



NOvA Data Quality Monitoring Framework

Jim Musser



DQ Organization

- Support for data quality related activities is provided by the Data Quality Working Group (DQG), which reports formally to the Run Coordinator, and informally to the Analysis Coordinator.
 - The DQG is currently co-convened by Mat Muether and Jim Musser.
- The DQG maintains a regular bi-weekly meeting schedule.
- A subset of the DQG, the Watchdog Group, meets weekly to review data quality metrics in detail, and provides a summary report at the DQG meeting. Members of this group are expected to scan DQ monitoring tools on at least a daily basis, and to keep up with operational activities. They provide an expert backup to the continuous data monitoring provided by Shifters.



DQ Group Deliverables

- The DQG developed and maintains the data monitoring tools used by the collaboration, in particular, shifters, and by the DQG itself.
- The DQG provides a weekly report to the Run Coordinators in support of maintenance activities.
- The DQG develops criteria for minimal acceptable data quality for analysis by sub-run and provides a list of sub-runs meeting those standards to the Production Group.



Data Quality Monitoring Tools

- Six principal tools are used to monitor data quality:
 1. **Online Monitoring** : Immediate monitoring of low level quantities
 2. **Nearline Monitoring**: Nearly immediate monitoring of higher level quantities.
 3. **Hardware Watch**: Tracks performance of front end hardware components, providing a maintenance list.
 4. **Time Server Monitor**: Monitors the state of the timing system.
 5. **File Transfer Checks**: validates data integrity through file transfers.
 6. **Offline Production Monitoring**: Validation of final data products.

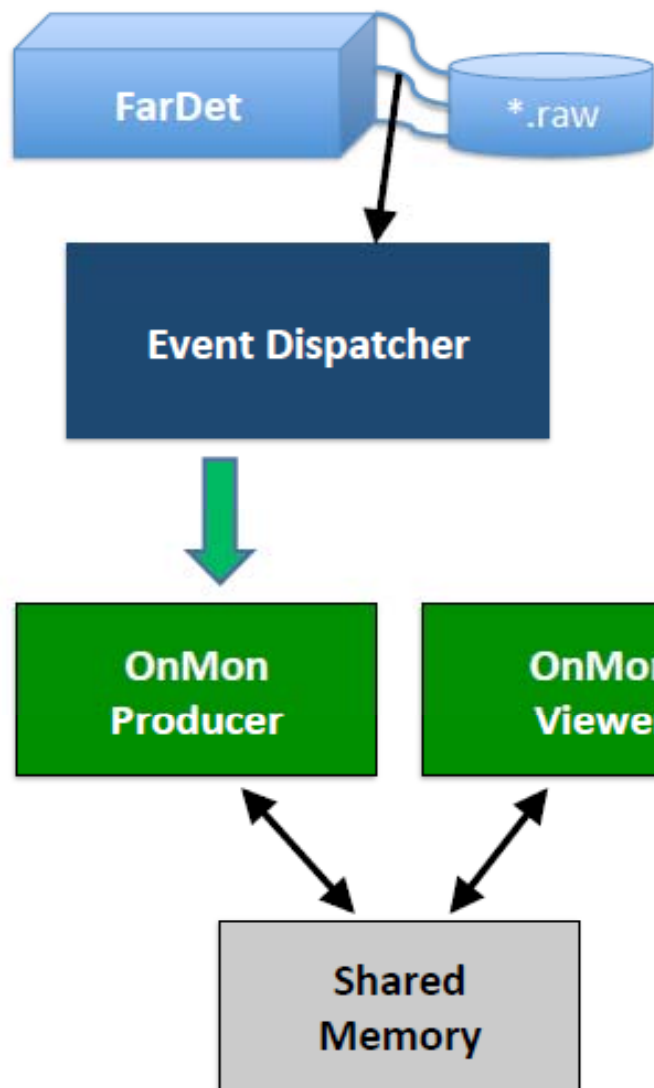


Online Monitoring

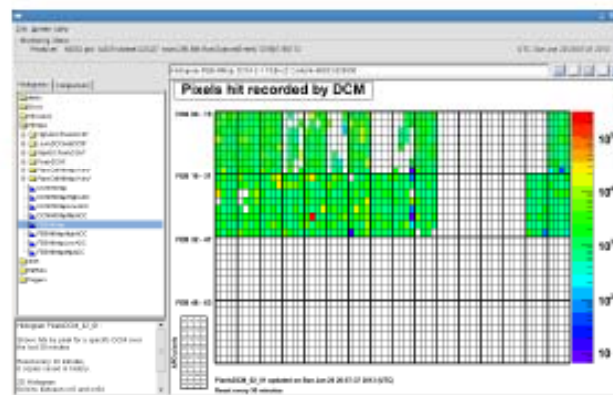
- Latency: ~seconds, continuous update.
- Time Period Covered: Sub-run.
- Tools: EVD, OnMon, data processed @ Ash River (FD).
- Quantities Monitored: pre-reconstruction cell-level rates, adc/PE/tdc distributions...
- Primary System Components Validated:
 1. DAQ functionality.
 2. Front end electronics/sensor functionality.
 3. Configuration (gain, thresholds, channel masking) .



OnMon Architecture



- The event dispatcher takes raw data as it flows out of a detector and serves it to clients.
- The OnMon producer is a nova job that makes histograms out of raw data.
- The OnMon viewer is a root GUI that displays the histograms made by the producer.

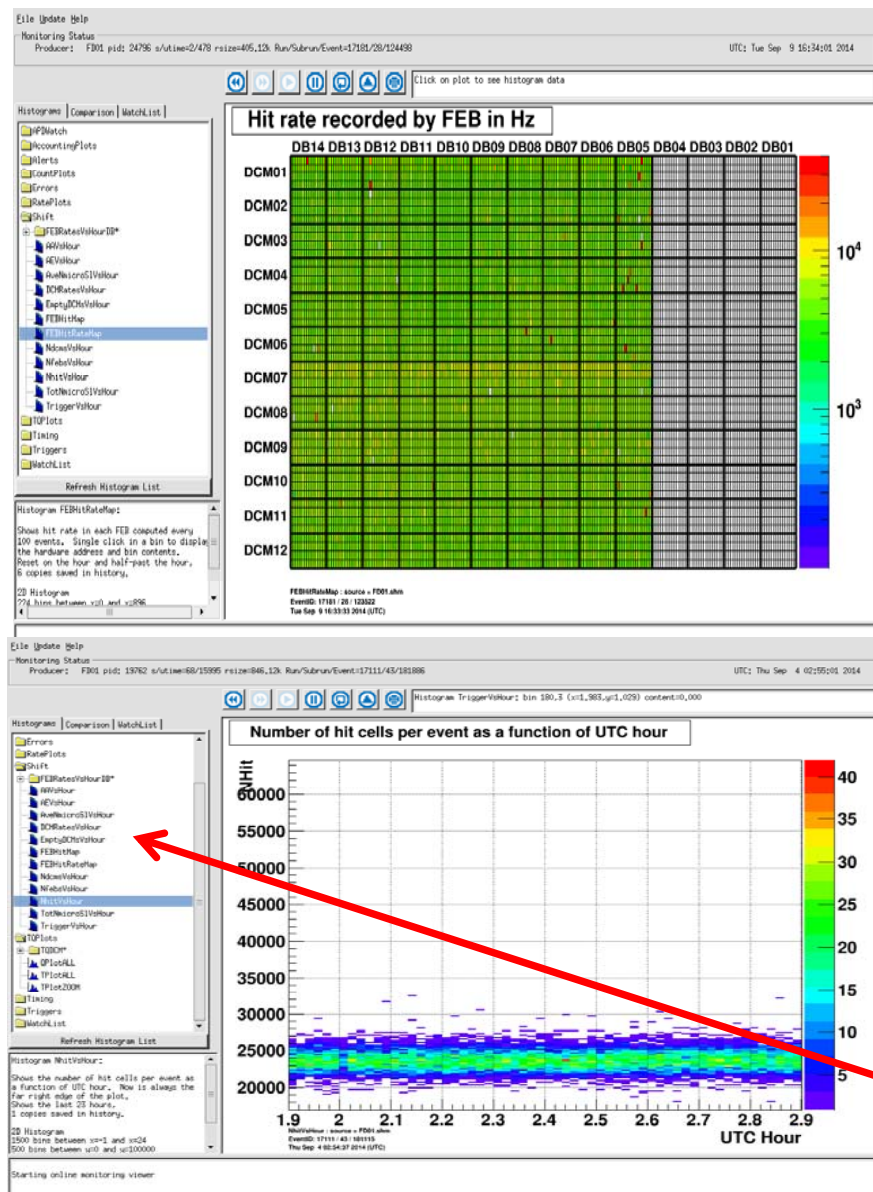




OnMon Outputs

Plot types:

- **Hit Maps:** shows total # of hits recorded at the various levels of detector granularity, mapped to hardware coordinates.
- **TQ Plots:** Shows time dependence of quantities such as rates, average adc by pixel, ...
- **Errors/Alerts:** Shows number of errors by type vs time and location.



Shifter Plot Checklist



Nearline Monitoring

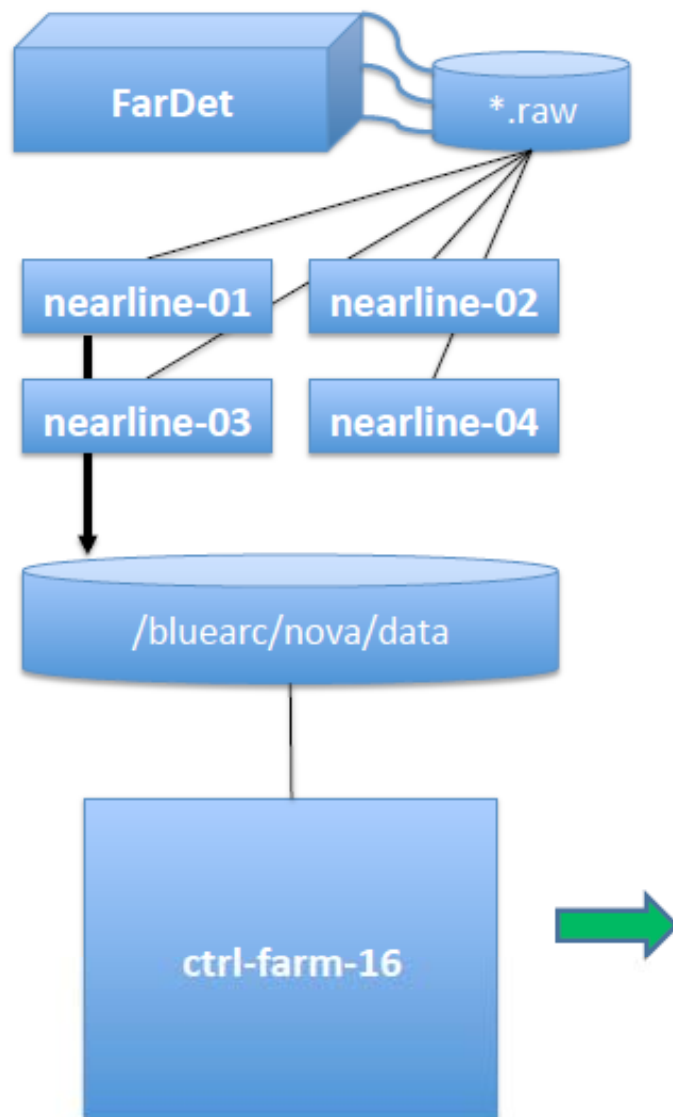
- Latency: ~30-60 min.
- Time Period Covered: daily, weekly, monthly
- Tools: Nearline, data processing @ Ash River (FD)
- Monitored: pre-reconstruction cell-level rates, adc/PE/tdc distributions...
- Web based: easily and universally available.

<http://nusoft.fnal.gov/nova/datacheck/nearline//nearlineFD.html>

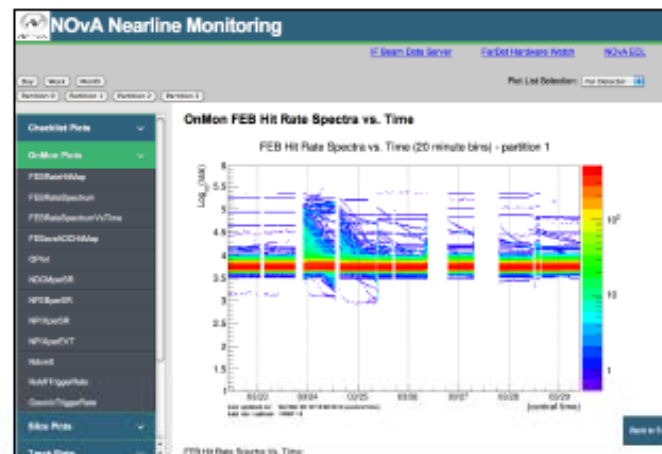
- Primary System Components Validated:
 1. All components validated by OnMon, plus....
 2. Low level reconstruction performance
 1. Slice count, size,... per trigger
 2. Tracking efficiency
 3. Fraction of data passing good run selection.
 4. DCM-level timing synchronization.



Nearline Architecture



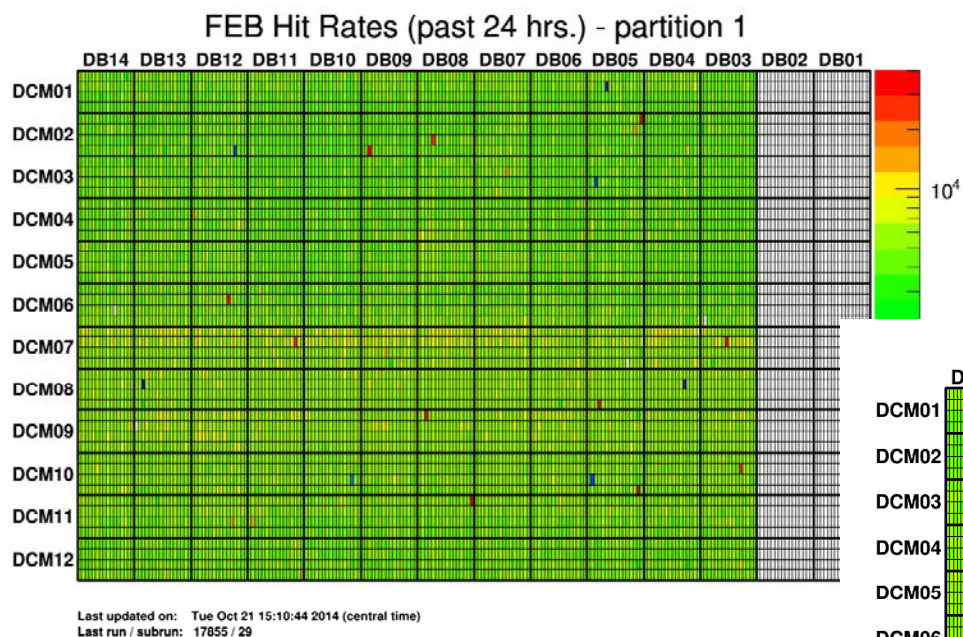
- Four machines process raw data after it is written to disk. They produce output files with both OnMon and Reco information.
- This data is transferred to FNAL.
- A machine at FNAL makes plots out of that data and posts them to the nearline website



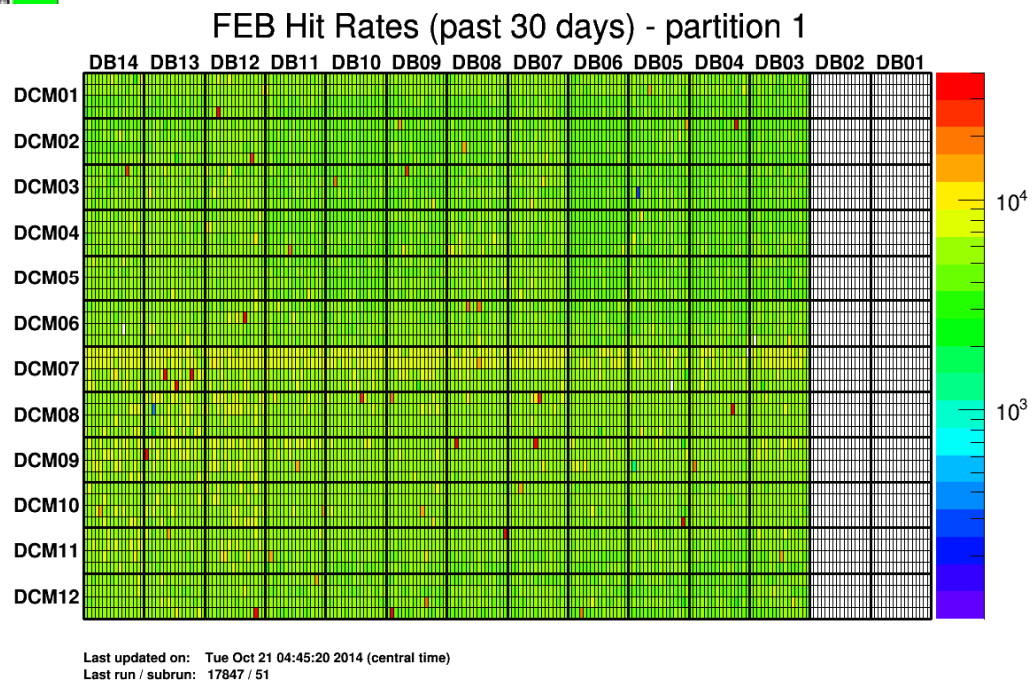


Nearline Monitor

- <http://nusoft.fnal.gov/nova/datacheck/nearline//nearlineFD.html>



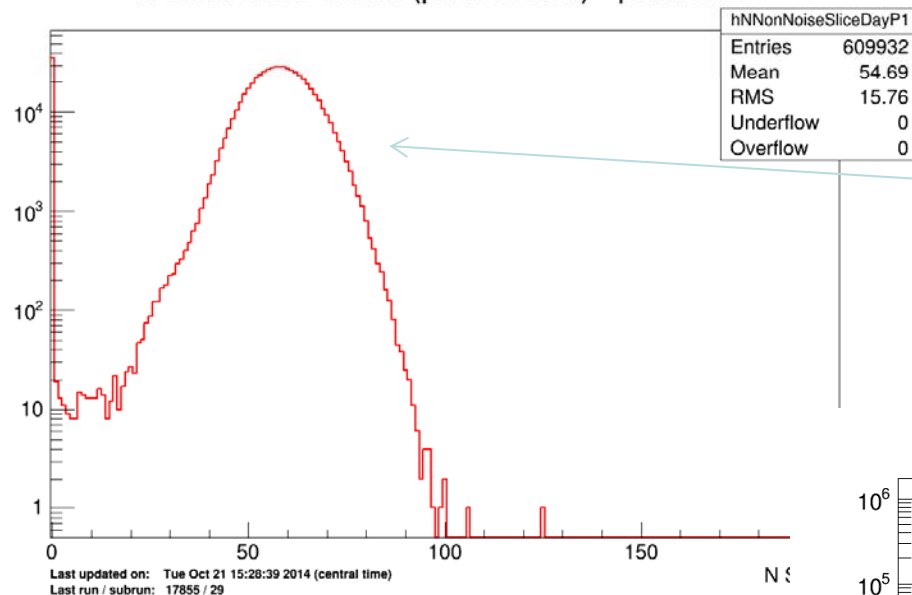
Provides OnMon-style plots over daily/weekly/monthly timeframes.





Nearline Monitor

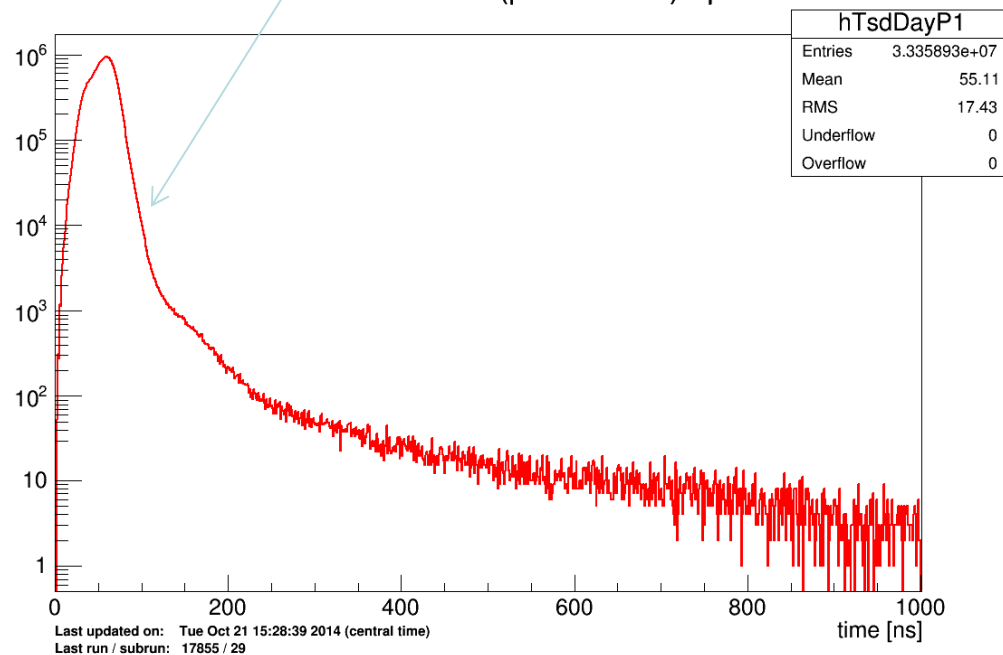
Non-Noise Slices (past 24 hrs.) - partition 1



Slice per trigger.

Slice Time Stand. Dev.

Slice Time Standard Dev. (past 24 hrs.) - partition 1

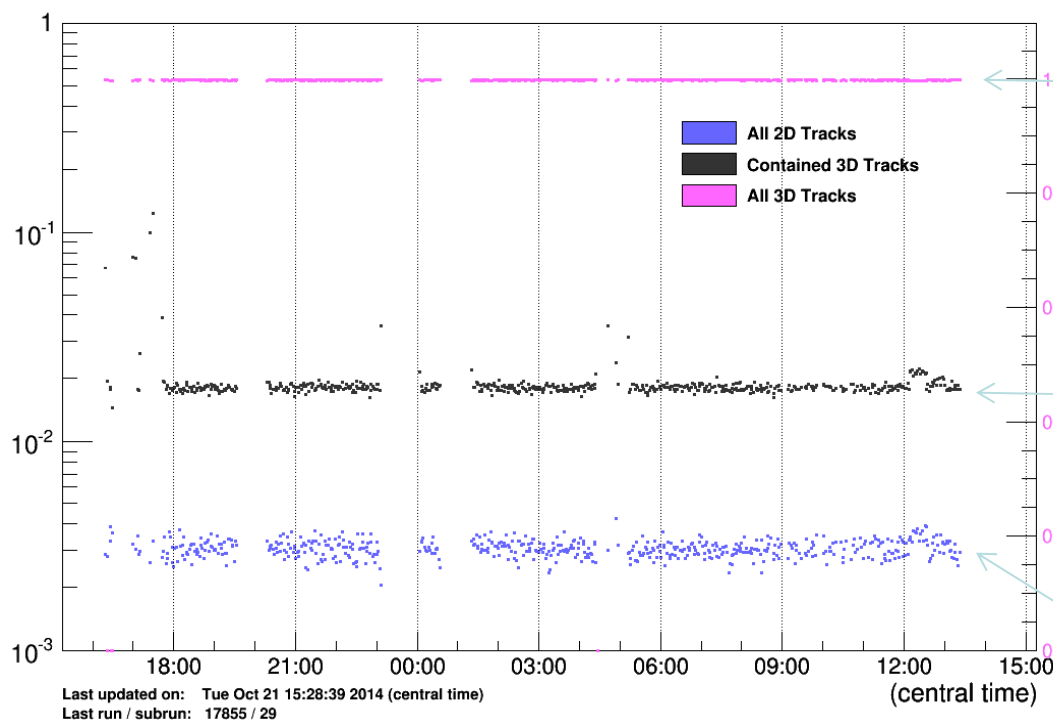


Provides monitoring based on low level reconstruction quantities



Nearline Monitor: Track Reco Efficiency

Track Fractions - partition 1



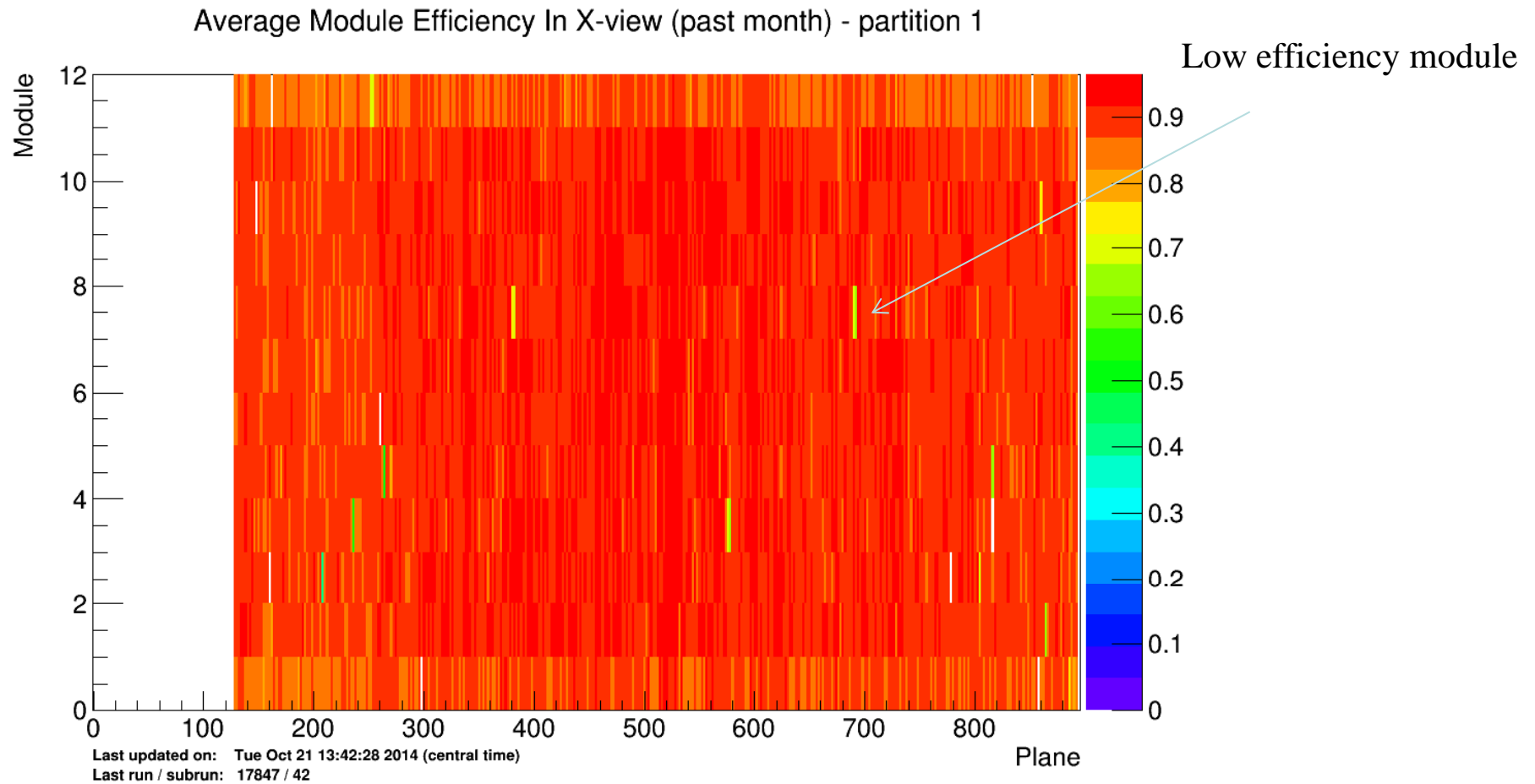
Fraction of tracks with full 3D reco. (right scale)

Fraction of tracks satisfying containment reqmt. (left scale)

Fraction of tracks with 2D reco. (left scale)

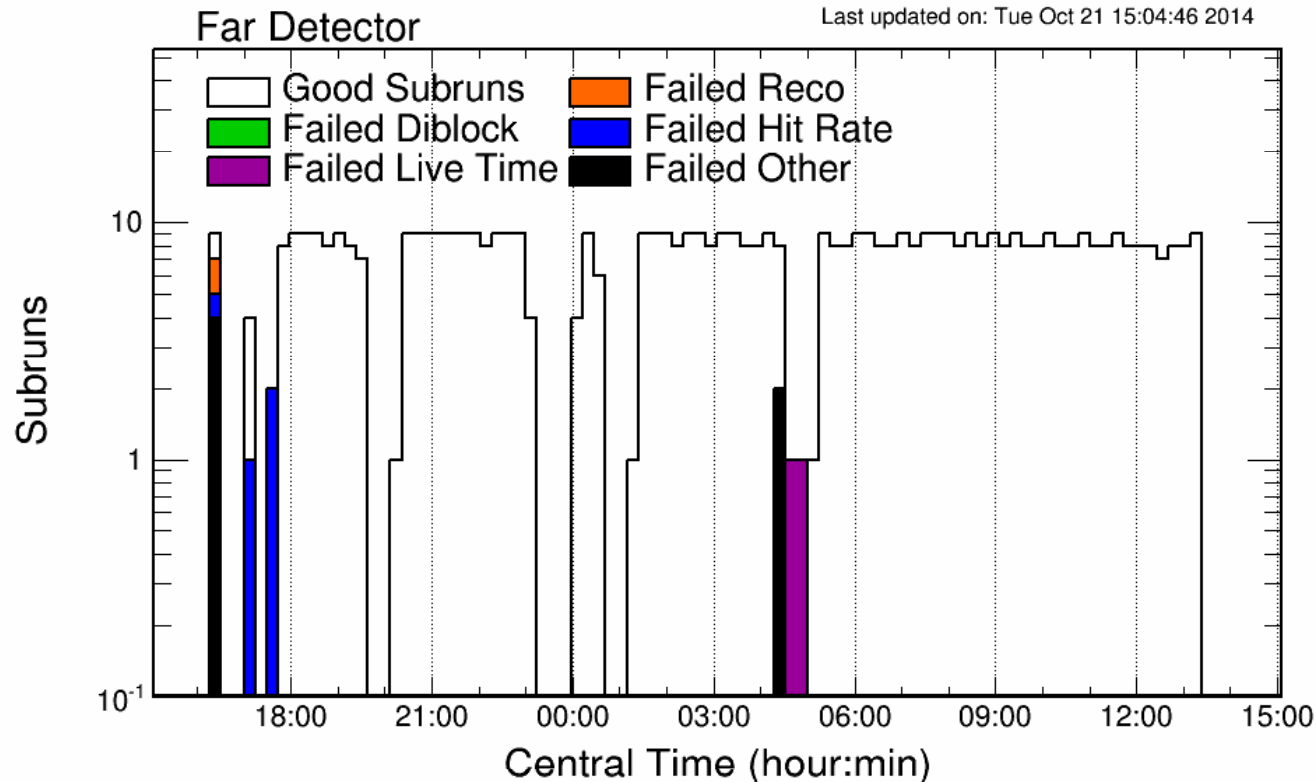


Nearline Monitor: Module Efficiency





Nearline Monitoring



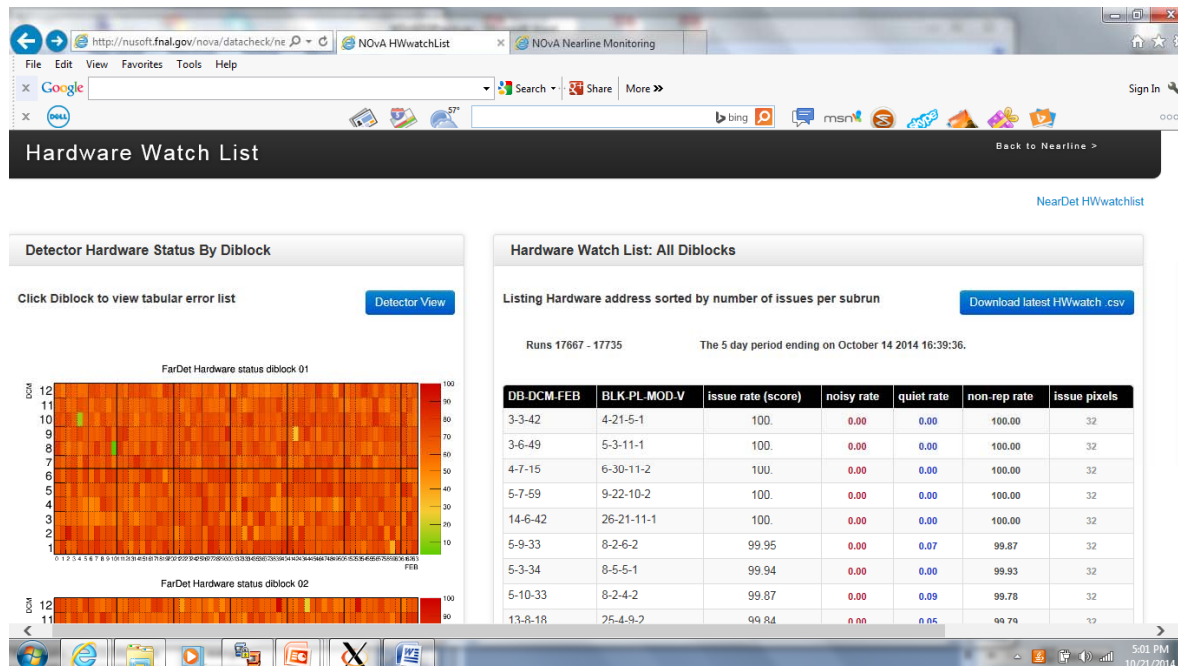
- The Nearline tracks good run selection efficiency in real time.
- This provides a simple single point check of overall data quality.
- Typical good run selection efficiency is $>98\%$.



Hardware Maintenance Support

- Hardware Watch monitoring tracks front end electronic performance based on rate and gain-based metrics. Reports of outliers needing maintenance are provided weekly, along with a continuously update web-based summary.

<http://nusoft.fnal.gov/nova/datacheck/nearline/HardwareWatchList.php?det=FarDet>





File Transfer Checking

- Raw files are transferred from Ash River to Fermilab for processing and storage in SAM.
- Data integrity through this process is checked by:
 - CRC comparison before and after file transfer, detecting any corruption of data occurring in the transfer.
 - Extraction of metadata: metadata is used to characterize the data file, and aid in future retrieval of desired datasets. The generation of metadata requires the successful unpacking of all data blocks without corruption, including successful calculation of a CRC on each block.

Errors (if any) are logged to a web page for expert review.



Offline Product Data Quality Monitoring

- Keep-up data production pipeline provides the data sets used for FD timing peak validation, and DQ validation using higher level reconstruction quantities not available at Nearline processing time.
- The FD timing peak monitoring is carried out by the DQ group using these keep-up data sets. To date, this process has involved a simple event pre-selection followed by event scanning (see Ryan's talk) . We are in the process of refining/automating this process, eliminating the scanning step. All the components for this are in hand.
- DQ validation at the post-processing level is carried out both within the DQ group itself, and with extensive support by analysis groups.



Summary

- The NOvA data quality tracking tools are fully developed and in place.
- These tools were employed during commissioning and so are mature and robust.
- We are ready for beam!