Chandra X-Ray Center

Transmission Grating Spectroscopy Data Processing Technical Reference

David Huenemoerder, John Houck

16 August 1999

1 Purpose

The purpose of this document is to provide a technical summary and snapshot of the existing Chandra transmission grating (TG) spectroscopy tools and documentation, and to additionally provide some examples of intended usage in data analysis. The audience is assumed to have experience with X-ray astrophysical science, data analysis methods, and software (e.g., a CXC staff scientist).

2 Tools

The following 7 programs will be referred to as the CXC Data System (CXCDS) Tools.

2.1 Event processing

`tgdetect` - detect zero-order sources and write a table of zero-order coordinates.

`tg_create_mask` - generate a spatial mask for filtering photons and "resolving" events.

`tg_resolve_events` - apply the spatial mask, apply aspect, assign diffraction coordinates; if ACIS is the detector, additionally resolve diffraction orders.

`tgextract` - bin diffracted events into 1D histograms.

2.2 Spectral measurement

`specdetect` - detect emission line features in a binned spectrum.

`specmeasure` - measure the features found by `specdetect` by fitting simple line-profile functions to them.

`specidentify` - apply user-specified model parameters and identify lines using a line-list data-base.

2.3 Usage Documentation

Each tool has an associated "help" file, which specifies the execution syntax, parameters, defines each parameter, and provides a synopsis of the algorithm, inputs, and outputs. These are written by the developer and reviewed by the specifier, and are integrated into the Data System. Copies of the help-files are attached. Each also has one or more associated parameter files which control the execution.

2.4 Spectral Analysis Systems

ISIS (Interactive Spectral Interpretation System) - perform detailed analysis of a line spectrum. ISIS is a self-contained interactive analysis environment. Its distribution package contains a technical reference manual.

CIAO: CXC Data Analysis tools exist for browsing, filtering, plotting, and fitting data.

2.5 File Formats

Output file formats are defined in the following Interface Control Documents:

- "Grating Data Products: Level 1.5 to ASC Archive ICD"
- "Grating Data Products: Level 2 to ASC Archive ICD"

Input files (Analysis Reference Data) are specified in detail by the following Interface Control Documents:

- "Analysis Reference Data for: Exposure Maps and Grating Levels 1.5–2 Tools"
- "Analysis Reference Data: Spectroscopy Database ICD"

Interface Control Documents are available at:
`http://icxc.harvard.edu/sds/development/ICD/index.html`.

2.6 Other Documents

Detailed descriptions of coordinate systems and transformations are given in:

- "Coordinate Systems (Rev 5.0)", Jonathan McDowell (document SDS-2.0) (`http://hea-www.harvard.edu/asclocal/sds/docs/jcmdocs/ps/SDS02.ps`)

A still-relevant overview, with graphical examples of the data, can be found in the HEAD-CC3 presentation:

- "AXAF Dispersive Spectroscopy Pipeline Overview",
  David Huenemoerder,
  HEAD-CC-3, 11 December 1998
  (`http://space.mit.edu/ASC/analysis/tg_headcc3/tg_headcc3.html`
  or
  `http://space.mit.edu/ASC/analysis/tg_headcc3/tg_headcc3.ps.gz`)

2.7 Significant external requirements

The CXC Data System tools are designed for use within the CXC Data System environments. This system is described elsewhere. Data products are FITS files, largely readable by a variety of software (FTOOLS and IDL, for examples).

3 Typical Data Flow

The Chandra transmission gratings, HETG or LETG, are typically used with the ACIS-S and HRC-S detector arrays, respectively. The data can be processed to a degree without any dependence upon the gratings. This is referred to as "Level 1": it is primarily a scaling and formatting of the event data and ancillary descriptive information about the observation. After Level 1, "Level 2" processing is defined as analysis of source properties, such as source detection, fluxes, or extent. Grating data require a bit of analysis in order to provide the basic spectral properties of each photon: its diffraction angle, wavelength, and if possible, a diffraction order. This is referred to as "Level 1.5", since it must be done before spectral analysis.

3.1 Level 1

During standard ACIS or HRC event processing, telemetry is processed by decommutation, scaling and formatting into FITS binary tables. The aspect solution is applied to give sky coordinates for the photons, but without knowledge of whether a photon has been diffracted by a grating. The result has columns in the data such as time, chip id, chip x, chip y, detector x, detector y, sky x, sky y, PHA, and PI. The output is a Level 1 events file.

3.2 Level 1.5 Processing

The order of use of tools, principal inputs and outputs, and major reference data are given here. More details are available in the help and parameter files.

`tgdetect`

Input: Level 1 events file
Output: source table

Before attributes pertaining to the diffraction can be assigned, the zero-order source position on the sky must be determined. This is the origin of the spectral coordinates, and so each source in the field will have its own origin for diffraction coordinates. The algorithm used is a simple cell-detect method, applied only near the aim-point (unless parameters are customized).

The output is a source table giving source positions in sky coordinates in the specified region. Since grating processing is done per observation interval, a source-matching algorithm is applied to create a merged table. Sources are found down to a specified signal-to-noise ratio fraction of the brightest via a parameter.

There are many heuristics which could be implemented to discriminate zero-order sources from emission lines (off-axis PSF vs line-spread function; symmetry about candidate zero-orders, broad PHA variance in zero-orders vs narrow in lines). Implementation of these advanced techniques has been deferred pending experience with real data and need based on frequency of multiple sources in the field.

`tg_create_mask`

Input: Level 1 events
Source table
Reference data: PSF width table
Output: region file

From purely geometrical considerations, given a few calibration parameters of the instrument, we can define spatial regions which identify distinct parts of the spectrum. For HETGS, the relevant parts for each source are: zero-order, MEG photons, HEG photons, or background photons. For LETG, there is only one grating type, so the parts are zero-order, LEG, or background. The High Energy Suppression Filter, or HESF, a.k.a. the Drake Flat, adds several more parts: straight-through, plus two more parts on each of the plus and minus order sides.

Error-budget calculations accomodate off-axis PSF (through the referenced PSF width table), astigmatism, which broadens the spectrum with diffraction angle, and roll-variation. All of these can be adjusted manually via the parameter set to make custom masks. The default masks are made to be many PSFs in size in order to provide lattitude for source and background spectral extraction in subsequent processing.

The zero-order and diffracted parts are calculated for up to 10 sources in the source table. The number 10 is somewhat arbitrary and is a compromise between expected maximum source density and implementation complexity.

`tg_resolve_events`

Inputs: Level 1 events file
Region file
Aspect offsets file
Alignment file
Reference Data: Response-matrix energy-limit table
Output: Level 1.5 events file

Attributes are assigned to each photon according the part of the region in which it falls: zero-order, HEG, MEG, LEG, HESF, or background. For photons in diffracted order regions, diffraction coordinates are assigned. These are in a spherical coordinate system, in which the pole is defined by the grating-node to zero-order centroid direction and the direction of the grating bars: the diffraction angle is analogous to a longitude, and the cross-diffraction direction to a latitude, with the central line of the diffraction being the equator (see the coordinates document for details). The diffraction and cross-diffraction angles are added to the event record in units of degrees, along with the spectral part and source number. Then the simple grating equation is applied ($m\lambda = P \sin\theta$) to assign $m\lambda$ (m is the integral diffraction order, λ the wavelength, P the grating mean bar period, and θ the diffraction angle).

If the detector is ACIS, the CCD pulse-height is used to discriminate diffraction orders, and m and λ are added to the event record. If photons lie in the diffracted spatial region but do not match a valid CCD PHA, they are labeled as background. To do ACIS-PHA order sorting, a special table (the reference data file) is used which is derived from the response matrices for each CCD. The table gives the low and high CCD energy (scaled PI) bounds at each energy for an *a priori* specified fractional enclosed probability (such as 0.95

or 0.98). Use of this table implies a scaling factor which has to be included in computing the effective area.

In the case of HRC, the values stored for m and λ are $+1$ or -1 and $m\lambda$, since order resolution is not possible with HRC pulse-heights.

If there are multiple sources in a field observed with HETG, then there will be spatial overlap of orders. It is still possible to resolve orders if PHA values are uniquely determined for each source in the overlap region. If they also collide in PHA space, then bits are set in a source-map; further resolution is left to the user, since it is a model dependent problem.

The output events file is identical to the input Level 1 file (unless the event definition parameter is customized), with the addition of the diffraction coordinates and other attributes. These are defined in detail in the Level 1.5 ICD. In summary they are:

`TG_SRCID` source id (from source table);

`TG_PART` the region to which the event belongs (1 → HEG; 2 → MEG; 3 → LEG; 0 → zero-order; 99 → background);

`TG_SMAP` a bitmap indicating all sources the event could be from, if unresolved to a unique part;

`TG_R` diffraction coordinate (degrees);

`TG_D` cross-diffraction coordinate (degrees);

`TG_MLAM` order times wavelength (Angstroms);

`TG_M` diffraction order (1 or -1 for HRC);

`TG_LAM` wavelength (identical to `TG_MLAM` for HRC) (Angstroms).

`tgextract`

Inputs: Level 1.5 events.
Output: Level 2 PHA spectrum.

`Tgextract` bins Level 1.5 events into "PHA" spectra (here, PHA refers to a file format, rather than a data coordinate). Parameters control the sources, grating parts (HEG, MEG, LEG, HESF), orders, regions, and units. By default, orders from -3 to +3 (excluding zero order) are binned in wavelength in a cross-dispersion region wide enough to contain the astigmatic profile at maximum diffraction angle (defaults are given in the help document).

Background regions adjacent to the source region are also binned.

The output product is by default a Type II PHA file (in the OGIP definition): it has a row per binned spectrum, and source and order attributes are included on the spectrum record. Parameters allow Type I files (single order and source), as well as other units (keV, eV, mm, degrees).

The region mask derived from `tg_create_mask` defines an upper limit to cross-dispersion binning regions. By default, the spatial selection region is generous so that there is freedom to adjust the cross-dispersion limits after diffraction coordinates have been calculated. Note that shrinking the cross-dispersion binning limits can alter both the PSF fraction at any wavelength (via truncation of the cross-dispersion profile) and alter the line-spread-function (LSF), by truncation of scattering halos. Hence, this region is crucial to both the effective area and redistribution matrix generation. The regions are appended to the spectrum file as FITS extensions.

`specdetect`

Inputs: PHA file (Type I or Type II)
continuum file
Outputs: peaks file

This program finds local maxima in one-dimensional grating data. A peak is defined as any bin where the counts are higher than at both neighboring bins. Detection of absorption features and edges are not currently implemented.

`specmeasure`

Inputs: peaks file
Outputs: line measurements file

This program fits line profiles to detected peaks using the information in the peaks file created by the tool `specdetect`, as initial guesses for fitting parameters.

There are currently 2 integrated (and 2 non-integrated) fitting models: `lorentzInt`, `xsgauss` (and `gauss`, `lorentz`). Integrated fitting should be used since the fits should be integrated over the bins while fitting the peaks on top of the continuum level. The continuum fitting is always done using an integrated polynomial (`polyInt`). Results of the fits, along with min/max fit ranges are written to the outfile.

`specidentify`

Inputs: line measurements file
Outputs: identified peaks file
Reference Data: linelist file

This program uses the information from the input fitted peaks file (created by `specmeasure`) to correlate the list of observed lines with the database line list. The results show the number of line matches within the allowed wavelength error tolerance. The user can apply a number of element/ion filters on the linelist, in order to match only a certain group. The user can also specify the source temperature and density, which helps narrow down the matching criteria for the lines.

ISIS ISIS is an interactive system to facilitate atomic database browsing and filtering in conjunction with a counts or flux-corrected spectrum. ISIS is still under development and features will grow with experience on flight data.

ISIS information can be obtained from:

`http://space.mit.edu/ASC/ISIS`.

4 Examples

Example usage may be found in the help files.

5 Sample Data

Sample data files are available upon request. These are necessarily simulations, which have both known and unknown approximations. Soon there will be flight data.

A collection of simulations may be found on the HEAD network in

/proj/asc/dph/TestData. At Mit, data are in /nfs/wiwaxia/d4/dph/L15-sims/ASCDSoutput. The file, Sims.doc, describes each simulation briefly. The file names do not follow flight conventions. "AS" refers to Acis-S/Hetg, "HS" to Hrc-S/Letg; "_L1", "_L1a", and "_pha2" refer to Level 1, Level 1.5, and Level 2 (binned spectra), respectively. The same root-name with an "_off" string refers to an aspect offsets file.

6 Issues and Caveats

Extended sources - Extended sources present special challenges to both the adopted Level 1.5 and Level 2 data structures and interpretation of results. For Hetg, large spatial masks will collide near zero-order. This is not handled in the data-structure. Two passes could be made, with wide Meg and Heg regions to produce "long-slit" spectra. The wavelengths derived are ambiguous, since the spectrum is the convolution of the spatial structure with the dispersion relation. Some degeneracy is removed by the order and grating multiplicity (but standard analysis techniques do not yet exist).

Off-axis sources - Off-axis sources will produce collisions of spectra with Hetg (if there is a source near the aim-point). Custom masks can be generated (with tg_create_mask) to include photon resolution with tg_resolve_events, but this is not done automatically.

crowded fields - There is a maximum of 10 sources imposed on event resolution. If there are more than 10 in the field, the user must decide which 10 to include.

Letg **orders -** There is no implemented algorithm for Letg order resolution. There are a couple of un-analyzed bootstrap methods proposed:

1. Forward bootstrap (formally an "LU-decomposition", in which there is only an "L" matrix): uses high-energy, uncontaminated portion of a spectrum to model and subtract successive higher orders; suffers from noise and resolution degradation.
2. Backward bootstrap: given a feature, look for features at lower or higher order locations until consistency is achieved.
3. Forward folding: the traditional, source-model-dependent method: given a grid of source models, iterate until minimizing residuals of actual with with predicted counts.

Coordinates tool: In principal, each event has $2N$ (Hetg) or N (Letg) sets of diffraction coordinates, where N is the number of sources. Some are more probable than others, since each source has a distribution which covers the entire field, but with higher probability in the zero-order and along the dispersion centroid (hence, the economizing use of spatial masks in standard processing). A tool to provide alternate coordinates would be useful for ambiguous (unresolved) photons, in order to apply statistical methods given source models.

Order sorting table: tg_resolve_events requires a table made from Ccd response matrices, and implies an *a priori* determined fractional enclosed power. The Bi Ccds (S1 and S3) have poorly known response, so the Ccd energy (Pi) region must be fairly broad to have relatively constant power along the spectrum. Too broad a region could cause order overlap in PHA-space. Optimum conditions are TBD by real data.

Dependence of LSF, ARF on regions (spatial and PI) RMF and ARF calculation depend on spatial and PHA regions applied.

Pileup: It is possible that bright sources (X-ray binaries) will pile-up in the diffracted spectrum. In CCD-imaging analysis, this is un-correctable without a source model. With grating data, there may be enough information to determine a unique, source-model-independent solution, but no analysis has been done yet.

Frame-transfer streak: During CCD readout, a streak is created in the parallel read direction (roughly perpendicular to the dispersion) by aliasing of `CHIPY` coordinates (the CCD is still exposing while the frame is transferred). For bright sources, there may be enough counts from the zero-order to analyze. There is no standard software for this.

CC-mode: CC-mode is viable with HETG. Order sorting will be possible, provided good event-grading is possible. This is included in the standard suite (`tgdetect`, `tg_create_mask`, `tg_resolve_events`), but has not been well tested.

Roll variation: Until flight observations are done, we will not know how stable roll is. It is expected to be small, but if it exceeds certain limits, the spectral regions (in aspect corrected sky coordinates) will be broadened. If variation is too large, the observation intervals will have to be sub-divided to provide optimal regions and L1.5 files. This is not provided automatically at this time.

Background: OAC has shown to date that both ACIS and HRC backgrounds have been higher than expected. If they don't decrease, their affect is TBD on the grating processing.

7 Known Bugs

Wavelength scale: There is a several bin discrepancy in the wavelength scale, as determined from simulations with marx. These could be due to inconsistent `marx` and `pixlib` parameters, or to as yet undetermined affects. There seems to be both a linear offset between plus and minus orders, and a non-linear trend with diffraction angle (as determined from no-scattering simulations).

Zero-order in the gap: The `tg_resolve_events` aspect algorithm uses `CHIP` coordinates. If the zero-order goes into a gap or off the detector, these are undefined and the event coordinate determination will fail. A fix has been generally identified (use `DET` coordinates, which cover the gaps), but has not been fully analyzed or implemented.