

Calstis4, calstis11, calstis12: Wavecal Processing in the STIS Calibration Pipeline

Phil Hodge, Stefi Baum, Melissa Mcgrath, Steve Hulbert, Jennifer Christensen

and the Spectrographs Group pipeline block: Ivo Busko (SSG), Paul Goudfrooij, J.C. Hsu (SSG), Rocio Katsanis, and Dick Shaw

April 21, 1998

ABSTRACT

Due to uncertainties in positioning by the mode select mechanism (MSM) and thermal drifts internal to STIS, an image may be shifted on the detector by several pixels from its nominal location. The purpose of wavecal processing is to determine the shift of the image on the detector in each axis. This is only used for spectroscopic data. It requires one or more wavecal (line lamp) observations, taken without moving the MSM from the setting used for the science data.

*This document updates and supersedes STIS ISR 96-19, and it describes the functionality that was implemented for **calstis** V1.7.*

1. Introduction

Wavecal processing is usually done automatically by *calstis0* while calibrating a science file because WAVECORR in the science file header has been set to PERFORM by Generic Conversion. In order to do wavecal processing independently of *calstis0*, several executables must be invoked, or the IRAF **wavecal** task may be used. The steps are as follows:

- (1) Basic 2-D processing (*calstis1*),
- (2) subtract science from wavecal if needed (*calstis11*),

- (3) 2-D rectification (*calstis7*),
- (4) find the shifts and save in the wavecal header (*calstis4*),
- (5) copy the shifts to the science file header (*calstis12*).

These are shown schematically in figure 1 and are described in detail in the next section.

Figure 1: WAVECAL Processing Flow

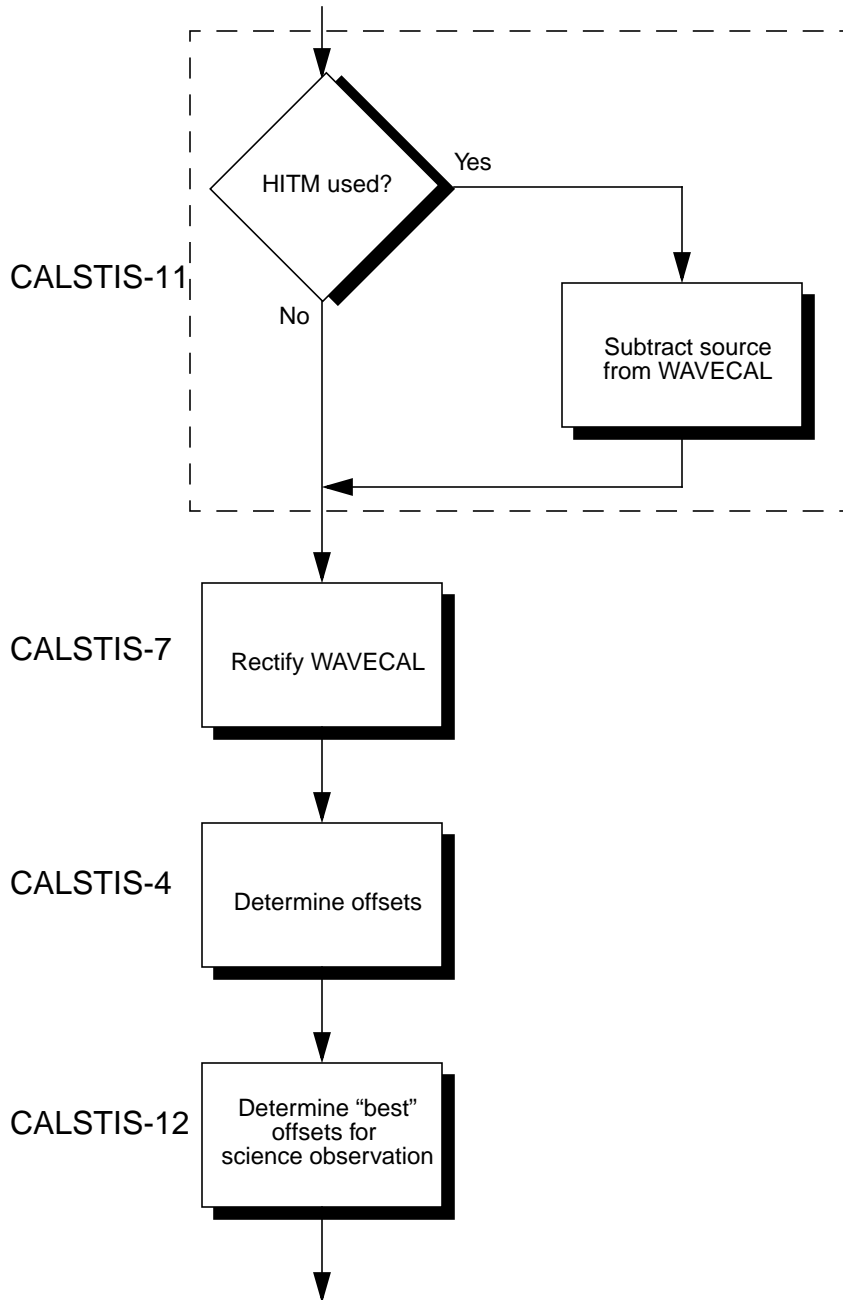


Table 1. WAVECAL Processing Flow—Alternate View

Calibration Step	CALSTIS	Input	Output
If needed, subtract source from WAVECALs taken with HITM	11	Flat-fielded WAVECAL and Flat-fielded science image	WAVECAL with source subtracted
Rectify WAVECAL	7	WAVECAL (with source subtracted, if appropriate)	Rectified 2-D WAVECAL
Determine position and wavelength offsets from WAVECAL	4	Rectified 2-D WAVECAL	Rectified 2-D WAVECAL with updated header keywords indicating computed offsets
Select the “best” offsets for the science observation	12	Rectified WAVECAL and Flat-fielded science image	Flat-fielded science image with update header keywords indicating computed offsets.

2. Discussion

Basic 2-D processing for the wavecal (*calstis1*)

This step is not absolutely essential except for the CCD. For CCD data the overscan must be stripped off before 2-D rectification.

Subtract science from wavecal (*calstis11*)

For CCD data taken with the hole-in-the-mirror (HITM) system, the external shutter is usually open, so the detector was simultaneously exposed to the science target and the wavecal. This case is identified by the primary header keywords DETECTOR and SCLAMP; there is no calibration switch. (This will need to be changed in the future to make use of a new keyword giving the status of the external shutter.) *calstis11* scales the flat fielded science image by the ratio of wavecal to science image exposure times, and then subtracts the scaled science image from the flat fielded wavecal. The output of this step is a temporary file with default suffix “_cwf_tmp”.

2-D rectification (*calstis7*)

This step is done so that the image can be collapsed along columns to get a spectrum or along rows to get an outline of the slit (in the cross dispersion direction). What we really need is the shift of the flat fielded wavecal in pixels, however, and 2-D rectification does remove a certain amount of distortion, both in the dispersion and spatial directions, so the rectified image is not ideal for this purpose. This is not a serious problem because the distortion is small.

When running *calstis0*, the 2-D rectification step will be performed for the wavecal regardless of the X2DCORR switch in either the science or wavecal file.

Find the shifts and save in the wavecal header (*calstis4*)

The input to *calstis4* must be a 2-D rectified wavecal, the headers of which will be modified in place. No change is made to the data. The keywords SHIFTA1 and SHIFTA2 in the SCI extension headers will be updated, WAVECORR will be set to COMPLETE, and history records will be written to the primary header. Note, however, that when this is called from *calstis0* the input is a temporary file which will (by default) be deleted when no longer needed. It will no longer be needed after *calstis12* has copied the keyword values to the science file headers (see step 5). If there is more than one imset in the wavecal, it is likely that not all shifts will be copied to the science file header, so there will be no permanent record of the shifts of individual imsets in the wavecal. However, the shifts are written to the trailer file (i.e. printed to the screen).

The shift in the first image axis (the dispersion direction) is written to SHIFTA1, and the shift in the second axis is written to SHIFTA2. If the shift in either or both axes cannot be determined, the header value will be set to -9999. Note that these keywords are in units of reference pixels, i.e. unbinned CCD or native format (1024x1024) MAMA, which need not be the same as the pixel size in the wavecal or science file.

Each imset in the input file is processed independently. There can be multiple imsets for two reasons, (1) more than one wavecal may have been stored in one file, or (2) for echelle data there will be one imset per spectral order because the data are 2-D rectified. For echelle data there are two complications. Some of the orders of the wavecal may have few or no lines, in which case it may not be possible to determine accurate shifts for those orders. The other complication is that the raw wavecal file may contain more than one exposure, so the 2-D rectified file can contain two or more imsets for each spectral order. These issues are dealt with by *calstis12*, not *calstis4*.

The three components of *calstis4* are as follows, and they are discussed below.

- (1) flag cosmic rays (CCD data only),
- (2) collapse the columns and find the shift in the first axis,
- (3) collapse the rows and find the shift in the second axis.

Wavecal data are not CR-split. If there are multiple imsets in a wavecal file, they were typically taken at different times, and the MSM shifts may be different, so that information would be lost if the imsets were combined. Thus we can't use *calstis2* to reject cosmic rays, and a different cosmic ray rejection function has been added to *calstis4*. It is called if the SCI extension keyword SDQFLAGS (serious data quality flags) includes 8192 (i.e. if (SDQFLAGS&8192) is nonzero); no calibration switch is involved. The logic here is that if SDQFLAGS does not include 8192 (DATAREJECT), the bit used to flag cosmic rays, then cosmic rays are evidently not considered "serious." When a pixel has been identified by *calstis4* as being affected by a cosmic ray hit, it is flagged by including 8192 in the DQ

array for that pixel. This is done in the DQ array in memory; the DQ extension in the file on disk is not modified.

Here are the details of the algorithm for identifying cosmic rays. Since the data have been 2-D rectified, values within a single column are all at the same wavelength, so the values should be fairly consistent in their brightness for a line lamp exposure. Cosmic rays (and hot pixels) will stand out above this brightness level. Actually, the histogram of a single column should have two peaks, one for pixels that were illuminated by the lamp and one for pixels that were obscured by an occulting bar or were beyond the ends of the slit. Since at the beginning of *calstis4* we know only approximately where the boundaries are between illuminated and obscured regions, if we erroneously think a pixel is obscured when it is actually illuminated, it would look like a cosmic ray because it would be too bright for an obscured region. The algorithm must therefore be biased against finding “cosmic rays” with a brightness near what is expected for an illuminated pixel.

For a given column of data, we find the mean and standard deviations separately for the illuminated and obscured regions. The two regions are distinguished (although imperfectly) by the DATAMASKED bit in the data quality array. Before computing the mean and standard deviation, values are rejected as outliers if they are more than three MADs from the median, where MAD is the median of the absolute values of the deviations from the median. A pixel in an illuminated region will be flagged as a cosmic ray if the value is more than three times the standard deviation from the mean, using values of mean and standard deviation for an illuminated region. For a pixel in an obscured region to be flagged as a cosmic ray, the value must not only be more than three times the standard deviation from the mean (using parameters for an obscured region), but it must also not be within two illuminated standard deviations from the illuminated mean. The latter condition is the attempt to protect against rejecting illuminated pixels in regions that are erroneously thought to be obscured.

In order to find the shift in the dispersion direction, the data are collapsed along columns to get a one-dimensional spectrum, and that 1-D spectrum is cross correlated with a template.

At the time of writing, it is not possible to perform either 2-D rectification or *calstis4* on data taken with the cross disperser gratings, i.e. with $\text{dispaxis} = 2$. It is anticipated that this will be handled in the future by transposing the input to the 2-D rectification step. If this is done, then the input to *calstis4* will have the dispersion along the first axis regardless of which grating was used.

The DATAMASKED bit is locally removed from `sdqflags` before trying to find the shift in either the dispersion or spatial direction, partly because the regions so flagged can be somewhat shifted from the actual masked regions (that’s why *calstis4* is being called,

after all), and because in the long slit case it is precisely these regions that we use to find the shift.

When collapsing data along columns, all the columns are used, but the sum over rows is restricted to a fraction of the second axis. Currently that fraction is 0.5 of the height of the image; note that the image is larger than the raw image because during 2-D rectification about 80 pixels are added all around to absorb shifts, such as MSM shifts and heliocentric velocity. An unweighted mean is taken over each column, excluding bad pixels.

The resulting 1-D spectrum may have anywhere from no bright lines to many lines. In order to reduce the effect of noise where there is little signal, the median of the spectrum is subtracted from each element of the spectrum, and the result is truncated below zero. The resulting spectrum is used to find the shift, and it is also saved to be used as weights for collapsing the data along rows.

The reference spectrum, read from the LAMPTAB, is a high resolution spectrum over a wide range of wavelengths, so a certain amount of processing is necessary to make the template for cross correlating with the observed spectrum. It is intended that the template be as similar as possible to the observed spectrum, except of course for the shift which is to be determined. The first step is to integrate the reference spectrum over wavecal pixels. The LAMPTAB contains an array of wavelengths, not necessarily uniformly spaced, and an array of fluxes. The wavecal is uniformly spaced in wavelength, so the wavelength at the edge of each pixel can be obtained from the coordinate parameters CRVAL1, CRPIX1, and CD1_1, which are read from the `_x2d` or `_sx2` file. For each pixel in the first axis of the wavecal, the fluxes in the reference spectrum are added up to give one pixel in the template spectrum. Unless the slit used for the wavecal was less than three pixels wide, the template spectrum is then convolved with a 1-D box to simulate integration over the slit width. The width of the 1-D box is the nearest odd number to the slit width.

The cross correlation will be done using offsets from -31 to +31 pixels, or a full range of 63 pixels. If the wavecal is a very small subarray, however, a range of 63 would be too large. In this case the range will be reduced to half the width of the wavecal. The range must be odd, and it must be at least three pixels. When computing the cross correlation, $(\text{range}-1)/2$ pixels will be ignored at each end of the observed spectrum, so that the same number of pixels will contribute to the sum regardless of the offset. Also, any pixel flagged as bad will be set to zero. The maximum value in the cross correlation is found, and a quadratic is fit to that value and the value on either side in order to get the shift between the observed and template spectrum. If the maximum in the cross correlation is at either end of the range, the message "Warning Peak in cross correlation is at end of range" will be printed. This could indicate that the shift really was of order 31 pixels, but it could also mean that there was essentially no signal in the wavecal, just noise.

In order to find the shift in the spatial (cross dispersion) direction, the data are collapsed along rows to get a one-dimensional image of the illumination along the length of the slit. The shift is found by comparing the locations of features in that 1-D image with where they were expected to be found.

When collapsing data along rows, all rows in the image are used, but the sum over columns is restricted to a fraction of the first axis length. Currently that fraction is 0.7 of the full width of the wavecal image. The full width is typically 1201, because of the border added by *calstis7*. The sum is weighted by the spectrum found previously when collapsing along columns. Recall that this spectrum, and therefore the weight, was truncated below its median, so regions with little signal will have zero weight.

The method used to find the shift in the spatial direction depends on the length of the slit, which is obtained from the aperture name, as given by the APERTURE primary header keyword (e.g. 52X0.1 is 52 arcseconds long). Any slit shorter than five arcseconds is referred to (in this document) as a short echelle slit, and any slit longer than seven arcseconds is a long slit. The medium length slit is six arcseconds long. For a short echelle slit, the 1-D collapsed image is cross correlated with a template of the slit, using the same cross correlation routine as was used for finding the shift in the dispersion direction. For medium length and long slits, the features used are the slit ends and the locations of the occulting bars respectively. The long slits have two occulting bars, and it is the edges of these bars that are used for finding the location. Edges are found by convolving with [1,0,+1]. The location of an edge is then either a minimum or maximum (depending on whether the edge is from bright to dark or from dark to bright), while constant regions are near zero. The peak in the cross correlation is obtained to subpixel level by fitting a quadratic to the three pixels nearest the extremum.

The shifts are measured in units of pixels of the wavecal image. If the wavecal is binned in either axis or is in high-res MAMA pixels, the shifts are then scaled to reference pixel size. The scaled shifts are written to the SCI extension header of the wavecal using keywords SHIFTA1 and SHIFTA2.

Copy the shifts to the science file header (*calstis12*)

The final step is to copy SHIFTA1 and SHIFTA2 from the wavecal to the science file and to update CRPIX1 and CRPIX2. A series of science images and wavecals may have been taken, with the wavecals interspersed in time among the science images. The MSM shift can vary with time, so it is necessary to select the most appropriate wavecal exposure for each science exposure. Currently, the wavecal exposure nearest in time to a given science exposure is the one selected.

The wavecal and science exposures may have been taken using different apertures. This will be the case, for example, when the science data were taken with a wide slit, because with such a slit, the emission lines in the wavecal would be broad, nearly flat-

topped, and overlapping, which would result in a poorly determined shift in the dispersion direction. The various apertures are not exactly centered with respect to each other; the offsets (from a reference aperture) are given in the aperture description table. The offset of the aperture used for an exposure is accounted for by modifying the dispersion coefficients when running 1-D extraction or 2-D rectification; this is the incidence angle correction. Therefore, when copying the SHIFTA1 and SHIFTA2 values from the wavecal to the science file, no adjustment needs to be made for the offset between the apertures.

For echelle data, in the 2-D rectified wavecal there will typically be a few dozen spectral orders, each a separate imset, corresponding to one wavecal exposure. *calstis4* will have determined a shift independently for each spectral order, so *calstis12* must average all these to get one shift in each axis for that wavecal exposure. If the raw wavecal contained more than one exposure, *calstis12* must also distinguish between the 2-D rectified imsets belonging to different exposures. This is done by using the time of the exposure, taken to be the average of EXPSTART and EXPEND. Some of the spectral orders will contain a sufficient number of emission lines to reliably determine the shift, but some may contain few or even no bright lines. Therefore, the averaging process must ignore cases where the shift was not found (i.e. the value was -9999), and it must ignore outliers where the shift that was found was not very accurate. If there are only two points, and they differ by less than two pixels, the average is returned. If there are more than two points but fewer than 10, the median is returned with no further attempt to reject outliers. If there are 10 or more points, however, the following procedure is used. The median is computed as a first step; then the median of the absolute values of the deviations from the median is computed (i.e. the MAD); then points farther than five times the MAD from the median are rejected; finally, the remaining points are averaged.

If the shifts (or the averages of the shifts) are not flagged as not found, they are then written to all three extension headers (SCI, ERR, DQ) of the science file for the current imset. If either or both shifts were not found (-9999), a value of zero will be written to the header keyword(s). This is so that *calstis6* or *calstis7* can be run on the science file without introducing spurious shifts.

calstis12 also adds SHIFTA1 and SHIFTA2 to the current values of CRPIX1 and CRPIX2 (the reference pixel). Actually, *calstis6* and *calstis7* do not use CRPIX1 and CRPIX2, so this is of little significance. In order to allow the user to run *calstis12* more than once on the same science file, what is actually done is the following. The current values of SHIFTA1 and SHIFTA2 and CRPIX1 and CRPIX2 are read from the science header, the current shifts are subtracted from the current reference pixel, then the new shifts are added, and the result is written to CRPIX1 and CRPIX2 in the science header. Note that this means that SHIFTA1 and SHIFTA2 should be initially zero in the flat fielded science header, which will normally be the case, since these keywords are set to zero in the raw files by Generic Conversion.

3. Keywords and column names

Aside from the keywords describing the instrument configuration that are read by most *calstis* modules (e.g. DETECTOR), the *calstis* modules also read or update some other keywords. These are listed in Table 1.

Table 2: Image Header Keywords Accessed during Wavecal Processing

Keyword	Description	Location	Referenced by	Output Updated?
WAVECORR	Updated to COMPLETE or SKIPPED	Primary header	<i>calstis4</i> , <i>calstis12</i>	Yes
LAMPTAB	Name of table containing template lamp spectrum	Primary header	<i>calstis4</i>	No
SCLAMP	Name of line lamp	Primary header	<i>calstis4</i> , <i>calstis11</i>	No
DISPAXIS	Dispersion axis (1 or 2)	Extension header	<i>calstis4</i>	No
SDQFLAGS	Serious data quality flags	Extension header	<i>calstis4</i>	No
SHIFTA _i	Shift (pixels) in each axis	Extension header	<i>calstis4</i> , <i>calstis12</i>	Yes
CRPIX _i	Reference pixel	Extension header	<i>calstis4</i> , <i>calstis12</i>	Yes
CRVAL _i	Coordinates at the reference pixel	Extension header	<i>calstis4</i>	No
CD _i _ _i	Pixel size, Angstroms by degrees	Extension header	<i>calstis4</i>	No
LTM _i _ _i	Size of reference pixel in image pixel units	Extension header	<i>calstis4</i>	No
EXPTIME	Exposure time (seconds)	Extension header	<i>calstis11</i> , <i>calstis12</i>	No
EXPSTART	Exposure start time (MJD)	Extension header	<i>calstis11</i> , <i>calstis12</i>	No
EXPEND	Exposure end time (MJD)	Extension header	<i>calstis11</i> , <i>calstis12</i>	No

calstis4 reads two reference tables, the APDESTAB (`_apd`) table for aperture description, and the LAMPTAB (`_lmp`) table for the template lamp spectrum. The column names that are used from these reference tables are listed in tables 2 and 3 respectively.

Table 3: Columns in the APD Reference File

Column Name	Data Type	Units	Description	Examples
APERTURE	A16	—	Name of aperture	52X0.1
WIDTH _i	D	arcsec	Width of aperture in each axis	0.094, 50.57
NBARS	I	—	Number of occulting bars	0, 1, 2, 3
BAR _i LOCN	D	arcsec	Location of occulting bar number <i>i</i>	-11.44,+11.62
BAR _i WIDTH	D	arcsec	Width of occulting bar number <i>i</i>	0.5, 0.86

Table 4: Columns in the LMP Reference File

Column Name	Data Type	Units	Description
SCLAMP	A12	—	Name of line lamp (e.g. LINE, HITM1, HITM2)
LAMPSET	A6	mA	Lamp current
NELEM	I	—	Number of elements in arrays
WAVELENGTH	D	Å	Array of wavelengths
FLUX	D	—	Array of fluxes

4. Problems, testing, debugging

When *calstis4* is trying to find the shifts, there are a number of things that can go wrong. One possibility, of course, is that the wavecal exposure is too short, so that the signal is too weak to reliably determine the shifts. For echelle data, this will often be the case for at least a few spectral orders. One common problem is an excessive number of pixels flagged as bad in the data quality extension. Bad pixels are not included in the sums when collapsing the data to 1-D arrays. If there is sufficient data this is OK, but for some modes there are only a few strong emission lines, and if there is a pixel flagged as bad in the wings of one of those lines, omitting it can have a significant effect on the shift that is obtained. If a pixel really is bad it should not be used (cosmic ray hits, for example). In many cases, however, pixels are flagged in reference files even though the condition has been corrected or would not be serious for the purposes of *calstis4*. Two examples are warm pixels that have been corrected by the dark image, and regions behind occulting bars in the flat field image. This problem can be worked around by modifying SDQFLAGS in the SCI extensions of the wavecal file. The value to set it to depends on which data quality flags you want to be considered as “serious.” A value of 4 should always be included, because this flags regions that are off the detector. For the CCD, 8192 should normally be included in order to exclude cosmic ray hits. To summarize, one could use SDQFLAGS = 4 for MAMA data, and SDQFLAGS = 8196 for CCD data. Another potential problem is with MAMA data using a long slit. The shift in the dispersion direction should be found without problem, but finding the shift in the spatial direction can be affected by one of the occulting bars being off the detector, due to the large image scale of the MAMAs.

It is intended that a debug mode be added to *calstis4* in the near future. This will be invoked by including “-d filename” on the command line. The output to the debug file will include the collapsed 1-D spectrum and the template spectrum resampled to the

same pixel spacing, the collapsed 1-D image of the illuminated slit, and the calculated locations of each of the features (e.g. edges) that were to be found.

One way to test that the wavelengths are being computed correctly is to use a wavecal exposure as both wavecal and science file. The wavelengths of emission features in the resulting `_x1d` or `_x2d` file must then agree with the wavelengths in the template spectrum (LAMPTAB) that was used for calibration. This is a simple idea, but there are a couple of steps that may not be obvious. Here is a procedure that will work. Suppose the wavecal is called `o47s01k9m_wav.fits`. First copy that to a temporary file, say `o47s_raw.fits`, which we will pretend is a science file, hence the suffix “_raw”:

```
cl> copy o47s01k9m_wav.fits o47s_raw.fits
```

Now edit the headers of `o47s_raw.fits` as follows:

```
cl> hedit o47s_raw.fits[0] wavecal o47s01k9m_wav.fits
```

```
cl> hedit o47s_raw.fits[0] wavecorr PERFORM
```

```
cl> hedit o47s_raw.fits[0] crcorr OMIT
```

```
cl> hedit o47s_raw.fits[0] fluxcorr OMIT
```

```
cl> hedit o47s_raw.fits[0] helcorr OMIT
```

```
cl> hedit o47s_raw.fits[0] x2dcorr PERFORM
```

Make the following changes for each SCI extension, not just the first:

```
cl> hedit o47s_raw.fits[sci,1] asn_mtyp TEST
```

```
cl> hedit o47s_raw.fits[sci,1] exptime "(exptime * 100.)"
```

Clobbering ASN_MTYP is necessary so that *calstis7* will not realize that `o47s_raw.fits` is really a wavecal, because when *calstis7* processes a wavecal it doesn't apply the SHIFTA1 and SHIFTA2 keywords. Changing the exposure time is only necessary for CCD data taken with the HITM; if the exposure time were left the same as in the original wavecal, then when *calstis11* subtracts the “science” from the wavecal, the result would be zero.

If 1-D extraction is desired:

```
cl> hedit o47s_raw.fits[0] backcorr OMIT
```

```
cl> hedit o47s_raw.fits[0] x1dcorr PERFORM
```

Then run *calstis0*:

```
cl> calstis o47s_raw.fits verbose=yes > o47s.log
```

5. Table of parameters for *calstis4*

There are a number of parameters involved in the *calstis4* processing which currently are just defined in the source code. In the near future we intend to create a new reference table for these parameters and to modify *calstis4* to read the parameters from this table.

6. References

Hulbert, S., Hodge, P., Baum, S., 1996, STIS Instrument Science Report 96-19