X-ray Data Analysis Overview

Randall Smith
Chandra X-ray Center

The text of this talk is available at
http://asc.harvard.edu/ciao/wrkshp.html

The fundamental unit in X-ray astronomy is the **event**.
Each **event** has a measured

- position on detector (x,y)
- pulse height (p)
- time of arrival (t)

Obviously, each of these is imprecisely known, and in fact the errors on each value are frequently correlated.

The event's detector position is then converted to a sky position using the telescope's aspect solution.

The event's pulse height (in a CCD/proportional counter/microchannel plate) is related to the X-ray's energy (in a complex fashion) via a detector's response matrix.

We address this by folding an assumed model of the source through the instrument response (so-called **forward-folding**) and comparing the measured and modeled results.

$$S_D(x,y,p,t) = T(x,y,t_\rho) \int dE \int d\vec{\rho} R(x,y,p,t,\vec{\rho},E,t_\rho) S(\vec{\rho},E,t_\rho)$$

where E (P) is the true (measured) energy of the photon, $t_\rho(t)$ the true (measured) arrival time, and $\vec{\rho}(x,y)$ the true (measured) direction. $T(x,y,t_\rho)$ is the "live time" of the instrument, and the function $R(...)$ is the telescope response, which includes the mirror area, point-spread function, detector efficiency and redistribution, and aspect solution.

So, the three most important issues are:

1. Calibration

2. Calibration

3. Calibration

Everything else is just manipulation of the event data–quite a powerful method of analysis since the data can be sliced or combined along any axis: position, energy, or time.

X-ray data is processed in a "pipeline", where the raw data is processed, calibrated, and evaluated. The steps in the Chandra pipeline will be discussed in more detail in a later talk, but just for a frame of reference:

Level 0: Nearly raw data: voltages, counts, frames.

Level 1: Scaled, calibrated, and aspect-corrected data. Each event now has an estimated sky position, time of arrival, and (possibly) an energy.

Level 1.5: (aka 1a; only for grating data) Each event has an associated dispersion from the zero-order image, and (when possible) an estimate of the order.

Level 2: Filtered, concatenated, and binned data. Only those data which are evaluated to be "good" are included; this includes selecting on event grades, position, and aspect times.

Then we get to the user-level data, where you can calculate lightcurve, make images and spectra, fit models, etc.

Basics of X-ray Astronomy: File Formats

Practically all datafiles are stored in FITS (Flexible Image Transport System) format. A FITS file consists of one or more blocks, each of which has a "header" and a "data" component. The header is simply a list of strings, which both describes the format of the data in the block and documents what the data contains:

...	...	...
HDUCLASS	OGIP	
HDUCLAS1	EVENTS	
HDUCLAS2	ALL	
ORIGIN	ASC	Source of fits file
MISSION	AXAF	Mission
TELESCOP	CHANDRA	Telescope
INSTRUME	ACIS	Instrument
DETNAM	ACIS-235678	Detector
GRATING	NONE	Grating
OBJECT	MBM 12	Source name
OBSERVER	DR. RANDALL SMITH	Principal Investigator
...	...	...

The data block can either be image or table data. Image data consists simply of an array of numbers. Table data has a format like a spreadsheet, with a some fixed number of named columns (such as **time, energy, ccd_id**) and a number of rows.

For historical reasons, the first block (known as the *PRIMARY* block) in any FITS file must contain image data. Frequently, this block is simply left blank (null) and the first "real" block is the second block.

The CIAO "Data Model" handles FITS files as well as files in the IRAF format QPOE in a context-sensitive manner.

Basics of X-ray Astronomy: The CALDB

The calibration database (CALDB for short):

- is a set of directories (approximately one per instrument)
- Each instrument directory subdirectories with basic calibration files (bcf) files and calibration product files (cpf). Together, these contains all the data needed to describe that instrument's response function.
- The data files are all in FITS format with descriptive headers
- These are used by the analysis software in order to create response matrices, telescope vignetting functions, and the like.

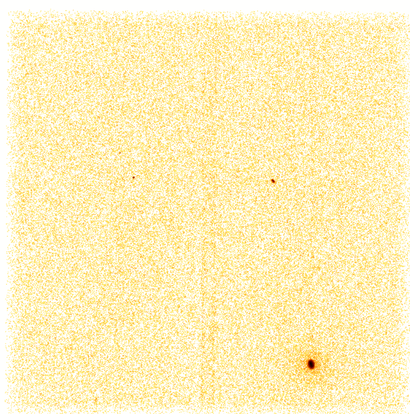
As the Chandra calibration effort continues, new versions of the CALDB are released. These may contain updated files, or all new files to handle new settings. For example, the ACIS focal plane temperature at launch was -90 C; currently, it is set at -120 C. Since the response of the CCDs is temperature-sensitive, different calibration files are needed. The most recent release of the CALDB is v2.3, as of March 7, 2001.

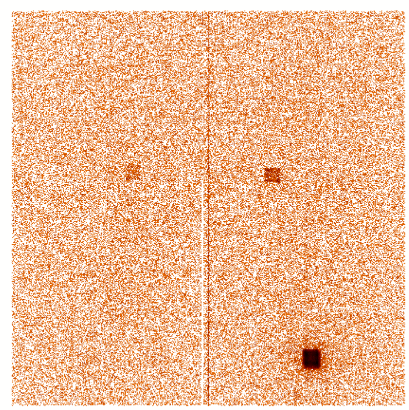
The CALDB concept is not Chandra-specific; other X-ray missions (ROSAT, ASCA, Einstein, XTE, etc) store their calibration data in the CALDB. However, the software to access this data has been mission-specific. The CXC has created the ARDLIB as a mission-independent way of accessing the CALDB.

Know Your Data

- Check the V&V report for any issues
- Check the Chandra webpage for any relevant caveats
- Make and examine the image of the entire event file
- Create and plot a lightcurve of the entire event file
- Check the aspect solution

For example, in this (early) dataset not all the bad columns were removed from the Level 2 file:

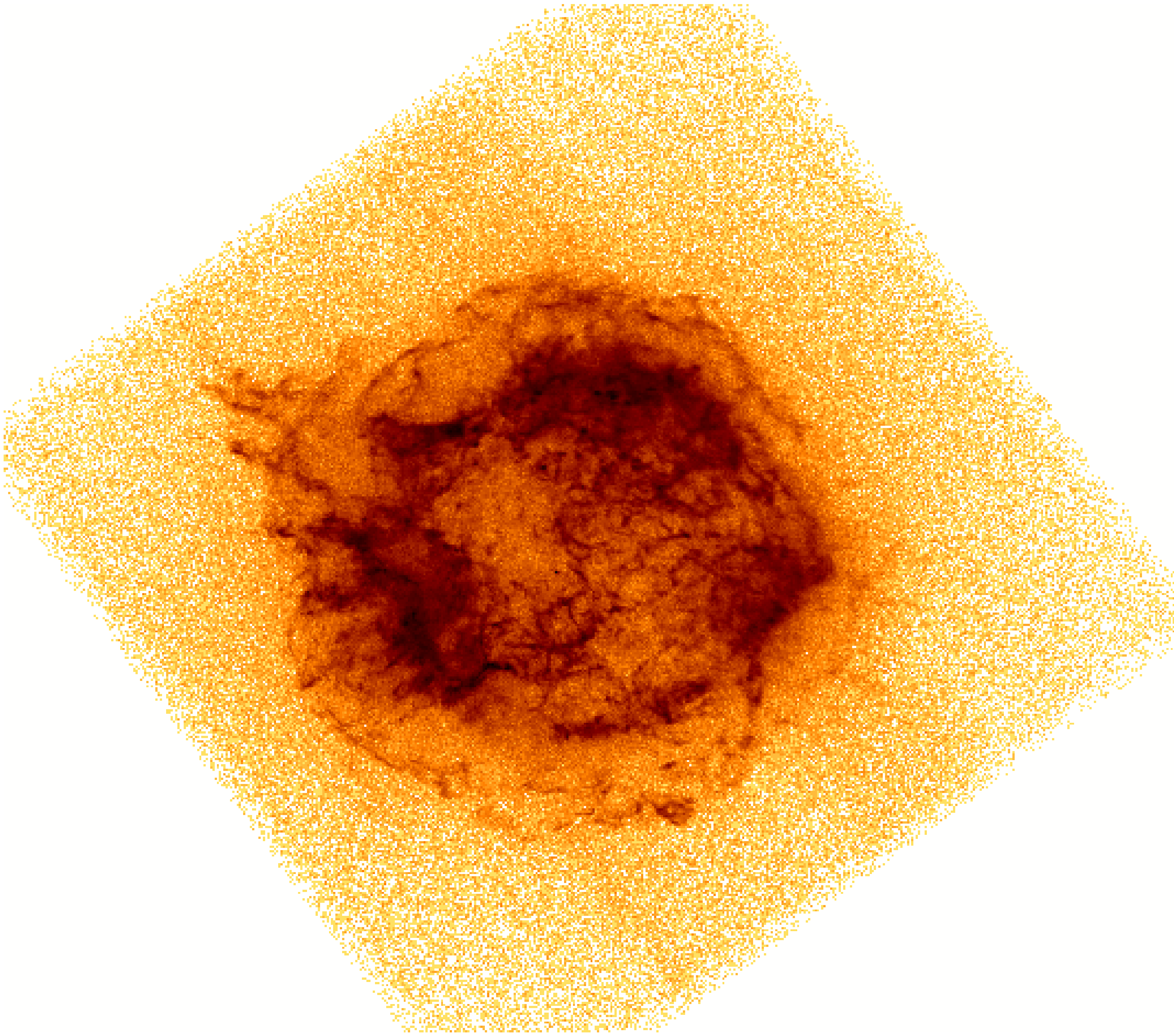




In this case, it is much easier to see the problem by viewing the file in detector (chip) coordinates, rather than sky coordinates. (Note: the command is `ds9 obsid_evt2.fits -bin cols detx dety`.)

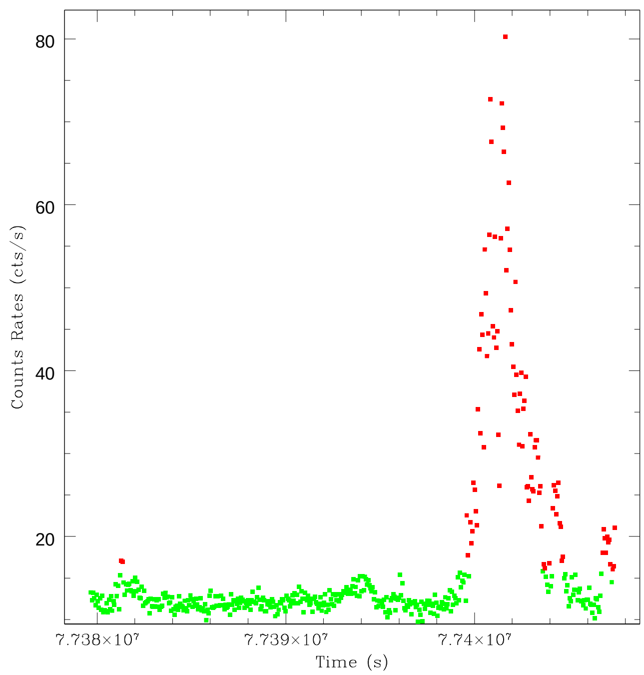
For a start, good files to examine are the *_evt2.fits, *_pha2.fits, *_aoff1.fits. This can be easily done with the **`firstlook`** tool, along with the oif.fits file for an observation. In general, the column names are given recognizable names.

Cas A



This image looks pretty good; no bad pixels can be seen. There may be a bad column or two, but the faint streaks could also be due to out-of-time events.

Checking the lightcurve for your observation is vital, as this is not in general checked by either the pipeline or in V&V. Solar flares can easily cause huge temporary flares in count rates. For point sources, this may not be important, but for extended sources the data taken during the flare should probably be removed:



Lightcurve for 3C273 Observation

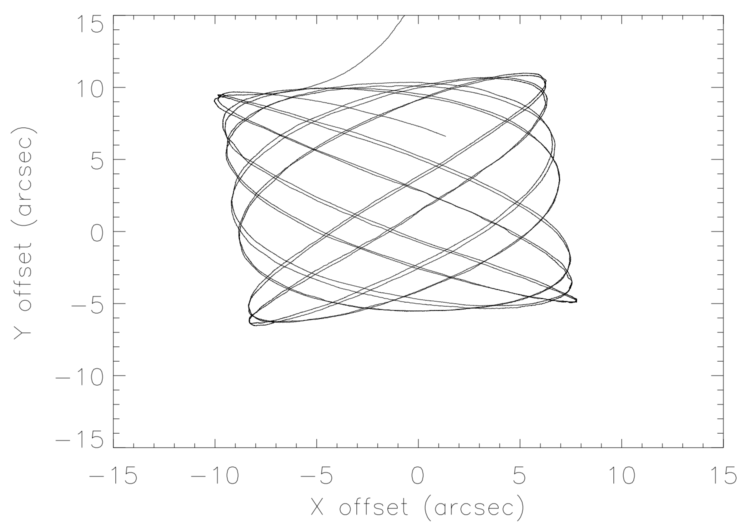
The **`analyze_ltcrv.sl`** script will attempt to identify flares in an automatic fashion, as shown above.

Chandra dithers about the pointing direction, in order to

- fill in chip gaps
- smooth over pixel-to-pixel variations
- limit radiation damage to pixels from ultra-bright sources

Effects of dither:

- Makes responses observation-specific
- Eliminates the need for pixel-level calibration
- Bad pixels appear as “dither patterns” in sky projection



Chandra’s pointing direction over a 10 ksec observation

(This aspect information is in the `obsid_aoff1.fits` file.)

Reprocessing: Necessary or Not?

Every X-ray mission goes through multiple iterations of processing, as the instrument calibration is improved, software is improved, and so on. Each Chandra dataset contains all the datafiles from Levels 1, 1.5 (if applicable), and 2. When deciding whether or not to reprocess your data, ask the following questions:

- Do the Level 2 images, lightcurves, spectra, and aspect solution appear reasonable? Are there obvious outliers that are not excluded? If so, reprocess from the Level 1 data.
- Has there been a new CALDB released with relevant updates since this data was processed? This can be checked on the Chandra main page (`http://asc.harvard.edu/`).

If you decide reprocessing is required, check out the Chandra threads at `http://asc.harvard.edu/ciao/documents_threads.html` to learn how.

In many cases, reprocessing from Level 1 is not required; instead, the level 2 can be filtered as necessary to destreak the data, or to detect sources, remove noisier pixels, and so on.

Spatial: (use `dmcopy` to make images)

- Either point or extended source
- Usually on the ACIS-I, ACIS-S3, or the HRC-I detectors

Spectral: (use `dmextract` or `tgextract` to make spectra)

- Low-resolution ($E/\Delta E \sim 10$): ACIS-I, ACIS-S3
- High-resolution ($\lambda/\Delta\lambda \sim 1000$): HETGS, LETGS

Temporal: (use `lightcurve` to bin events)

- Low-resolution: ACIS-TE, HRC-I
- High-resolution: ACIS-CC, HRC-S

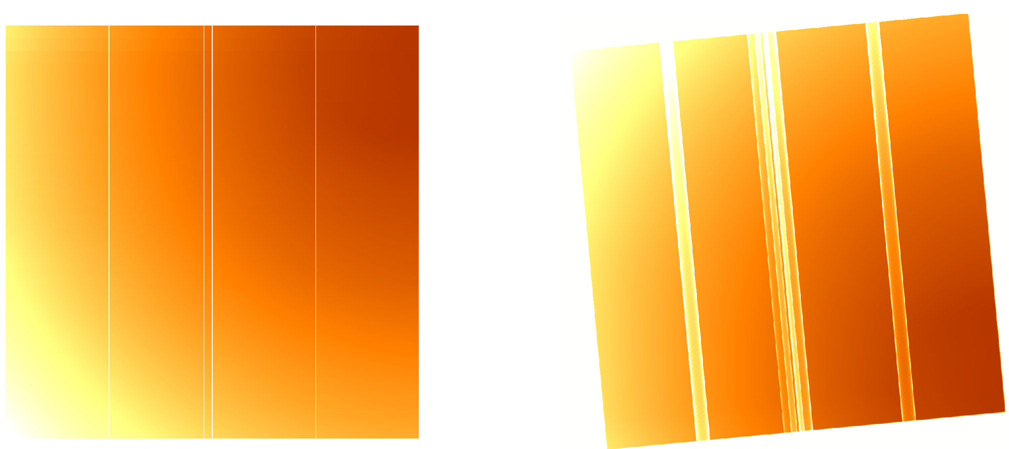
To interpret any of these, the telescope response must first be calculated–which is usually trickier than making an image, spectrum, or lightcurve.

The telescope response for spatial data is called the **exposure map**, which has units of area × time (cm^{2}s). This keeps the spatial information, at the expense of the spectral:

$$\int_{E-\Delta E}^{E+\Delta E} S(E,\vec{\rho})dE \approx \frac{C(p,x,y)}{A(E,\Delta E,\vec{\rho})}$$

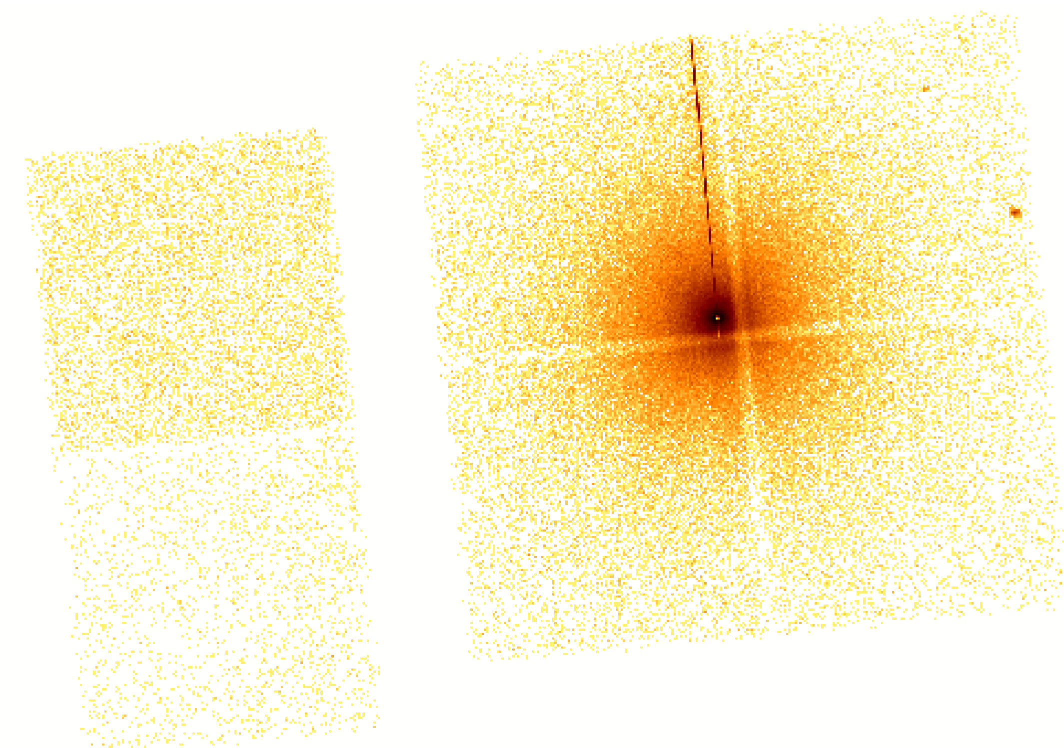
where $S(E,x,y)$ is the true source flux (ph/cm^2/s) at energy E and positions (x,y), $C(p,\Delta p,x,y)$ is the number of counts (events) with pulse heights in the range $(p-\Delta p, p+\Delta p)$, at position (x,y), and $A(E,\Delta E,x,y)$ is the effective area integrated over energies $(E-\Delta E, E+\Delta E)$ at position $\vec{\rho}$.

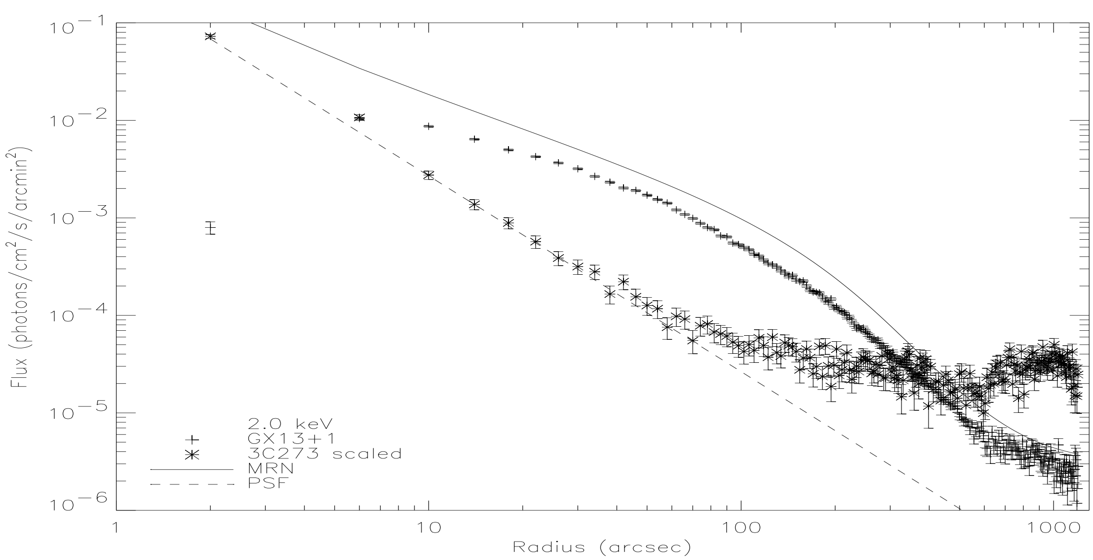
Exposure maps for Chandra are made in two stages: first, the instrument map (which includes the mirror area and the detector QE over the chosen energy band) is calculated, and then the exposure map, which includes aspect information. They can be made for a single energy (delta function) or can be integrated over a source spectrum.



(a) Instrument Map for ACIS-I3 (b) Exposure Map for ACIS-I3

The low-mass X-ray Binary GX13+1:





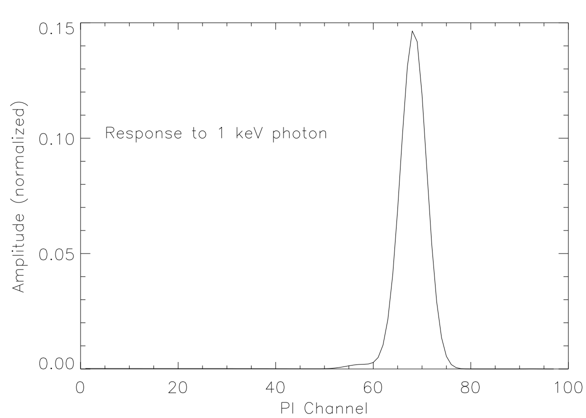
The ACIS CCDs provide low-resolution spectral data. Each event has an associated **PHA, PI**, and **grade**. The **PHA**, a value between 0 and 4095, is the "pulse height amplitude" of the event, which is directly related to the number of electrons transferred from the CCD pixel. The **PI** is the "invariant pulse height," with a value between 0 and 1023, which has been corrected for any positional or gain variations. The ACIS **grade** is a number between 0 and 7; good events have grades 0, 2, 3, 4, or 6. (Grade 0 events are the "cleanest" and grade 6 events the "noisiest"). Either PHA or PI values can be used to make a spectrum; PI is recommended, however.

The relationship between the "true" spectrum and the measured one is non-linear:

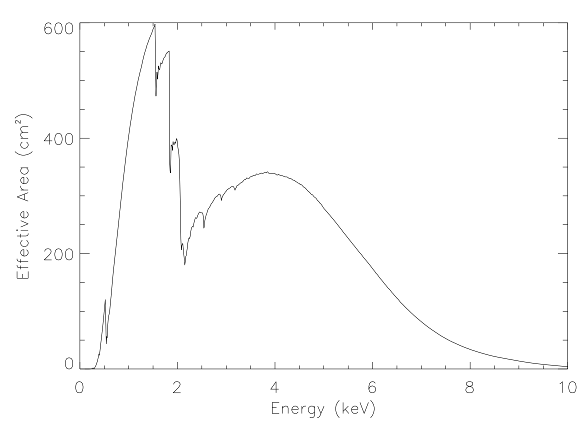
$$C_\Omega(p) = \tau_{eff} \int dE D_R(p, E) A_\Omega(E, \vec{\rho}) S(E)$$

where $C_\Omega(p)$ is the number of counts at pulse-height p in region Ω, τ_{eff} is the effective observing time, and $S(E)$ the true spectrum.

- $D_R(p, E)$ is the Redistribution Matrix Function(File), or RMF
- $A_\Omega(E, \vec{\rho})$ is the Auxiliary Response Function(File), or ARF. This is the effective area of the telescope at energy E to photons from $\vec{\rho}$ in region Ω.



(a) RMF at 1 keV for a point on ACIS-I



(b) ARF for same position

Additional Notes

PHA file Pulse-height spectra (of either PI or PHA) can be made with the `dmextract` tool.

RMF, ARF creation These are made with the **mkrmf** and **mkarf** tools, respectively.

Spectral Modeling Once you have a PHA file, and an RMF and ARF for your source, you can use Sherpa (or XSPEC, SPEX, or ISIS, . . .) to create a spectral model, fold it through the RMF/ARF, and compare the result to the PHA file.

Backgrounds Especially for dim sources, a "background" region should be selected and a PHA spectrum created. This can then be subtracted from the source PHA, or modeled separately and summed with the source model.

Extended Sources For ACIS, both the RMF and ARF are somewhat position-dependent. Methods for created "averaged" ARF and RMFs are in progress.

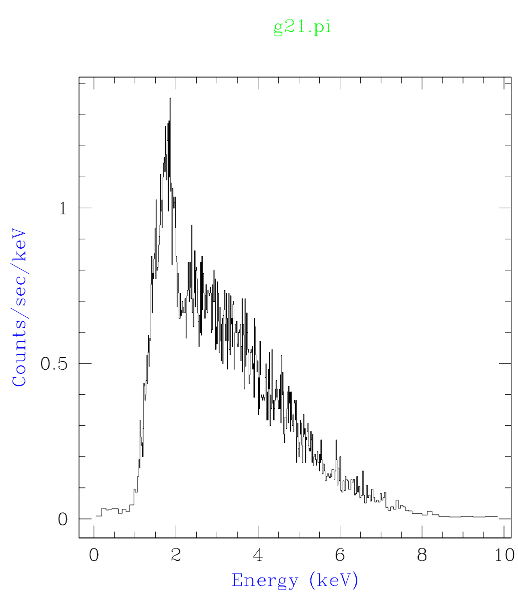
Pile-Up CCDs are subject to pile-up, if two photons strike the same pixel in the same frame. Two 1 keV photons in the same pixel and frame, for example, may appear to be a single 2 keV photon. Pile-up percentages can be calculated from the source's flux and spectrum; if more than 2-3% of photons are "piled-up," care should be used. ISIS (`http://space.mit.edu/CXC/ISIS`) has a pile-up model in this case.

Yet More Notes

Energy & PI There is a linear relationship between PI and energy, which is useful (if the limits on the CCD resolution are kept in mind). Each event file has an **Energy** column (units of eV) that can be used for filtering.

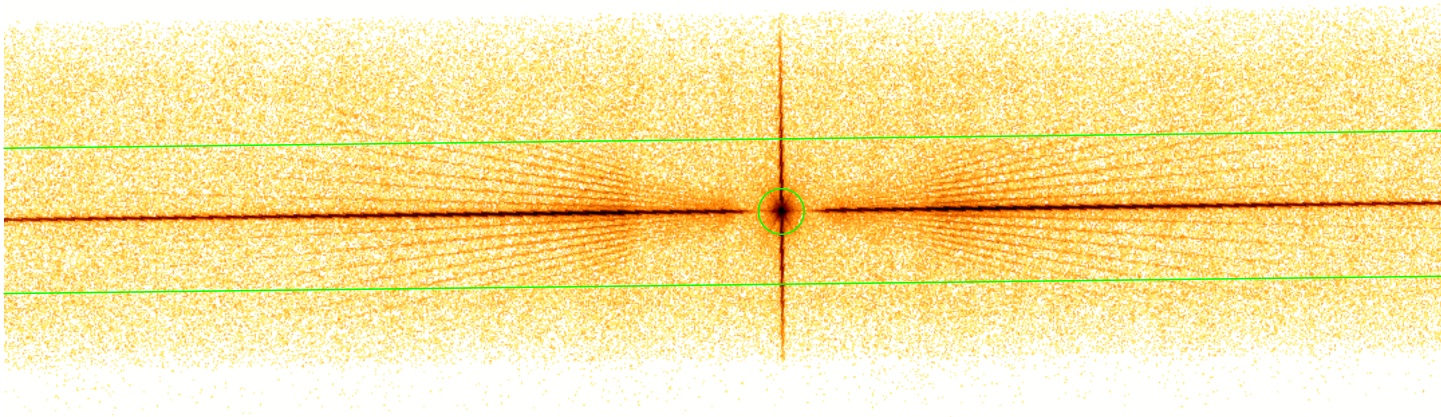
Grouping Spectral must be binned before it is fit. This binning does **not** have to have constant spacing. Bins can be grouped together with **dmgroup** in order to increase the number of counts per bin.

Fitting
- χ^2 statistics systematically underestimate the flux of low-count data, because the errors are underestimated.
- Use χ^2 if the errors are Gaussian distributed (*i.e.*, $\gtrsim$ 10 counts/bin).
- For low-count data, either (1) bin the data, (2) use an approximate error estimate, such as CHI GEHRELS, or (3) use a Bayesian method, such as the CASH statistic.



A 20,000 count spectrum from the SNR G21.5

High-resolution (grating) spectra on Chandra cover a huge range of wavelengths: from 1.2-170Å, over two orders of magnitude. Note here that we switch from energy to wavelength as the natural unit, since all the high-resolution data are from gratings.

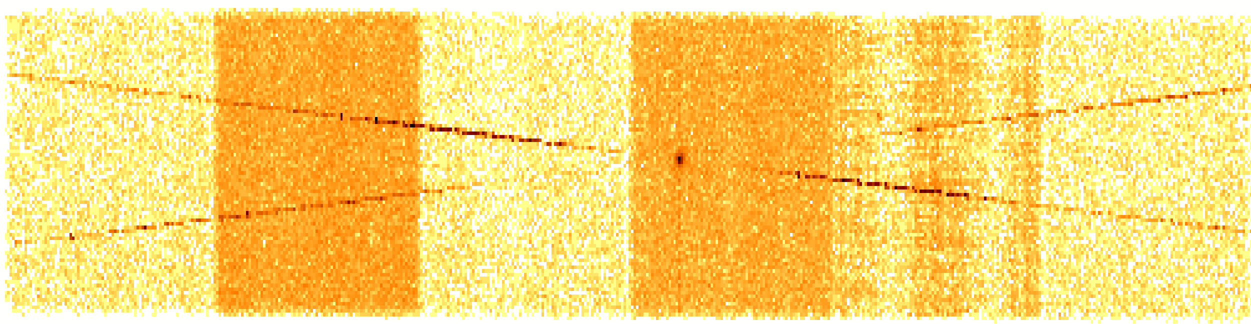


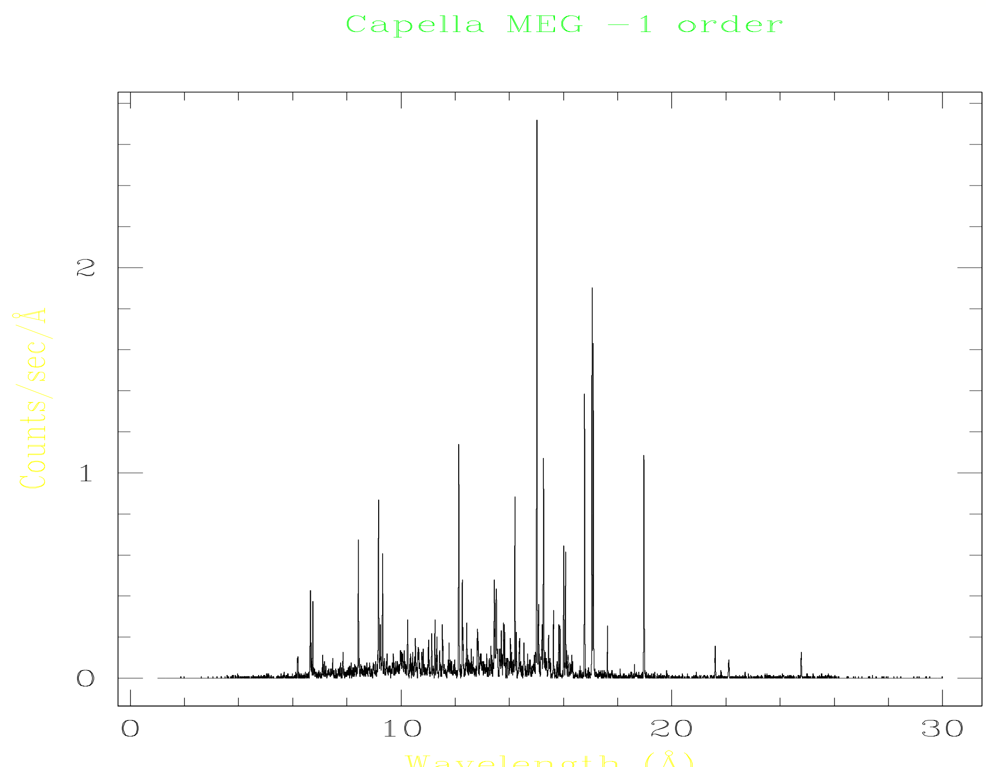
LETG/HRC-S Observation of NGC6624

Extracting a spectrum in this case requires finding the zero-order image, measuring each event's position relative to it, and using the dispersion relation to calculate $m\lambda$, the order$\times$wavelength. If ACIS is the detector, the CCD resolution can be used to distinguish between different orders; on the HRC, this must be modeled.

Clearly, the spatial and spectral elements are tightly coupled. If the zero-order image is slightly displaced (as can easily happen with heavily piled-up sources), the $\pm$ order wavelengths will be offset from each other. (If this occurs, you may wish to reprocess with a new position measured by hand.)

Capella with the HETG/ACIS-S detector:





Spectrum from the MEG -1 order

Grating PHA For ACIS grating data, 12 spectra are created ($\pm 1, 2, 3$ orders for both the HEG and MEG). For HRC data, 2 spectra are created ($\pm \sum_n i$ orders).

Grating ARF This can be created for each order using `mkgarf`; it gives the effective area for each order.

Grating RMF Optional; this is the Line Spread Function (LSF) for grating data, but this can be modeled using gaussians.

In Sherpa, grating spectra can be jointly fit to a model. Or, each PHA can be extracted and divided by the ARF to create a fluxed spectrum, which can be fit using any fitting program. Plus and minus orders can be co-added to increase count statistics, although it will inherently decrease resolution.

Currently, CIAO has few analysis tools for temporal data. The main tool is `lightcurve`, which can be used to bin event data by exposure time per bin, counts per bin, or rate. `lightcurve`'s output is a FITS file, which can be plotted with `chips`.

ACIS-TE Fundamentally limited by the 3.2 s frame time.

ACIS-CC Much higher time resolution (3 msec) at the expense of one spatial dimension.

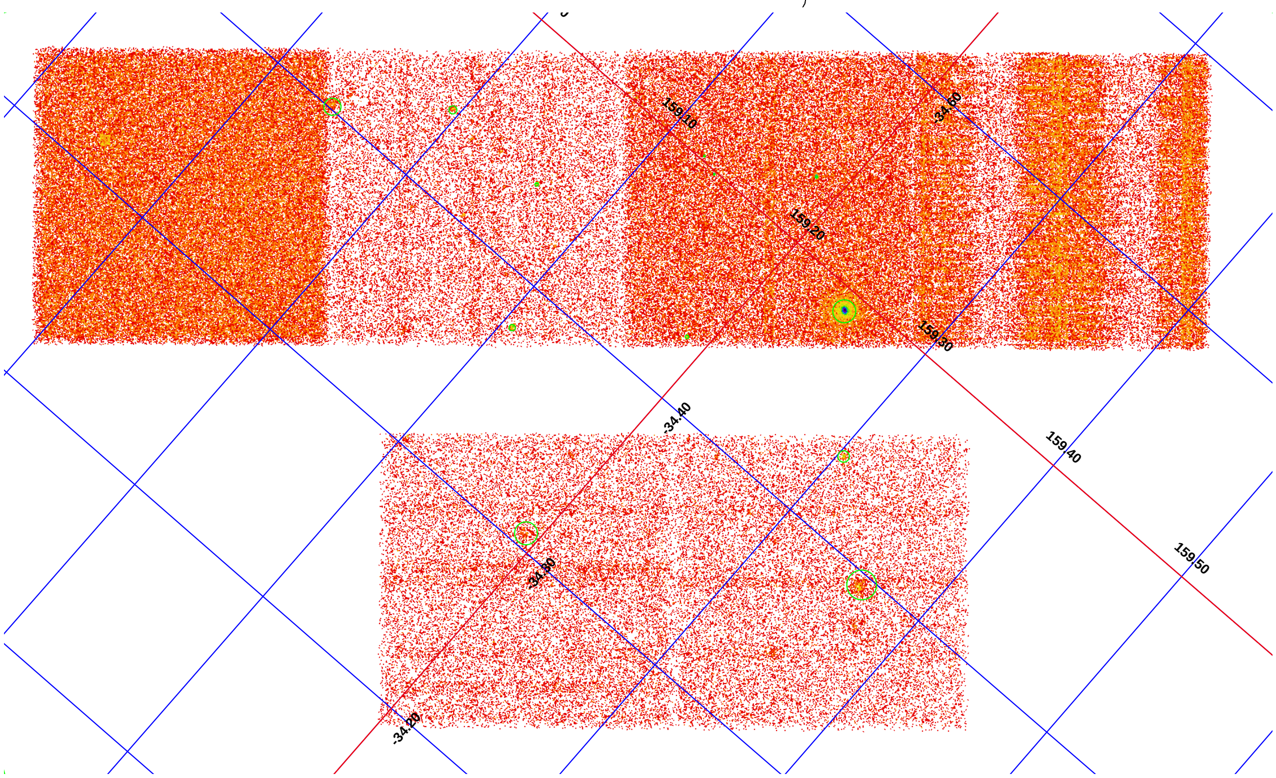
HRC-S Can achieve 16 μsec timing accuracy if only the central plate is on and telemetry not saturated.

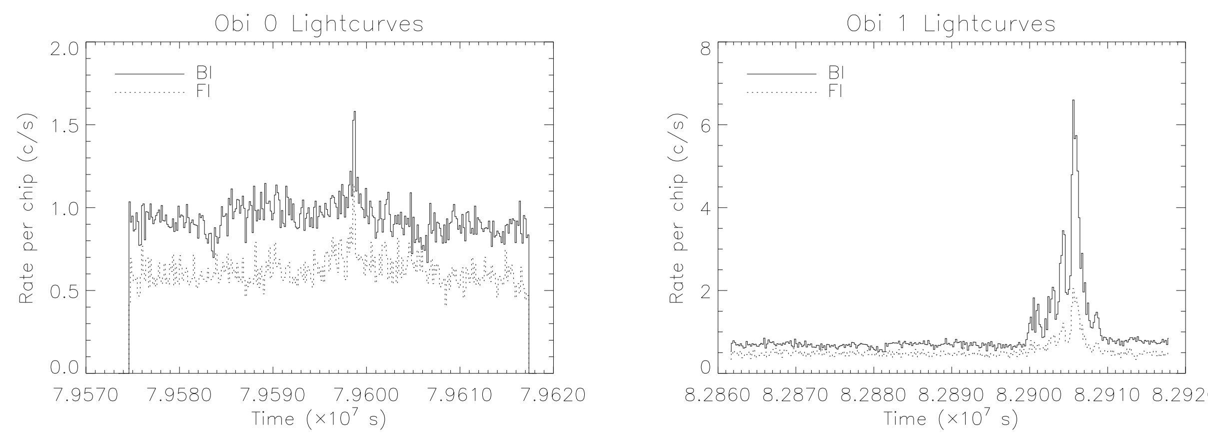
The lightcurve "rate" can then be divided by the effective area (ARF) to get a flux.

The Chandra dither will smooth out most variations in the effective area (ARF), and so a time-resolved ARF will likely not be needed (unless the source is on a chip gap and has variability on the scale of the Chandra dither ($\sim$ 1 ksec).

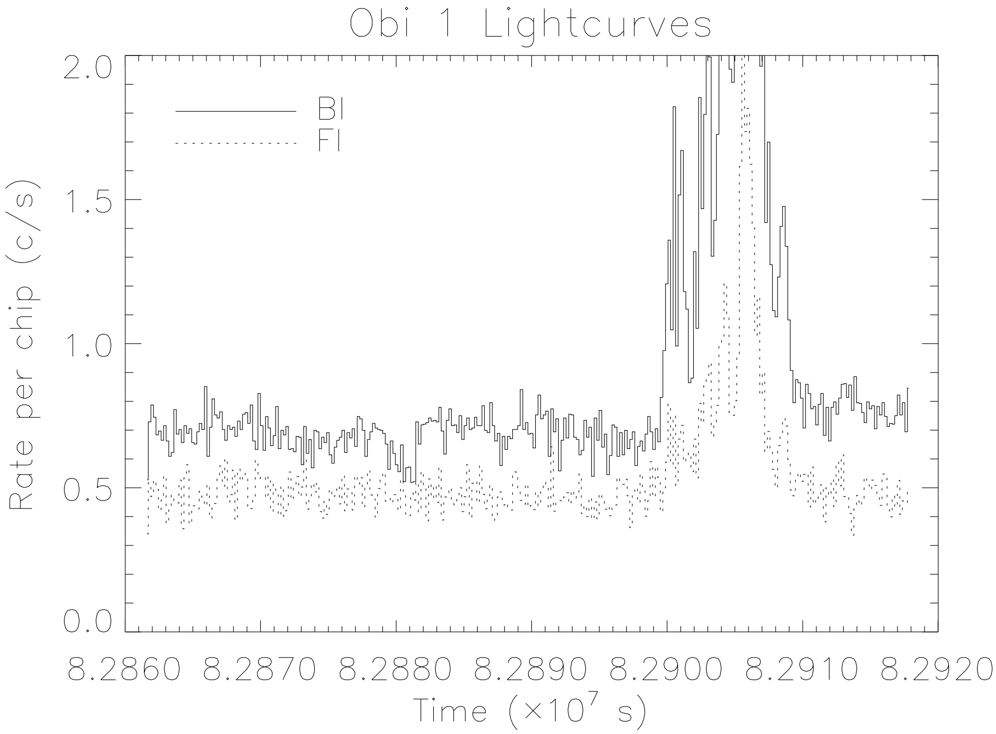
An important use of the `lightcurve` tool is to find flares in the data, which may be from the source or from more local problems, such as the Sun.

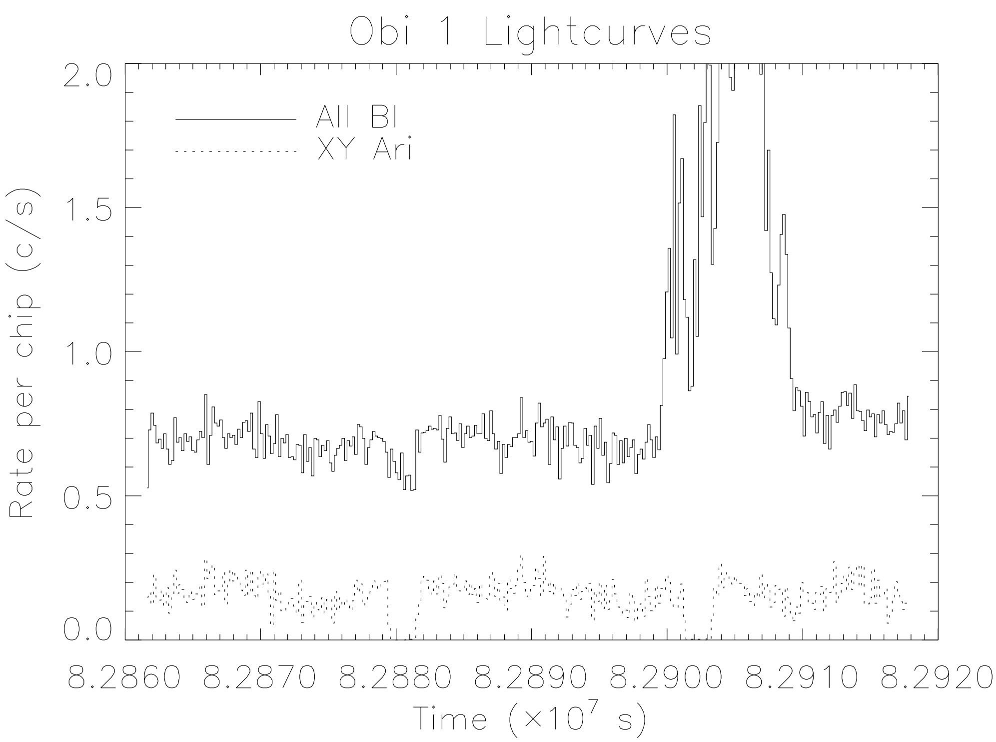
The nearest molecular cloud, MBM 12:





Lightcurves from each observation


Obi 1 Lightcurves
BI
FI
Rate per chip (c/s)
2.0
1.5
1.0
0.5
0.0
8.2860
8.2870
8.2880
8.2890
8.2900
8.2910
8.2920
Time (×10^7 s)


Obi 1 Lightcurves
All BI
XY Ari
Rate per chip (c/s)
2.0
1.5
1.0
0.5
0.0
8.2860
8.2870
8.2880
8.2890
8.2900
8.2910
8.2920
Time (×10^7 s)

In low-resolution spectra, the number of counts will determine what can be done:

$\sim$ **100:** Enough for a hardness ratio

$\sim$ **1000:** A poor spectrum that may be able to distinguish between a power-law and a thermal source

$\sim$ **10,000:** Some strong individual lines can be measured.

In high-resolution spectra, the following are important:

Continuum: Many thermal sources have far more lines than our models include. Be careful to fit the continuum in line-free regions only.

Line Shapes: Lines can be modeled as gaussians, or as delta functions if an RMF is used. If line shape (broadening) is important, a RMF should be used.

Adding Orders: Will increasing counts, at a cost in resolution. Joint fitting will likely give the best results, after a co-added analysis is done to guide the eye.

Know your Data: The Event File is your Friend

unix% dmlist acis_evt2.fits cols

Columns for Table Block EVENTS

Col	Name	Unit	Type	Range	Description
1	time	s	Real8	7.7e7:7.8e7	S/C TT corresponding to mid-exposure
2	ccd_id		Int2	0:9	CCD reporting event
3	node_id		Int2	0:3	CCD serial readout amplifier node
4	expno		Int4	0:2147483647	Exposure # of CCD frame
5	chip(chipx,chipy)	pixel	Int2	1:1024	Chip coords
6	tdet(tdetx,tdety)	pixel	Int2	1:8192	ACIS tiled detector coordinates
7	det(detx,dety)	pixel	Real4	0.50:8192.50	ACIS detector coordinates
8	sky(x,y)	pixel	Real4	0.50:8192.50	sky coordinates
9	pha	adu	Int4	0:36855	total pulse height of event
10	energy	eV	Real4	0:1.e6	nominal energy of event (eV)
11	pi	chan	Int4	1:1024	pulse invariant energy of event
12	fltgrade		Int2	0:255	event grade, flight system
13	grade		Int2	0:7	binned event grade
14	status[4]		Bit(4)		event status bits

unix% dmlist hrc_evt2.fits cols

Columns for Table Block EVENTS

Col	Name	Unit	Type	Range	Description
1	time	s	Real8	6.9e7:7.0e7	time tag of data record
2	rd(tg_r,tg_d)	deg	Real4	-2.0: 2.0	Grating angular coords
3	chip(chipx,chipy)	pixel	Int2	1:4096	Chip coords
4	tdet(tdetx,tdety)	pixel	Int4	1:49368	Tdet coords
5	det(detx,dety)	pixel	Real4	0.50:65536.50	Det coords
6	sky(x,y)	pixel	Real4	0.50:65536.50	Sky coords
7	chip_id		Int2	1:3	
8	pha		Int2	0:255	
9	pi		Int2	0:255	
10	tg_m		Int2	-62:62	Diffraction order (m)
11	tg_lam	angstrom	Real4	0: 400.0	wavelength (lambda)
12	tg_mlam	angstrom	Real4	-400.0:400.0	Order times wavelength (m * lambda)
13	tg_srcid		Int2	0:32767	source ID, index from detect table
14	tg_part		Int2	0:99	HEG, MEG, LEG, HESF regions
15	tg_smap		Int2	0:32767	source map; flags for up to 10 sources
16	status[4]		Bit(4)		event status bits