Technical Note on the ICC Work Package: Deglitching

SPIRE-UOL-NOT-002205

Prepared by:

Peter Davis (University of Lethbridge)
peter.davis@uleth.ca

Todd Atkinson (University of Alberta)
todda@ualberta.ca

Trevor Fulton (University of Lethbridge)
trevor.fulton@uleth.ca

Andres Rebolledo (University of Lethbridge)
andres.rebolledo@uleth.ca

Approved by:

David Naylor (University of Lethbridge)
naylor@uleth.ca

Document History:

Issue	Date
0.1 draft	5 March 2004
Version 1.0	May 13, 2004
Version 1.1	November 18, 2004

Table of Contents:

1. Distribution List

Dave Clements (ICL)

~~Matt Fox (ICL)~~

Mattia Vaccari (ICL)

~~Ken Ganga (IPAC)~~

Bernhard Schulz (IPAC)

Steve Serjeant (Kent)

Toshi Takagi (Kent)

Glenn White (Kent)

Jean-Paul Baluteau (LAM)

~~Didier Ferrand (LAM)~~

Christophe Ordenovic (LAM)

Christian Surace (LAM)

Steve Guest (RAL)

Ken King (RAL)

Tanya Lim (RAL)

Sunil Sidher (RAL)

Bruce Swinyard (RAL)

Sarah Leeks (ESTEC)

2. Reference Documents

ID	Title	Author
RD1	SPIRE ICC Consolidated WorkPlan, 1.0 draft 2	Ken King
RD2	Operating Modes for the SPIRE Instrument v3, SPIRE-RAL-PRJ-000320	Bruce Swinyard
RD3	Minutes of the OBST / ISDT workshop, 6TH and 7TH April, Imperial College London	Dave Clements
RD4	DCU Design Document v0.3, SAp-SPIRE- FP-0063-02	Frederic Pinsard

3. Applicable Documents

ID	Title	Author
Crill 2001	A Measurement of the Angular Power Spectrum of the Cosmic Microwave Background with a Long Duration Balloon-borne Receiver, PhD Thesis	Brendan Crill (CalTech)
Tahic 2003	Wavelet Analysis, PPT presentation	Margaret Tahic (UoL)
Nieminen 2004	Herschel Space Radiation Environment and SREM	Petteri Nieminen (ESTEC)

Fulton 2004	Glitch Simulation, v0.1	Trevor Fulton (UoL)

4. Acronyms

CALT	Calibration Team
CSA	Canadian Space Agency
DRCU	Detector Read-out and Control Unit
ESTEC	European Space Research and Technology Centre
ICC	Instrument Control Centre
ICL	Imperial College London
IPAC	Infrared Processing and Analysis Center
ISDT	Instrument Software Development Team
ISO	Infrared Space Observatory
JAC	Joint Astronomy Centre
JPL	Jet Propulsion Laboratory
LAM	Laboratoire d'Astrophysique de Marseille
OBST	Observations and Data Processing Team
RAL	Rutherford Appleton Laboratory
SMEC	Spectrometer Mechanical Unit
TBC	To be confirmed
UoL	University of Lethbridge
ZPD	Zero Path Difference

5. Objective of this document

This document provides a comprehensive and detailed overview of the processes that go into the delivery of the ICC Work Package Deglitching. It aims to collect and consolidate contributions from groups and individuals.

6. Objective of Work Package

"Provide data processing steps and software used to identify and remove (or flag for removal, TBC) detector data that is affected by 'glitches' – assumed to be caused by cosmic ray hits on the detector." (RD1, 33)

It is currently agreed that the Deglitching work package identifies glitches in the photometer data whereas the Time Averaging work package removes the identified glitches. The work package will identify **and** remove glitches for the spectrometer data.

The performance goal for detection algorithms is (1) to correctly identify >95% of 3-sigma glitches and (2) to flag <2% false positives per dataset, i.e. ~500 data points. The minimal performance requirement for the detection algorithms is to achieve the above mentioned performance level for 5-sigma glitches.

~~There are currently two major uncertainties concerning this work package:~~

~~One question is whether and how data, flagged as glitch signature, are removed. We propose to remove glitches by the same task that identifies the glitches. Glitch removal should be performed by the glitch identification module as different ways of identifying glitches will lead to different ways to remove glitches. A switch will be made available to the user to set how data will be removed. This will allow users to cut out the glitches, retain as much data as possible, or not remove glitches at all. Possibly, there will be more options. By default, glitches will be removed.~~

~~A second question concerns the various stages of deglitching. When are deglitching schemes used and at what stage in the data processing pipeline are they implemented?~~ We propose to use a robust and computationally straightforward deglitching scheme at the first stage and ~~a~~ more complicated schemes (wavelet analysis or such) ~~after demodulation~~later on.

Comments:

~~oThis work package deals with the deglitching of both, photometer, and spectrometer data.~~

- This work package will have to be revisited once SPIRE is in space to adjust for actual physical conditions for Herschel/SPIRE (cosmic ray flux, impulse response function).
- **The proposed implementation focuses of glitches that have been caused by cosmic ray hits. Only funnies with this characteristic shape will be flagged.**
- There will be different tasks to deal with the different operating modes POF1-4 & SOF3/4, POF5/6, and SOF1/2 respectively.

7. Input

Data input for the spectrometer will be the timeline from the detectors (SDT) and the stage.

Data input for the photometer will be Phot Detector Timeline (PDT).

It is also assumed that additional metadata will be available, such as the operating parameters of the photometer/spectrometer.

Comments:

- The input/output interfaces depend on where in the pipeline this task is located.

8. Output

Data output for the spectrometer will be the ~~unchanged~~ deglitched spec detector timeline (SDT)~~ with a Quality Control bitmap data points which are affected by glitches~~.

Data output for the photometer will be the unchanged Phot Detector Timeline (PDT) with an additional ~~Quality Control~~ data bitma~~p~~sk to indicate which data points ~~which~~ are affected by glitches plus information used for the handling of these data points.

~~Caveat: Whether the data remain unchanged depends on the scope of this task (see earlier).~~

Information on the cosmic rays will be made available such as number, tallest peak, area under curve, duration, or the complete cut-out signatures. For each data point flagged as a cosmic ray hit, the tasks will produce the normalized RMS to indicate how confident the scheme was about that particular set of points being due to a cosmic ray.

~~Comments:~~

~~oThe input/output interfaces depend on where in the pipeline this task is located.~~

9. Quality Control

Possible indicators of the data quality from the work package Deglitching are:

- Number of samples affected by glitches divided by the number of samples.
- The noise-level of the baseline data, measured by the square root of the median of the squared differences to the median.

~~The task is designed specifically to manage data quality. A bitmap to flag and remove or treat data points will be produced as the main product of this task (see above).~~

~~The percentage of data points affected by cosmic rays could be given for each observation.~~

10. Milestones

May 15, 2004: First version of the data product definition

July 2004: Definition of Data Products

March 2005: Definition of deglitching schemes

~~January~~ September 2005: Delivery of Processing Modules

~~January~~ September 2005: Delivery of Conversion Data/Tables

11. Involved parties

11.1. SPIRE teams

- Other users of input data: DRCU, in particular photometer bolometer arrays, spectrometer bolometer arrays, and SMEC ~~(JPL ???; LAM, Jean-Paul Baluteau, …)~~
- Other users of output data: DRCU, in particular bolometer arrays ~~(JPL ???; IPAC, LAM, Jean-Paul Baluteau, …)~~

- OBST (ICL, ~~Matt Fox~~Mattia Vaccari, Dave Clements)

~~□Quality Control (ICL, Dave Clements)~~

- Trend Analysis (RAL, Sunil Sidher?)
- CALT (RAL, Tanya Lim)
- ISDT (RAL, Steve Guest)

11.2. Institutions

- ICL: ~~Matt Fox (OBST, photometer deglitching),~~ Mattia Vaccari, Dave Clements (OBST, Quality Control)
- Kent: Steve Serjeant, Toshi Takagi, Glenn White (deglitching methods)
- IPAC: ~~Ken Ganga,~~ Bernhard Schulz (photometer deglitching)
- LAM: Jean-Paul Baluteau et al. (spectrometer deglitching, SMEC data)
- RAL: Tanya Lim (CALT), Steve Guest (ISDT), Sunil Sidher? (Trend Analysis)

12. Glitch characterization

The development of powerful glitch detection and removal routines depends on knowledge about the frequency and signature of glitches due to cosmic rays.

12.1. Signature of cosmic rays

Trevor Fulton has modeled the signature of a cosmic ray:

The following assumptions were made: The cosmic ray is a Dirac-spike, i.e. a certain amount of heat gets deposited instantaneously in the bolometer. The signature of a cosmic ray hit on the detectors depends on the thermal behavior of the bolometers and the electrical filtering of the read-out electronics.

The thermal response of the SPIRE bolometers to cosmic rays can be modeled as an RC-low-pass-filter. The average time constant of the bolometers has been measured to be ~~10 – 20~~18 ms for PLW. A time constant of ~~2 –~~4.2 ms is expected for the bolometers in the spectrometer array.

The specifics of the read-out electronics are given in RD4. The photometer arrays are read out through a four poles Bessel filter and the spectrometer arrays through a six poles Bessel filter. The following pass-band filter described in RD4 throws the modeled cosmic ray signature out of whack and has been left out until technical details have been discussed with CEA.

Under these assumptions, the signature of a cosmic ray can be characterized as follows:

Spectrometer

$$y(t) = h(t) = Ae^{-\omega_t t} + 2|(B)|e^{-\alpha_1 t}\cos(\beta_1 t + \angle B)$$

$$+ 2|(C)|e^{-\alpha_2 t}\cos(\beta_2 t + \angle C) + 2|(D)|e^{-\alpha_3 t}\cos(\beta_3 t + \angle D)$$

with:

- A = 4.4475, $\omega_t = 55.56$
- B = 15.583, $\angle B = 138.116^{o}$
- $\alpha_1 = 236.432$, $\beta_1 = 80.518$
- C = 1.1109, $\angle C = -120.205^{o}$
- $\alpha_2 = 147.382$, $\beta_2 = 264.775$
- D = 9.937, $\angle D = 0.5917^{o}$
- $\alpha_3 = 211.604$, $\beta_3 = 153.242$

Photometer

$$y(t) = h(t) = Ae^{-\omega_t t} + 2|(B)|e^{-\alpha_1 t}\cos(\beta_1 t + \angle B) + 2|(C)|e^{-\alpha_2 t}\cos(\beta_2 t + \angle C)$$

with:

- A = 28.458, $\omega_t = 55.56$
- B = 27.891, $\angle B = -124.551^{o}$
- $\alpha_1 = 42.326$, $\beta_1 = 14.021$
- C = 0.52675, $\angle C = +67.722^{o}$
- $\alpha_2 = 4.190$, $\beta_2 = 39.071$

The curves in the following figure are simulations of the output signal for the spectrometer (black) and photometer (red) systems in response to a simulated cosmic ray event.

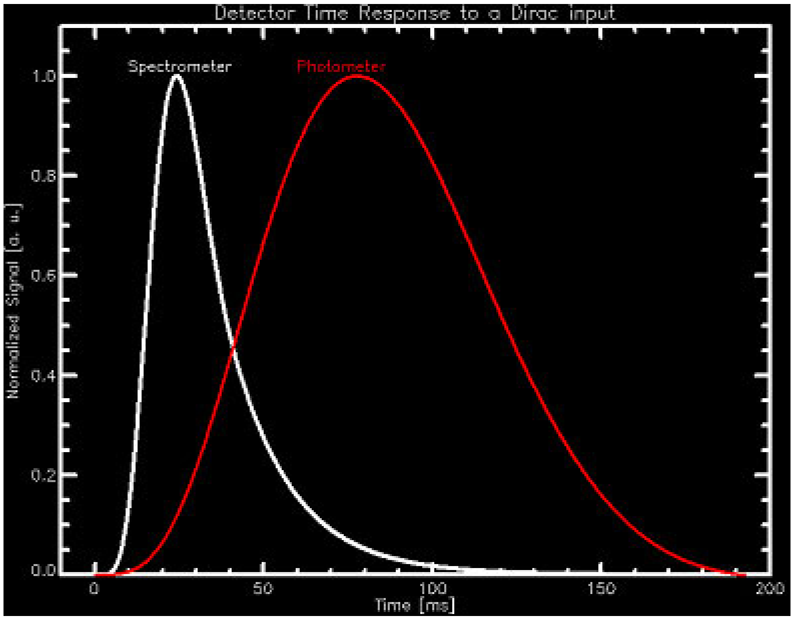


Figure 1 Time Response of SPIRE Detectors to a Simulated Cosmic Ray Event

12.2. Cosmic ray flux

The uncertainty about the frequency of bolometers catching cosmic rays is high. First, the cross section of spider-web bolometers for cosmic rays has not been characterized. Second, the cosmic ray flux at L2 is unknown. Third, the shielding of the bolometers through the Herschel satellite is not known.

~~Sarah Leeks is doing research in this area.~~No conclusive results on this issue are expected before Herschel actually gets to L2. The qualification of the glitch rate should be analyzed as early as possible.

13. Implementation options

The deglitching work package can be broken down into ~~four~~ three parts: Glitch (1) identification and (2) removal for the spectrometer. Glitch (3) identification ~~and (4) removal~~ for the photometer. The sections below specify the implementation options for all four areas.

The operating modes for SPIRE (RD2) need to be taken into consideration when developing glitch identification and removal routines. The spectrometer operates in two qualitatively different modes: Continuous scan and step and integrate. As far as deglitching is concerned, the photometer operates in three qualitatively different modes: Chopped with pointed telescope, scanning telescope, and chopped with scanning telescope.

13.1. Spectrometer – glitch identification

The glitch identification routine for the step and integrate operating modes SOF3/4 will follow the schemes for the photometer as data from individual stage positions are qualitatively similar to data from the photometer operating mode POF1. In the following section, glitch identification is discussed for the continuous scan modes SOF1/2.

13.1.1. Wavelet Analysis

Wavelet analysis provides the opportunity to inspect the detector data in the time-domain. This should make the identification of glitches easier as cosmic rays have a signature that can be characterized by its time-evolution and not – such as the interferogram – based on spatial modulation. Also, any kind interpolation prior to searching for glitches may introduce unwanted artifacts, or smear out, i.e. dampen the effect of the cosmic rays and therefore make it harder to identify the glitches.

However, as of now, the feasibility of this option has not yet been proven. There is no first hand experience on deglitching interferograms based on wavelet analysis available. The University of Lethbridge has explored this option using the (commercial) IDL wavelet package. A Powerpoint presentation (Tahic 2003) has been compiled to show the strength of wavelet analysis for the inspection of interferograms. LAM is contacting researchers who use wavelet deglitching in high-energy physics.

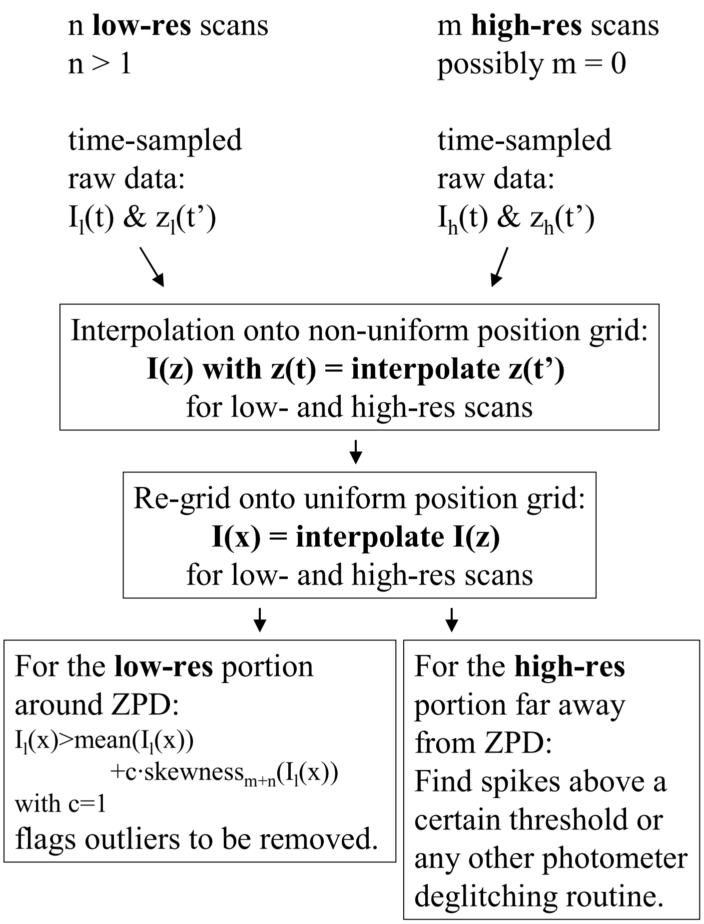
~~This option requires further investigation into~~

~~(a)the availability of suitable wavelet analysis packages in Java.~~

~~(b)the definition of criteria to flag a glitch.~~

13.1.2. Compare Interferograms

A better known route to identifying glitches is to average interferograms from the same source and look for outliers. The following describes a respective process:



It should be noted that this deglitching approach will have to keep up and down-scans separate as the phase-shift of the interferograms depends on the scan-direction.

The comparative algorithm for the low-res part has been implemented and tested for full-length interferograms from post-vibration CQM2 testing. When compared to visual inspection, this algorithm is more sensitive: It finds all the glitches a person has found and more. 3/8. 13/26, 6/15 glitches were flagged only by the algorithm. A single pass through the interferograms was sufficient. The sensitivity of this algorithm is still tbd.

13.1.2.1. Central Peak Region, n = 2

If there will be a case where only 2 interferograms have been taken of a specific target, the following algorithm will find glitches:

1. Interpolate both interferograms onto the uniform position grid.

2. Calculate the difference between interferograms. All data analysis is done on this difference array NOT the actual interferograms.
3. The difference in interferograms should be zero plus small noise and the glitch. There will be more noise around ZPD because of the larger magnitude of the numbers there but the glitch should still stand out.
4. Use a steep slope test to identify glitch points, checking for monotonic recoveries of 6 or more points.
5. Glitch stops when the difference is greater than or equal to zero (assuming a negative spike) since it has recovered back to the original interferogram.

13.1.2.2. Wing Analysis

Beyond the double-sided part of the interferogram, glitch identification can be based on data from a single interferogram. The following algorithm has been developed and tested:

1. Cut the single-sided wing from I(t').
2. Fit a Gaussian envelope to the wing and divide the data by the Gaussian function.
3. Compute the three-point difference function $\Delta_i = d_i - 0.5(d_{i-1} + d_{i+1})$ to measure curvature for each point in the wing.
4. For a window (200 points wide, tbc.), sliding through the wing: compute the standard deviation for the slopes. Flag any points as glitches, where Δ_i is greater than 1.25 (tbc.) standard deviations of the slope.
5. Replace all flagged glitches by interpolating between the adjacent points.

Preliminary results for this glitch identification algorithm are given in Figure 2:

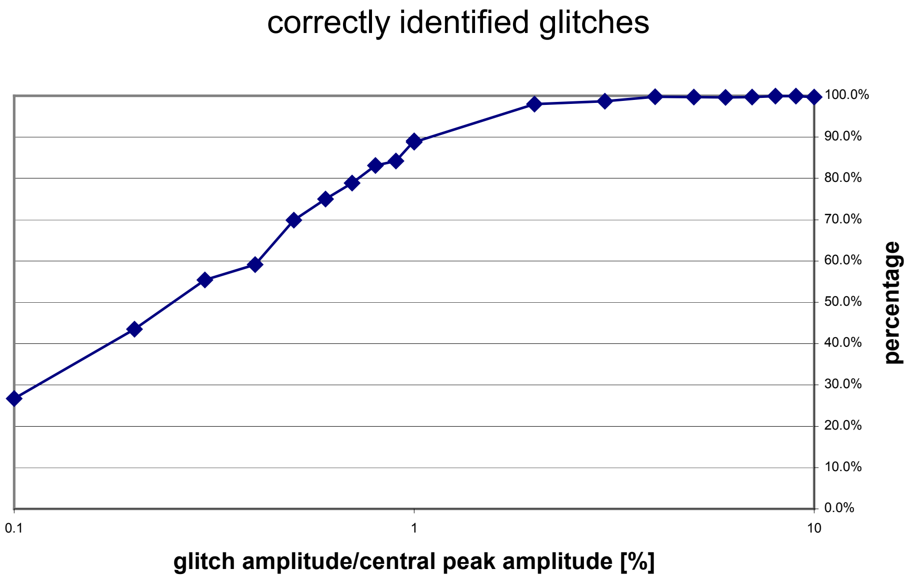


Figure 2: Percentage of correctly identified glitches in the interferogram wing as a function of glitch amplitude, measured as a fraction of the central peak amplitude

The number of false positives stays well below 1% per interferogram (see Figure 3). However, it should be noted that the chance to catch a false positive increases with increased glitch frequency and amplitude:

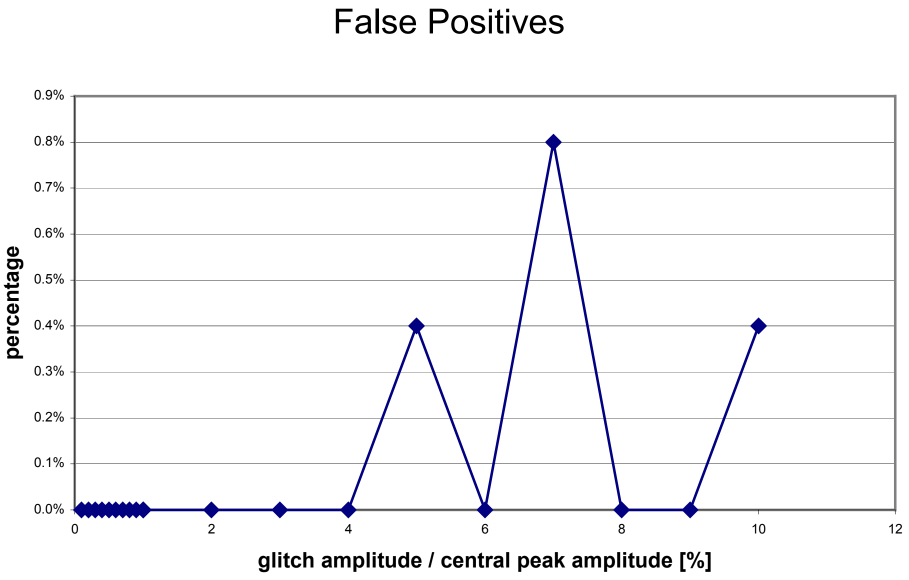


Figure 3: Percentage of false positives per interferogram

13.1.3. Other Approaches

Other approaches to identifying glitches in interferograms have been proposed and used for deglitching of the UoL-FTS at the JAC, Hawaii. They have proven to be inadequate:

- Comparison of interferograms to a predefined envelope.
  × Is too coarse as a generic envelope has to be rather large and specific envelopes are too difficult to provide.
- Search for large slopes in the interferogram.
  × Cannot deal with the region around ZPD where the signal covers a large range.
- Calculate the running next-neighbour average and look for strong deviations.
  × Cannot deal with the region around ZPD where the signal covers a large range.

13.2. Spectrometer – glitch removal

Very little work has been done on glitch removal as of now~~ since the response of the detectors has not been determined yet~~.

- Linear fit to cover affected data points.
  × Works OK when done manually, automatic removal has faces the problem of finding a balance between getting rid of the glitch completely but not more than is necessary.

- Remove a fitted exponential (one or two time constants)

  × Probably not feasible as large amounts of deposited energy will lead to non-linear behavior of the bolometer.
- Fit data for affected data points based on surrounding signal shape
- Don't remove glitches, but throw data out completely (only feasible if cosmic rays occur rarely enough).
- Fill in averaged data points from clean interferograms.

It is expected that it will be necessary to patch up regions that have been affected by cosmic rays in order to guarantee the integrity of the interferogram for the FT.

13.3. *Photometer – glitch identification*

The glitch identification routines for the photometer depend on the operating mode in use and when the deglitching task is called as part of the data analysis pipeline:

POF1-4 'Chop Without Jiggling' leads to quadrature-like data ~~at the early stage (Deglitch1 in RD3)~~ which is demodulated into one-level data streams ~~later on~~in the demodulation task~~ (Deglitch2 in RD3)~~.

POF5/6 'Scan Map without/with Chopping' make it harder to tell glitches from scanned point sources in time-ordered data. It should be possible to differentiate bright point sources and glitches by the point spread function (PSF) of sources which should be Gaussian and recognizably different from the steep rise and exponential decay for cosmic ray glitches. An analysis of time-ordered data across pixels also seems to be possible. This analysis would have to take into account how the telescope is scanning the sky.

~~The following sections discuss generic options to identify glitches. The analysis is~~An early analysis of glitch detection routines was based on data from BLAST. More detailed development of algorithms was done with synthetic data to match the properties of data from the various operating modes.

~~The following d~~Deglitching schemes were first tested on BLAST data and then further developed in algorithms suitable to deal with data from SPIRE:~~whether they could find exactly those cosmic ray signatures.~~

13.3.1. POF1-4: Chopped Stare

The chopped stare operating modes of SPIRE (POF1-4) will produce the most pristine data of all the operating modes: The signal should be dominated by the source with instrument noise plus glitches. Three methods have been tested on data that resembles data from these operating modes at a sampling rate of 16Hz.

Median outlier detection.

Three point difference function.

Fitting the impulse response function to potential glitches.

13.3.1.1. Median outlier detection.

A generic median outlier algorithm has been developed to distinguish glitches from instrument noise. It does not rely on knowledge on the physics of the instrument and will identify funnies largely independent of a specific glitch characteristic.

1. **Demodulation**
   Separate the chopped signal into upper and lower levels. This is currently done by calculating the average value of the signal which, assuming two reasonably constant signal levels, should be approximately halfway between the two levels. Therefore any signal values greater than the average are stored in the 'upper' array and all values less than the average are store in the 'lower' array. The two arrays are then analyzed separately. For operation, the separation into two levels will be able to rely on instrument data and will not depend on a difference in voltage between the two BSM positions.
2. **Computation of the median and the RMS of the deviations from the median**
   In order to separate glitches from noise with an outlier scheme it is necessary to determine the level of the data baseline and its spread around that baseline – regardless of any glitches apparent in the data. The average and standard deviation of the measured data are not suitable in this respect since they increase significantly with even a small number of glitches present in the data. The median and the square root of the median of the squared differences to the median (σ_m) are considerably less sensitive to small numbers of glitches in the data.
3. **Outlier detection**
   All points that are $\geq 6.25\ \sigma_m$ below median are identified as glitches.
4. **Tracing the glitch**
   In order to flag not only the lowest points of the glitch but all samples that are significantly affected by the glitch, samples adjacent to identified glitches are subject to a more sensitive outlier detection. They are identified as glitches if they are $\geq 3.5\ \sigma_m$ below the median. The tracing of the glitch stops when this criterion yields no additional results.

This algorithm was developed and tested with synthetic data that was built around two arbitrary baselines (the levels are separated by 10%) with white noise (standard deviation $\sigma = 0.04$ as measured from SPIRE CQM testing) and impulse response functions as modeled in Fulton 2004. Data sets of 500 points each are generated. They are seeded with a stochastic distribution (probability of a glitch is $p = 0.1\%$) of glitches of random amplitude (between 1 and 100 σ). Data sets are generated until 1000 glitches have been inserted into the data. These tests have shown that all glitches of 5σ amplitude or above are identified (see Figure 4).

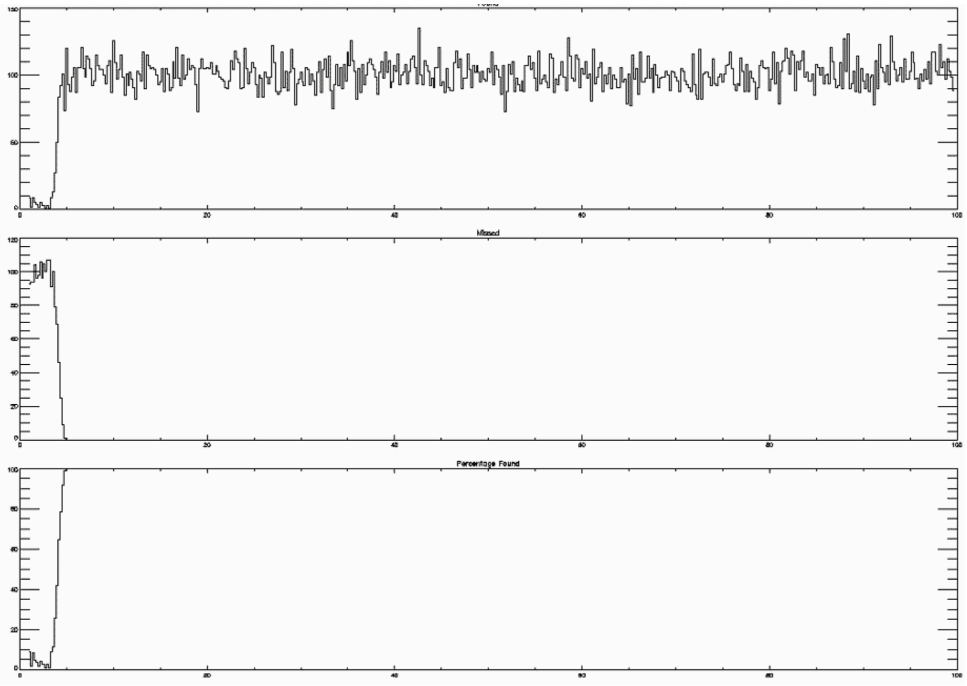


Figure 4: Number of detected glitches (top), number of missed glitches (middle), and percentage of detected glitches (bottom) as function of glitch amplitude.

The number of false positives identified is independent of the number and amplitudes of glitches present and in the range of 1.3 ± 0.3 %. The performance of the algorithm degrades sharply between 3 and 5σ.

Glitch Tracing

Once a glitch has been (hopefully correctly) identified, there is another problem for the deglitching process: It must be ensured that only those points are removed that are significantly affected by a cosmic ray hit. Without any further treatment of the glitches to remove all points affected by the same cosmic ray hit, the following results (see Figure 5) for the correct identification of glitches have been obtained:

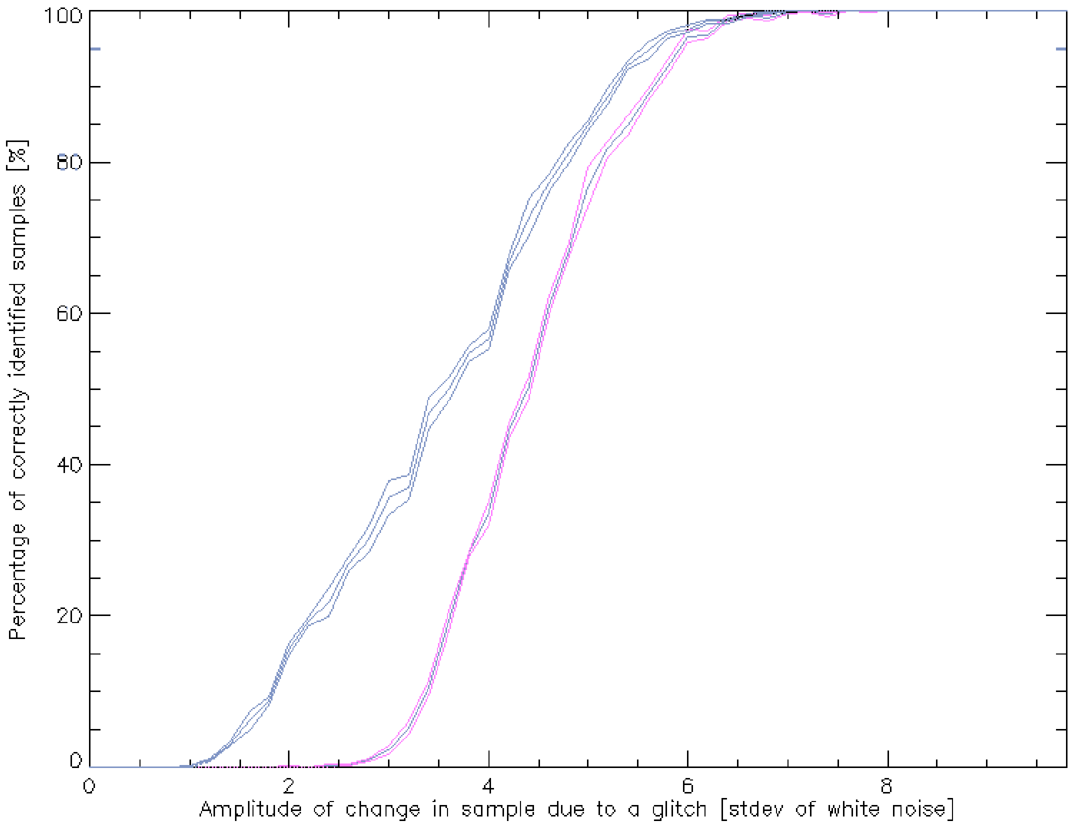


Figure 5: Percentage of correctly identified samples as a function of the sample's change in amplitude due to cosmic ray hits, measured in standard deviations of the introduced white noise. Success rate of the detection without (red) and with (green) glitch tracing. Included are also the standard deviations of the results.

A more sensitive outlier detection has been introduced to remove a greater proportion of affected samples. The same principle (outliers from the median) is used, however, with a decreased threshold of 3.5σ. While the percentage of false positives increased slightly to 1.7%, this tracing of the glitch significantly improved the percentage of correctly identified glitches, particularly in the 2 – 6 σ region.

Interface Comments

- The first step is an ad-hoc demodulation of the data stream. Mattia Vaccari is responsible for a WP Demodulation to do that.
- Tracing the glitch can either be done in the work package Deglitching or in Time Averaging.

13.3.1.2. Three point difference function

Calculate the three point difference function $\mathrm{delta}_i = d_i - 0.5\,(d_{i-1} + d_{i+1})$ of the time-ordered data, where σ refers to the standard deviation of the data. This function measures essentially the curvature at any given point. Results are very promising (see Figure 6, Figure 7, and Table 1):

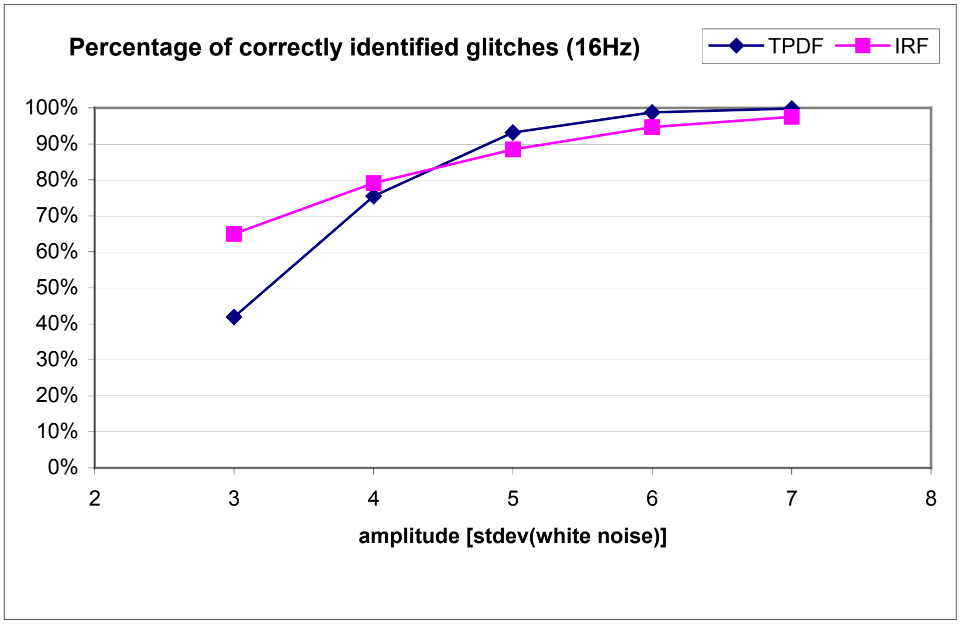


Figure 6: Percentage of correctly identified glitches, with a three-point difference function, and by fitting the expected impulse response function; with a 16Hz sampling rate

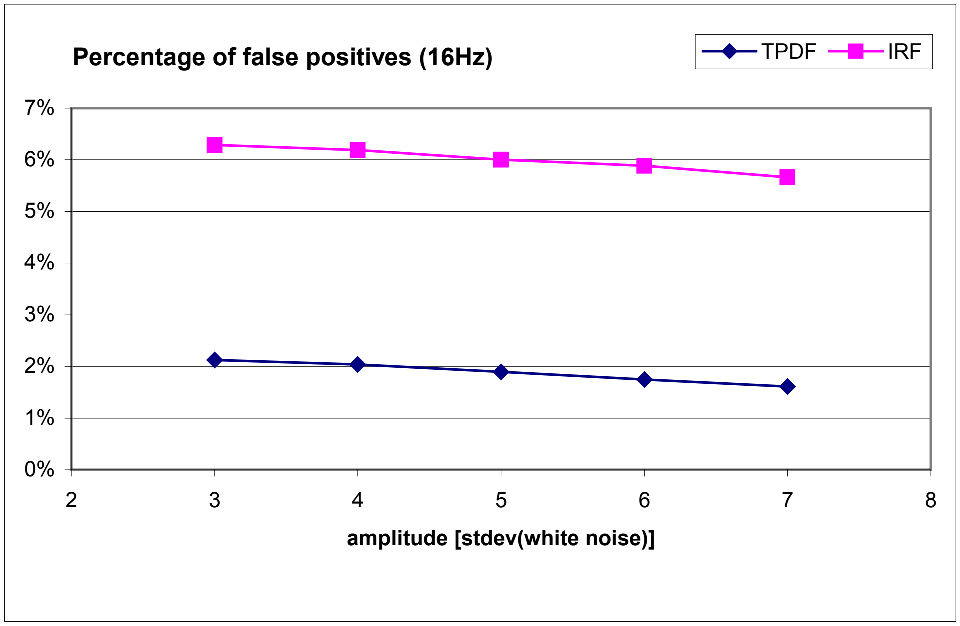


Figure 7: Percentage of false positives, with a three-point difference function with a three-point difference function, and by fitting the expected impulse response function; with a 16Hz sampling rate

Amplitude	3	4	5	6	7
Correctly identified	41.94%	75.47%	93.15%	98.80%	99.82%
False positives	2.13%	2.04%	1.90%	1.75%	1.61%

Table 1: Performance of a three-point difference scheme at 16Hz

13.3.1.3. Fitting the impulse response function

A very conservative steep slope criterion (1 sigma outliers) is used to create a pool of potentials. In simulations, this pool includes most of the glitches. However, a large number of false positives (~14% of all data points) are flagged as well (see Table 2).

Glitch amplitude	3	4	5	6	7
Correctly identified	97.64%	99.63%	99.82%	100.00%	99.88%
False positives	14.86%	14.52%	14.25%	13.78%	13.37%

Table 2: Creating a pool of potential glitches through a 1 sigma, steep slope criterion.

The percentage of false positives can be reduced in a number of ways. Fitting the expected impulse response function to a number of points around the identified potentials has shown the best results so far. If the normalized RMS of the fit is less than 0.15, the potential is accepted as a glitch. For both, the percentage of correctly identified and false positives, the performance goal is not achieved (see Figure 6 and Figure 7). The goal of keeping the percentage of false positives below 2% is not practical with this approach.

Glitch amplitude	3	4	5	6	7
Correctly identified	65.01%	79.10%	88.42%	94.67%	97.55%
False positives	6.28%	6.19%	6.00%	5.89%	5.66%

Table 3: Performance of glitch identification via impulse response function fitting; normalized RMS must be less than 0.15 for a potential to count as a glitch.

13.3.1.4. Comparison and evaluation

The median outlier detection is close to the detection goal: Almost 90% of 5 sigma glitches are detected while the percentage of false positives is low with 1.7%.

The three-point difference function satisfies the 2% criterion for false positives while detecting close to 95% of 5 sigma glitches.

The fitting of the expected IRF yields results which are considerably worse (5 to 6%) than the goal of 2% false positives. At the same time, the level of correctly detected glitches is around 90% for 5 sigma glitches.

Between the first two detection algorithms, the three-point difference function is preferable as it shows a significantly better performance for small glitches (3 and 4 sigma) than the median outlier algorithm which falls off sharply.

13.3.2. POF5: Scan

The scanning operating mode of SPIRE (POF5) will produce a 25Hz timeline of data samples that reflects a continuous line of sight of the telescope while it is scanning across the sky. Sudden changes in the signal may be due to bright point sources or cosmic ray hits. Ideally, data from a single timeline can be used to discriminate between these two cases based on the different signal profiles of a point source and a cosmic ray glitch. If this cannot be done from data from a single pixel, a considerably more complicated scheme will have to be developed that looks for the bright point source as it comes into the field of view of several pixels.

13.3.2.1. Median outlier detection

A straightforward median outlier is not feasible for the scanning operating mode since any modulation during a scan, e.g. due to a bright source, would be tagged as a glitch even in the complete absence of cosmic ray impacts.

13.3.2.2. Three-point difference function

The same algorithm as for the chopped stare operating modes has been implemented. For the faster sampling rate of 25Hz, this algorithm is performing rather poorly (see Figure 8 and Figure 9) as the glitches are not essentially spikes any more, but several sampling points are recorded while the glitch affects the data stream.

13.3.2.3. Fitting the impulse response function

A very conservative steep slope criterion (1 sigma outliers) is used to create a pool of potentials. In simulations, this pool includes most of the glitches. However, a large number of false positives (~14% of all data points) are flagged as well (see Table 4).

Glitch amplitude	3	4	5	6	7
Correctly identified	93.43%	98.84%	99.64%	99.67%	99.84%
False positives	14.83%	14.67%	14.37%	14.11%	13.67%

Table 4: Creating a pool of potential glitches through a 1 sigma, steep slope criterion.

The percentage of false positives can be reduced in a number of ways. The length of the recovery period of the glitch has been shown to give the best results so far for the 25Hz sampling. If the recovery period includes more than 2 slopes, the potential is accepted as an actual glitch. However, for the percentage of correctly identified as well as the percentage of false positives, the performance goal is not achieved (see Figure 8 and Figure 9). The goal of keeping the percentage of false positives below 2% is not practical with this approach.

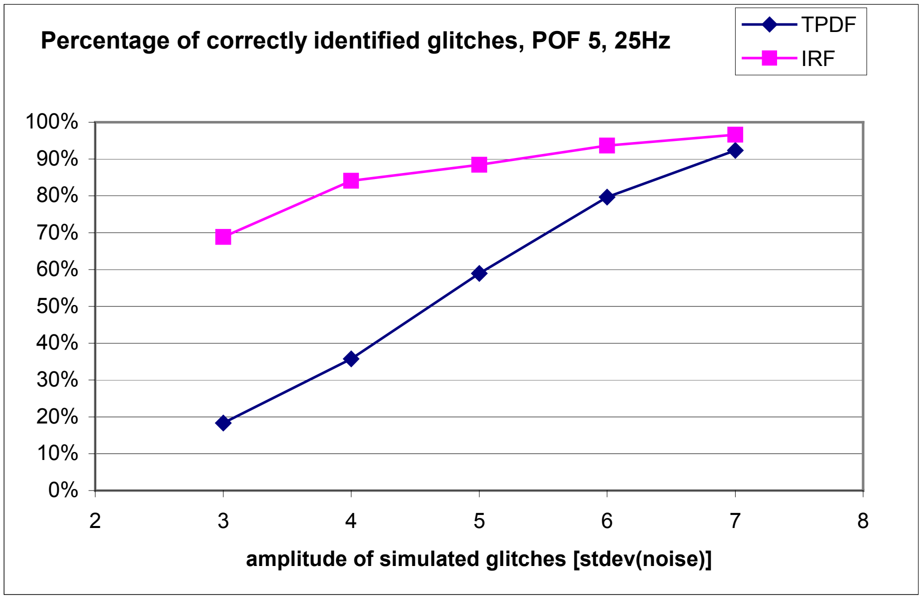


Figure 8: Percentage of correctly identified glitches, with a three-point difference function, and by rejecting glitches with short recovery periods, with a 25Hz sampling rate

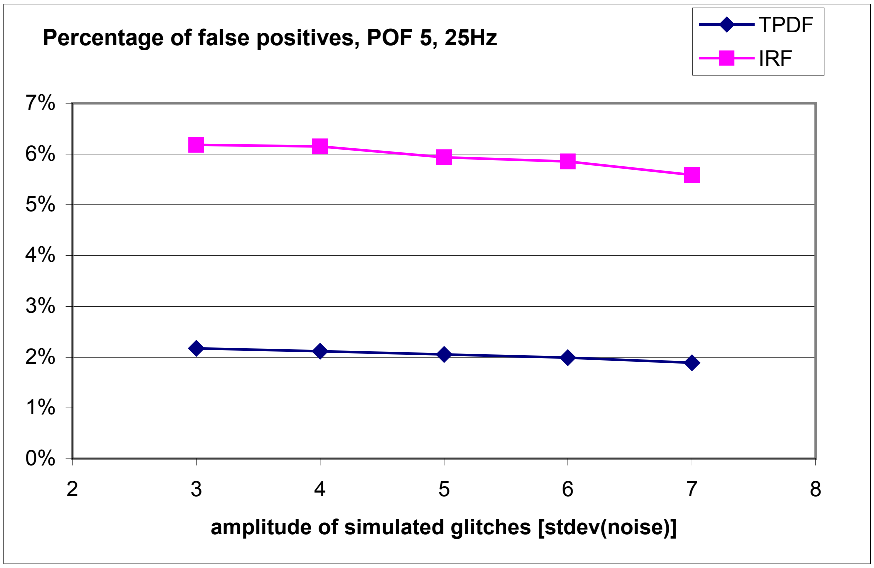


Figure 9: Percentage of false positives, with a three-point difference function, and by rejecting glitches with short recovery periods; with a 25Hz sampling rate

Glitch amplitude	3	4	5	6	7
Correctly identified	68.84%	84.08%	88.47%	93.65%	96.61%
False positives	6.23%	6.15%	6.04%	5.87%	5.74%

Table 5: Performance of glitch detection based on the recovery period of the glitch.

13.3.2.4. Comparison and evaluation

Both algorithms described above fail to meet the target performance. While the search for an extended rising slope yields about 6% false positives, it's superior performance in detecting actual glitches recommend it as the best available option as of now.

13.3.3. POF6: Chopped Scan

No work has been done specifically on this operating mode. It will be the most challenging in terms of deglitching. It should be possible to use techniques similar to the ones for the un-chopped scan mode.

~~13.4.Photometer – glitch removal~~

~~It seems that, in the case of the photometer, glitch removal would be simply the omission of the glitch data. Details need clarification such as whether data points are to be removed from the timelined data or whether the average of the neighboring data points is to be inserted.~~

14. Open Issues

- This work package needs the SMEC data in addition to detector data to identify glitches for the spectrometer!
- Q: ~~What~~ How do you determine which one is the best deglitching method?
  A: The one that flags ~~all~~ the most cosmic ray signatures and the fewest false positives~~nothing else~~.
- Q: What are the criteria to decide which method is best?
  A: The degree to which routines flag cosmic ray signatures of varying amplitude at arbitrary points in the data stream.
- Q: Will the base-level of the bolometers change due to a cosmic ray?
  A: Based on CQM data from ~~BLAST~~ the SPIRE PLW this seems highly unlikely. Even large transient events lead back to the same base-level.
- May require the merging of signal and position data into interferograms, i.e. modules from the work package Fourier Transformation.
- Can we anticipate kinds of glitches that are not caused by cosmic rays?
- Cosmic rays may also affect the read-out electronics of the stage. In this case, the position of the stage will have to be reconstructed by other means.
- Test data from the various operating modes will be necessary to develop and test the various tasks.
- If quality control relies on a bitmap per data point, information will be lost when data points are simply cut out.

15. Checkpoints

Other users/OBST:

- Have all providers of input data agreed on its format?
- Have all users of the end product agreed on its format?

Quality Control:

- Are adequate flags for data quality and processing quality provided?

Trend Analysis:

- Are adequate flags for long-term trends in the data quality and processing quality provided?

CALT:

- Have all necessary calibration tables been identified?
- Can all necessary calibration tables be provided by the scheduled tests/operations?

ISDT:

- Are all software requirements (coding conventions, exception handling, test procedures, documentation, benchmarking) met?