CUC April 2007

SCIENCE DATA SYSTEMS

INTRODUCTION

1) Chandra Source Catalog

2) Data processing status

3) CIAO: Schedule and plans

- CIAO3.4 released
- CIAO4-Beta1 imminent
- Future releases; Chips/Sherpa

4) Development work (Mike Nowak)

CHANDRA SOURCE CATALOG

Since the last CUC:

- The SDS/DS team has made significant progress on a number of science issues and improvements to the prototype pipeline
- We have developed a much more detailed and realistic schedule and assigned extra resources
- There remain significant schedule risks

I will address science progress, schedule issues and challenges

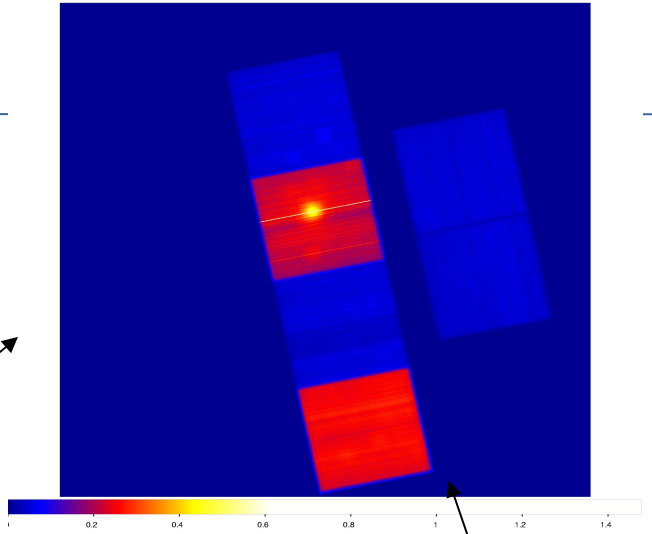
Science Highlights since last CUC meeting

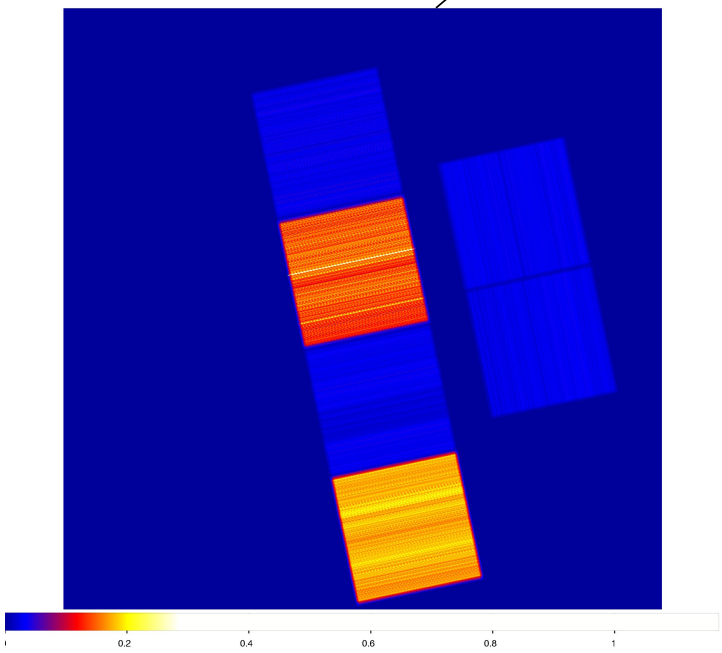
- Requirements and specifications delivered
  - ACIS low frequency background and readout streak background specifications (excluding chip edge and exposure map refinements)
    - Modified Poisson mean using pixel value histogram on 128x128-pixel sliding patches as low pass filter.
    - Recombine result with streak map
    - Normalization is currently biased, losing some sensitivity: working to improve
  - Aperture flux specifications
    - confirmed wavedetect regions are robust (small variation in aperture correction)
  - Hardness ratio specifications

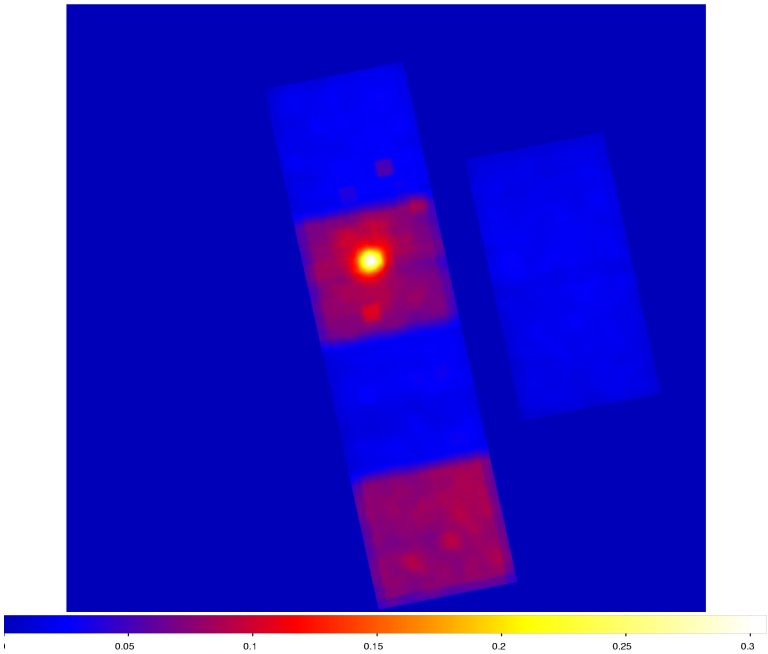
    Using Bayesian BEHR algorithm (received prototype code from Calibration group)
  - Intra-observation time variability analysis specifications
  - Merge pipeline specifications

    Source matching and source property combination (excluding errors)
  - Sensitivity map specifications
  - User interface requirements (revision)
  - Delivered Chandra Source Catalog Requirements document version 0.5(maintained by Ian)

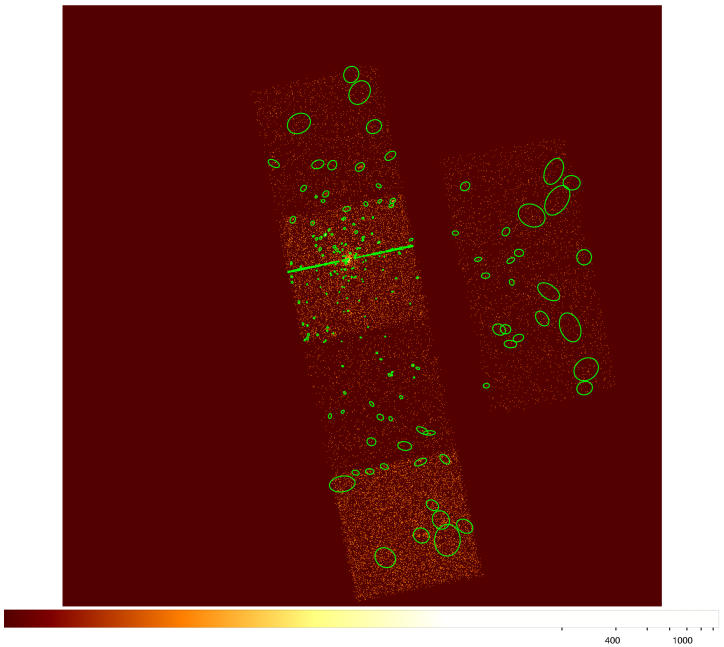
High Freq and Low Freq maps combine to give final background map

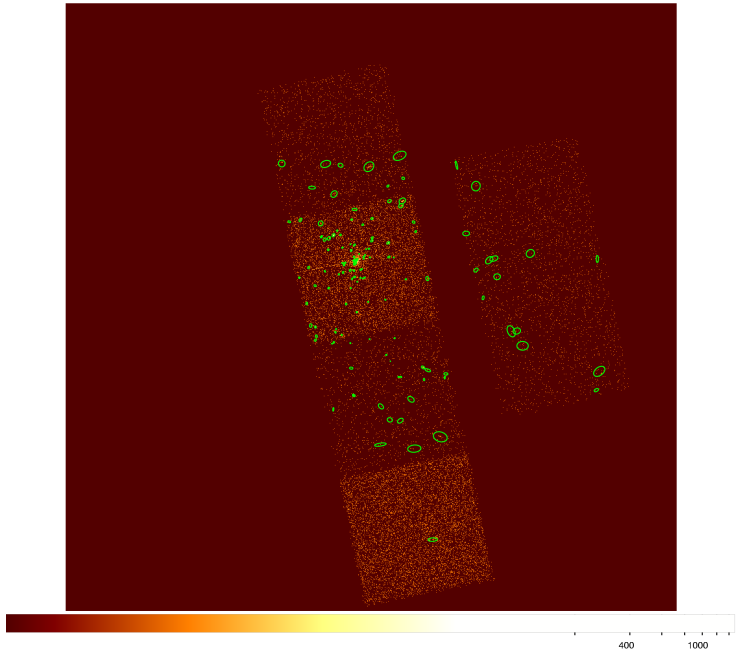




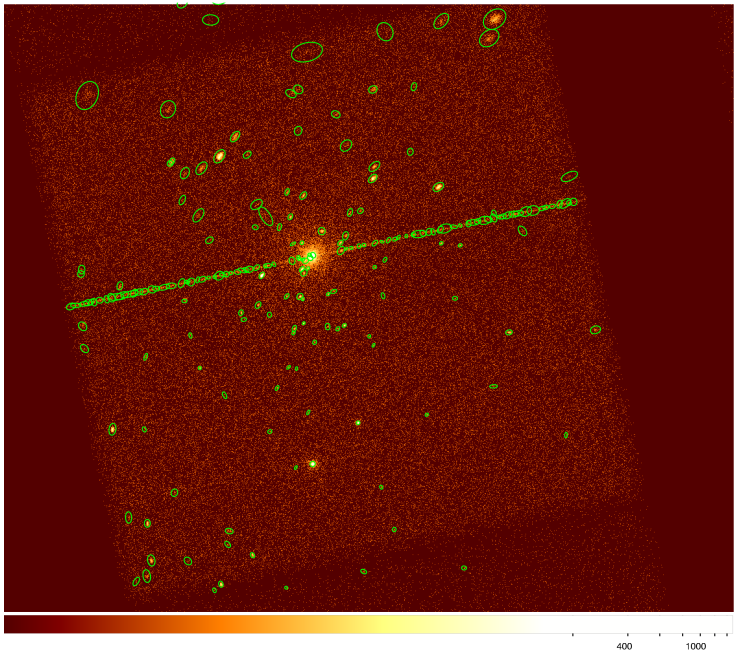


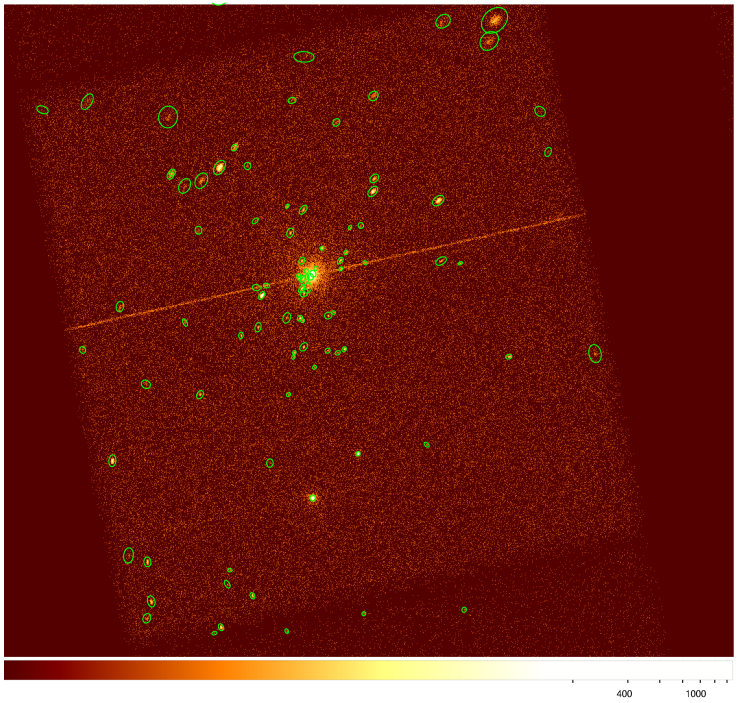
Incorporating Background in Wavdetect suppresses false sources





Streak Map is very efficient at removing spurious sources due to readout effects





Chandra Source Catalog Progress

Science Highlights since last CUC meeting

- Science studies
  - Characterizing and evaluating `wavdetect` outputs
    - Confirmed that source regions are useful for photometry (PSF correction is small)
    - Position errors agree with empirical relation derived from ChaMP statistical study
    - Investigating anomaly in source properties associated with combining scales
  - PSF model study
    - Evaluated number of rays needed for PSF models
    - Concluded that monochromatic PSFs in each band are adequate
    - In both cases, statistical errors dominate
  - Simulated source properties
    - Completed architecture to easily add simulated sources to real data
    - Using to evaluate position and extent errors

Infrastructure Highlights

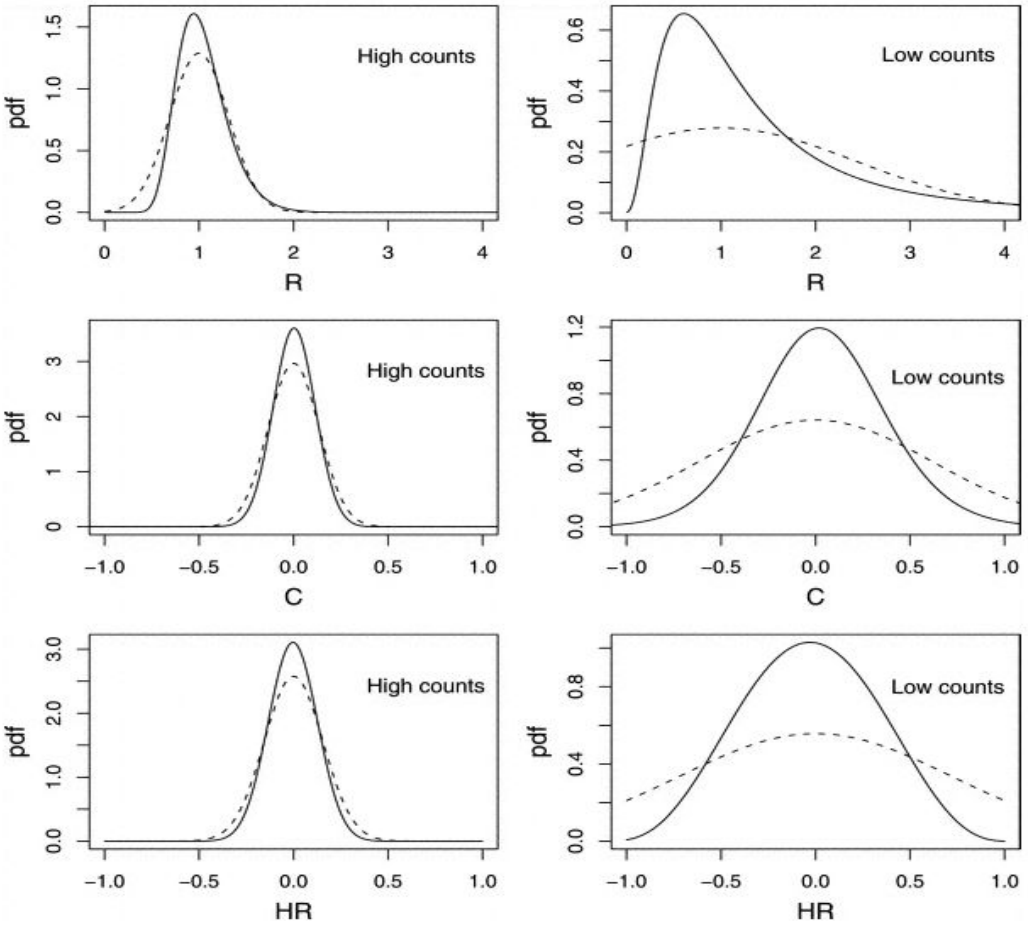
Revised archive data volume estimates: 5 TB per copy at end of mission

Updated data products list; working on quality assurance plan

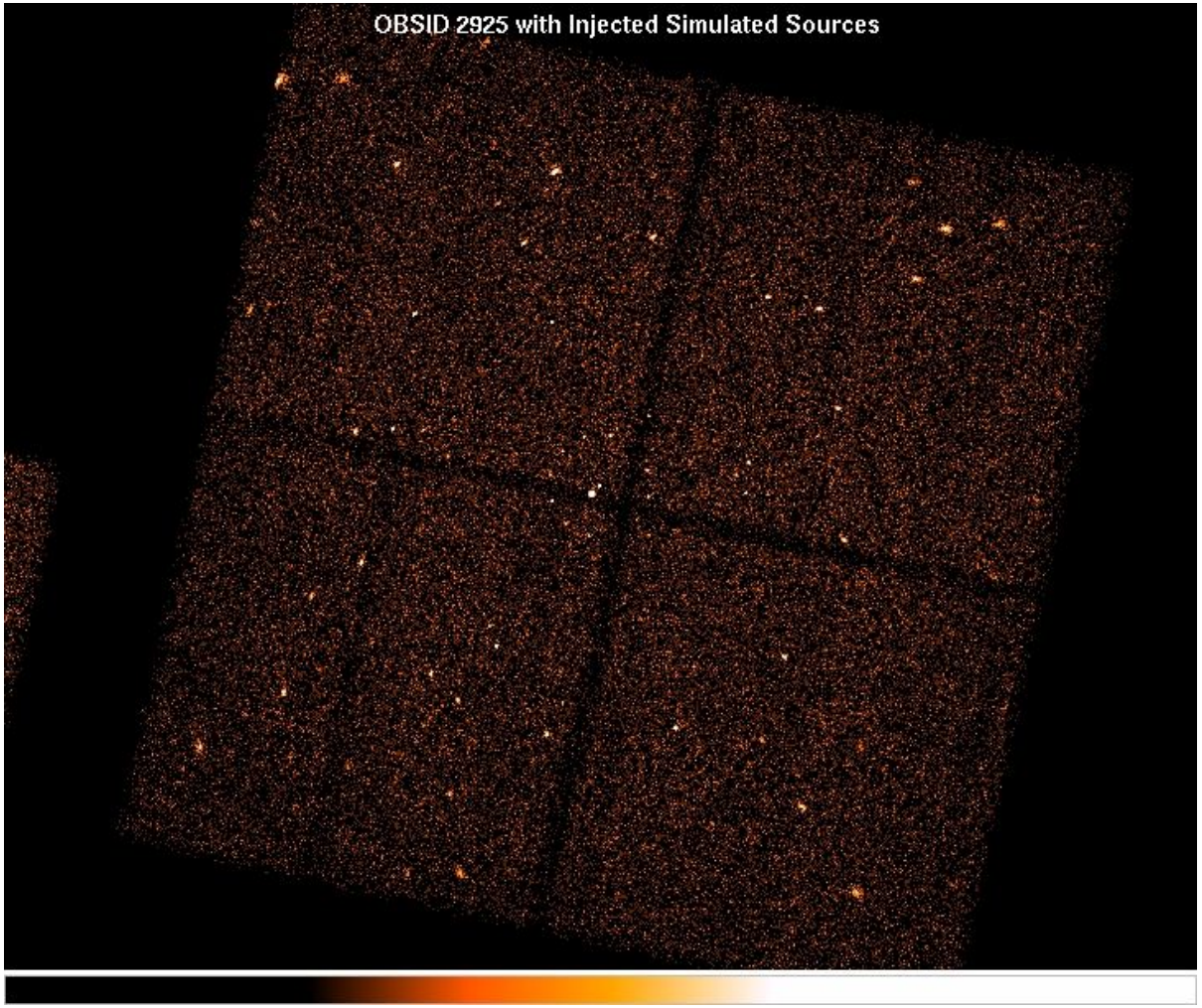
Developed plan for data product linkages and versioning;

requirements for database views

Comparison of the shape of the posterior distribution for hardness ratios $\mathcal{R}$, C, and $\mathcal{HR}$ with the classical Gaussian approximation for high counts (case I; *left column*) and for low counts (case II; *right column*). The solid lines represent the posterior distributions of the hardness ratios, and the dashed lines, a Gaussian distribution with the mean and standard deviation equal to the classical estimates. A prior distribution that is flat on the real line ($\phi = 1$; see § 5.1) is adopted for the Bayesian calculations. Note that as the number of counts increases, the two distributions approach each other. At low counts the two distributions differ radically, with the classical distributions exhibiting generally undesirable properties (broad and extending into unphysical regimes).

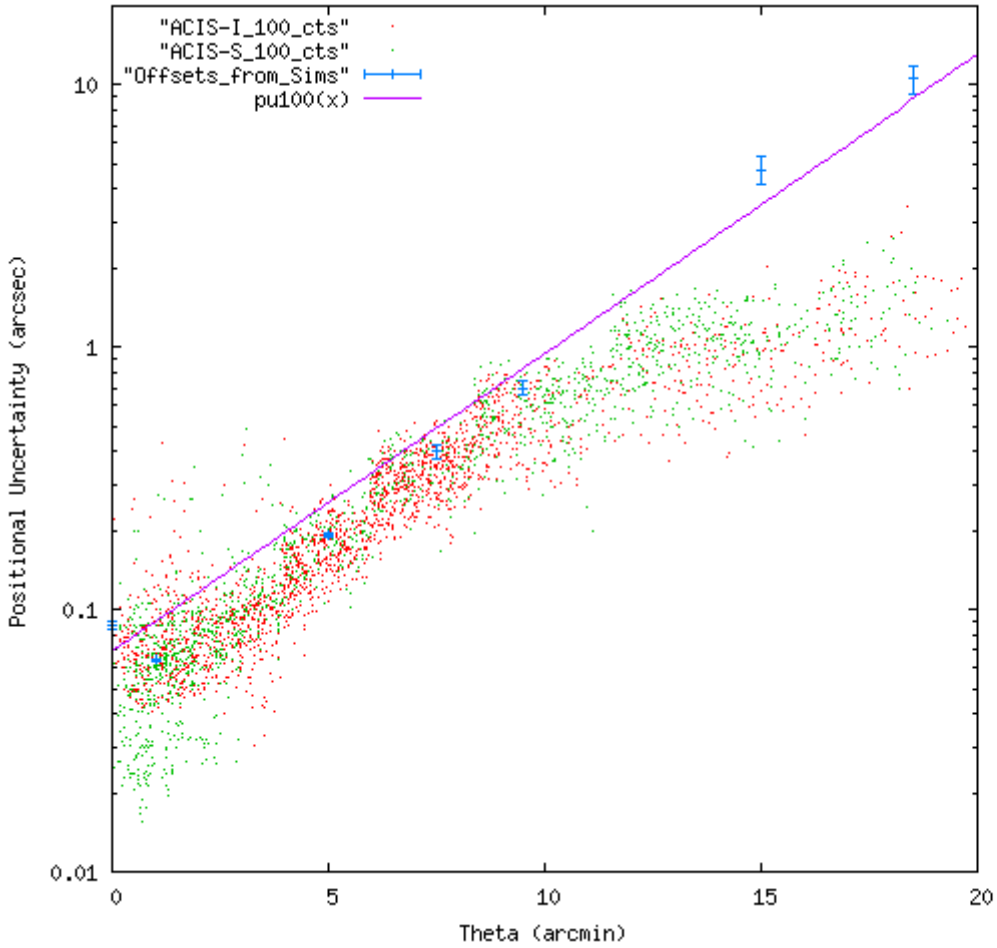


Inject simulated sources into data



Simulations calibrate wavedetect pos'n errors

CXC



Chandra Source Catalog Schedule

Schedule

- Developed detailed task list and schedule leading to first catalog release
  - SDS and DS scientists, software engineers
  - Detailed breakdown of science tasks
  - Mapping of pipelines and science issues to software tasks
  - Checking interdependencies
  - All (almost all) near and medium term tasks assigned to individuals
  - Tasked individuals gave input and feedback on completeness and time/work estimates
- Detailed schedule being maintained and will be made available publicly
  - Hardcopy available for inspection
  - Still figuring out how to use the schedule tool!
- Schedule updates on 2 week timescale; report to staff meeting
  - Ian Evans (DS) maintains schedule
  - Report to NASA-Marshall each month in telecon
- New resources added from SDS (Mike Nowak, John Houck, John Davis, Aneta Siemiginowska) working specific tasks

Chandra Source Catalog Schedule

Schedule Highlights

- Production processing for mission-to-date begins early-March, 2008
- First public access to catalog database and products late-April, 2008
  - ~1/3 of public imaging observations processed and available
  - Preliminary statistical characterization of catalog properties available
- Complete catalog formal release-1: mid-August, 2008
  - Includes all public mission-to-date imaging observations
  - Complete statistical characterization of catalog properties available

Chandra Source Catalog Progress

Pipeline releases

- CAT 2.4 build (complete)
  - Beowulf, SAOTrace, archive interface test; run with 200 OBSIDs
  - Sizing and performance tests
- CAT 2.5 (July)
  - Improved background tool; Source merging between OBSIDs
  - Prototype databases
- CAT 2.6 (Sep)
  - Source merging improved, some quality assurance checks, UI prototype
- CAT 2.7 (Nov)
  - Production pipelines, archive ready to support catalog release
- CAT 3.0 (Feb 2008)
  - Revise pipelines based on lessons learned, improved UI
  - Support for ongoing automatic processing operations

Chandra Source Catalog Schedule Components

Schedule Components

- Schedule broken into several key areas
  - Science processing
    - Algorithm analysis/evaluation, specification, and software implementation
    - Split into observation, source, and merge processing pipelines
  - Science characterization
    - Statistical analysis/evaluation of extracted source and catalog properties
  - Quality assurance
    - Techniques to verify integrity of extracted science properties and validate processing, specification and software implementation
  - Processing software infrastructure
    - Software design and implementation to integrate catalog and Beowulf-specific processing into existing processing infrastructure and operations software
  - Archive and database infrastructure, and catalog user interface
    - Software design and implementation of file archive and databases, support for product ingest and versioning, queries, operational interfaces, and release infrastructure
    - Design and implementation of catalog user interfaces to support science use cases
- All components share a common schedule, and complete at about the same time

THERE ARE NO LONG POLES

(or alternatively, all of the poles are equally long!)

Chandra Source Catalog Schedule Risks

Schedule Risks

- Science algorithm development takes longer than planned
  - What is the minimum for making the product scientifically useful?
  - Source existence; Positions, rates/fluxes, ERRORS.
  - Must have confidence that these are reasonably good and that the catalog is not dominated by contaminating false sources etc. This is the long pole in the science.
  - Other catalog properties are jettisonable for release 1, but are not that hard to add
  - Intermediate catalog software releases provide opportunities to test and fine-tune algorithms, software, and infrastructure prior to production
- Software implementation takes longer than planned
  - risk where science specs not yet finalized
  - using conservative estimates and detailed flowdown of science tasks to s/w components
- Science algorithms or software are inadequate
  - early science and operational testing, 6 week pre-production test with 1/3 of mission data
- Operational problems
  - Beowulf performance is good; new archive server hardware needed to keep up on ingest
  - Pre-production test allows operations shakeout, staff training
- Resource availablity
  - Strong priority from management; but mission events may override; SDS and DS tight on resources, key personnel availability always a risk

Data Processing Status

Pipelines and Internal Software

Releases

- Dec 15 DS7.6.9.1 for Cycle 9 support (RPS etc)
- Jan 9 DS7.6.10, CTI improvements in pipeline, minor telemetry changes
- Jan 23 DS7.6.10.1 PIMMS bug fix
- Feb 15 DS7.6.10.2 Improvements to DDT/TOO proposal support (grids) and update to CHART
- Mar 26 DS7.6.10.3 Upgrade software on archive server

Processing

- Repro 3 will complete this summer
  - Only year 2000 data is left
  - Will not reprocess 1999 data at this time (Repro 3 only for -120C data)
- Standard data processing continues to achieve 30-32 hour turnround from end of observation to user delivery

CIAO

CIAO releases

CIAO3.4

- Released Dec 19
- CTI upgrades including serial CTI for S3
  - Bugfixes (including wavdetect improvements)
- hrc_build_badpix, hrc_dtfstats
- reproject_aspect to fix up astrometry
- mkarf dead area correction (off by default)

CIAO 3.4.1, 3.4.1.1

- Mac PPC patch for problem with Prism
- Mac Intel support under Rosetta

CIAO plans

CIAO4

- CIAO3.4 uses old compilers, OTS – we are pretty stuck taking it to the latest operating systems: needed major scrub
- CIAO4.0 move to new compilers and OTS, support for source build, full support for Intel Mac; phased release; modular packaging
- GPL compliant
- May 15 – CIAO 4 Beta-1 (Tools)
  - Tools, AHELP, Prism-with-no-plotting: No ChIPS/Sherpa
- Aug – CIAO 4 Beta-2 (ChIPS and DM Ascii Kernel)
  - GPL compliant plotting
  - Python support
  - Manipulating text tables using DM syntax
- Oct – CIAO 4 Beta-3 (Sherpa)
  - New Sherpa
- Dec – CIAO 4.0 Full release
  - Also available as separate Sherpa, Tools, ChIPS packages
  - Intent is that Sherpa may be useful for non X-ray projects

Why the rewrite?

- License issues – SM was not free
- New approach is GPL compliant
- New package has improved capabilities to manipulate individual pieces of a plot
- Used in the data processing pipeline, especially for validation and verification. Supporting new automated V&V was a major driver for the rewrite.
- Needed to revise architecture to fit in with overall infrastructure improvements

What level of resources?

- New ChIPS is basically done except for interface improvements
- SDS focus on user interface and testing

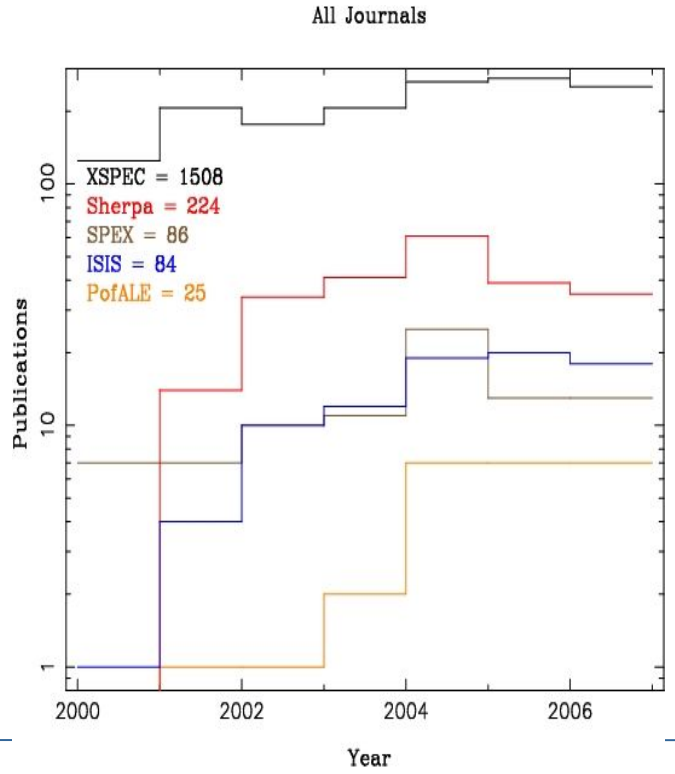
Sherpa user interfaces

- In beta release, main UI is a python-based scripting interface
- Side effect of using python as the application glue for the C++ libraries; we also retain S-lang interface
- Allows capability for sophisticated scripting, needed for advanced analysis; easier for SDS to provide new capabilities
- Not the right interface for typical astronomers: will provide higher level interface in final release
  - Survey shows community want a simple interface: we will support this
  - Object-oriented methods available for advanced users

```
for y in meth0:
    set_method(y)
    print get_method()
    for x in stat0:
        set_stat(x)
        try:
            m1=get_method().name
            s1=get_stat().name
            print m1
            print s1
            print >>file0, 'METHOD = %s' % m1
            print >>file0, 'STATISTIC = %s' % s1
            abs1.nh=0.1
            mek1.kt=0.1
            mek1.norm=0.1
            mek2.kt=2.
            mek2.norm=0.5
            fit
            projection
            covariance
            r1=get_fit_results
            pr1=get_proj_results
            print >>file0, "statval = %.5f" % r1.statval
            print >>file0, ["%.5f" % x for x in r1.parvals]
            print >>file0, ["%.5f" % x for x in pr1.parmins]
            print >>file0, ["%.5f" % x for x in pr1.parmaxes]
file0.close()
```

Sherpa is in use for real science

- SDS performed full text literature search of refereed journal articles
- Looked for XSPEC, Sherpa, ISIS, SPEX, PINTofALE
- Manually filtered out irrelevant entries (eg other packages called ISIS)
- SPEX usually used purely for atomic data, not fitting
- Results show Sherpa in 2nd place well behind XSPEC



Sherpa Progress

- In science testing phase, closer testing and evaluation loop between programmers and scientists=> immediate feedback
- Typically bi-weekly updates of new features and bug fixes
- Two demos at ADASS and AAS
  - ADASS - interest from Python community
  - AAS - interest from astronomers

Sherpa Status

- Since the last CUC meeting work on:
  - Core Functionality:
    - Convergence, Parameter Estimation
  - User Interface:
    - Data access functions for both Python and Slang users
  - 2D support:
    - PSF functionality, WCS, Region filtering, Imager
- Core functionality finalized by mid-June
- Evaluation and testing over the summer
- Beta ready for release in the Fall 2007 as CIAO Beta3 package - building from source on all release platforms
- Sherpa independent of CIAO for Python and S-lang users

SDS Development Work

SDS Priorities for ACIS

- SDS held internal discussions, and then approached ACIS team (hardware, calibration) with priorities
- Identified Three Areas for Further Assessment:

  Graded Mode for Frontside Chips

  - 18% of all HETG

  CC-mode

  - 12% of all HETG, 4% of ACIS-S Imaging

  0^{th} Order Response

  - Increasingly important with deeper looks & extended fields (e.g., AGN, galaxies)

Graded Mode

- Well calibrated for backside chip (S3), e.g., Cas A
- Frontside S-array (i.e., HETG)

  Lack of CTI correction leads to order sorting issues

  Uses old “FEF” (mkrmf) paradigm for response, as opposed to newer “p2resp” (mkacisrmf) files

  Forces us to maintain separate code base

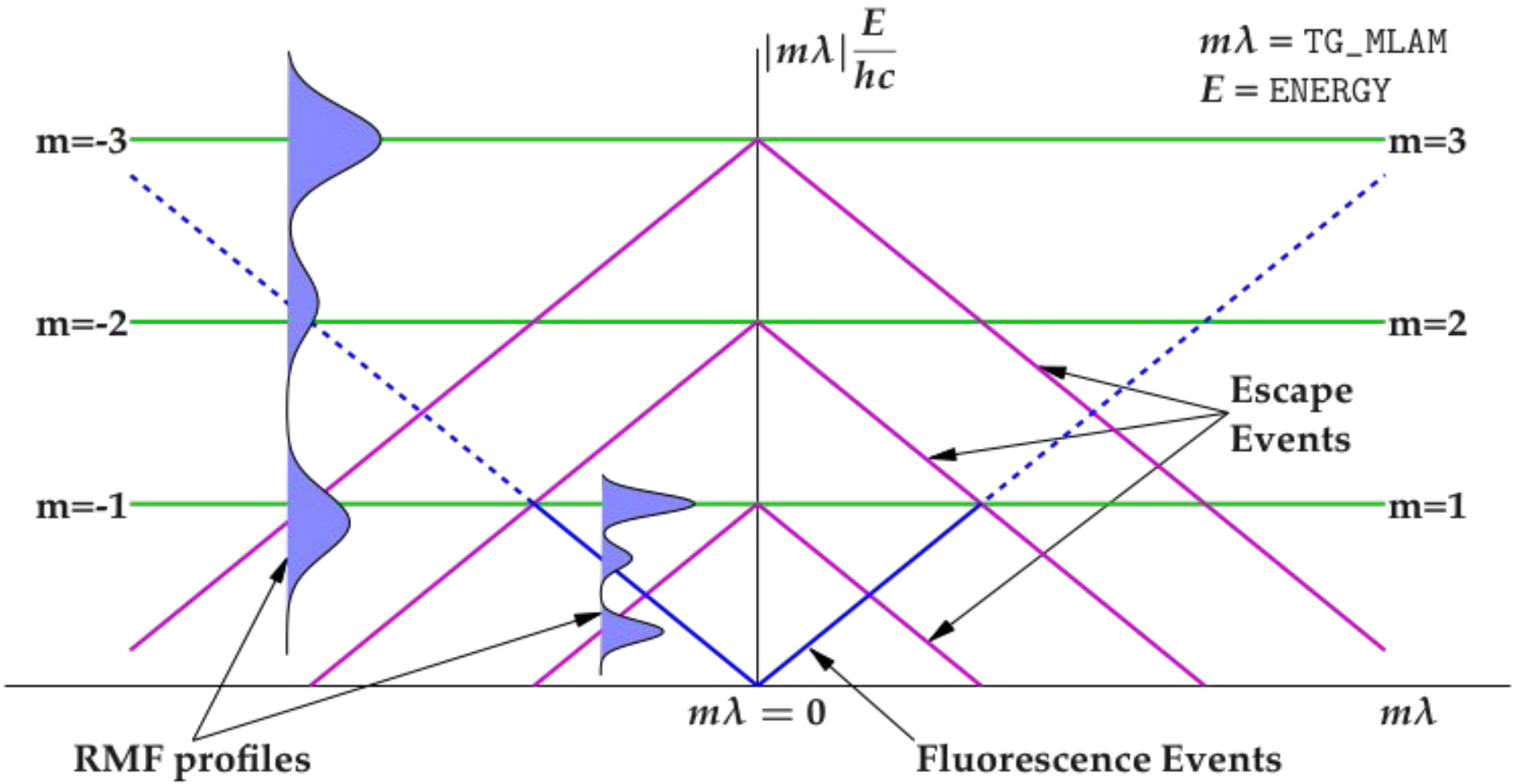
  Updates to CTI-corrected products (p2resp based responses) not carried over to graded mode

  Generation of new graded mode products from calibration data a long task

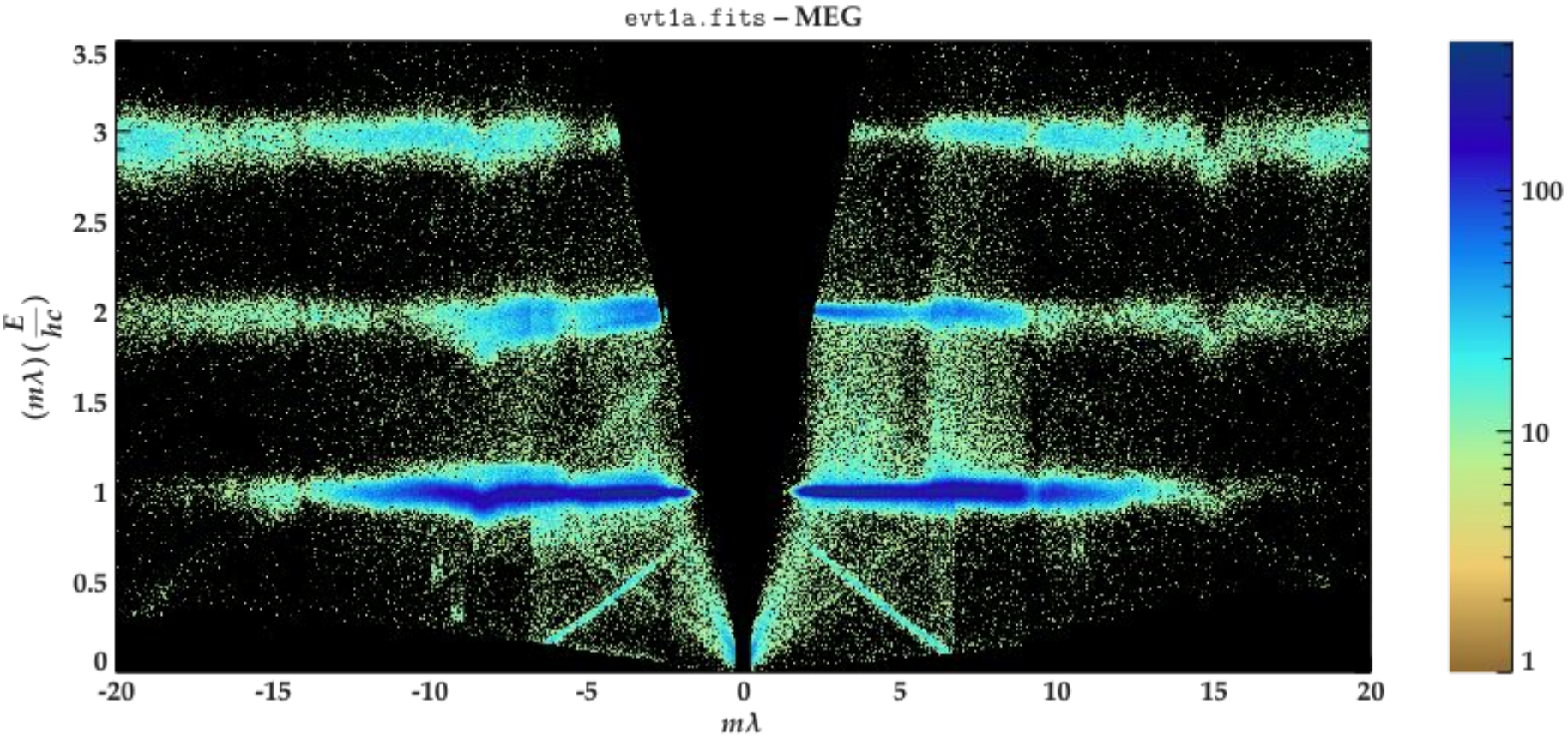
Graded Mode

- SDS concept (developed by John Davis) - Create an empirical “CTI Correction” for Graded Mode
- Use Standard “p2resp products” for “Corrected” Graded Mode Data
  - Eventually phase out FEF products (single code base)
  - Updates to p2resp automatically apply to graded mode
- Preliminary Results very Promising
- Further Enhancement and Refinement by Alexey Vikhlinin of the ACIS Team
  - Incorporates CCD trap maps/node dependence

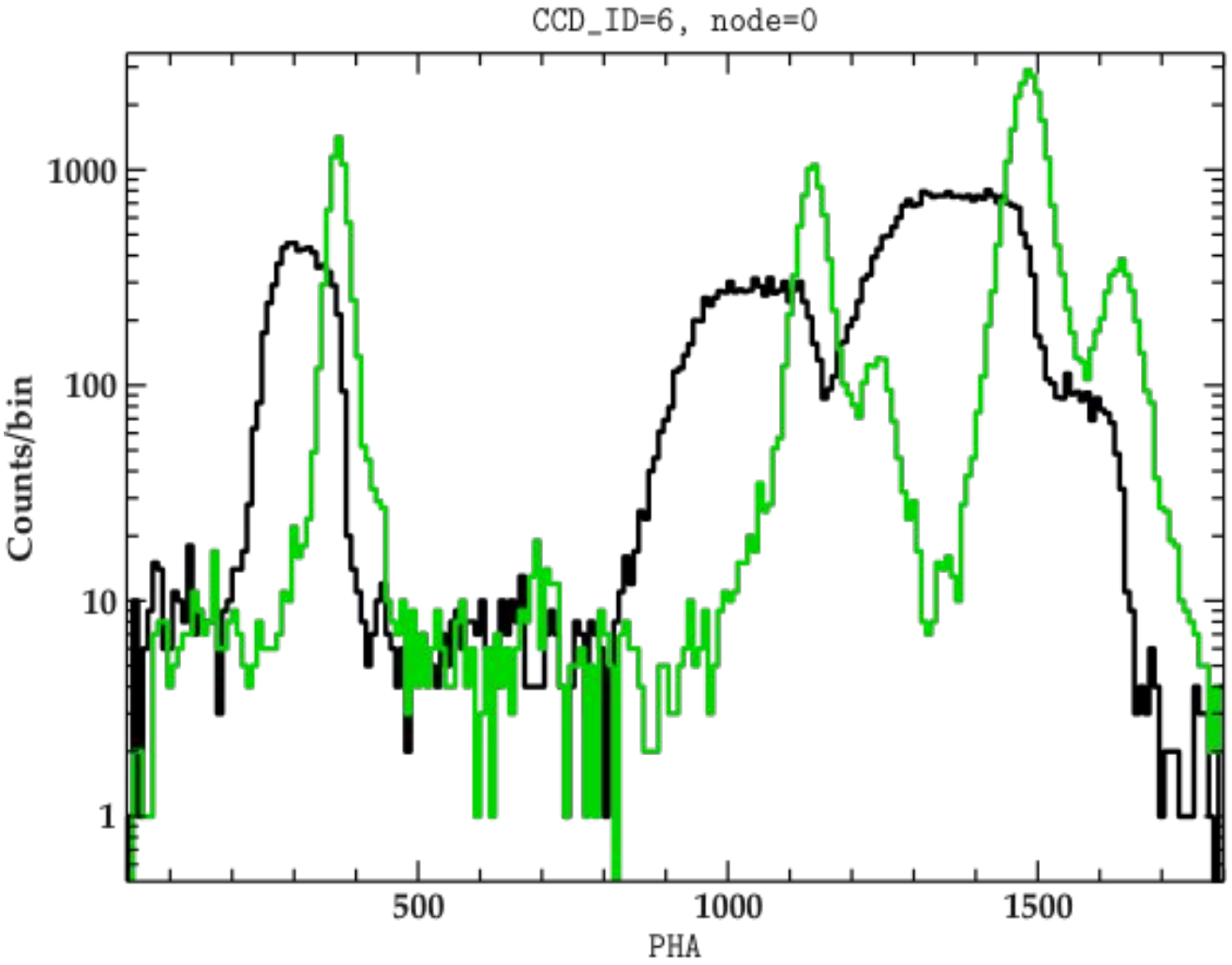
Order Sorting Basics:



FEF/No CTI Correction



CTI Corrected Cal Source:

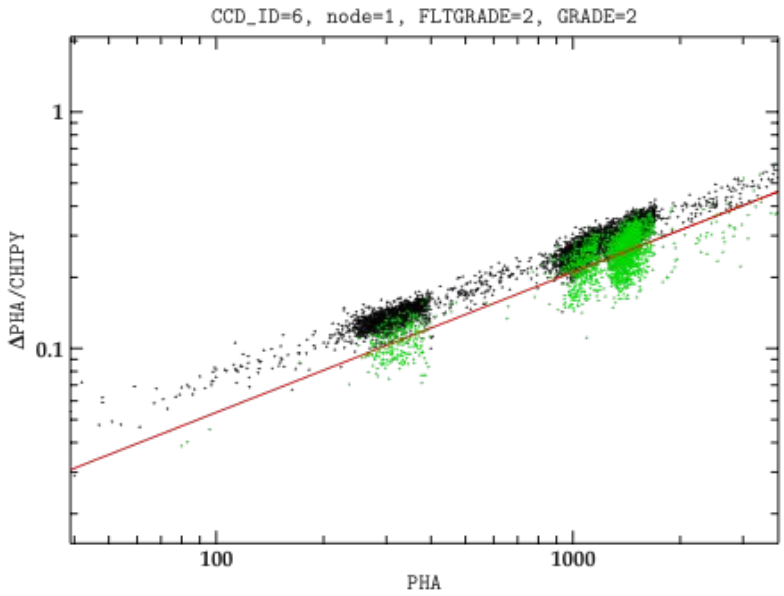


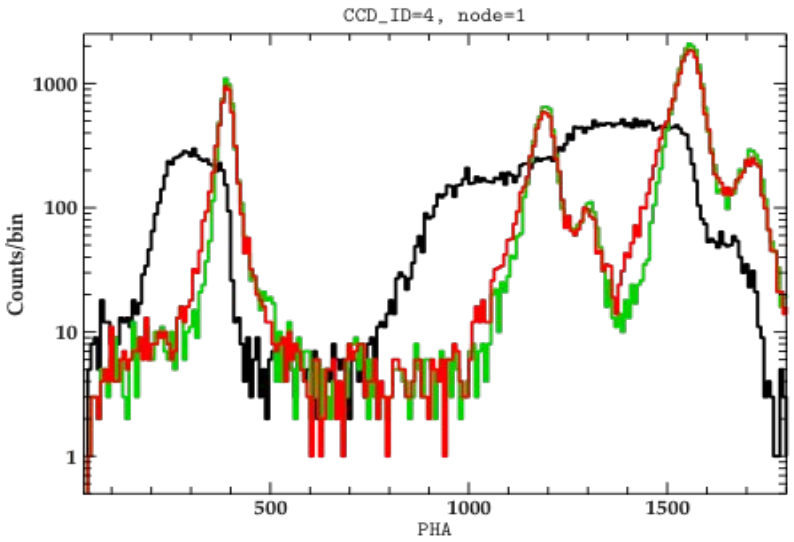
Adjust the pulse height of the event (for grade g, Chip y):

$$h_g \longrightarrow h_g + \Delta h_g(y),$$

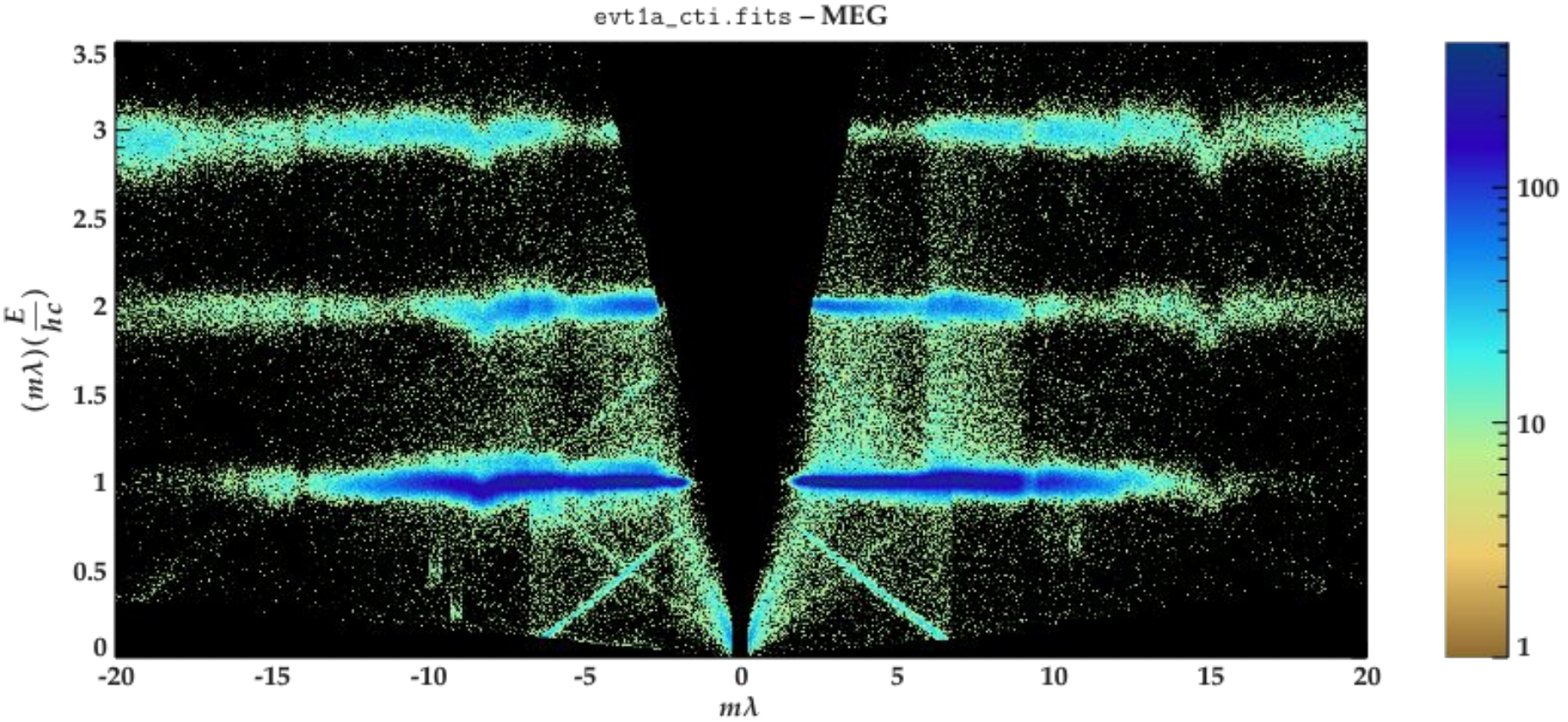
$$\Delta h_g(y) = \alpha_g h_g^{\beta_g} y$$

with time dependent coefficients



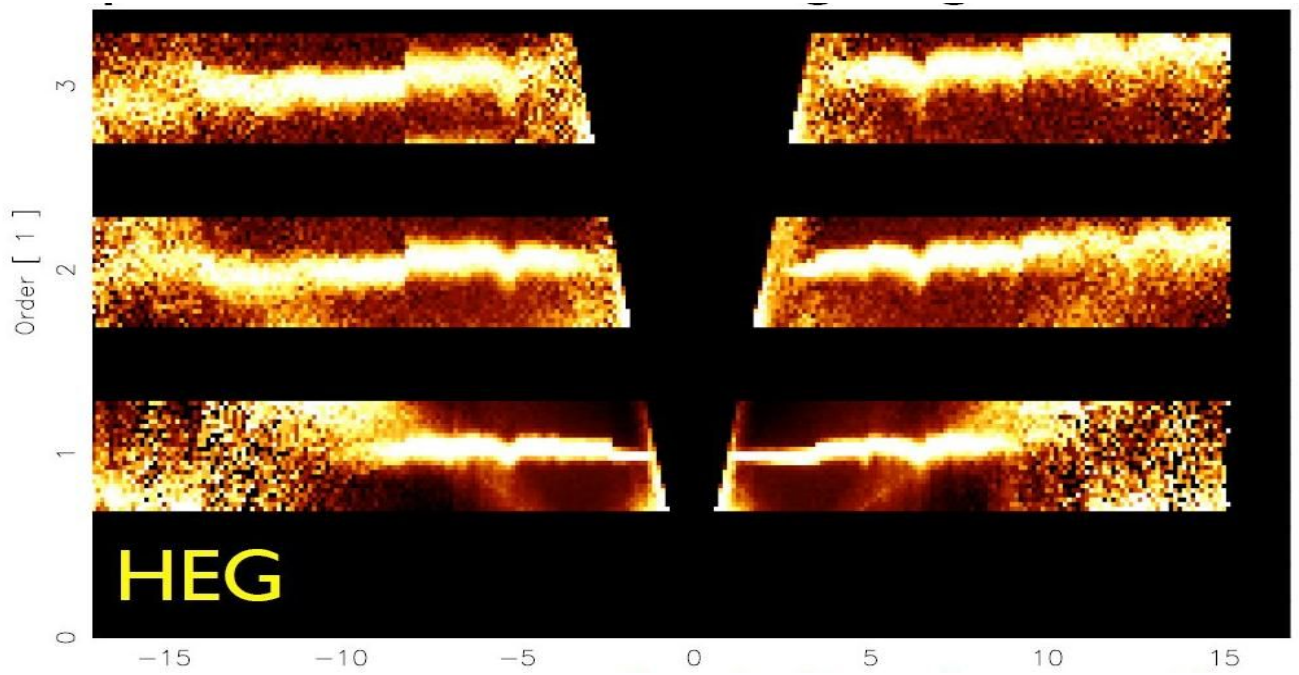


p2resp/CTI Correction



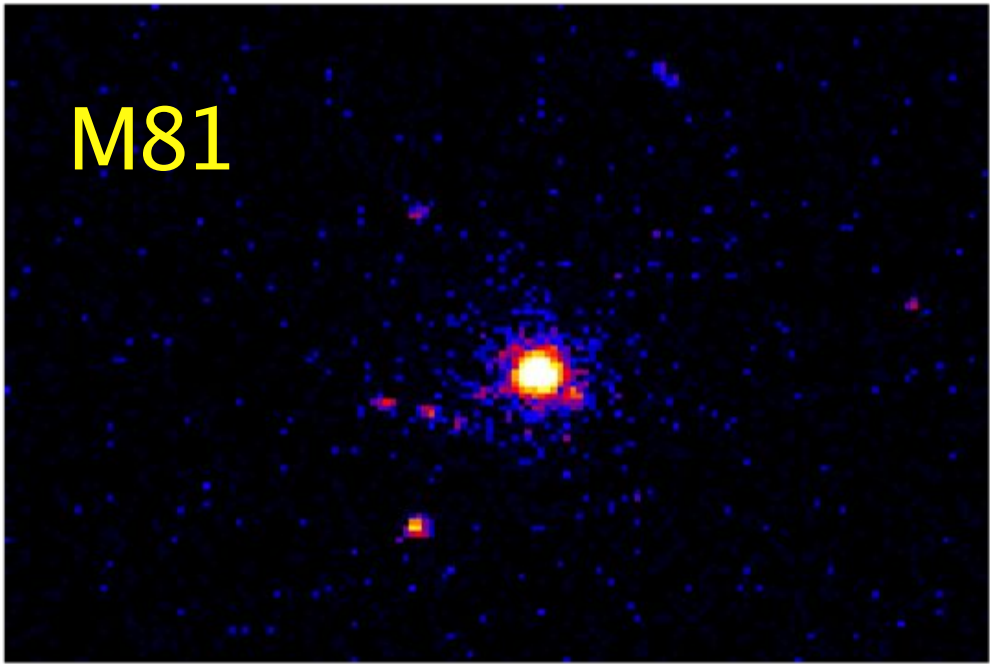
CC-Mode

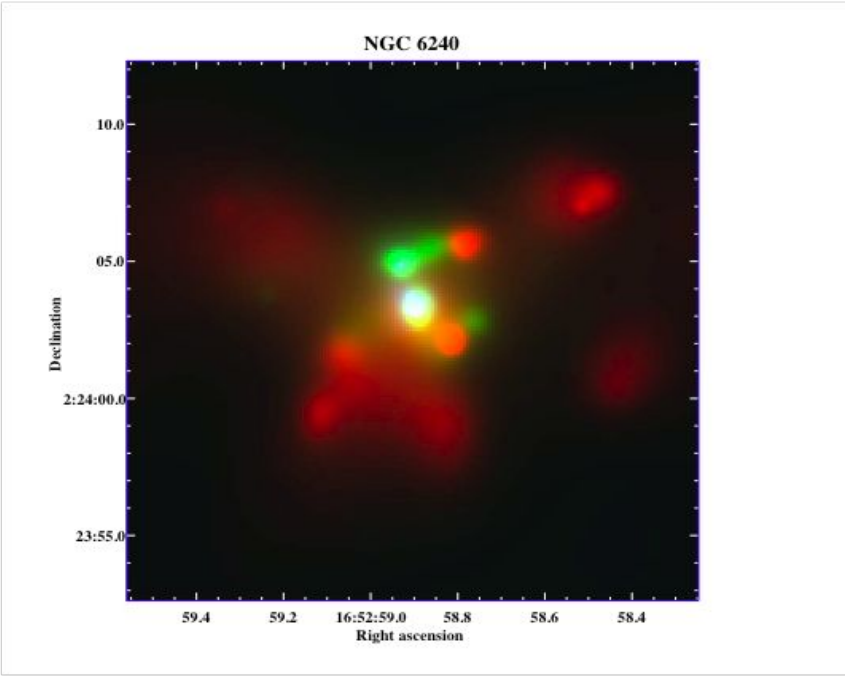
- SDS in the Process of Examining CC Mode Data
- For Bright Sources, We've Seen Some Sources with odd Order Sorting Plots, +/- Order Disagreements
- No Explanation Yet - An Ongoing Research Project



0th Order Response

- Longer Looks Becoming More Prevalent
- Increasing Catalog of Observations with 0th Order Structure (e.g., M81*, NGC 6240, Orion, ...)





0^{th} Order Response

- First Goal is to Assess the Quality of the Current Calibration Products
- Searching for Good HETG Data Sets Where:
  - 0^{th} Order isn't Piled Up
  - Gratings Have Sufficient Statistics
- One Promising Data set is the GTO Data of SN 1987a
- Plan to Have Preliminary Assessment in the Next Few Months
  - Current Products May Be Sufficient

New Tool: tgextract2

- Primary motivation: allow large, asymmetric regions for better determination of background
- Secondary motivation: Improve the tool efficiency & simplify the interface
- Available for Analysis (but not put in pipeline)

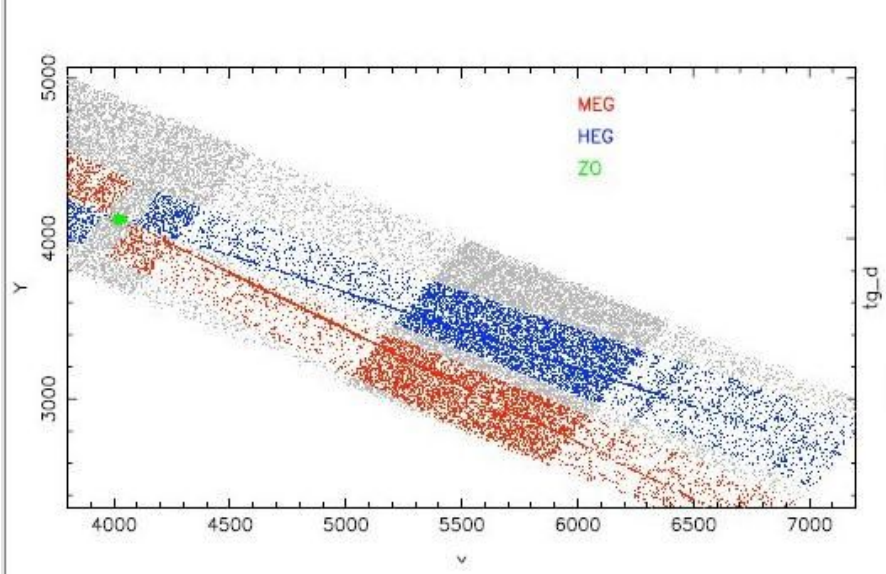
  Region-driven, with non-rectangular backgrounds

  - Parameter-specified regions still allowed

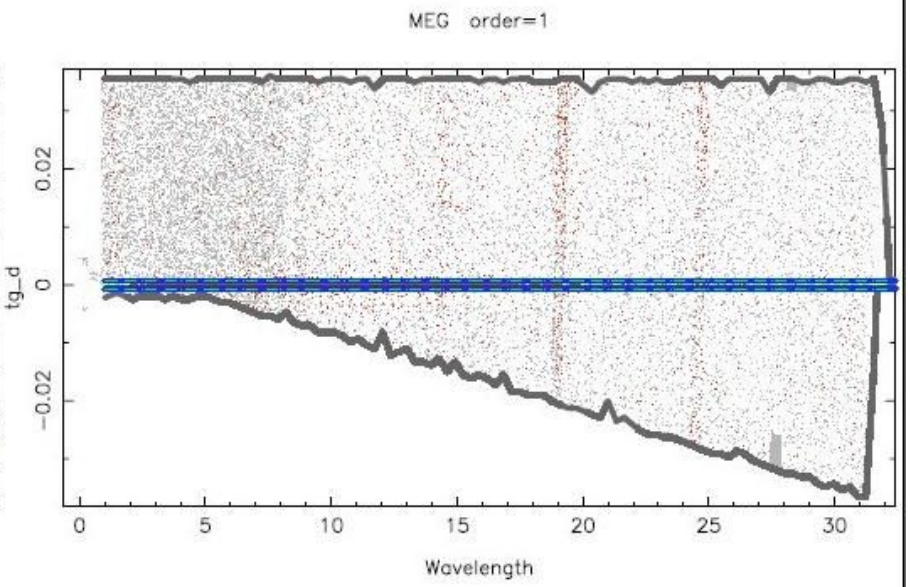
  Backscale computed as an array

  - Accounts for detector edges. Backscale computed from region or events.
  - _up, _down background arrays summed

tgextract2 Example



Selected Regions



Background Region in Cross Dispersion Profile

Update: acis_build_badpix

- Progress made in giving users finer control of bad pixel files (CUC Oct 2006)

- Specified enhancements for acis_build_badpix
  - bitflag parameter – allows customization by class of bad pixel (e.g. hot pixel, afterglows)
    - control over whether or not to flag adjacent pixels
  - usrfile parameter – finer control, can exclude or add specific regions or pixels to the output file.
- In testing for release in a later CIAO (to be determined)
- Further enhancements considered for later specification:
  - user specified count rate threshold for 'badness'
  - search for bad columns in CC mode files