# MIT Kavli Institute          Chandra X-Ray Center

**MEMORANDUM**
November 16, 2015

**To:**       Jonathan McDowell, SDS Group Leader; Kenny Glotfelty, DS Tools Lead

**From:**     David Huenemoerder, SDS

**Subject:** tgextract2 Specifications

**Revision: DRAFT** 1.3 (31 Aug 2006)

**URL:**      http://space.mit.edu/CXC/docs/docs.html#TBD

**File:**     ∼dph/h3/CXC/ICD/Specs_tgextract2-1.0/tgextract2.tex

## Relevant Documents:

1. "PHA (Spectrum) Files"
   http://cxc.harvard.edu/ciao/data_products_guide/pha_descrip.html

2. "Grating Data Products: Level 2 to CXC Archive Interface Control Document" (revision
   1.4—28 May 1999) (http://space.mit.edu/cxc/docs/ICD_TGL2-1.4.ps.gz)

3. "Grating Data Products: Level 2 to CXC Archive Interface Control Document" (revision
   2.1—25 July 2006) (http://space.mit.edu/cxc/docs/ICD_TGL2-2.1.ps.gz)

## 1   `tgextract2`: Interface Specifications & Algorithm Changes

The primary motivation for modifications to `tgextract` was to allow large and asymmetric
regions for better determination of background counts. Changes have already been made in
`tg_resolve_events` to support arbitrary polygonal spatial filters which can be used to pass events
through wider and asymmetric sky region filters. This is more important as longer observations
are made of fainter sources, or as archival analysis attacks weakly exposed features.

Secondary motives were to improve the efficiency of `tgextract` by reducing the number of
passes through the event file, and to simplify the parameter interface.

The new tool is not back-compatible, since the interface has been changed. The output file also
has a change in that `BACKGROUND_UP` and `BACKGROUND_DOWN` columns are no longer written, so a
few downstream applications will have to change. The background scaling information, `BACKSCAL`,
has also been generalized into an array, as required by possibly non-uniform regions. Some names
and conventions have been adopted from `dmextract` where appropriate in order to be as uniform
as practical with `dmextract` products.

## 1.1 Summary of Changes

1. Binning regions may be polygons. Regions are always from a `TGMASK2` file (such as the bow-tie already in `CALDB` for use with LETGS). Box-type regions will also be supported.

2. Background regions may be determined from the intersection of the region and the (dithered) detector boundary, or optionally from the region alone. Only one background counts array is written, and is the sum of _UP and _DOWN regions.

3. `BACKSCAL` can be computed from the region rather than read from a tabulation in the `TGMASK2` file, since it may be altered by the detector boundary or be of non-uniform width. For non-uniform polygonal regions, it cannot be easily tabulated in the mask file.

4. Backscales are written to each row for source and background regions. `BACKSCAL` is the source region cross-dispersion width in degrees. The background region's backscale will be called `BG_AREA` to be compatible with `dmextract` output, and will be the total cross-dispersion width of the background regions. If no backgrounds are extracted, then only the source region backscale will be written to output. *Contrary to prior CXC conventions,* `BACKSCAL` *will no longer be the ratio of background to source regions. Software which reads background data will be required to look for* `BACKSCAL` *(keyword or array),* `BG_AREA`, *associated exposures, and perform the proper background counts scaling.*

5. Any time filter on the input virtual file specification is used to modify the `EXPOSURE` and `EXPOSUR`$n$ in the header.

6. If determining the extent of the background from the events, statistics can be improved by including bad events, which are at least good for marking the detector boundary. Hence, input events for ACIS should not be filtered on grade or status or energy on input (that is, input is "Level 1.5", not "Level 2"). A parameter can be used to specify a filter to be applied before binning, since any virtual file specification on input pre-filters those events. For short exposures with poor background statistics, or for cases when it is known a priori that the region alone should be used to define backscale, a parameter setting will bypass the implicit method. If the implicit method is required but statistics are poor, the wavelength resolution used for `tg_d` limit-determination can be lowered via a parameter setting. HRC-S has sufficient background even after filtering that the pre-binning filter is not necessary.

7. Efficiency has been improved by limiting the number of passes through the event file. (The prior version made a pass for each order to be binned.)

8. The "error" parameter added to correspond to `dmextract` usage.

## 1.2 Parameter Interface

**infile:** Input `evt1a` file, as written by `tg_resolve_events`. This has the events to be filtered and binned. `tgextract` will filter on source, grating, orders, and cross-dispersion region. Any other data-model filter specifications can be applied to the input file as long as they are not in conflict with the implicit binning columns or with the `evt_filter` (see below).

**evt_filter** Filter to apply to events before binning into counts spectra. If determining region limits from the events, having bad events ("Level 1.5") helps to mark the detector. Example: ``grade=0,2,3,4,6,status=0''. This is generally unimportant for HRC-S, but can be significant for ACIS-S. (Default: "none").

**outfile:** Output `pha` file name, or rootname if requesting multiple file output via the "opt" parameter. (Required)

**opt:** `pha2` or `pha1`, for `pha` Type II vs `pha` Type I output, respectively. If "opt=pha1", then multiple files will be output for each order, grating, and source, using "outfile" as a rootname.

(Default: `pha2`)

**tg_srcid_list:** "all", or comma-separated integer list. Specifies `tg_srcid` filter. Multiple source processing can produce very complicated source-background region dependencies. It is up to the user to ensure that any overlaps are handled by filtering and binning regions in `tg_resolve_events` or here in `tgextract`. This program merely filters on the `tg_srcid` column in the input event file and loops over the source id when binning.

(Default: 1)

**ignore_source_id:** "yes" or "no"; If "no", then match `region_file`'s `tg_srcid` with events' `tg_srcid` column.

If "yes", then the first region found is used for all sources given by parameter `tg_srcid_list`, regardless of the region file's `tg_srcid`. If sources are well separated, then in diffraction coordinates regions are and any source-ID can be safely ignored. That is, one region suffices for all sources, rather than having a multi-source region file for which source and region must be paired.

(Default: `yes`)

**tg_part_list:** HETG | HEG | MEG | LETG | `header_value`; grating parts to process (`tg_part` column of event file; 1 for HEG, 2 for MEG, or 3 for LEG). HETG means both HEG and MEG.

(Default: `header_value`)

**tg_order_list:** Grating diffraction orders to process: "`default`", $comma-list$, $range-list$, or $@file$. Orders are specified by non-zero integers (use `dmextract` for zero order). Each order is assigned to a row in the output `pha` Type II file, or to a separate Type I file.

(Default: `default`, which means "-3, -2, -1, 1, 2, 3" for ACIS-S, and "-1, 1" for HRC-S.)

**region_file:** CALDB or $filename$. If `CALDB` then look-up a standard HETG or LETG region file. Otherwise, use the specified file, which must be a FITS region file with the same `CONTENT` as the `CALDB` files.

A value of "`none`" means that appropriate rectangular defaults will be chosen for HETG or LETG with ACIS-S or for LETG/HRC-S. These defaults can be overridden with parameters which specify the minimum and maximum `tg_d` coordinates for the background and source

3

regions. See below for parameters and default values. The format is described in detail in the help file or the "Grating Data Products: Level 2 to CXC Archive Interface Control Document" (`http://cxc.harvard.edu/ciao/data_products_guide/pha_descrip.html`). Regions may be created with tool (`mktgmask2` **TBD**). The file is of FITS `CONTENT` type of `TGMASK2`, which defines a region in grating diffraction coordinates: dispersion angle, `tg_r` and cross-dispersion angle, `tg_d` (both in units of degrees). Regions are either `POLYGON` or `BOX`, and may differ for grating type or order sign, but are identical for higher orders of the same sign. That is, -1 and -3 MEG regions are spatially identical, and hence only -1 and +1 order region geometry need to be specified.

The default (`CALDB`) files will be equivilant to CIAO 3 `tgextract` HETG parameters (rectangles), or the "bow-tie" (LETG, e.g. `letgD1999-07-22regN0002.fits`).

(Default: `CALDB`)

**wav_grid:** "use_header" or indirect parameter reference (of form "`wav_grid = )`*param_name*") to point to pre-defined grids (below). Alternately an explicit grid specification, given as `min:max:step`, or `min:max:#bins`.

If the value is "`use_header`" and the header's `GRATING` keyword value is `HETG`, then the `wav_grid_heg` and `wav_grid_meg` parameters (see below) are used to specify the grids.

It is possible to specify HEG and MEG grids which evaluate to a different number of bins. In this case, to maintain simplicity in the FITS file (fixed length arrays), the shorter grid will be extrapolated to have the same number of bins as the longer. For example, if HEG and MEG grid specifications were `1.0:21.48:0.0025` and `1.0:21.48:0.005`, respectively, then the MEG grid would be extrapolated to a longer-wavelength end point as `1.0:41.96:0.005`.

(Default: `use_header`.)

**wav_grid_heg:** "1.0:21.48:0.0025"; pre-defined standard HEG grid.

**wav_grid_meg:** "1.0:41.96:0.0050"; pre-defined standard MEG grid.

**wav_grid_leg:** "1.0:205.80:0.0125"; pre-defined standard LETGS grid.

**wav_grid_leg_acis:** "1.0:103.4:0.0125"; pre-defined standard LETG/ACIS-S grid.

**extract_background:** "yes" or "no"; Extract the local background spectrum from the background region specified in the input region file.

(Default: `yes`)

**backscale_method:** "EVENTS | REGION"; Use events and region, or region-only for background backscale computation. If using events, then then the maximum absolute extent of the region is determined by finding the maximum absolute `tg_d` in coarse wavelength bins, which are then interpolated to the resolution of the specified binning grid. This is so that if large background regions extend beyond the detector boundaries, `BACKSCAL` will be computed correctly.

A disadvantage of using events is that the background limit will have noise. Hence it is recommended to keep bad events up to this point (bad grades, bad status), so that they can be used to improve statistics in the geometric determination (especially for ACIS-S; generally

4

not necessary for HRC-S). The effect of noise is more likely for narrower background regions. (Bad events and grades should not, of course, be included in the output histograms; see `evt_filter`). Note that the source region backscale is always computed from the source region, and will be written even if "`extract_background = no`".

(Ignored if "`extract_background = no`".)

(Default: `EVENTS` )

`backscale_resolution:` Amount in pixels (as implied by `wav_grid`) to bin in wavelength to improve statistics in determination of background backscale from events (see `backscale_method`). The default is 64. (This does not change the output histogram grids.)

(Ignored if "`extract_background = no`".)

(Default: 64)

`error:` "gaussian" (default), or "gehrels"; method for determination of uncertainties. If "gaussian", then the keyword, `POISSERR` is set to `TRUE` and `_ERR` columns are omitted. If the value is "gehrels", then `POISSERR = FALSE` and uncertainty columns for both source and background are computed and written to output:

$$\mathtt{STAT\_ERR} = 1 + \sqrt{\mathtt{COUNTS} + 3/4}$$

$$\mathtt{BG\_ERR} = 1 + \sqrt{\mathtt{BG\_COUNTS} + 3/4}$$

(Default: `gaussian`)

`min_tg_d` The minimum `tg_d` value to include in the source spectrum, in degrees (at grating focal length). Default for ACIS-S is -6.6e-4 deg, and for HRC-S, -2.00e-3 deg. Ignored if a region file has been given. Use default if value is "`default`"

`max_tg_d` The maxmimum `tg_d` value to include in the source spectrum, in degrees (at grating focal length). Default for ACIS-S is 6.6e-4 deg, and for HRC-S, 2.00e-3 deg. (Ignored if a region file has been given.) Use default if value is "`default`"

`min_upbkg_tg_d` Specifies the minimum `tg_d` value to include in the "upper" background (`tg_d > 0`) side of the spectrum. If set to "default", the value of `max_tg_d` will be used. (Ignored if a region file has been given.)

`max_downbkg_tg_d` Specifies the maximum `tg_d` value to include in the "down" background (`tg_d < 0`) side of the spectrum. If set to "`default`", the value of `min_tg_d` will be used. (Ignored if a region file has been given.)

`max_upbkg_tg_d` Specifies the maximum `tg_d` value to include in the "upper" background (`tg_d > 0`) side of the spectrum. If it is set to "default", the values of 6.6e-3 and 1.33e-2 will be used for ACIS-S and HRC-S, respectively. (Ignored if a region file has been given.)

**min_downbkg_tg_d** Specifies the miniimum `tg_d` value to include in the "down" background (`tg_d` < 0) side of the spectrum. If it is set to "`default`", the values of -6.6e-3 and -1.33e-2 will be used for ACIS-S and HRC-S, respectively. . (Ignored if a region file has been given.)

**geompar:** "geom"; Parameter file for `pixlib` geometry files.

**clobber:** "yes" or "no"; If "yes", then overwrite any existing output file.

**verbose:** 0-5; Verbosity level. 0 for no display, except of errors.

**mode:** "ql"; parameter interface mode switches.

## 1.3 Changed or Removed Parameters:

1. The following have been deleted because they are always filled with default vaules. They are required by the OGIP file format and so will still be written to the output file. When response and/or background files have been calculated, then these can be set by by external programs (such as `dmhedit`):

   ```
   ancrfile
   respfile
   backfile
   rowid
   ```

2. The following parameter has been replaced with the "**opt**" parameter so that it is similar to `dmextract`:

   ```
   outfile_type
   ```

3. The following has been removed, since units are always given in Angstroms:

   ```
   bin_units
   ```

4. The following binning parameters have been replaced with grid specifications:

   ```
   min_bin_leg
   max_bin_leg
   bin_size_leg
   num_bins_leg
   min_bin_meg
   max_bin_meg
   bin_size_meg
   num_bins_meg
   min_bin_heg
   max_bin_heg
   bin_size_heg
   num_bins_heg
   ```

6

5. New parameters:

```
evt_filter                                                      |
ignore_source_id
opt
wav_grid
wav_grid_heg
wav_grid_meg
wav_grid_leg
wav_grid_leg_acis
backscale_method
backscale_resolution
error
```

6. Renamed parameters:

```
region_file  ( formerly inregion_file )
```

# 2  Output File Specification:

The output file is a standard CXC `pha2` Type I (scalar columns) or Type II (array columns, multiple orders) file. The full specification for the Level 2 PHA (binned spectrum) file will be given in the Interface Control Document (Document 3).

The region extension is of content type `TGMASK2`, and must tabulate the source binning region, since the cross-dispersion width is required by `mkgrmf` to compute the grating response matrices' efficiency factor. The background region is optional (**TBD**) starting in CIAO 4, since it may be complex, if determined from events. By default, the output background regions will be the same as the input, unmodified by any intersection with the detector boundary and so may be inconsistent with the event-based determination of `BACKSCAL`.

The output region format is similar to the input region format, except that the coordinates are in `TG_LAM`, `TG_D` (MTYPE=wavpos), whereas the input region was in `TG_R`, `TG_D` (MTYPE=dispang) coordinates.

Here we list the differences from the previous version.

## 2.1  New Columns:

BACKSCAL Source region `BACKSCAL`, cross-dispersion width (`tg_d`) of the source region in degrees (at grating plate scale) as a function of wavelength (on same grid as source counts).

BG_AREA Background region `BACKSCAL`, in total cross-dispersion width (`tg_d`) of the background regions (at the grating plate scale) as a function of wavelength (on same grid as source counts).

BG_COUNTS Background counts (sum of _up and _down parts of background regions, if multiple), on same grid as spectrum.

**BG_ERR:** Background uncertainty array (if "`error=gehrels`").

Since the regions may now be specified by non-uniform width (in `tg_d`) polygons, the backcales are not necessarily constant. Hence, the backscale values become arrays and are tabulated with the counts.

## 2.2   Deleted Columns

The following background components have been replaced by their sum in `BG_COUNTS`:

```
BACKGROUND_UP
BACKGROUND_DOWN
```