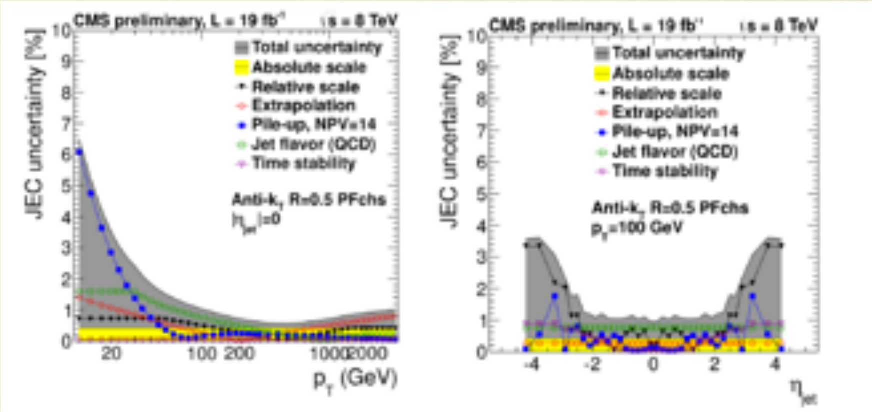
Z->e^+e^- mass stability before and after LM corrections

Analysis Level Calibrations: jet energy corrections

- Jet energy corrections are extracted only once the data have been fully reprocessed.
  - Specific samples are needed to determine them.
  - By construction they take a long time to be determined and they change only with the reprocessing version (once or twice a year).



- what do we correct? we correct reconstucted jets back - on average - to particle level
- PileUp Correction to correct for offset energy
- Correction to particle level jet response vs pt and eta (from Simulation)
- Only for data: small residual correction (relative corrections) to flatten the response in eta, and absolute to compensate the remaining difference between data and simulation

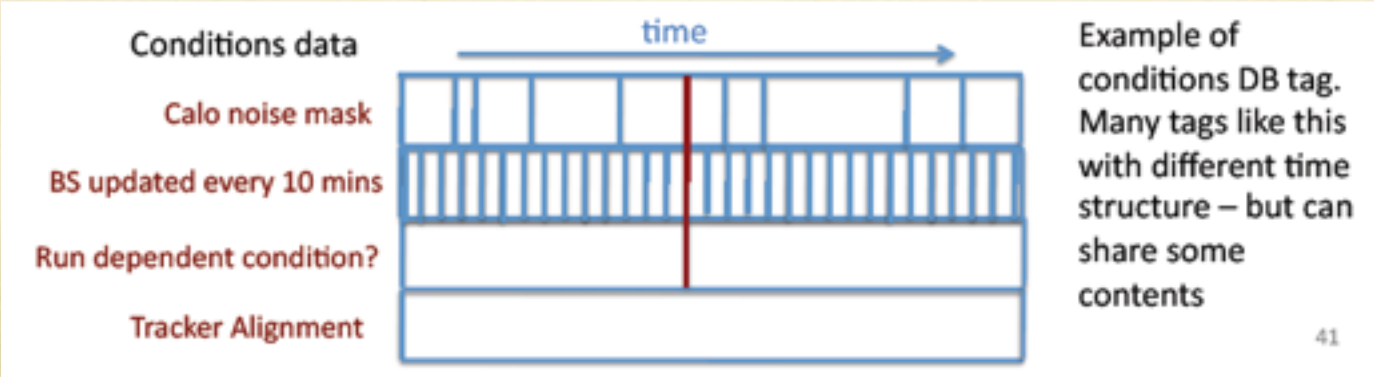


- Full validation cycle before being used!

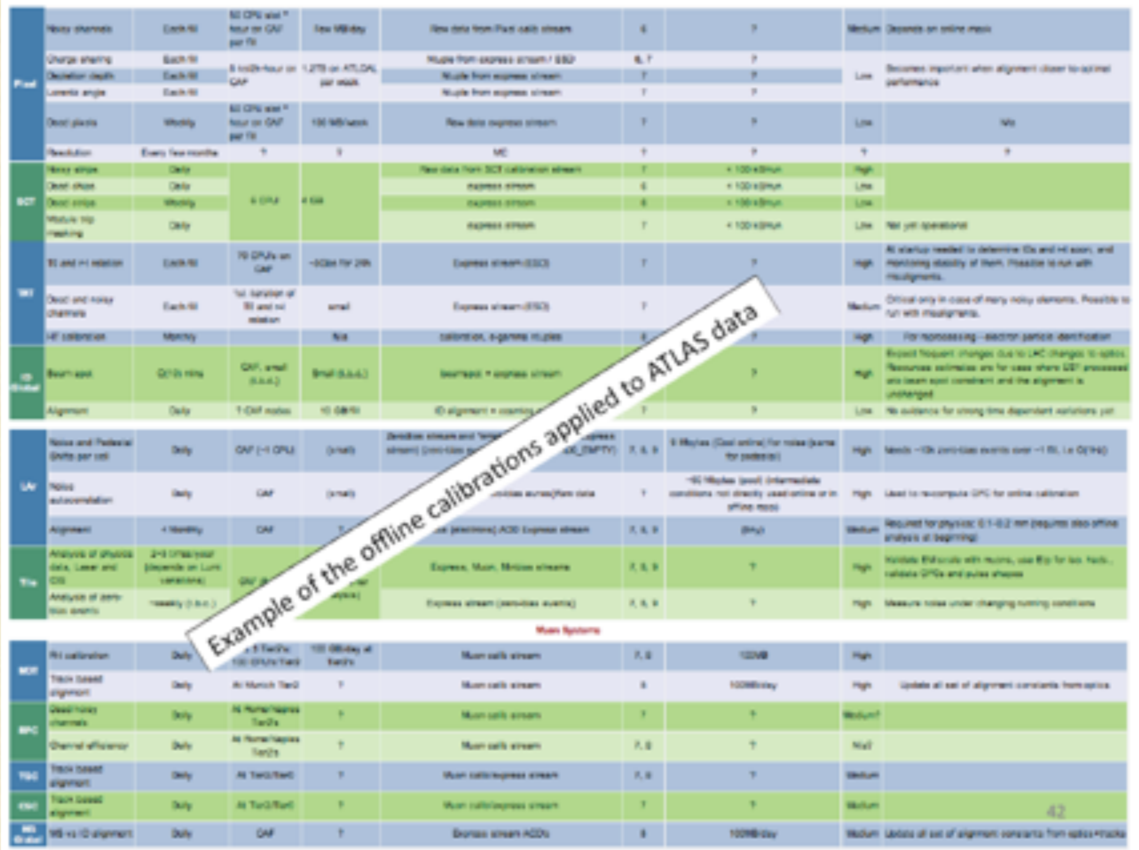
vendredi, 30 août 13

Time dependence of Calibrations/Conditions

- Calibration/Conditions come with a specified «lifetime» call «interval of Validity» (IOV)
  - some of them change rapidly, even within a run
    - beam-spot, noisy channels..
  - some change very slowly with time:
    - radiation damage calibration, alignement
  - some change only with a specific version of the Reconstruction or the geometry of the detector(Upgrades)
  - some are defined only for the Data, or the MC, some affect both
- Sophisticated DataBase structure to keep track of all this.
  - Sometimes big changes are seen when one condition is changed even for an identical reconstruction version (Validation!)
  - Need to have versioning, reproducibility, evolution schema including the Upgrade
  - make it easy for the analyser to use them

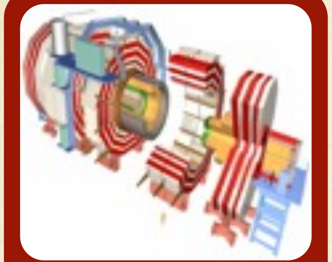


Organizing the calibration information



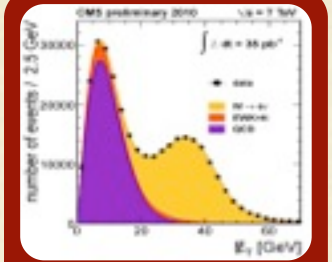
- New challenges for the future: now experiments need to follow up an increasing number of «detector configurations»: RunI, Run II, Phase I , Phase II, analysis!
    - for each of these configuration there is the complete set of calibration/conditions!
- Significant developments needed to make the system flexible and robust, and most of all easy to be used by the physicsist doing analysis.


Collisions!


Detector Response

Trigger

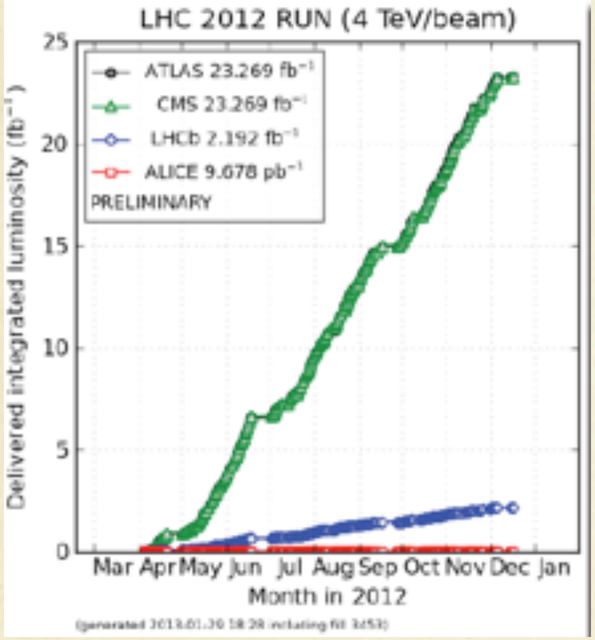

Event Reco.


Physics Analysis

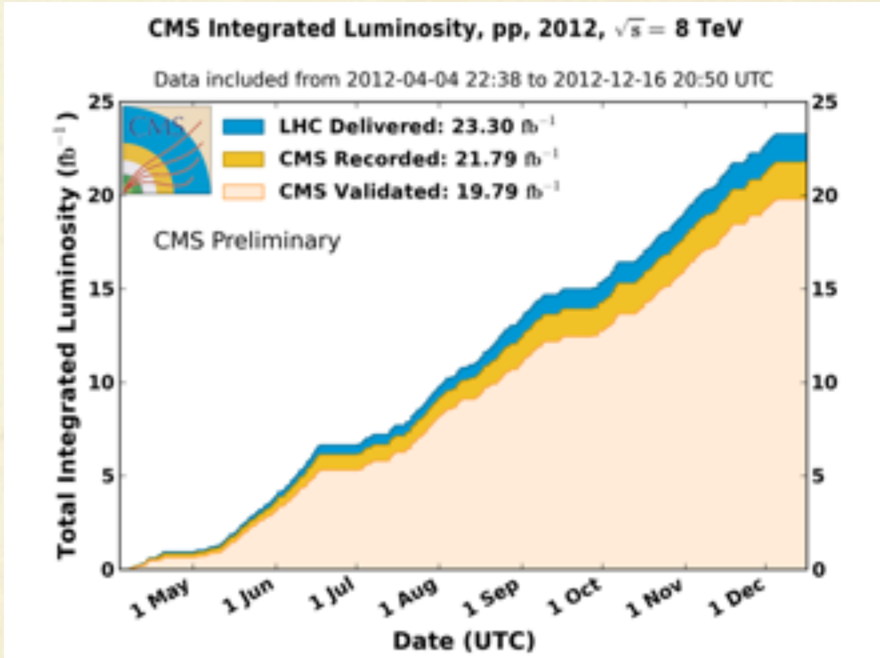
Data Quality & Certification

Luminosity definitions

DELIVERED Luminosity: depends on the LHC performance (varies with the interaction points)



RECORDED Luminosity: depends on the fraction of time when the detector (CMS) was taking data



Lots of time&effort spent to understand how to optimize the fraction of Recorded and Validated: more data for Physics!

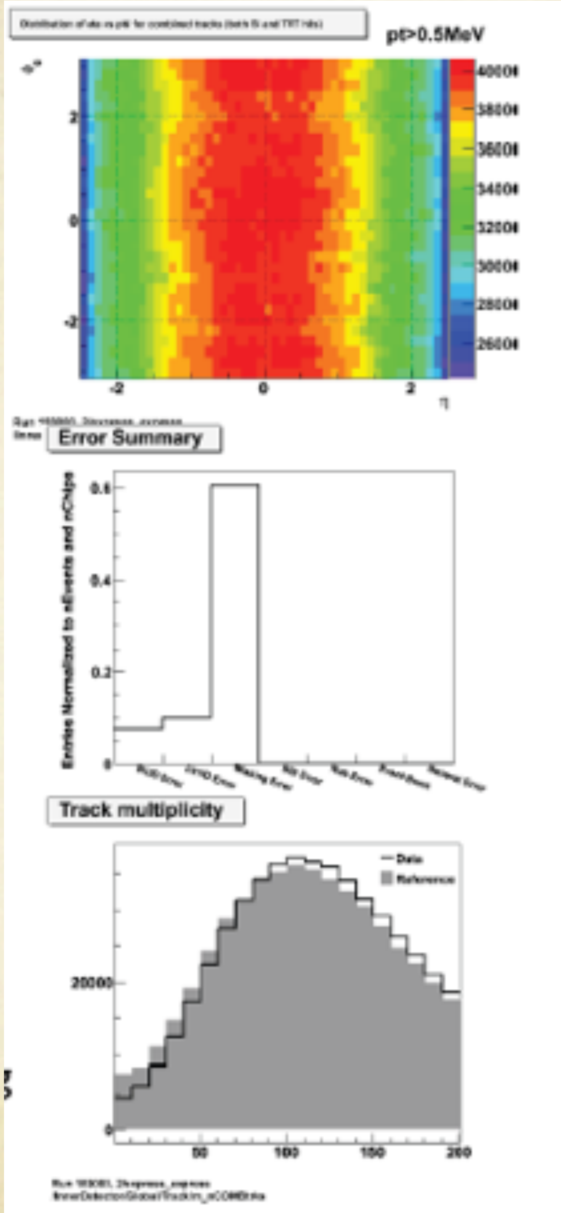
VALIDATED Luminosity: fraction of luminosity checked to be good for Physics (in this plot «Golden» fraction)

Monitoring the data quality

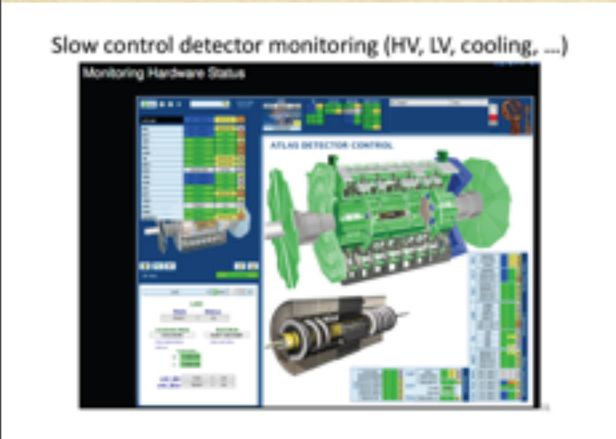
- **ONLINE:** Need fast monitoring of detector performance during data taking. There is even a dedicated event stream
- **OFFLINE:** need to monitor the performance of physics objects reconstruction in different instances:
  - Express Reconstruction: fast turnaround for data used in on-line calibration
  - Prompt-Reconstruction: continuous monitoring (24h-7d shifts) + certification
- Framework that provides sets of histograms for all the interesting quantities and tools to display them on a GUI and to compare them.
  - Along with these is very important to have a flexible way to record the status (OK/NOT) of the various components with a granularity of a LS (1m@ATLAS, 23s@CMS)
    - these Databases (RunRegistry/Defects) allow a simple way to extract the good run list given specific requirement from the analysis.
    - Allow to modify the status of a component in case of fixes/reprocessing/review
    - allow to optimize the usable luminosity for the analysis given the granularity of the information (a muon analysis might not care for a malfuction in the Calorimeters)

Quantities monitored

- **Local reconstruction:** hits multiplicity maps (in eta/phi or hardware space)
  - look for dead/noisy regions
  - extremely useful for quick feedback (note >100M channels!)
- **Errors in the data stream:**
  - Counting the DAQ errors
  - Reconstructions code errors
- **Global reconstruction:** object multiplicities (tracks, jets, muons) and related quantities (hits on tracks...), quality resolution and efficiencies
  - can be complex as Z->ee tag&probe analysis (???)
- **Noise monitoring** uses special EMPTY triggers (i.e. no collisions)
- **Granularity? Input streams?**
  - need to be granular to be able to minimize the data to mask as bad
  - need to have enough statistics to see an effect
  - The definition of which selections are used for the data that feed the various monitoring code very important.
- **Reference histograms:** essential to set automatic alarms. Need to be kept up to date with data type and running conditions
  - for example: changing the pileup...

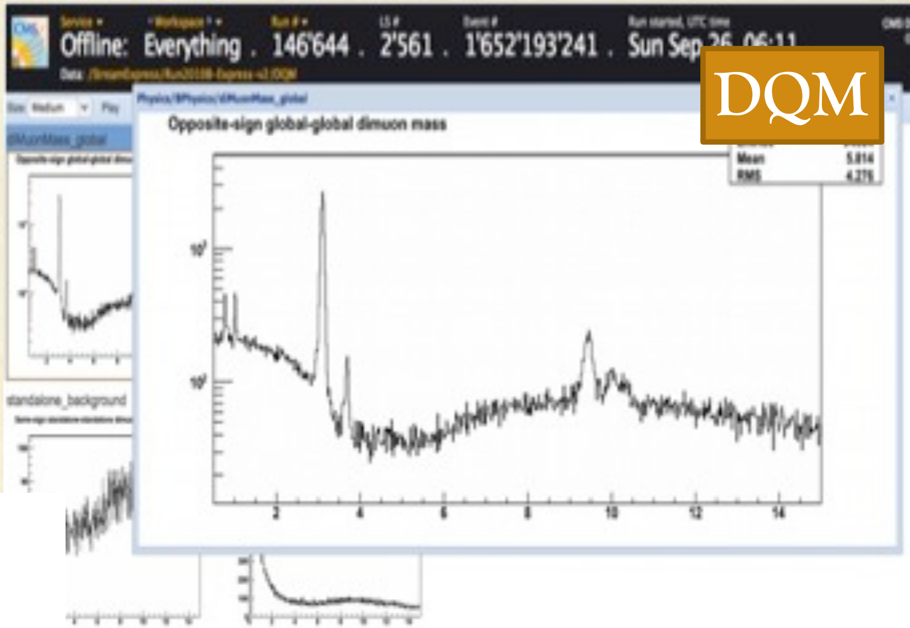


Detector slow controls

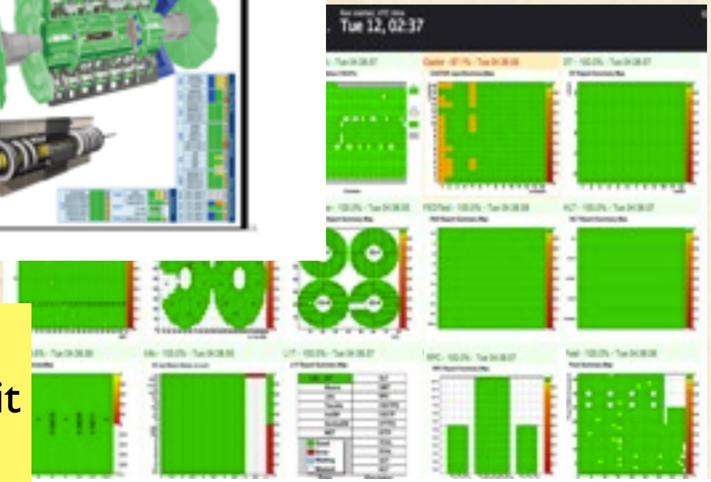


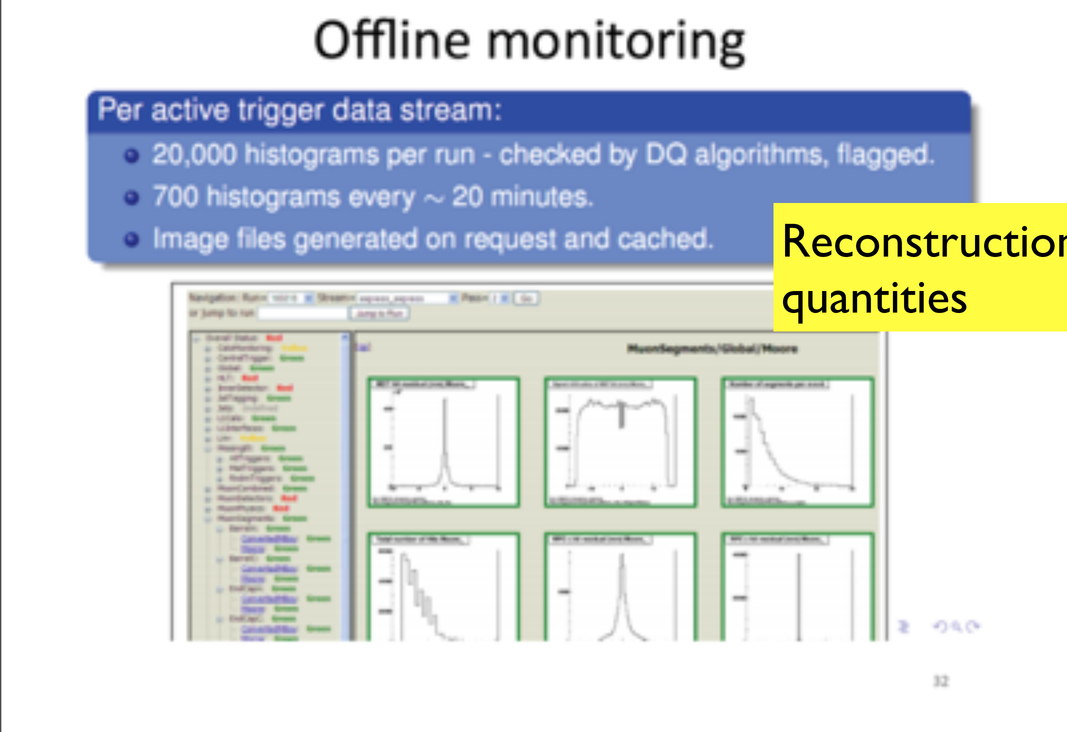
Physics quantities: M(mumu)

DQM



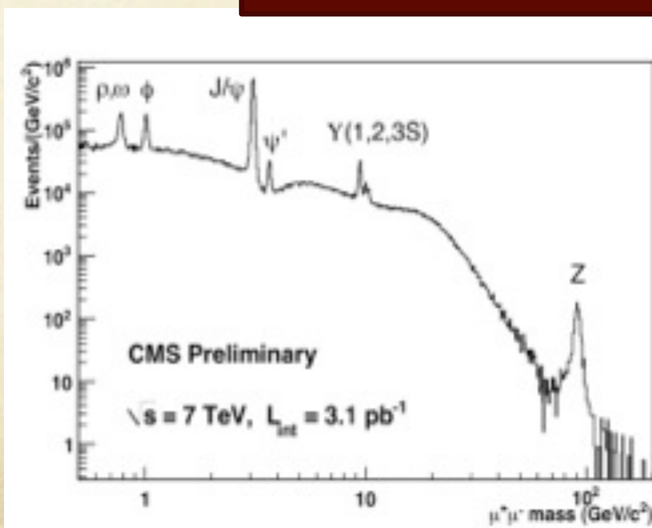
Detector local information (hit occupancy, ...)



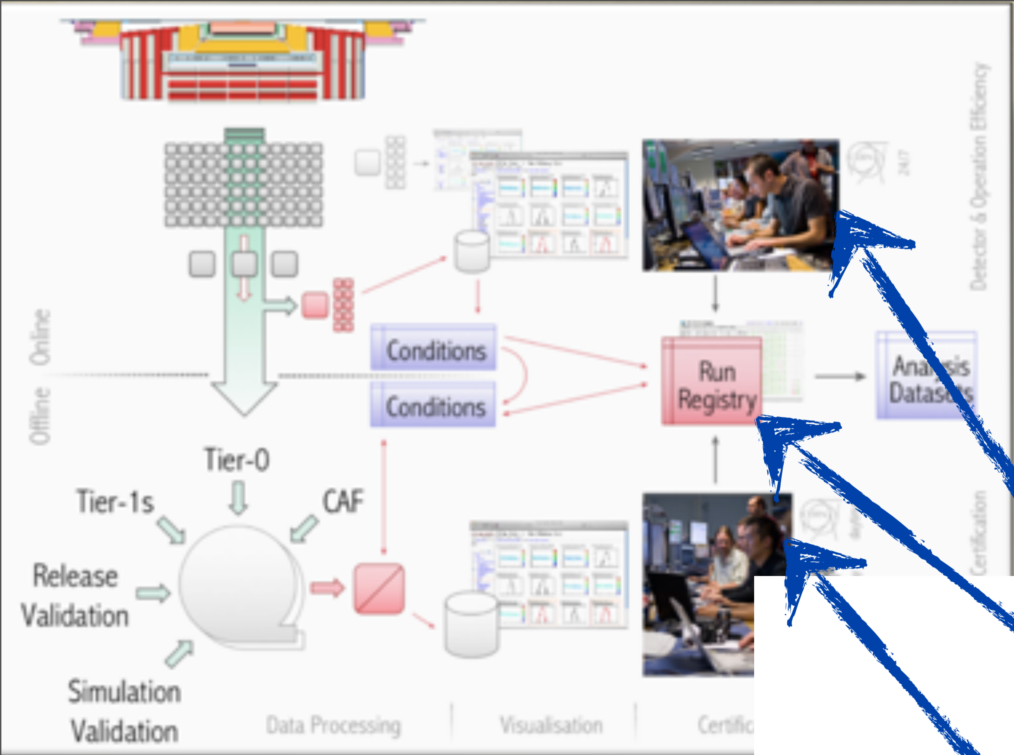


Reconstruction quantities

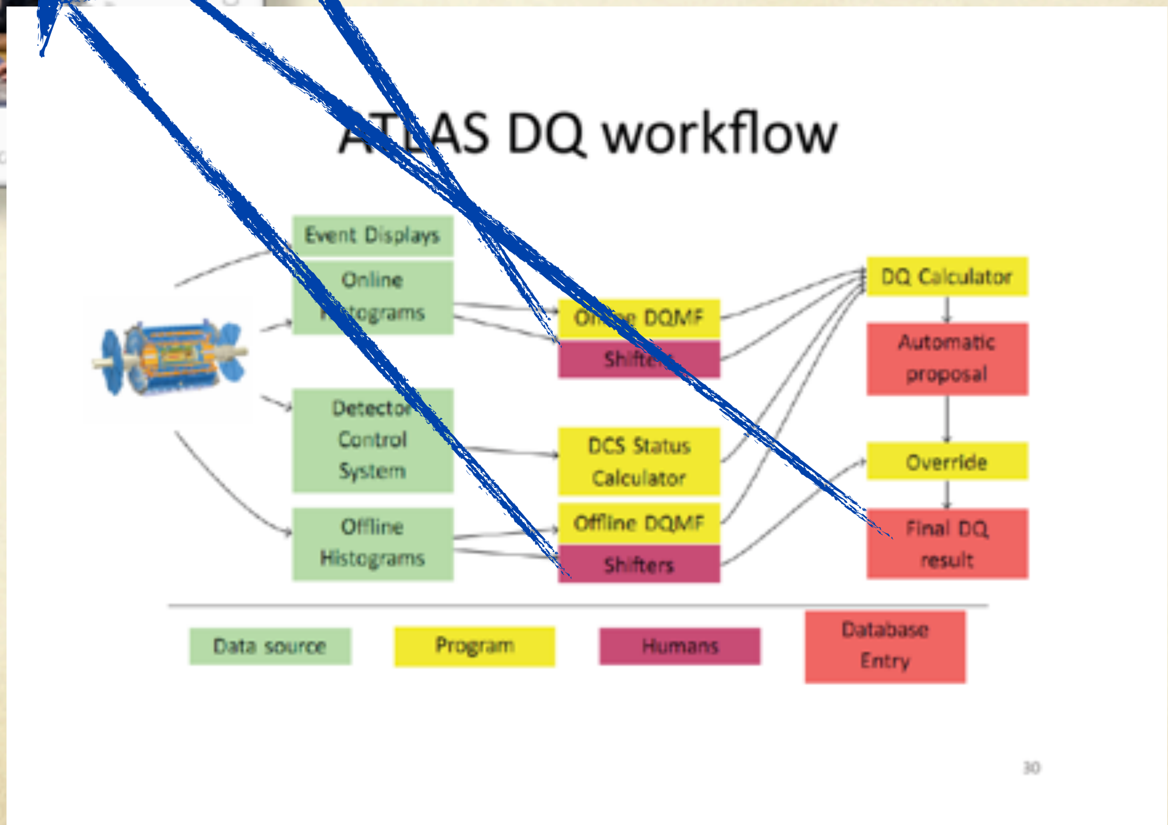
FULL ANALYSIS



Data Quality Workflows



- Very similar structure in differenc experiments
- Mix of automatic and human intervention
- Results in DB



Certification Workflow

- Shifts ongoing 24/7, in addition to the shifters and the automatic checks there are also checks performed by the «detector experts» that can modify by hand an entry in the DB.
- Some entries in the DB are automatic: those that come from the detector conditions (DCS) such as HV, Cooling...etc
- Usual turn around for signoff is weekly. In case of exceptional need (such as upcoming Conferences) the management can require a speed up of the procedure (min 4 days).
- Infrastructure to allow users to (re)create the Good Run list to use for their analysis and calculate properly the corresponding Luminosity.

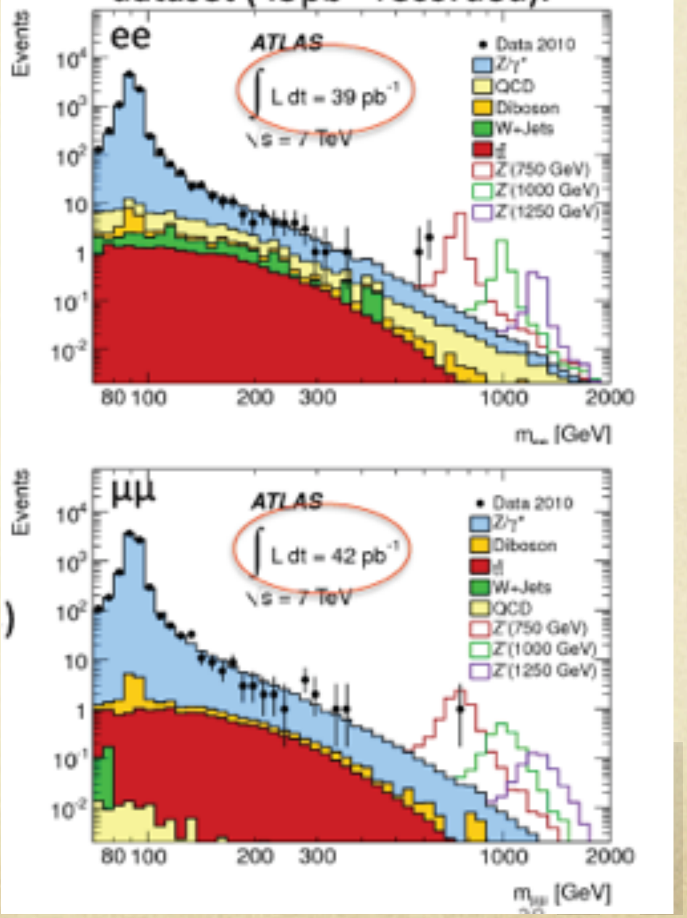
Defining the Good Run List(s)

Example from 2010: now difference is negligible

- Granularity of the «defects/problem» information is at the subdetector level and lumisection level.

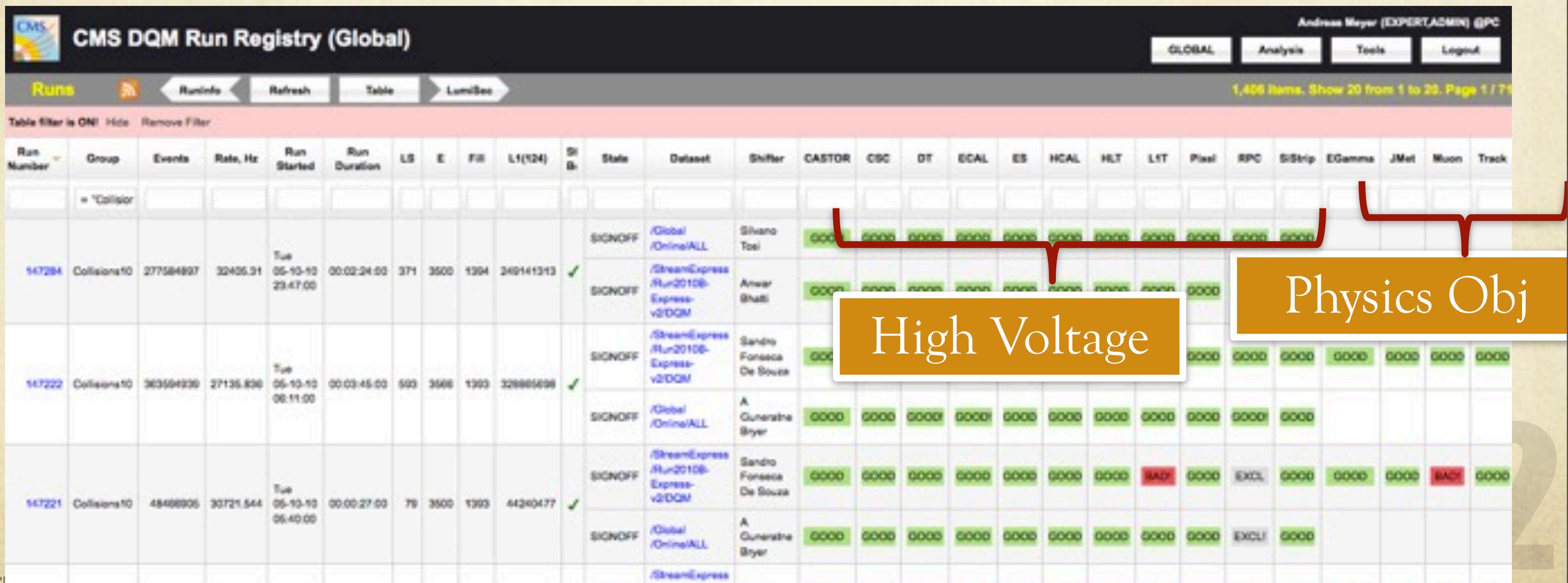
- All experiments mantain different good run lists for different analysis, the minimal set is
  - «Golden» that means that everything is perfect
  - «muon» for those analysis that rely on the muon information and are not dependent on the Calorimeters
  - however there are specific analysis that might require very specific characteristics and select of more/less/different events
  - However, the quality of the data is so high that at the end of Run I ATLAS released a single «Golden» good run list.

ATLAS Z' plots from full 2010 dataset ($45 pb^{-1}$ recorded).



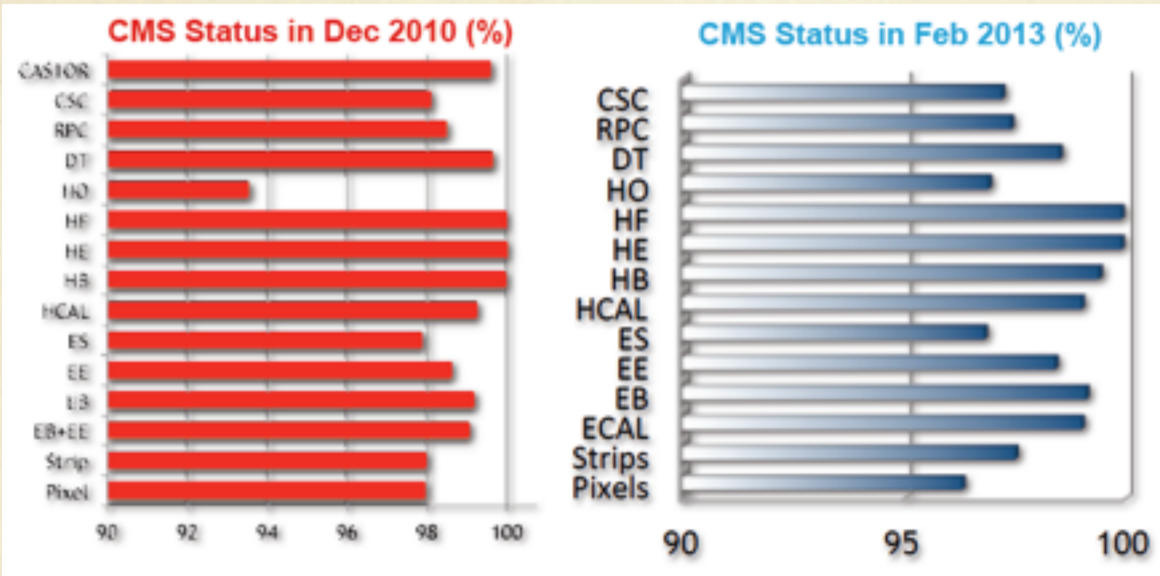
A single place for the information

- All the information on the data quality from the detector level up to the physics objects is saved in a database (RunRegistry or Defects DB).
  - The granularity is of a LumiSection (23s/1m)
- **It is the only source to define the good run list for the analysis. Allows full reproducibility of the luminosity calculation**



how do we lose Recorded luminosity?

- There are dead times during data taking: warm up of the detectors, raising the HV, trips that need to be recovered, stops in between two trigger table changes, magnet trips...
  - the responsability for this data loss lies in the hands of the Detector+Trigger



No degradation and even some recovery!

- However, most of the time these detectors perform beautifully. There are some other things that can still go wrong and make the data that have been collected not «good for Physics»
  - the responsability for these data loss lies in the hand of Detectors, Trigger but also Software, Computing and Calibration. It is one of the biggest jobs for the «Data Preparation» groups.

How can we optimize the luminosity for Physics when things go wrong?

Software/ conditions failure

Recoverable

It is possible to recover the data fixing the problems with a reprocessing.
No impact on the simulation of the detector!

Hardware Failure

Recoverable

Evaluate impact on physics to decide if the data should be kept/ thrown out.
There could be consequences on the simulation of the detector

Unrecoverable

Evaluate impact on physics
Improve Reconstruction
Adapt simulation

Improving the certification efficiency

ATLAS p-p run: April-December 2012										
Inner Tracker			Calorimeters		Muon Spectrometer				Magnets	
Pixel	SCT	TRT	LAr	Tile	MDT	RPC	CSC	TGC	Solenoid	Toroid
99.9	99.1	99.8	99.1	99.6	99.6	99.8	100.	99.6	99.8	99.5
All good for physics: 95.5%										

Luminosity weighted relative detector uptime and good quality data delivery during 2012 stable beams in pp collisions at $\sqrt{s}$=8 TeV between April 4th and December 6th (in %) – corresponding to 21.3 fb^{-1} of recorded data.

Improvement of 7% more good data for physics due to the reprocessing.

ATLAS 2011 p–p run												
Inner Tracking			Calorimeters				Muon Detectors				Magnets	
Pixel	SCT	TRT	LAr EM	LAr HAD	LAr FWD	Tile	MDT	RPC	CSC	TGC	Solenoid	Toroid
99.9	99.8	100	89.0	92.4	94.2	99.7	99.8	99.7	99.8	99.7	99.3	99.0

Luminosity weighted relative detector uptime and good quality data delivery during 2011 stable beams in pp collisions at $\sqrt{s}$=7 TeV between March 13th and June 29th (in %). The inefficiencies in the LAr calorimeter will partially be recovered in the future. The magnets were not operational for a 3-day period at the start of the data taking.

New code allowed to apply an «event veto» to events containing LAr noise bursts

ATLAS 2011 p–p run												
Inner Tracking			Calorimeters				Muon Detectors				Magnets	
Pixel	SCT	TRT	LAr EM	LAr HAD	LAr FWD	Tile	MDT	RPC	CSC	TGC	Solenoid	Toroid
99.9	99.8	100	96.3	98.6	98.9	99.7	99.8	99.8	99.8	99.7	99.3	99.0

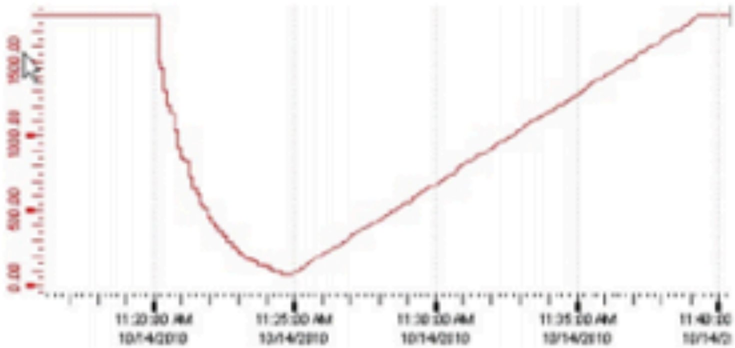
Luminosity weighted relative detector uptime and good quality data delivery during 2011 stable beams in pp collisions at $\sqrt{s}$=7 TeV between March 13th and June 29th (in %).

Example of improvements to increase DQ efficiency

HV trips in ATLAS LAr calorimeter cause loss of good data.
Ramping up the HV causes noise in the detector.
Full trip + ramp-up takes ~20mins (bad data quality).

Trip detection and autorecovery implemented – now 2mins of data with bad data quality

LAr trip, ramp to zero and recover - takes 20 minutes

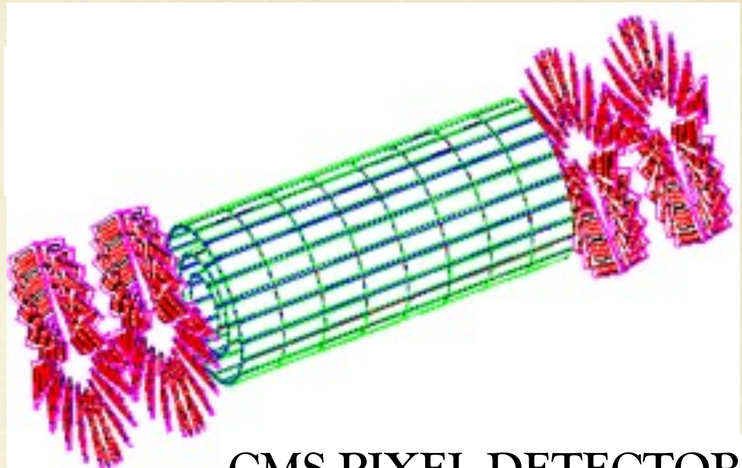


14 oct

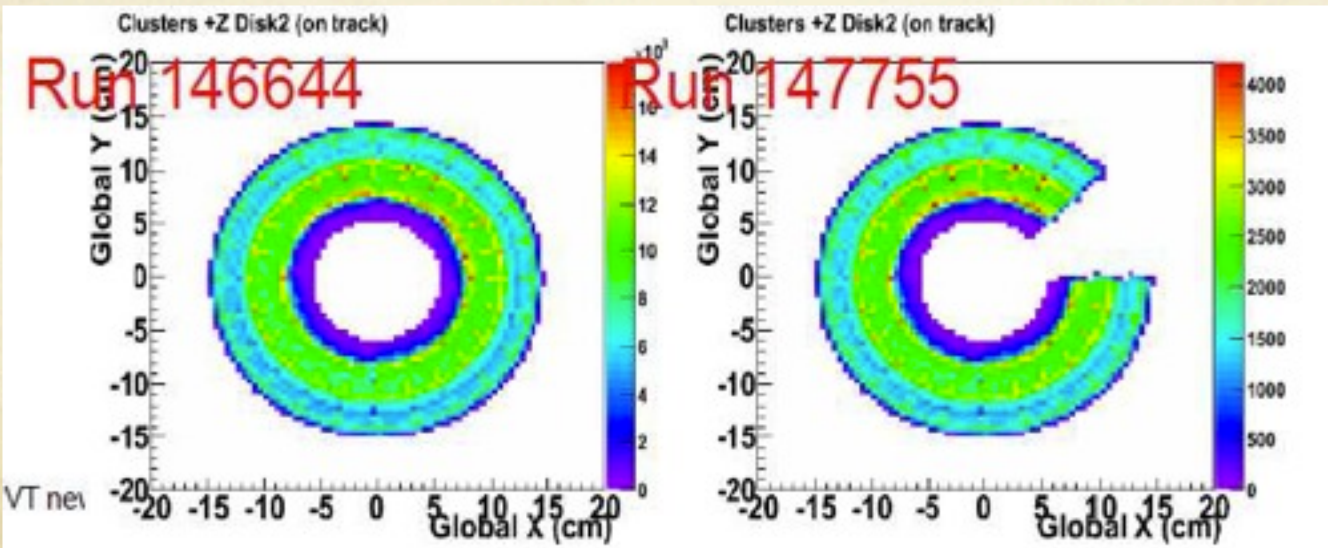
26 oct

what to do in case of hardware accident

- On Sunday October 10th 2010 a few channels of the CMS Pixel Endcap detector stopped functioning.
- This problem was currently defined as “permanent” and it will be fixed during thelong shutdown.
- Clearly defining all the data «bad» for this reason was not an option.

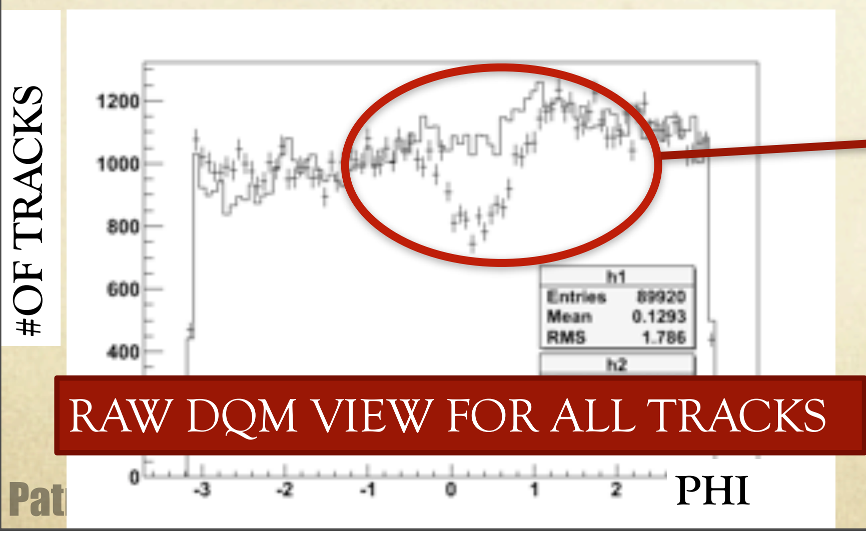


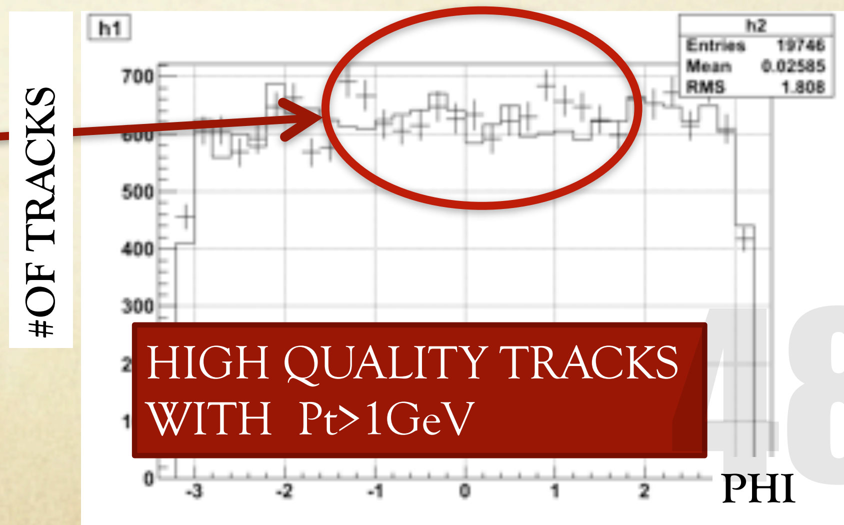
CMS PIXEL DETECTOR



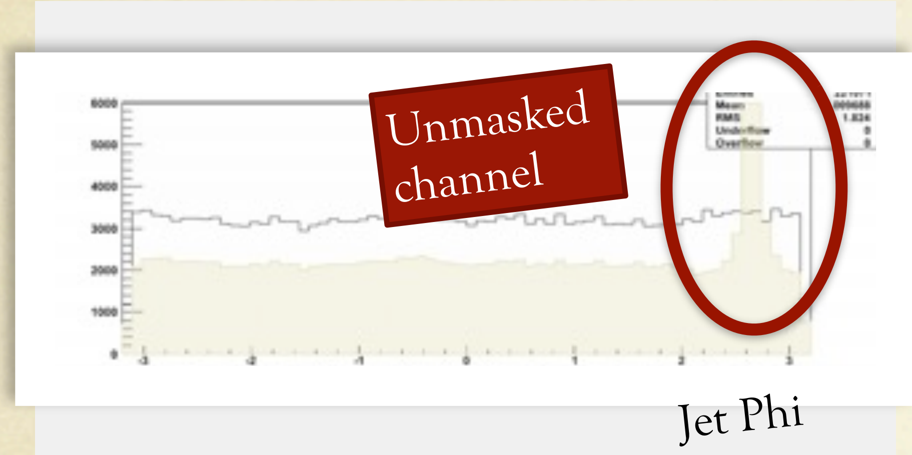
Dealing with the consequences

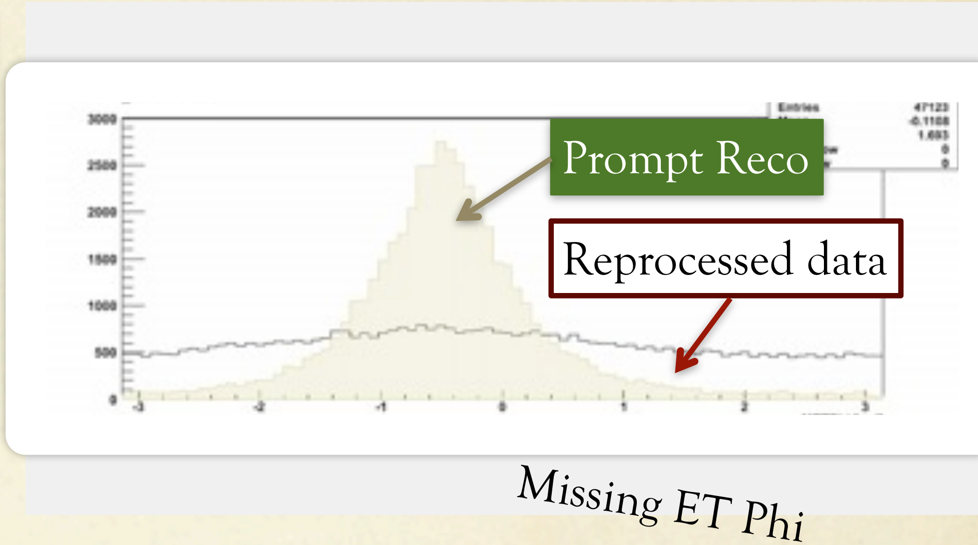
- Data taken after that accident have to be considered “good” for the Tracker.
  - Simulation will reflect the real conditions of the detector (eventually run by run)
- Study the real impact on the Physics:
  - Apply quality and analysis cuts: i.e. problem dominated by low Pt tracks.
  - Cuts in track reconstruction and quality have been set based on a perfect detector → reduction in efficiency and increase in fake rate
  - *Increase robustness:* keep into account the damaged region, refrain to use cuts on the total number of hits, but rather on a ratio.





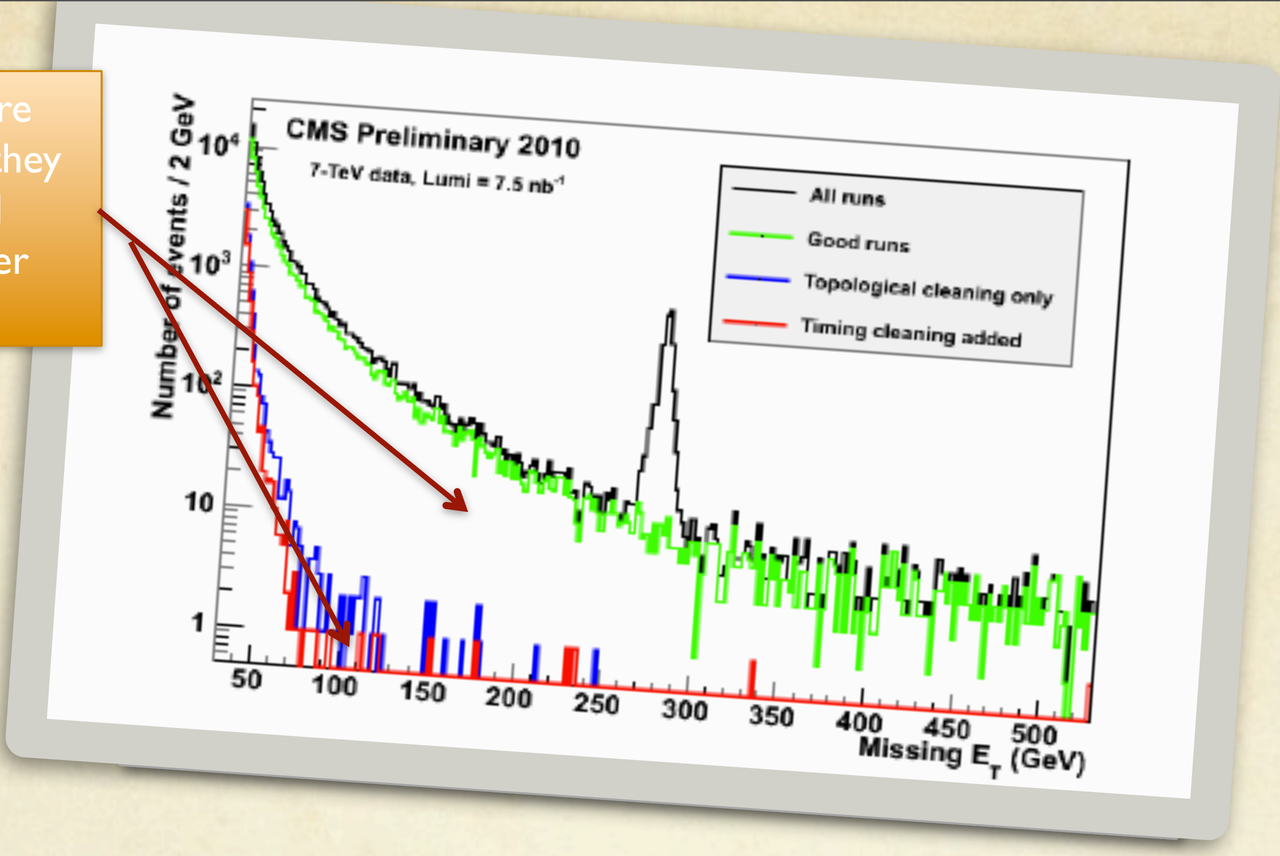
HCAL channel (un)masking





- **Another example of data recovery with a reprocessing**
  - By human mistake some of the HCAL channels usually masked got unmasked at some point…
- Effect visible immediately on trigger rates and other basic variables show here (plots from DQM monitoring)
- Histogram shows the data behaving nicely after the reprocessing with the proper masking enabled again.

Note: events are not removed, they are just moved down to smaller values of MET



Impact on good runs on Physics - MET

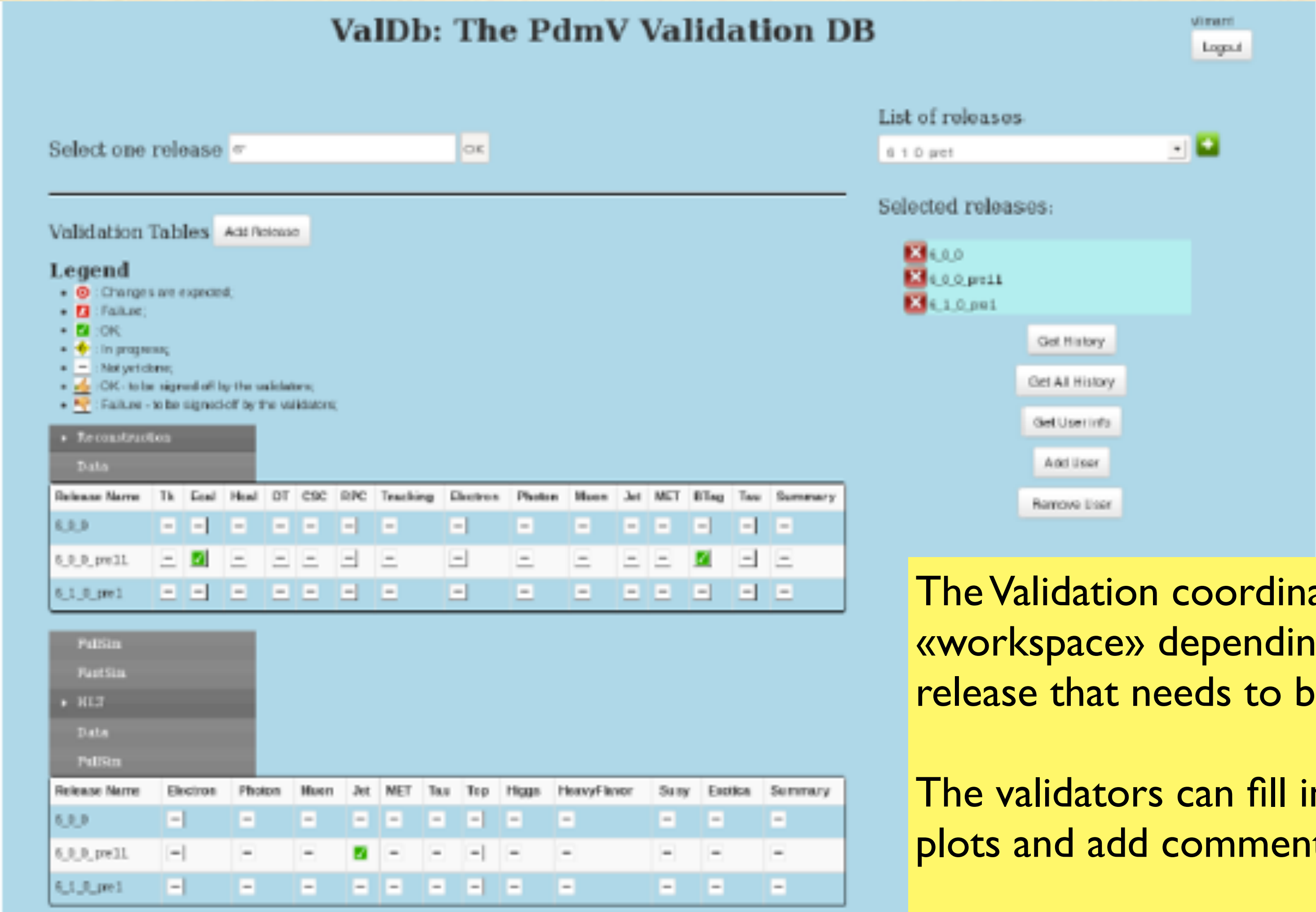
- **Effect of choosing the certified list of good runs on the Missing Energy distribution.** The peak corresponds to a hot tower. This represent an a-posteriori validation of the certification procedure. And shows the very good quality of the data.

- The blue and red curve show the improvement adding the topological cleaning for hot towers and then the effect of the use of timing information.

Beyond «Certification», what is «Validation»?

- The same frwamework structure used for «certification and Good Run list» is also used for «validation».
- Validation of:
  - release for T0 processing or for MC production
  - conditions A vs condition B before a reprocessing or MC production
  - release-n vs release n-1 during development phase
  - FastSim vs FullSim
  - data vs MC
- Different meanings:
  - **Basic sanity check: no changes applied, no changes expected**
    - For instance checks at the hit level response in the detector
  - **Known changes behave as expected**. This is particularly true for improvements or fixes
    - For instance this is true of efficiency or fake rates for more complex objects (tracks, muons, electrons, jets)
  - **Good for Physics: the changes or improvements are visible or have an effect on complete physics analysis**
    - Effects on overall selection efficiency for complete analyses
    - Improving agreement of data and MC

Validation DB for regular Release validation



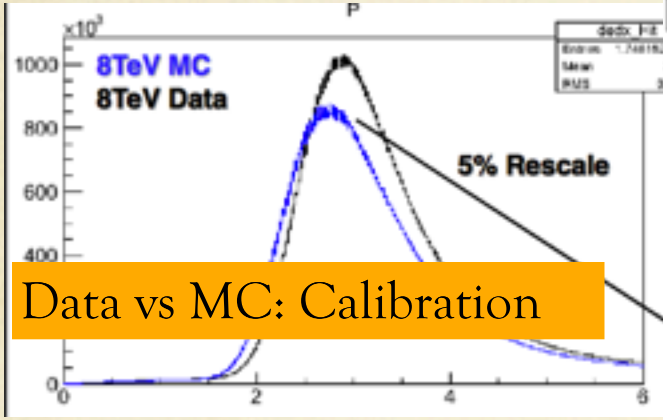
The Validation coordinators prepare the «workspace» depending on the release that needs to be validated.

The validators can fill in their results, link plots and add comments.

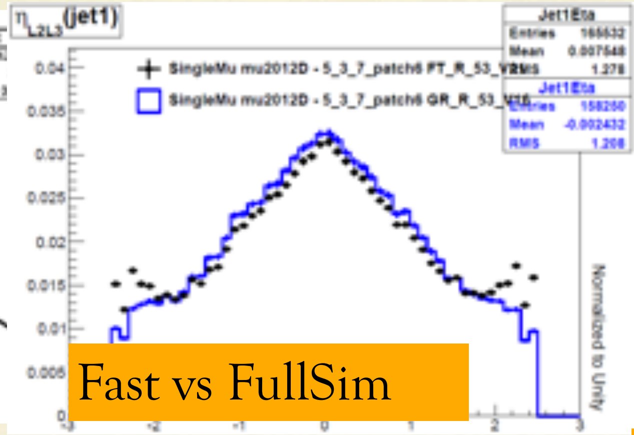
Big improvement given the large amount of validation campaigns: RunI, RunII, Upgrade Releases

Data vs MonteCarlo (and more ...)

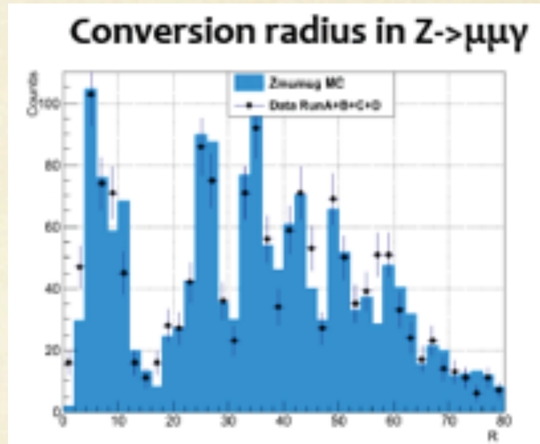
- Comparisons with different scenarios very efficient way of validating: Data vs MC, Fast vs Full, PU vs noPU
  - Effort to automatize paradigms and tools (usually done by hand or during analysis)
  - particular effort in defining appropriate data samples and workflows.
- this will set the bar much higher for the next run! foresee a great help from this during commissioning times!



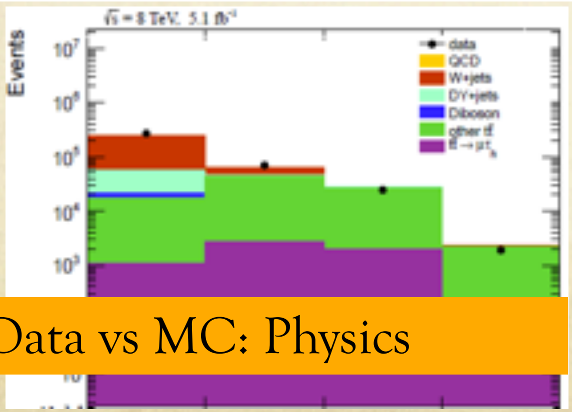
Data vs MC: Calibration



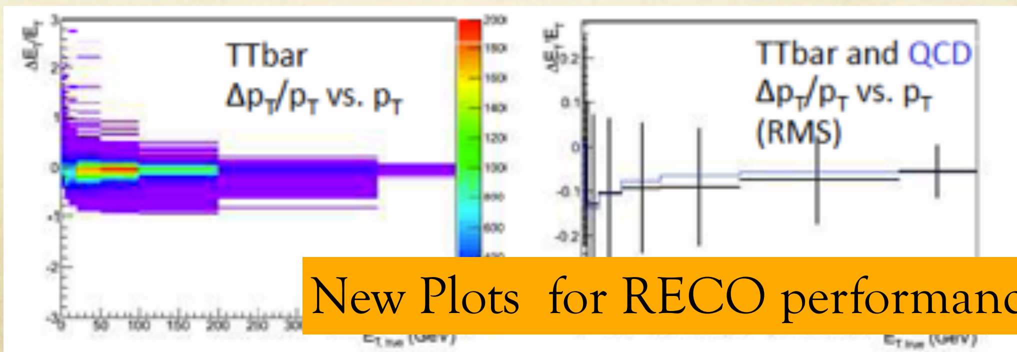
Fast vs FullSim



Data vs MC: Physics



Data vs MC: Physics

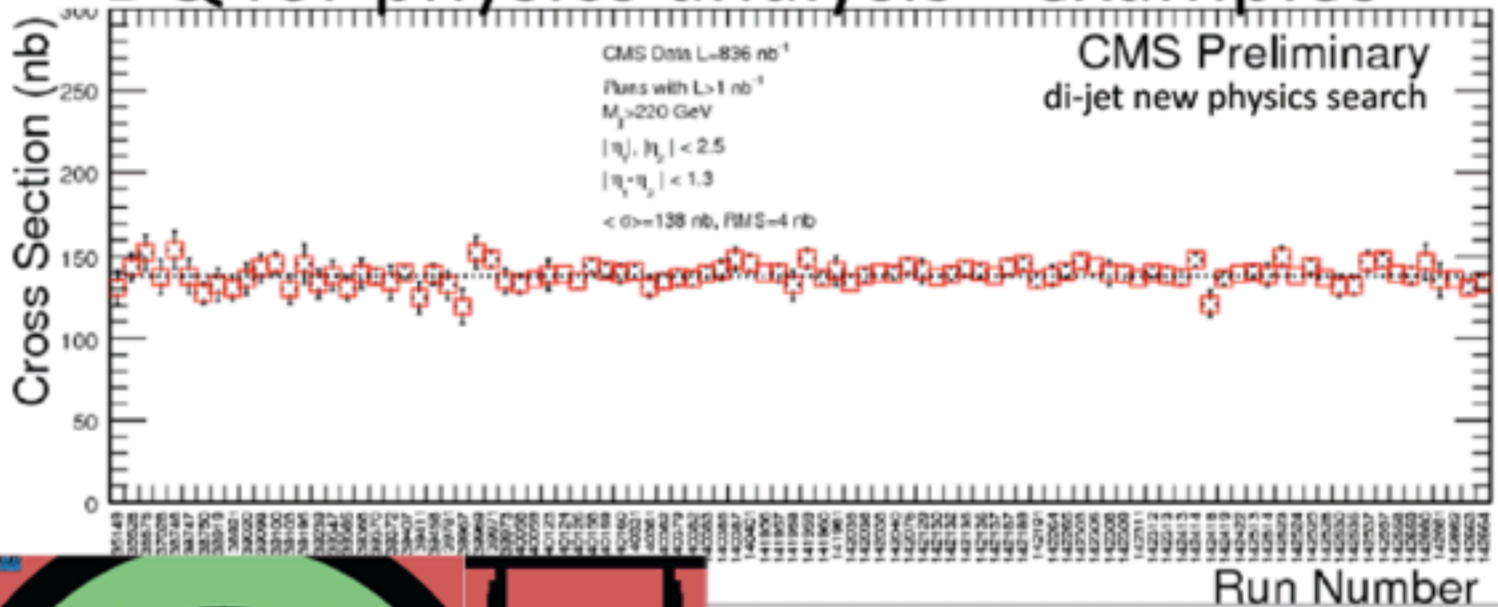


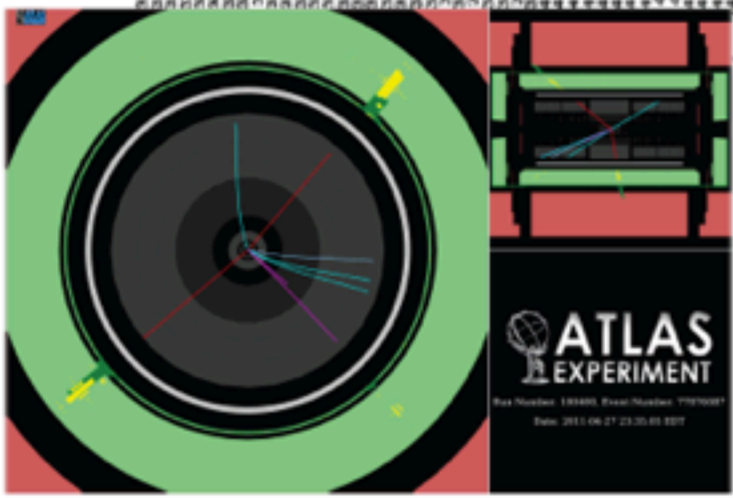
New Plots for RECO performance

Data quality and Physics Analysis

- *It has to be clear to everyone that all this work and checks performed by the «Data quality and Certification» groups does not mean that one should use the events blindly in the analysis.*
- Apart from the obvious that not everything can be spotted, it might well happen that your particular analysis is biased toward selecting events with specific problems!
- This is «typical» of new physics searches and the analyst need to keep a very professional approach to check everything without compromising a possible discovery.
- «Blinding» techniques are very popular but they require an enourmous care in making sure the definition of the control regions is correct to spot problems not related to the signal that it is sought for.
- Few examples of checks that should be done ALWAYS:
  - plot yelds/luminosity as a function of time/run number: might spot problematic runs, helps validate the luminosity estimate
  - Plots eta-phi distribution of the selected objects: compare with expectation. Sometimes selections can enhance detector issues.
  - Check visually (with the event display) the events selected in the tails of the distributions (for instance very large MET) to spot possible reconstructio/detector issues.

DQ for physics analysis - examples





Event display of highest mass di-electron event from ATLAS Z'->ee search

38

Summary and conclusions

- During data taking, between those making sure that the detector work and those that make beautiful analysis plots there is a large fraction of physicist that:
  - Worry about taking data: Trigger Menus and Datasets definition
  - Worry about calibrating the data: online (for the Trigger immediate decision making), quasi-online (before the data are Promptly Reconstructed), and then offline (to make sure the reprocessed data contain the best conditions)
  - Worry about validating and certifying every step of the way: online, offline, software releases, to deliver the good run list to be used for analysis and the best calibration and software for the data processing and MC production
- Last but not least there is a lot of technical tool development to make sure that everything that can be automatized will be and the human talens are best used for all the rest.
- and not forget the planning of all this! More on the next chapter....