This chapter describes the pipeline calibration system developed at STScI. This system provides observers with NICMOS data after various instrumental signatures are removed, conversions to flux units are performed, and patterns of exposures are combined. Several enhancements to the HST ground system have been made to support NICMOS, including the concept of associations of datasets and an improved file format for data storage and distribution. A detailed description of the analysis of HST data in general and NICMOS data in particular can be found in the *HST Data Handbook*.

**Overview**

All data taken with NICMOS are automatically processed and calibrated by a suite of software programs known as the *pipeline*. The purpose of pipeline processing is to provide data products to observers and the HST Data Archive in a form suitable for most scientific analyses. Pipeline processing is also applied to engineering data and calibration data.

The basic sequence of steps in the STScI pipeline system (also known as OPUS) is:

1. Assemble data received from HST into datasets.
2. Perform a standard level of calibration of the science data.
3. Store both the uncalibrated and calibrated datasets in the Archive and populate the Archive database catalog to support StarView queries.
The pipeline must also handle exceptions (e.g., incomplete data) and perform a general data evaluation and quality control step. Final delivery of data to observers is accomplished by the data distribution mechanisms of the Archive system.

The calibration step has several goals:

- Remove the known instrumental signatures (e.g., flat field and dark current).
- Correct the data for non-linear behavior and convert to physical units (e.g., gain and flux calibration).
- Flag degraded or suspect data values and provide estimates of the statistical uncertainties of each pixel.

While a calibration pipeline may not be able to provide the optimal calibration for a specific observation (which may, in fact, not become available until some time after the data were obtained and calibrated), the goal is to provide data calibrated to a level suitable for initial evaluation and analysis for all users. Observers frequently require a detailed understanding of the calibrations applied to their data and the ability to repeat, often with improved calibration products, the calibration process at their home institution. Further, certain types of image artifacts can appear in NICMOS data, which require processing with specialized tools to remove. To support these goals, the calibration software is available within the IRAF/STSDAS system and the calibration reference files (e.g., flat fields) are available from the HST Archive via StarView so that observers have the ability to repeat and customize the calibration processing to meet the specific needs of individual observations.

**Associations**

To improve the utility of the pipeline processing for the second generation science instruments—NICMOS and STIS—several significant changes were made to the structure of the calibration pipeline. The largest of these changes was to enable the combination of multiple observations during the calibration process. This permits the pipeline to both generate a combined product and to use calibrations obtained contemporaneously with the science observations. This capability is designed to support the cosmic ray event removal, mosaicing, and background subtraction for NICMOS observations. As discussed in Chapter 11, mechanisms exist for compactly requesting such observations in the Phase II proposal.

**Concept**

The basic element in the HST ground system has historically been the *exposure*. The first generation HST science instruments were commanded to generate single exposures, which result from a recognizably distinct
sequence of commands to the instrument. This creates a flow of data which is assembled into a single dataset. Each dataset is given a unique 9 character identifier (an IPPPSSOOT in STScI terminology) and is processed by the pipeline, calibrated, and archived separately from all other datasets.

An illustrative (partial) counter example to this procedure is the WFPC2 CRSP LSTM proposal instruction. This results in two WFPC2 exposures from a single line on the exposure logsheet (the way in which observers specify commands for HST). However, the HST ground system treats a CRSP LSTM as two distinct exposures which are commanded, processed, calibrated, and archived separately. The pipeline does not combine these two images (datasets) to create the single image without cosmic ray events which was the observer’s original intention. Currently, the observers (and any future archival researchers) are left to perform this task on their own.

The second generation instruments present many instances in which the combination of data from two or more exposures is necessary to create a scientifically useful data product. Both NICMOS and STIS need to combine exposures to remove cosmic rays and to improve flat fielding (by dithering). For NICMOS, the HST thermal background contributes a significant signal at wavelengths greater than 1.7 \( \mu m \). Multiple exposures (dithered for small targets and offset onto blank sky—chopped—for extended targets) are necessary to measure and remove this background.

**Usage**

Associations exist to simplify the use of HST data by observers. This starts from the proposal phase, continues with a more complete calibration process than would be possible without associations, carries into the archiving and retrieval of associated data, and includes the use of HST data by observers within the IRAF/STSDAS system.

An association is a set of one or more exposures along with an association table and, optionally, one or more products. We define the following terms:

- An exposure is the atomic unit of HST data.
- A dataset is a collection of files having a common rootname (same IPPPSSOOT).
- A product is a dataset derived from one or more exposures.

The first generation instruments all had a one-to-one correspondence between exposures and datasets. They do not have products. NICMOS and STIS use the association structure as a meta-dataset. Further, they use the information in multiple exposures during the calibration process to create products.

From a high level, an association is a means of identifying a set of exposures as belonging together and being, in some sense, dependent upon one another. The association concept permits these exposures to be
calibrated, archived, retrieved, and reprocessed (within OPUS or STSDAS) as a set rather than as individual objects. In one sense, this is a book-keeping operation which has been transferred from the observer to the HST data pipeline and archive.

Associations are defined by optional parameters on a single exposure logsheet line. That is, there is a one-to-one correspondence between proposal logsheet lines and associations (although it is possible to have exposures which are not in associations).

Observers may obtain one or more associations at each of one or more positions on the sky using the NICMOS proposal grammar. Typically usage will be:

- To obtain a sequence of slightly offset exposures (dithering) to improve the flat fielding, avoid bad pixels and cosmic rays, and, for sufficiently compact targets, to remove the thermal background signal.
- Mapping of targets larger than the NICMOS detector's field of view.
- To obtain a sequence of observations in which the telescope is chopped between the target and one or more offset regions of (hopefully blank) sky.

A set of predefined patterns are provided in the proposal instructions for these types of observations or a combination of both types (Chapter 11). The Institute ground system will expand the observer’s single line request into multiple exposures each with its own identifying name (IPPPSSSOOT) and populate the necessary database relations to describe this association for the OPUS system.

### Re-engineering

For the second generation science instruments the format of the data products from the pipeline is FITS (Flexible Image Transport System) files with image extensions. The IRAF/STSDAS system was modified to operate directly on these files. Each NICMOS image is expressed as a set of five image extensions representing the image, its variance, a bit encoded data quality map, the number of valid samples at each pixel, and the integration time at each pixel. This structure is used at all stages of the calibration process which permits the re-execution of selected elements of the pipeline without starting from the initial point. Finally, the calibration code itself is written in the C programing language (rather than IRAF’s SPP language). This greatly simplifies the modification of the pipeline code by users and the development of new NICMOS specific data processing tasks.
The NICMOS calibration task is divided into two stages: **calnica**, which is used for every individual exposure, and **calnicb**, which is used after **calnica** on those exposures which comprise an association.

**Static Calibrations—calnica**

The first calibration stage, **calnica**, (Figure 12.1) performs those calibrations which can be done to a single exposure using the configuration information from its telemetry and the Calibration Data Base. Such calibrations are derived from the calibration program (see Chapter 13) and typically change on time scales of months. This is analogous to the WFPC2 calibration process (**calwp2**). **Calnica** performs the following steps:

- Correct for signal present in the zeroth read (in MULTIACCUM data)
- Subtract the bias level (in MULTIACCUM data).
- Flag known bad pixels in the data quality array.
- Calculate a noise model for each pixel.
- Subtract dark current.
- Correct for non-linearity.
- Flat field to bring each pixel to a common gain.
- Convert the image data to count rate units.
- Calculate various image statistics (e.g., median).
- Store photometric calibration information in image header keywords.
- Correct for cosmic ray events and pixel saturation (in MULTIACCUM data).
- Calculate estimates of the background.
- Analyze the internal engineering telemetry for potential problems with the observation.
Observers will be given both the uncalibrated (raw) data and the processed data for each exposure. For MULTIACCUM observations, partially calibrated data for each readout will be generated (which excludes the cosmic ray and saturation corrections), in addition to a final single image.

To recalibrate NICMOS data, observers will need the calnica software (included in the STSDAS distribution) and the necessary calibration reference files (available from the HST Data Archive using StarView and from the NICMOS WWW pages).

The data processing flow chart for normal imaging and spectroscopic images is shown in Figure 12.2.
Figure 12.2: Calibration Steps of the calnica Pipeline

<table>
<thead>
<tr>
<th>Input Files</th>
<th>Processing Steps</th>
<th>Keyword Switches</th>
<th>Calibrated Output Files</th>
</tr>
</thead>
<tbody>
<tr>
<td>RAW Science Images</td>
<td>Zero-Read Signal Correction</td>
<td>ZSIGCORR</td>
<td>CAL</td>
</tr>
<tr>
<td></td>
<td>Subtract Zero-Read Image</td>
<td>ZOFFCORR</td>
<td></td>
</tr>
<tr>
<td>MASKFILE</td>
<td>Mask Bad Pixels</td>
<td>MASKCORR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wrapped Pixel Correction</td>
<td>BIASCORR</td>
<td></td>
</tr>
<tr>
<td>NOISFILE</td>
<td>Compute Statistical Errors</td>
<td>NOISCALC</td>
<td></td>
</tr>
<tr>
<td>DARKFILE</td>
<td>Dark Current Subtraction</td>
<td>DARKCORR</td>
<td></td>
</tr>
<tr>
<td>NLINFILE</td>
<td>Linearity Correction</td>
<td>NLINCORR</td>
<td></td>
</tr>
<tr>
<td>FLATFILE</td>
<td>Flat Field Correction</td>
<td>FLATCRR</td>
<td></td>
</tr>
<tr>
<td>PHOTTAB</td>
<td>Convert to Count rates</td>
<td>UNITCRR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Photometric Calibration</td>
<td>PHOTCALC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cosmic Ray Identification</td>
<td>CRIDCALC</td>
<td></td>
</tr>
<tr>
<td>BACKTAB</td>
<td>Predict Background</td>
<td>BACKCALC</td>
<td></td>
</tr>
<tr>
<td>SPT</td>
<td>User Warnings</td>
<td>WARNCALC</td>
<td></td>
</tr>
</tbody>
</table>
Contemporaneous Observations—calnicb

While previously it has been possible to execute multiple exposures from a single proposal logsheet line (e.g., WFPC2, CR-SPLIT, and NEXP=n constructs), this capability was significantly expanded to support requirements of the second generation science instruments. Typical examples include the removal of cosmic rays, the construction of a mosaic image, and the subtraction of the sky background from a sequence of on-target and off-target observations. These observations are distinguished by the fact that their calibration and processing depends upon other observations obtained at the same time.

The calnicb part of the pipeline carries out the calibration and merging of associated data frames, each of which has first been processed by calnica. In the case shown in Figure 12.3, the associated set has 3 individual datasets that are combined into one merged and calibrated dataset.

We refer to these sets of exposures as associations. The calnicb task operates on an entire association, and produces one or more products from that set (Figure 12.3). In the case of dither patterns, calnicb reads from the headers of the individual exposure files what the telescope offsets were. It then identifies sources in the images, and, starting with the telescope pointing information from the headers as an initial guess, determines what the pointings actually were. It then combines the images into a final mosaic, rejecting from the output any cosmic rays that had not been detected in the individual exposures when they were processed by calnica.
The \texttt{calnicb} code uses all the data quality information generated by \texttt{calnica} to avoid propagating identified cosmic rays, bad pixels, or saturated pixels, into the output mosaic. In the case of chopped images, if multiple images were obtained at each chop position, \texttt{calnicb} generates a mosaic for each background position and produces an output product image for each of those. It then combines each of the background images to generate an average background for the target position, removes this background from the mosaic it has generated for the target position, and then saves the result of this operation as the final, background subtracted mosaiced image of the target. When \texttt{calnicb} is calculating the offsets between images, it starts with the telescope pointing information as its first guess. If it is unable to match the various images by adjusting this pointing information by more than some limit, the code reverts to using the pointing information alone, on the assumption that there are no sources bright enough to detect in the individual images, or that there are instrumental artifacts that are confusing the offset calculation, or that the telescope suffered a loss of guide star lock during one or more of the exposures resulting in a potentially very large offset being introduced. To date, we have found that all of these things happen very rarely indeed.

\section*{NICMOS Data Products}

\subsection*{Standard NICMOS Dataset Structure}

NICMOS data are represented by five data arrays for each readout. These arrays contain the:

- Science image.
- Error array.
- Quality flags array.
- Samples array.
- Integration time array.

Each downlinked readout is always represented by these five data arrays. In the basic NICMOS data file format, shown in Figure 12.4, each of these data arrays is stored as a separate image extension in the FITS container file. The \texttt{MULTIACCUM} mode (Figure 12.5) produces multiple images. The file structure for such a dataset consists of a lattice of the 5 data arrays that are created for each readout, as shown in Figure 12.5. Compact FITS representations are used to store arrays in which all elements have the same value.
Science Image (SCI)

The science image contains the information recorded by the NICMOS detector. The data may be represented as counts (i.e., data numbers) or as count rates (i.e., data numbers per second). Generally the latter is desirable since it is easier to interpret in mosaiced datasets and corresponds closely to flux.

Error Array (ERR)

The error array contains an estimate of the statistical uncertainty at each pixel. It is expressed as a real number of standard deviations. This is a calculated quantity based on a noise model of the instrument and its environment.

Data Quality Flags Array (DQ)

The data quality flags array provides 16 independent flags for each pixel. Each flag has a true (set) or false (unset) state and is encoded as a bit in a 16 bit (short integer) word. Users are advised that this word should not be interpreted as an integer.

Samples Array (SAMP)

The samples array is used for one of two purposes:
• For data where multiple samples of the array were obtained during the integration (MULTIACCUM mode), the samples array denotes the number of samples for each pixel in the corresponding science image.

• When multiple integrations are combined to produce a single image, the samples array will contain the number of samples retained at each pixel. Note that this implies that the original number of samples information is not propagated forward into combined images.

Integration Time Array (%TIME)

The integration time array contains the total integration time at each pixel. While initially a simple parameter in some observing modes, combining datasets into composite mosaics and using the information obtained by multiple non-destructive readouts during an integration requires us to keep track of the actual exposure time for each pixel. This array is useful for simple conversions between counts and count rates.

Figure 12.5: MULTIACCUM Data Format

IRAF Access

The physical format of NICMOS datasets is FITS with image and table extensions. Data is delivered to observers and used within IRAF and STSDAS in this format. The FITS image kernel that became available in IRAF v2.11 supports the reading and writing of this data format directly.
This permits the use of NICMOS data without conversion (i.e., the \texttt{strfits} task is \textit{not} necessary).

Individual FITS image and table extensions are accessed from IRAF tasks by appending to the file name either the index number of the extension, or the combination of the extension name (e.g., SCI, ERR, DQ, SAMP, or TIME) and the extension version number (see Figure 12.4 and Figure 12.5).