TECHNICAL REPORT

SPACE TELESCOPE SCIENCE INSTITUTE

Operated for NASA by AURA

Title: Level 2a Pipeline Processing of NIRCam Data		Doc #: JWST-STScI-002082, SM-12 Date: May 17 2010 Rev: -
Authors: M. Robberto	Phone: 410-464-0883	Release Date: 1 November 2011

1.0 Abstract

This document describes the processing steps that compose the Data Processing Level 2a (CALWEBB) for NIRCam. It is intended to provide enough information to allow setting the requirements for the development of the pipeline, including the needed calibration data, keywords and products. It envisions 10 steps, which include the initial definition of Data Quality and Error arrays, reference pixels, non-linearity, saturation, dark current, cosmic rays (both in time and spatial-domain), persistence and optimal ramp fitting. I propose a number of modifications to the current WFC3 pipeline, including the need of an extra Data Value in the JWST Data Quality images needed to properly account for Cosmic Ray events happening within a group of averaged samples. This document may provide a convenient baseline against which the various instrument teams may find commonalities and differences with regard to Data Processing Level 2a.

2.0 Introduction

The general goal of the JWST science data pipeline is to process science data to remove instrumental signatures and observing artifacts, yielding measurements in physical units ready to be analyzed for science (JWST DMS Concept Document, JWST-STScI-001526).

The structure of the JWST science data pipeline, originally defined in MO-602 of the MOCD, is based on the concept of "Science Data Processing Levels". The sequence of algorithms composing the pipeline have been grouped in 5 Data Processing Levels. Each Data Processing Levels implements a set of passages needed to a produce Science Data Products (fits files) of increasingly higher level. Science Data Products of level *n-1*, together with the needed auxiliary data, are given as input to Data Processing Level *n*, which in turns returns Science Data Products of level *n*. This architecture gives to the pipeline a convenient modular structure with well defined interfaces[1].

The nomenclature for the 5 Data Processing Levels has been recently redefined in the JWSTS DMS Concept Document, JWST-STScI-001526. They are copied here for reference:

[1] These definitions do not prevent the possibility of extracting fits files at intermediate level of processing for calibration purposes.

- Level 0: Recorded Science Data – Downlinked science data files containing packets of compressed pixel data, as stored on the Solid State Recorder (SRR). Data from one exposure with one detector may be spread across multiple SRR files. An SRR file may include data from more than one detector or more than one exposure.
- Level 1: Raw Exposure Data – FITS files containing up-the-ramp image data for one detector and one exposure, which may consist of multiple integrations. Relevant image data packets are extracted from recorded science data files, uncompressed and assembled into one image sequence per integration. Engineering keywords from the telemetry are also added to the fits header, creating the so-called Level 1a data; the further addition of keyword information from the Proposal and Engineering databases results in Level 1b data. The DMS will perform this conversion as a single step and the resulting data product is simply referred to as Level 1 data.
- Level 2: Calibrated Exposures
  - Level 2a: Count Rate Images – Up-the-ramp image sequences (raw exposure data) are fitted to create a count rate image, an uncertainty image and a data quality image for each integration. Count rate images are corrected for cosmic rays, bias drifts and dark current.
  - Level 2b: Processes Science Exposures – Count rate images are processed to remove instrumental signatures, to extract spectra, and to convert to physical units. Measurements from contemporaneous calibration exposures (e.g. science target location in a NIRSPEC reference image) are used when creating processed science exposures. Data from multiple integrations are combined where needed.
- Level 3: Combined Science Exposures – Associated science exposures (e.g. dithered images or spectra, coronographic images at different roll angles) are combined to form the definitive data product of an observation.
- Level 4: High-Level Science Products – "are designed to allow research on the entire JWST catalog (MO-602 in MOCD). Data from many JWST observations are processed to create catalogs, mosaics, or other data products that span many observations. This is equivalent to the High Level Archive for HST data.

This document describes the processing steps composing Data Processing Level 2a, i.e. from Raw Exposure Data to Count Rate Images. It should contain enough information to allow setting the requirements for the development of the pipeline, including the needed calibration data, keywords and products. At the same time, this document may provide a convenient baseline against which the various instrument teams may find commonalities and differences with regard to Data Processing Level 2a.

3.0 Data Processing Level2a: General Strategy

The fact that the NIRCam detectors are read out non-destructively allows monitoring the signal during the integration. This provides a powerful diagnostic tool to disentangle various types of anomalies as they appear during the data acquisition process and/or at different levels of the electronic chain. Since each anomaly affects the information that was available at the end of the previous level of the readout chain, it is clear that the optimal method of ramp processing should remove the system anomalies following their reverse order of appearance. Let us therefore look to the ideal structure of the signal chain.

At the beginning of the integration each detector diode (pixel) is biased (reset) to a relatively random level, due to the presence of KTC noise. The photon flux is then measured sampling the resulting de-biasing of the diode. Dark current is always present and initially contributes to the de-biasing process. Eventually, when the diode is pushed in forward-bias mode by integration of photoelectrons (and/or by the dark current itself), it is still a dark current term that balances any further flux of photogenerated charges producing "saturation". Charges trapped during previous bright exposures are also released and detected ("latency" or "persistence"). Cosmic rays of various energies hit the pixels, further de-biasing the junction and possibly causing saturation. Overall, the response to these combined effects is non-linear, due to the progressive reduction of the junction bias (Figure 1).

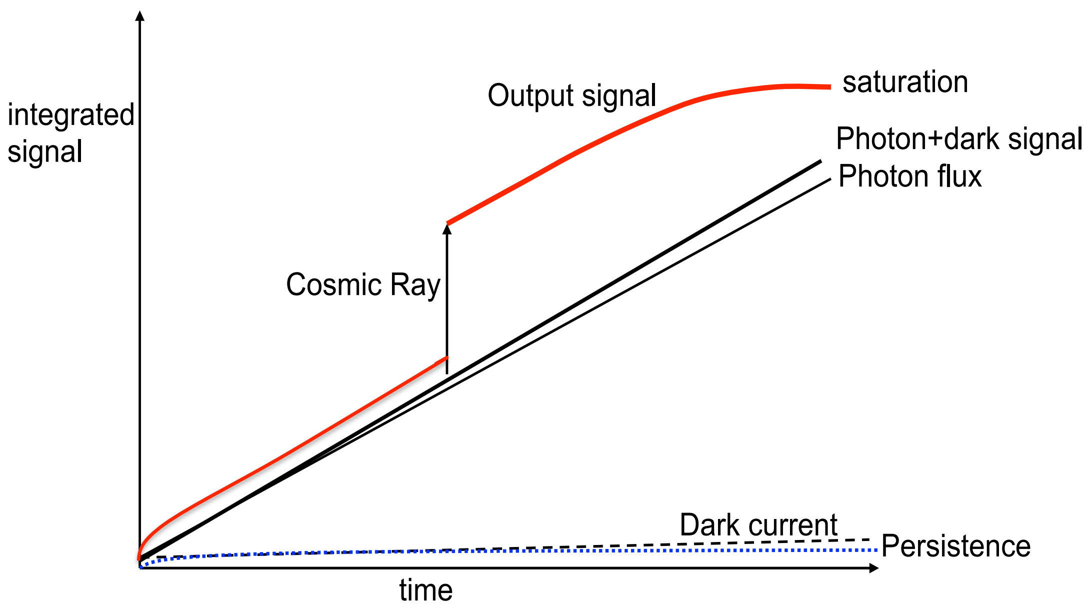

Figure 1 Schematic layout of ramp affected by the typical detector effects
For demonstration, an extreme case is shown with persistence dominating the signal at the beginning of the ramp, therefore causing an initial decay of the integrated symbol.

The integrated signal is extracted from the diode through a trans-impedance amplifier and further amplified by a readout chain. This adds random "readout" noise and drifts may appear in the DC level. The drifts can be corrected by comparison with the light insensitive reference pixels present on the detector. However, reference pixels are not sampled simultaneously with the signal and therefore they cannot correct for high frequency noise. The amplified signal voltage is finally converted into counts by an Analog-to-Digital converter and made available for data processing to recover the original photon flux.

The best strategy for data processing requires walking through this process in reverse order. One has therefore to implement the following main steps:

1. Remove the DC drifts by subtracting the appropriate level derived by the reference pixels
2. Re-establish the linearity of the detector response, which affects all signal components.
3. Deal separately with the four signal contributions, typically in the following order:
    a. Remove cosmic ray events, which cause discontinuity and possibly ramp saturation.
    b. Remove dark current, which may be different before and after a cosmic ray hit.
    c. Remove the possible presence of persistence
    d. Fit the ramp of the linearized signal to derive the photon flux.

4.0 Relation with CALWEBB and current NICMOS and WFC3 pipelines

4.1 CALWEBB

Robberto et al. (2008) discussed the pipeline commonalities between observing modes and instruments showing that they could naturally lead to envision a common pipeline architecture shared by all instruments. Robberto et al. (2008) concentrated on the Data Processing Level 2a and presented a possible scheme, called CALWEBB, based on a modified version of the current NICMOS and WFC3 pipeline (Figure 2).
The steps proposed in Section 5 of this report are similar to those envisioned in the CALWEBB concept document. This is reflected in our nomenclature, where we use for each procedure the prefix **CW_** to indicate that this is a CALWEBB (i.e. Level 2a) procedure.

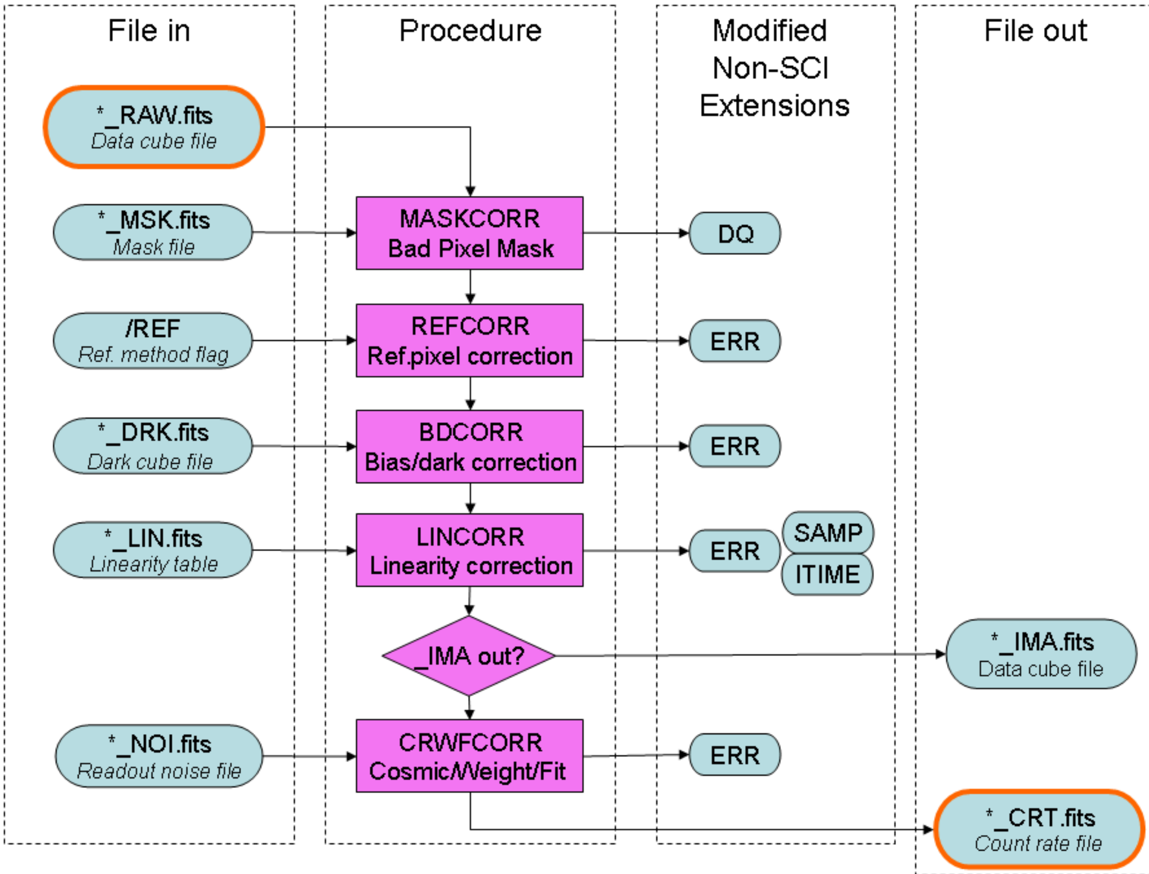

Figure 2 NIRCam Version of the Original CALWEBB (from Robberto et al 2008)

4.2 Exposure Terminology

Table 1 Exposure Terminology

Term	Definition
Sample	One digitized measure of the number of electrons in a pixel. MIR detectors can average 10 consecutive samples for a pixel before proceeding to the next pixel.
Frame	One set of measured values (obtained from 1 or more samples) for every pixel in a detector or in a commanded subarray.
Group	The average of 1, 2, 4, 8 or 16 frames. Groups are stored on the Solid State Recorder (SRR) and downlinked to the ground.
Integration	The sequence of groups between detector resets. Charge accumulated in pixels and is sampled up-the-ramp over the course of an integration.
Exposure	A sequence of repeated integrations with identical parameters, separated by detector resets but no other changes in instrument or observatory configuration.

4.3 Ancillary Data Arrays: current WFC3 status

Before detailing the pipeline, it is important to remember that Science Data Products include not only science data (groups) but also associated ancillary arrays. Building on the NICMOS and WFC3 IR pipeline, we can assume that the science data are associated with ancillary arrays providing, for each pixel, the data quality (DQ) array and the error (ERR) array, ignoring for the moment the Number of Samples (SAMP) and the Integration Time (INT) Array as they can be reconstructed from the DQ arrays. We report hereafter the definitions adopted by WFC3-IR.

4.3.1 WFC3 Data Quality Array (DQ)

This array contains 16 independent flags indicating various status and problem conditions associated with each corresponding pixel in the science image. Each flag has a true (set) or false (unset) state and is encoded as a bit in a 16-bit unsigned integer word. Table 2 lists the WFC3 data quality flags.

Table 2 WFC3 Data Quality Flags

Flag Value	BIT Setting*	Data Quality Condition
0	0000 0000 0000 0000	OK
1	0000 0000 0000 0001	Reed-Solomon Decoding Error
2	0000 0000 0000 0010	Data Replaced by Fill Value
4	0000 0000 0000 0100	Bad Detector Pixel
8	0000 0000 0000 1000	Unstable in Zero-Read
16	0000 0000 0001 0000	Hot Pixel
32	0000 0000 0010 0000	Unstable Response
64	0000 0000 0100 0000	Warm Pixel
128	0000 0000 1000 0000	Bad Reference Pixel
256	0000 0001 0000 0000	Full-Well Saturation
512	0000 0010 0000 0000	Bad or Uncertain Flat Value
1024	0000 0100 0000 0000	(Reserved)
2048	0000 1000 0000 0000	Signal in Zero-Read
4096	0001 0000 0000 0000	Cosmic Ray Detected by Multidrizzle
8192	0010 0000 0000 0000	Cosmic Ray Detected During Up-the-Ramp Fitting
13384	0100 0000 0000 0000	Pixel Affected by Ghost/Crosstalk

* The most significant bit is on the left

4.3.2 WFC3 Error Array (ERR)

This is a floating-point image that contains an estimate of the statistical uncertainty associated with each corresponding science image pixel. It is expressed as a real number of signal units or signal rates (as appropriate for the units of the science image). The values of this array are calculated during calibration with the CALWF3 task, combining detector read noise, Poisson noise in the detected signal and uncertainties from applied calibration reference data.

4.3.3 Comment on Ancillary Data Arrays: difference between JWST and WFC3

Both the DQ and ERR arrays may maintain the same structure and basic conventions used by WFC3, with one exception:
One of the main peculiarities of the JWST data is that they are averaged in groups. This has a unique consequence in what concern Cosmic Rays. In the case of WFC3, a CR hit happens always between "samples" (in the WFC3 jargon). The DQ code with Flag Value 8192 (bit 14) has therefore a unique, non ambiguous meaning: it splits the ramp in two parts, before and after, and no "sample" is lost. In the case of JWST, however, a CR hit may happen also within a group. In this case it is not enough to specify "cosmic ray detected during up-the-ramp fitting", as one must also specify if the group has to be rejected or not. In other words, there are two parameters associated to a CR hit: 1) where it happens and 2) if there is a group to reject. These two keywords can be appropriately coded in the DQ array by adding an extra bit definition (flag). This may require expanding the DQ array to 32 bit.

5.0 NIRCam Data Processing Level 2a Pipeline

5.1 CW_DQI: IR Data Quality Arrays Initialization

PERFORMED TASK
CW_DQI initializes the data quality (DQ) arrays of all IR groups by combining them with a table of known bad pixels for each detector. A DQ array specifies which pixel has to be neglected from further analysis and why.

RATIONALE
Each detector has a fraction of pixels that show anomalous response and cannot be recovered. They must be flagged early-on in the pipeline to avoid biasing data processing e.g. with wrong spatial averages or to prevent numerical errors in the algorithms.

METHOD
This procedure initializes the Data Quality extension of each group by looking at the current reference static DQ tables and telemetry log.

NOTES
1. Different types of bad pixels can be flagged using a code describing the nature of the discrepancy. WFC3 table in Section 4.3.1 provides an excellent example.
2. This function also checks for transient problems occurring e.g. during the readout or data downlink. If so, pixels in the affected readouts are flagged as bad and will not be used in the following steps.

INGREDIENTS
- INPUT DATA:
    - Raw science data ramp

  - Telemetry information
- REFERENCE FILES:
  - Bad pixel static mask

PREVIOUS CALIBRATION ACTIVITIES
Static bad pixels are identified and maintained through instrument monitoring programs and appropriate calibration procedures.

PRODUCTS

1. Raw science data ramp (unchanged)
2. Associated Data Quality (DQ) maps populated with initial values.

5.2 CW_Noise: IR Error Array Initialization

PERFORMED TASK
CW_Noise creates the initial noise arrays accounting for the estimated errors associated with the raw science data.

RATIONALE
Knowledge of the noise associated to each pixel and group is needed for cosmic ray identification and ramp fitting with optimal weighting.

METHOD
This procedure initializes the Noise Map (data cube) and associates to each pixel and group an initial noise value, typically given by the readout noise (single sample, derived e.g. by the Double Correlated Noise divided by square root of 2 and multiplied by the square root of the number of samples in a group).

NOTES
Poisson noise is also present in the data, adding correlation between samples and within groups. It is possible to calculate the combined contribution of readout and Poisson noise, accounting for the covariance terms. This requires some knowledge of the signal rate and will be added later in the pipeline. Other noise sources (e.g. 1/f) may also need to be taken into account and added, at different stages, to the noise map.

INGREDIENTS

- INPUT DATA:
  - Raw science data ramp
- CALIBRATION DATA:
  - Table of readout noise values for each pixel

PREVIOUS CALIBRATION ACTIVITIES
Dark current frames can be used to derive the readout noise for each pixel.

PRODUCTS

1. Initial Noise data cube

5.3 CW_RefPix: IR Bias Level Correction using Reference Pixels

PERFORMED TASK

CW_RefPix removes the group-to-group variation of DC level using the reference pixel located in the four external rows/columns of each HgCdTe SCA.

RATIONALE

Consistently with the assumption that the data reduction pipeline must remove detector anomalies "backward", electronic drifts of the electronic chain are removed first through comparison with reference pixels.

METHOD

This procedure:

1. Reads a table listing the good reference pixels (static mask)
2. Searches and rejects spurious values (CR events) in the reference pixels
3. Calculates an appropriate mean value of the "good" reference pixels, with the corresponding uncertainty, for each group;
4. Subtracts the mean value from each group
5. Adds a constant term to restore the correct dynamic range of the original signal.
6. Updates the noise map of each group.

NOTES

1. The mean value of the reference pixels may depend on the sector ("quadrant"), i.e. on the output channel of the SCA. Higher order corrections accounting for high frequency terms could be needed.
2. The mean value subtracted from each group is also affected by an error which has to be propagated into the error map of the group. The error may depend on the number of reference pixels considered (the inner channels therefore giving larger errors) and on the number of cosmic rays falling on reference pixels (which increases with the integration making this error larger with time).
3. The constant term to be re-introduced to preserve the dynamic range is needed for linearity correction and saturation check. It can be
    a. the average value measured in the first exposure. This can be the very first sample of the ramp, non-coadded in the first group, which by default is saved as a separate frame. In this case the reference pixels track the offset relative to the very first sample which therefore represents a variable "zero" level.
    b. a value fixed a-priori. This can be a constant or, more properly, a quantity depending on the temperature of the device, as it has been shown that the DC level of the output signal depends on the temperature.

Regardless on the strategy adopted, the constant term should be recorded in a keyword (e.g. like MEANBLEV) in the SCI header, (possibly one for each readout channel). All groups of an integration, after the CW_RefPix passage, will have the same mean value of the reference pixels. Large departures from the typical mean level should be flagged and notified, as they may be indicative of system malfunction.

4. It may be convenient to maintain the reference pixel in the frames throughout the rest of the processing, working on images which maintain the original 2048x2048 format until they are combined in a count-rate image. A specific keyword on the data quality may allow the calibration algorithms to deal with (or simply ignore) them.

INGREDIENTS

- INPUT DATA:
  - Raw science data ramp
  - Data quality array
- CALIBRATION DATA:
  - Static map of reference pixels to be used
  - Expected value of reference pixels (average and standard deviation)
- PROCEDURES:
  - CR-rejection procedure

PREVIOUS CALIBRATION ACTIVITIES

A correlation between reference pixel values and temperature may be present. It may need to be calibrated for each detector and channel and accounted for in the algorithm.

PRODUCTS

1. Reference pixel corrected ramp
2. Updated noise map

5.4 CW_Nlin: IR non-Linearity Correction

PERFORMED TASK

CW_Nlin corrects the integrated counts in the science images for the non-linear response of the detector.

RATIONALE

The detector response to the photon flux is intrinsically non linear. A linearity correction has to be applied before ramp fitting.

METHOD

This procedure multiplies the counts measured in each pixel by a factor that depends on

the counts themselves. The factor can be interpolated from a table or derived from an equation.
It is often assumed that the non-linearity correction can be expressed through a polynomial law of the type:

$$F_c = (1 + c_1 + c_2 \times F + c_3 \times F^2 + c_4 \times F^3) \times F \qquad (0.1)$$

where F and F_c are the counts (DN) measured in the pixel before and after the correction and c_1 , c_2 , c_3, and c_4 are the correction coefficients.

NOTES

1. For NICMOS and WFC3, the current form of the correction uses a third-order polynomial, as shown here. In general the algorithm can handle an arbitrary number of coefficients, together with their error estimates, as well as other functional forms.
2. Non linearity correction can be meaningfully performed only up to some extent. It is usual to assume that departures less than 5% from a linear fit to the initial part of the ramp can be corrected. At high signal levels the detector response shows larger deviations that cannot be corrected to the required level of accuracy. This linearity correction threshold level occurs before the physical saturation of the detector. In many cases it is convenient to assume that the saturation level is defined by the amount of departure from linearity that cannot be reliably corrected.
3. Pixels above their saturation values receive no correction, are flagged as saturated in the DQ array for their corresponding readout and ignored in later processing. Examination of the ramp for saturation may proceed from the last group to the first one, stopping when a group less than full is reached.
4. The saturation threshold analysis performed at this point is not complete (see next section).

INGREDIENTS

- INPUT DATA:
    - Reference pixel corrected ramp
    - Data quality array
- CALIBRATION DATA:
    - A table containing the coefficients and their associated errors, for each pixel R
    - Linearity and saturation thresholds for each pixel

PREVIOUS CALIBRATION ACTIVITIES

1. Take flat field ramps up to the saturation threshold in condition of excellent stability of the illuminating source. Take repeated flat fields to improve statistics. Special care must be taken to minimize latency effects, e.g. by taking dark current frames between each saturated ramp.

PRODUCTS

1. Linearity corrected ramp.
2. Updated noise map.
3. Updated data quality table specifying, for each pixel, the reads rejected due to saturation.

5.5 CW_Sat: Saturation Check

CW_Sat identifies saturated data values and flags them, preventing them from being used further in the data reduction process.

RATIONALE

There are two reasons for a specific Saturation Check algorithm: 1) the saturation threshold may be reached between groups averaged into a group. In this case the group has to be rejected even if its value (average) is below the threshold. 2) the saturation threshold may be reached because of a cosmic ray rather than because of the integrated signal. Since CRs affect nearby pixels, it is important to flag the major events properly for CR spatial analysis. Thus a pixel-sample which becomes saturated due to a CR must be recognized as both "CR affected" and "saturated" in the DQ map.

METHOD

This procedure analyzes the raw counts of each pixel and their first difference (derivative) to establish when the saturation threshold has been passed and if a CR is responsible for saturation.

NOTES

1. This procedure is more refined than a simple check on the integrated counts, accounting for the following issues:
    a. In the vicinity of saturation some of the individual frames may have been taken above the saturation threshold. This means that even if the group average remains below saturation, it is not usable. This case can be treated by taking into account the slope of the integrating signal.'
    b. Cosmic Rays suddenly saturate a pixel by releasing a quantity of charges not compatible with the previously accumulated signal.
    c. More generally, the linearity correction is applied to the average of different samples forming a group, but the integrated signal follows a non-linear law. Depending on the number of samples in a group, one could consider having a more accurate (iterative) correction that accounts for the underlying non-linear distribution of individual samples.

INGREDIENTS

- INPUT DATA:
    - Linearity corrected ramps

    - Data quality array
- CALIBRATION DATA:
    - A table containing the coefficients and their associated errors, for each pixel
    - Linearity and saturation thresholds for each pixel

PREVIOUS CALIBRATION ACTIVITIES

1. Needs the same dataset used for the initial linearity correction

PRODUCTS

1. Updated noise map.
2. Updated data quality table specifying, for each pixel, the reads rejected due to saturation

5.6 CW_Dark: IR Dark Image Subtraction

PERFORMED TASK

CW_Dark subtracts the detector dark current from the science data.

RATIONALE

Remove "dark current", including additive spurious contributions to the signal (e.g. thermal, light leaks, etc).

METHOD

This procedure:

1. Opens a "super-dark" calibration file appropriate for the operating temperature and possibly bias of the detector.
2. Reconstruct, if needed, the dark frame to be subtracted by combining adjacent reads averaged in groups.
3. Subtracts the dark current returning a dark subtracted ramp.

NOTES

1. Due to the intrinsic non-linearity of the dark current, as well as to possible reset related effects in the first one or two reads of an exposure, the dark current subtraction may not applied by simply scaling and subtracting a generic dark current rate for each pixel.
2. Step 2 above may be skipped, and the procedure made more efficient, if a library of dark current ramps is maintained for each of the available predefined group sequences (as well as for each of the available sub-array readout modes).
3. This procedure is applied before the ramp has been cleaned of cosmic rays because the dark current is intrinsically non-linear, depends on the detector bias level. Since dark current dominates at low signal level, this is generally ignored and the dark current is measured only at zero level signal, i.e. maximum width of the diode depletion region. However, cosmic ray events may de-bias the junction. Even in the presence of low signal, the dark current before a CR hit the may

therefore be different from the dark current after the CR hit, resulting in a measurable change of slope of the ramp. Ideally the dark current correction should take into account the level of the integrated signal, i.e. it should measure at different bias levels.
4. Error estimates on the subtracted dark current are propagated into the ERR images of all readouts.

INGREDIENTS

- INPUT DATA:
  - Linearity corrected ramps
  - Data quality array
- CALIBRATION DATA:
  - Dark current reference files.

PREVIOUS CALIBRATION ACTIVITIES

1. Take dark current reference files covering all exposure modes, possibly for a range of detector temperatures and detector bias values.

PRODUCTS

1. Dark current subtracted ramps.
2. Updated noise map.

5.7 CW_CRid_TD: IR Cosmic-Ray Identification- time domain

PERFORMED TASK

CW_CRid analyzes each individual pixel ramp to identify and flag pixels suspected of containing cosmic-ray (CR) hits. It looks for large jumps in the ramps due to electronic glitches, which may be both positive a negative in value and may affect only one read.

RATIONALE

Remove cosmic ray events from the accumulating signal.

METHOD

This procedure:

1. For each pixel, all valid groups are analyzed to flag CR events. At least two possible algorithms can be envisioned, depending on the dominating source of noise:
   a. Iteration of a linear fit to the accumulating counts-versus-exposure time relation, flagging major adjacent outliers with opposite sign. This is the type of procedure currently adopted by e.g. the NICMOS pipeline.
   b. Calculation of the first differences and iterates to flag outliers.

NOTES

1. In general, the groups of a multi-accum ramp have a correlation term due to the integrated signal common to the various groups; vice-versa, the set of first differences is uncorrelated in what concerns the signal term but correlated in what

concerns the readout noise, since each readout noise term appears in two differences (except for the first and last group). Therefore, from a statistical point of view, it may be more appropriate to use the fit-to-the-ramp approach in the case of readout noise limited signal, and the first difference analysis in the case of Poisson noise limited signal.

2. To discriminate between the best procedures to use, a simple and robust signal estimate can be made using the difference between the last and first data, divided by the elapsed time. The associated Poisson noise can be compared with the readout noise associated to that pixel, stored in the noise map/table. It will be necessary to read also the electronic gain to convert electrons in ADUs.
3. Regardless on the type of algorithm adopted to detect CR events, the detection threshold t should account for increased Poisson noise and therefore could be set equal to $t = kN\sigma$, with N equal the number of groups and k a parameter specifying the number of standard deviations; some analysis and fine tuning may be needed to fix the optimal value of k, which should also be user selectable. If the appropriate algorithm is adopted, one may assume that the data point follow a gaussian error distribution and therefore the standard deviation has $N-1$ (in the case of fit to the ramp) or $N-2$ (in the case of fit to the first differences) degrees of freedom. Let us assume for example a standard deviation calculated on a group of 20-1 =19 first differences, with 20-2=18 degrees of freedom. At $N = 3$ the confidence level is ~0.3%, i.e. 3 out 1000 cases, or 1 out 333 cases, fall above this threshold. With only 18 groups, we have 18/333=5.4% probability of rejecting a valid data point.
4. The algorithm must also take into account the possibility of having data "partially" contaminated by a cosmic ray. This may happen when the samples, averaged together into a group, have a CR in the middle which biases the group to a value intermediate between the signal before and after the CR event. In this case the ramp will pass from the pre-CR to the post -CR level through an intermediate "step". Linear fitting may find this intermediate step quite close to the optimal "mid point", especially if the CR falls right in the middle of the group. The algorithm should to be robust against this type of occurrence. In the case of first differences, the intermediate step may reduce the amplitude of the jump to the point that it falls below the detection threshold. Simulations will be required to find the best strategy for removing this type of event, knowing that some compromise will be needed to preserve computational efficiency, speed and low probability of rejecting valid data.
5. A group affected by a cosmic ray is entirely lost and must be ignored in the successive line fitting process. Vice-versa, if the CR falls between groups, the ramp will only be truncated, as both groups, before and after the CR, can be used for linear fitting.
6. The program should returns all information required to characterize the two semi ramps defined by a CR event. The ID code can be used to flag CR events.
7. A pixel that, after received a CR hit, becomes saturated should also be qualified in the error map as "CR-affected". This is needed to allow the procedure that

searches spatially for CR hits to analyze this pixel, that otherwise would be skipped.

8. A CR may hit the ramp before the first group is taken. In this case the signal of that pixel will be anomalously higher than the typical zero read value, and the ramp will show an offset at t=0. The algorithm must be able to flag also these cases.

INGREDIENTS

- INPUT DATA:
  - Linearity corrected ramps
  - Data quality array
  - ID Codes
  - Flag indicating the type of CR algorithm to be used
  - Rejection threshold
- CALIBRATION DATA:
  - none

PREVIOUS CALIBRATION ACTIVITIES

none

PRODUCTS

1. Updated DQ codes, flagging the presence of a CR in the pixel.

5.8 CW_CRid_SD: IR Cosmic-Ray Identification- space domain

PERFORMED TASK

CW_CRid_SD analyzes the ramps of pixels next to pixels affected by a cosmic ray. It looks for not-so-large jumps that may have been missed by the first "time-dependent" CR finding algorithm due to high threshold values.

RATIONALE

Remove cosmic ray events from the accumulating signal performing a spatial search

METHOD

This procedure:

1. Points to the subset of pixels affected by CR found through time domain analysis;
2. Analyzes the ramps of the eight adjacent pixels in correspondence of CR event using a more conservative rejection threshold
3. Flags the pixels-groups that appear to be affected by the CR event.

NOTES

1. The routine may look for more than the adjacent 4 or 8 pixels, trying e.g. to flag CR entering nearly parallel on the detector.
2. Knowledge about the Inter-Pixel Coupling (IPC) may allow predicting the counts expected in the pixels adjacent a CR event and thus refine the threshold definition.

INGREDIENTS

- INPUT DATA:

    - Linearity corrected ramps
    - Data quality array
    - ID Codes
    - Flag indicating the type of CR algorithm to be used
    - Rejection threshold
    - Information about IPC
- CALIBRATION DATA:
  none

CALIBRATION ACTIVITIES
Analysis of CRs events (or other methods) to quantify IPC

PRODUCTS
1. Updated DQ array.

5.9 CW_Pers: Persistence analysis

PERFORMED TASK
CW_Pers analyzes the ramp for persistence (latency) effects and flags pixels to be excluded from the ramp fitting process.

RATIONALE
Persistence affects mostly the early groups of a ramp, which can therefore be neglected to improve the reliability of the ramp fitting process.

METHOD
At the moment this procedure is not implemented in the NICMOS/WFC3 pipeline, so we only envision possible methods. Two approaches seem most viable:
1. If a persistence map can be obtained, e.g. on the basis of a dark frame taken immediately before beginning the integration or from the previously acquired datasets, it may be possible to pre-determine the number of initial groups to be excluded. This requires an absolute calibration of the relation between decay time vs. illumination, for each pixel.
2. If a persistence map cannot be obtained, or if the calibration map cannot be made accurate enough, one may fit the early groups of the ramp with a curve instead of the (nearly) straight line predicted by theory.

NOTES
1. In the first method a persistence map may be difficult to obtain for the very first ramp of an observing program for a variety of reasons. In this case, one may assume that enough time has elapsed from the previous program that persistence is negligible, or take by default an adequate number of darks in order to model the decay time in the following frames.

INGREDIENTS
Depending on the method, one may need table files of decay times, polynomial

coefficients for the decay, etc.

PREVIOUS CALIBRATION ACTIVITIES
Analysis and definition of the optimal strategy requires a large variety of images taken in conditions of different illuminations (flat fields, stellar fields) and darks.

PRODUCTS

1. Update DQ map table flagging initial pixels to be neglected for ramp fitting.

5.10 CW_RampFit: Ramp fitting

PERFORMED TASK
CW_RampFit calculates the slope for each pixel, assuming that the ramp (or set of subramps)
has been linearized.

RATIONALE
Extract the count rate image and associated noise. This completes the first part of the pipeline.

METHOD
This procedure:

1. Look at the table of the valid pixels and isolates the relative sub-ramps.
2. Perform a slope calculation with optimal weighting through a linear fit to the ramp.
3. The formal error on the slope is calculated and the slopes averaged with the correct weights to determine the final values.

NOTES

1. The noise associated to each group is a combination of detector read noise and Poisson noise in the signal, plus all the errors introduced by the calibration steps. A proper estimate of the Poisson noise depends on the signal rate, and a correct estimate must account for the correlation of the signal between successive reads. See Robberto (2009) for a treatment of the total error in the case of grouped samples.
2. The ramps or semi-ramps are individually reduced using an estimate of the uncertainties based on the analysis of Fixsen et al (2000) and Regan (2007).
3. In the case of multiple slopes, the slope errors are used as optimal weights to derive the final estimate of the slope, following Regan (2007) and Robberto (2008).
4. The procedure also calculates the effective integration time for each pixel, estimated on the basis of the valid data points used for the fit. A TIME frame may be produced mapping the exposure time.
5. This procedure should be able to return as fits files also the groups BEFORE the linear fitting. It is important to have the capability of extracting groups before the

linear fit in order to create e.g. the dark frames needed by the pipeline.

INGREDIENTS

- INPUT DATA:
  - Linearity corrected ramps
  - Data quality array
- CALIBRATION DATA:
  - none

CALIBRATION ACTIVITIES

none

PRODUCTS

1. Slope image
2. Intercept image
3. Noise image.
4. Data Quality map remains unaffected
5. Bad pixels map remains unaffected
6. Saturated groups map remains unaffected

6.0 Summary

The processing steps envisioned in this report and the Calibration files/tables.

6.1 Data Processing Level 2a: list of NIRCam (CALWEBB) procedures

1. CW_DQI: IR Data Quality Arrays Initialization
2. CW_Noise: IR Error Array Initialization
3. CW_RefPix: IR Bias Level Correction using Reference Pixels
4. CW_Nlin:- IR Non-Linearity Correction
5. CW_Sat: Saturation Check
6. CW_Dark: IR Dark Image Subtraction
7. CW_CRid_TD: IR Cosmic-Ray Identification - time domain
8. CW_CRid_SD: IR Cosmic-Ray Identification - space domain
9. CW_Pers: Persistence analysis
10. CW_RampFit: Ramp fitting

6.2 Data Processing Level 2a: list of Calibration files.

1. Bad Pixel Mask (DQ static)
2. Saturation Level Array
3. Non linearity coefficients and thresholds
4. Library of dark Arrays
5. Readout noise array
6. Reference pixel offset level
7. Persistence coefficient arrays

7.0 Conclusion

This document describes the processing steps that compose the Data Processing Level 2a (CALWEBB) for NIRCam. It is intended to provide enough information to allow setting the requirements for the development of the pipeline, including the needed calibration data, keywords and products. It envisions 10 steps, which include the initial definition of Data Quality and Error arrays, reference pixels, non-linearity, saturation, dark current, cosmic rays (both in time and spatial-domain), persistence and optimal ramp fitting. I propose a number of modification to the current WFC3 pipeline, including the need of an extra Data Value in the JWST Data Quality images needed to properly account for Cosmic Ray events happening within a group of averaged samples. This document may provide a convenient baseline against which the various instrument teams may find commonalities and differences with regard to Data Processing Level 2a.