

The STScI NICMOS Pipeline: CALNICA, Single Image Reduction

Howard Bushouse, Chris Skinner, John MacKenty
June 10, 1996

ABSTRACT

This ISR describes CALNICA, the single-image reduction portion of CALNIC, which produces individual calibrated images. We include a brief reminder of the NICMOS operational modes and data formats, then a description of the algorithmic steps in CALNICA, and then the reference files used by CALNICA.

1. Introduction

In this ISR we describe the steps in CALNICA, the portion of the NICMOS pipeline which reduces and calibrates single images. The output of CALNICA is an image that has had all instrumental signatures removed and in which the photometric header keywords have been populated. NICMOS data file formats and contents used in the pipeline are described in a general way in ISR NICMOS-002. The overall structure and functionality of the CALNIC pipeline, which will comprise of two parts (CALNICA, described here, and CALNICB, described in a forthcoming ISR), is described in ISR NICMOS-008.

The remainder of this ISR briefly reviews the various NICMOS operational modes and the data they produce, the NICMOS data file contents, and CALNICA design issues. It also describes the individual algorithmic steps in CALNICA and the calibration reference files and tables used by the algorithms.

2. NICMOS Operational Modes

NICMOS operates in four modes: ACCUM, MULTIACCUM, RAMP and BRIGHTOBJECT.

In ACCUM mode, a single exposure is taken. The array is initially read out (non-destructively) some number of times n , defined by the observer, and for each pixel the mean of the resulting values is calculated on-board. The array then integrates, for a time chosen by the observer from a pre-defined menu. Finally, the array is again read out non-destructively n

times and the average calculated on-board. The difference between initial and final values is the signal which will be received on the ground. This is the simplest observing mode.

MULTIACCUM mode allows only a single initial and final read. However, a number n of intermediate non-destructive reads are made during the course of the exposure. The temporal spacing of the readouts may be linear or logarithmic, and is freely selectable by the observer. In this mode, the initial read (known hereafter as the zeroth read) is *not* subtracted on-board from the subsequent reads, so this will have to be done on the ground. The total number of reads *after* the zeroth read may be up to 25 (and it should be noted that the number n defined by the observer is the number *after* the zeroth read). Note that the final read will contain the signal from the *entire* exposure time, *not* just the interval between the last and the penultimate reads. The result of each of the $n+1$ reads will be received on the ground.

RAMP mode makes multiple non-destructive reads during the course of a single exposure much like MULTIACCUM, but only a single image is sent to the ground. The results of each successive read are used to iteratively calculate a mean count-rate and an associated variance for each pixel. A variety of on-board processing can be employed to detect or rectify the effects of saturation or Cosmic Ray hits. The data sent to the ground comprises a single image, plus an uncertainty and the number of valid samples for each pixel. For each pixel the science data image contains the mean number of counts per ramp integration period.

BRIGHTOBJECT mode provides a means to observe objects which ordinarily would saturate the arrays in less than the minimum available exposure time (which is of order 0.2 seconds). In this mode each individual pixel is successively reset, integrated for a time defined by the observer, and read out. Since each quadrant contains 16,384 pixels, the total elapsed time to take an image in this mode is 16,384 times the exposure time for each pixel. The result is a single image just like that produced in ACCUM mode.

3. NICMOS Data File Contents

The data from an individual NICMOS exposure (where MULTIACCUM observations are considered a single exposure) is contained in a single FITS file. The data for an exposure consists of five arrays, each stored as a separate image extension in the FITS file. The five data arrays represent the science (SCI) image from the FPA, an error (ERR) array containing statistical uncertainties (in units of 1σ) of the science data, an array of bit-encoded data quality (DQ) flags representing known status or problem conditions of the science data, an array containing the number of data samples (SAMP) that were used to calculate each science image pixel, and an array containing the effective integration time (TIME) for each science image pixel.

This set of five arrays is repeated within the file for each readout of a MULTIACCUM exposure. However, the order of the images is such that the result of the longest integration time occurs first in the file (*i.e.* the order is in the opposite sense from which they are obtained).

In RAMP mode, as the data from successive integrations are averaged to calculate the count rate,

so the variance of the rate for each pixel is calculated. The data sent to the ground includes the count rate and also this variance. The variance as calculated on-board will be read into the error array in the Generic Conversion stage on the ground, and must eventually be converted to a standard deviation for consistency with data from other modes. It should be noted that the variance calculated in RAMP mode will not always be consistent with the manner in which uncertainties are calculated on the ground in other modes. For instance, in the case where the number of RAMP integrations in an exposure is rather small, the variance calculated in this manner may not deliver a good estimate of the statistical uncertainty of the data. In the first build of CALNICA we will simply propagate the on-board variance as our uncertainty in RAMP mode, but in future builds further consideration will be given to the possibility of making the uncertainties consistent with those which we calculate in other modes.

4. CALNICA Design Goals

All processing steps of CALNICA will treat the five data arrays associated with an exposure as a single entity. Furthermore, each step that affects the science image will propagate and update, as necessary, the error, data quality, number of samples, and integration time arrays.

In order to easily facilitate the common treatment of the five data arrays within any step of the pipeline, the format of all input and output data files - including most calibration reference images - will be identical. The commonality of input and output science data file formats also allows for partial processing and the reinsertion of partially processed data into the pipeline at any stage. In other words, the pipeline will be fully 're-entrant'.

MULTIACCUM exposures will generate a single raw data file containing $5(n+1)$ extensions, where n is the number of readouts requested by the observer. When this is processed by CALNICA, there will be two output files. The first will consist of a set of $5(n+1)$ extensions, each of which may have had some processing, and can be considered as an intermediate file. The second will contain just 5 extensions, like the output from all other modes, and will represent the result after the final or n th read. An exception will be the case where the 'Take Data Flag' is lowered during a MULTIACCUM exposure. This may indicate, for example, that the spacecraft has lost its lock on guide stars and is no longer pointing at the desired target. In this case NICMOS will continue to take data, as experience with other instruments has shown that useful data may still be acquired when a TDF exception occurs. The data will all still be processed and written to the intermediate file. However, the final output file will be the result of the exposure only up to the time of the last readout *before* the TDF exception occurred.

Processing control will be provided by a set of "switches" consisting of keywords in the science data file header. Two keywords associated with each processing step will be used to 1) determine whether or not a given step is to be performed, and 2) record whether or not a step has already been performed. This means that CALNICA will be aware of what processing has already been done to a given set of data, and so will be able to issue warnings if the user is requesting the same processing step to be done more than once (which would often, but not always, be undesirable).

The CALNICA code will consist of a series of reusable modules, one for each significant step as defined in “Algorithmic Steps in CALNICA” on page 4 (e.g. flat-field correction, photometric calibration, data I/O...). The code will be written in ANSI C, which will allow the code to be portable to a number of different programming environments (e.g. IRAF, IDL, MIDAS, Figaro...), as well as rendering it more easily maintainable. Infrastructure, to tie the modules of the code together in a pipeline, will be written for the IRAF environment, so that CALNICA will run as part of the STSDAS package as with all the other HST instruments. We anticipate that the NICMOS Instrument Definition Team (IDT) will probably write an equivalent IDL infrastructure, so that the same code will run as a pipeline in IDL. We are developing all the C code modules in collaboration with the IDT, so that although the pipelines will be run from different environments by STScI and the IDT, both groups should be using identical algorithms and code. I/O to and from the pipeline code is being minimised, and restricted as far as possible to the beginning and end of run time, in order to maximise the ease of portability of the code to different environments.

5. Algorithmic Steps in CALNICA

Figure 1 shows the flow of CALNICA processing steps. Table 1 gives a summary list of the steps, the name of the “switch” keyword that controls each step, and the reference file keyword (if any) associated with each step. Each step is described in more detail below.

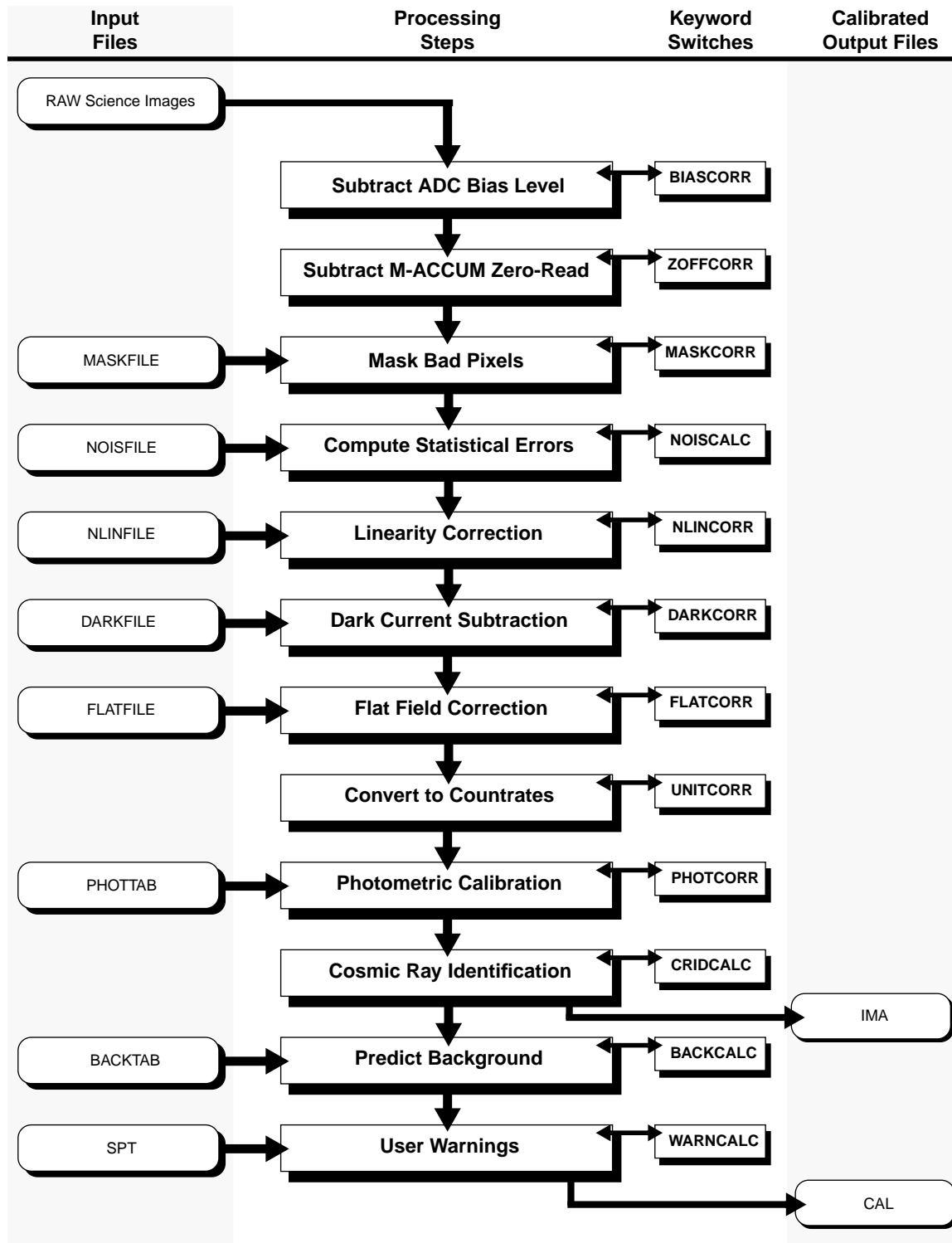
DATA I/O

Pipeline processing obviously must always commence by reading in the data file. This step will also entail tests of the validity of the data. Some error conditions encountered here, such as if the data are found to all be negative (which would imply an error condition in the SI) will cause the pipeline to terminate. Others will generate warning messages or prompt specific non-standard behaviour. If CALNICA determines upon reading in a data file that a given science data extension consists entirely of zeroes, appropriate warnings will be issued, and no calibrated output generated. So, for MULTIACCUM observations, n extensions of each type (science data, error, DQ, samples and time) will be written into the calibrated output file, and the first extension encountered in the output file will be the result of the last *valid* readout.

BIASCORR

NICMOS uses 16-bit Analog-to-Digital Converters (ADCs), which convert the analog signal generated by the detectors into signed 16-bit integers. Because the numbers are signed, a zero signal will not generate a zero number in the output from the instrument, but instead will generate a large negative number. Just how large this number is can be adjusted from the ground, although it is not accessible to GOs. This number is the ADC zero level, or bias, which must be removed before any processing is done. This removal is carried out in the BIASCORR step, which is essential unless the image was taken in MULTIACCUM mode and the zeroth-read subtraction (ZOFFCORR) is carried out, in which case BIASCORR is not necessary - applying it anyway will not be harmful

Figure 1: Pipeline Processing by CALNICA



though.

No reference files are used by this step. The value of the ADC zero level is determined from the “ADCZERO” header keyword.

ZOFFCORR

A NICMOS image is formed by taking the difference of two non-destructive detector readouts. The first read occurs immediately after resetting each pixel in the detector. This records the starting value of each pixel and defines the beginning of an integration. The second non-destructive read occurs after the desired integration time has elapsed and marks the end of the exposure. The difference between the final and initial reads results in the science image.

As mentioned before, in all instrument observing modes except MULTIACCUM, the subtraction of the two reads is performed on-board. In MULTIACCUM mode the zeroth read is returned along with all subsequent readouts and the subtraction must be performed in the ground system pipeline. The pipeline will subtract the zeroth read image from all readouts, including itself. The self-subtracted zeroth read image will be propagated throughout the remaining processing steps, and included in the output products, so that a complete history of error estimates and DQ flags is preserved.

No reference files are used by this step.

MASKCORR

Flag values from the static bad pixel mask file are added to the DQ image. This uses the MASKFILE reference file, which contains a flag array for known bad pixels. There will be one MASKFILE per detector, containing the bad pixel mask in the DQ image of the file.

NOISCALC

Statistical errors will be computed in order to initially populate the error image. The error values will be computed from a noise model based on knowledge of the detector characteristics. The noise model will be a simple treatment of photon counting statistics and detector readout noise, i.e.:

$$\sigma = \sqrt{\sigma_{rd}^2 / nread + counts \cdot adcgain / adcgain}$$

where σ_{rd} is the read noise in units of electrons, $nread$ is the number of initial/final readouts, $counts$ is the signal (in units of DN), and $adcgain$ is the electron-to-DN conversion factor.

This calculation is not needed for RAMP mode data because a variance image computed on-board is recorded in the raw error image by the OPUS generic conversion process. In this case the NOISCALC step simply computes the square-root of the raw error image so that the variance values become standard deviations.

This step will use the NOISFILE reference file, which will be an image containing the read noise (pixel-by-pixel) for a given detector. There will be two NOISFILEs per detector, one for each read rate. Data quality (DQ) flags set in the NOISFILE will be propagated into the science data.

Table 1. CALNICA Calibration Switches and Reference Files

Switch Keyword	Processing Performed	File Keyword	Reference File Contents
BIASCORR	Subtract ADC bias (zero) level	N/A	N/A
ZOFFCORR	Subtract M-ACCUM zero read	N/A	N/A
MASKCORR	Mask bad pixels	MASKFILE	bad pixel flag image
NOISCALC	Compute statistical errors	NOISFILE	read-noise image
NLINCORR	Linearity correction	NLINFIL	linearity coefficients image
DARKCORR	Dark subtraction	DARKFILE	dark current images
FLATCORR	Flat field correction	FLATFILE	flat field image
UNITCORR	Convert to countrates	N/A	N/A
PHOTCORR	Photometric calibration	PHOTTAB	photometric parameters table
CRIDCALC	Identify cosmic ray hits	N/A	N/A
BACKCALC	Predict background	BACKTAB	background model parameters
WARNCALC	User warnings	N/A	N/A

NLINCORR

The linearization correction step corrects the integrated counts in the science image for non-linear response of the detectors. The correction is represented using a piece-wise continuous function for each pixel, where each of the three pieces has a quadratic form. Coefficient error matrix values are used to estimate the uncertainty of the corrected pixel values. For a quadratic linearity correction of the form

$$C' = a_1 + a_2 \cdot C + a_3 \cdot C^2$$

the error estimate is

$$\sigma^2 = \epsilon_{11} + C^2 \cdot \epsilon_{22} + C^4 \cdot \epsilon_{33} + 2 \cdot (C \cdot \epsilon_{12} + C^2 \cdot \epsilon_{13} + C^3 \cdot \epsilon_{23})$$

where C is the uncorrected science pixel value, C' is the corrected value, a_n is the n th polynomial coefficient, and ϵ_{nm} is the n by m term of the error matrix. This uncertainty is propagated (in quadrature) with the input error value for each pixel. Pixel values that are above the saturation limit will be flagged with the saturation DQ value.

This step uses the NLINFIL reference file, which consists of a set of images containing function

coefficients at each pixel location and does not, therefore, conform to the standard NICMOS data file format. The coefficients will be stored as a set of image arrays, using one image per coefficient. For a three-piece quadratic function, the NLINFILE will contain a total of nine coefficient images. Three variance and three co-variance arrays are needed for each function piece, yielding a total of eighteen error arrays in the NLINFILE. One data quality array will be used for each function piece in order to warn of poor quality fits for any pixel. Three arrays will also be used to store the function crossover (or “node”) values. Two of these node arrays will contain the crossover values from the first to the second, and the second to the third function pieces, while the third array will contain the saturation value for each pixel. This gives a grand total of thirty-three image arrays in the NLINFILE (nine coefficients, eighteen errors, three data quality, and three node arrays). There will be one NLINFILE per detector.

DARKCORR

The detector dark current will be removed by subtracting a dark current reference image appropriate for the exposure time of the science data. This step will be skipped for BRIGHTOBJ mode observations because the very short exposure times result in insignificant dark current.

A simple scaling of a single dark reference image to match the exposure time of the science data is, unfortunately, not possible due to the non-linear behavior of the dark current as a function of time. Therefore a library of dark current images will be maintained, covering a range of “typical” exposure times and observers will be encouraged to obtain observations at these exposure times as much as possible. When the exposure time of a science image matches that of one of the library darks, then a simple subtraction of the reference image will be performed. When the science image exposure time does not match any of those in the reference library, a suitable dark current image will be constructed in the pipeline by interpolation of the reference images. The type of interpolation that will be necessary has not yet been determined, but will most probably eventually be a spline due to the highly non-linear behaviour of the dark current. However, the initial release of CALNICA will use linear interpolation, and the form of the interpolation will be altered in subsequent releases as we receive better information about the behaviour of the dark current from the IDT. In the case of observations with integrations longer than the longest library dark, a simple linear extrapolation of the temporal behaviour of the dark current will be adopted. The IDT’s current understanding of the detectors suggests that for long integration times this should be entirely adequate.

There will be one reference file (DARKFILE) per detector, which will contain the entire set of dark images (at various exposure times) for that detector. The pipeline will determine whether or not a dark image exists that matches the science image exposure time and, if not, will interpolate amongst the images to form an appropriate dark image.

FLATCORR

In this step the science data are corrected for variations in gain between pixels by multiplying by

an (inverse) flat field reference image. This step will be skipped for observations using a grism because the flat field corrections are wavelength dependent. This step uses the FLATFILE reference file, which contains the flat field image for a given detector and filter (or polarizer) combination. There will be one FLATFILE per detector/filter combination.

UNITCORR

The conversion from raw counts to count rates will be performed by dividing the science image data by the exposure time. No reference file is needed. For RAMP mode the exposure time will be the time per ramp integration, rather than the total exposure.

PHOTCORR

This step provides photometric calibration information by populating the photometry keywords PHOTFLAM, PHOTZPT, PHOTPLAM, and PHOTBW with values appropriate to the camera and filter combination used for the observation. The photometry parameters will be read from the PHOTTAB reference file, which will be a FITS binary table containing the parameters for all observation modes.

CRIDCALC

This step will identify and flag pixels suspected to be affected by cosmic ray hits. The algorithm is not yet defined for single (ACCUM/BRIGHTOBJECT/RAMP) images, but for MULTIACCUM data CRs can be recognised by seeking large changes in count rate between successive images (as is done on-board in RAMP mode). Such an algorithm is being implemented in the second build of CALNICA.

For single-image modes, identification of a cosmic ray hit will only result in setting the cosmic ray DQ flag value for that pixel in the output calibrated (“cal”) image. The science (SCI) image pixel values will *not* be modified. In MULTIACCUM mode, identification of a cosmic ray hit will similarly result in the setting of a DQ flag in the appropriate group of the intermediate (“ima”) output file, but no science image pixel values will be modified in the intermediate file. In the process of creating the final calibrated (“cal”) image for MULTIACCUM mode, however, only good (unflagged) pixels will be used. The samples (SAMP) and integration time (TIME) arrays in the MULTIACCUM calibrated (“cal”) file will reflect the actual number of samples, and their total integration times, used to compute the SCI image and the DQ image in the “cal” file will also reflect the fact that only good data was used.

BACKCALC

This step will compute a predicted background (sky plus thermal) signal level for the observation based on models of the zodiacal scattered light and the telescope thermal background. The results will be stored in header keywords. This step will use the BACKTAB reference file, which will be a FITS binary table containing model background parameters.

USER WARNINGS

In this step various (TBD) engineering keyword values will be examined and warning messages to the observer will be generated if there are any indications that the science data may be compromised due to unusual instrumental characteristics or behavior (such as unusual excursions in some engineering parameters during the course of an observation). No reference files will be used.

STATISTICS

Statistics of the SCI image data will be computed for all CALNICA output files. The quantities computed will be the mean, standard deviation, minimum, and maximum of the good (unflagged) pixels in the whole image, as well as each quadrant. The computed values will be stored as SCI image header keywords.