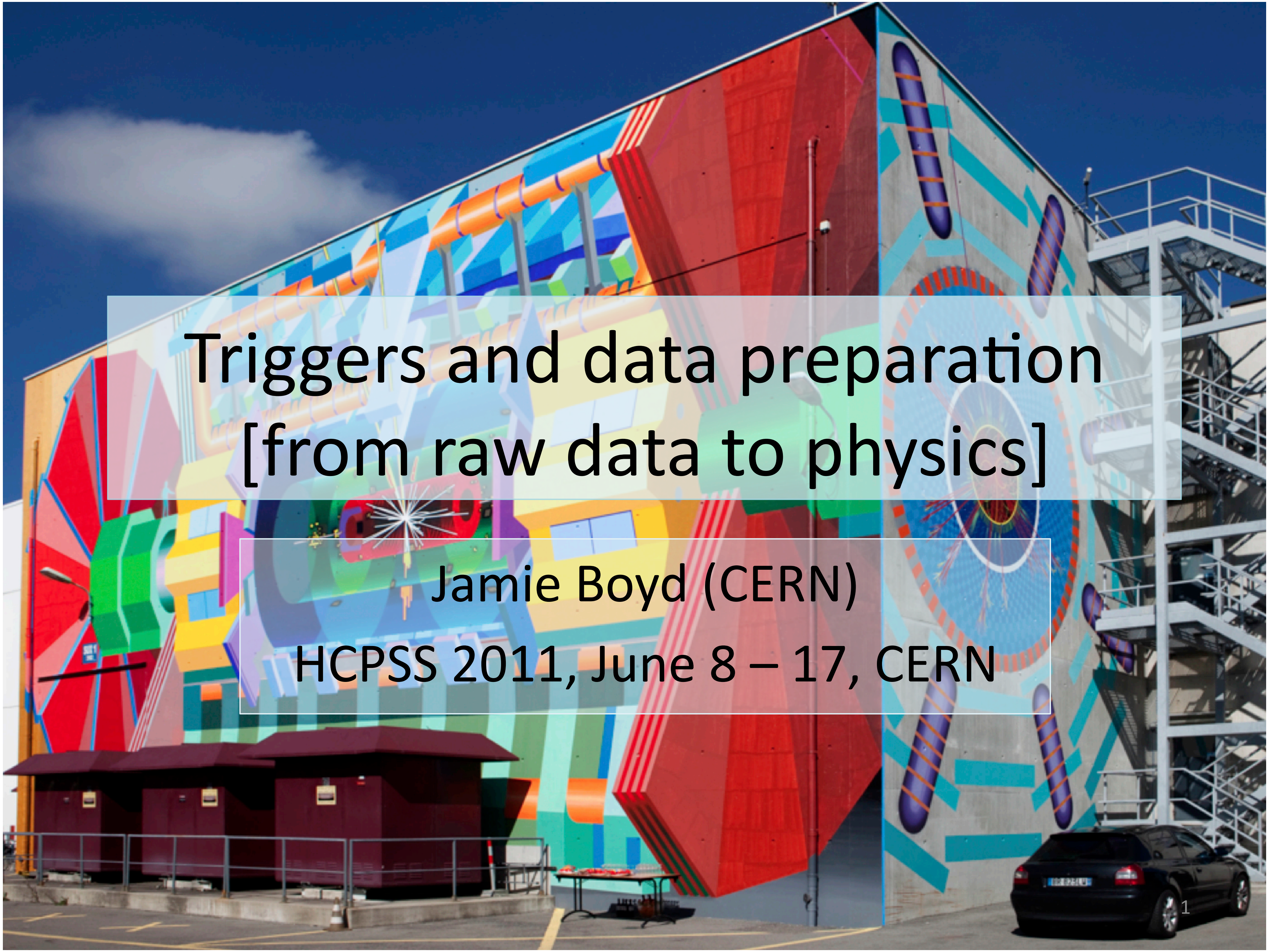
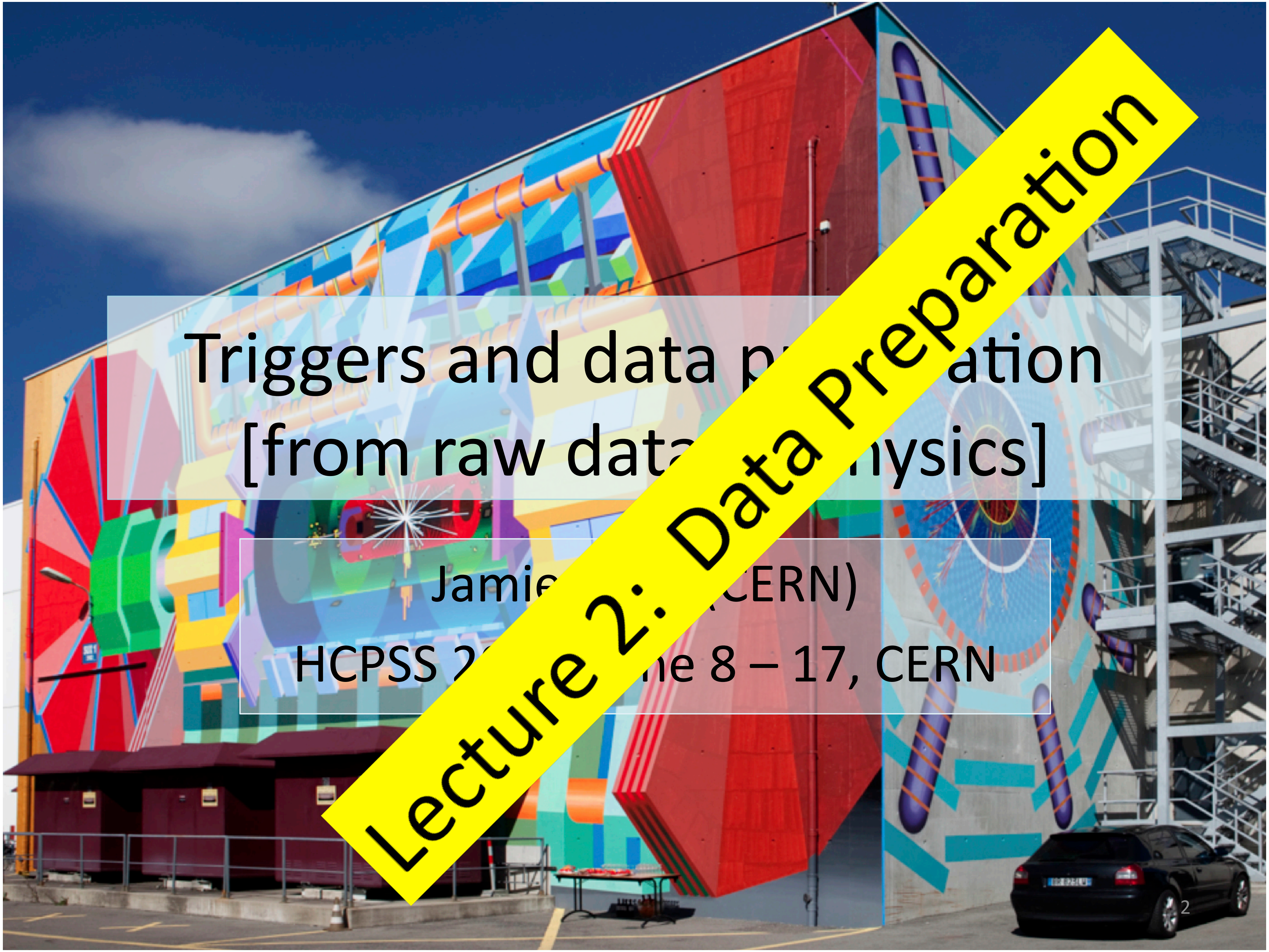


Triggers and data preparation [from raw data to physics]

Jamie Boyd (CERN)

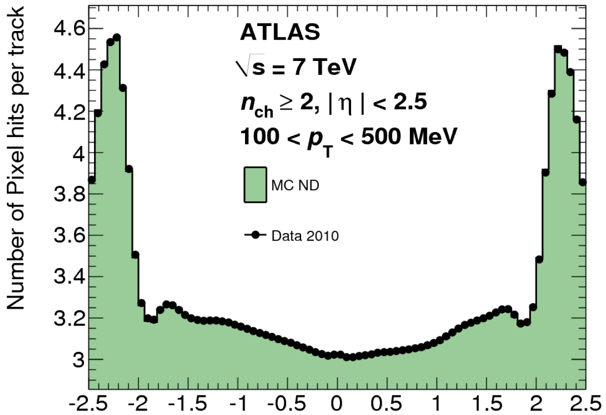
HCPSS 2011, June 8 – 17, CERN

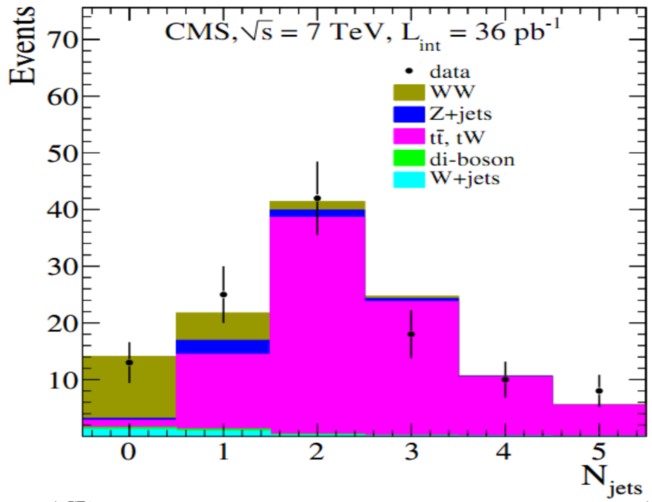

Triggers and data p ation
[from raw data hysics]
Jamie (CERN)
HCPSS 2 ne 8 – 17, CERN
Lecture 2: Data Preparation

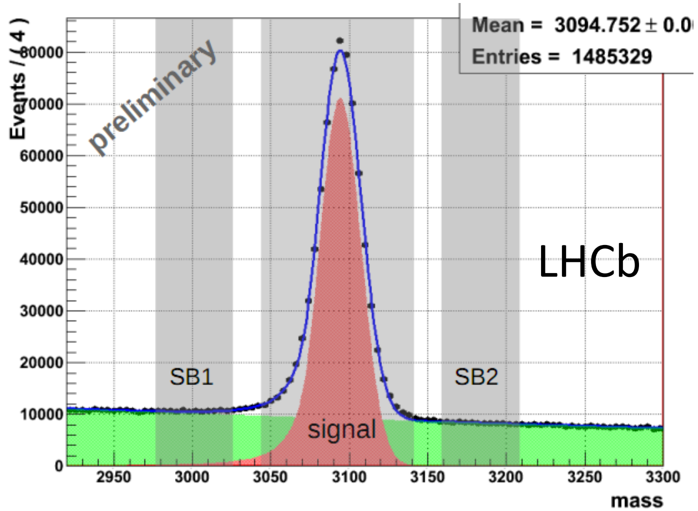
Introduction

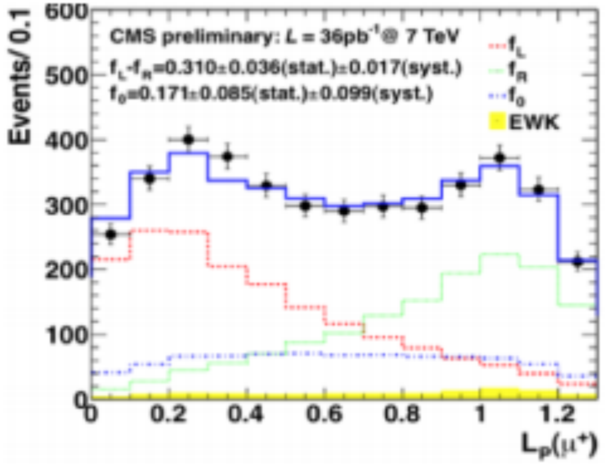
- As I am sure you have seen we have many LHC experiment plots showing that the MC is describing the data with amazing accuracy
  - This is not for free
- Many years of test beam studies looking at detector response
- Plus:
  - Data Quality selection
  - High quality data reconstruction and calibrations
  - Reconstruction ‘tricks’ to be robust against detector issues
  - Careful modeling of the detector/conditions by the MC
- I want to talk about these last points
  - with some examples

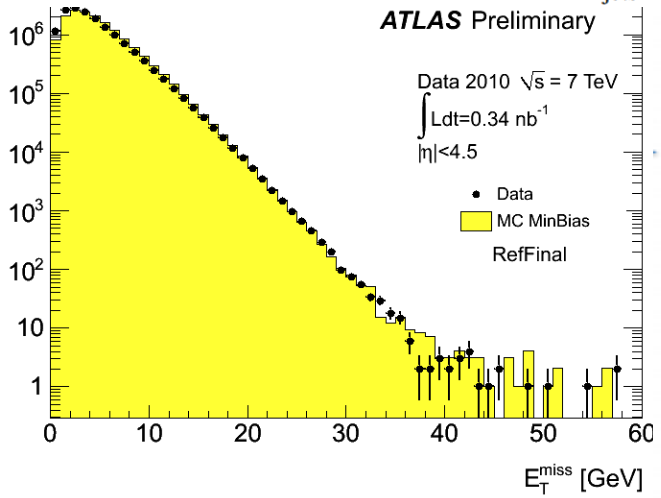
Above areas are an important part of the work of an experimental particle physicist

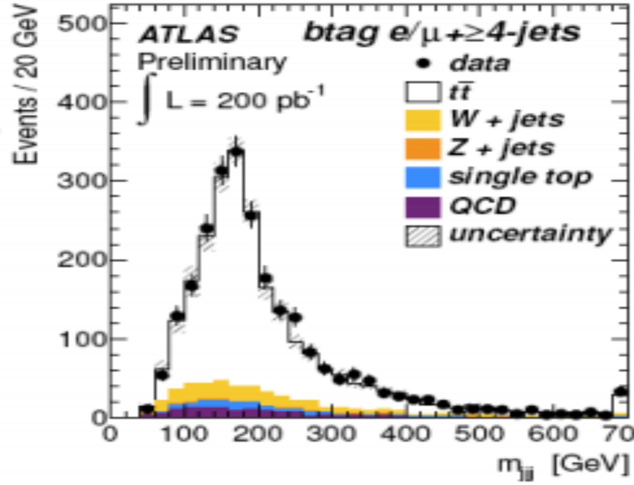

ATLAS
$\sqrt{s}$ = 7 TeV
$n_{ch} \geq 2$, $|\eta| < 2.5$
$100 < p_T < 500$ MeV
MC ND
Data 2010
Number of Pixel hits per track

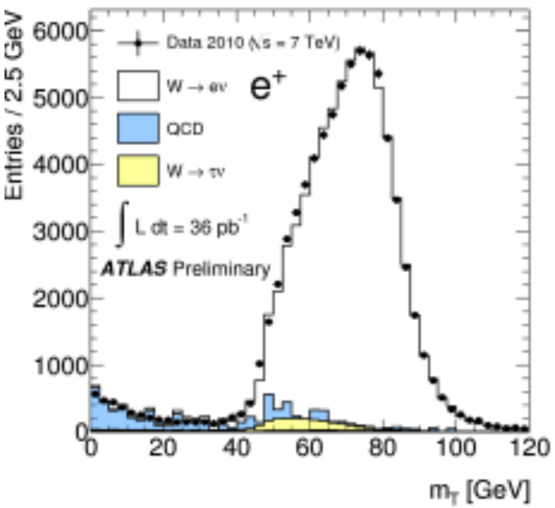

CMS, $\sqrt{s}$ = 7 TeV, L_{int} = 36 pb^{-1}
data
WW
Z+jets
$t\bar{t}$, tW
di-boson
W+jets
Events
N_{jets}

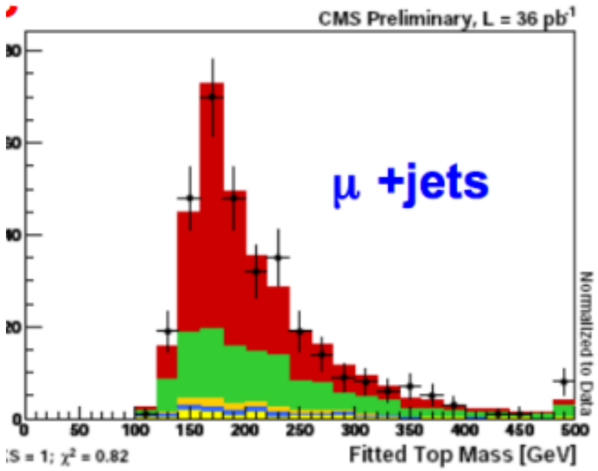

Mean = 3094.752 ± 0.0
Entries = 1485329
preliminary
LHCb
SB1
SB2
signal
mass

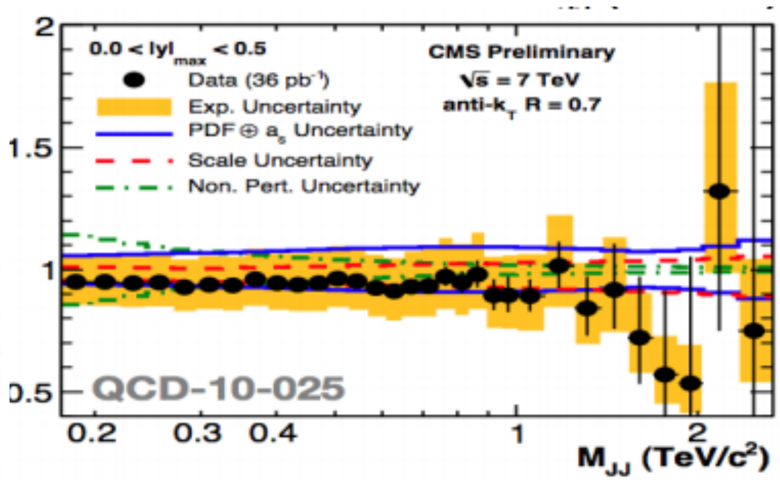

CMS preliminary: L = 36pb^{-1}@ 7 TeV
f_L-f_R=0.310±0.036(stat.)±0.017(syst.)
f_0=0.171±0.085(stat.)±0.099(syst.)
f_L
f_R
f_0
EWK
Events/ 0.1
$L_p(\mu^+)$


ATLAS Preliminary
Data 2010 $\sqrt{s}$ = 7 TeV
$\int$Ldt=0.34 nb^{-1}
|η|<4.5
Data
MC MinBias
RefFinal
E_T^{miss} [GeV]


ATLAS
Preliminary
btag e/μ+≥4-jets
$\int$ L = 200 pb^{-1}
data
$t\bar{t}$
W + jets
Z + jets
single top
QCD
uncertainty
Events / 20 GeV
m_{jjj} [GeV]

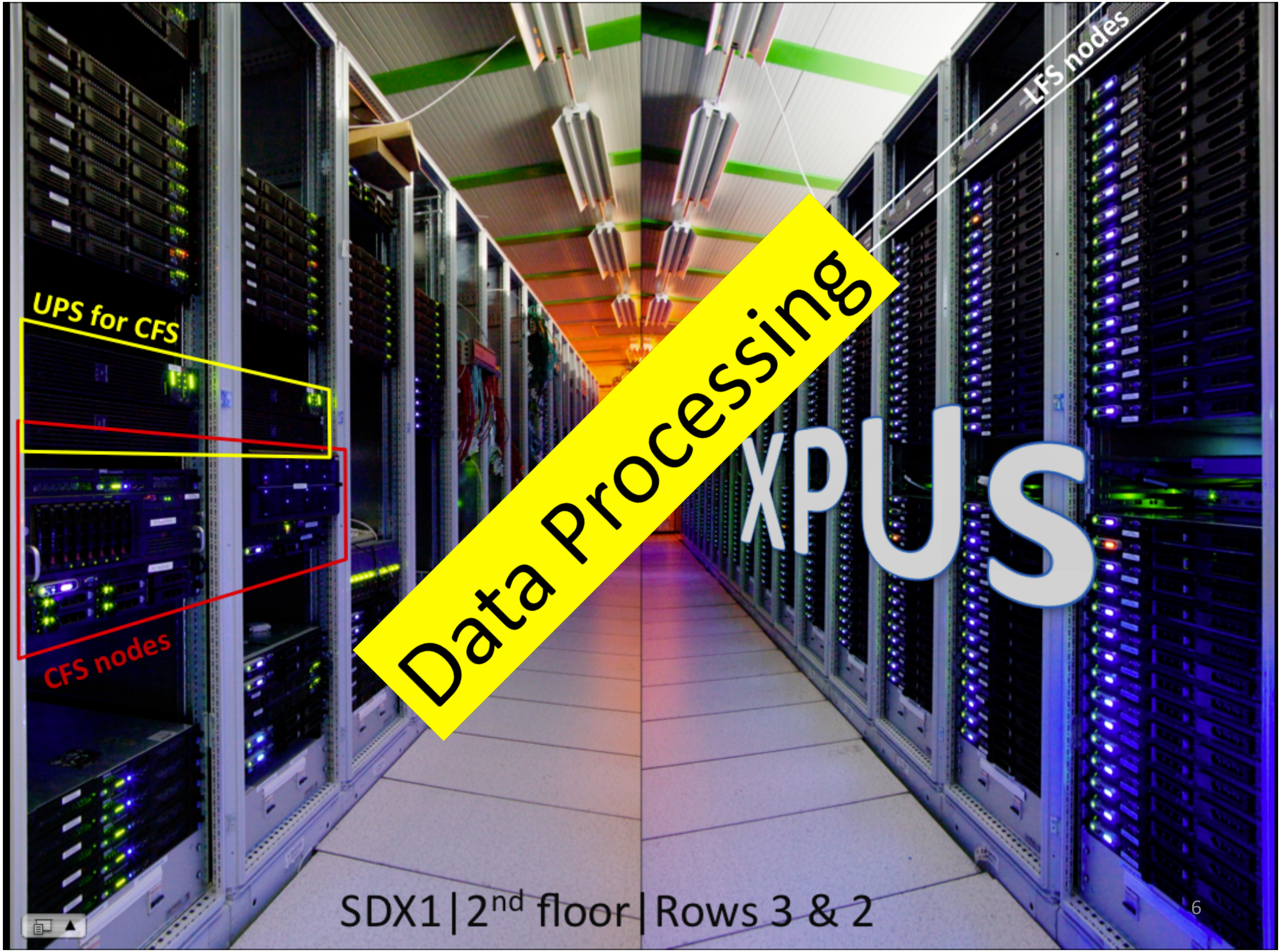

Data 2010 ($\sqrt{s}$ = 7 TeV)
W → eν
e^+
QCD
W → τν
$\int$ L dt = 36 pb^{-1}
ATLAS Preliminary
Entries / 2.5 GeV
m_T [GeV]


CMS Preliminary, L = 36 pb^{-1}
μ +jets
Normalized to Data
Fitted Top Mass [GeV]


0.0 < |y|$_{max}$ < 0.5
CMS Preliminary
$\sqrt{s}$ = 7 TeV
anti-k_T R = 0.7
Data (36 pb^{-1})
Exp. Uncertainty
PDF ⊕ α_s Uncertainty
Scale Uncertainty
Non. Pert. Uncertainty
QCD-10-025
M_{JJ} (TeV/c^2)

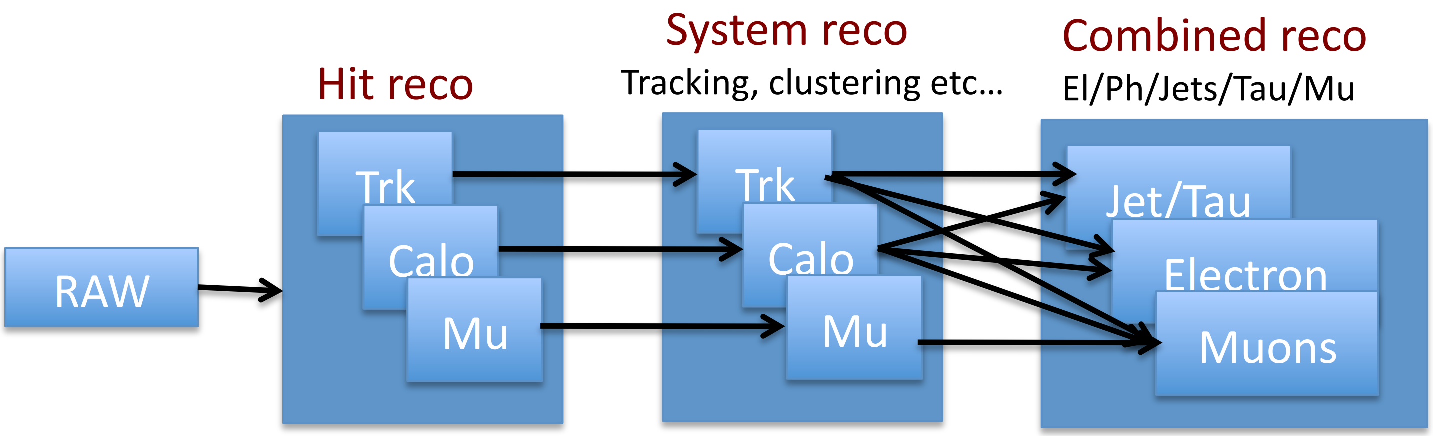
Contents

- Data processing
  - Prompt calibration loop
- Data Quality
  - Infrastructure
  - Examples
- Calibrations & Alignment
- Some specific examples
  - Dead channels
  - MC reweighting
  - Reconstruction robustness
  - Detector description


UPS for CFS
CFS nodes
LFS nodes
Data Processing
XPUS
SDX1 | 2nd floor | Rows 3 & 2

Data processing - reconstruction

- The RAW data of triggered events are written to disk/tape
- This data is processed to produce outputs for physics analysis
  - The processing 'reconstructs' the data
    - RAW ADC counts -> detector 'hits'
    - Track and cluster finding
    - Physics object reconstruction (combining information from different detectors)
    - Applying calibrations and alignment in many of these steps



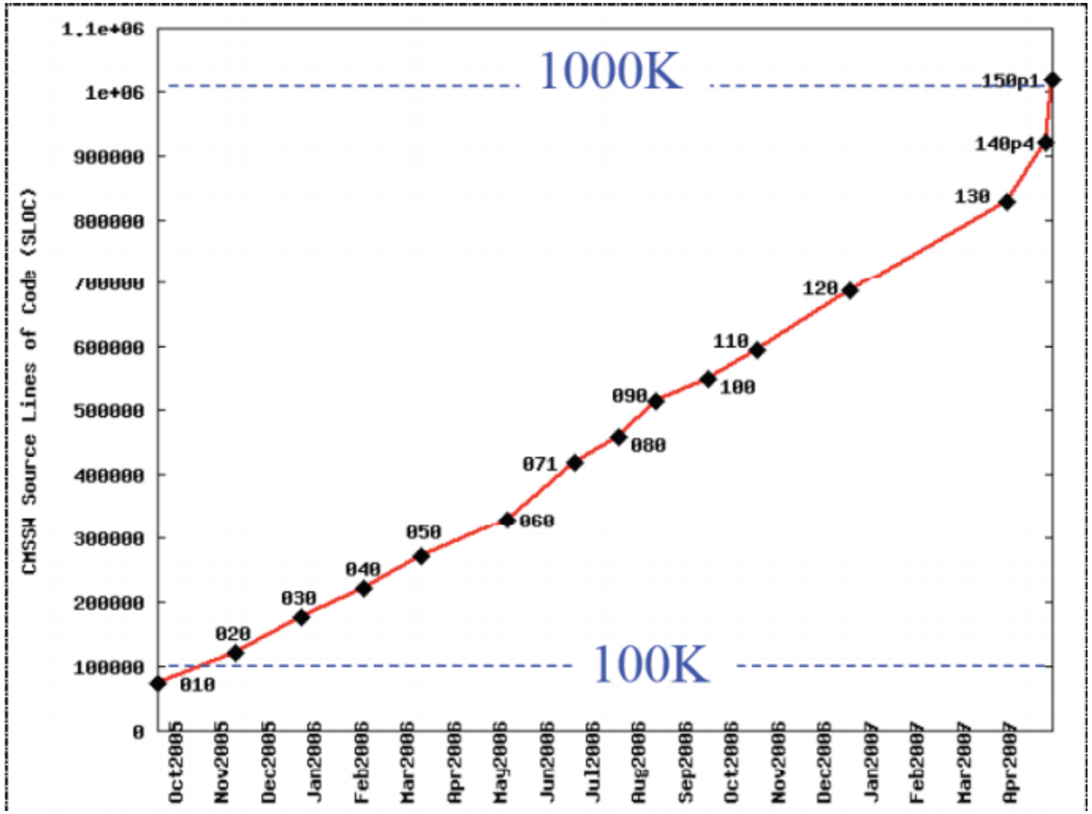
- Often the data is processed promptly at the Tier-0 and then reprocessed at a later time (with improved software and/or calibrations)

Reconstruction Software

Reconstruction software can be very complex.
Many algorithms, complex configurations.
e.g. >1M lines of code in CMS reconstruction software in 2007.

Speed of reconstruction software and size of outputs is one of the main limiting factor for the output trigger rate!

In ATLAS full reconstruction takes ~13 s/evt (2011 conditions).
Increases with pileup.

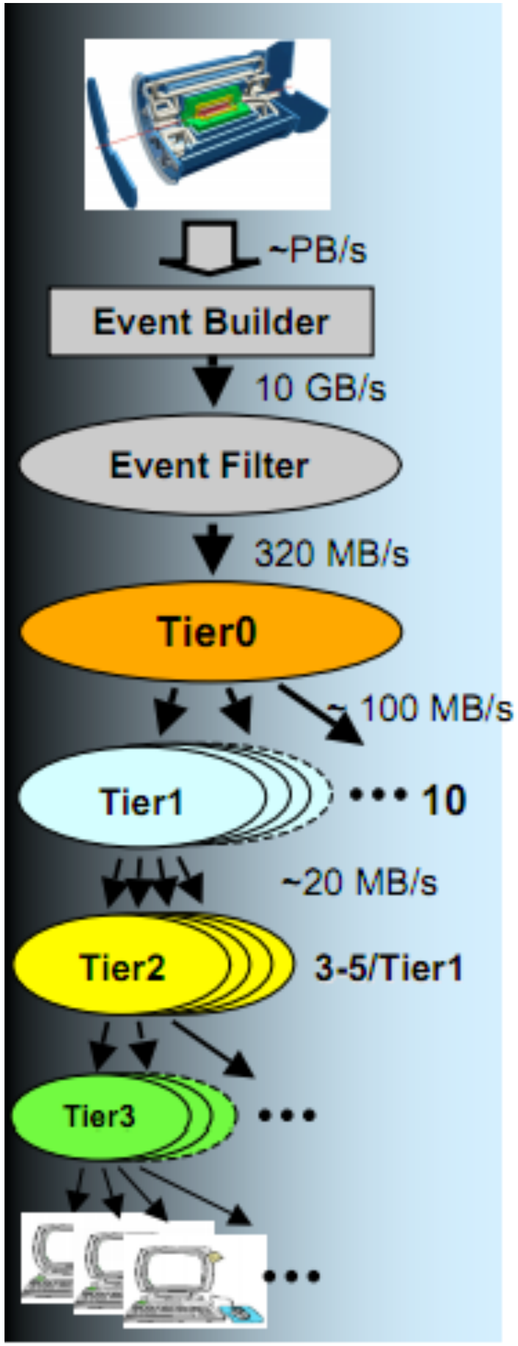


Reconstruction Stability

- Must keep the **reconstruction software stable** for long periods
- **Use same software to reconstruct MC** simulation as used for data
    - Need to compare data and MC
    - Need to combine data from different data-taking periods
- In ATLAS we have a formal 'frozen' Tier0 software policy
- If a real bug in the code is found it is not obvious if you should fix it in the Tier0 or wait until the next full data/MC reprocessing
    - Decision taken case-by-case in consultation with physics management

Because of this **validation of the software is paramount**.
Need to check the physics performance and the technical performance (CPU time, memory usage, memory leaks).

Reconstruction data types (ATLAS)



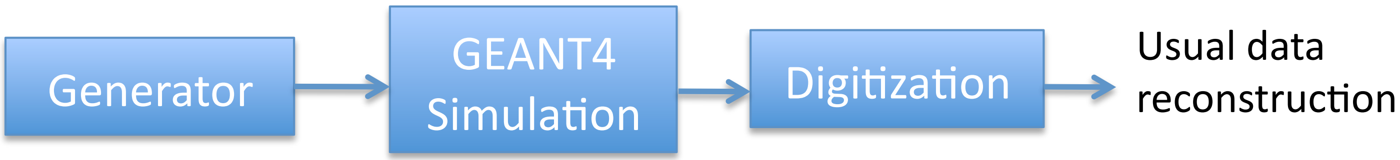
- Reconstruction dataflow (central and user) :

RAW → ESD → AOD → NTUP

- RAW:
  - "ByteStream" format ~1.5MB/evt
- ESD (Event Summary Data)
  - Full output of reconstruction (object POOL/ROOT format)
    - Tracks (& their hits), Calo cells, Calo clusters, combined reconstruction objects
    - Nominal size ~1MB/evt
    - In updated computing model no longer kept in longterm
- AOD (Analysis Object Data)
  - Summary of event reconstruction with "physics objects"
    - Electrons, muons, jets etc..
    - Nominal size ~100kB/evt (now ~180kB/evt)
- Final ntuples usually apply
  - Skimming (throw out non-interesting events)
  - Thinning/Slimming (throw away objects, and object detail info)
- Need to reduce data size in order to facilitate analysis

Monte Carlo Simulations

- Monte Carlo simulations **critical for physics analysis**
  - Designing selections
  - Evaluate acceptance
  - Study backgrounds (hope background estimate will be data driven in the end – but will always rely on simulation to guide study!)
  - Unfolding
  - Setting limits for new physics models
- Need the MC to **model the data as accurately as possible** - Many complications
  - Modeling of the physics processes
  - Modeling of the detector response
  - Modeling of the material, alignment, calibrations, dead channels, noise, etc…
- Monte Carlo simulation data flow



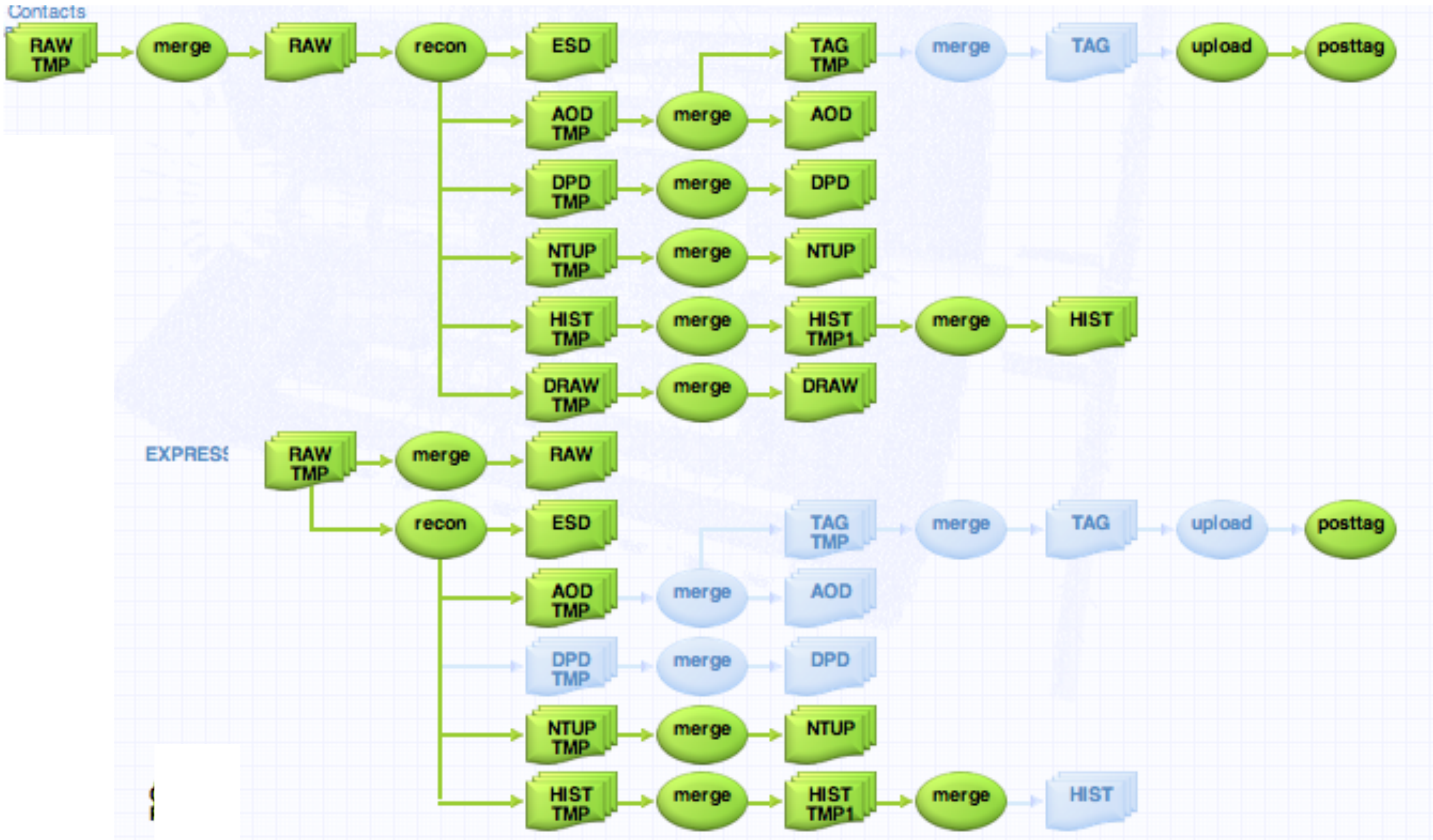
- For high multiplicity LHC events the simulation step can be very CPU intensive (~24hr's for 1 lead-lead central collision event!)
- Output of digitization is the same as the real data coming out of the detector (except also includes 'truth' information)

Some ATLAS Tier0 numbers…

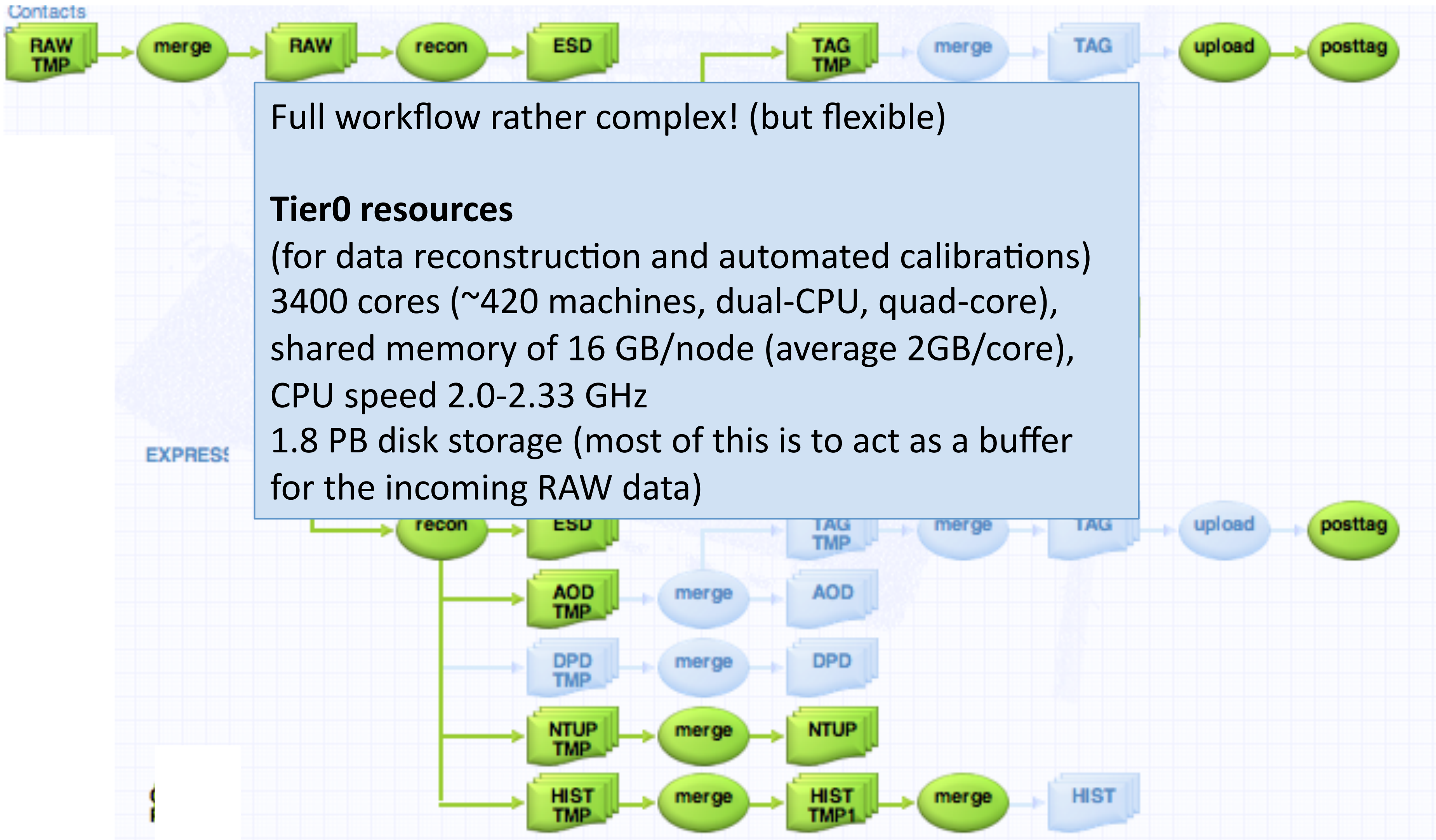
	2010 7TeV data #jobs	#evts	2011 7TeV data #jobs	#evts
Reco	1.4M	1.3B	0.7M	0.6B
	2010 7TeV data #files	**data volume**	**2011 7TeV data #files**	**data volume**
RAW (unmerged)	2.4M	1.6PB	1.5M	0.5PB
RAW (merged)	0.9M	1.6PB	0.5M	0.5PB
ESD	1.3M	1.5PB	0.7M	0.5PB
AOD	87k	130TB	42k	71TB

Processed nearly 2 billion events.
Combined file sizes >4 PetaBytes (>4000 000 GB).
In the workflow there is lots of merging of intermediate files.

ATLAS Tier0 workflow



ATLAS Tier0 workflow



Full workflow rather complex! (but flexible)

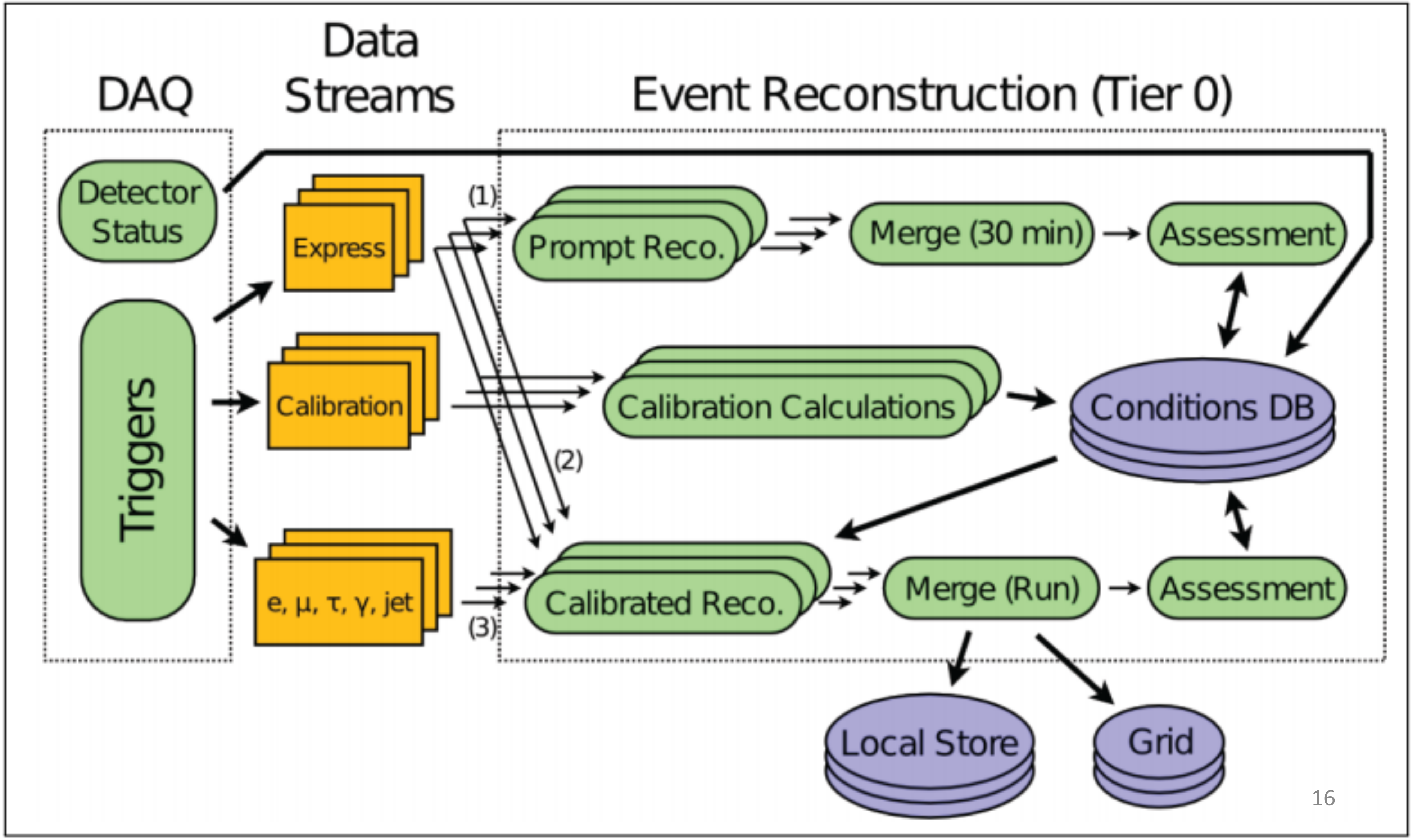
Tier0 resources
(for data reconstruction and automated calibrations)
3400 cores (~420 machines, dual-CPU, quad-core),
shared memory of 16 GB/node (average 2GB/core),
CPU speed 2.0-2.33 GHz
1.8 PB disk storage (most of this is to act as a buffer for the incoming RAW data)

ATLAS prompt data processing model

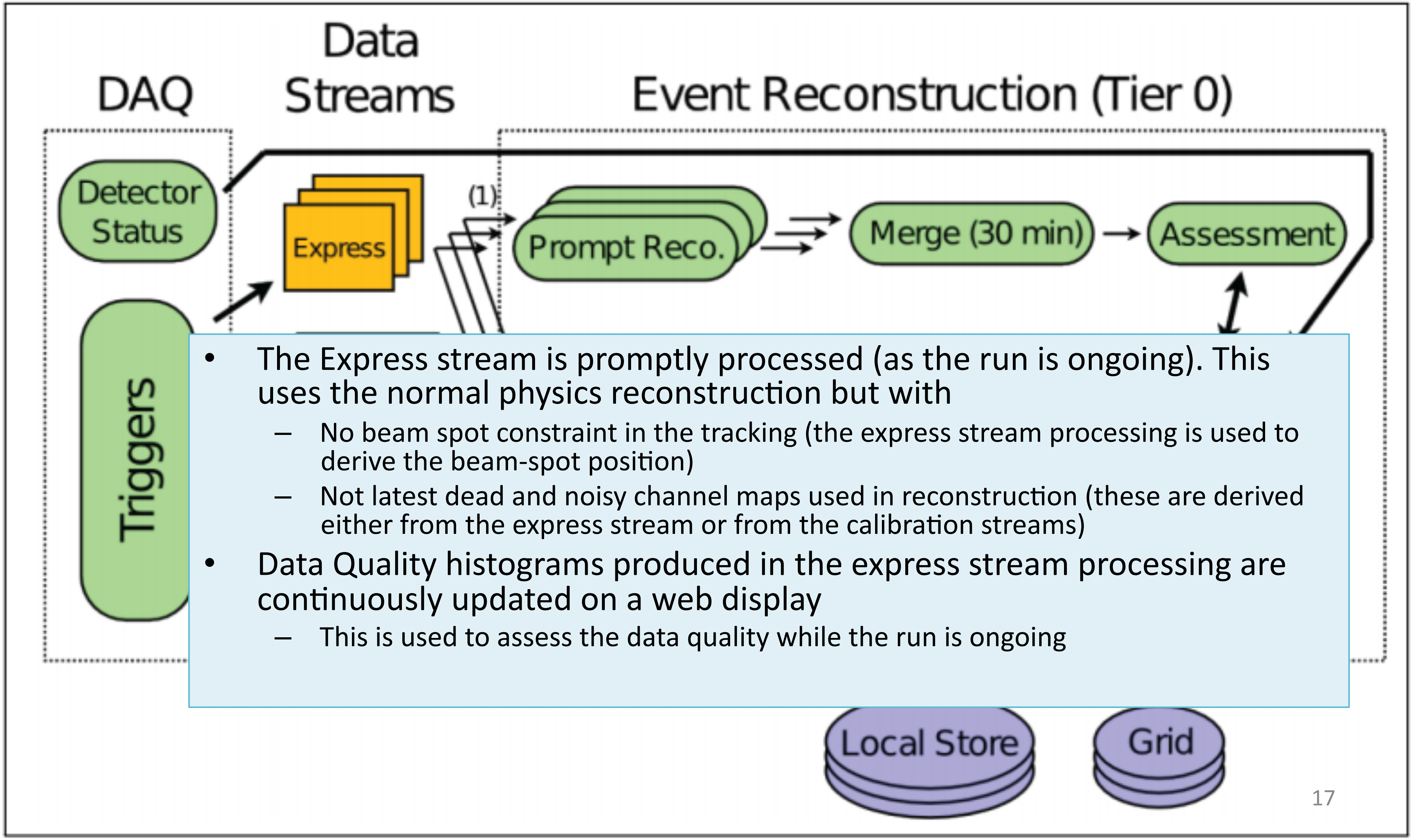
- The ATLAS trigger system produces 3 types of data streams
  - **Physics streams** (Egamma, Jet, Muons) used for physics analysis (~100Hz each)
  - **Express stream** used for data quality monitoring and calibrations (~10Hz)
  - **Calibration streams** – dedicated streams for calibrations – often partial events (rates and event sizes vary a lot)
- The data is promptly processed at the Tier-0 using the workflow described on the next page

Express stream contains a mix of prescaled triggers useful for monitoring
and calibrations eg. Triggers for J/psi, Z, W, Jets, Taus, photons
Events in the express stream are also in the physics streams.
Express stream should not be used for physics analysis!

ATLAS prompt processing model

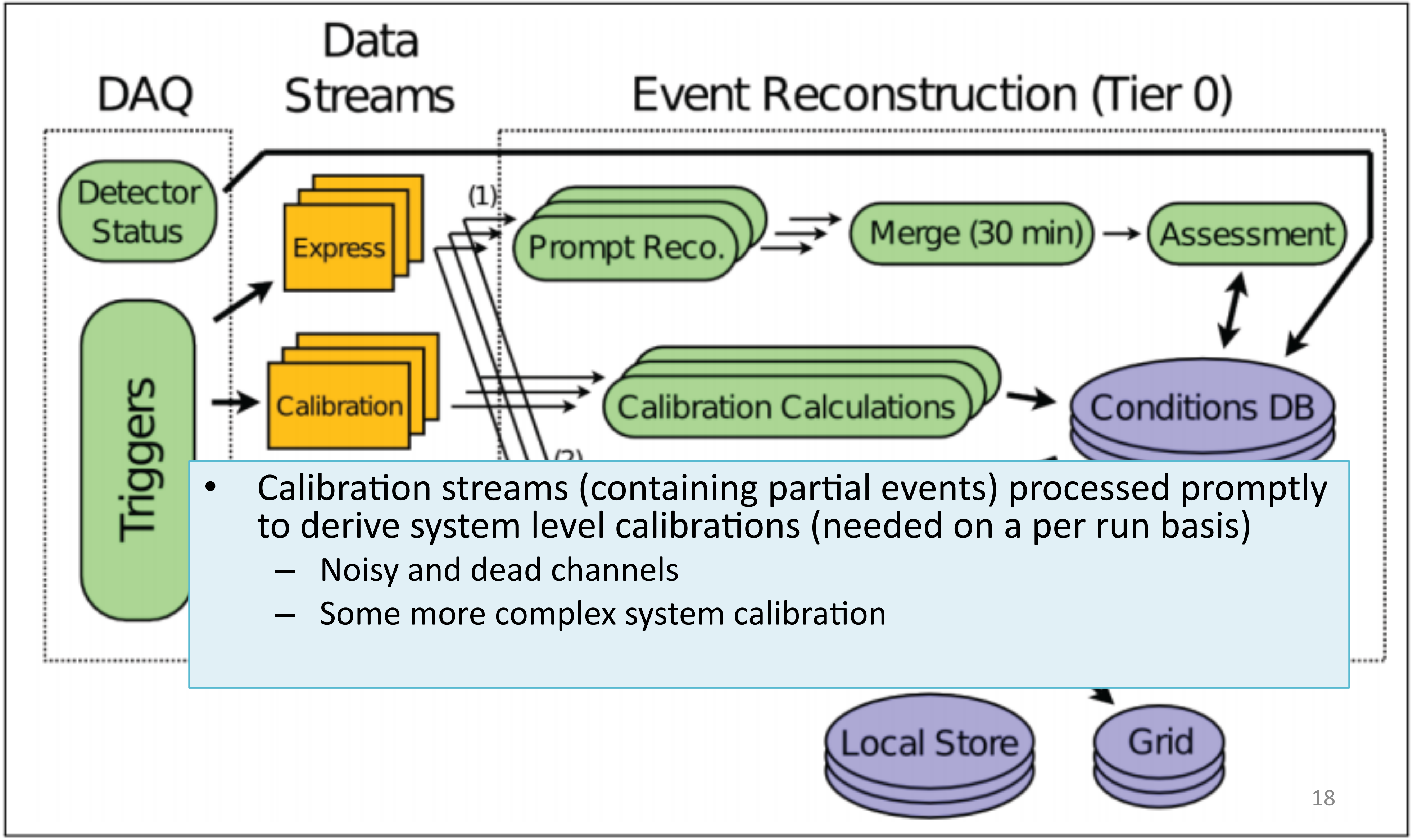


ATLAS prompt processing model



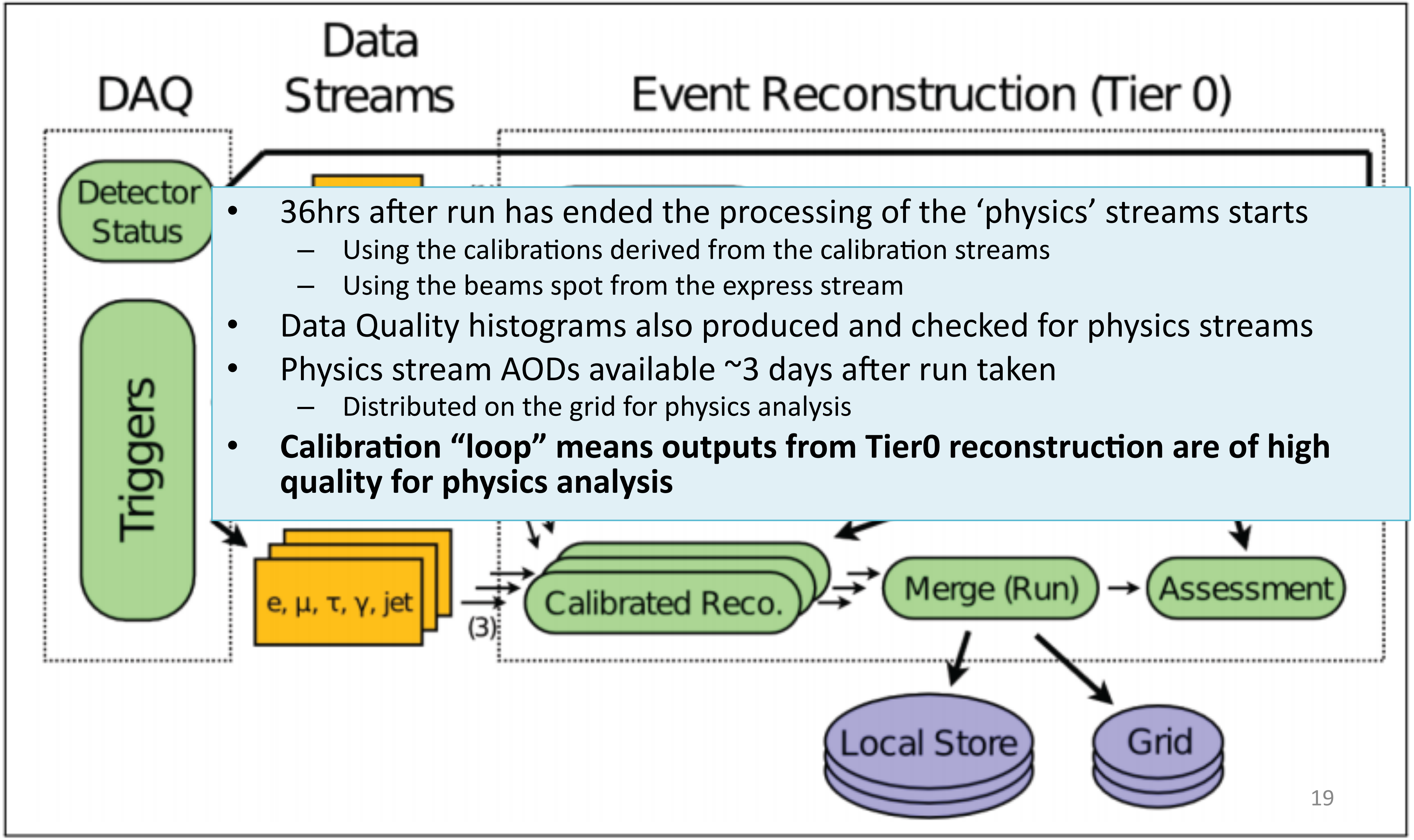
- The Express stream is promptly processed (as the run is ongoing). This uses the normal physics reconstruction but with
  - No beam spot constraint in the tracking (the express stream processing is used to derive the beam-spot position)
  - Not latest dead and noisy channel maps used in reconstruction (these are derived either from the express stream or from the calibration streams)
- Data Quality histograms produced in the express stream processing are continuously updated on a web display
  - This is used to assess the data quality while the run is ongoing

ATLAS prompt processing model



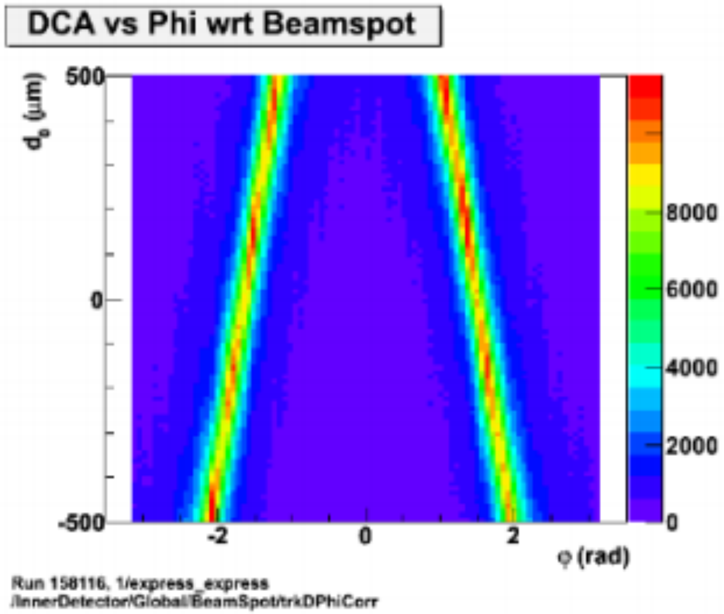
- Calibration streams (containing partial events) processed promptly to derive system level calibrations (needed on a per run basis)
  - Noisy and dead channels
  - Some more complex system calibration

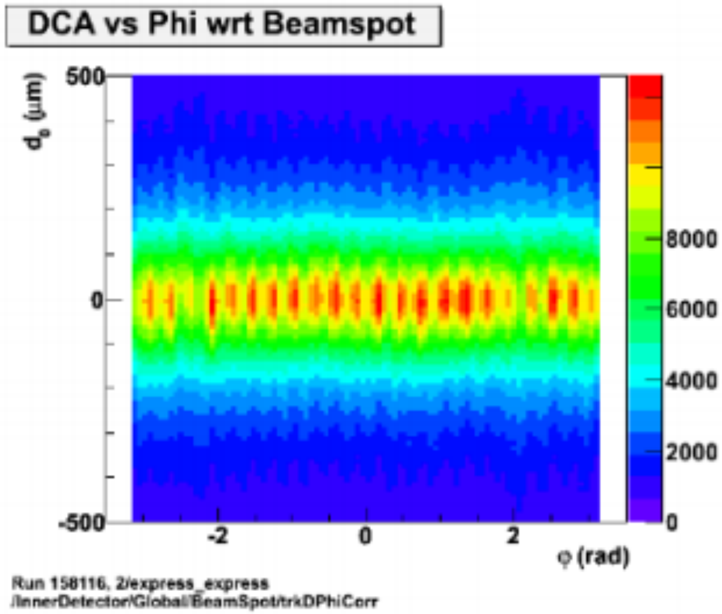
ATLAS prompt processing model



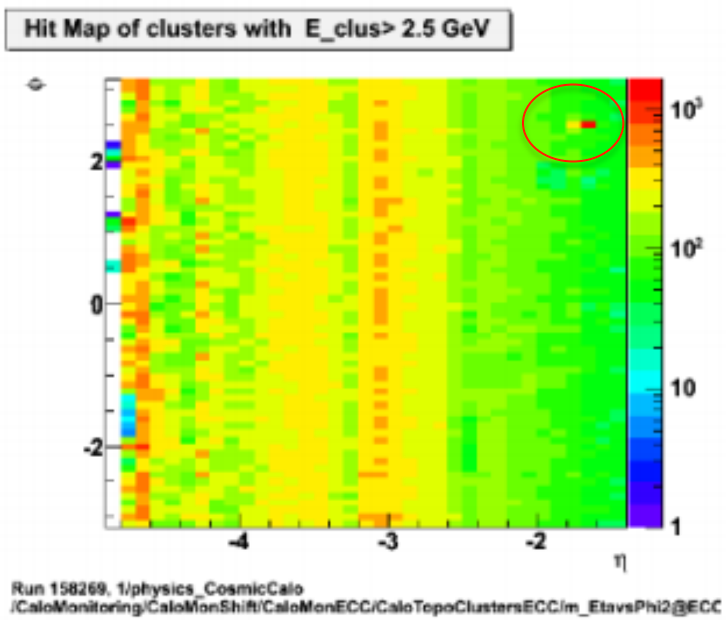
- 36hrs after run has ended the processing of the ‘physics’ streams starts
  - Using the calibrations derived from the calibration streams
  - Using the beams spot from the express stream
- Data Quality histograms also produced and checked for physics streams
- Physics stream AODs available ~3 days after run taken
  - Distributed on the grid for physics analysis
- **Calibration “loop” means outputs from Tier0 reconstruction are of high quality for physics analysis**

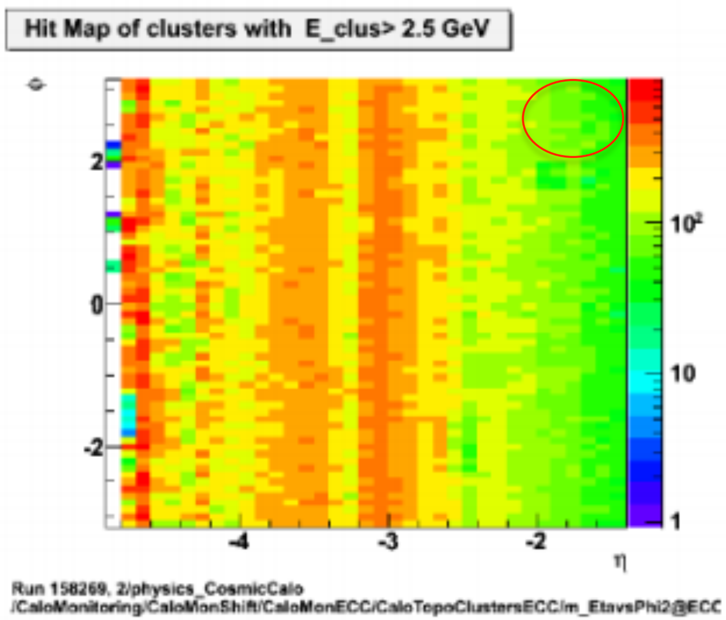
(ATLAS) 36 hr calibration loop



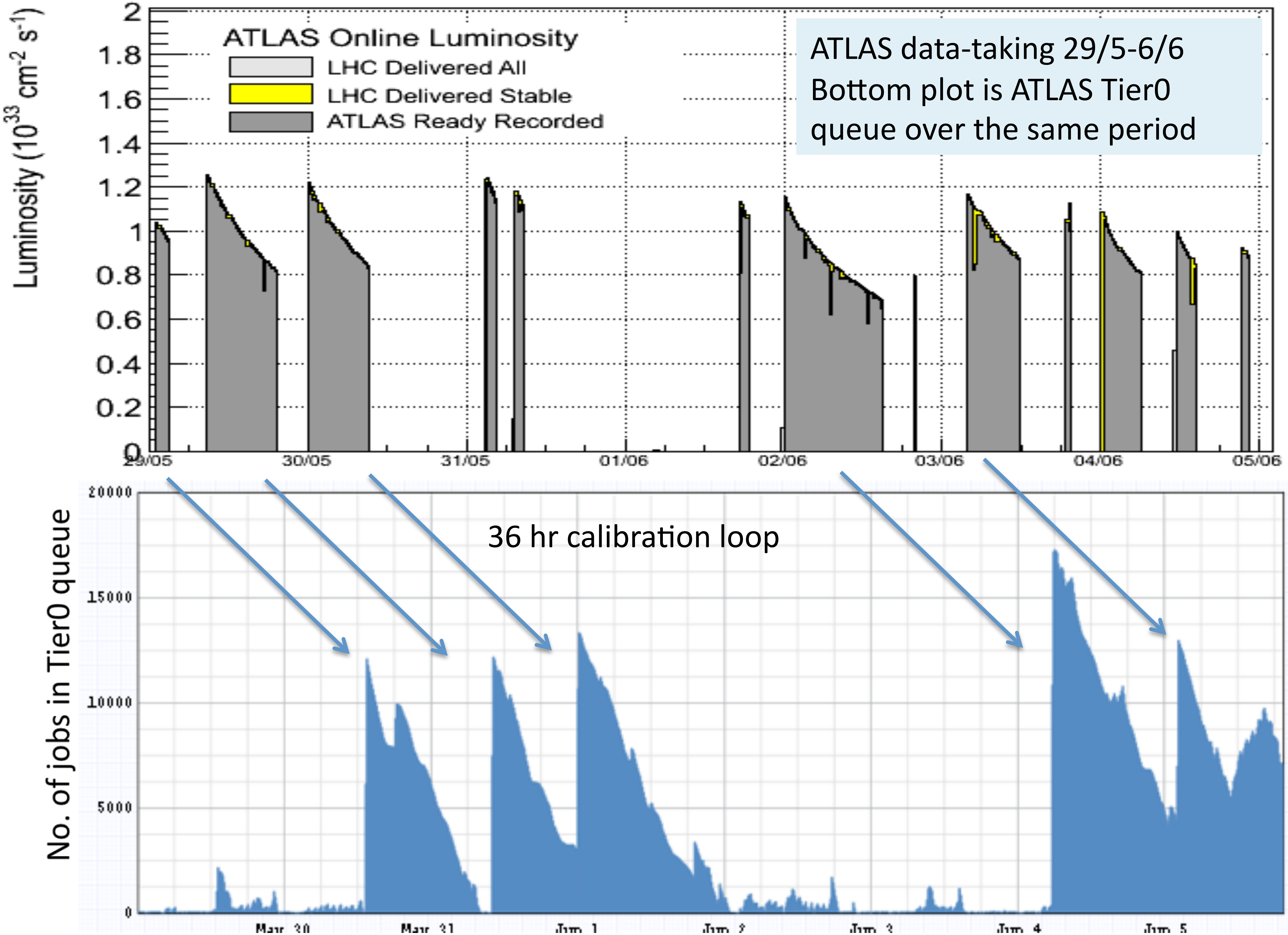


Beam spot before/after calibration loop





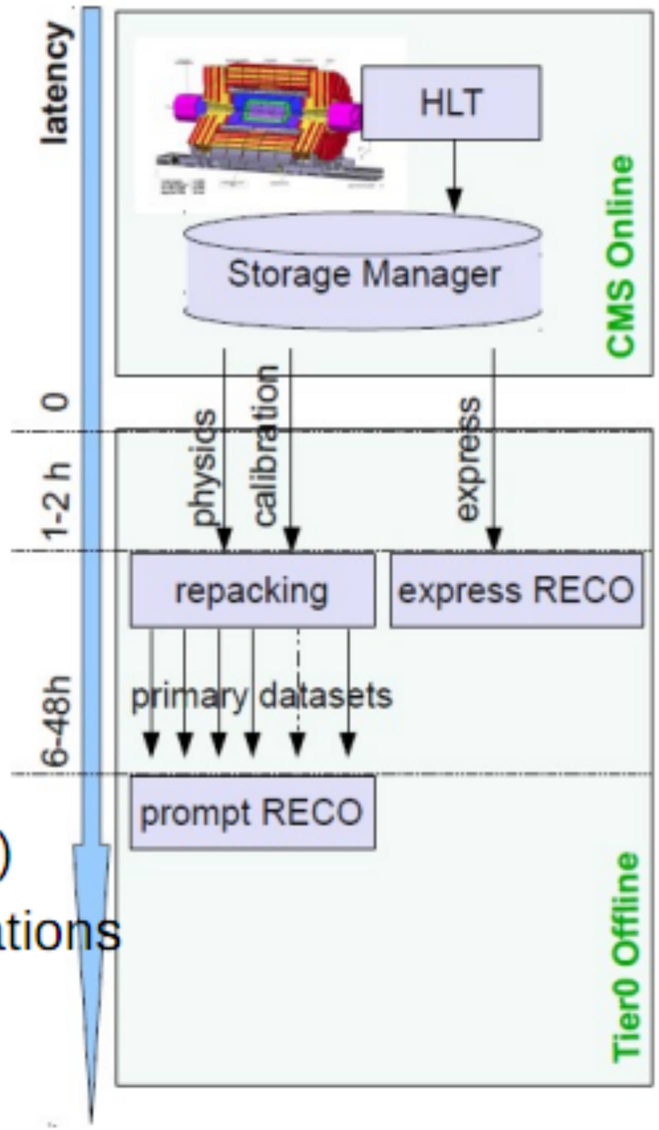
Hot channel in the calorimeter is masked in the physics stream processing.



ATLAS Online Luminosity
LHC Delivered All
LHC Delivered Stable
ATLAS Ready Recorded
Luminosity (10^33 cm^-2 s^-1)
29/05
30/05
31/05
01/06
02/06
03/06
04/06
05/06
ATLAS data-taking 29/5-6/6
Bottom plot is ATLAS Tier0 queue over the same period
36 hr calibration loop
No. of jobs in Tier0 queue
20000
15000
10000
5000
0

CMS model

- Data streams & Tier0 workflows → specialized for different tasks
- Depending on the latency
  - express → prompt feedback & calibrations
    - short latency: 1-2 hours
    - ~40Hz bandwidth shared by:
      - calibration (½)
      - detector monitoring (¼)
      - physics monitoring (¼)
  - Alignment & Calibration (AlCa) streams
  - bulk data → sample for physics analysis (prompt reconstruction)
    - split in Primary Datasets (using High Level Trigger (HLT) decision)
    - will be delayed of 48h → get latest calibrations
    - writing ~300Hz

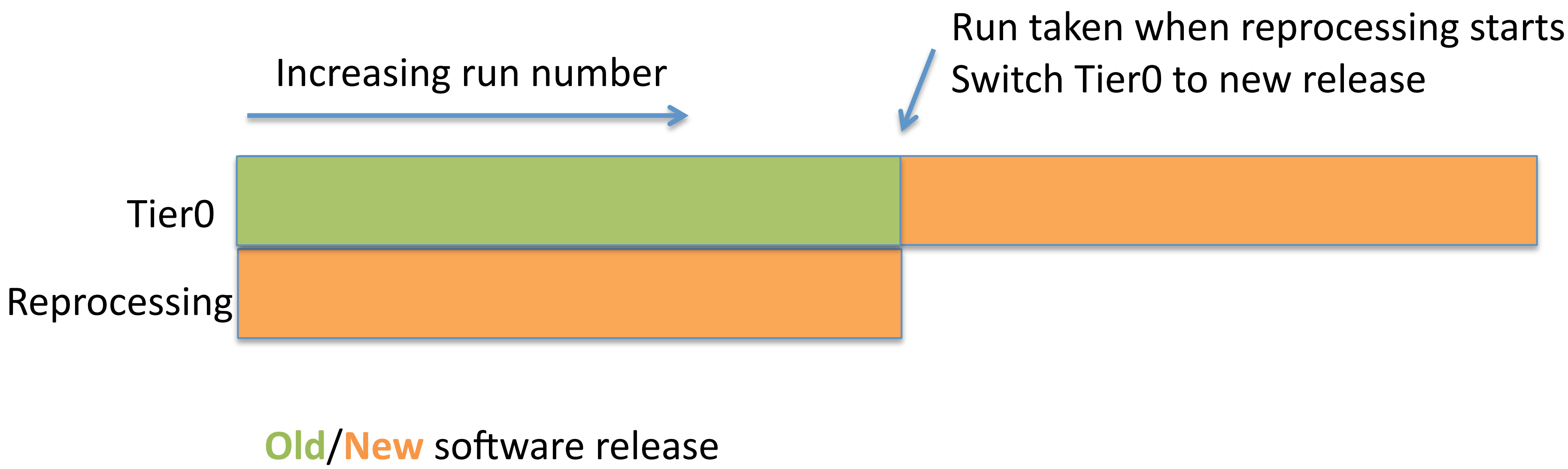


CMS Tier0 workflow very similar (this is from ICHEP 2010 so maybe a bit outdated!)

Reprocessing

Data reprocessed with **updated reconstruction software** and **improved calibrations and alignment** ~1-2 times a year. (usually targeting a major conference).
Reprocessing uses same dataflow as Tier0 reconstruction (RAW->ESD->AOD)
Validation of the new configuration (s/w and conditions) is a major undertaking.
Need to **reprocess the MC** with the updated configuration too (often just re-digitize/re-reconstruct).
Calibrations for reprocessing have much more time to be developed
- Simple calibrations like noise maps usually the same as from prompt reconstruction
- More sophisticated calibrations (like alignments) can be improved for reprocessing

Reprocessings run on the GRID at large computing centers around the world.
When reprocessing starts switch Tier0 to use new release so have a consistent dataset with the new software.

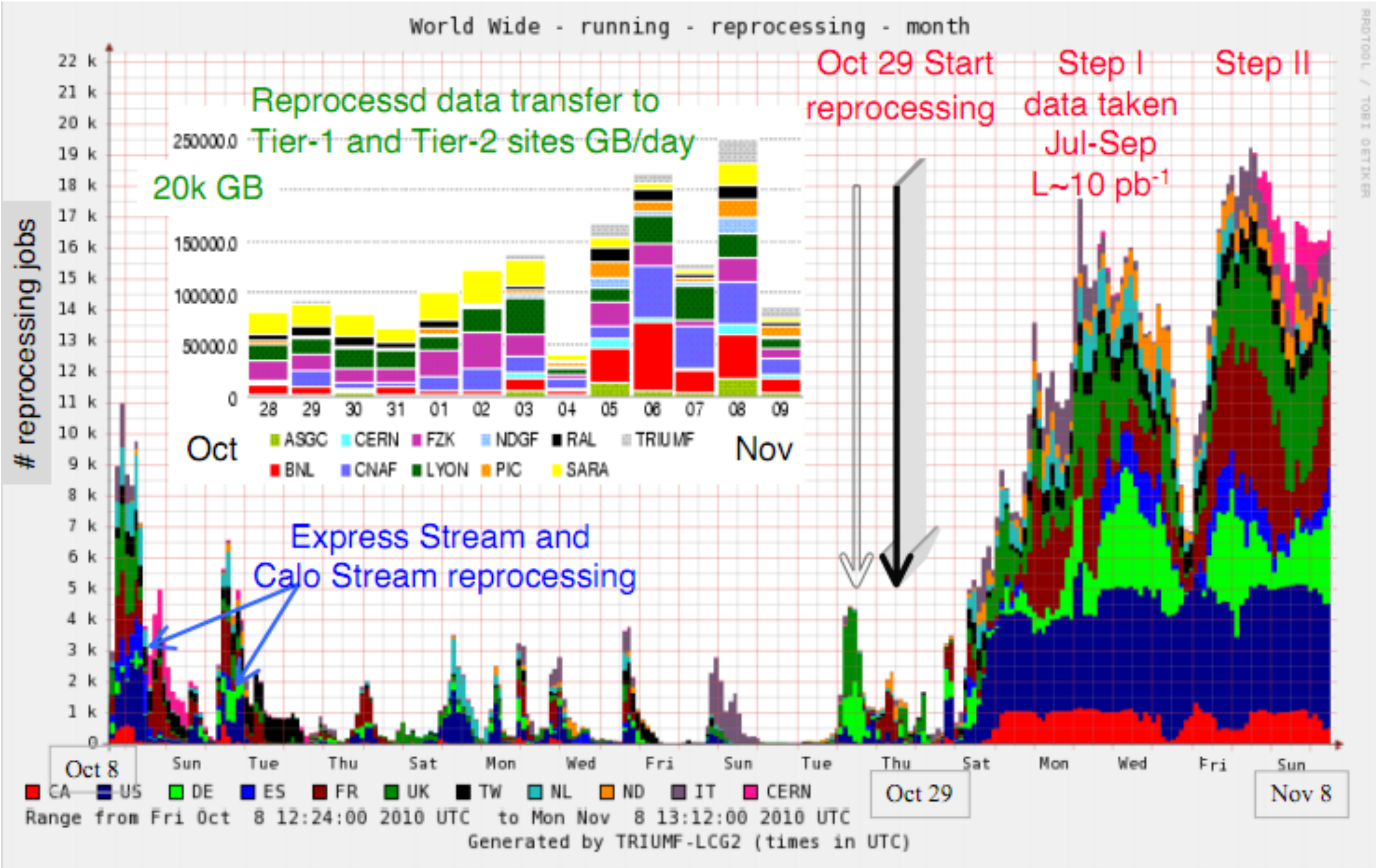


Reprocessings in ATLAS

Plan is to reprocess data 1-2 times a year.
Since s/w and calibrations still being commissioned during early data-taking more reprocessings carried out. So far reprocessed the pp data:
December 2009, February 2010, May 2010, October 2010
Currently planning a large reprocessing for Sept 2011.

ATLAS reprocessing of full 2010 pp data in Oct 2010.

Reprocessed data undergoes updated DQ assessment.
Sometimes possible to recover data with bad DQ by reprocessing.



Data Processing Summary

ATLAS/CMS have similar data processing model

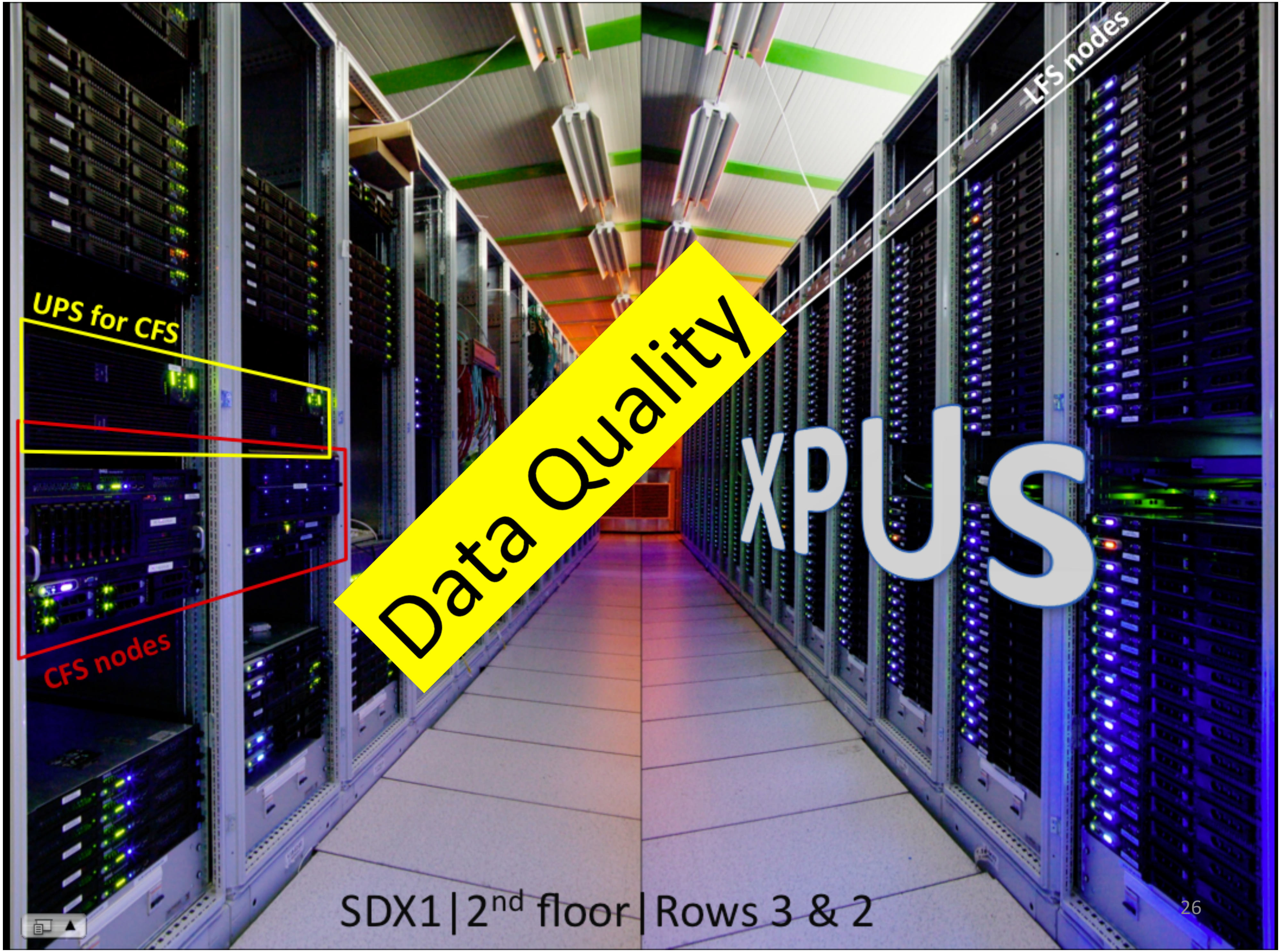
Prompt reconstruction using a calibration loop

- Processing of the physics data at Tier0 delayed by ~36/48hours to allow use of calibrations in the processing
- Means the output of prompt reconstruction is of high quality and can be used for physics analyses
  - Many physics papers published promptly processed data
  - in long term when the luminosity is stable and when we have more sophisticated calibrations may want to only publish papers based on reprocessed data
- Promptly processed data available for physics a few days after the data is taken

Data reprocessed with improved software and calibrations 1-2 times a year

In order to have consistent data and MC samples the MC needs to be reconstructed with the same release as the data

Stability of software very important to facilitate physics analysis


UPS for CFS
CFS nodes
Data Quality
XPUS
LFS nodes
SDX1|2nd floor|Rows 3 & 2

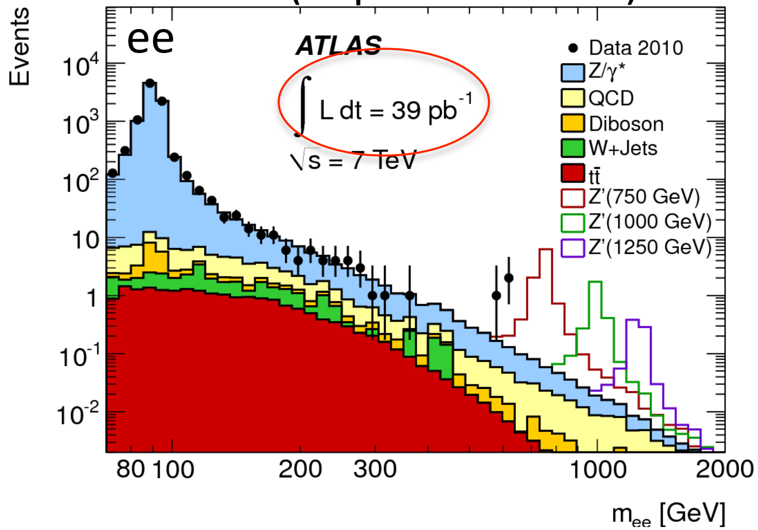
Data Quality - Introduction

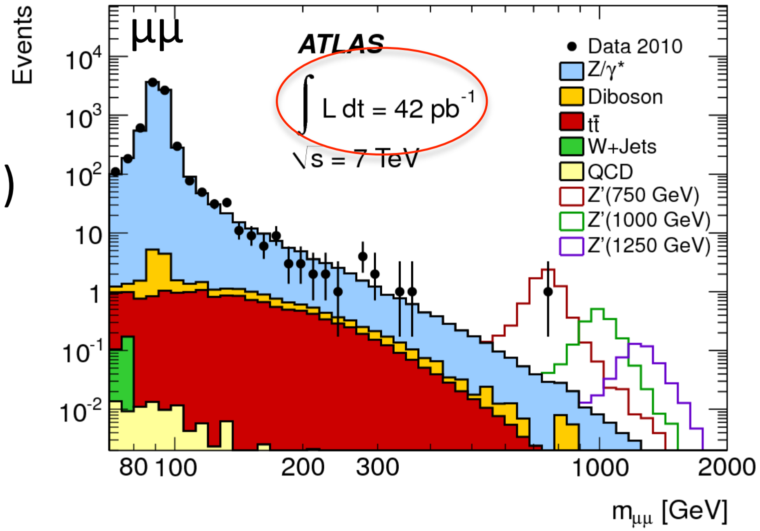
- Data Quality (DQ) is the system for telling people **what data to use for physics analysis**
  - DQ also maintains a 'known problems' database
- Data can have bad quality because
  - Detector problem (dead channels, noise, data corruption)
  - Trigger / DAQ problem
  - Bad calibration / Reconstruction problem
- Data time granularity
  - ATLAS data is divided into 1 minute luminosity blocks (LB)
  - CMS use 23s lumi sections
- This is the time unit used for DQ and luminosity measurement
  - Eg. If a detector has a problem for 5 mins the corresponding LBs will be marked bad for physics for that detector
- DQ recorded for different systems separately
  - Can have a LB good for muons but not good for calorimeter
- DQ includes offline reconstruction and calibrations
  - Can recover some DQ efficiency in future data reprocessings

DQ general

- What granularity to apply to DQ depends on what kind of problems occur and at what frequency and how they effect physics analysis
  - Mark endcaps and barrel separately?
- The system has to be **flexible** to be able to cope with whatever problems may occur
- Both ATLAS and CMS use different *Good Run Lists* for analyses that use only muons, compared to those that use all systems
  - Removing lumi blocks where the calorimeter has noise is not needed for Z' -> μμ analysis
- Can also remove **events** which suffer from a certain type of problem
  - e.g. DAQ problems or very short lived problems (ms)
  - Lose much less data by veto'ing events
  - Need to correct luminosity for event loss
  - Make sure this doesn't bias any physics analysis

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

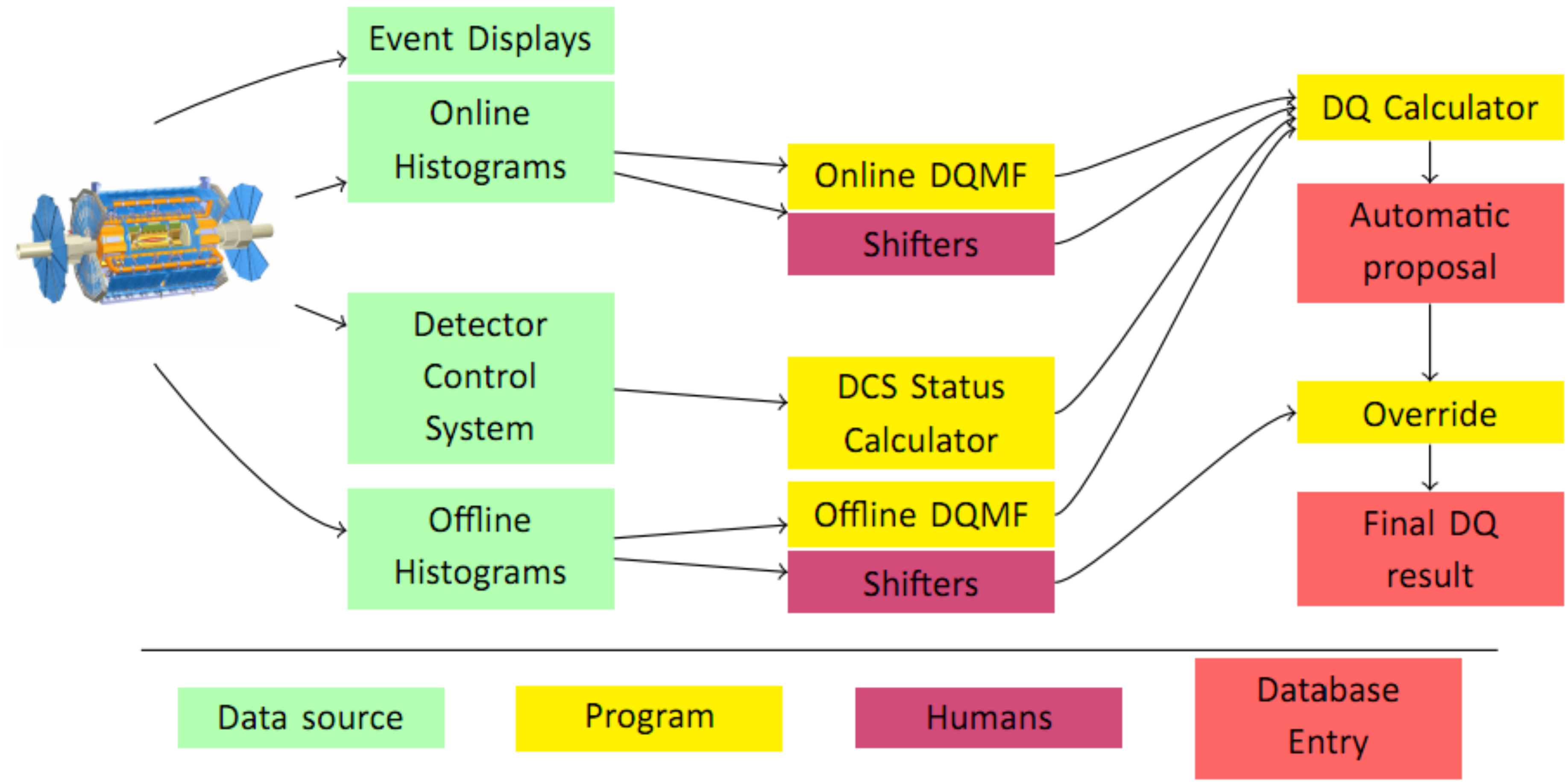




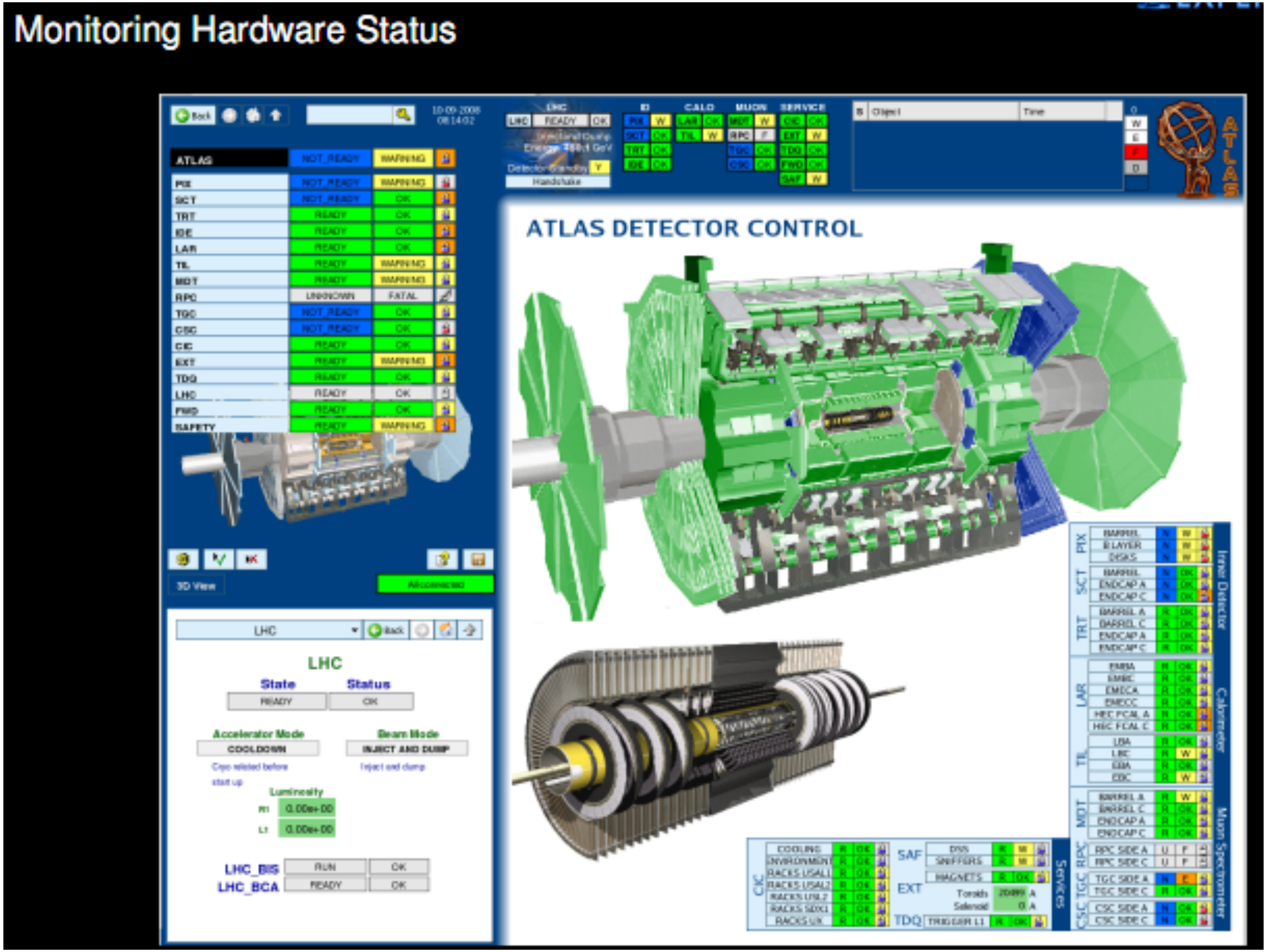
DQ assessment

- DQ assessment requires a lot of manpower
  - In ATLAS >10 shifters at one time, DQ meeting every week day, ….
- DQ assessment procedure based **on automated checks** and **manual checks** by shifters
  - Online and Offline histograms checked
  - Offline DQ assessment initially based on histograms from the express stream processing then the physics stream processing histograms are also used
- The Detector Control Software automatically marks LBs as bad when there are problems
  - e.g. HV, LV, cooling problems in more than a predefined fraction of the detector
- Final **DQ sign-off** of a weeks worth of data **once per week**
  - Procedure can be exceptionally accelerated on request from physics management
  - Minimum time from run being taken to DQ signoff is ~4 days
- DQ information is stored in database with versioning to allow reproducibility of results
- **Infrastructure to create Good Run List** (list of LBs passing a DQ selection) and to apply this in physics analyses (calculate lumi for a given GRL)
  - For luminosity determination very important to be able to keep track of which LBs were used for an analysis (even if no event is selected from that LB)

ATLAS DQ workflow



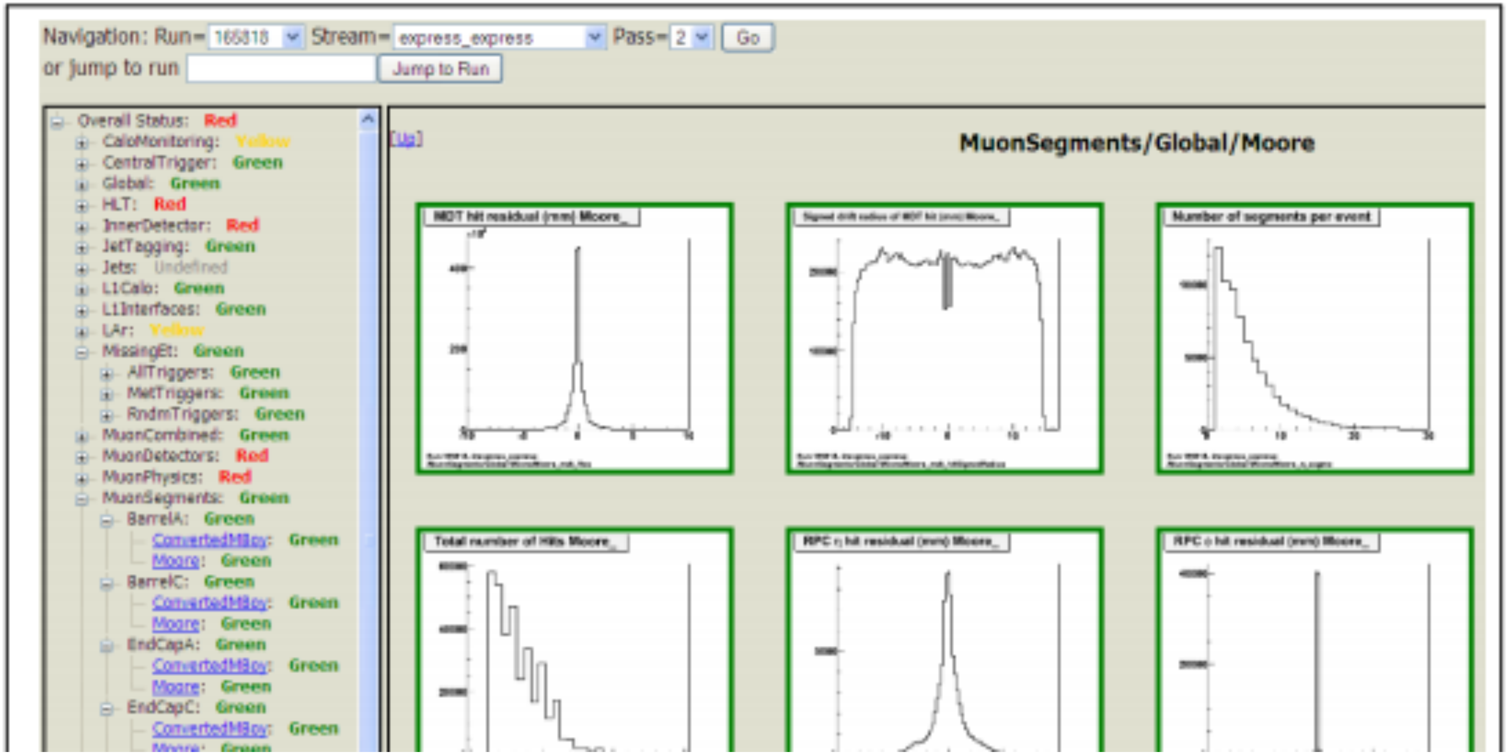
Slow control detector monitoring (HV, LV, cooling, …)



Offline monitoring

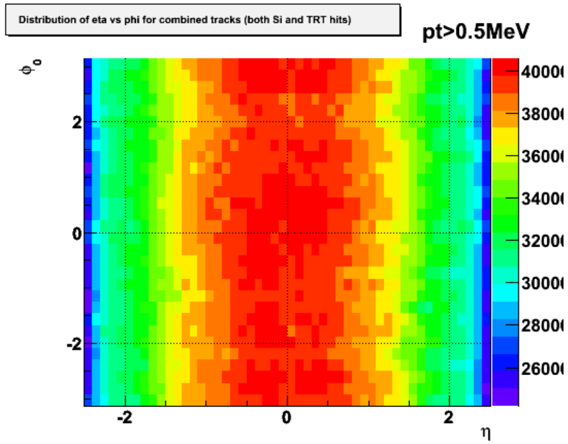
Per active trigger data stream:

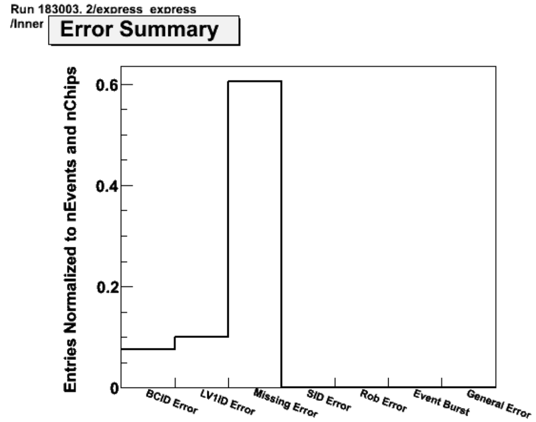
- 20,000 histograms per run - checked by DQ algorithms, flagged.
- 700 histograms every ∼ 20 minutes.
- Image files generated on request and cached.

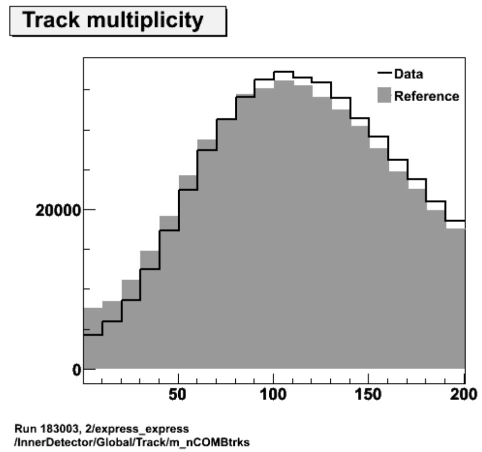


DQ quantities monitored

- Hit multiplicity maps (in eta/phi, or hardware space)
  - Look for dead / noisy regions
  - Extremely useful (difficult when you have ~100M channels!)
- Errors in the data
  - Counts of DAQ errors etc..
  - Reconstruction errors
- Object multiplicities, quality, resolutions and efficiencies
  - Can be quite simple: Nmuons, trks in eta/phi, hits on track
  - Or complex like full Z->ee tag and probe analysis
- Noise monitoring can use EMPTY triggers
  - No colliding bunch in the detector
- Time granularity of such plots requires some thought
  - Want to be fine grained enough in order to be able to mark the minimum amount of data as bad
  - But need sufficient statistics to see an effect (& don't want to have to look at too many histograms)
- Reference histograms very useful for being able to spot problems – but references need to be kept upto date
  - need to be with the same detector and beam conditions (e.g increasing pileup)







DQ efficiency

Inner Tracking Detectors			Calorimeters				Muon Detectors			
Pixel	SCT	TRT	LAr EM	LAr HAD	LAr FWD	Tile	MDT	RPC	CSC	TGC
99.1	99.9	100	90.7	96.6	97.8	100	99.9	99.8	96.2	99.8
Luminosity weighted relative detector uptime and good quality data delivery during 2010 stable beams in pp collisions at $\sqrt{s}$=7 TeV between March 30th and October 31st (in %). The inefficiencies in the LAr calorimeter will partially be recovered in the future.										

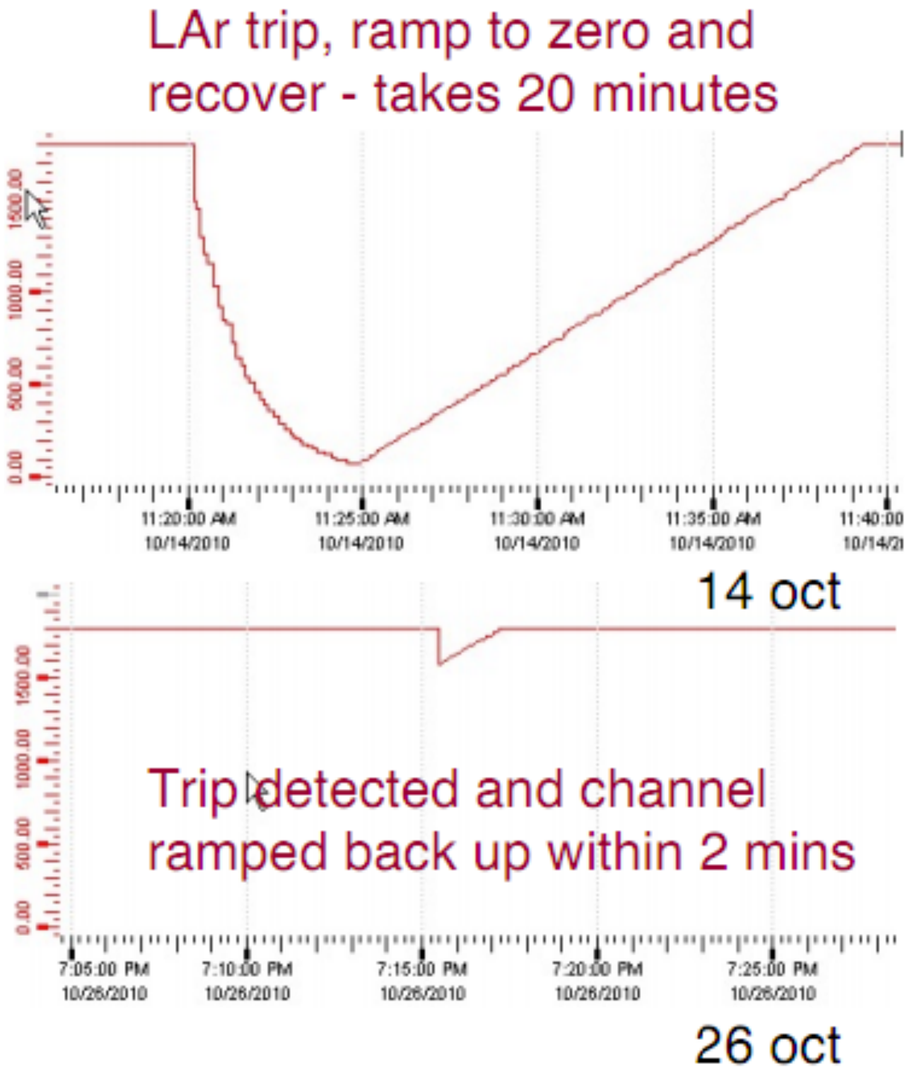
DQ efficiency for the different ATLAS detector systems for the 2010 datasets.
~10% loss of efficiency in the LAr calorimeter is due to:
- Sporadic noise bursts
-HV trips (data marked bad when voltage is ramping as induces noise in other channels)
Lots of effort going on calorimeter group to improve the DQ efficiency

-For 2011 data DQ efficiency for all systems to be good is ~85%

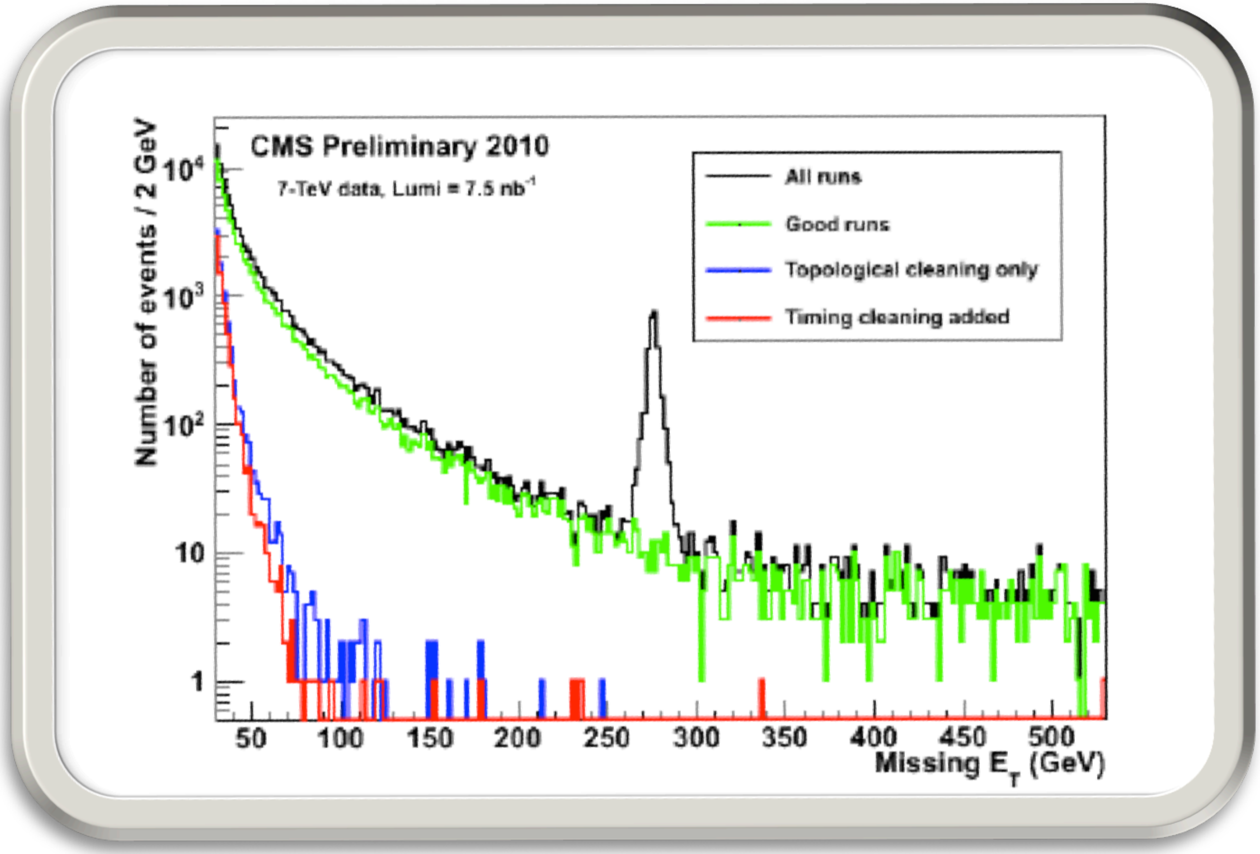
Example of improvements to increase DQ efficiency

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



Why a good run list is needed!



Missing E_T is generally the quantity most sensitive to bad DQ.
Noisy or dead regions anywhere in the calorimeter can make MET bad!

DQ for physics analysis

Despite the thorough DQ assessment that the data goes through it is still very important for **the physics analysis to thoroughly check the data** makes sense. DQ checks can not spot all issues.

Example of checks which should be done at the analysis level (of course which tests depends on the exact analysis):

- Plot the yields/luminosity as a function of time to make sure no run is wildly off

- could indicate a problem in one run
- helps validate luminosity of the dataset is correct

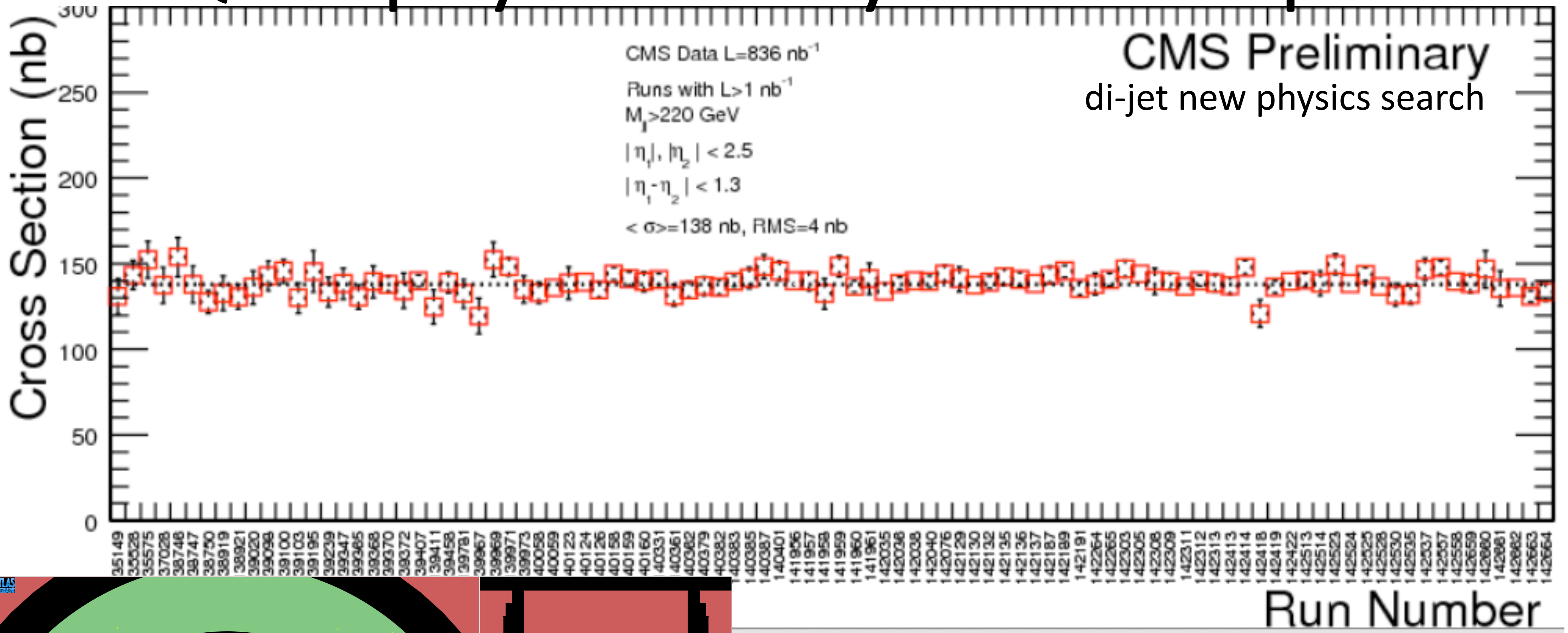
- Plot the eta-phi maps of analysis selected objects

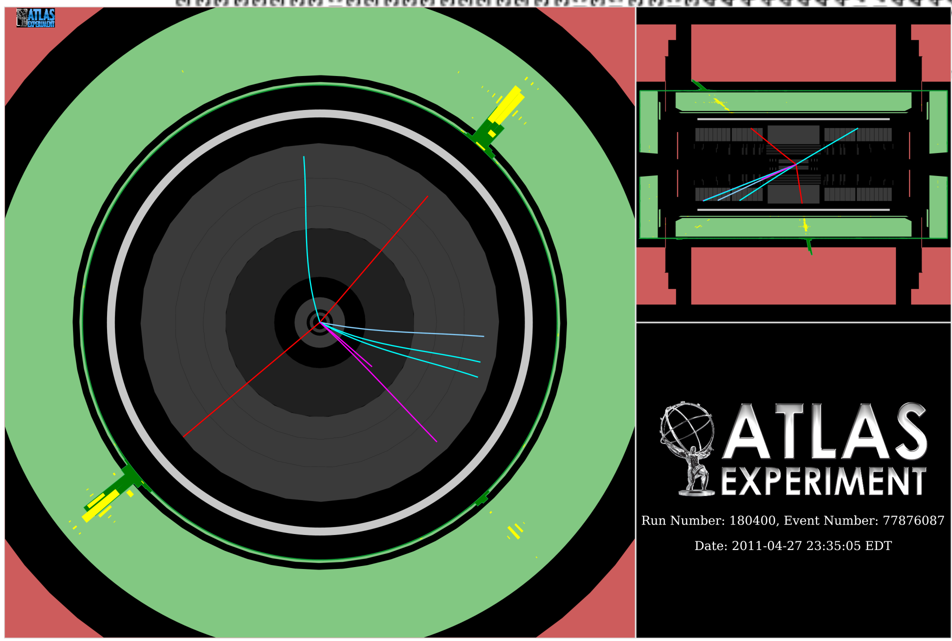
- to check this looks as expected – flat in phi, no hot-spots

- For search analyses which have a few events in the signal region it's a good idea to **look at the events in the event display** and to study the physics object quality for these events in detail

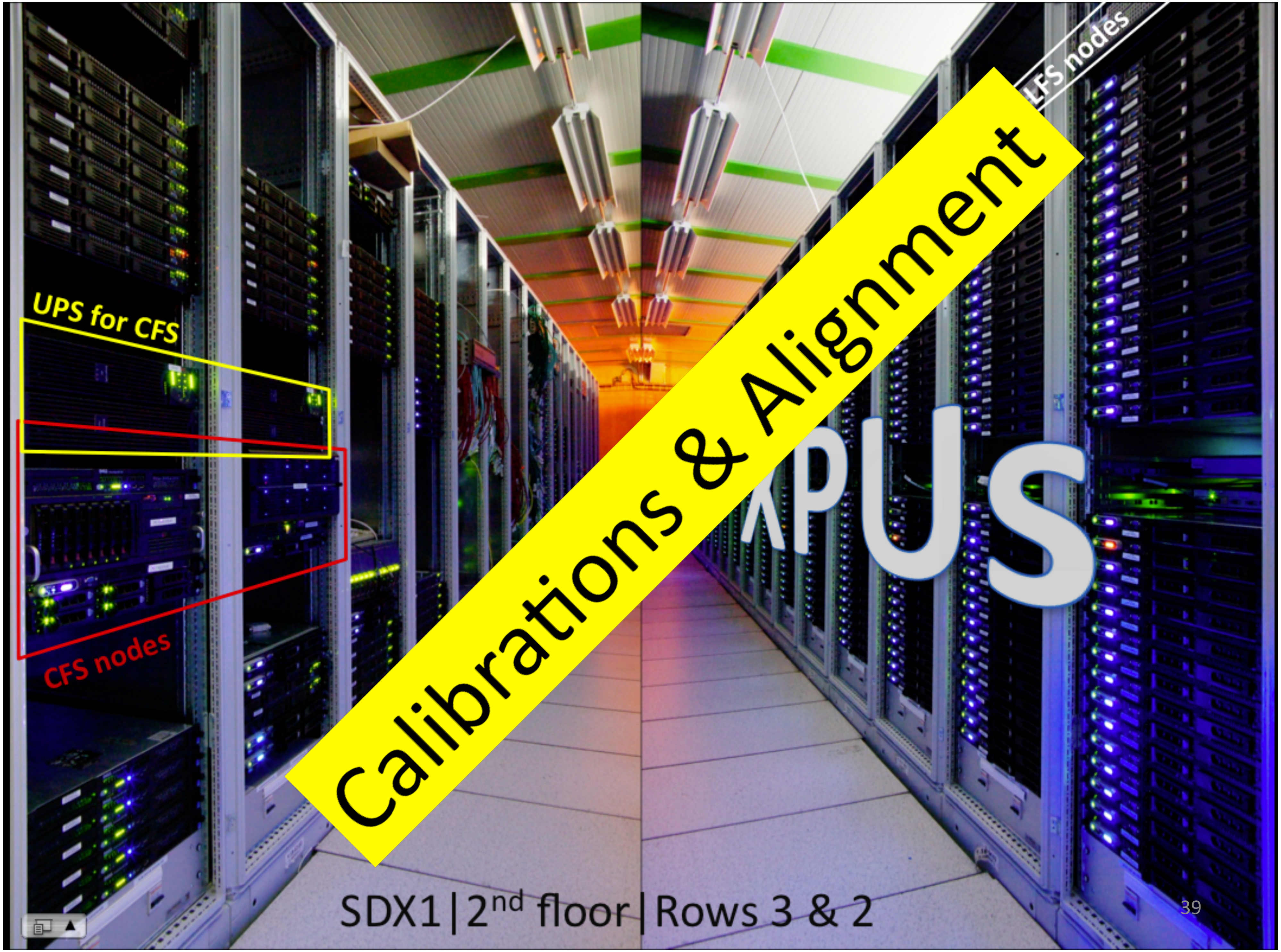
The above checks are an integral part of doing a physics analysis!

DQ for physics analysis - examples





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


UPS for CFS
CFS nodes
LFS nodes
Calibrations & Alignment
xPUs
SDX1|2nd floor|Rows 3 & 2

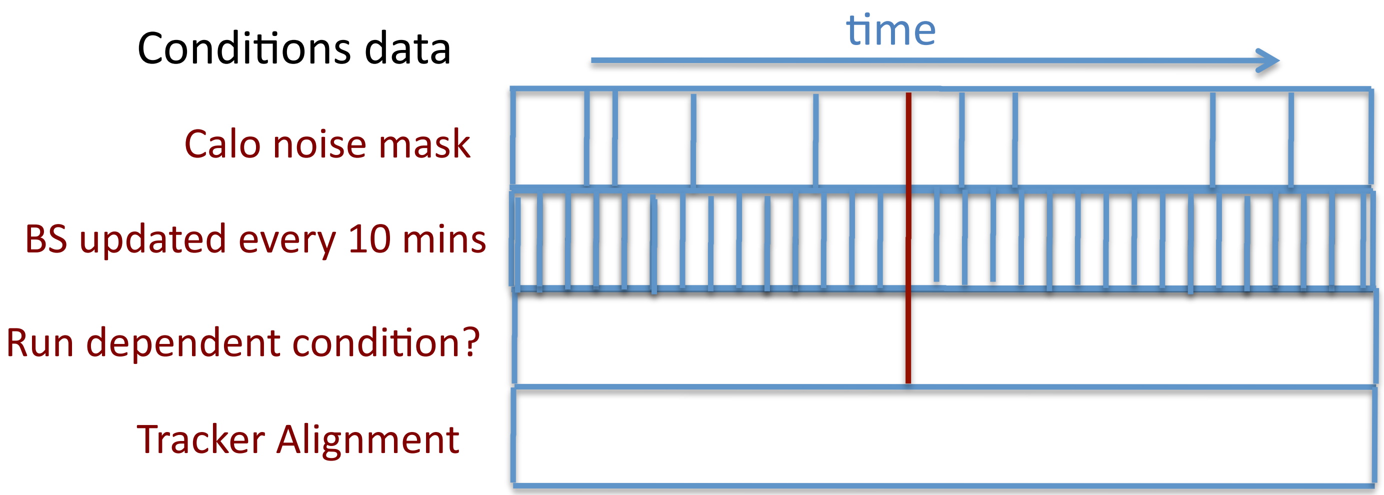
Calibrations and alignment

- **Many** offline **calibrations and alignments** applied in **reconstruction**
  - Tracker & Muon system internal alignments
  - Tracker, Muon system and Calorimeter global alignment
  - Beam-spot*
  - Dead/Noisy channel map*
  - Magnetic field map(s)*
  - Energy calibration in calorimeters
  - Drift-time calibrations in gas detectors (Tracking/Muons)
  - Many physics object calibrations
    - Jets, Egamma, B-tagging, Tau, …
    - Adds flexibility if these can be reapplied during analysis
- Which ones to apply to MC or not is not always obvious
  - Eg. Usually have perfect alignment in MC and try to make the data as perfect as possible (if you know how to make the MC misaligned like the data – you can just correct the data)
  - Quite often in MC digitization apply a correction eg. f(E) only to apply in reconstruction $f^{-1}(E)$
  - Different experiments deal with this in different ways

* - some of these things are true calibrations as such but they are treated in the same way in the reconstruction so I list them here and call them calibrations throughout this lecture!

Calibrations and alignment – time dependence

- Different calibrations have different **Interval-Of-Validity** (**IOV**)
  - Some are changed within a run
    - Beam-spot, noisy channel mask etc…
  - Some change slowly with time
    - Detector alignment
  - Some are linked to reconstruction version or the material map / geometry of the detector in the s/w
- When to apply new calibrations
  - Do they change with time (on what timescale?)
  - Can we get better with more statistics (for physics object calibrations…)
  - Condition versioning – need to be able to reproduce results later
- Need sophisticated database structure to be able to deal with the time varying nature of these calibrations
  - Different database tags for MC production, Tier0 reconstruction, reprocessings, HLT etc..
  - Bookkeeping is very important



Example of conditions DB tag. Many tags like this with different time structure – but can share some contents

Example of the offline calibrations applied to ATLAS data

Inner Detector

Pixel	Noisy channels	Each fill	50 CPU slot * hour on CAF per fill	Few MB/day	Raw data from Pixel calib stream	6	?	Medium	Depends on online mask
Pixel	Charge sharing	Each fill	5 ksi2k-hour on CAF	1.2TB on ATLCAL per week	Ntuple from express stream / ESD	6, 7	?	Low	Becomes important when alignment closer to optimal performance
Pixel	Depletion depth	Each fill			Ntuple from express stream	7	?		
Pixel	Lorentz angle	Each fill			Ntuple from express stream	7	?		
Pixel	Dead pixels	Weekly	50 CPU slot * hour on CAF per fill	100 MB/week	Raw data express stream	7	?	Low	N/a
Pixel	Resolution	Every few months	?	?	MC	?	?	?	?
SCT	Noisy strips	Daily	6 CPU	4 GB	Raw data from SCT calibration stream	7	< 100 kB/run	High	
SCT	Dead chips	Daily			express stream	6	< 100 kB/run	Low	
SCT	Dead strips	Weekly			express stream	6	< 100 kB/run	Low	
SCT	Module trip masking	Daily			express stream	7	< 100 kB/run	Low	Not yet operational
TRT	T0 and r-t relation	Each fill	70 CPU's on CAF	~5Gbs for 24h	Express stream (ESD)	7	?	High	At startup needed to determine t0s and r-t soon, and monitoring stability of them. Possible to run with misaligments.
TRT	Dead and noisy channels	Each fill	1st iteration of T0 and r-t relation	small	Express stream (ESD)	7		Medium	Critical only in case of many noisy elements. Possible to run with misaligments.
TRT	HT calibration	Monthly		N/a	calibration, e-gamma ntuples	8	?	High	For reprocessing - electron particle identification
ID Global	Beam spot	O(10) mins	CAF, small (t.b.d.)	Small (t.b.d.)	beamspot + express stream	[illegible]	?	High	Expect frequent changes due to LHC changes to optics. Resources estimates are for case where ES1 processed w/o beam spot constraint and the alignment is unchanged
ID Global	Alignment	Daily	? CAF nodes	10 GB/fill	ID alignment + cosmics s[illegible]	7	?	Low	No evidence for strong time dependent variations yet

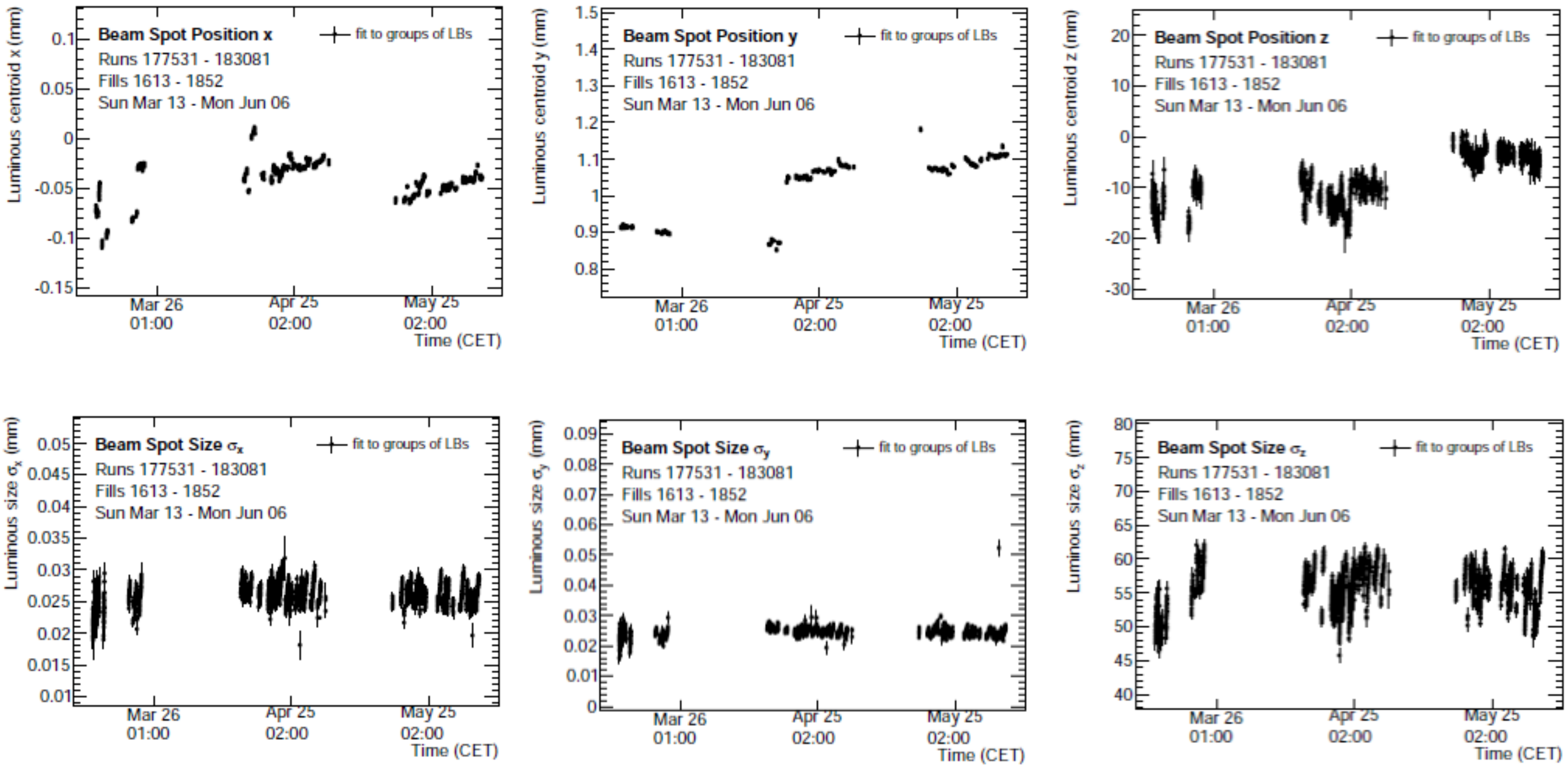
LAr	Noise and Pedestal Shifts per cell	Daily	CAF (~1 CPU)	(small)	ZeroBias stream and "empt[illegible] express stream) (zero-bias ev[illegible] D0_EMPTY)	7, 8, 9	9 Mbytes (Cool online) for noise (same for pedestal)	High	Needs ~10k zero-bias events over ~1 fill, i.e O(1Hz)
LAr	Noise autocorrelation	Daily	CAF	(small)	[illegible] o-bias evnes)Raw data	?	~60 Mbytes (pool) (intermediate conditions not directly used online or in offline reco)	High	Used to re-compute OFC for online calibration
LAr	Alignment	< Monthly	CAF	?	[illegible] cs (electrons) AOD Express stream	7, 8, 9	(tiny)	Medium	Required for physics: 0.1–0.2 mm (requires also offline analysis at beginning)
Tile	Analysis of physics data, Laser and CIS	2–3 times/year (depends on Lumi variations)	CAF (8 [illegible]	[illegible] ter [illegible] alysis)	Express, Muon, Minbias streams	7, 8, 9	?	High	Validate EM scale with muons, use E/p for iso. hads., validate OFCs and pulse shapes
Tile	Analysis of zero-bias events	~weekly (t.b.c.)			Express stream (zero-bias events)	7, 8, 9	?	High	Measure noise under changing running conditions

Muon Systems

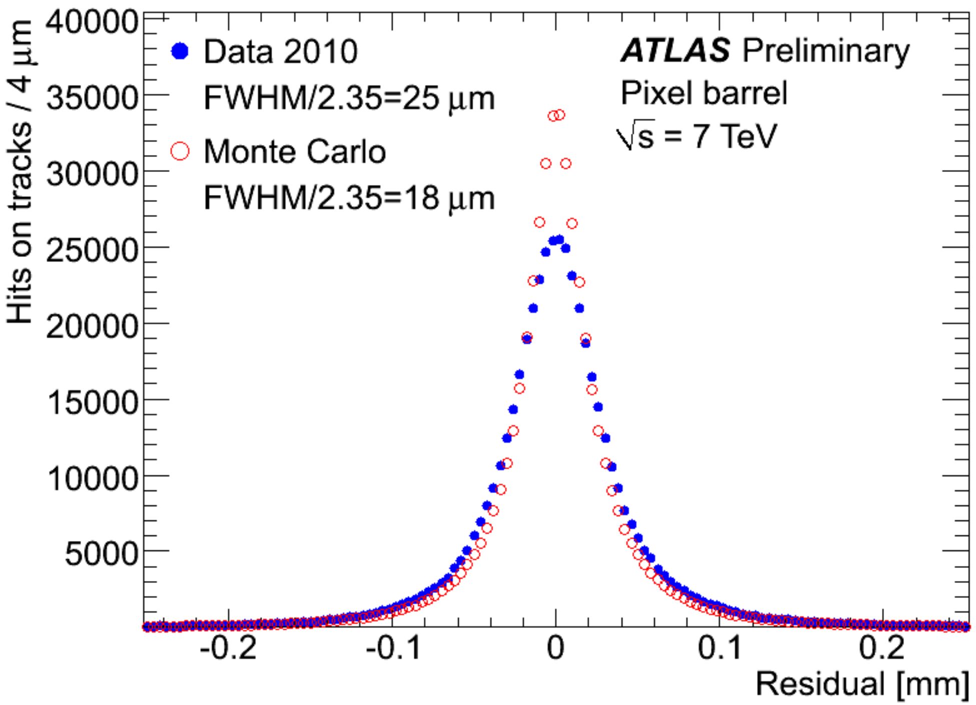
MDT	R-t calibration	Daily	[illegible] 3 Tier2's: 100 CPU's/Tier2	100 GB/day at Tier2's	Muon calib stream	7, 8	100MB	High	
MDT	Track based alignment	Daily	At Munich Tier2	?	Muon calib stream	8	100MB/day	High	Update all set of alignment constants from optics
RPC	Dead/noisy channels	Daily	At Rome/Naples Tier2's	?	Muon calib stream	7	?	Medium?	
RPC	Channel efficiency	Daily	At Rome/Naples Tier2's	?	Muon calib stream	7, 8	?	N/a?	
TGC	Track based alignment	Daily	At Tier2/Tier0	?	Muon calib/express stream	7, 8	?	Medium	
CSC	Track based alignment	Daily	At Tier2/Tier0	?	Muon calib/express stream	7	?	Medium	
MS Global	MS vs ID alignment	Daily	CAF	?	Express stream AOD's	8	100MB/day	Medium	Update all set of alignment constants from optics+tracks

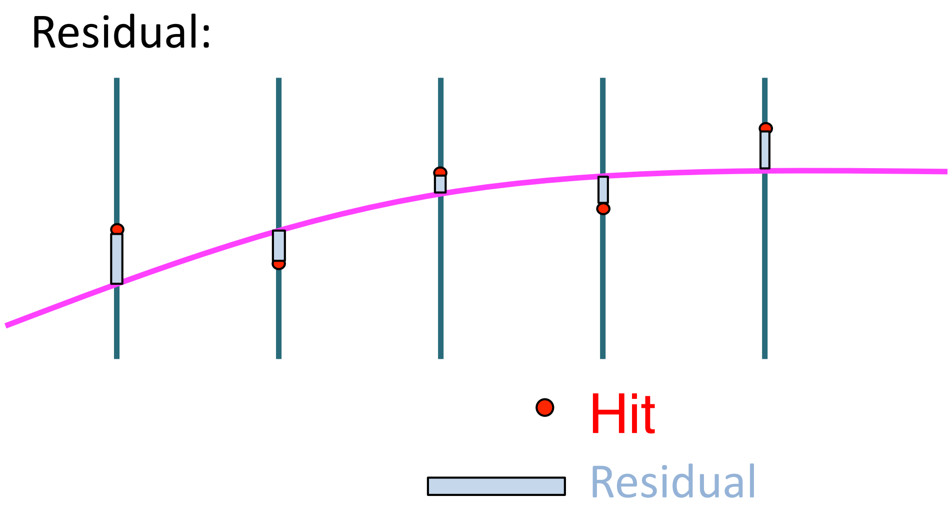
Beam-spot

- Calculated in the prompt calibration loop from reconstructed primary vertices (using tracks found without a beam-spot constraint)
  - There is also the online beam-spot which is determined and used in the HLT (eg. For b-jet triggers)
- Expect the beam spot to change from fill to fill
  - Can also change within the fill (emittance growth)
  - In ATLAS we derive the BS every 10mins
- Needs to be re-derived in a reprocessing if the tracker alignment changes



Tracker alignment





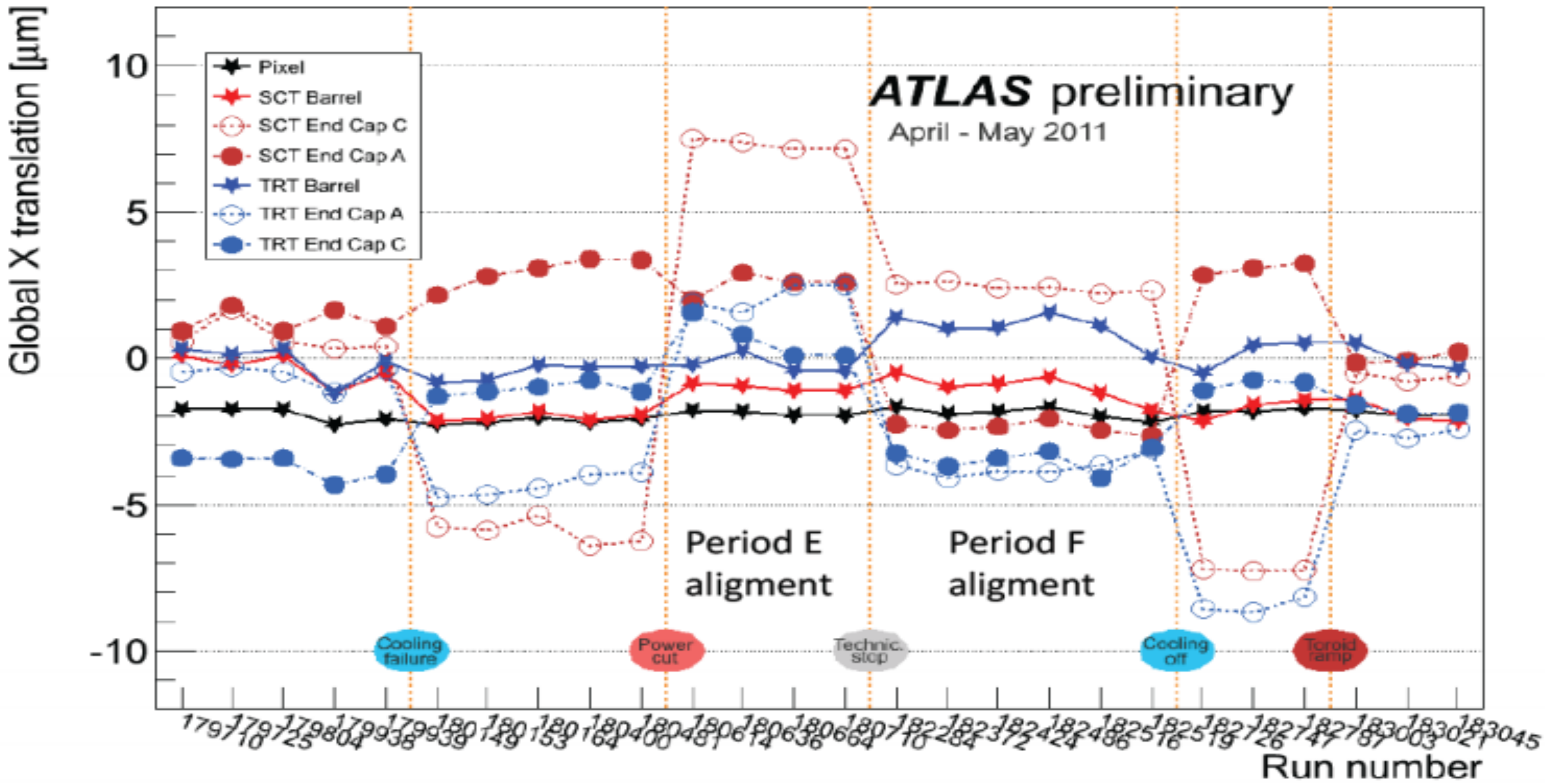
ATLAS pixel x-residual improvement with updated alignment.

Alignment derived from high P_T tracks in physics streams – but can use calibration stream (with only tracker information readout) in future.

So far ATLAS tracker alignment has only been done every few months and applied for the next reprocessing.

However recently evidence that the alignment changes when cycling the cooling (power cuts etc…) so the alignment will now be continuously derived and applied in prompt reconstruction if it changes significantly.

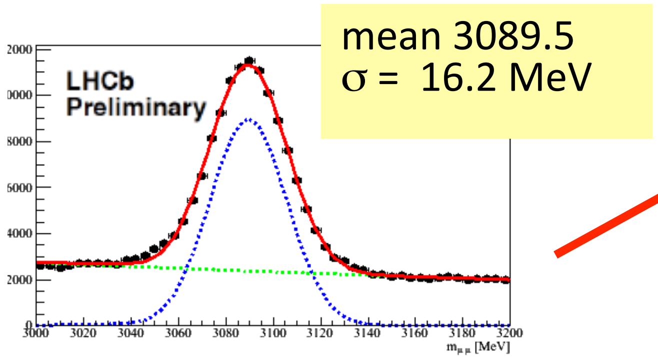
Tracker alignment



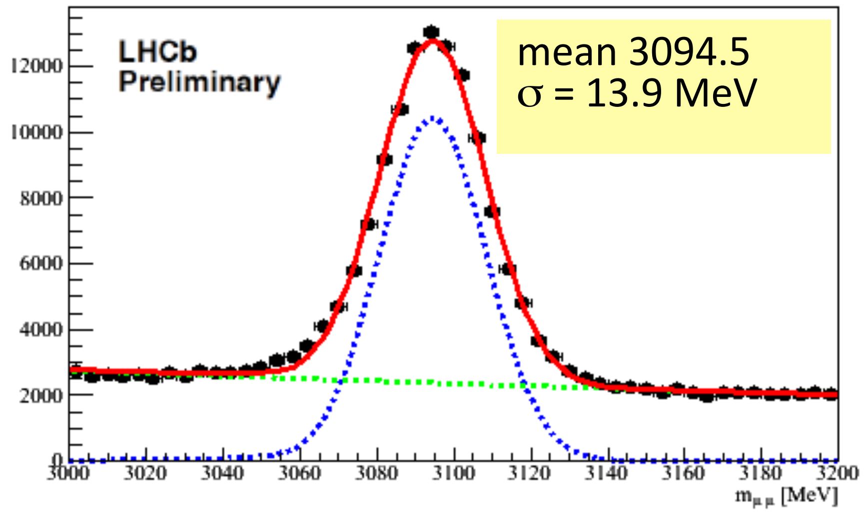
However recently evidence that the alignment changes when cycling the cooling (power cuts etc…) so the alignment will now be continuously derived and applied in prompt reconstruction if it changes significantly.

J/ψ and Υ → $\mu^+\mu^-$

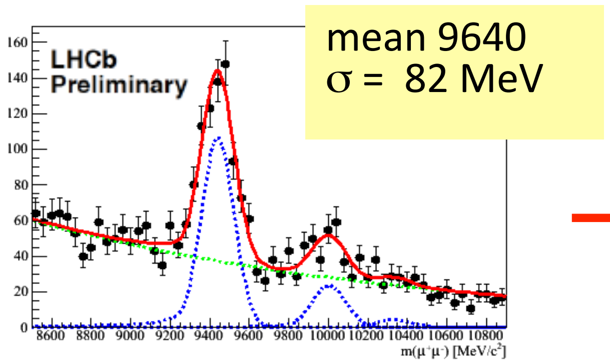
- LHCb improved alignment
- (LHCC meeting September 2010)

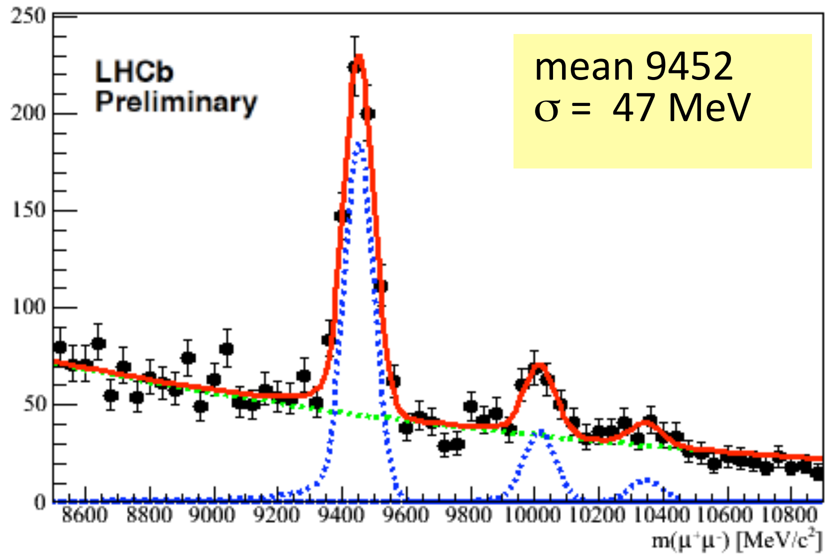


J/ψ PDG mass 3096.916 ± 0.011 MeV

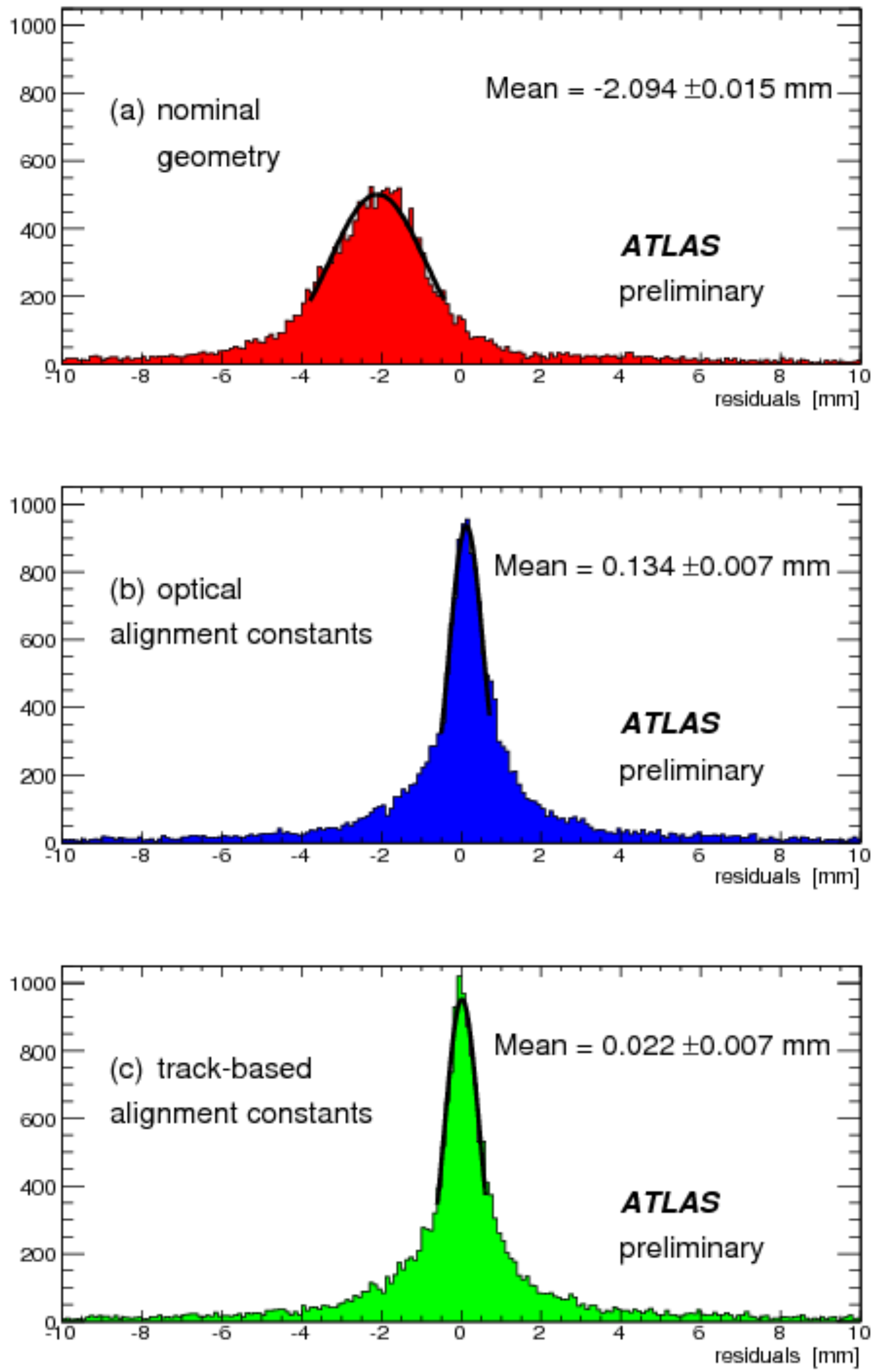


Υ(1S) m = 9460.30 ± 0.26 MeV, (2S) and (3S) states resolved





Better mass resolution directly feeds into better signal to background ratio in an analysis!



Muon system alignment

Muon system alignment using:

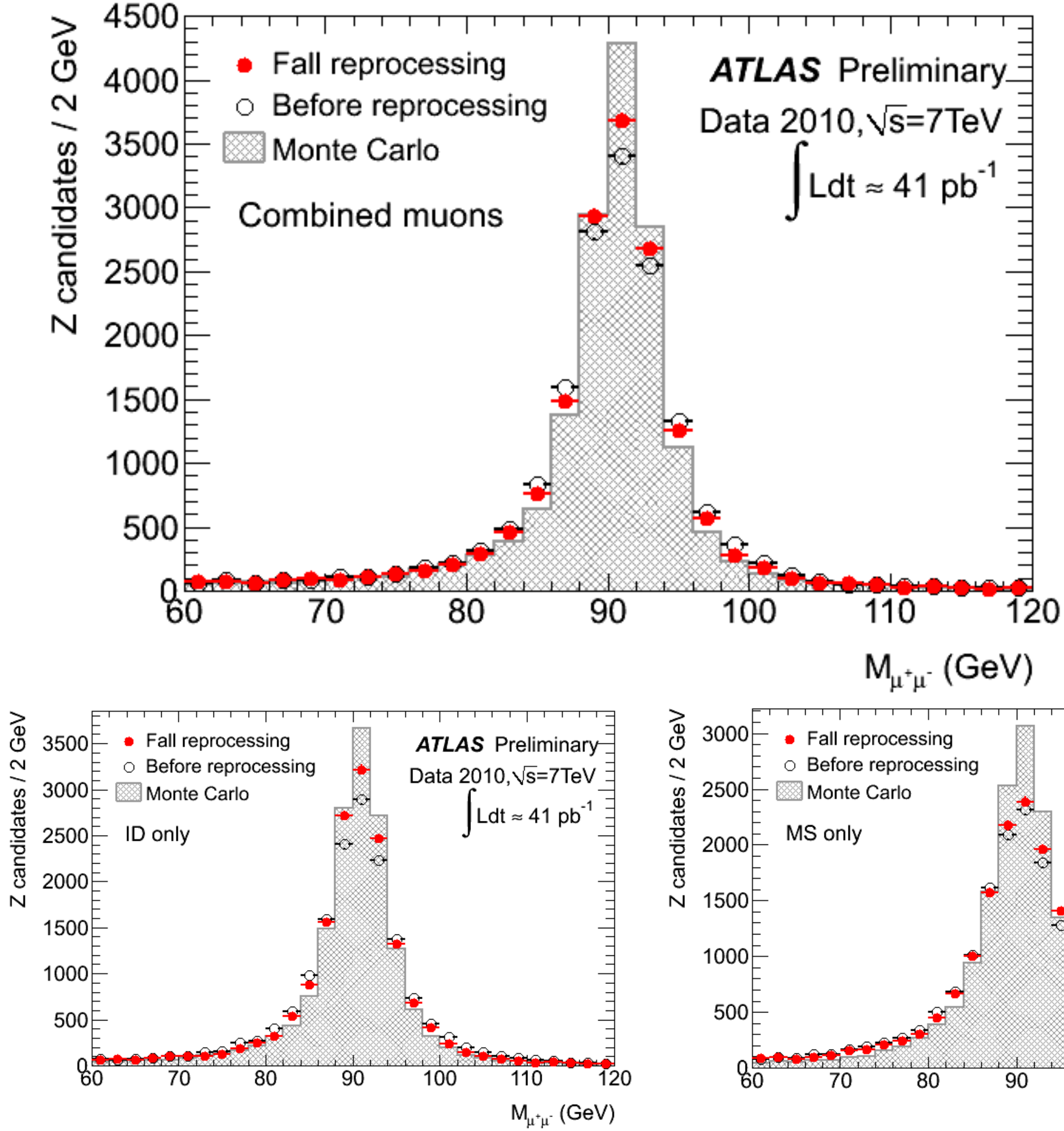
i) Optical alignment system
ii) Straight tracks (from toroid off run)

Give ultimate precision.

For 10% momentum resolution on a 1TeV muon need sagita uncertainty of 50µm – so need to align the muon chambers to better than 50µm
(Toroid off, solenoid on run allows one to disentangle muon alignment from magnetic field effects)

Muon system moves when toroid is ramped/dumped alignment needs to follow these movements.

Z->μμ improvement



Z->μμ in ATLAS has contribution from tracker alignment and from muon spectrometer alignment.


UPS for CFS
CFS nodes
Some more specific examples...
SDX1 | 2nd floor | Rows 3 & 2

Dealing with detector problems

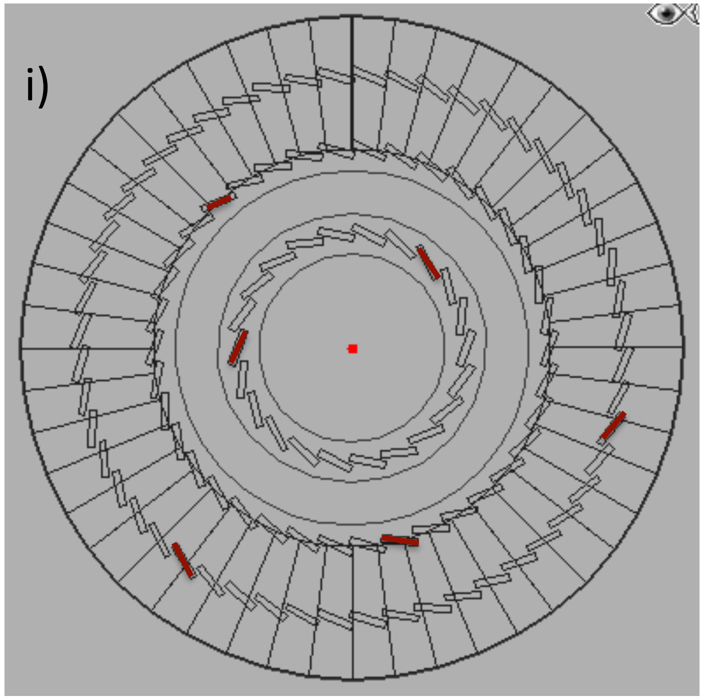
- Need to be able to cope with imperfect data from the detector from physics analysis
  - Long runs at the LHC don't allow certain problems to be fixed very often
- How one deals with a detector problems is not obvious – e.g a dead region in the detector
  - If possible to quickly fix it
    - Stop run / reconfigure / restart – but this procedure takes time which means luminosity is lost. For a very small region may not be worth doing
  - If there for a longtime can model it in the MC which means acceptance should be correct in MC. But of course still lose physics acceptance
    - Time varying MC that samples lumi weighted detector conditions is an option
- Very related to Data Quality
  - If 5% of calorimeter is dead for ~1% of data would probably mark that data as bad
  - But if the same 5% is dead for ~50% of the data will have to use that data (mark as good for physics)
    - Requires more work in physics analysis to be able to deal with such problems
  - **No easy answer** – system needs to be flexible
- Reconstruction techniques can try to minimize effect of dead regions of the detector

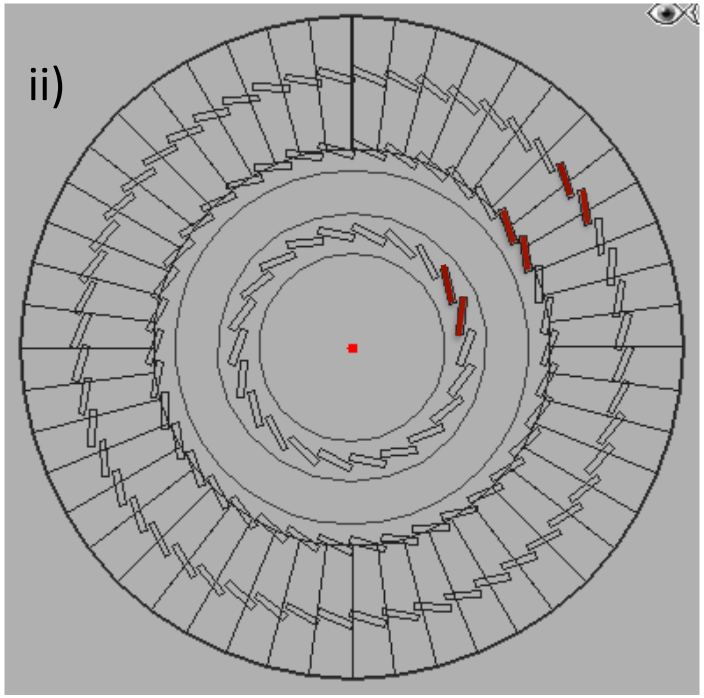
Dead regions - Toy example

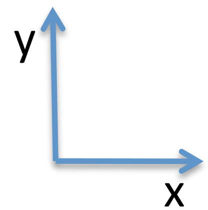
r-ϕ schematic of the ATLAS pixel detector.
Showing toy example of 6 dead modules.

i) Randomly scattered across the detector
ii) Concentrated in a specific region

Probably i) would count as good data quality and ii) would not as the tracking efficiency would be effected in this case.

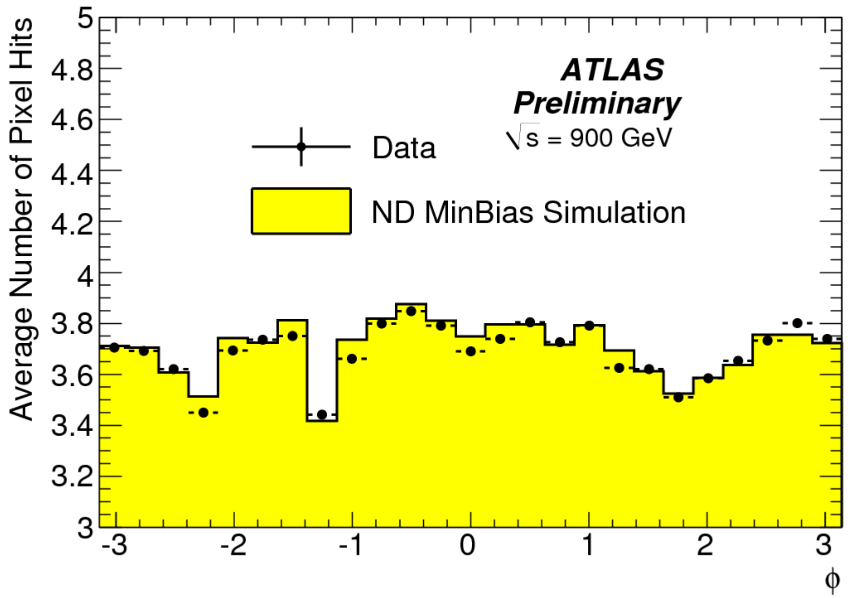


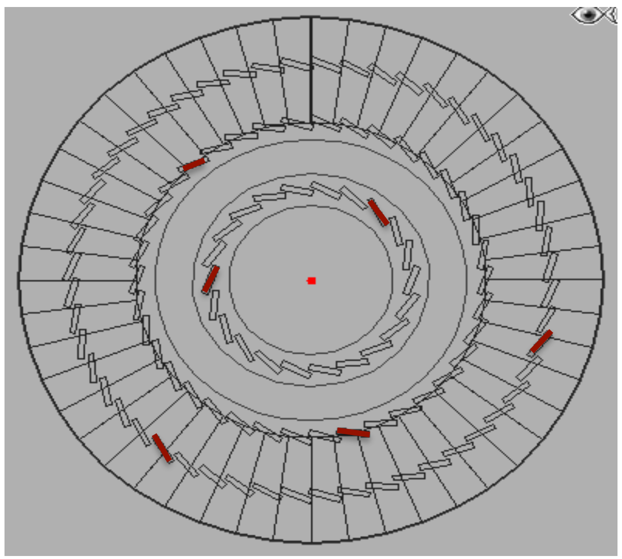




Not such an unlikely situation as often electronics associated with groups of modules which are physically close together

Realistic MC modeling





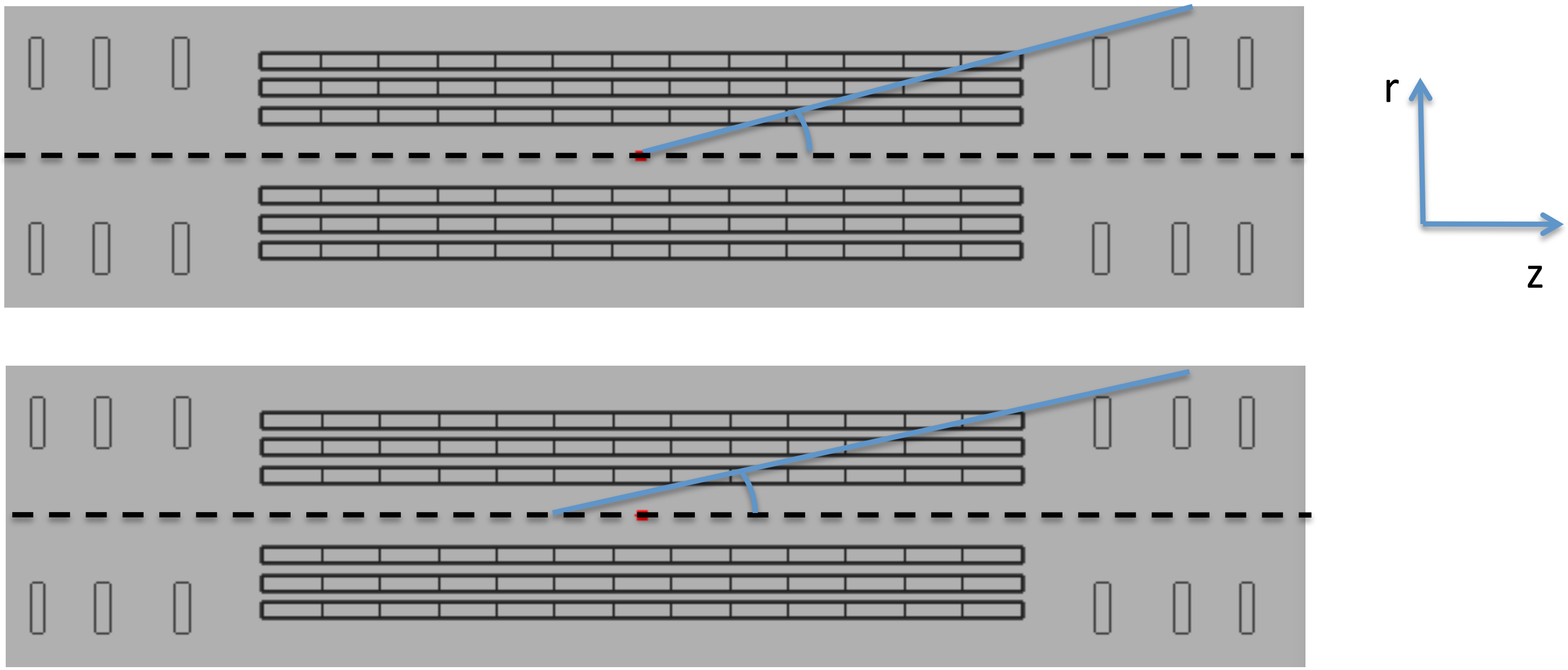
Often very important to model a realistic description of the detector data-taking configuration in the MC.

Eg. Number of pixel hits on track, as a function of track phi. Dips in distribution are due to a small number of pixel modules out of the readout.
MC made to model the data well, by excluding these modules also in the MC.

If the modules out-of-readout are changing with time this becomes hard to get right in the MC.

Reweighting MC

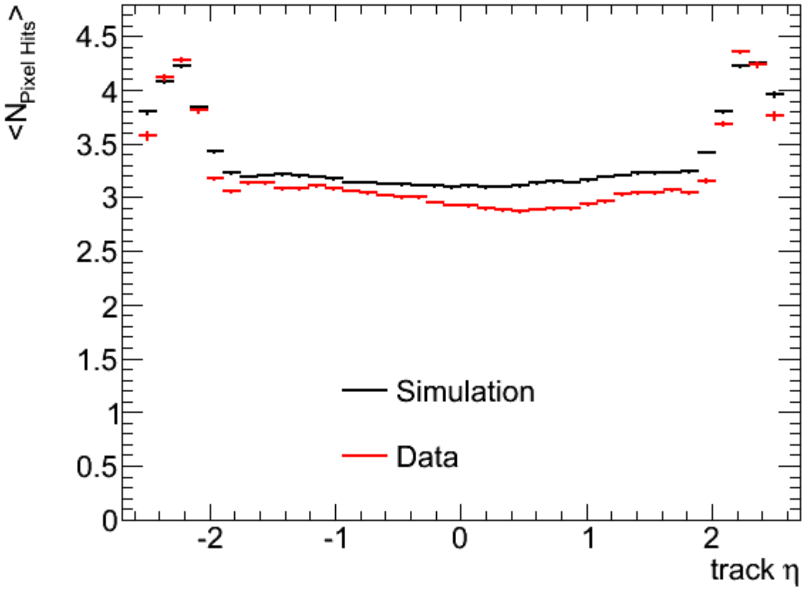
- Often when making MC can not guess exactly how the data will look
  - Can get round this by reweighting some quantity in the MC to look-like the data
- Good example Beam Spot reweighting

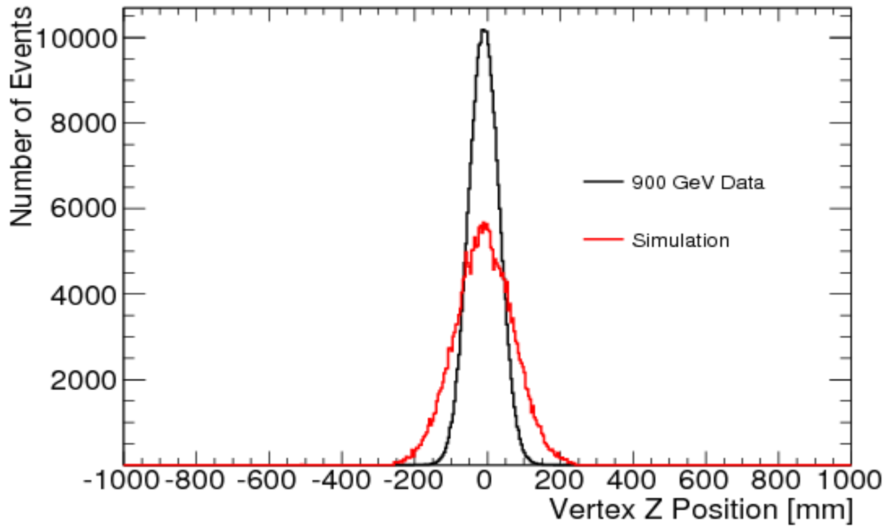


η of track to just hit first endcap disk depends on the track origin (ie. the beam-spot z-position)

Reweighting MC – Beam-spot

- Often when making MC can not guess exactly how the data will look
  - Can get round this by reweighting some quantity in the MC to look-like the data
- Good example Beam-spot reweighting

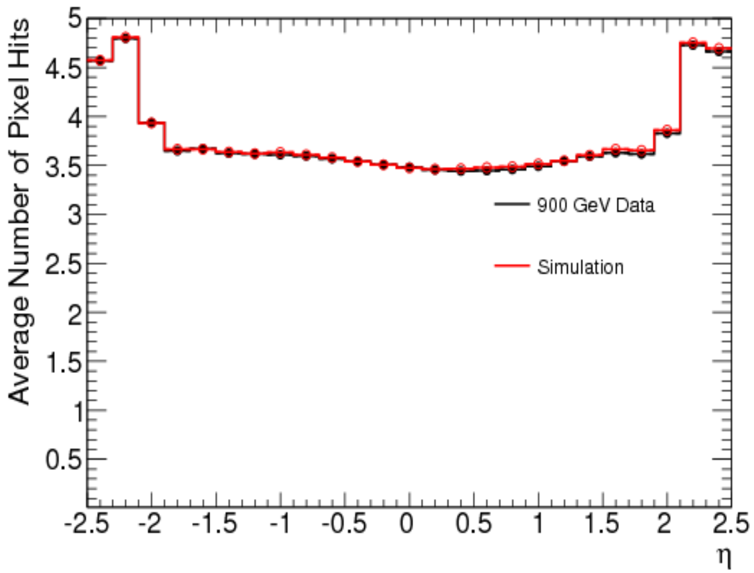


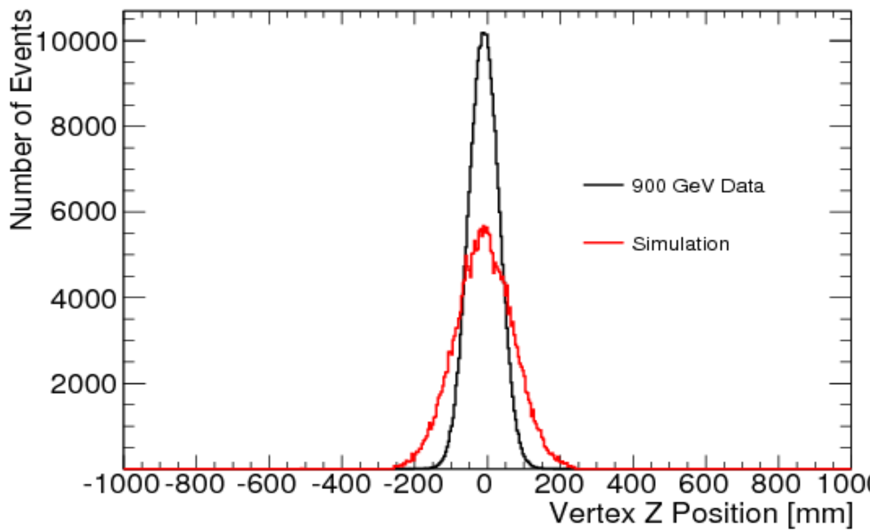


Data/MC comparison of the number of pixel hits versus η. Sensitive to the beam-spot z-position in the MC compared to the data. This is hard to get right in the MC as it changes with time.

Reweighting MC – Beam-spot

- Often when making MC can not guess exactly how the data will look
  - Can get round this by reweighting some quantity in the MC to look-like the data
- Good example Beam-spot reweighting

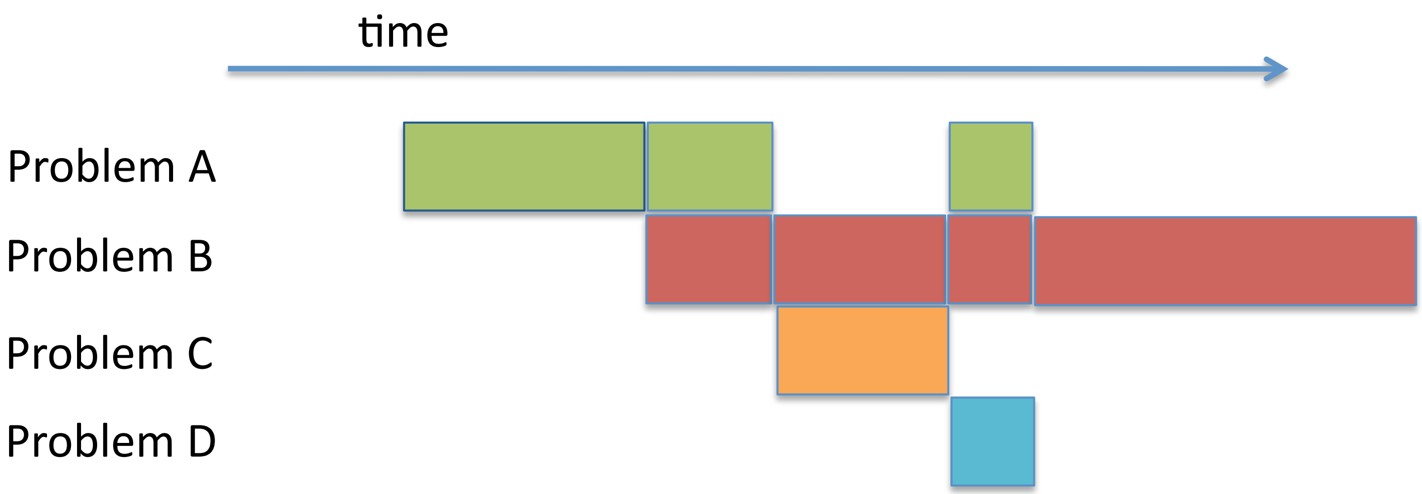




Can reweight the MC beam-spot z-distribution to mimic the data -> greatly improves the MC description of the data. Want weights to be close to 1 otherwise lose MC statistics. Similar reweighting techniques can be used for other variables to improve MC description of the data! (eg. vertex multiplicity reweighting for pileup)

Time varying MC

- A possible solution is to use time-varying MC
- Use real data conditions to model the dead channels in the MC in a time dependent way
  - Make your MC realistically model the luminosity weighted detector conditions
  - Takes into account the correlation between different detector problems
- However technically challenging
  - Use the real data conditions database in the MC reconstruction
- Both ATLAS & CMS are considering this for the future
  - Requires making MC to compare to data already taken - with known conditions
  - So far at LHC we have always been producing MC to compare to (mostly) future data which so we don't know the detector conditions of this data



As soon as there are a few different problems with different time structure - producing separate MC samples for each configuration becomes very messy!

Reconstruction Robustness

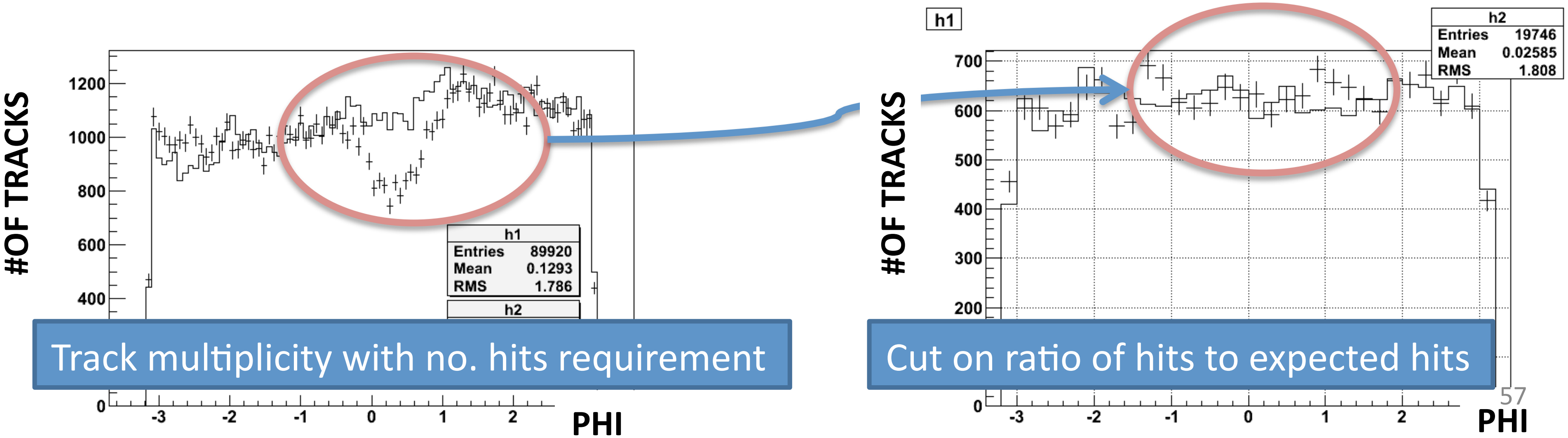
- **Can improve the robustness of the reconstruction to detector problems** in various ways

Example:

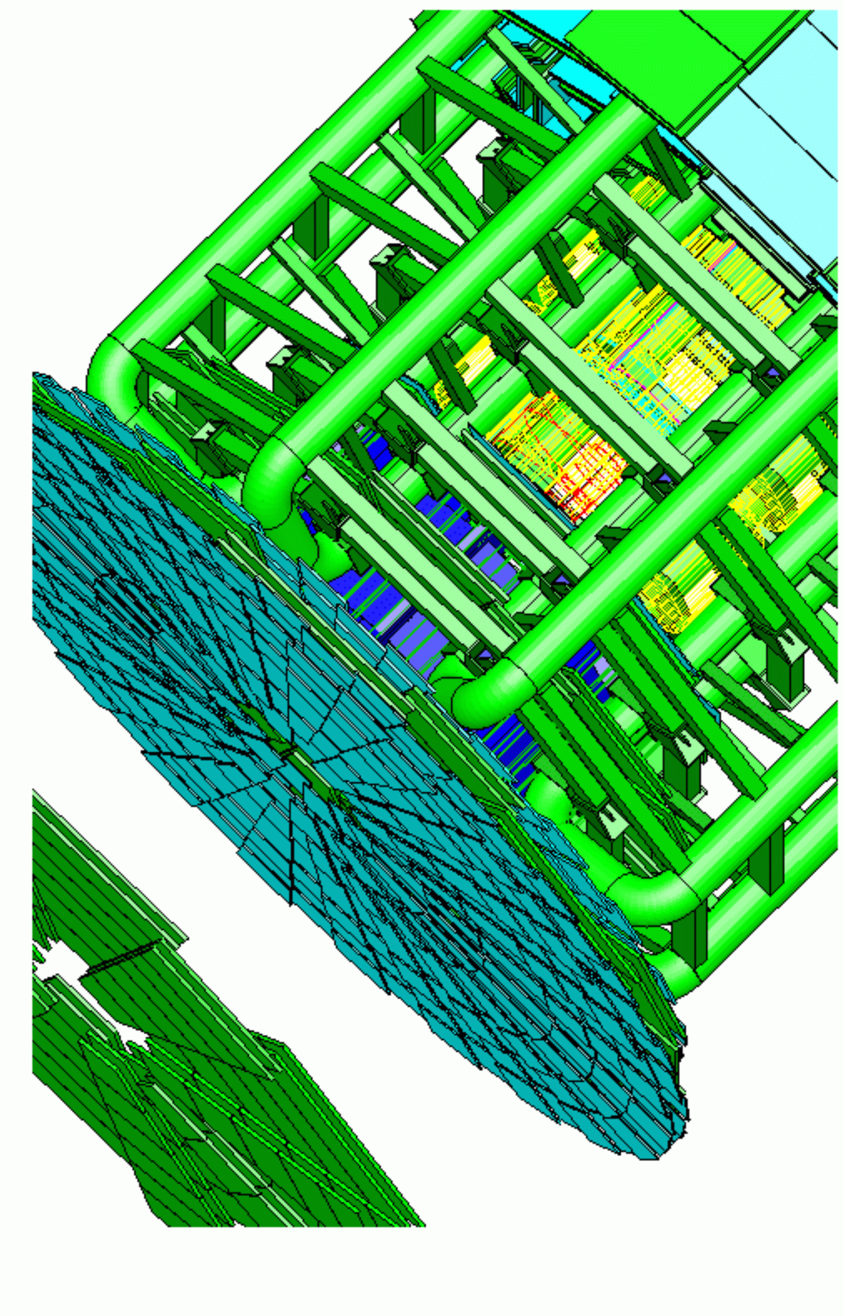
- Track reconstruction can use knowledge of which modules are in the read out (which can vary with time) in order to know what hit multiplicity cuts to apply
  - If there is a missing module require less hits on a track
- Much more robust against time-varying detector problems
  - Improves the physics quality of the data

Example of dead region of the detector

- Causes inefficiency if cut on number of hits
- No inefficiency if cut on $N_{HITS}/N_{EXPECTED\text{-}HITS}$

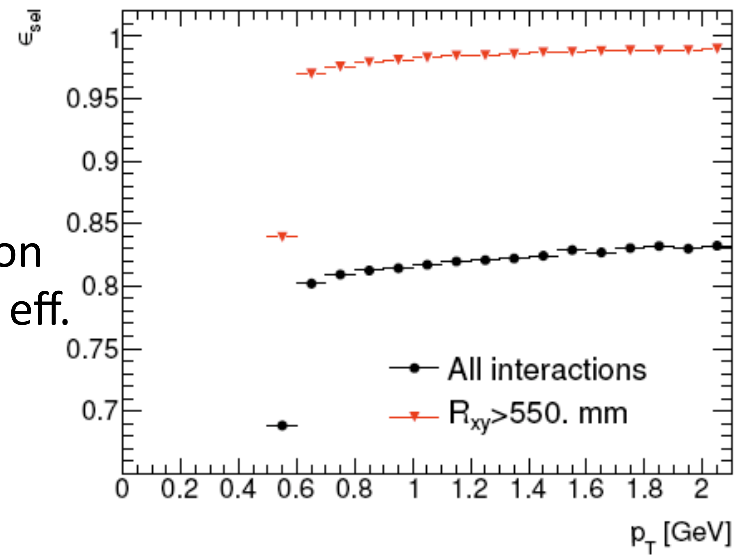


Material description in MC and reconstruction



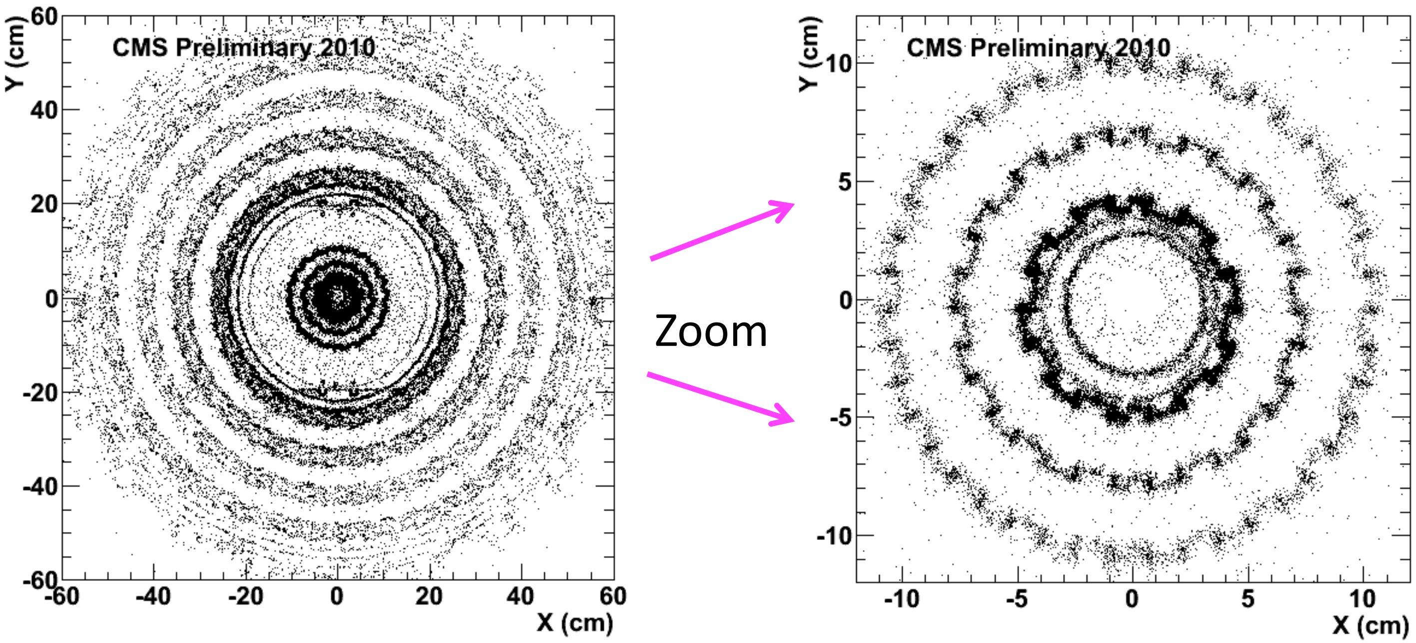
- A map of the material distribution inside the detector volume is needed in the simulation and the reconstruction
- Very important to have this as realistic as possible
  - Tracking efficiency dominated by material interactions in the tracker volume
  - Also effects calorimeter energy reconstruction
  - Muon momentum resolution dominated by material in muon system

ATLAS simulation Tracking eff.



Material description - validation

- Careful weighing of the detectors before installed
- Can use the following to check the consistency of the material map in the simulation compared to the data
  - Photon conversions
  - Nuclear interactions
  - Track extension studies
- In general, compare real data with detailed GEANT 4 simulation based on design, and gradually refine the material map in the simulation



Conversion material 'map' of CMS tracker

Conclusions…

- Data preparation **critical for physics results**
- Consistent reconstruction and calibration of data and MC essential
  - Both in prompt reconstruction and in reprocessings
  - Stability very important
- Calibration loop allows high quality data to be available for physics analysis in short time
- Proper detection, flagging, bookkeeping and application (GRL) of DQ assessment results crucial for reliable physics output
- Clever tricks in reconstruction and MC simulation can help us use imperfect data for physics analysis
- Offline processing is a complex system
  - Interplay between different conditions, data and software
  - Flexibility important to be able to deal with changing conditions
- High quality physics results from experiments proves the system is working well
- Challenge is continuing smooth operation
  - With increased luminosity and pileup
  - With long operation of experiments without shutdowns (problems can occur that cant be fixed)

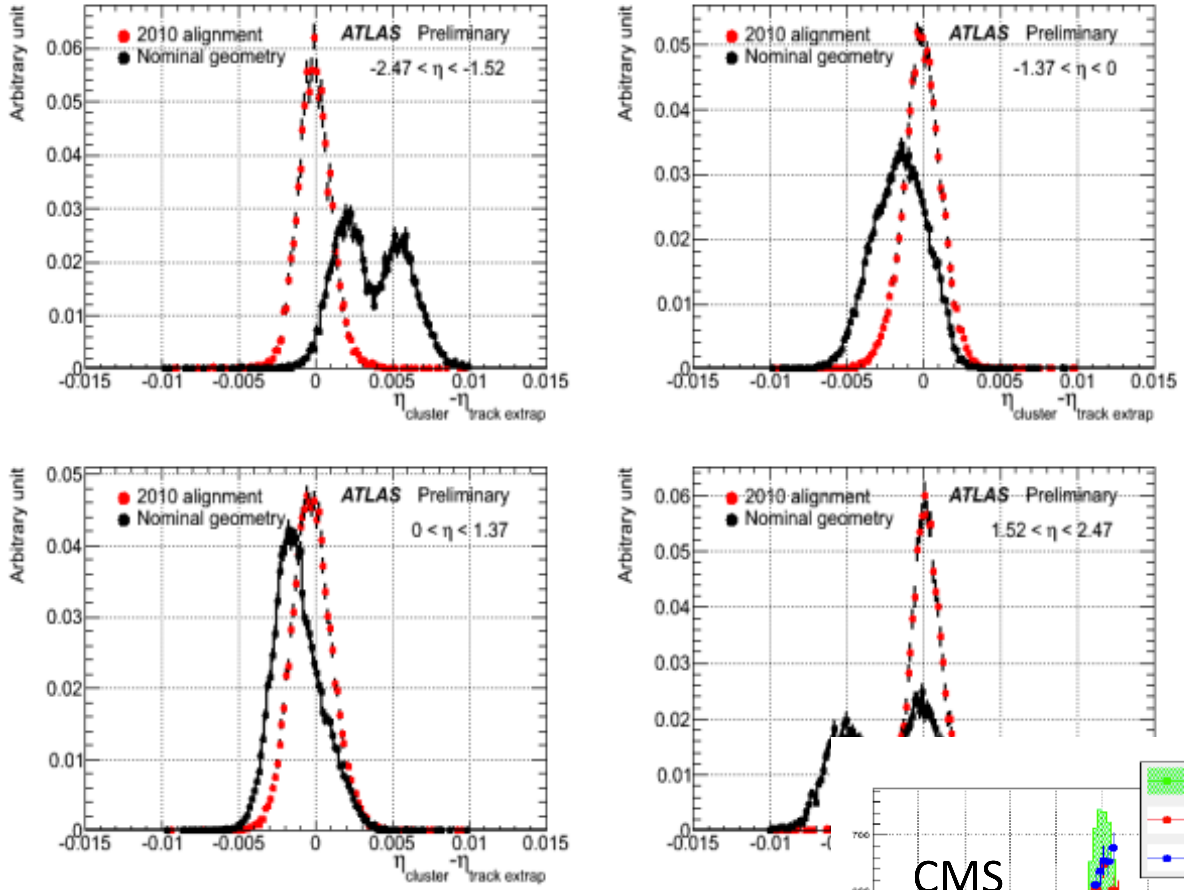
Final words….

- There are a number of constants themes between the trigger system and the offline data preparation
- For both there is a tension between **improvements** and **stability/robustness**
- **Validation** is extremely important
  - The trigger decision really can not be re-done and therefore this is more critical, but for LHC data volumes offline reprocessing of the full dataset is extremely resource intensive (both human resources and computational resources) and so can not be redone easily
- Propagation of information to the physics users is very important
  - Trigger prescales/menu
  - DQ status
- Both the trigger system and the offline processing design has proved to work after a successful commissioning period
  - Both are **complex systems**
- Now the challenge is **robust and efficient running** with increasing luminosity and pileup
- **Flexibility** is key to be able to deal with future challenges

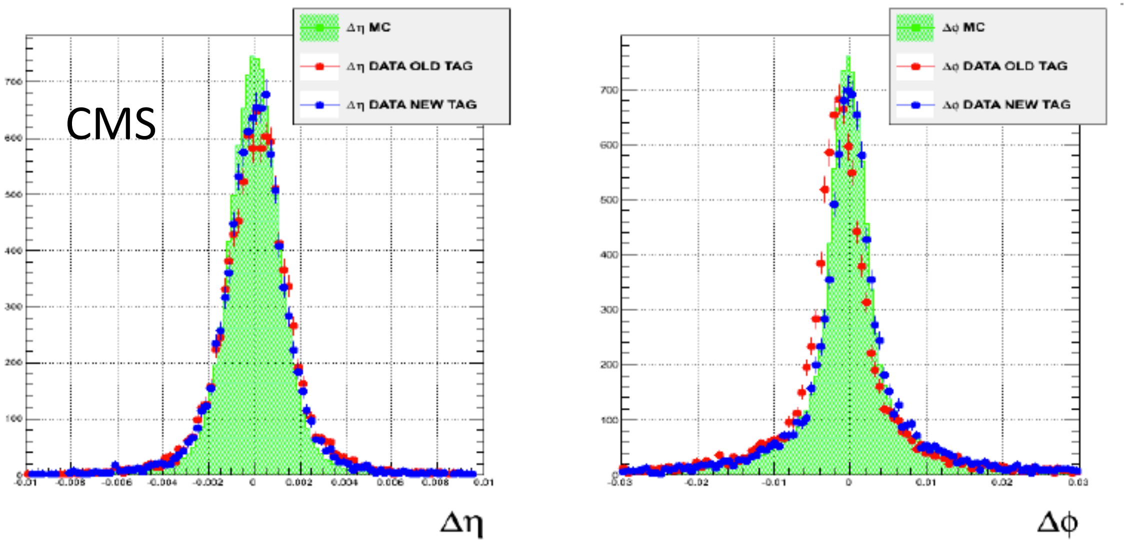
Thanks to

- Patrizia Azzi, Beate Heinemann, Andreas Hoecker, Pippa Wells, Heather Gray, Armin Nairz..

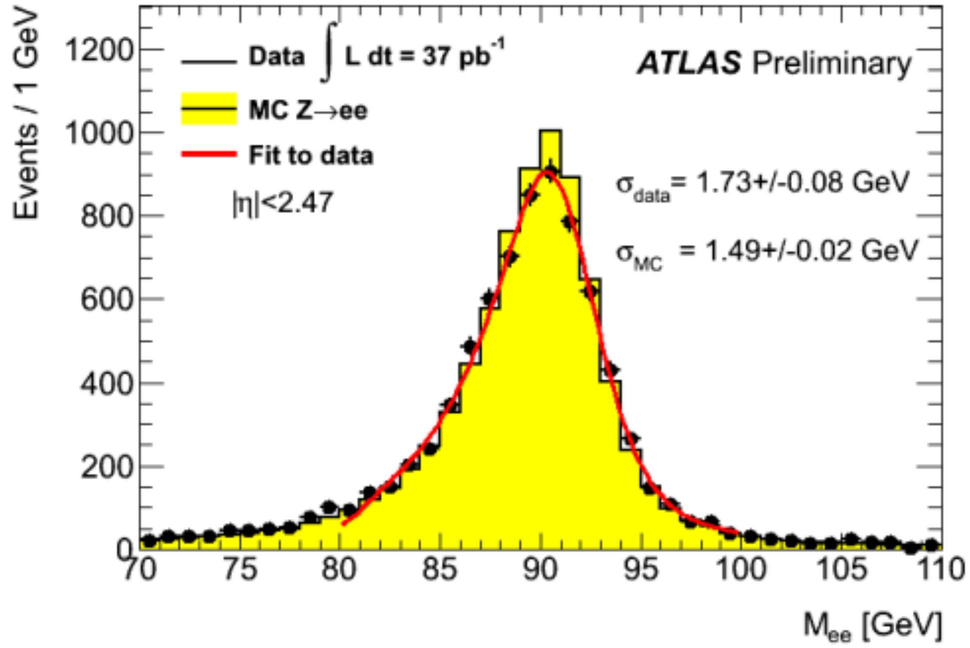
Calorimeter alignment



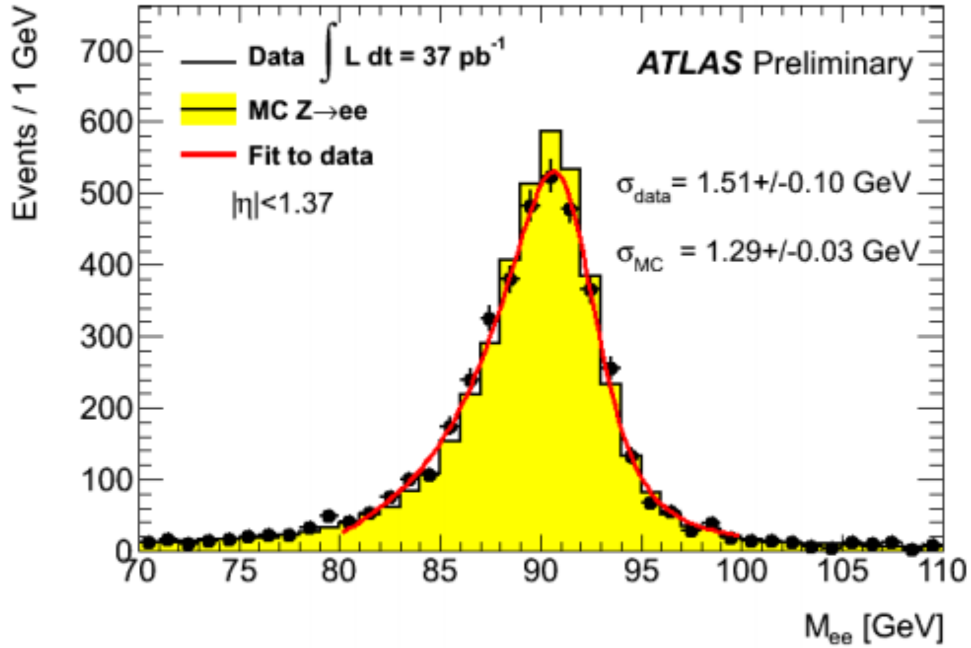
Alignment of the calorimeter with respect to the tracker.
Very important for El and Ph ID (calo/track match, E/p etc…).
Calibration from 300k inclusive El candidates (pt>10GeV).
These constants applied at reconstruction.
Calibration only changed when the detector is opened in a shutdown.



EM calorimeter calibration



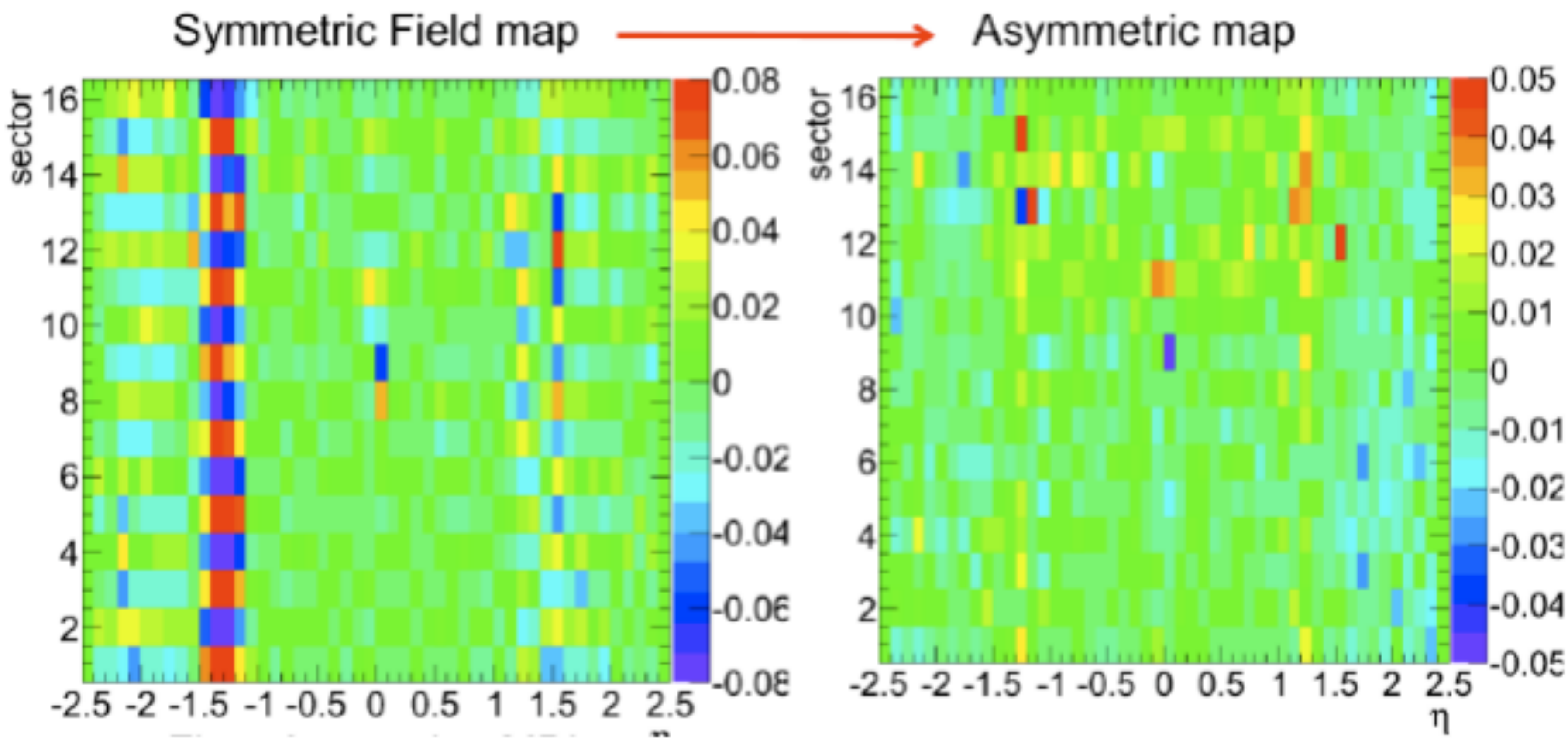
All pairs



Barrel-Barrel only

EM calibration done by constraining di-electron pairs to follow the Z-lineshape (from MC). Calibration carried out in 28 regions of the calorimeter.
Average ~-1% for the barrel and ~2% for the endcaps (consistent with expectations from test beam – where the original scale was set with a ~3% uncertainty due to temperature sensitivity).
This calibration is not expected to change with time, and so is only recalibrated when there is a large increase in statistics which allows a finer grain calibration to be applied.

ATLAS Toroid Field Map



Difference between muon momentum in tracker and in muon spectrometer.
Can see a ~5% bias in muon momentum at eta~-1.5 (for all phi) in the old map.
Much improved with new Asymmetric field map.
(Use new map in both data reconstruction and simulation generation & reconstruction)

When something goes 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

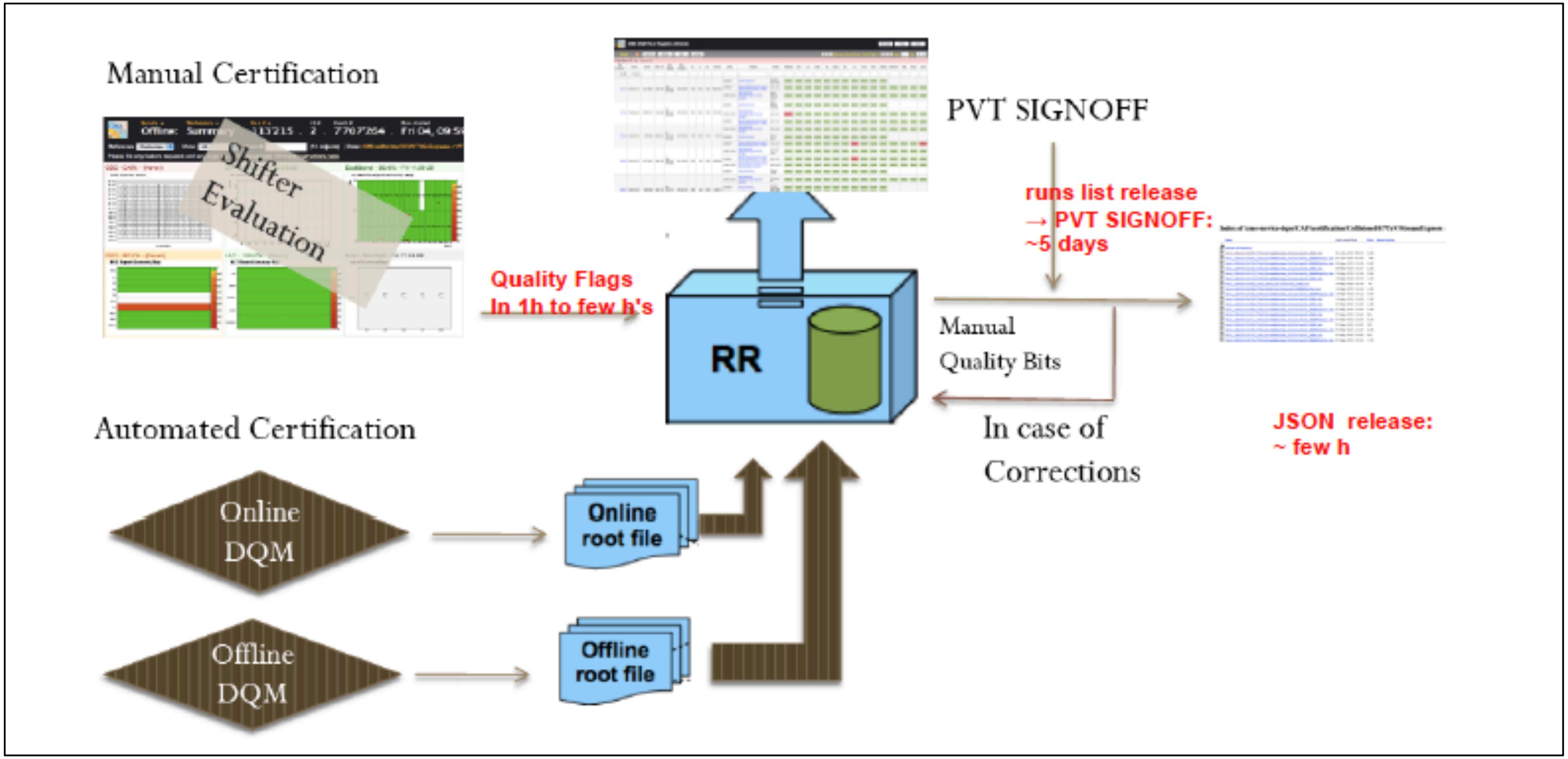
Recoverable

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

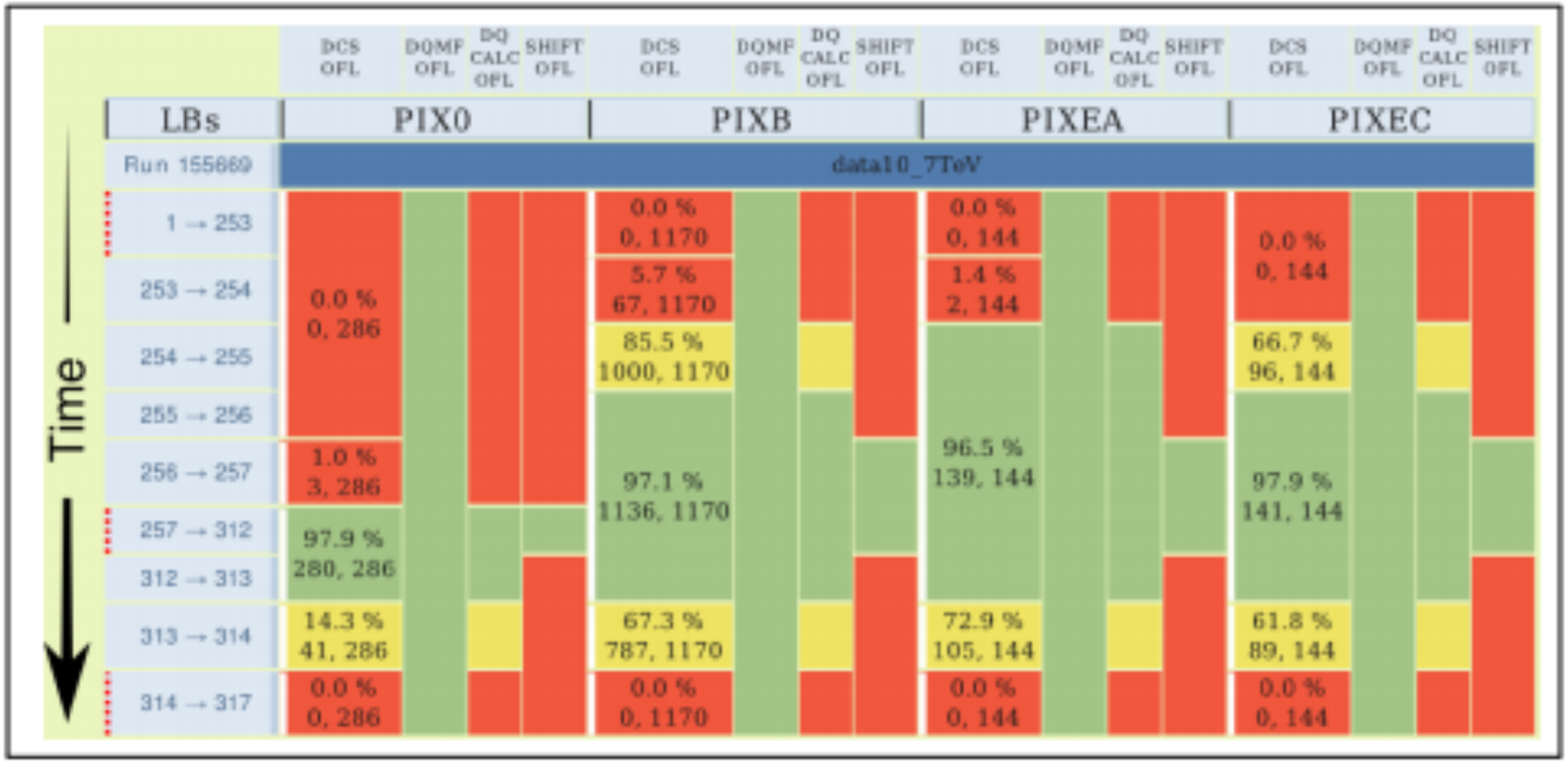
Unrecoverable

Evaluate impact on physics
Improve Reconstruction
Adapt simulation

CMS Data Quality workflow



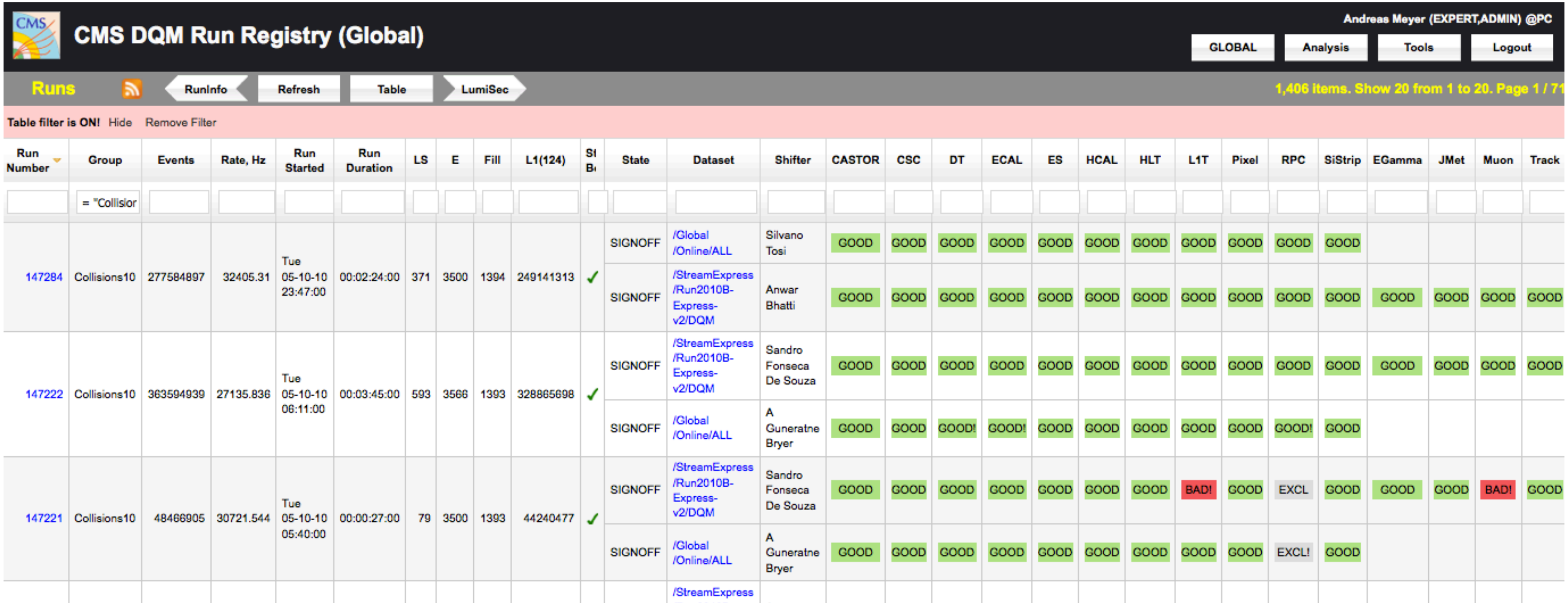
Example DQ luminosity block structure



For the ATLAS pixel regions (layer0, pixel barrel and A & C endcaps) you can see the lumi block structure for good (green), bad (red) DQ. (yellow is still to be decided after expert consultation).

Visualizing DQ results

Important to be able to check the DQ results (sanity checks very important)
e.g. CMS run registry (ATLAS runquery is a similar tool)



CMS DQM Run Registry (Global)

Andreas Meyer (EXPERT,ADMIN) @PC

GLOBAL | Analysis | Tools | Logout

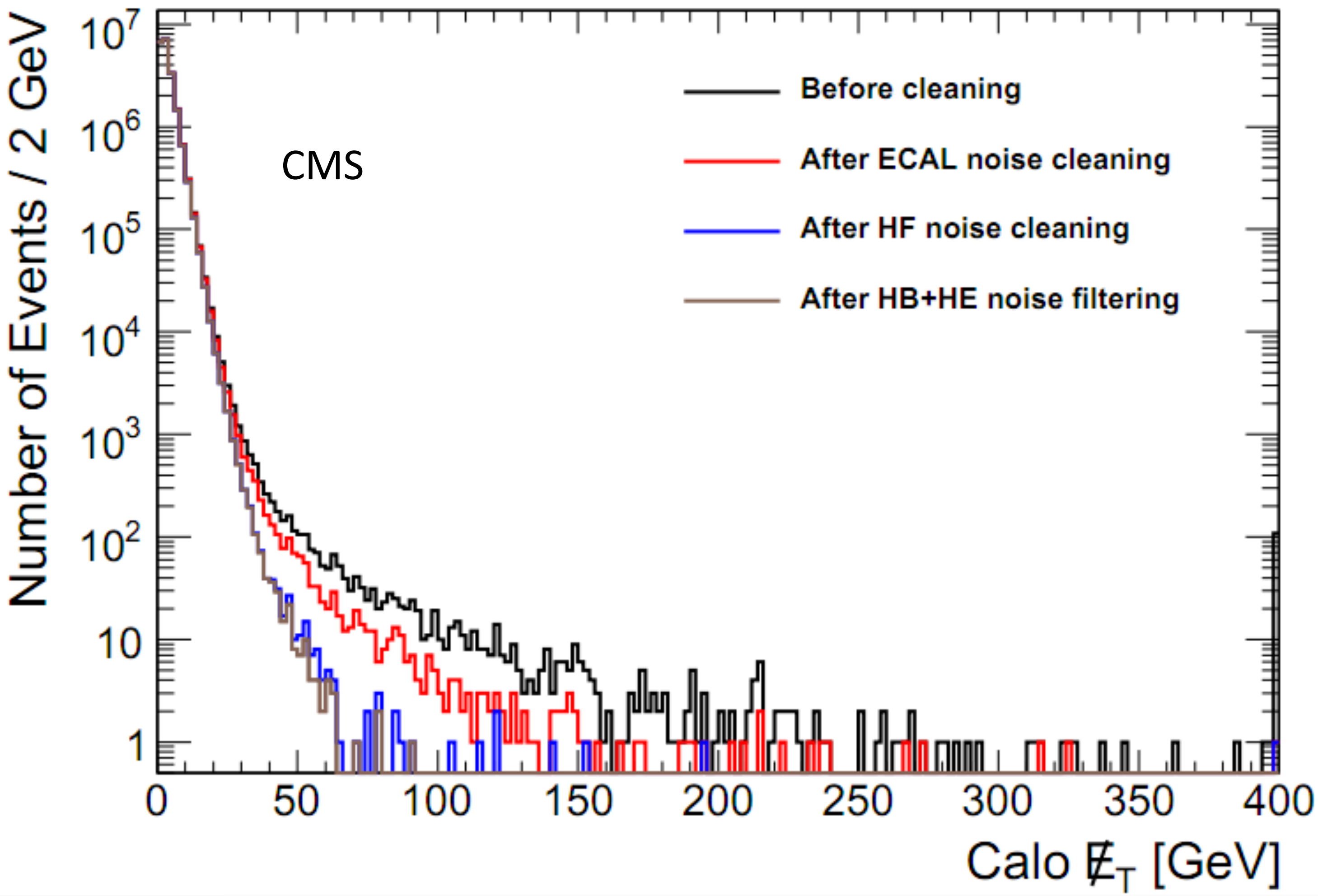
Runs | RunInfo | Refresh | Table | LumiSec

1,406 items. Show 20 from 1 to 20. Page 1 / 71

Table filter is ON! Hide Remove Filter

Run Number	Group	Events	Rate, Hz	Run Started	Run Duration	LS	E	Fill	L1(124)	St B	State	Dataset	Shifter	CASTOR	CSC	DT	ECAL	ES	HCAL	HLT	L1T	Pixel	RPC	SiStrip	EGamma	JMet	Muon	Track
	= "Collisior																											
147284	Collisions10	277584897	32405.31	Tue 05-10-10 23:47:00	00:02:24:00	371	3500	1394	249141313	✓	SIGNOFF	/Global /Online/ALL	Silvano Tosi	GOOD	GOOD	GOOD	GOOD	GOOD	GOOD	GOOD	GOOD	GOOD	GOOD	GOOD				
											SIGNOFF	/StreamExpress /Run2010B-Express-v2/DQM	Anwar Bhatti	GOOD	GOOD	GOOD	GOOD	GOOD	GOOD	GOOD	GOOD	GOOD	GOOD	GOOD	GOOD	GOOD	GOOD	GOOD
147222	Collisions10	363594939	27135.836	Tue 05-10-10 08:11:00	00:03:45:00	593	3566	1393	328865698	✓	SIGNOFF	/StreamExpress /Run2010B-Express-v2/DQM	Sandro Fonseca De Souza	GOOD	GOOD	GOOD	GOOD	GOOD	GOOD	GOOD	GOOD	GOOD	GOOD	GOOD	GOOD	GOOD	GOOD	GOOD
											SIGNOFF	/Global /Online/ALL	A Guneratne Bryer	GOOD	GOOD	GOOD!	GOOD!	GOOD	GOOD	GOOD	GOOD	GOOD	GOOD!	GOOD				
147221	Collisions10	48466905	30721.544	Tue 05-10-10 05:40:00	00:00:27:00	79	3500	1393	44240477	✓	SIGNOFF	/StreamExpress /Run2010B-Express-v2/DQM	Sandro Fonseca De Souza	GOOD	GOOD	GOOD	GOOD	GOOD	GOOD	GOOD	BAD!	GOOD	EXCL	GOOD	GOOD	GOOD	BAD!	GOOD
											SIGNOFF	/Global /Online/ALL	A Guneratne Bryer	GOOD	GOOD	GOOD	GOOD	GOOD	GOOD	GOOD	GOOD	GOOD	EXCL!	GOOD				
												/StreamExpress	A															

MET / Jet cleaning



Also important to reject beam background (halo) events and cosmic ray showers leaving large energy deposit in the calorimeters.

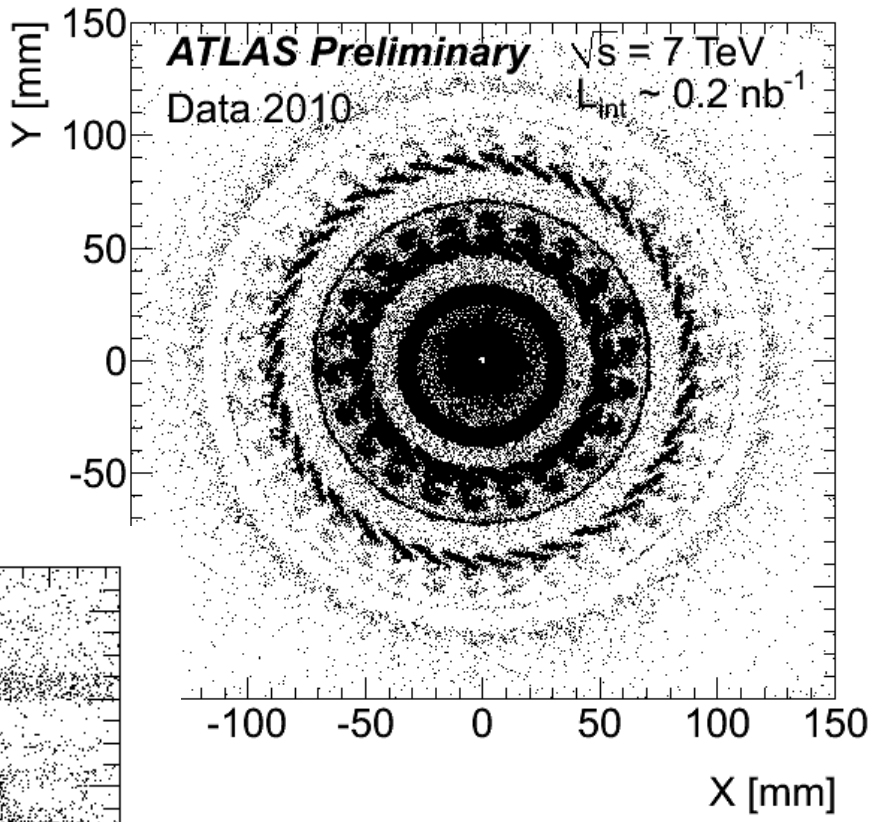
CMS missing energy distribution after various levels of noise cleaning:
Identify & reject anomalous signals based on unphysical charge sharing between neighboring channels in space and/or depth, as well as timing / pulse shape information.
Tail in missing energy clearly reduced by this (otherwise this can look just like 'new physics' like SUSY!)

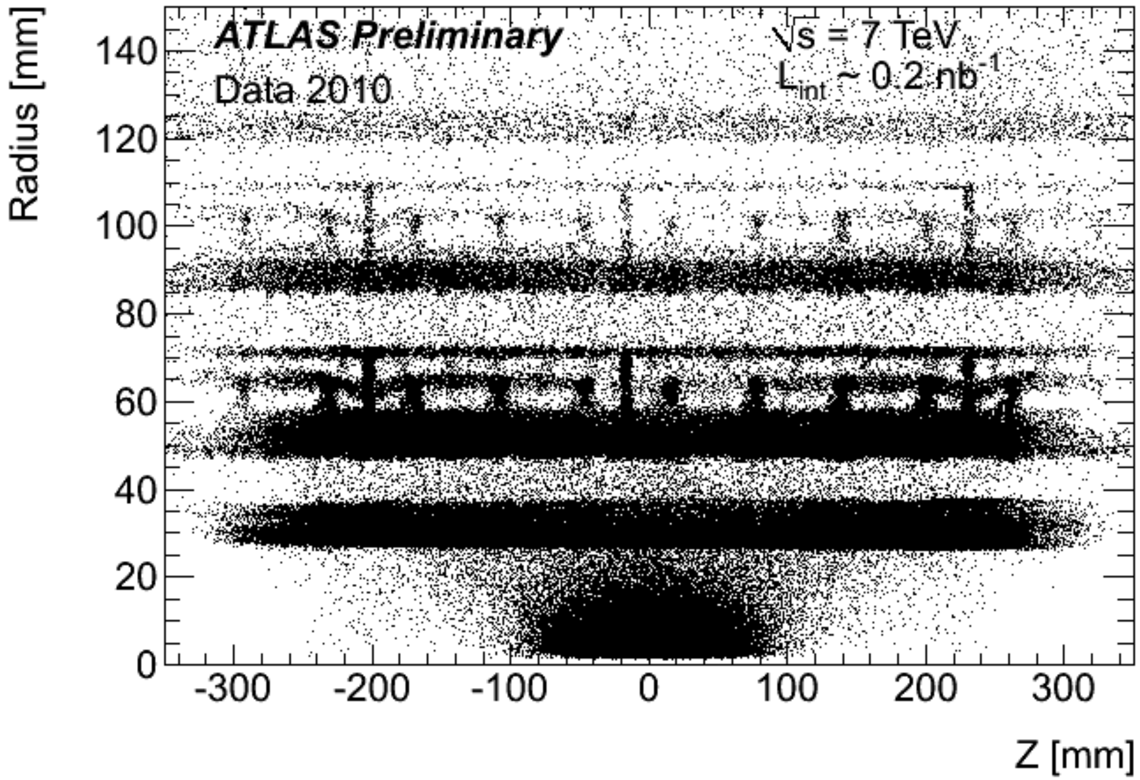
(ATLAS) 36hr calibration loop

- The processing of the physics streams is delayed by 36hrs after the end of the run
- In this time new conditions are derived for the run
  - These are derived either from the express stream processing, or from dedicated processing at calibration centres of the calibration streams
  - Conditions that are updated include
    - Beam spot
    - Dead and noisy channels list (these are used in reconstruction)
    - RT calibrations in gas detectors
- After 36hrs the physics stream reconstruction starts using the updated conditions
  - Occasionally the physics stream processing is delayed to wait for a calibration if there were problems with that calibration procedure
- Data Quality histograms also checked for the physics stream reconstruction (to check that the new conditions are having the desired effect)

Nuclear interactions

- ATLAS example
  - Tracks with d_0>2mm w.r.t PV
  - Form secondary vertices



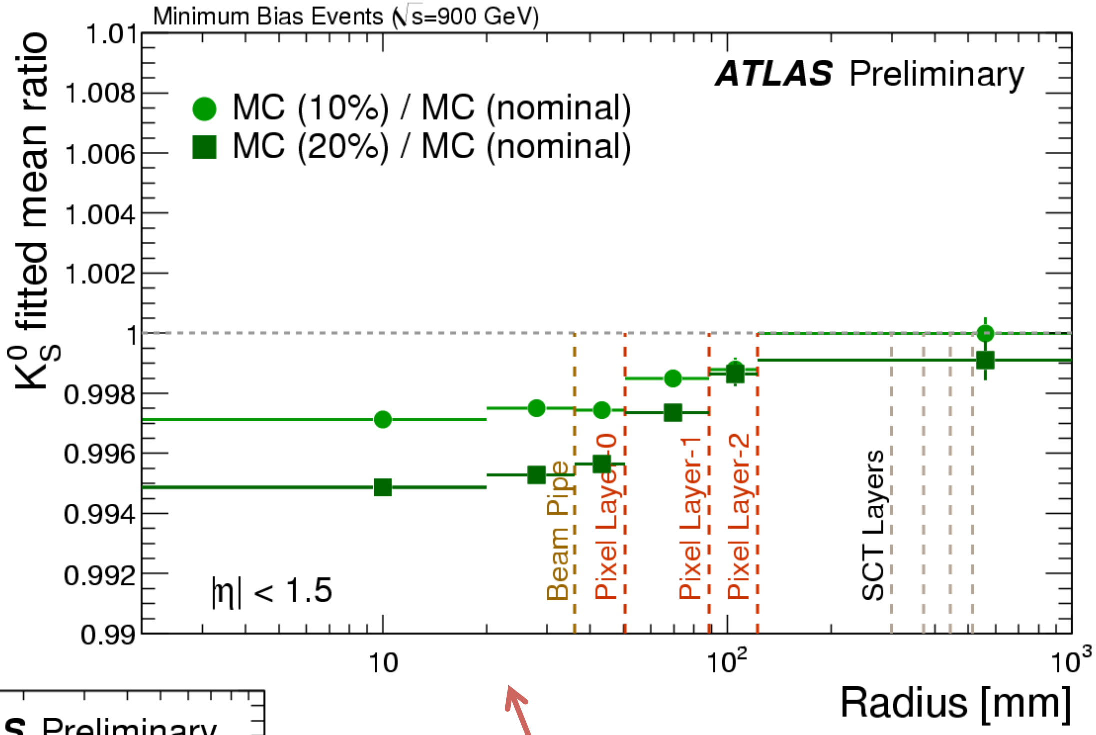


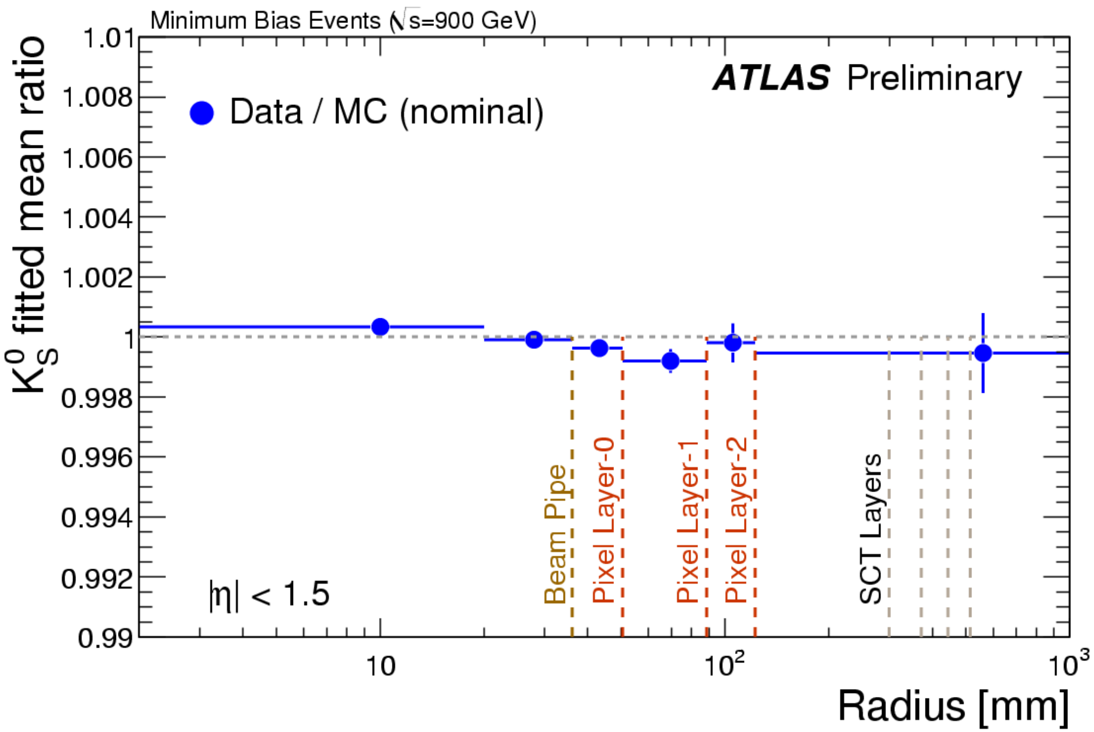
- x-y view for |z|< 300mm
- Sensitive to interaction lengths

ATLAS-CONF-2010-058

K^0_s and material

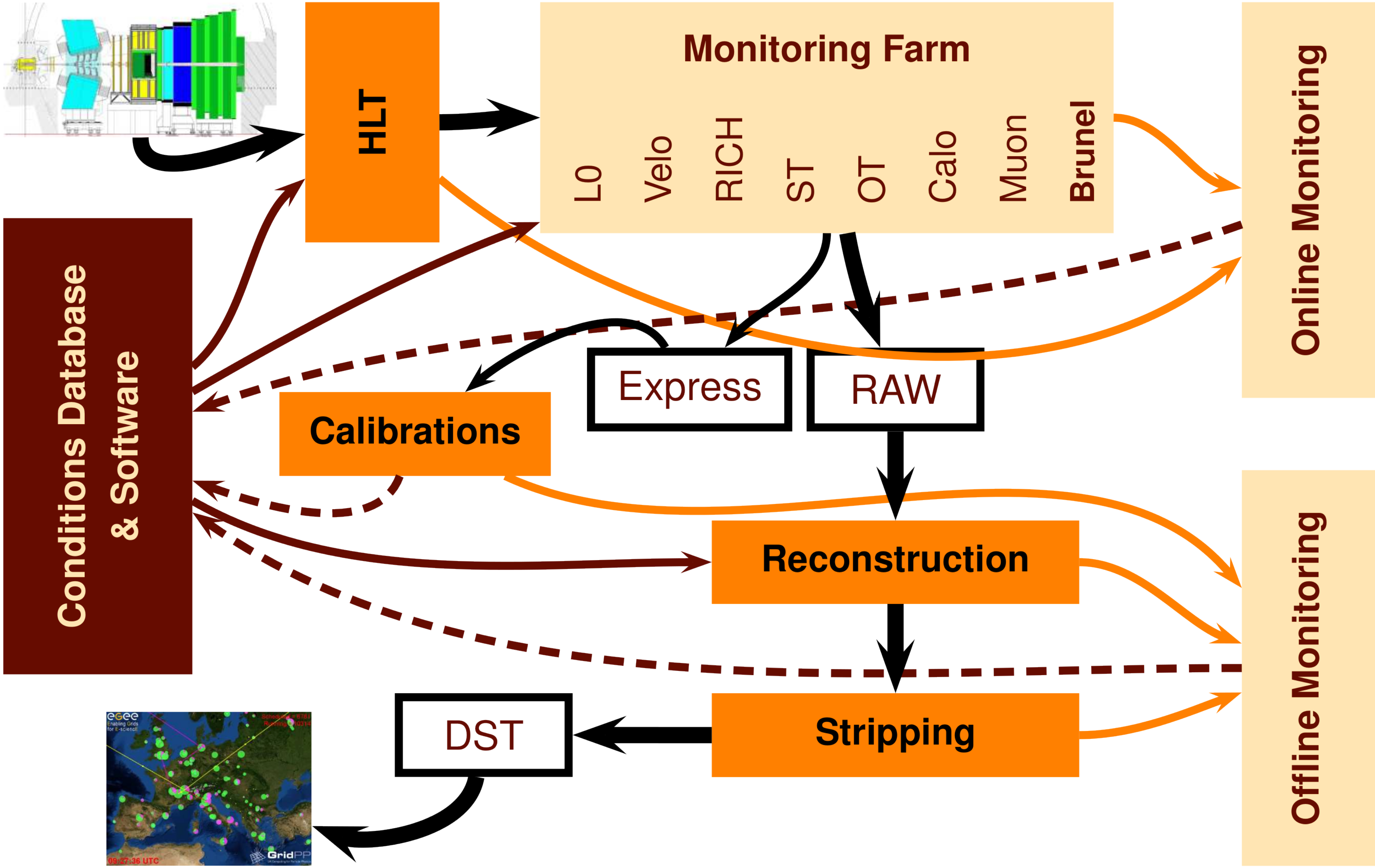
- Look at fitted mass as a function of decay radius
- Data consistent with nominal MC



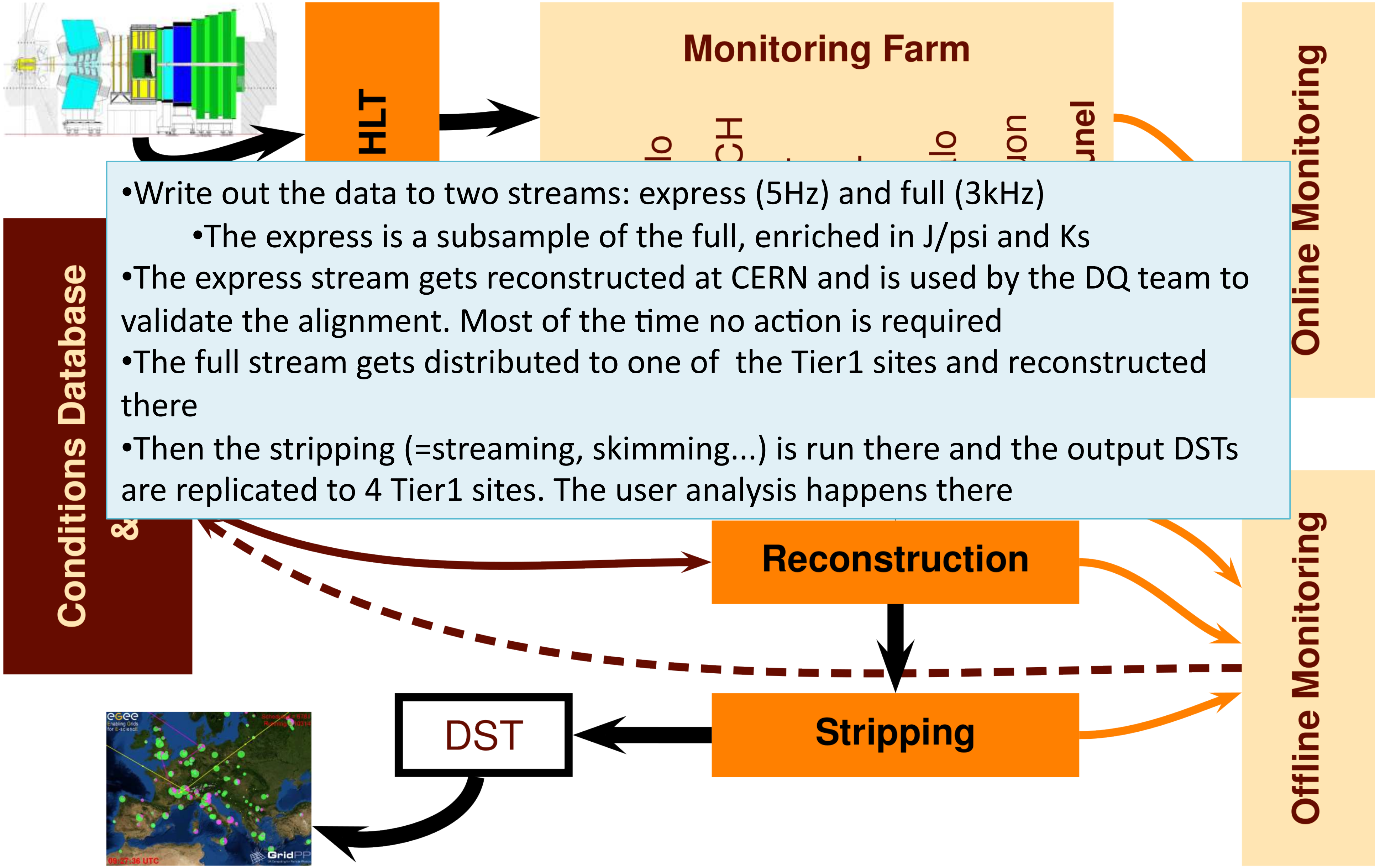


- MC with 10% or 20% extra material predicts much bigger deviations

LHCb workflow…



LHCb workflow…



- Write out the data to two streams: express (5Hz) and full (3kHz)
    - The express is a subsample of the full, enriched in J/psi and Ks
- The express stream gets reconstructed at CERN and is used by the DQ team to validate the alignment. Most of the time no action is required
- The full stream gets distributed to one of the Tier1 sites and reconstructed there
- Then the stripping (=streaming, skimming...) is run there and the output DSTs are replicated to 4 Tier1 sites. The user analysis happens there