Chapter 9

Low-Dispersion Flux Extraction (*SWET*)

Spectral data are extracted from the two-dimensional (2-D) low-dispersion resampled images (SI) using a signal weighted extraction technique (*SWET*). This technique provides an increase in the signal-to-noise ratio (S/N) of the extracted spectrum, as compared to simple "boxcar" extractions, by utilizing information on the cross-dispersion spectral profile, while preserving the total flux. The output of this technique also includes an error estimate associated with each point in the extracted spectrum and cosmic ray hits are, in many cases, automatically rejected from the calculation of extracted fluxes. The technique, which has been developed for low-dispersion *IUE* data, is based on the weighted slit extraction code originally developed by Horne (1986) for use with optical CCD data. A detailed explanation of the adaptation of this software for use with *IUE* data processed using IUESIPS is given by Kinney et al. (1991) who supplied the software to the *IUE* Project. Various threshold values in the code were modified by the *IUE* Project to account for the very different flux scales of NEWSIPS and IUESIPS.

Briefly, the processing steps involved in the weighted extraction of a one-dimensional (1-D) spectrum from the low-dispersion SI are as follows:

- The background flux as a function of wavelength in the low-dispersion SI is characterized by a sixth-order Chebyshev polynomial.
- The spectral signal is located within the image and the average signal strength is evaluated in order to determine whether or not a spectral profile can be computed empirically.
- If there is sufficient signal strength, the cross-dispersion instrumental flux profile is derived for each wavelength sample. A default profile shape is used for spectra that are too faint to allow an empirical determination.
- The cross-dispersion profile is used in combination with a model of the detector noise characteristics to identify anomalous data points which are to be excluded from the

computation of extracted fluxes.

- The good data points are weighted according to their positions on the cross-dispersion profile and according to their relative noise in order to calculate the total net flux and its associated one-sigma uncertainty at each wavelength sample.
- A 1-D ν flag spectrum is calculated to show where flagged pixels may have precluded the calculation of a reasonable flux value.

9.1 Noise Models

Two portions of the low-dispersion extraction procedure rely upon an estimate of the detector noise in order to perform their functions. First, the cross-dispersion profile fitting routine utilizes the estimated S/N of each wavelength sample in order to calculate appropriately weighted spline fits to the data. Second, the extraction procedure uses the noise model information to derive an error estimate for each point in the extracted spectrum. The noise models are derived empirically for each camera by measuring the scatter in the flux numbers (FN) around the mean FN in the background regions of several hundred science and flat-field images taken at a variety of exposure levels. Since the sigma as a function of FN is wavelength-dependent, the analysis is performed within 20 equal-sized wavelength bins (∼54Å wide for the SWP and ∼85Å wide for the long-wavelength cameras) in the low-dispersion SI. For each wavelength bin the standard deviation in FN versus mean FN is represented by a third-order polynomial. The wavelength-dependence of the four coefficients of this polynomial are then each represented with a third-order polynomial to allow a determination of the expected standard deviation of any pixel given its wavelength and observed FN.

9.2 Background Flux Extraction

The background flux that is present in *IUE* images is composed of contributions from several sources, including the null pedestal, particle radiation, radioactive decays within the detector phosphor, halation within the UV converter, background skylight, and scattered light. The integrated effect of the last three sources varies in a complicated manner across the target depending on the spectral flux distribution of the object observed, whereas the general radiation and null components vary slowly across the vidicon tube. In practice, the average background flux in the low-dispersion SI is computed within two regions on either side of the spectrum and this average flux as a function of wavelength is fitted with a sixth-order Chebyshev polynomial.

The two background regions that are examined are each 7 pixels wide and, for large-aperture data, begin at a distance of 13 pixels from the predicted center location of the spectrum. For small-aperture data the two regions begin at a distance of 8 pixels from the center of the spectrum (see Figure 9.1). The narrow large-aperture extraction region corresponds to the 13-pixel point source slit height, while the broad region is the 23-pixel extended, trailed, multiple, and flat-field source slit height. The entire wavelength space of the low-dispersion SI is included in the analysis. Before calculating the average flux as a function of wavelength for each of the two regions, each spatial image line within the two regions is examined individually in order to identify flagged pixels. The FN values of flagged pixels are temporarily replaced with the FN value of the closest unflagged pixel. Telemetry dropouts in the background regions are identified during this process and ν flag values of -4 are set in the output 1-D flag spectrum as well as retroactively set in the 2-D low-dispersion resampled flag image. This flag replaces the original value of -8192.

At this point a difference spectrum is computed from the average FN as a function of wavelength for each of the two background regions. This difference spectrum is used to detect the presence of cosmic ray hits that may occur in either of the two regions. The Chebyshev polynomial that will be fitted to the background data is sensitive to broad features such as grazing cosmic ray hits. Since we are only concerned with relatively broad features, the difference spectrum is smoothed with a 7 pixel wide boxcar function. The mean and standard deviation of the smoothed difference spectrum is then computed and any pixel locations that deviate by more than 2σ from the mean have their weights for the polynomial fitting step set to zero. All remaining good pixel locations are given equal non-zero weights in the polynomial fit. Pixel locations that are beyond the long-wavelength edge of the camera target area are also excluded from the polynomial fit. Once the Chebyshev fit has been calculated, the values in the 1-D background spectrum that correspond to locations outside the camera target area are replaced with the fitted value of the last valid pixel.

9.3 Spectrum Location and Signal Level

The location of the spectrum in the low-dispersion SI (in the cross-dispersion direction) is found by computing the wavelength-averaged flux-weighted centroid of the background-subtracted flux levels within a search region that is centered on the predicted spectrum location. For SWP images the region of the low-dispersion SI below 1233Å is excluded from this calculation so as to avoid contamination from geocoronal Lyα. If insufficient signal exists from which to determine reliably a location for the spectrum, the predicted center position is used instead and a warning message to this effect appears in the processing history log.

If the spectrum centroid can be determined empirically, then the dispersion line that contains the maximum net flux (summed over wavelength) is also determined and is referred to as the spectrum peak location. The wavelength-averaged net flux of this dispersion line is computed and compared against the threshold value of 5 FN in order to determine whether the spectral signal is strong enough to attempt an empirical profile fit. For the SWP camera, the region below 1233Å is excluded from this computation so as to avoid contamination from geocoronal Lyα. If the image is processed as a point source, then a warning message appears if the location of the spectrum peak is found to be more than one image line away from the spectrum centroid, since this may indicate that the spectrum is not characteristic of a point source but rather may be spatially extended.

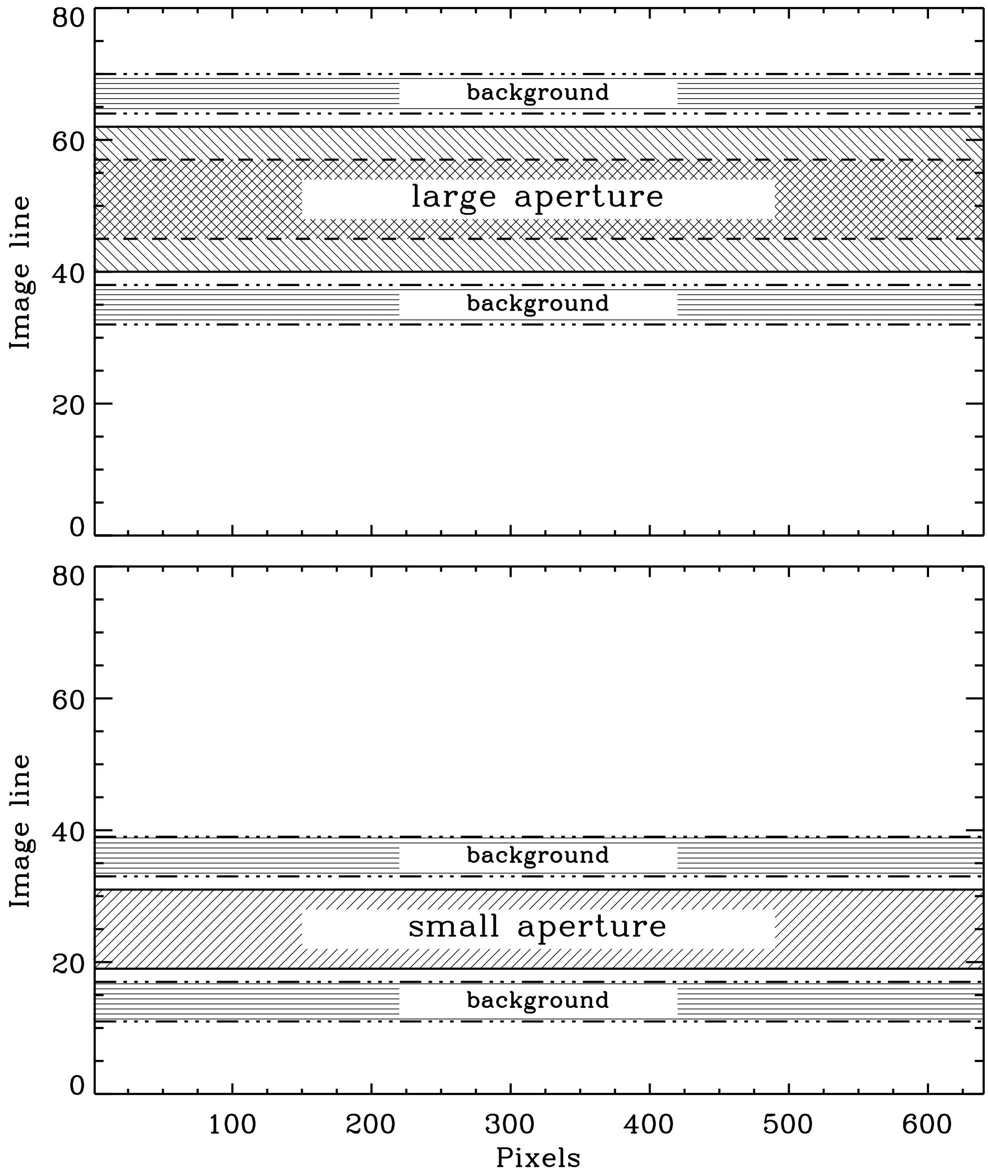

Figure 9.1: Locations of the extraction and background fitting regions within the low-dispersion SI.

If the average FN of the peak line of the spectrum is below the threshold value, then an empirical fit is not attempted. Instead, a default profile shape is used to compute the extracted fluxes. There are three default profiles for point sources, one for each camera. Each default profile has been derived from well-exposed observations of *IUE* standard stars, by averaging the profile fits for these observations over wavelength and renormalizing, thus producing a profile that is constant in wavelength. Since the spectral profile for *IUE* data is not repeatable from image to image, this average profile provides a rough weighting that is useful in very low signal images and provides a lower noise spectrum than attempting to fit the profile from the image. For extended and trailed sources, the default profile has a boxcar-like shape where all spatial image lines receive equal weights. Warning messages appear in the processing history log when a default profile is used.

A warning message is also issued if the location of the spectrum centroid is found to be more than two image lines away from the predicted spectrum center. This situation may arise if the location of the spectral format within the raw image space does not follow, for whatever reason, the normal behavior of the spectral format as a function of time (observation date) and camera temperature (THDA) as may be the case for objects poorly centered in the aperture. If this is the case, then the time and temperature-corrected dispersion constants that are computed for this particular image may be called into question. Note, however, that this warning only indicates that the location of the spectrum in the cross-dispersion (i.e., spatial) direction has been found to be unusual. We have no information as to whether or not a similar shift may have occurred in the dispersion direction of the image.

The search for the spectrum center and the determination of the average peak FN is skipped entirely for images classified as flat-field (i.e., nulls and other floodlamp only images), since they contain no spectral data. In this case the predicted spectrum center location and the default profile for extended/trailed sources are used to perform the computation of extracted fluxes.

9.4 Profile Fitting

The cross-dispersion profile is determined from the background-subtracted net flux data. Each line of data in the low-dispersion SI that is contained within the spatial limits of the extraction slit (13 lines for point sources, 23 lines for extended, trailed, multiple, and flat-field sources) is temporarily binned in wavelength according to the S/N, with bin sizes typically ranging from 1 to 10 pixels. A warning is issued if the bin size exceeds 4 pixels since this may indicate either weak or noisy spectral data or that a large fraction of the data is unusable due to conditions such as saturation or telemetry dropouts. The area under the cross-dispersion profile at each wavelength bin is then normalized to one to divide out the spectral information. After normalizing, the fraction of flux in each line of data (relative to the total flux in each wavelength bin) is smoothed by fitting splines with anywhere from 2 to 15 nodes in each line.

The number of spline nodes is determined dynamically based on the total of the square of the S/N of the spectral data. The placement of the nodes in wavelength space is also determined dynamically so that the data located between each pair of nodes have an equal total of $(S/N)^2$. Thus regions of higher S/N have more spline nodes placed more closely together than regions with lower S/N. All lines of data in the cross-dispersion direction use the same number and placement of spline nodes and there is always a minimum of two nodes placed at either end of the spectrum. However, when only two spline nodes are used, the spectral data is fitted with a default extraction profile. The spline fits for each line of data are computed iteratively, rejecting data bins that have residuals to the fit greater than 3.5σ. The iteration cycle ends when no new bins are rejected.

In the procedure outlined above, those wavelength bins that have negative or zero total flux are rejected from the computation of the spline fits. If bins on either end of the spectrum are rejected, they are replaced by the average fraction of light values for the last 10 good bins closest to the end. This is done to avoid extrapolation of the high order spline fits into wavelength regions that have no constraining data. Rejected bins that are not at either end of the spectrum but are located in the interior of spectrum are simply excluded from the computation of the spline fit and are therefore essentially interpolated over. This extrapolation/interpolation process provides a means of dealing with discontinuous data.

The cross-dispersion profile at each wavelength sample is constructed by evaluating the spline fits, clipping the negative values that may occur in the wings of the profile, and renormalizing to ensure a sum of unity at each wavelength. Thirteen spline fits (one for each of the 13 lines of the low-dispersion SI containing source flux) are used to specify the cross-dispersion profile for point source data, and likewise 23 are used for extended, trailed, or multiple sources. This profile is used when computing the extracted fluxes to weight the data so that the pixels with the largest fraction of light are given the most weight. Because the profile is derived by spline fitting each spatial line of data independently of other lines, a wide variety of profiles can be accommodated. Specifically, the algorithm is very capable of handling spectra obtained through either the large or small spectrograph apertures, trailed sources, and extended objects with complicated source structures which may cause the profile to vary dramatically as a function of wavelength. However, because the algorithm produces a single output spectrum, spectra of several sources in the aperture (multiple spectra) will not be separated. Such images will require custom extraction.

9.5 Extraction of Flux and Cosmic Ray Removal

The computation of the extracted flux and corresponding identification and rejection of bad pixels is performed iteratively. The total flux at each wavelength is found by performing a weighted sum of all "good" (see next paragraph) pixels at the given wavelength according to:

$$FN_{total} = \frac{\sum FN(i) \times p(i)/\sigma(i)^2}{\sum p(i)^2/\sigma(i)^2}$$

where $FN(i)$ is the net flux of the i^{th} spatial pixel at this wavelength, and $p(i)$ and $\sigma(i)$ are the corresponding profile and estimated noise values for that pixel. Similarly, the estimated error in the extracted flux is:

$$error(i) = \sqrt{\frac{1}{\sum p(i)^2/\sigma(i)^2}}$$

which is hereafter referred to as the “sigma” spectrum. Because the sum of the profile values is involved in the calculation of the total flux, individual bad pixels can be excluded from the sum so that a reliable estimate of the total flux can be obtained based on only the remaining good pixels.

At the outset of the extraction process, “good” pixels are those with ν flags greater than -256. Hence, pixels which have been assigned ν flags of -256 or more negative by the earlier processing stages are excluded from the extracted flux computation. However, pixels which were not flagged previously but which are found during the course of the extraction to have anomalously high FN values (hereafter referred to generically as *SWET* cosmic ray pixels) are also identified and excluded during calculation of extracted fluxes. This is accomplished by rejecting the most extreme outlying pixel at each wavelength sample if the value of that pixel is greater than a predetermined sigma level from the scaled profile at that wavelength. The threshold values for this rejection are 6, 5, and 4σ for the LWP, LWR, and SWP cameras, respectively. The rejection thresholds for each camera were determined empirically so that most cosmic ray pixels are removed without removing good data points. When such a pixel is identified, it is assigned a ν flag value of -32 which is written back into the low-dispersion ν flag image (SF). Each time a new cosmic ray pixel is identified and excluded, an updated value for the total flux at that wavelength is computed. This process is repeated until no more pixels are rejected or until less than 30% of the flux in the profile at that wavelength sample remains.

The total numbers of pixels excluded as bad because they had already been assigned ν flags of -256 or more negative before the extraction procedure and those excluded as cosmic ray pixels by this routine are reported to the processing history log. In the event that either of these values exceeds 10% of the total number of pixels available to the extraction process, a warning message is issued in the processing history log.

9.6 One-Dimensional ν Flag Spectrum

In deriving the 1-D ν flag spectrum, the cross-dispersion profile and the flag values in the low-dispersion SF are examined. For each wavelength sample the contribution to the spectral profile from unflagged pixels is calculated. If this contribution amounts to at least 45% of the total flux (as represented by the associated profile values), the wavelength sample will contain no flags in the output spectrum. If less than 45% of the total flux in the profile is from unflagged pixels, then each flag value is examined individually to determine what fraction of the total contribution from all pixels is due to one particular flag condition. For each flag condition that affects pixels representing more than 15% of the spectral profile, that flag value will appear in the output flag spectrum.

Situations can exist in which the contribution by pixels with one particular flag condition account for 45% of the flux at a given wavelength, for example, and a second flag condition is associated with pixels that account for only 11%. Here the total contribution from flagged pixels exceeds the 55% threshold, but only the flag value corresponding to the first condition will appear in the output flag spectrum. The threshold values were determined empirically so that problem conditions that definitely have an adverse affect on the computation of extracted fluxes are always flagged in the output spectrum, while, at the same time, conditions that only have a marginal impact are not flagged. Note that in marginal cases where only one or perhaps a few pixels that contribute in a relatively minor way to the total flux at a given wavelength are flagged and hence excluded from the computation of the extracted flux, the associated error estimate for that extracted flux value will be increased accordingly even though no flags will appear in the output spectrum. Thus the sigma spectrum serves as an additional indicator of the relative accuracy of the extracted fluxes.

9.7 *SWET* Output

The main output data product produced during *SWET* is the low-dispersion merged extracted image FITS file (MXLO). The MXLO contains the net (background-subtracted) integrated flux spectrum, the Chebyshev characterized background spectrum, the ν flag spectrum, the absolutely calibrated net flux spectrum, and the calibrated sigma spectrum stored in a FITS binary table extension. The net flux and background spectra are in units of FN, the flag spectrum is in the normal bit-encoded unitless values, and the calibrated flux and sigma spectra are in physical units of ergs sec^{-1} cm^{-2} Å^{-1}. The sigma spectrum, originally computed in FN units, undergoes identical calibration steps (including all corrections for sensitivity degradation, etc.) as the net flux spectrum. Chapter 11 details the process of absolute flux calibration and camera degradation correction.

The background flux spectrum is scaled to the same integrated slit length as the net flux spectrum and is evaluated over the entire wavelength space of the camera target region contained in the low-dispersion SI (1050Å to ∼2000Å for SWP and 1750Å to ∼3400Å for LWP and LWR). The background values for pixel locations past the long-wavelength end of the camera target are replications of the last valid background value. The net flux and sigma spectra are computed over the entire wavelength space of the low-dispersion SI, but locations that are outside the camera target area will have net fluxes equal to zero because these regions do not contain any valid image data. The edge of the camera target areas are encountered at wavelengths greater than (approximately) 3395, 3425, and 2000Å for the LWP, LWR, and SWP cameras, respectively, for large-aperture spectra. Due to the curvature of the target boundary, small-aperture spectra reach the target edge at different wavelengths: 3380, 3400, and 2015Å for the LWP, LWR, and SWP cameras. The absolute calibrations are valid over wavelength limits of 1150–1980Å for SWP and 1850–3350Å for LWP and LWR. Beyond these limits the calibrated net flux and sigma spectra are set to values of 0 and −1, respectively and have a ν flag value of −2. The valid wavelength limits of the calibrated spectra are consequently somewhat truncated as compared to the net flux spectrum.

The *SWET* module writes the following information to the HISTORY portion of the image label:

- noise model version number,
- aperture-dependent extraction/calibration information:
  - predicted and found line positions of spectrum center, line position of the peak flux, and average peak FN,
  - cross-dispersion profile information (blocksize, total number of blocks, and number of rejected blocks),
  - number profile nodes and sigma rejection,
  - line position of profile centroid,
  - line numbers of flux extraction window,
  - pixel rejection threshold,
  - total number of extracted pixels, number of pixels rejected as cosmic rays, and number of pixels rejected as bad,
  - absolute flux calibration version number,
  - absolute calibration mode (i.e., point or trailed),
  - absolute calibration epoch,
  - camera rise time,
  - effective exposure time,
  - THDA of image,
  - reference THDA,
  - temperature-dependent sensitivity correction coefficient,
  - temperature correction factor,
  - time-dependent sensitivity degradation correction version number,
  - time-dependent sensitivity degradation correction mode,
  - sensitivity degradation calibration epoch, and
  - observation date.