XI International Conference
on Calorimetry in High Energy Physics

Calor 2004

DATA QUALITY MONITORING FOR THE D0 CALORIMETER

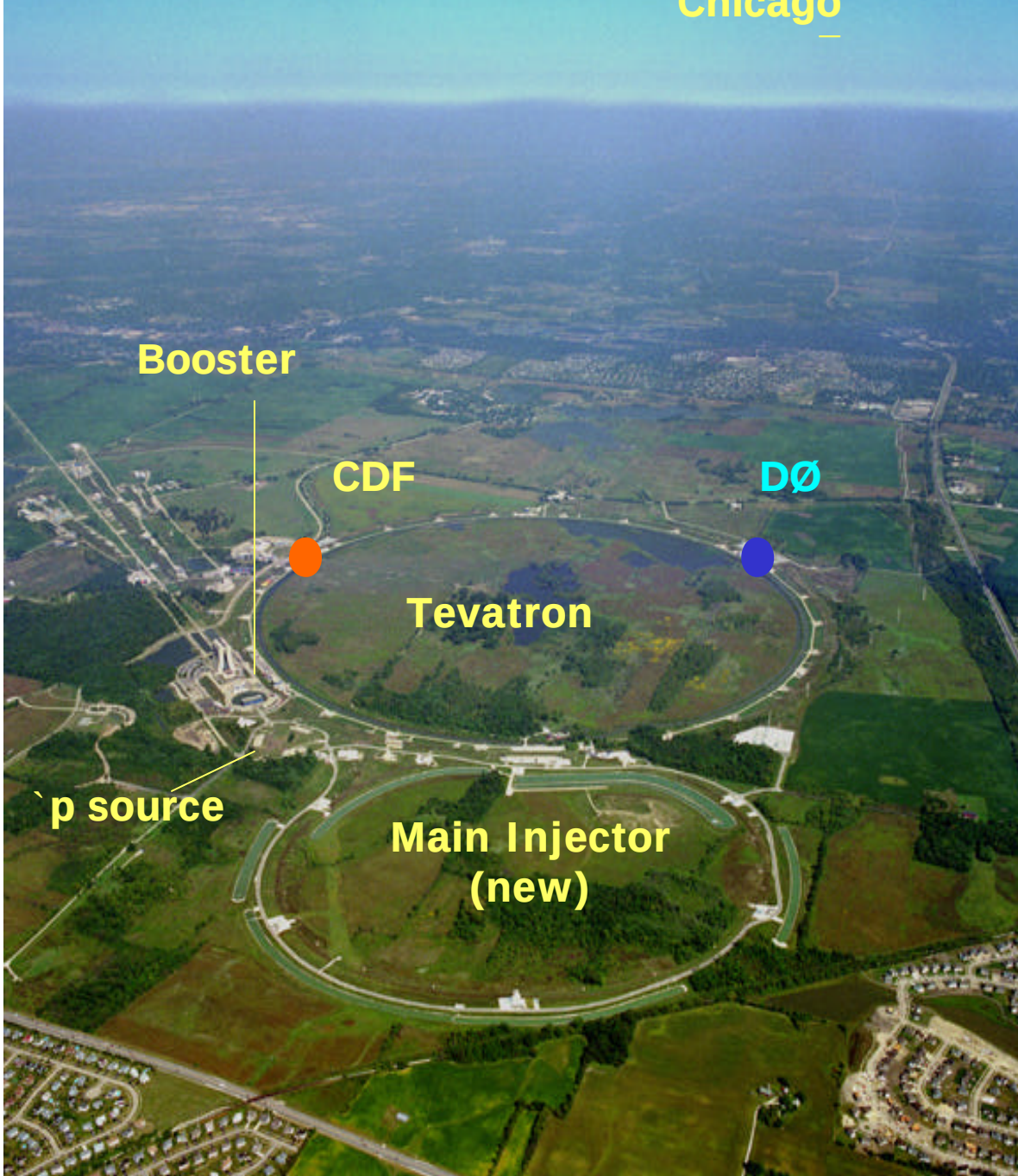
Viacheslav Shary

Laboratoire de l'Accélérateur Linéaire

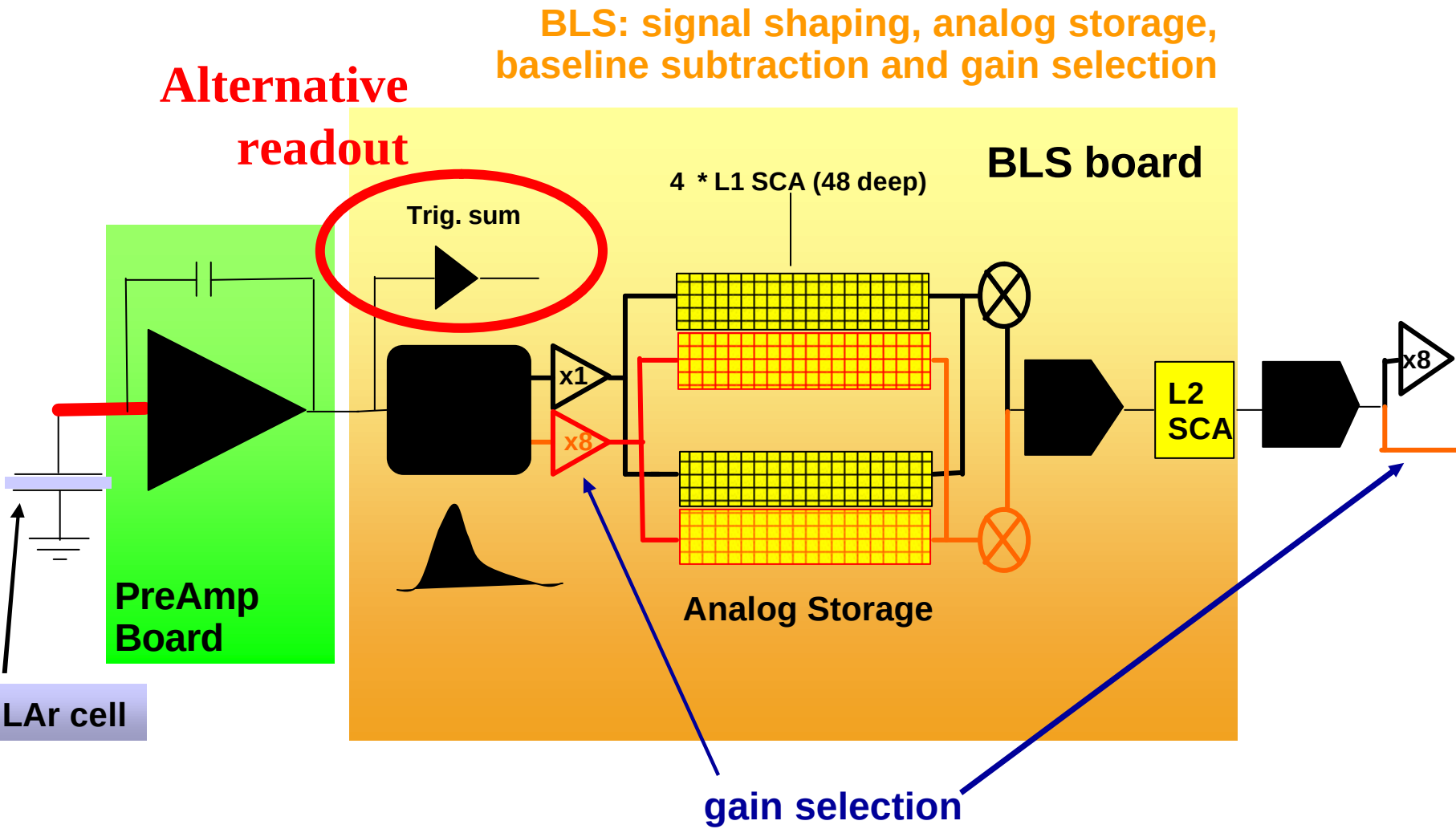
FOR THE D0 COLLABORATION

CONTENTS

- Problems
- DQ strategy
  - Online / offline monitoring
  - Corrections
- Actual examples



ELECTRONICS



SIGNAL

- Baseline subtraction $\Rightarrow$ negative signals
- Hardware pedestal subtraction.
- 1.5σ hardware 0-suppression, 2.5σ offline 0-suppression $\Rightarrow$ reduce number of cells in the event from ~ 50 000 to ~2000.
- Non-linearity correction
- Layer dependent ADC counts $\Rightarrow$ GeV conversion

DATA QUALITY TIMING

- If a problem appears in the run
  - Channel(s) can be masked ⇒ 5 – 15 min of beam time is lost; this run affected; cell(s), tower(s) are lost
  - If need a calibration ⇒ end of store (10 – 20 hours)
  - If need a hardware fix ⇒ request an access (few days)
- If a problem is found off-line
  - Many runs may be affected
  - Need and offline correction for the collected data

DATA QUALITY ISSUES

- Missing cells, towers, BLS.
- “Hot” cells, towers, BLS.
- External noise in calorimeter
- “Coherent noise” in calorimeter
- “Non trivial” calorimeter problems

"HOT" CELLS / TOWERS, MISSING TOWERS

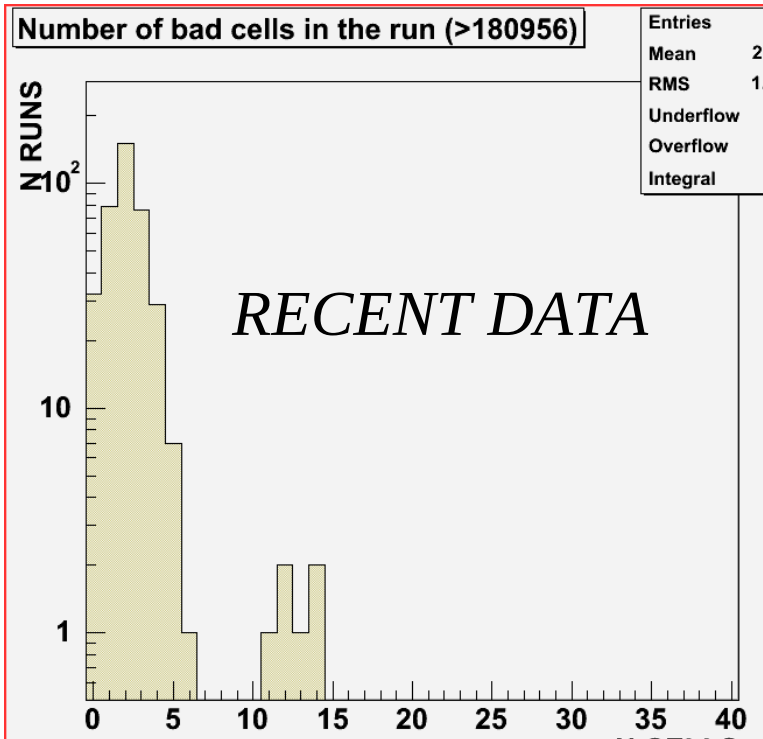
"Hot" cells/towers appear due to the hardware failure or a pedestal drift.

Can affect only one cell or all cells in a tower coherently.

Can produce fake jets, electrons.

To repair need a calibration or hardware replacement.

The permanent online monitoring exists.

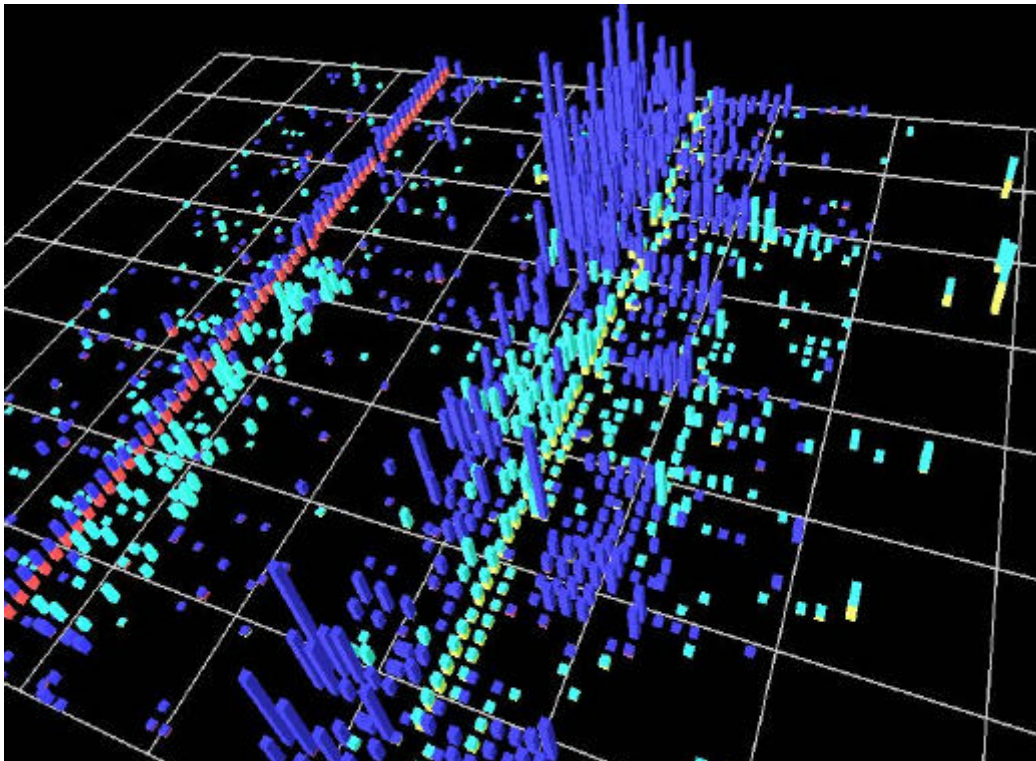


EXTERNAL NOISE IN CALORIMETR

Particularly visible in specific region of calorimeter due to HV distribution layout.

Comes in burst. One needs to identify it and remove a "luminosity block".

This type of noise was seen often at the early stage of experiment. The noise path has been found and fixed.

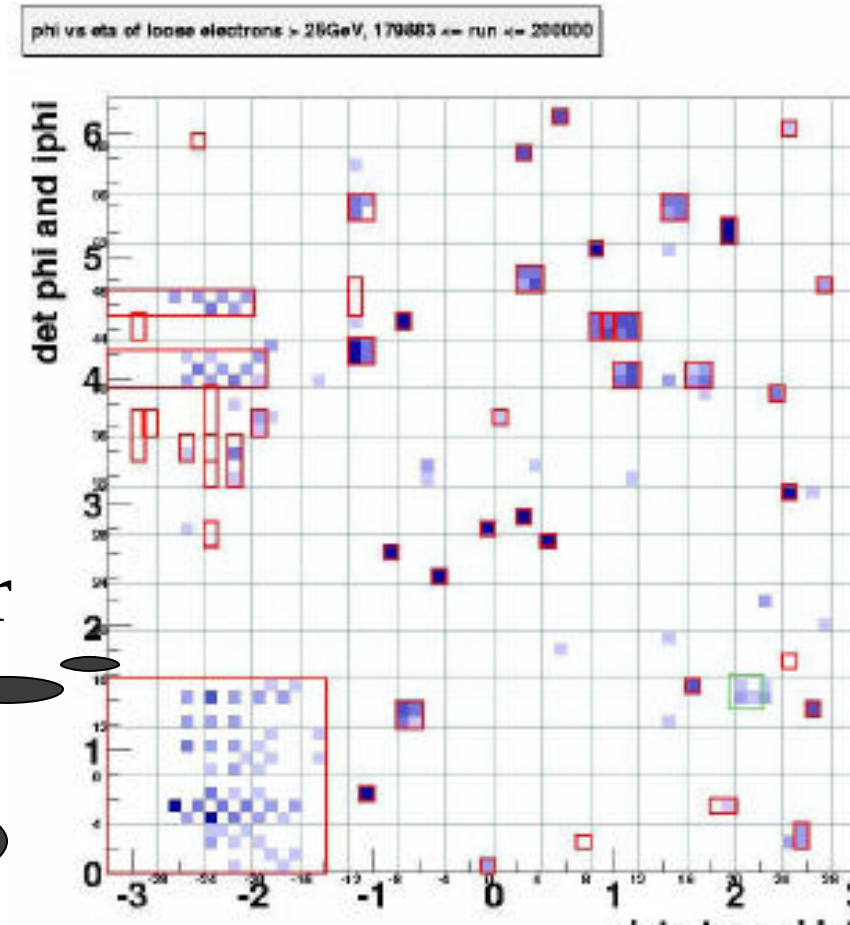


luminosity block" – the smallest part for the luminosity calculation. Duration near 1 min, 2500 events in average, luminosity ~1 nb^{-1}

“NON TRIVIAL” PROBLEMS

These problems were detected for the first time in analysis, mainly in the high energy electron distributions.

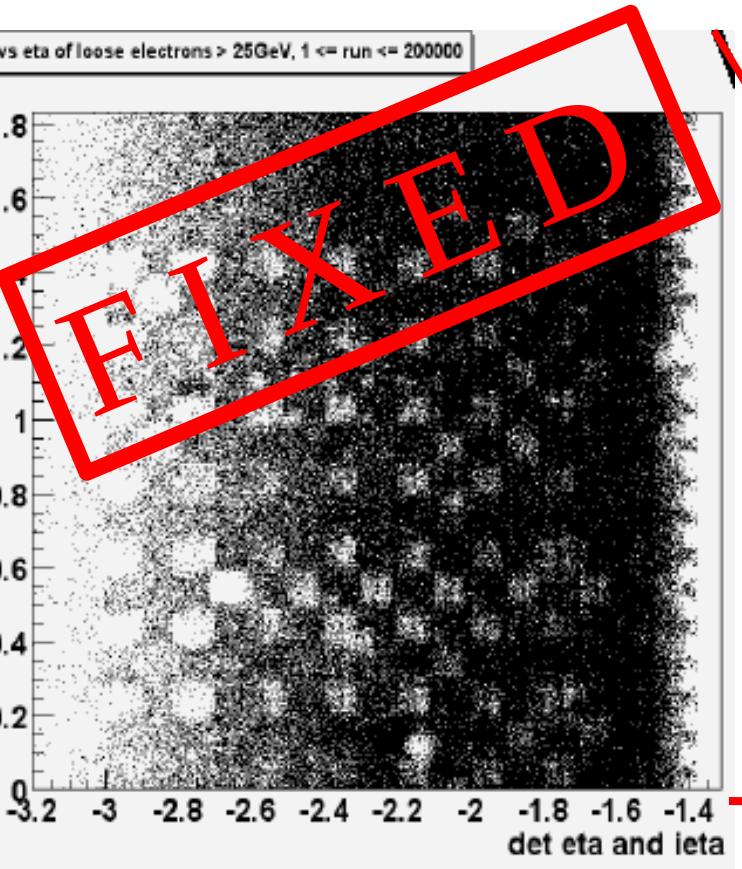
They was corrected in hardware. The software correction exists for the collected data and is being used for the reprocessing of the full dataset to be finished before the summer conferences.



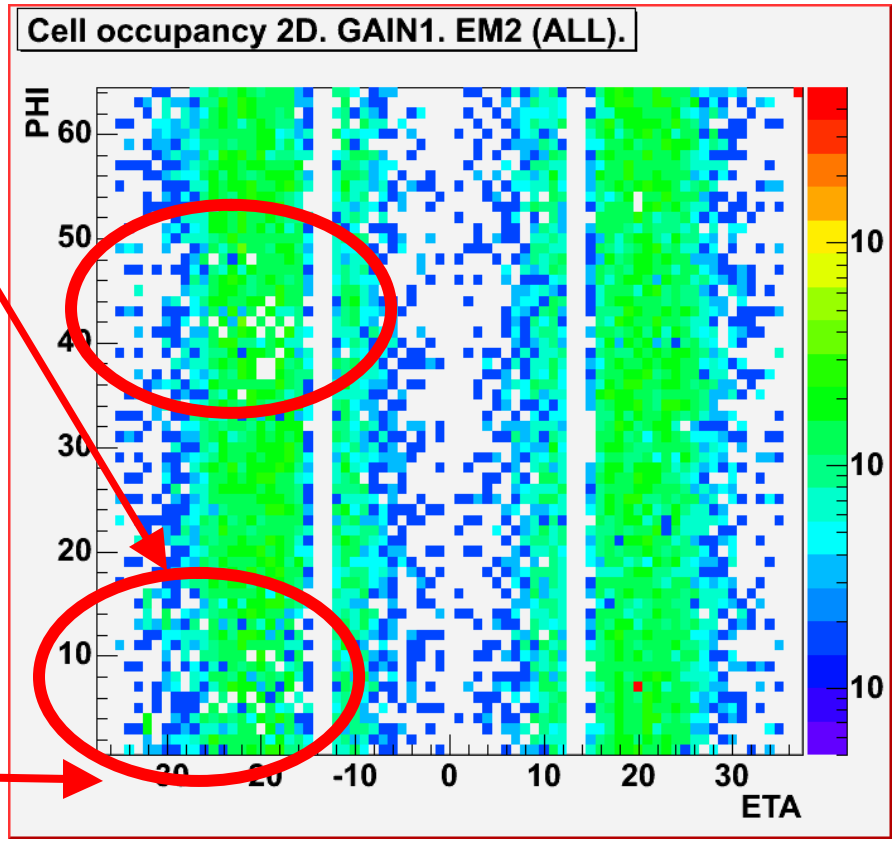
All these problems were detected and fixed

EXAMPLE OF THE FEEDBACK FROM OFFLINE TO ONLINE

HI-ETA DISTRIBUTION FOR THE LECTRON WITH E_T>25 GEV



CELL OCCUPANCY MEASURED ONLINE FOR THE SIGNALS WITH GAIN 1



mall effect on the physics since it is affecting high energy electrons t high eta, not yet used in the analysis

STRATEGY OVERVIEW

- **To detect problems**
  - Online monitoring with integrated data quality controls
  - Offline monitoring using reference histograms
  - Event quality check online / offline ⇒ event quality flag.
  - Physical analysis.
- **To correct problems**
  - Hardware repairing online
  - Algorithms to remove isolated “hot cells”
  - Offline corrections
  - The event quality flag is used in analysis.
  - Good run / luminosity blocks selections

ONLINE CALORIMETER MONITORING

ROOT based monitoring program. ~10% of recorded events are used

Different types of problems affect different triggers $\Rightarrow$ 4 different streams in use.

Control the channels signals and missing E_T quantities.

Automatic data quality control

- ✓ Missed towers
- ✓ "Hot" cell / towers
- ✓ Noise measurements
- ✓ The comparison with L1 readout very powerful for detecting "non trivial" problems (work in progress)

The quality of the run based on listed above quality checks is reported. This quality is used in analysis to remove bad runs (runs with severe hardware problems, 3% of the data so far, getting smaller)

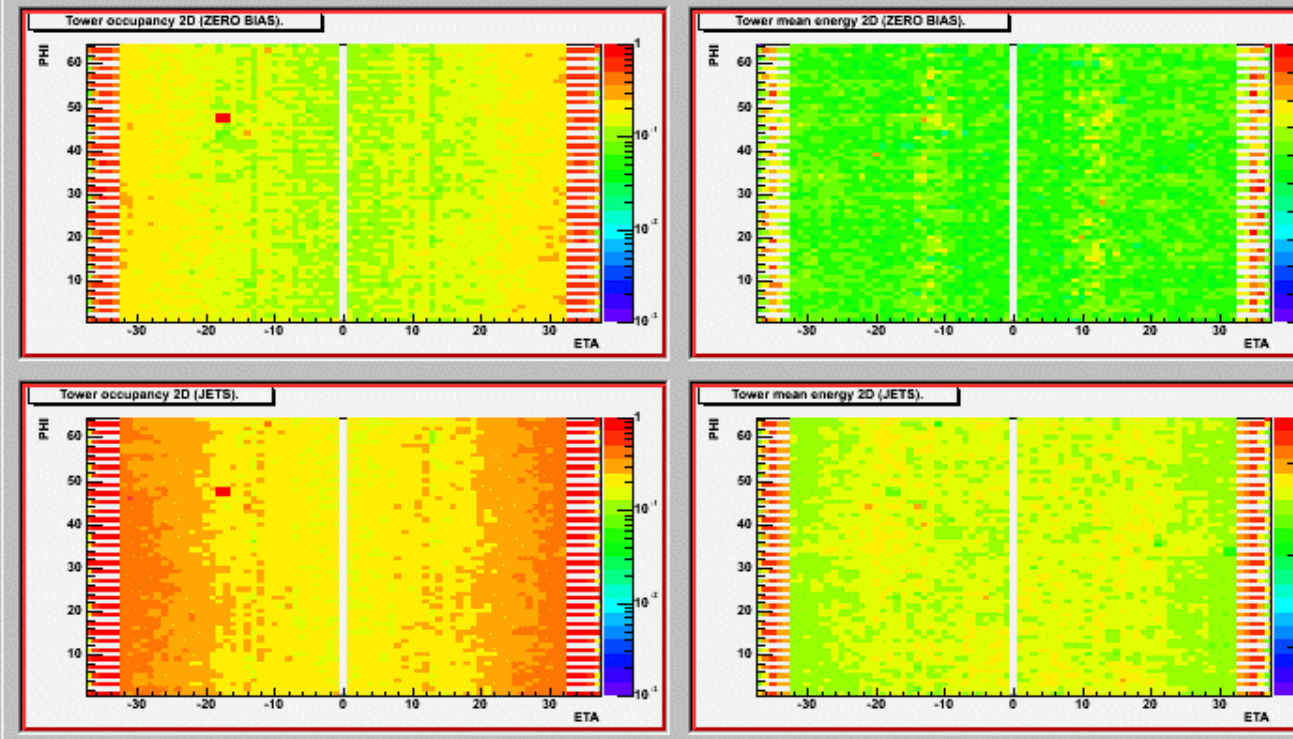
Mode Options Runs Html

X x Y: 2x2 INFO REPO

Last (current) from: dq_calo_190369.root to: dq_calo_190369.root Reload

SHIFT
- Tower occupancy 2D (ZERO BIAS).
- Tower mean energy 2D (ZERO BIAS).
- Tower occupancy 2D (JETS).
- Tower mean energy 2D (JETS).
- METX: no CH (ZERO BIAS)
- METY: no CH (ZERO BIAS)
- METX: no CH (JETS)
- METY: no CH (JETS)
- L1 EM tower occupancy 2D (JETS, THRESHOLD 3 GeV
- L1 HAD tower occupancy 2D (JETS, THRESHOLD 3 Ge
- L1 EM tower occupancy 2D (ALL, THRESHOLD 3 GeV
- L1 HAD tower occupancy 2D (ALL, THRESHOLD 3 GeV

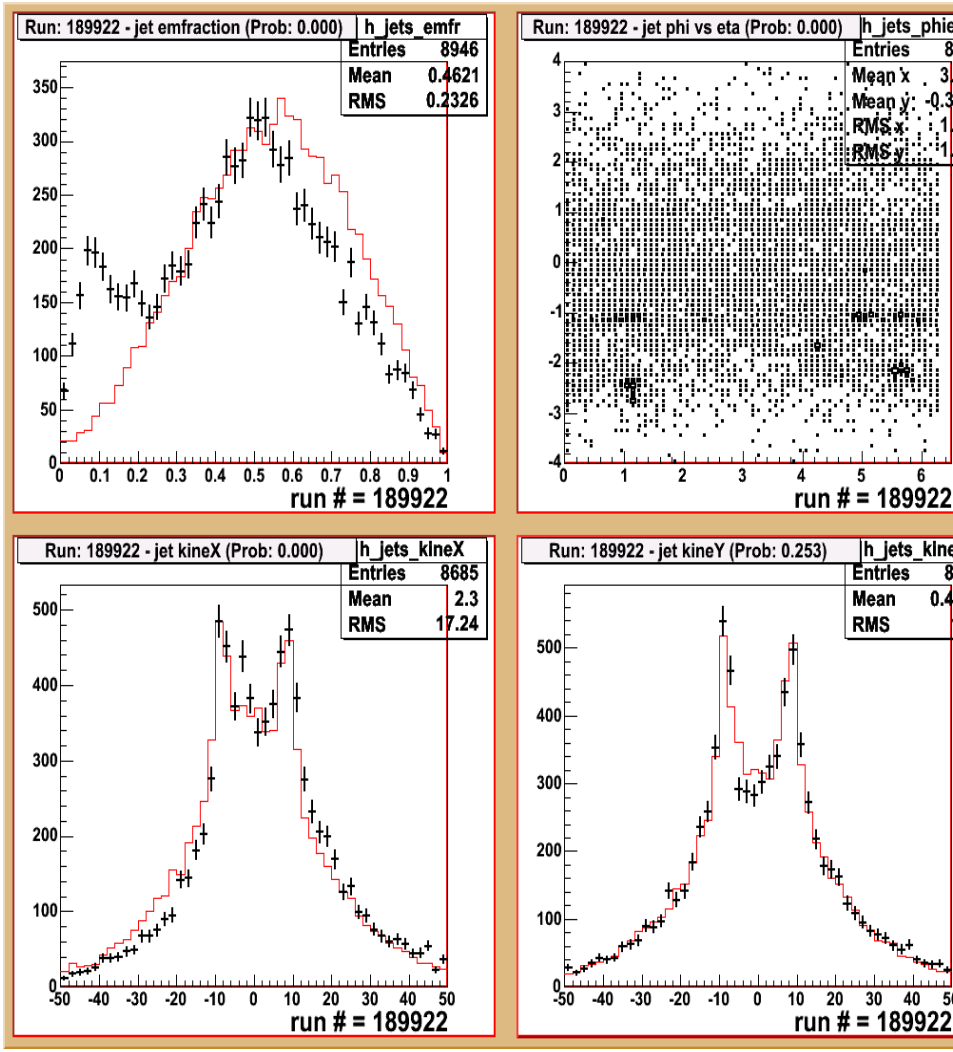
EXPERT
- ZERO BIAS
- UNSUPPRESSED
- JETS
- ALL




```
BER OF "JETS" EVENTS:          7946
BER OF "ALL OTHER" EVENTS:     19685
-------------------------------------------------------------
mpty towers
ese towers are:
ta,iphi): (14, 59) CRATE/ADC/BLS/TOWER: 6/4/3/2
-------------------------------------------------------------
0% of noisy events in total:
0% of "ring of fire",  3.15% of "coherent noise",  0.05% of "missing crates"
0_BIAS:    3.18% of noisy events:
0% of "ring of fire", 2.89% of "coherent noise", 0.29% of "missing crates"
S:         3.96% of noisy events:
0% of "ring of fire", 3.96% of "coherent noise", 0.00% of "missing crates"
 OTHER:    3.01% of noisy events:
0% of "ring of fire", 2.96% of "coherent noise", 0.05% of "missing crates"
-------------------------------------------------------------
bad" cells
ese cells are:
```

“GLOBAL MONITORING”

- We also monitor data online using L1, L2, L3 and reconstructed objects (jets, electrons, muons). This allows to find the problems not visible with detector quantities.
- We monitor data offline using the reconstructed objects. Results are available 2 – 3 days after data taking.
- This kind of monitoring help understand the impact on physics comparing the histograms with reference set.



EVENT QUALITY

- Special package is designed to detect external noise, “coherent noise” and other similar calorimeter problems for each event.
- It is used online in the run selection.
- It is used offline for the good luminosity blocks selection and in the analysis.
- The typical fraction of events marked is a few percent.

ALGORITHMS

- To remove isolated hot cells (because of electronics or uranium noise): “NADA” algorithm compares cell to its neighbors.
- To reduce influence of noise and improve resolution in missing E_T: “T42” algorithm
  - remove all cells with negative signals
  - select a cell with high signal ($> 4\sigma$) and keep their neighbors ($2.5 – 4\ \sigma$).

OFFLINE CORRECTION

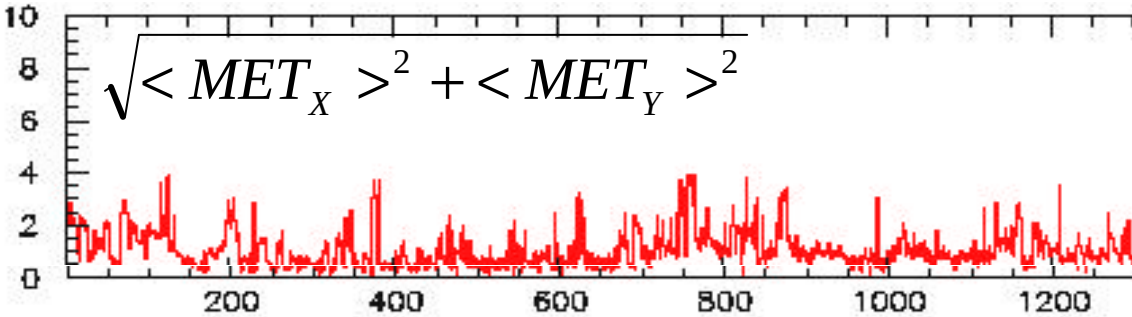
- Subtle problems are found offline and affected many runs, need to go back and reprocess in some way to improve the data quality.
- The trigger readout plays a key role. This alternative readout usually is not affected by the same problem, so it can be used to determine the correction.

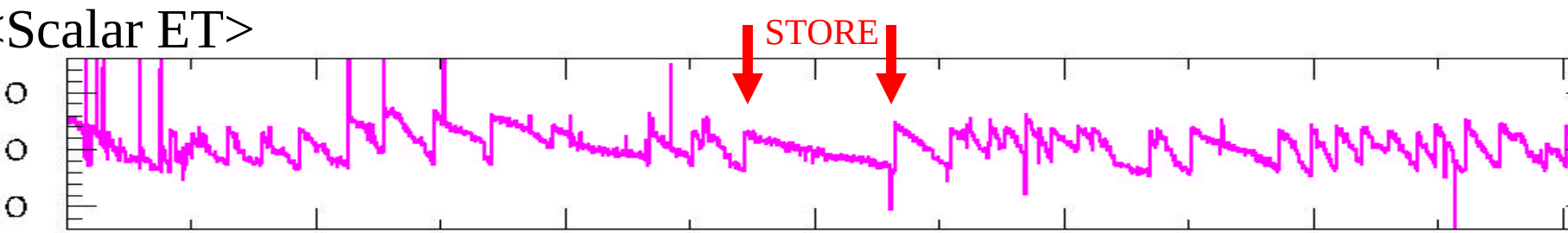
RUN / LUM BLOCKS SELECTION

Not all problems can be corrected, so the selection of good data is needed.

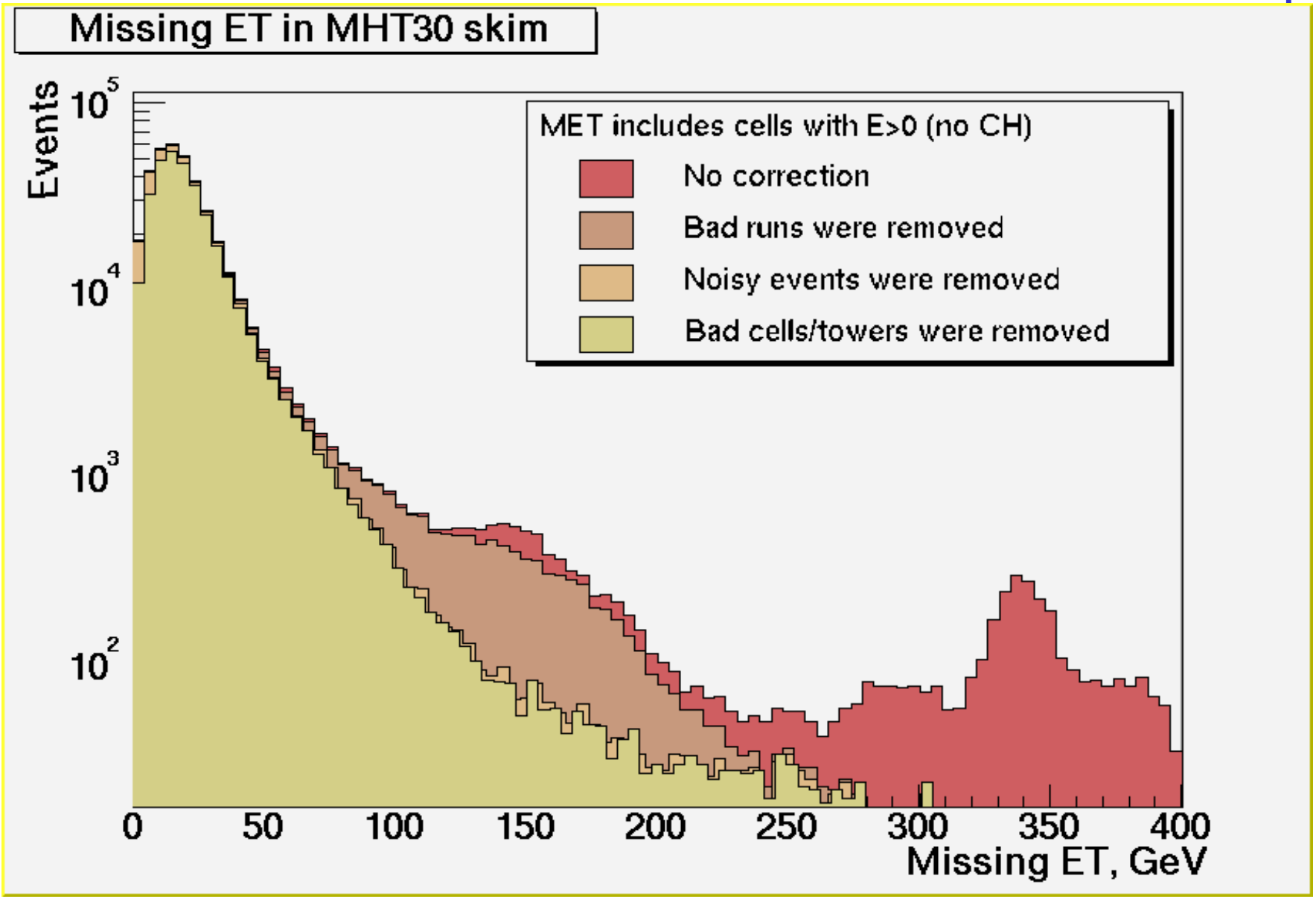
- Severe hardware problems: need good run selection
- Noise problems: need luminosity block selection
- Unknown problem: can be controlled using missing E_T variables (projection on X and Y axes).
- In recent data 5% have been marked bad, of which 2% for not yet studied reasons.

Examples of parameters are used for data selection versus run number



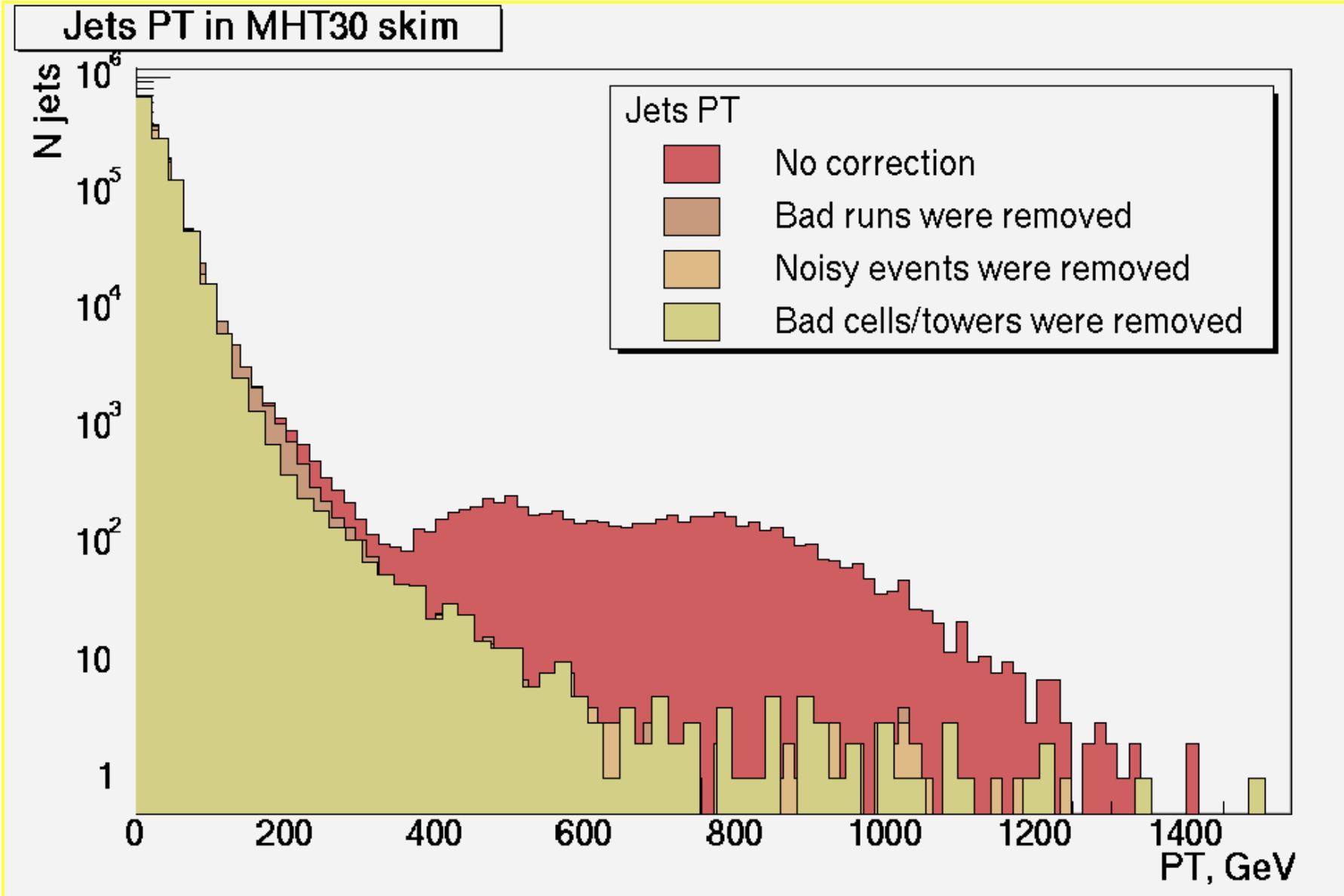


PHYSICAL EXAMPLES: MISSING E_T



problems are here enhanced by the MET trigger selection

PHYSICAL EXAMPLES: JETS P_T



problems are here enhanced by the MET trigger selection

D0 CALORIMETER DATA QUALITY STATUS

- Less than 5% percent of the recent data is marked as “bad” due to calorimeter problems.
- The “bad” channels reduce the efficiency by less than 0.5%.
- But precision physics and new phenomena need a perfect treatment of the known calorimeter problems.

“PESSIMISTIC” CONCLUSION

- Origin of the problems:
  - Modern setups have sophisticated electronics with big number of channels and … hardware failures. Better electronics can reduce number of failures, but I do not believe that one can have failure free setup.
  - Bugs in the software and in the firmware. Difficult to imagine software without bugs.
  - Noise.
- The tails of missing E_T and jets P_T are extremely sensitive to the data quality. To study new physics a good data quality strategy should be applied.

“OPTIMISTIC” CONCLUSION

- You can not expect that shifter will notice all problems, hence automatic data quality check needed online.
- Some problems can be seen only with specific triggers. Need to control separately different set of triggers.
- Comparison between independent trigger readout and precision readout is very useful to detect problems. Trigger readout also can be used in corrections.
- A good data quality strategy including the online / offline monitoring, noise removing, good data selection will assure the quality of the physics results.